An assessment of Radar, Gauge and Kriged Gauge Rainfall Data in Free State, South Africa

Tadiwanashe Simpson Gutsa
(217081368)

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

Supervisor: Dr. M Kumarasamy
Co-Supervisor: Prof. GGS Pegram

December 2019

"You cannot connect the dots looking forward; you can only connect them looking backwards. So you have to trust that the dots will somehow connect in your future"
Steve Jobs.

Declaration

I, Tadiwanashe Simpson Gutsa, declare that;

All work contained herein is original and my own intellectual property and has in no way been submitted partially, or in whole to any other university. Where the work of others has been included to any extent within this dissertation, all necessary acknowledgements, references and credit have been given. All research relevant to this dissertation was carried out at the University of KwaZulu-Natal, Howard College Campus within the Centre for Research in Environmental, Coastal and Hydrological Engineering at the School of Civil Engineering. This dissertation was supervised by Dr. M Kumarasamy and Prof. GGS Pegram.

Signed

....
Tadiwanashe Simpson Gutsa   Date

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

....
Dr. M Kumarasamy   Date

....
Prof. GGS Pegram   Date

<div align="right">

Tadiwanashe Simpson Gutsa
(217081368)
December 2019

</div>

Acknowledgements

Abstract

In the last decades, flooding has caused significant damages and death in South Africa. Flooding is associated with heavy rainfall events which vary largely in space and time. The forecast of such phenomena requires quality rainfall data to generate output with a reasonably high degree of confidence. This study aims to document the difference of rainfall estimates derived from radar, rain gauge and a kriged rain gauge network. A review on the use of weather radar in hydrological studies shows that there is an opportunity of using radar estimates in near-real time flood forecasting and warning applications. The quality of radar rainfall estimates is assessed using pluviometer rain gauges and a daily kriged rain gauge surface in the Vaal sub-catchment in the Free State, South Africa. The study uses data from the MRL5 S-band radar located at Bethlehem and from rain gauges within weather stations operated by South African Weather Services (SAWS). The analysis assesses the effect of rainfall seasonality, radar range dependencies and storm variation on the quality of estimated rainfall accumulations. In addition, gauge density is analysed to determine the effect it has on the performance of kriging estimation. During the research notable spatial rainfall variation and areas with quality radar estimates have been identified. The results show that there exists a seasonal bias between radar and rain gauge estimates with the radar underestimating low intensity gauge rainfall pronouncedly during the winter rainfall events by an average of 31%. The underestimation of rainfall by the MRL5 radar increases as we move away from the radar tower. During the summer rainfall events the radar estimates are almost similar to point gauge estimates especially during late summer (February and March) . Radar underestimation of winter rainfall is probably due to overshooting of the tops of stratiform rainfall by the radar beam. Correlations are high between MRL5 radar and rain gauges during summer rainfall events ranging between 0.7 and 1. Results from this study provide information to guide on the application and selection of rainfall estimation techniques.

Table of contents

List of figures

List of tables

Nomenclature

Greek Symbols

γ complex function

λ Kriging weight

μ numerical mean

ε linear function

Subscripts

i, j, 0 subscript index

Other Symbols

$\forall$ for all

$\in$ in

$\supset$ super set

$<$ less than

$>$ greater than

Acronyms / Abbreviations

ACRU   Agricultural Catchments Research Unit

AP   Anomalous Propagation

BLUP   Best Linear Unbiased Predictor

CAPPIs   Constant Altitude Plan Position Indicators

CM   Conditional Merging

CSAG   Climate Systems Analysis Group

CSIR   Council for Scientific and Industrial Research

DISPLACE   Digital Signal Processing for Logarithmic, Linear and Quadratic Responses

DSD   Drop Size Distribution

DWS   Department of Water and Sanitation

EDA   Exploratory Data Analysis

GHCN Global Historical Climatology Network

IDW   Inverse Distance Weighting

KED   Kriging with external drift

MAE   Mean Absolute error

MDV   Meteorological Data Volume

METSYS Meteorological Systems and Technology Section

NDMC National Disaster Management Centre

NETCDF Network Common Data Format

NPRP National Precipitation Research Programme

NWRN National Weather Radar Network

OK   Ordinary Kriging

QPEs   Quantitative Precipitaion Estimates

RDAS   Radar Data Acquisition System

RMSE   Root Mean Square Error

SAFFG South AfFlash Flood Guidance System

SAWS   South African Weather Services

SK   Simple Kriging

SKvm   Simple Kriging with local varying mean

TITAN Thunder Indentification, Tracking, Analysis and Nowcasting

TP   Thiessen Polygons

UK   Universal Kriging

WMO   World Meteorological Organization

WRC   Water Research Commission

Z-R   Reflectivity to Rain-rate algorithm

Chapter 1

Introduction

1.1 Background

In the history of Southern Africa disasters as a result of floods have exacted a heavy toll, producing significant damages in property and casualties (Davis and Vincent, 2017). Worldwide, records show that over the last 30 years, more than 200 000 people have lost their lives due to flooding and 2.8 billion others affected (Jakubicka et al., 2010). Floods are the most common natural disaster experienced in South Africa. The observed climate change is an issue worldwide and has led to an increase of extreme rainfall events in South Africa (Ziervogel et al., 2014). In 2011, flooding left more than 120 people dead and 20 000 in need of assistance in some parts of South Africa (Eric, 2011). The South African insurance industry reported a cost of R1 billion in a calendar year (2014/2015) from claims related to high rainfall events. The costs of damages caused by floods nationwide have been covered by the South African Government and have shown to cause a significant strain on the economy (Davis and Vincent, 2017).

Flooding is the temporary overflow of water on normally dry land due to excess rainfall runoff (Dingman, 2015). While flooding is influenced by several factors which include meteorological, hydrological and human-instigated factors, rainfall tends to be the most influential factor that governs the magnitude and frequency of floods (Takeuchi, 2001).

The commonly used tools for predicting floods are hydrological models. Hydrological models conceptualize the physical characteristics of a catchment and are used to examine the response of flow rates and levels of water bodies in relation to rainfall events. Application of hydrological models in flood forecasting requires detailed information which describes the rainfall characteristics over a region. The quality of a hydrological model relies heavily on the accuracy and reliability of the rainfall data to serve as a valuable flood mitigation tool (Zhu et al., 2013). Output from these tools is important for design, decision and policy making in water resource management. Furthermore, accurate rainfall data are a requirement for calibrating models to obtain parameters which match the catchment characteristics. Cole and Moore (2008), Beven (2011), Berne and Krajewski (2013) and Xu et al. (2013) show that accurate rainfall fields are required to reproduce hydrological scenarios. The quality of the input rainfall data is directly linked to the consistency of the model and it influences the decision-making process carried out using the model outputs (McMillan et al., 2011). In particular, for small catchments, precise spatially continuous rainfall data is critical for forecasting and/or simulating streamflow during flooding. This has driven efforts to improve the scale and quality of rainfall measurements using different techniques and methodological approaches (Gabriele et al., 2017).

The South African Weather Services (SAWS) and Council for Scientific and Industrial Research (CSIR) are the main institutions in South Africa which are responsible of monitoring

rainfall events. They mainly focus on rain gauge and radar rainfall products, which are the most widely used rainfall sensors for measuring rainfall accumulations (Berne and Krajewski, 2013). Rain gauges have been the established rainfall measuring tool since the birth of hydrology (Clothier, 2011). Rain gauges record the actual amount of rain falling at a point, but their spatial representation of rainfall variability is limited especially during convective events. This brings about limitations in forecasting where data at both larger spatial and temporal resolution are required. The distribution of point rainfall measurements can be achieved using Geostatistical interpolation methods. These methods combine earth sciences and statistical tools to spatially distribute rain gauge measurements. The interpolated rainfall fields can provide information on the spatial structure of rainfall that is useful for modelling purposes (Gabriele et al., 2017). Nevertheless, considering interpolated rainfall fields are calculated from rain gauge data and the resulting estimates will always be a smoothed version of the gauge data plus sampling errors.

High-resolution precipitation data derived from the weather radar might be favourable in forecasting applications although it still lacks quantitative precision due to several sources of error (Berne and Krajewski, 2013). The weather radar is described as a remote sensing tool that measures the reflectivity of precipitation at a given height above the ground (Goudenhoofdt and Delobbe, 2009). The applicability of weather radar in quantitative estimation of rainfall has been intensively investigated (Collier, 1989; Clothier and Pegram, 2002; Goudenhoofdt and Delobbe, 2009 and Xu et al.,2015). Ciach et al. (2007) report that weather radar overestimates rainfall accumulation in the summer months and significantly underestimates it at longer ranges from the radar tower (>100 km) during winter in Oklahoma, United States which is also the case in the Free State, South Africa. In New Mexico, Xie et al. (2006) show that there is a substantial difference (64%) between rainfall from NEXRAD radar pixels and interpolated rain gauge grid cells. Recent studies have examined the application of radar estimates in streamflow predictions (Einfalt et al., 2004; Hamidi et al., 2017 and Thorndahl et al., 2017). Kneis et al. (2017) report that using radar estimates increases the lead time for forecasting flash floods in small basins (400 km^2), in the Phillippines. Hydrological clients require better information to make decisions on the best rainfall estimation method to use and have some knowledge of each method's limitations. For example, SAWS and the National Disaster Management Centre (NDMC) collaborated in 2009 to start a project for implementing a flash flood warning system known as South African Flash Flood Guidance system (SAFFG) (De Coning and Poolman, 2011). The SAFFG utilizes radar rainfall estimates for streamflow forecasts, therefore, there is a need to evaluate rainfall estimates derived from radar and to consider approaches to improve them were necessary.

Furthermore, the application of weather radar, gauge and interpolated rainfall products is influenced by location-based operational and environmental factors, which affect the accuracy and performance of each method in an operational context. Radar rainfall estimation is affected by storm characteristics, topography and proximity of the radar tower, while catchment characteristics and gauge density are the main influencing factors in rain gauge networks. Studies by Kitchen and Blackall (1992), Michelson and Koistinen (2000), Moral (2010) and Berne and Krajewski (2013) show that these influencing factors significantly affect the accuracy of rainfall

estimation techniques by reducing the consistency of the rainfall estimates. How and to what extent each factor influences the selection of a suitable rainfall estimation method needs to be taken into consideration. This prompts the need to investigate the performance of each rainfall estimation technique in estimating rainfall accumulations for varying cases and to quantify the effect of each factor on the resulting rainfall product.

1.2   Aim

Until fairly recently, rainfall estimation in hydrology has been done using point rain gauge stations. However, the high spatial variability exhibited by rainfall events limits the use of point rainfall estimates because a typical rain gauge just covers an area of 0.02 m^2. For the past decade, most hydrological institutions have adopted methods which estimate rainfall at higher temporal and spatial resolutions. Unfortunately, few studies reveal the quality contained in rainfall fields produced by these methods. The uncertainties and complexity associated with the rainfall estimation process of each individual method have limited their widespread application. The aim of this study is to assess the performance of radar in quantifying daily accumulations using pluvial rain gauges and a network of Ordinary Kriged rain gauges. The study will use rain gauge and radar data provided by the SAWS and Computing Centre for Water Research (CCWR) from a part of the upper Vaal sub-catchment near Bethlehem in the Free State Province, South Africa, as a case study. The emphasis of this thesis is to research, understand, describe and use methods of estimating spatial rainfall from the combination of gauge and radar data. The core of the thesis demonstrates the application of these tools using the valuable data collected by SAWS.

1.3   Objectives

The goal of the thesis is achieved by satisfying the following key objectives:

(i)   Conduct a comprehensive literature review to indicate the derivation and application of radar, rain gauge and geostatistical interpolated rainfall estimates, including the sources of error associated with each estimation method;

(ii) Examine the daily rainfall characteristics near Bethlehem, Free State province, a summer high-veld rainfall region;

(iii) Generate spatial daily rainfall fields from rain gauges and weather radar; and

(iv) Evaluate the impact of operational and environmental factors on the accuracy of radar, rain gauge and Kriged gauge rainfall products.

1.4   Structure of the Dissertation

This thesis was written following the classical monograph format. After the first chapter, a comprehensive review describing rainfall estimation methods including their application and limitations is presented. The third chapter describes the methods used for obtaining spatial rainfall estimates and comparison statistics. The last two chapters (Chapter 4 and Chapter 5) present the findings from the research and a conclusion to the findings, respectively.

The Chapters which make up the thesis are outlined below:

Chapter 2: provides a detailed literature review on rainfall estimation techniques.

Chapter 3: provides a description of the study area, rainfall data, Ordinary kriging algorithm and comparison statistics.

Chapter 4: presents the key findings of the study, focusing mainly on the effect of operational and environmental factors on spatial rainfall estimation.

Chapter 5: provides the summary and conclusion of the findings in the previous chapter, as well as recommendations for future research.

Appendices: provide additional information of the National Weather Radar Network (NWRN), python programming codes for the Ordinary kriging algorithm, kmeans clustering algorithm and the algorithm for extracting radar data.

Chapter 2

Literature review

2.1   Introduction

This chapter is a comprehensive review of spatial rainfall estimation using weather radar and geostatistical interpolation methods. The aim is to satisfy the following objectives:

1. provide a review and description of how rainfall estimates are derived from radars and rain gauges, including the limitations of each instrument;

2. describe geostatistical interpolation methods and their application in spatial estimation of rainfall; and

3. describe the operational and environmental factors which influence the performance of rainfall estimation techniques.

The application of radar and rain gauge rainfall estimates in hydrology has been extensively studied. This literature review will aid in identifying applications and limitations of radar, rain gauge and kriged rain gauge data in operational contexts.

2.2   Rainfall Estimation: Radar

2.2.1   Types of Weather Radar

Hydrological clients usually do not have an option to choose among different types of weather radars. However, they should be aware of the basic differences between the radars. Weather radars are distinguished according to emitted wavelengths of electromagnetic waves. Three bands are in operational use today for radar-hydrometrics namely C-band, S-band and X-band (Table 2.1). The size of wavelength used is mainly determined by the typical rain drop size, since there is a characteristic ratio between the size of the radar wavelength and water droplet (Berne and Krajewski, 2013). An ideal size ratio guarantees maximum detection of rainfall, but is seldom achieved in practice because of the range of size and shape of the hydrometers. Ground clutter and variations within the reflectivity profile are the main limitations when operating large wavelength radars while short wave length radars are hindered by attenuation, which increases as the wavelength gets shorter. Therefore, the choice of wavelength to use is a compromise. X-band radars tend to be the most affected by attenuation but are suitable for determining urban hydrology tasks and analysing high resolution vertical profiles (Einfalt et al., 2004). S-band radars are the least affected by beam attenuation and are mainly used when radar data is required over large areas (Meischner, 2005). However, S-band radars do not detect light

rain well compared to X-band and C-band radars (WMO, 2008). In South Africa, the National Weather Radar Network (NWRN) operates a combined coverage system of X-band, C-band and S-band radars. These radars operate at a usable range of 120-150 km and a maximum range which exceeds 200 km. At long ranges above 250 km the elevation of the radar beam above the ground is too high to capture rainfall estimates appropriate for small scale hydrology.

Table 2.1 Bands of weather radar

Band	Wavelength (cm)	Frequency (GHz)	Typical scanning interval rates (min)
S	5.70 - 19.30	2 - 4	5 - 10
C	4.84 - 7.70	4 - 8	5 - 10
X	2.75 - 5.77	8 - 12	0 - 2

2.2.2 Historical Perspective of Weather Radar and Developments in South Africa

The technology of radar developed from the theoretical predictions of J.C Maxwell but Heinrich Hertz is regarded as the "Father of radar technology" for his experimental work of demonstrating Maxwell's predictions (Hertz and Kelvin, 1893). Radar was firstly used as a security alert system to detect aircraft by military defenses during 2nd World War in 1940. The acronym radar was adopted in the 1940's, standing for radio detection and ranging (Van Heerden and Steyn, 1999). After the war, extensive research and major developments on non-military use of radar technology were achieved (Doviak et al., 2006). It is reported that rain showers were first observed by radar in Britain in 1941 and recognition of the establishment of radar meteorology is given to Ryde (1941).

Dr Basil Schonland of the Bernard Price Institute of Geophysical Research at the University of Witwatersrand pioneered the setting up of the first South African radar. It is reported that this radar received its first echo from the Northcliff Water Tower, Johannesburg in December 1939 (Van Heerden and Steyn, 1999). The first role of the weather radar in meteorological studies in South Africa was to monitor cloud seeding experiments (Hall, 1952). The radar marked its first use in hydrological studies in 1975, when the (CSIR) radar observations were used to predict extreme weather conditions in the Vaal River catchment (Alexander, 1990). During the 1980's, there was a notable use of the radar in hydrological studies which resulted in some published literature (Rosenfeld and Gagin, 1989; Rosenfeld and Mintz, 1998 and Rosenfeld et al., 1993). The establishment of the National Precipitation Research Programme (NPRP) in 1990 accelerated developments in radar related research. This was embraced with achievements such as the birth of the Digital Signal Processing for Logarithmic, Linear and Quadratic Responses (DISPLACE) method by Terblanche (1996). By the end of the 20th century, the communication of radar rainfall data to clients in the hydrological field was plausible (Van Heerden and Steyn, 1999). In 2001 the South African Weather Services (SAWS) was formed and since then it has been responsible for most of the Radar precipitation developments. The Meteorological

Systems and technology Section (METSYS) established in SAWS oversees the operation of the National Weather Radar Network (NWRN). METSYS, in collaboration with the Water Research Commission (WRC) have contributed to applications of weather radars in South African hydrology (Van Heerden and Steyn, 1999; Clothier and Pegram, 2002; Pegram, 2003; Sinclair and Pegram, 2005; Pegram et al., 2007; and Becker, 2016;).

2.2.3   Radar Quantitative Precipitation Estimation

2.2.3.1   Reflectivity Estimation

Estimation of precipitation using radar requires careful consideration of reflectivity measurements. Radar measures rainfall indirectly by sending brief pulses of electromagnetic waves at a high frequency into the atmosphere through an antenna, and then listens to the backscattered echo from rain droplets. Some of the electromagnetic energy propagated is absorbed by the raindrops whilst the echo of the unabsorbed energy is reflected to the antenna. The intensity of reflected echo depends on the size of the raindrops and increases with rainfall intensity (Van Heerden and Steyn, 1999). Time measured between the transmitted pulse and received echo is used to calculate the range of the rain droplets. The weather radar operates by "binning" the corresponding signal of reflectivity (Z) from the continuous echo received at the antenna. The unit of measure for the reflectivity signal is the decibel (*dBZ*). This signal is organized into a set of bins by sampling the signal at short discrete phases of time and then averaging it over the sampling interval (Clothier, 2011). Through analysing the strength of the received signal in the bins, it is possible for the radar to arrange the reflectivity across different bins, matching different ranges. Since the returned echo contains a wide range of reflectivity, setting a minimum and maximum threshold to filter noise and exclude reflectivity from wet hail cloud particles respectively is a prerequisite (Austin, 1987). If not accounted for these may lead to an overestimation of rainfall.
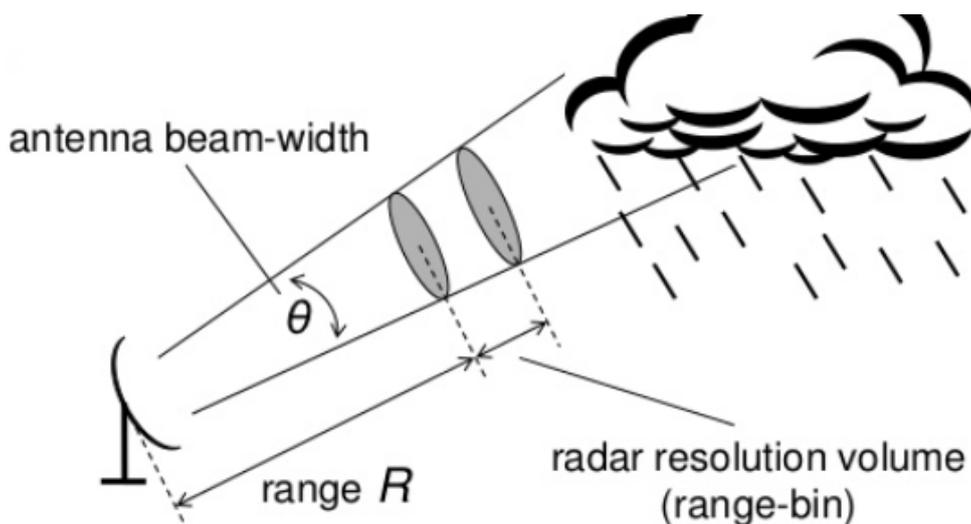
Figure. 2.1 Weather radar beam (Otto, 2012).

Weather radar transmits a beam in the form of a cone which is perpendicular to the antenna as shown in Figure 2.1. The beam width depends on the size of the antenna and is proportional to the wavelength of the transmitted wave. The focusing ability of the radar is also determined by the size of the antenna. To achieve a tightly focused beam with long wavelength radars, a large antenna is required. Reflectivity measurements are done at all azimuths (side to side) by rotating the antenna to produce a 2-dimensional reflectivity field. Furthermore a 3-dimensional field can be obtained by changing the antenna elevation (up and down) after each complete revolution. The configuration of the radar is set depending on its intended operational use. However, the base scan (all azimuths, with a fixed elevation) and volume scan modes (all azimuths and elevations) are the common radar configurations. The SAWS operate radars on volume scan mode which are then converted to 2-dimensional reflectivity fields called Constant Altitude Plan Position Indicators (CAPPIs). CAPPIs show reflectivity data on a Cartesian grid space extracted from the polar coordinate form of volume scans. The CAPPIs are extracted using the DISPLACE method proposed by Mittermaier and Terblanche (1997). This method uses an interpolation technique to generate point reflectivity data in a Cartesian space by means of interpolating radar bins at two different elevations (Figure 2.1).

Figure. 2.2 Diagram illustrating a point in a Cartesian grid space and the eight radar data points used to interpolate the point (Mittermaier and Terblanche, 1997).

The DISLACE method is efficient and produces images which are fine, that is, without discontinuities in the reflectivity field (Clothier, 2011). Radar reflectivity in CAPPI form facilitates the merging of radar fields from radars and increases the computation efficiency of rainfall rate (Terblanche et al., 2001). Also, the TITAN software used by the SAWS to visualize radar reflectivity requires the data in CAPPI format (Jager, 2017).

2.2.3.2   Conversion of reflectivity to rainfall rate

The next stage after extracting CAPPIs and obtaining the reflectivity measurements is to convert the reflectivity to rain rate. Several methods to convert reflectivity into rain rate have been proposed in the past which include the Z-R relationship, Area-time integral, Polarimetric methods and Statistical methods (Doviak et al., 2006). However, the rain rate in this study was derived using the Z-R relationship based on the work of Marshall and Palmer (1948). Their work showed that the rain rate (R) could be related to radar reflectivity (Z) by means of a single power law transformation known as the Z-R relationship. This relationship assumes an exponential drop size distribution (DSD) which best describes average drop size spectra. A body of research work done using the MRL5 S-band radar at Bethlehem confirms the applicability of the Z-R relationship in that region (Mittermaier and Terblanche, 1997; Deyzel et al., 2004 and Clothier, 2011). They are factors which should be considered before applying an appropriate Z-R relationship because no universal relationship is suitable for all rainfall events (Suzana and Wardah, 2011). This is mainly due to the variation in rainfall patterns and drop sizes. The main factors affecting the selection of a Z-R relationship are the geographical location and precipitation types (Chen et al., 2008). Battan (1973) gives an account of modifications to the Z-R relationship obtained using various storm types in different geographical locations. Uijlenhoet (2001) then demonstrated that the Z-R relationship gives sensible results which are comparable to all the 69 relationships suggested by Battan. The classical empirical Marshall and Palmer relationship is defined as:

$$Z = aR^b \tag{2.1}$$

where Z is the observed reflectivity in decibels, R is the rain-rate in (mm/h), a and b are constants traditionally chosen as 200 and 1.6 respectively. The resulting instantaneous rainfall is averaged into hourly radar rainfall fields and further to 24-hour radar rainfall fields.

2.2.3.3   Rain Gauge Adjustment

The need to adjust radar rainfall estimates was proposed as early as in the 1950's by Hitschfeld and Bordan (1954), when they considered calibrating the radar against rain gauge measurements. The rain-gauge adjustments are applied in accordance to an assumption that radar rainfall estimates are influenced by a constant multiplicative error (Goudenhoofdt and Delobbe, 2009). This error can be a result of varying coefficients adopted in the Z-R relationship or erroneous electronic calibration. An adjustment factor is then consistently applied to the radar estimates to

take out these errors. The adjustment factor is given by the following equation:

$$F = \frac{\sum_{i=1}^{n} G_i}{\sum_{i=1}^{n} R_i} \tag{2.2}$$

where n is the number of radar and rain gauge pairs, G_i is the ith rain gauge value, and R_i represents the radar value corresponding to i^{th} gauge. The purpose of the rain gauge adjustments is to maintain the accuracy of 'ground truth' data provided by rain gauges whilst retaining the enhanced spatial and temporal resolutions of the radar product (Wang et al., 2013). A summary of the radar quantitative rainfall estimation process adopted by the SAWS from the Australian Bureau of Meteorology is given in Figure 2.3.
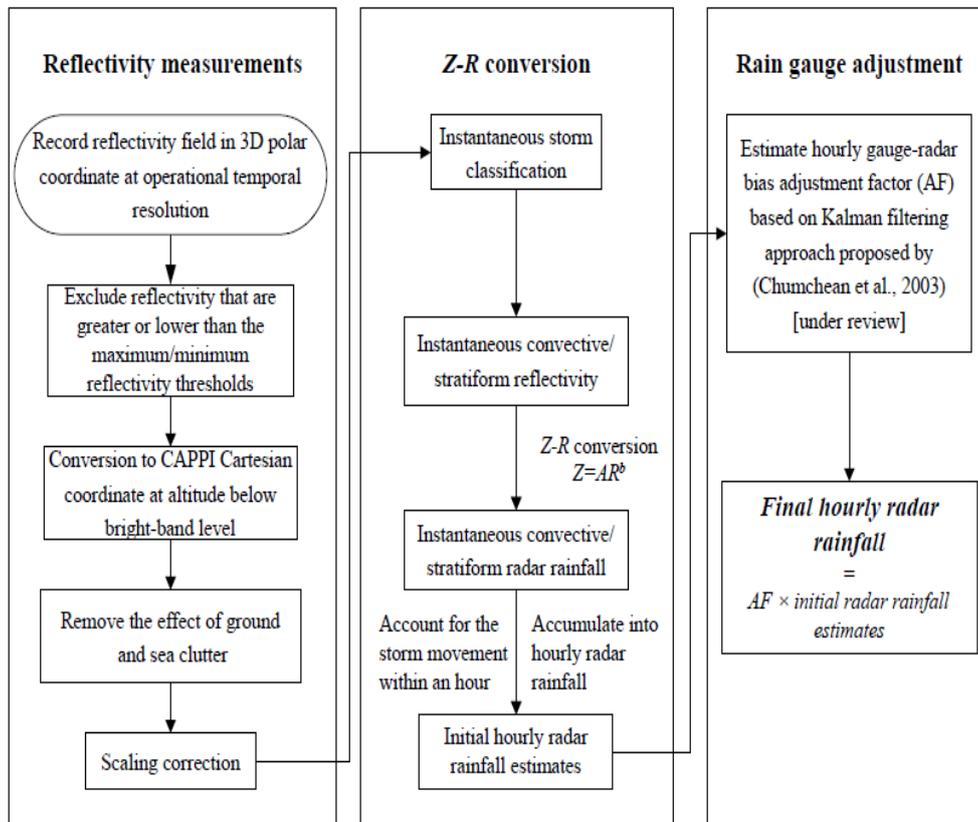
Figure. 2.3 Process for estimating 1-hour rainfall accumulation proposed by the Australian Bureau of Meteorology (Chumchean et al., 2006).

2.2.4 Application of Radar Rainfall Estimates

Since the early 1960s, the application of radar data in hydrology has been viewed to have great potential because of its high spatial and temporal resolutions (McKee, 2015) . Wilson and Brandes (1979) were among the first authors to report an account of quantitative estimation of rainfall using radar over a catchment in Oklahoma, United States of America. These authors used radar Quantitative Precipitation Estimates (QPEs) for flood forecasting and indicated benefits of the spatial resolution of radar in flash flood warning systems over rain gauge data. Similarly,

Van de Beek et al. (2010) examined the application of radar for rainfall estimation over a small catchment in the Netherlands. Their results show that radar data is appropriate for monitoring rainfall events and improves the accuracy of hydrological models. The hydrological models are driven by real-time rainfall predictions. Spatially distributed models which divide catchments into sub-units tend to benefit the most from radar QPEs . Kneis et al. (2017) used radar QPEs to build up components of a flood warning system for a small basin in the Philippines. They observed that radar can significantly extend run-off forecasts produced by heavy rainfall events by 40 minutes. This makes time for decision making processes in hydrological operations. In urban hydrology, radar QPEs have facilitated the analysis of extreme rainfall events. Several countries in Europe have used radar data as input for modeling rainfall-runoff in small urban catchments with a specific goal to improve pipe flow in sewer networks (Einfalt et al., 2004). An investigation on the spatial variability of extreme rainfall events in New York City using gridded stage IV radar QPEs was carried out by Hamidi et al. (2017). Radar rainfall fields clearly highlighted the spatial differentiation of extreme rainfall events within the districts in the City. Furthermore, both hourly and daily accumulations showed that extreme rainfall events in the summer season have a smaller areal extent compared to the winter season. Despite these advantages, the measurement of rainfall by radar has uncertainties associated with it (Golding, 2009 and McKee and Binns, 2016) which are discussed in the next sub-section.

2.2.5 Limitations of Radar Rainfall Estimation

Errors are introduced into radar QPEs during the reflectivity measurement and rain-rate conversion process. Efforts to improve radar rainfall measurements have been proposed but despite significant improvements, considerable errors still exist in the estimation process (Gabriele et al., 2017). The sources of error described in this section affect the application of radar QPE's in hydrological modeling according to (Meischner, 2005 and Villarini and Krajewski, 2010).

2.2.5.1 Variation of the Z-R Relationship

Generally for each radar station the Z-R relationship is calibrated once such that the corresponding coefficients remain constant throughout its operation (Steiner and Smith, 2002). The conversion of reflectivity to rainfall rate is dependent on the DSD. However, the DSD varies according to the type of rainfall and with geographical location. This variation leads to fluctuation of coefficients in the Z-R relationship; for example, *a* in Equation 2.1 increases during a convective storm and decreases for stratiform events (Villarini and Krajewski, 2010). These fluctuations result in underestimation and/or overestimation of rainfall. Studies on the variation of DSD and Z-R relationships in different types of storm events are presented in Uijlenhoet (2001), Habib et al. (2008) and Smith et al. (2009). Lee and Zawadzki (2006) further explain the effects of applying a single Z-R relationship for different storm types and emphasize that the physical processes responsible for the variability of DSD's are a major drawback in radar rainfall estimation.

2.2.5.2   Beam Blocking

Beam blocking is a result of radar beams being obstructed by objects in their path. When the focal part of the radar beam strikes a large object, for example a tall building or mountain, or even a heavy cumulus cloud, it fails to detect rainfall beyond that point. Blocking can be classified as partial or total beam blocking depending on the extent of obstruction (Bech et al., 2003). Partial beam blocking results when a fraction of the beam is blocked leading to an under estimation of precipitation whereas total beam blocking arises when the whole beam is blocked in a way that no precipitation beyond that point is recorded. Reflectivity volume scans at lower levels are mostly affected by blocking and paradoxically, at the same time these are the significant scans in estimating rainfall at ground level (Collier, 1989). Germann et al. (2006) show that in mountainous areas it is difficult to completely avoid beam blocking.

2.2.5.3   Bright Band

When ice particles, typically clusters of snow flakes and graupel, drop through the 0º isotherm during precipitation they begin to melt from the outside inward. This results in the formation of a thin coat of water that surrounds the ice particles. The coated ice particles are observed by the radar as large rainfall drops of higher reflectivity and are visible on the radar rainfall field as a ring of high rainfall intensity known as the bright band, caused by the conical scan of the radar intercepting the almost level melting layer. Ryde (1941) reported the first account of the bright band, so named because of its brighter appearance on the display of early radars. The bright band appears just below the 0º isotherm (freezing level) and can overestimate reflectivity values by a factor of 2-5 (Tabary, 2007). The bright band is mostly observed in stratiform precipitation which presents an evident distinction in particles as we move from one layer to the other in the atmosphere (Deyzel et al., 2004). Snow and ice particles have low reflectivity compared to water droplets, therefore where the radar beam is above the bright band, the result is an underestimation of rainfall at ground level. In South Africa, it is generally observed above the 2 *km* CAPPI level because of the high temperatures which accompany the summer (Pegram et al., 2006). Borga (2002) indicates that S-band radars are much more influenced by bright band errors than other radars. Zhang et al. (2016) show the effects of bright band on QPEs from the KLWX radar in Oklahoma, USA before and after the application of an apparent vertical profile reflectivity correction (AVPR). Before the correction the radar QPE to rain gauge ratio (Q/G) was high (1.75) when compared to the rain gauge measurements. The ratio then decreased to 1.07 after application of the correction.

2.2.5.4   Ground Clutter and Anomalous Propagation

When the side lobes of the radar beam intersect the ground during its propagation, a strong echo is reflected to the radar. These reflected echoes are known as ground clutter. Clutter echoes usually result from steep slopes, mountains and structures immediately around the radar environment. It is easy to point out ground clutter from visualizing a series of instantaneous images since it is stationary, hence it can be easily distinguished from rainfall. Meischner (2005)

investigates the effect of applying an algorithm by Steiner and Smith (2002) to remove ground clutter on reflectivity scans. The authors show that clutter can extremely magnify precipitation estimates by 50-70%. The effects of ground clutter can be minimized by using a Doppler radar and selecting a CAPPI altitude which is preferably 2 km directly above the ground (Becker, 2016). Rinehart (1991) defined anomalous propagation (AP) as the extended detection of objects on the ground by the radar. When the radar transmits waves into the atmosphere it is assumed that the waves move in a straight path to the target object and reflected back along the same path. However, the propagated electromagnetic waves can be refracted (bent) due to the change in density of the atmosphere in which they are travelling through. An in depth description on anomalous propagation of the radar beam in the atmosphere is provided in Doviak et al. (2006). This phenomenon is usually experienced when there is high humidity or in the event of a temperature inversion in the atmosphere. In extreme cases the beam is refracted such that it strikes the earth's surface and reflects off the ground back to the radar in the same path. The radar observes the reflected echo as a high reflectivity point and it is also difficult to pick up in a radar rainfall field since it is not stationary. Harrison et al. (2000) show that eliminating AP can lead to a 30% reduction in error between radar and gauge rainfall estimates. This can be done on a pixel by pixel basis using an integrated quality control system as illustrated by Friedrich et al. (2006). However, AP is still a problem especially when it constantly appears within the radar rainfall echoes (Seo, Krajewski and Mishra, 2015).

2.2.5.5   Attenuation

Attenuation is the reduction of signal strength of the radar beam during transmission in the atmosphere. Attenuation arises when the beam encounters particles (e.g. gases or clouds) which absorb part of its energy and results in the continuing waves getting weaker along the original path of transmission. As the range to the target increases the propagated waves become weaker. Rainfall attenuation is a topic of concern because it is most frequently experienced (Villarini and Krajewski, 2010). The effects of attenuation are typically limited to short wavelength radars such as C- and X bands (Uijlenhoet and Berne, 2008). In the event of extreme attenuation, it is possible for the radar to significantly underestimate the rainfall beyond an intense storm cell. Radome wetting is the main cause of attenuation during rainfall estimation. The effects of radome wetting even reduce reflectivity from nearby storm cells, causing underestimation of precipitation. An investigation to determine attenuation due to a C-band radar radome was carried out by (Kurri and Huuskonen, 2008). Their study shows that radome wetting results in a 3-dB dual transmission loss at a moderate rain rate of 15 mm/h.

2.2.6   Effect of Uncertanties in Radar Rainfall Estimates on Flood Forecasting Studies

Studies indicate that the uncertainties in radar rainfall measurements have an impact on the accuracy of hydrological forecast outputs. Borga (2002) shows that radar rainfall estimates with errors related to beam attenuation and fluctuations in the Z-R relationship can result in stream-

flow simulations with the same magnitude of error as rain gauge-based simulations. A similar study by Kouwen (1988), focusing on errors in modelling streamflow fed by radar-rainfall for flood forecasting, due to anomalous propagation and ground clutter in Ontario, Canada, shows that the sources of error in the radar estimates led to a 10% over-estimation in the predicted flows. Krajewski et al. (2010) used the same but upgraded radar network as Wilson and Brandes (1979) in Oklahoma, USA, to show that developments associated with the technical operation of radars have improved the accuracy of rainfall estimates published by radar stations. However, there are still concerns about the use of raw radar data. Jayakrishnan et al. (2004) and Neary et al. (2004) suggest that the data must be corrected and adjusted before being applied in hydrological simulations. The effect of radar uncertainties in flood forecasting also depends on the catchment characteristics. Bell and Moore (1998) and Vehvilainen et al. (2004) observed that there is an improved accuracy in forecast results in small catchments (<500 km) when radar data is used, while no substantial improvements in accuracy are observed in large catchments. During an extreme rainfall event in Toronto, Canada, a flash flood resulted in devastating damage of property estimated at a cost of $1 billion (Milrad et al., 2015). During this period, a C-band radar stationed in the region estimated 22 mm of rainfall, while a rain gauge station in the same area recorded 126 mm of rainfall (Boodoo et al., 2015). This big difference in rainfall estimates was suggested to be a result of radome wetting and beam attenuation effects. This example shows the possible extent to which radar errors can develop and the resulting consequences when used for forecasting.

2.3   Rainfall Estimation: Rain Gauge

Traditionally, QPEs used in hydrological studies and water resource management have been provided by rain gauges. Currently, the use of rain gauges is still popular and remains one of the most trusted rainfall accumulation measurement method (Chaudhary et al., 2017). A rain gauge is an instrument that measures the rainfall depth during a given time interval at a specific point. In this context, "point" means that the rain gauge typically has an interception area of 0.02 m^2. Generally, the rainfall estimate from the rain gauge is taken as the rainfall representative measurement of the area in which it is located (WMO, 2008). Daily rain gauge, tipping bucket rain gauge, siphon-type rain gauges, optical rain gauge and weighing rain gauge are the commonly used rain gauges in hydrological studies. The SAWS operates practically 1500 rain gauges across the nation and the majority of these are tipping bucket, siphon-type and daily rain gauges (Figure 2.4).

(a) Tipping bucket gauge   (b) Daily rain gauge

(c) Siphon rain gauge
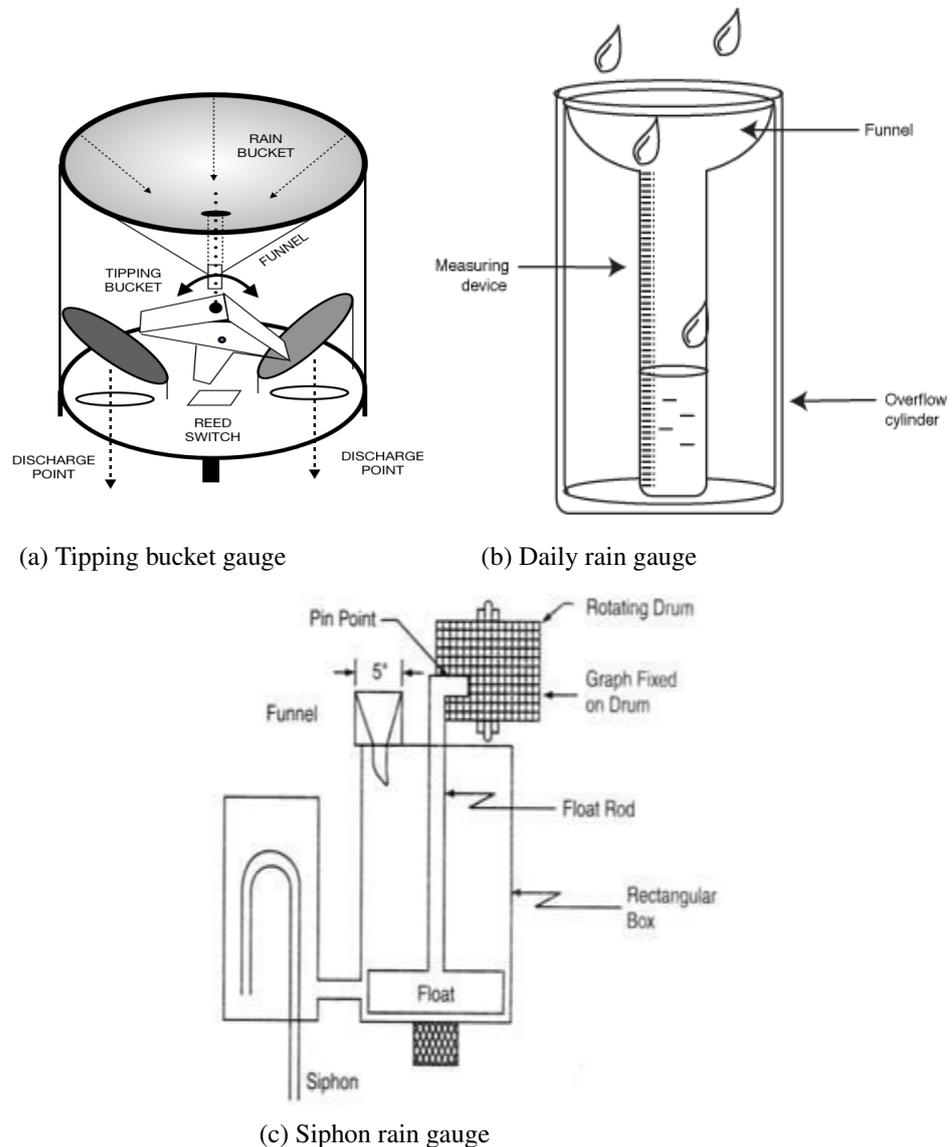
Figure. 2.4 Types of rain gauge (Strangeways, 2006).

2.3.1   Application of Rain Gauge Rainfall Measurements

The first continuous rainfall measurements were done in the late 1670's by Towneley in Lancashire, Britain. These records were used to analyse the variation of the mean annual rainfall in the region (Strangeways, 2006). Since then, rain gauges have been an established instrument for measuring rainfall. Frei and Schär (1998) constructed a precipitation climatology for the European Alps using daily rainfall accumulations from a rain gauge network comprised of 6 600 gauges. The design successfully met the requirements of mesoscale climate dynamics. It was shown from their study that the primary characteristics of precipitation for the region could be determined using the daily rainfall accumulations attained from the rain gauges. Similarly, Lynch (2004) developed a database of annual, monthly and daily rainfall for Southern Africa

using 13 251 rain gauges across the region. The database is used to support hydrological simulation models in the region including the ACRU modelling system in South Africa. The rain gauges have proved reliable at providing daily rainfall extremes (Chaudhary et al., 2017). Despite their ability in providing accurate point rainfall records, the rain gauge measurements have limitations especially on a spatial scale.

2.3.2   Limitations of Rain Gauge Rainfall Measurements

According to Wilson and Brandes (1979), the limitations of rain gauges in estimating rainfall are due to systematic and calibration errors; and their low spatial coverage. Systematic and calibration errors include wind drifts, evaporation losses and splash losses from the collector funnel. These losses affect the amount of rainfall recorded. Wind drifts are the most significant, causing rain gauges to underestimate or overestimate the rain falling on an area by about 3% (Frei and Schär, 1998). A study by Strangeways (2004) on wind effects at varying wind speeds shows that it is difficult to improve the design of rain gauges to eliminate wind drift. This is due to the complex flow of wind in a realistic context. Evaporation losses tend to be sensitive to seasonal changes and surfaces of the collector funnel. Evaporation losses from the storage container itself have been found to be negligible. The latter also applies to splash losses which are only significant during hail storms (WMO, 2008). Rain gauges are mainly limited by their inability to capture the spatial variability of rainfall (Sinclair and Pegram, 2005). This has restricted their use in hydrological studies which often require distributed spatial rainfall input. During a storm in Southern Alberta, Canada, Barge et al. (1979) observed that a rain gauge recorded a rainfall amount which would have triggered a flood warning if the flood prediction model had been based on rain gauge data only. However, radar observations and an analysis of the corresponding stream flow data revealed the localization of rainfall above that specific rain gauge. Considering this, networks which consist of several rain gauges distributed across a catchment are routinely used to determine the spatial variability of rainfall (McKee and Binns, 2016).

2.3.3   Effect of Uncertainties in Rain Gauge Measurements on Flood Forecasting Studies

The high variability of rainfall has demonstrated significant effects in forecasting hydrological scenarios. Schilling and Fuchs (1986) and Faurès et al. (1995) studied the effect of rainfall variability on stream-flow by varying the rain gauge density in a catchment in Arizona, USA. Their results show that in catchments where extreme rainfall events with a large spatially variability are dominant, rain gauge density is a big influencing factor and can lead to inaccurate results in the predicted flows. Although Strangeways (2004) mentions that systematic and calibration errors only contribute 5% to the error in rain gauge estimates, Habib et al. (2008) highlight that this error is transferable and magnified if the estimates are used in forecasting models. Using a catchment in Louisiana, USA, they demonstrate that effects of systematic and calibration errors in gauge estimates result in variations to the order of 15% in the peak flow

estimation. The errors associated with rain gauge estimates reduce the confidence of using rain gauge data alone as input in rainfall-runoff models for flood forecasting. During a flood event in Alberta, Canada, a late flood warning was issued due to inaccurate results from a hydrological model which resulted in serious flooding. Prior to the flood, engineers ran a simulation and the model estimated a peak flow rate of 650 m^3s^{-1} which was not capable of flooding the town. However, hours into the event, the estimated flow rate increased to 985 m^3s^{-1} resulting in complete flooding (McClure and Howell, 2013). The reason behind this erroneous prediction was inaccurate rain gauge estimates and some missing rain gauge records.

2.4   Geostatistical Interpolation of Rainfall

To facilitate the assessment of the MRL5 radar reflectivity scans which are composed of range bins, using rain gauge estimates, the rain gauge data was interpolated using geostatistical methods. It would be infeasible and costly to have a rain gauge for each radar bin. Spatial Interpolation Methods (SIMs) are usually categorized into deterministic and geostatistical methods (Li and Heap, 2011). Traditional interpolation methods such as Inverse Distance Weighting (IDW) and Thiessen Polygons (TP) are the most frequently applied deterministic methods in rainfall interpolation (Ly et al., 2011). These assume that the estimated rainfall at an ungauged location is influenced by a limited number of neighbouring rain gauge measurements near it. However, traditional interpolation methods produce strongly smoothed areal distributions of rainfall which mischaracterize the actual spatial rainfall structure (Berndt et al., 2014). In addition, they do not provide a measure of error. This has driven the quest to improve spatial rainfall estimates using geostatistical interpolators. Previous studies show that the use of traditional interpolators is associated with large estimation errors compared to geostatistical interpolators (Delbari et al., 2013; Mair and Fares, 2010 and Cheng et al., 2017). Geostatistical interpolation methods combine statistics and earth sciences for optimal rainfall estimation. This section provides a review of the common geostatistical interpolation methods used in rainfall interpolation studies. The aim of the review is to provide an understanding of geostatistical interpolators, and their suitability and capability in estimating daily rainfall accumulations. This is made possible by describing their individual methods of operation and application in rainfall interpolation studies. A brief literature summary of Geostatistics is introduced first to help explain the methods.

2.4.1   Introduction to Geostatistics

Geostatistics was originated in the mining industry by a South African statistician and mining engineer named Danie Krige (1951). The experimental work of Krige was used to predict mineral resources from spatially dependent data in gold mines. In the 1960's, Krige's work was formalised by Matheron (1965) and he termed it "Kriging" in his honour. Today, Kriging is a general name used by the geostasticians to describe several interpolation algorithms which are based on least-squares regression.

2.4.1.1 General statistics

Statistics is a mathematical science of collecting, organising, analysing and interpreting quantitative data (Liptser and Shiryaev, 2013). In rainfall interpolation studies, statistical theory considers that a rainfall field can be represented by a subset of rainfall observations acquired in that region. A subset qualifies as a good representative if it shows characteristics of the source rainfall field. Usually, rainfall observations which make up the representative subset are sampled from various locations within the region of interest. Using the representative subset statistical inference is done which leads to conclusions about the rainfall field. Classical descriptive statistics is used to analyse the behaviour of the rainfall observations without considering their geographical distribution. Descriptive statistics are used to summarize continuous data and these include measures of central tendency and measures of spread. However, these are not discussed in this study, as an assumption that the reader has contextual knowledge about statistics has been made. The statistical summaries which are provided in tabular and graphical format can be used to understand the rainfall distribution and trend.

2.4.2 Theory of Regionalized Variables

Matheron's theory of regionalized variables is the backbone of geostatistical interpolation, and it states that *"A mineralized phenomenon can be characterised by the spatial distribution of a certain number of measurable quantities called regionalized variables"* (Matheron, 1971). Matheron (1965) suggested that variation of physical processes should be treated as if they are random. The concept of random properties is translated into a mathematical function called a random process and can be formalized by the following notation (Oliver and Webster, 2014):

(i) Let us say that $\mathbf{z}(x)$ is one of an infinity of values of a random variable $\mathbf{Z}(x)$ at any location x. It is called the "realization" of a process;

(ii) The set of random values at all locations x within a region is called a random process; and

(iii) The random values are spatially dependent to a certain scale.

Rainfall is considered as a spatial random variable because it is randomly distributed in space. The rainfall value $\mathbf{z}(x)$ at any location x, is a realization of a random variable $\mathbf{Z}(x_i)$, while the set of rainfall values defining a rainfall field is a single realization of the random process $\{\mathbf{Z}(x_i), \forall x_i \in \supset\}$, which translates as: the variables $Z(x_i)$ are located at points x_i, all of which are sampled from a superset $\supset$. Considering the above assumptions geostatistics allows us to estimate the value $\mathbf{z}(x)$ at a location x where there is no data available. This estimation is based on the spatial dependence of the variable in question.

2.4.3 Spatial Dependence and the Semivariogram

The concept of spatial dependence can be explained using a simple example. Picture a network of randomly distributed rain gauges over a catchment. Generally, assuming all other variables

being constant, one might expect that the rainfall values measured by the two closest rain gauges would be similar, while those measured by gauges which are far apart would be somehow different. This can be supported by the fact that rain falling on the surface within very short distances (< 50 m) originates from the same clouds and as the distance increases up to tens of kilometres, the rainfall may start to vary because of physical factors such as topography and temperature. At large distances (hundreds of kilometres) the rainfall amount differs greatly because of variations in the average rainfall patterns, in South Africa from one province to the other (Ngoepe, 2016). What has been explained above is a theoretical understanding that describes the spatial dependence of rainfall. Therefore, it can be concluded that spatial dependence is the tendency of a variable to display similar or different values depending on the separation distance (lag) of the spatial locations at which it is measured. However, it is possible for the spatial dependence of rainfall to differ from that described above. Understanding the spatial dependence in rainfall interpolation studies is important because:

- It demands an investigation of the underlying processes that result in the observed rainfall spatial pattern,

- It provides some detail about the spatial structure of rainfall, and

- It forms a solid base from which the estimation of rainfall at ungauged locations can be done accurately.

2.4.3.1   Measuring Spatial Dependence using the Semivariogram

The most common tool used for measuring spatial dependence in geostatistics is the semi-variogram also referred to as variogram in some literature (Webster and Oliver, 2007 and Wackernagel, 2014). It is generated from the measured variable values within a study area, which are daily rainfall accumulations in this context. The semivariogram measures half the average squared difference between pairs of variable values separated by given distances (lags) on the ground using Matheron's method-of-moments estimator:

$$\gamma(h \pm d) = \frac{1}{2|N(h \pm d)|} \sum_{N(h \pm d)} (z_i - z_j)^2 \tag{2.3}$$

where $\gamma(h \pm d)$ is the semivariance at lag h (plus or minus some bin width tolerance range d), h is the distance between observed data points on the ground (lag), where $N(h \pm d) \equiv \left\{ (x_i, x_j) : |x_i, x_j| = (h \pm d); i, j = 1, \ldots, N \right\}$, is the set of all observed data points in each bin and $(z_i - z_j)$ represents the difference between values of random variables (z) at locations i and j.

In computing and applying the semivariogram, an assumption is made that the spatial process responsible for the observed spatial pattern in the variable is uniform throughout the area of interest. This means that the variable is assumed to have a constant mean and variance over the entire study area. If a change occurs in the underlying process throughout the entire study area, which is not detected, it leads to an invalid interpretation of the semivariogram (Cressie and Wikle, 2015). In some circumstances the semivariogram might vary depending on the direction

along which the variable was measured. However, in this study an *isotropic* semivariogram was assumed due to the lack of relevant data (i.e. the semivariance only depends on the lags). Figure 2.5 shows an example of a semivariogram and its characteristics. The semivariogram has three
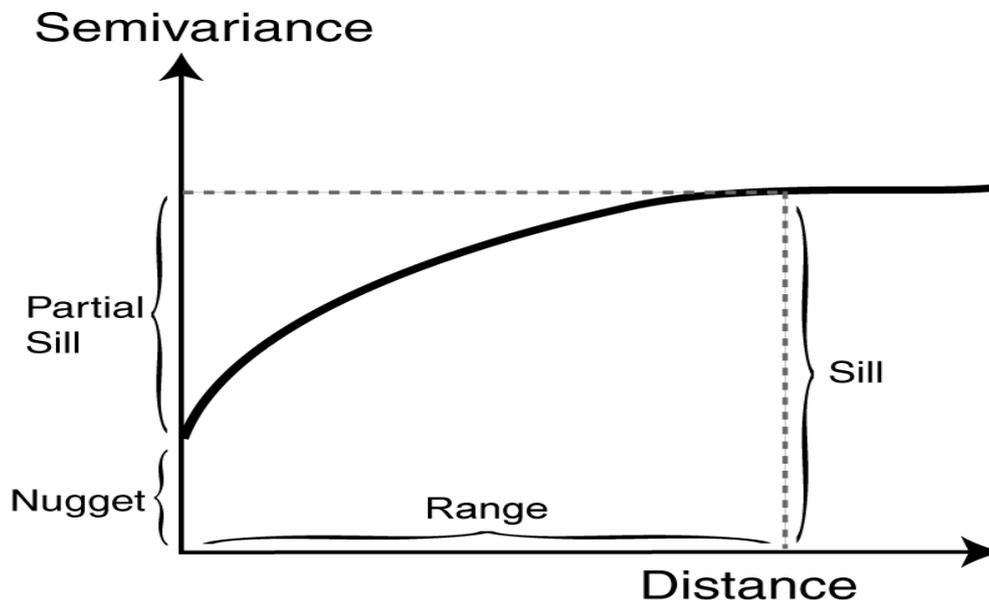
Figure. 2.5 Classic semivariogram model and its features (Anselin et al., 2006).

main characteristics; a *nugget*, a *sill* and a *range*. The *nugget* represents the random component of the semivariogram and is identified by a vertical jump on the y-axis from the origin. It describes the variability at very short lags, for example it can be observed when two close rain gauge measurements are examined, and different rainfall values are obtained. This shows that even at short lags there can exist a difference between the variable values (Houlding, 2000). The nugget effect is usually a result of rain gauge measurement errors and micro variability in rainfall. The *sill* is the asymptotic limit that shows the maximum semivariance in the observed data, while the range describes the distance at which it is reached. At distances above the *range*, spatial dependence between observed data points no longer exists. Therefore, the range marks the zone of influence of one observed value to another within a study area.

2.4.3.2   Modelling the Semivariogram

Usually the experimental semivariogram is defined empirically using Matheron's method-of-moments estimator Equation 2.3. However, in Kriging practice, several theoretical models are fitted to the experimental semivariogram. This is done because the Kriging algorithm will require access to semivariogram values apart from those used to construct the experimental variogram. In most circumstances experimental semivariograms tend to be noisy. Fitting an appropriate model to the experimental semivariogram can be achieved using different methods although it is has remained a controversial topic since the origins of Geostatistics (Oliver and Webster, 2014). Researchers suggest that fitting semivariogram models is more of an art than a

science. The three theoretical models mainly used for rainfall interpolation studies are Gaussian, spherical and exponential models (Wagner et al. (2012) and Xu et al., (2015)).

2.4.4   Kriging

Kriging methods make use of sparsely sampled environmental variables to estimate their values at unsampled locations over a random field. In rainfall estimation studies, Kriging uses a weighted average of neighbouring rainfall observations to estimate the unknown rainfall values at given locations. The Kriging weights are calculated based on the spatial structure of the observed data that is represented by the semivariogram model and are applied to the observed data according to Equation 2.4. The weights are calculated in such a way that the closest observations to the unsampled location carry more weight than observations that are further away. Rainfall estimation using Kriging can be done at any point between observations without bias and with a proper interpretation of the errors. The fact that Kriging methods have a measurement of error assists engineers in pre-design practice for hydrological applications and models. Most Kriging methods use a common estimator known as the "Kriging predictor" which is given as:

$$\hat{Z}(x_0) = \sum_{i=1}^{n} \lambda_i Z(x_i) \tag{2.4}$$

where $\hat{Z}(x_0)$ is the estimated rainfall at location x_0 , $Z(x_i)$ is the observed rainfall at location x_i, λ_i is the relative weight (Kriging weight) and n represents the number of observations used in the estimation. The Kriging predictor is a linear unbiased estimator and provides an exact interpolation, meaning that the prediction error is minimized at every point of the estimated values. The sub-sections below describe the common Kriging methods used for rainfall estimation.

2.4.4.1   Simple Kriging (SK)

Simple Kriging is the most basic Kriging method. It is used to predict residuals from a known mean of a stationary process given *a priori* (Wackernagel, 2014). The mean of the stationary process is calculated as the average of the observed data over the field of interest and is assumed to be the same across the whole field. Considering the above-mentioned conditions, the mean is used to improve the estimation of variables at unsampled locations. The SK estimate is computed by adding a constant mean to the Kriging predictor in Equation 2.4 as follows:

$$\hat{Z}_{sk}(x_0) = \mu + \sum_{i=1}^{n} \lambda_i^{sk}[Z(x_i) - \mu] \tag{2.5}$$

where μ is the mean and $(Z(x_i) - \mu)$ represents the residual. The simple Kriging weights λ_i^{sk} are determined by solving the simple Kriging system:

$$\sum_{i=1}^{n} \lambda_i^{sk} C_r(x_i, x_j) = C_r(x_i, x_0) \qquad\qquad for\ all\ i, j = 1, 2, \ldots, n \tag{2.6}$$

Where $C_r(x_i, x_j)$ is the covariance between the observed data points; and $C_r(x_i, x_0)$ is the covariance between the observed point and unknown point. Since the SK method assumes a constant mean across the whole study region, it cannot be used for rainfall estimation because the rainfall mean varies with location. This makes SK a typical textbook model which is seldom used in a realistic context.

2.4.4.2 Simple Kriging with local varying mean (SKvm)

SKvm is categorized under multivariate Kriging methods. It is an extension of the simple Kriging model where the constant mean is substituted by locally varying means. The skvm method uses secondary variables observed in the environment of the primary variable to estimate the varying means (Goovaerts, 2000). The varying means are estimated prior to the Kriging process using linear regression. Radar or altitude data can be used as secondary variables in estimating rainfall and they must be measurable at each rain gauge location. The linear regression is carried out at each rain gauge location using n nearest observations to obtain estimates which are treated as the local varying means. Replacing the constant mean (μ) in Equation 2.5 with the local varying mean $m(x)$ gives the skvm estimate:

$$\hat{Z}_{skvm}(x_0) = m(x_0) + \sum_{i=1}^{n} \lambda_i^{sk}[Z(x_i) - m(x_0)] \qquad (2.7)$$

Skvm weights are determined using the same Kriging system as in the SK method. Lloyd (2005) incorporated elevation as a secondary variable for spatial estimation of monthly rainfall in Britain. His study shows that the method provides accurate rainfall estimates only when there is a linear relationship between the rainfall amount and elevation. In South Korea, Park et al. (2017) investigated the accuracy of applying Skvm using multiple linear regression functions to estimate the varying means. The authors show that Skvm combined with multiple linear regression exhibits greater accuracy compared to gauge only interpolation methods and it increases the success rate of indicating rainy days from non-rain days.

2.4.4.3 Ordinary Kriging (OK)

Ordinary Kriging is a surface fitting technique used for spatial interpolation of random variables using given measurements or "known" observations at a set of locations in a region. This operation requires a lot of judgement and can be thought of as spatial forecasting because it is essentially combining available information to derive something for an observation we do not yet observe. The method assumes a constant unknown mean across a rainfall field and does not require any secondary variables in its estimation. The OK estimate is calculated using the Kriging estimator:

$$\hat{Z}_{ok}(x_0) = \sum_{i=1}^{n} \lambda_i^{ok} Z(x_i) \qquad (2.8)$$

where $\hat{Z}_{ok}(x_0)$ is the estimated rainfall at location x_0, $Z(x_i)$ is the observed rainfall at location x_i, λ_i^{ok} is the Ordinary Kriging weight and n is the number of observations used in the estimation.

OK weights are determined using a Kriging system of equations equal to the number of rain gauges as will be used in Section 3.5. The OK estimator is labelled as the Best Linear Unbiased predictor (BLUP) (Rabiei and Haberlandt, 2015), best in terms of its ability to minimize the estimated variance during the estimation process and linear because it gives a prediction based on the weighted linear sum of the known observations (Adhikary et al., 2015). According to Li and Heap (2011) OK is more often used in rainfall interpolation studies compared to the other Kriging methods described in this section. Xu et al. (2015) used the OK method to interpolate daily rainfall from 43 rain gauge stations during the rainy season in Sichuan, China. Their study showed that the OK method is optimal for interpolating daily rainfall as it recorded lower Mean Average Error (MAE) and Root Mean Square Error (RMSE) than deterministic methods (IDW and TP). Lloyd (2005) compared OK with Skvm and Kriging with External Drift (KED) and found that KED provided more accurate rainfall estimates than OK in some circumstances but OK performed well overall.

2.4.4.4 Kriging with External Drift (KED)

KED was developed in the mining industry in the early 1970s (Delhomme, 1978). The method has since been considerably extended and applied to solve practical problems in soil conservation and hydrological studies. A detailed description of the method is presented in Wackernagel (2014). It allows the estimation of an unknown variable (Z) at a set of locations using another variable (s) which is sparsely sampled in the same region. $Z(x)$ is chosen as the random function to model Z, and s is represented by a deterministic function $s(x)$. The two variables Z and s are assumed to be linearly related:

$$E[Z(x)] = a_0 + a_i s(x) \tag{2.9}$$

Where a_0 and a_i are regression coefficients. The deterministic function $s(x)$ is used as the mean and its shape improves the prediction of the unknown variable. The deterministic function introduces an additional constraint to the Kriging system of equations used to calculate the weights such that on average the weight must be consistent with the exact interpolation of $s(x)$:

$$s(x_0) = \sum_{i=1}^{n} \lambda_i s(x_i) \tag{2.10}$$

where $s(x_i)$ is the external drift at location i, $s(x_0)$ is the estimated external drift at an unsampled location and λ_i represents the Kriging weight. The universal Kriging equations are used to determine the weights in KED. Recently KED has been applied to combine radar and gauge information to improve rainfall estimation for hydrological applications (Jewell and Gaussiat, 2015). Gabriele et al. (2017) used KED to merge weather radar and gauge rainfall fields in Calabria, Italy and analysed the merged product using a rainfall-runoff simulation model. The addition of weather radar data improved the spatial estimation of rainfall fields especially for small catchments $\sim 1km^2$ in which the difference between the average areal rainfall recorded by gauges only and KED reached a peak of 40%.

2.4.4.5 Conditional Merging (CM)

Conditional merging also referred to as "Kriging with error based correction" was proposed by Sinclair and Pegram (2005). Conditional merging estimates the errors associated with an ordinary kriged rain gauge network using radar rainfall fields. The kriged rain gauge network is then corrected accordingly using radar observations, thus improving the spatial structure of the final rainfall field. KRE is carried out in the following steps:

1. Radar pixels at each rain gauge location are used to obtain a kriged rainfall field based on the radar values (R_k),

2. An error field is obtained by subtracting the radar-based kriged field from the original observed radar field as follows:

$$\varepsilon(x_i) = R(x_i) - R_k(x_i),\tag{2.11}$$

3. The error field is combined with the ordinary kriged field of rain gauges to get a corrected rainfall field using the following expression according to Sinclair and Pegram (2005):

$$C(x_i) = G_k(x_i) + \varepsilon(x_i)\tag{2.12}$$

where $C(x_i)$ is the corrected rainfall field, $G_k(x_i)$ kriged rain gauge field and $\varepsilon(x_i)$ represents the error correction field. Goudenhoofdt and Delobbe (2009) examined methods of combining radar and rain gauge products for daily interpolation of rainfall and his results show that CM performs well as KED, reducing the error by 40% with respect to gauge only data. In a similiar study by Pettazzi and Salsón (2012) during a flood in Galicia, Italy, CM also showed a greater accuracy than raw radar estimates.

2.4.5 Justification for using Ordinary Kriging in this Study

In this study the OK method is selected as the suitable geostatistical interpolator based on an extensive literature review. A study by Li and Heap (2008), in which 34 spatial interpolators for environmental variables were reviewed showed that out of 51 comparative studies, OK had the highest frequency and recommendation for spatial estimation with 37% and 8% respectively. Goudenhoofdt and Delobbe (2009) examine the performance of radar-gauge merging methods for estimating daily rainfall using OK as reference method for evaluation purposes, in which they show that the performance of OK with rain gauge data only was close to that of the geostatistical methods (KRI and KED) and it outperformed the original radar data. Mair and Fares (2010) estimated daily rainfall for wet and dry seasons using traditional interpolators (IDW and TP) and geostatistical interpolators (OK and Skvm) in O'ahu island, Hawaii. In their analysis, OK produced more accurate rainfall estimates, and it could indicate the regions with the highest rainfall shortfalls. The use of multivariate Kriging methods introduces additional problems in relation to data management and computational needs. Considering computational needs is

important when selecting geostatistical interpolators. The daily accumulations of the MRL5 radar data are available in the MDV format which requires high storage efficiency. Handling of this data combined with rain gauge data is computationally intensive.

2.5   Operational and Environmental Factors Influencing Spatial Rainfall Estimation

Spatial rainfall estimation can be carried out using several methods, but the difficulty lies in selecting the optimal method for a particular purpose. Reliable estimation techniques must provide rainfall estimates which meet specific demands such as: accuracy, applicability to large datasets, flexibility to describe various types of precipitation and computational efficiency. It is challenging to find a technique that satisfies all these demands for dissimilar geographical locations. Using an inappropriate technique can lead to inaccurate information which results in misinterpretation of hydrological scenarios. Wagner et al. (2012) show that inaccurate rainfall input affects parameter estimation and compromises the simulations of stream flows. This section describes common operational and environmental factors which influence the reliability of radar and rain gauge derived rainfall data. The reader is referred to Li and Heap (2008) and Berne and Krajewski (2013) for an extensive review.

2.5.1   Effect of Topography

Operational radar errors in catchments with mountains arise mostly due to the variability of precipitation with altitude and beam blocking. Mountains are a source of ground clutter and beam blocking as recorded by the radar, therefore to avoid contamination, radar beams in such areas are transmitted at a relatively larger elevation. Usually in a flat terrain, the radar beam is propagated at elevations less than $1.5°$ but in mountainous areas it is operated above $1.5°$. The propagation of beams at higher elevation sweeps increases the observation altitude of the weather system, hence when there is high variation in the nature of precipitation relative to height, the the radar observations might not be representative of ground rainfall (Gourley et al., 2002). Villarini and Krajewski (2010) and Lin et al. (2016) highlight that besides beam blocking and un-detection of low-level precipitation being the main sources of radar errors in mountains, the limited availability of rain gauges to correct the Z-R relationship is also a significant source of error. Rain gauge stations in mountains are usually located in valleys, so this tends to bias the rainfall information towards low areas (Sanchez-Moreno et al., 2014). Furthermore, precipitation in mountains is characterized by strong winds which influence the spatial distribution of rainfall. Due to the erratic and turbulent nature of the wind, it is difficult to characterize the rainfall using rain gauge stations. Borga et al. (2000) adopted radar rainfall estimates for flood estimation using the TOPMODEL over a mountainous Basin in Northern Italy. They compared the observed hydrographs with those simulated from rain gauges and conclude that radar data hinders the accuracy of rainfall simulation. The inaccuracy was blamed on the radar scans done at a higher level with the initial purpose of reducing topography interception.

2.5.2   Effect of Storm Type and Seasonal Changes

Radar reflectivity is sensitive to rain drop diameter which determines the DSD. Ulbrich and Atlas (2002), Zawadzki (2006) Villarini and Krajewski (2010) and Seo, Dolan, Krajewski, Rutledge and Petersen (2015) discuss how storm type affects radar rainfall estimation. The DSD of stratiform and convective rain are different because of the distinct physical processes responsible for each rainfall type. An investigation on the difference between DSD derived from stratiform and convective rainfall carried out by Yuter and Houze Jr (1997) shows that both stratiform and convective rain can be comprised of large and small drop spectra. The authors concluded that the two rainfall types cannot be distinguished using drop size spectra and recommended the use of an appropriate Z-R relationship for the two different rainfall types. Contrastingly, Atlas et al. (2000) suggest that stratiform rain is made up of smaller rainfall drops compared to convective rain on average, hence there is a significant difference between Z-R relationships derived from both rainfall types. Figure 2.6 shows the average DSD relationships derived from stratiform and convective rains.
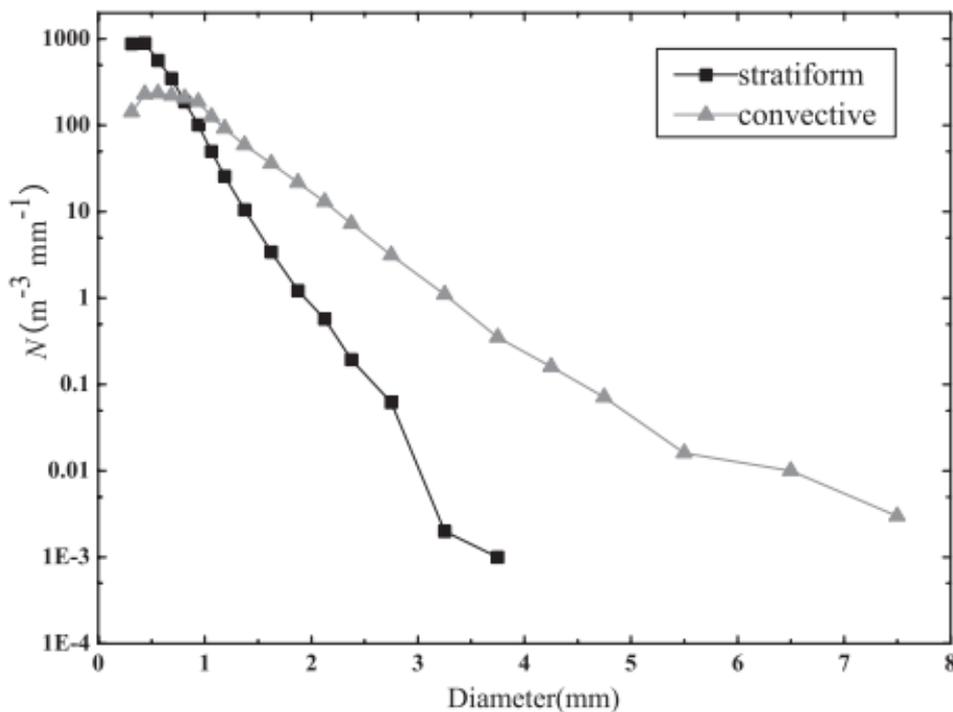
Figure. 2.6 DSD for stratiform and convective rains (Niu et al., 2010).

The difference between stratiform and convective rain is mainly responsible for the seasonal effects in radar rainfall estimation in South Africa. In most regions the summer season is dominated by convective rain whilst stratiform rain is received in the winter. Rain gauges miss more information during convective rainfall events than stratiform events (Smith et al., 2007). This is because convective rainfall is localized with high rainfall intensities at short time scales which are usually misrepresented by rain gauge networks, whereas stratiform rainfall is more smoothly and uniformly distributed at low intensities for long durations.

2.5.3   Effect of Radar Range

The main source of radar range biases are beam broadening and elevation of the sampling volume. Seo et al. (2011) describes range related errors in radar rainfall estimates. The match between radar and rain gauge data decreases as the range from the radar increases. As we move away from the radar location, the beam height above the ground increases due to the earth's curvature, this leads to overshooting of the weather system. In such circumstances there is a higher probability that the observed precipitation aloft is different from ground rainfall. Evaporation and advection are the common factors responsible for this observed difference. In addition, the radar sampling volume at long ranges is large, therefore there is higher chance that precipitation will not uniformly fill up the beam. Burcea et al. (2012) examined the radar range effects for 6 convective storms over the Moldavian Plateau in Romania. Their study reported that radar estimates are reliable within a 120-150 km range as demonstrated by the decrease of radar-gauge correlations as we move away from the radar. However, the spatial coverage provided by the radar is still beneficial. Peleg et al. (2018) examined the spatial characteristics of rain gauge and radar products under heavy rainfall events over a catchment in Northern Israel. Their results show that weather radar performed well compared to the dense rain gauge network of 200 rain gauges. Also, the radar could observe rainfall values of above 25 mm from a spatial coverage of 2000 km^2 while the rain gauges recorded a maximum value of 22 mm.

2.5.4   Density of the Rain Gauge Network

Rain gauge density is a key factor in assessing the spatial variability of rainfall. Flood forecasting requires a dense rain gauge network that is able to capture localized rainfall. The mis-representation of rainfall during high intensity rainfall events leads to substantial error in predicted flows (Golding, 2009). The WMO suggests that the choice of rain gauge density must match the catchment size and type (for example, in mountainous catchments they recommend one rain gauge per 250 km^2). Volkmann et al. (2010) recommend a rain gauge density of one rain gauge per 70 km^2 to obtain sampling errors below 5% for daily rainfall accumulations. However, Smith et al. (2007) suggest that it is difficult to achieve an optimal rain gauge density in catchments. Adhikary (2017) mention that economic and practical factors are mostly responsible for poor gauge networks in hydrological catchments. When using interpolation methods, Biggs and Atkinson (2011) observe that increasing the rain gauge density results in a greater estimation accuracy only up to a given point. Their results show that using six or ten rain gauges produced similar accuracy for radar adjustments over a catchment size of 2000 km^2 in the Severn Uplands, England. The use of less than six rain gauges resulted in a decrease in the interpolation accuracy, indicating the impact of the rain gauge density on interpolation methods. It can be concluded that interpolation of rain gauge data is dependent on the network density and spacing.

2.6   Summary

This chapter provided a review of rainfall estimation using radar, rain gauge and geostatistical interpolation methods. The review assists in the assessment of rainfall estimation tools considered for this study in operational conditions. While research indicates that spatial rainfall estimates derived from radar are more accurate compared to spatial interpolation using point rain gauge estimates, few South African studies have compared the quality of radar data using geostatistical interpolated rain gauge data at a catchment scale. Such research is very important to develop the application of both radar and rain gauge estimates in an operational context.

Chapter 3

Description of the Study Area, Rainfall data, Ordinary Kriging and Comparison Statistics

3.1   Introduction

This chapter describes the study area, radar and rain gauge rainfall data, OK method and error statistics. A sub-catchment of the Vaal, located near Bethlehem in the Free State Province was selected to aid the research purpose. This sub-catchment was selected on the basis of terrain, availability of daily rainfall records and its water supply contribution in South Africa. The rainfall data acquired from both the weather radar and rain gauges for this study are assumed to be reasonably accurate and obey the theory of regionalised variables. The study focuses on 24-hour accumulations recorded at weather stations during the 1994-2003 period. The last part of this Chapter focuses on comparison statistics used to assess the difference between radar estimates and rain gauge derived data.

3.2   Description of the Study Area

A sub-catchment of the Vaal Catchment located near Bethelehem in the Free State Province was selected for this study. It is located approximately between latitudes $27.5°$ and $29.0°$ and longitudes $-27.5°$ and $-29.0°$, as shown in Figure 3.1. The Vaal catchment is of important strategic interest with respect to water resource management and the control of floods in South Africa. Droughts are common in this region since it is in the sub-tropical high-pressure belt, however periods of drought are frequently broken by disastrous flooding events. The Department of Water and Sanitation (DWS) is responsible for managing the water resources in the Vaal catchment. The catchment is comprised of an area about 38 505 km^2 and is one of the main sources of stored water for Johannesburg and other surrounding cities. There is a high demand of water in this region, as it plays host to large scale mining, industrial and agricultural activities. For example, the coal power generation industry in this region generates 80% of the countries electricity. The SAWS operate two radars in the Vaal catchment, an S-band radar near Bethlehem and a C-band radar at Ermelo, which are used to monitor rainfall over the region.

Bethlehem experiences a semiarid summer rainfall highveld climate and lies at an altitude about 1700 metres. The weather in Bethlehem is characteristic of an interior plateau with a higher percentage of rainfall during the summer (October-April) and chilly dry winters. The annual average rainfall received is 650 mm, in a region where convective rainfall is dominant. Despite the mountain range in Lesotho which is less than 60 km away South East of Bethlehem, the topography is generally flat hence there are minimal orographic effects . The average high
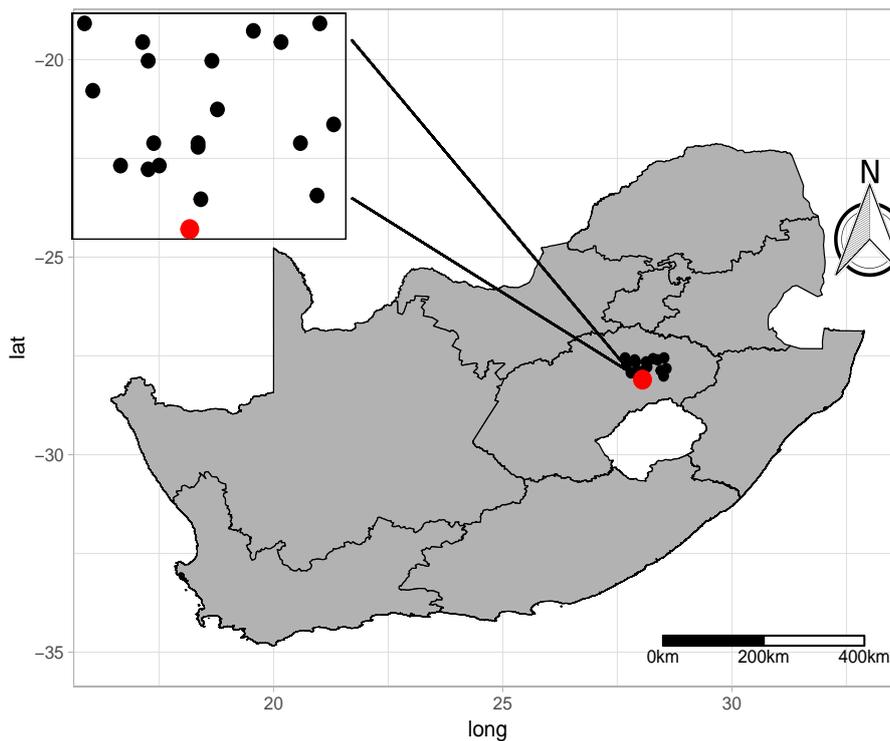
Figure. 3.1 The diagram shows the location of the MRL5 radar (red dot) and the spatial distribution of the study rain gauges (black dots) in the Free State, South Africa.

temperature is about $22°C$ and the average low temperature is about $6°C$ with January being the hottest month, while June is the coldest.

3.3   Description of Rainfall Data

3.3.1   Radar Data

The National Weather Radar Network (NWRN) operated by the SAWS consists of 14 operational radars. Figure 3.2 shows the NWRN operated by the SAWS with respect to provinces in South Africa. The black rings represent the useful range of radars operating in the C-band frequency and the S-band frequency radars are represented by the green rings. The radars with a maximum coverage of 150 *km* range are indicated by the smaller rings and the bigger rings indicate S-band radars operating at a range of 300 *km*. A detailed summary of the NWRN is given in Appendix A. The MRL5 radar at Bethlehem was used for this study. The S-band radar was obtained from Russia by the WRC in 1994 and rainfall data obtained from the radar station since 1992 are available. The radar dome is built on top of a control centre which has a developed Radar Data Acquisition System (RDAS) which utilizes the DISPLACE method by Terblanche (1996) to process the digitalized logarithmic receiver output. There is also a backup power supply available at the radar site to prevent data loss during electricity power cuts.

Figure. 3.2 Coverage of the 14 operational weather radars in the National Weather Radar Network (Jager, 2017).

Table 3.1 gives details of the MRL5 radar at Bethlehem. The MRL5 radar collected reflectivity data in 600 m bins along the beam up to a range of 150 *km*. When operating in volume scan mode the radar collects 224 bins in one complete revolution at an incremental azimuth of 1º. The acquired radar data is transmitted to the Weather Bureau in Bethlehem where it is processed and transformed into CAPPI format at an altitude of 2 km. Before being published, the radar data is corrected to remove clutter, beam blocking, attenuation and anomalous propagation errors. The MRL5 radar data is made available by the SAWS.

Table 3.1 Characteristics of the MRL5 radar at Bethlehem

Radar Station	MRL5
Location	Bethlehem, Free State Province
Latitude	28.098368 degrees
Longitude	28.163235 degrees
Measurement cycle	5 minutes
Frequency band	S (10 cm)
Doppler mode	Yes
Altitude	1722 m

The daily accumulations used in this study were available in Network Common Data Format (NETCDF classic) designed, and then provided to the author, by Dr Scott Sinclair of Civil Engineering, UKZN. NETCDF is a set of software libraries that has data access functions for storing and extracting data in the form of arrays. It is specifically designed for storing climatological data and allows the re-usability of software for array-oriented data management, display and analysis. Its goal is to provide efficient access to subsets within large datasets. The daily radar accumulations were accessed using a Python Netcdf4 package. Each file contains

the rainfall depth, units, accumulation period, number of radar scans and dimensions. The daily rainfall accumulations are extracted and displayed on a regular 1x1 km grid as shown in Figure 3.3.
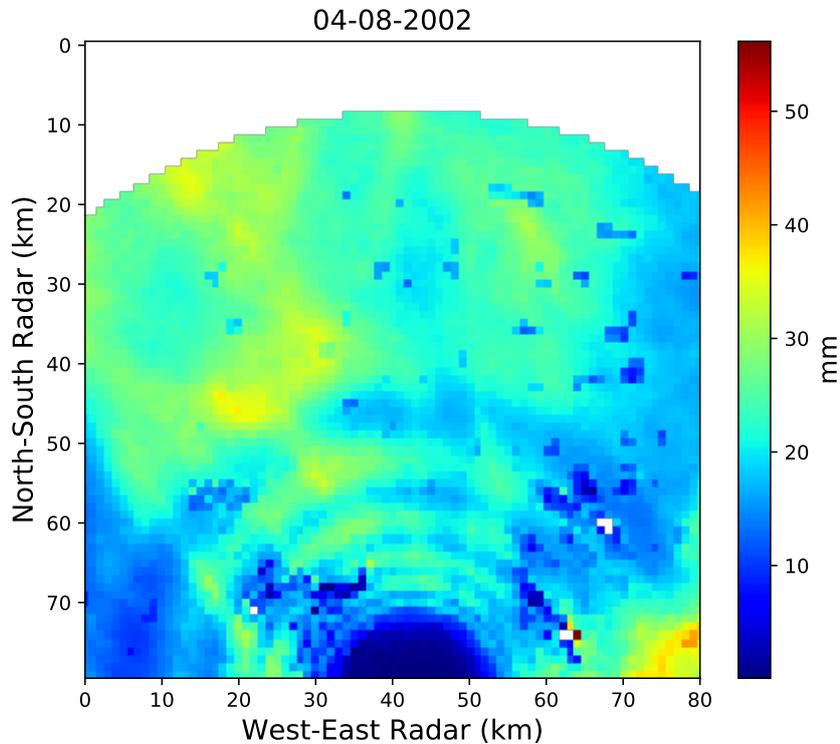
Figure. 3.3 MRL5 radar coverage of the study area shown on a 1x1km grid.

3.3.2   Rain Gauge Data

The SAWS operate more than a thousand gauge rainfall stations nationwide, of which 26 of these are installed at official weather offices. These rainfall stations include both manual and automatic stations. Collection of rainfall data is done via an SMS service for the automatic stations and through an observer report for the manual stations. The SAWS climatological rainfall database is mostly comprised of the graduated cylinder rain gauge, which is read manually every day. The tipping bucket rain gauge network is found at automated rain stations which have wireless systems of communication, which report to the main server every 15 minutes during a rainfall event and once a day when there is no rain. The weather stations are classified into three groups which are: first order, second order and third order stations. This classification is based on the number of observations that are taken per day as shown in Table 3.2 (Becker, 2016).

Table 3.2 Classification of weather stations

Class	Number of readings per day	Times SAST
First order	3	08:00, 14:00 and 20:00
Second order	2	08:00 and 14:00
Third order	1	08:00

A network of rain gauges is used to record and monitor rainfall accumulations in the study region. The network of rain gauges consists of daily read national gauges which are maintained by local volunteers and the SAWS respectively. The rain gauge data have a progression of over 80 years. The gauges used for this study are uniformly spread on an area of about 4 600 km^2. The raw rainfall dataset used in this study was pre-processed by the Climate Systems Analysis Group (CSAG) of the University of Capetown. The database was cleaned using a methodology developed by the Global Historical Climatology Network (GHCN) (Durre et al., 2010). The automated quality control methodology involves nineteen carefully assessed tests that sense meteorological outliers, duplicate information and several anomalies. Although the quality control procedure is automated, it has been manually validated by approaches suggested by Durre et al. (2008) to guarantee the best possible performance. Details of the methodology used in the quality control procedure are not discussed in this study, but the result is a dataset with good quality data over the Vaal catchment since 1960. The study uses data from 22 rain gauges recorded between 1994 and 2003. The rain gauges are stationed within a 8000 km^2 area covering part of the MRL5 radar domain (Figure 3.1). The rainfall data is made available in the form of text files, with each text file providing all the daily records for a specific rain gauge, the date, the geographical coordinates of the rain gauge and the station name. The data is extracted from the text files using Python v3.6 for a modeling quality check. The quality control procedure adopted by CSAG flags records with erroneous data, missing data and extreme maximas but does not remove the records from the dataset. Therefore the modeling quality check removes flagged rain gauges during the quality control procedure and organizes the data into a format most suited for programming purposes.

3.4   Exploratory Data Analysis (EDA)

Exploratory data analysis is a prerequisite for any geostatistical work or programming related project. EDA is the analysis of data using visual and quantitative methods with the aim of understanding and summarizing the dataset without considering any assumptions made about the contents of the dataset. In rainfall interpolation, EDA is done prior to the actual modelling because it provides critical information required to develop an appropriate interpolation model and it also assists the interpretation of the results. The basic concept behind EDA is that we need to have knowledge about the make-up of our data before applying any predictive algorithms. EDA carried out in this study is a combination of the following methods:

- Descriptive statistics and graphical visualization of each field in the rainfall dataset.

- Statistics for assessing the relationship that exists amongst the study rain gauges and clustering of similar rain gauge observations into groups.

3.4.1 Descriptive Statistics and Data Visualization

The rain gauge observations are used to draw conclusions about the rainfall distribution in the study region. The Python Seaborn package was used for the statistical analyses and visualization of the rainfall data. The descriptive statistical concepts used are measures of central tendency, measures of central variability and measures of symmetry. Measures of central tendency included the arithmetic average of the rainfall data (mean), central value of the rainfall data (median) and the most frequently occurring value in the dataset (mode). From these, information on the range and variability of the rainfall can be deduced, that is, the spread of rainfall data between the minimum and maximum observations, and the spread of rainfall values around the mean respectively. Skewness was considered as the measure of symmetry of the rainfall data. Skew is a measure of the extent to which distribution of a particular dataset departs from symmetry.

Usually rainfall data exhibit significant skewness, however the Kriging practice tends to produce optimal results when the data is normally distributed (Mukhopadhaya, 2016). To further describe the rainfall data, graphical presentations were done using histograms, box plots and scatter plots. The descriptive statistics and graphical presentations were used to examine data consistency and outliers. Rainfall measurements which significantly differ from neighbouring observations or general spread of values were flagged out. These observations were removed from the study rainfall dataset as they were considered invalid. Scatter plots showing the positions of the study rain gauges also referred to as "posting" illustrate the sampling characteristics (Webster and Oliver, 2007). This aids in identifying areas where there is oversampling, which may introduce bias in the estimated results. Only rainfall events where more than 50% of the study rain gauges received rainfall above 3 *mm* were selected since the end purpose of the rainfall estimates in this study could be for flood forecasting applications. Traces of rainfall (rainfall less than 3 *mm*) have no considerable effect on stream flows and furthermore, OK is sensitive to low values because it is difficult to estimate a robust semivariogram for low rainfall events (Soenario et al., 2010).

3.4.2 Probability Statistics and Clustering

The assumption of normality is important when using the Kriging method because the data observations are assumed to be from a normally distributed population to get the best possible results from the Kriging estimation. The Q-Q plot was used to assess whether the observed rainfall data follows a normal distribution. This is done by plotting two quantiles against each other, in this case the quantiles of the observed rainfall values are plotted against quantiles from a theoretical normal distribution. A Q-Q plot, or quantile-quantile plot, is a graphical tool used to assess if a set of data plausibly came from some theoretical distribution such as a Normal or exponential. On the Q-Q plot a 45° line is drawn which is the reference for normality in the

observed data. If the observed data comes from a normal distribution the data points should lie near the reference line. It is also common for observed data to be biased, for example a cluster of data might be present at a specific area which leads to over-sampling of a statistical population. K-means clustering (MacQueen et al., 1967) was used to classify the rainfall stations by their statistical summaries.

K-means clustering is an exploratory analysis technique used to group data points into homogeneous classes. The K-means method was applied to obtain an insight on how the rainfall data behaves and group the rainfall stations with similar patterns. The resulting clusters are natural rainfall data groups based on specific statistical combinations. K-means clustering is implemented in several sequenced steps, which in this study were carried out in Python using the *sklearn* package. The K-means algorithm aims to minimize an objective function known as the squared error which is defined as (MacQueen et al., 1967):

$$K(m) = \sum_{i=1}^{c} \sum_{j=1}^{c_i} (|x_i - x_j|)^2 \tag{3.1}$$

where $|x_i - x_j|$ is the euclidean distance between gauge stations x_i and x_j, c_i is the number of gauge stations in the ith cluster and c represents the number of centroids (or cluster centers). Let $G = \{g_1, g_2, g_3, \ldots, g_n\}$ be a set of rain gauges and $M = \{m_1, m_2, m_3, \ldots, m_c\}$ a set of centroids. In centorid-based clustering, clusters are represented by a central vector or location, which may not necessarily be a member of the data set. When the number of clusters is fixed to c, K-means clustering gives a formal definition as an optimization problem: find the c cluster centers and assign the objects to the nearest cluster center, such that the squared distances from the objects to their cluster centre are minimized. After all gauge stations have been assigned, the centroids are recalculated using the following equation:

$$M_i = (\frac{1}{c_i}) \sum_{j=1}^{c_i} g_i \tag{3.2}$$

where c_i is the number of gauge station in the ith cluster. The rain gauge stations are reassigned to their closest centroids which in some cases result in some stations shifting from one cluster to the other. Updating and assignment of rain gauge stations is repeated until convergence, that is, until there is no shifting of points between clusters also referred to as "stabilization". The initial number of clusters was selected using the elbow method where average distances between the centroid and the rain gauge stations are plotted across a range of clusters. The curve in this plot falls rapidly until the right number of clusters k to begin with is reached, and then it continues to fall gradually. The point where the curve starts to fall gradually is selected as the initial number of clusters to use. Figure 3.4 illustrates the elbow method for data with two clusters.
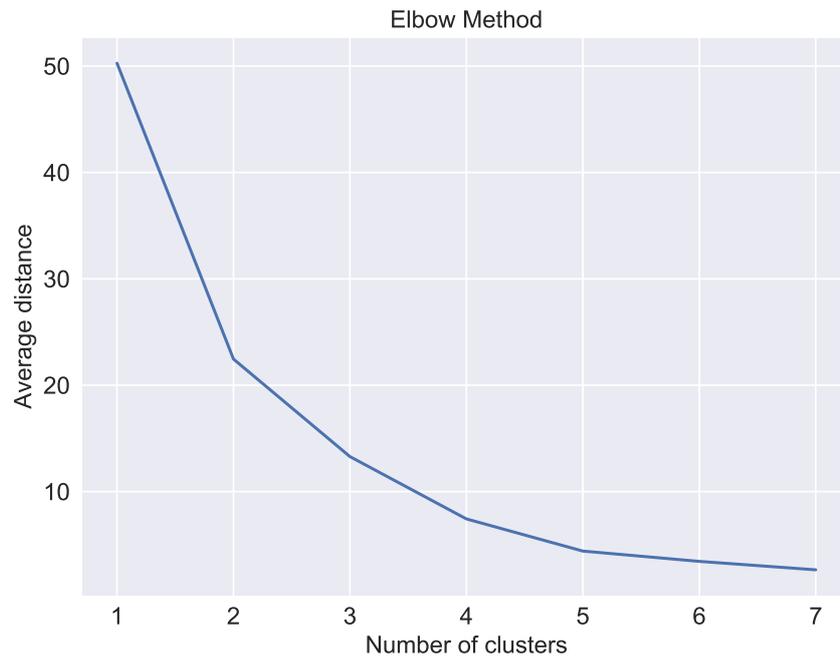
Figure. 3.4 Graphical representation of the elbow method

3.5   Ordinary Kriging Method (OK)

Geostatistical interpolation of rainfall using Ordinary Kriging was selected for this study based on the extensive literature review, summarised in section 2.4.5.The geostatistical work was done using Python v3.6 in two steps, which are: structural analysis and spatial estimation.

3.5.1   Structural Analysis

3.5.1.1   Estimating the Semivariogram

Ordinary Kriging is based on real data and it considers the spatial structure of the rainfall (Pegram et al., 2006). The spatial correlation between rainfall observations was determined using the semivariogram. This is mainly because of robustness of the semivariogram and the ability it has to filter the influence of the mean which varies spatially (Cressie and Wikle, 2015). In this application, the spatial correlation was assumed to be isotropic, hence the semivariogram was estimated considering the magnitude of the distance vector $|h|$ between rain gauge locations. The classical semivariogram in Equation 2.3 is influenced by uncharacteristic observations. These observations have a negative effect on the semivariogram function since they are repeatedly used during computation at different lag intervals (Cressie and Wikle, 2015). Therefore, a "robust" estimation of the semivariogram, as suggested by Cressie and Hawkins (1980), was used to fit

the normally transformed data as shown in Figure 3.4, which is defined as:

$$2\hat{\gamma}(h) = \frac{\left\{ \frac{1}{|N(h)|} \sum_{N(h)} |z_i - z_j|^{\frac{1}{2}} \right\}^4}{0.457 + \frac{0.494}{|N(h)|}} \tag{3.3}$$

Equation 3.3 reduces the effect of uncharacteristic observations and it is most suited for contaminated normal distributions (Lark, 2000). By altering the lags (separation distances) between daily gauge rainfall observations, a well-ordered set of semivariances were obtained and used to construct an experimental semivariogram. The gauge rainfall data pairs were pooled into bins of average lag distances since the rain gauges are irregularly distributed. This was done by selecting a set of lags (h) at random constant increments such that each rainfall data pair that falls in a bin can be used to estimate $\gamma(h)$. The bin interval was chosen as the average lags between nearest neighbours. This effectively weights the small distances more highly when determining the semivariogram, as proposed by Kitanidis (1997).

3.5.1.2   Modelling the Semivariogram

A number of methods for fitting the semivariogram have been proposed, for example the maximum likelihood, least squares and visual assesment method (Lloyd, 2005; Webster and Oliver, 2007 and Moral, 2010). However, there is in no fitting method which can be defined as the best because of the absence of vigorous tests for determining the goodness character of models. The semivariogram in this study was fitted using visual inspection and the following Weighted Least Squares (WLS) criterion such that differences between the experimental $\hat{\gamma}(h_k)$ (h_k being the ordered sequence of all K distinct distances) and theoretical model $\gamma(h_k)$ semivariogram values are minimum:

$$WLS = \sum_{k=1}^{K} w(h_k)[\hat{\gamma}(h_k) - \gamma(h_k)]^2 \tag{3.4}$$

The weights ($w(h_k)$) were determined as $N(h_k)/[\gamma(h_k)]^2$ to give more weight to the first lags and the ones calculated using more data pairs.

The first stage in modelling the semivariogram was to plot the the experimental semivariogram. Then a model with the right shape and detail to characterize the structure of the experimental variogram was selected from three theoretical models namely: the exponential, Gaussian and spherical models introduced in Section 2.4.3.

- Exponential model

$$\gamma(h) = p. \left(1 - \exp\left(\frac{-3r}{d} \right) \right) + n \tag{3.5}$$

- Gaussian model

$$\gamma(h) = p. \left(1 - \exp\left(\frac{-3d^2}{r^2} \right) \right) + n \tag{3.6}$$

- Spherical model

$$\gamma(h) = \begin{cases} p.(1.5(\frac{d}{r}) - 0.5(\frac{d}{r})^3) + n & d \leq r \\ p + n & d > r \end{cases} \tag{3.7}$$

where d is the distance value used to compute the semivariogram, p is the sill and r represents the range. The three models are demonstrated in Figure 3.5, without a nugget n, all with the same parameters.
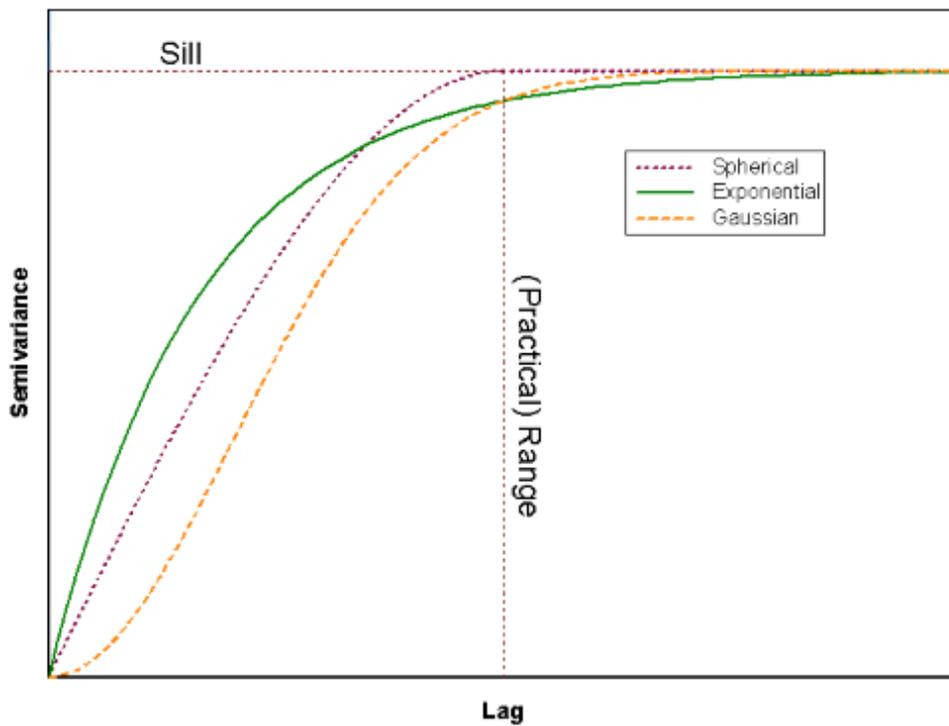
Figure. 3.5 Representation of the variogram fitting models all with the same parameters p, r, d and n (Falah et al., 2017).

The theoretical models are then fitted by minimizing the squares between values on the experimental semivariogram and the fitted model. Lastly, the fitted theoretical model and experimental variogram are plotted on the same pair of axes for visual inspection. Most interpolation studies have been done on monthly and annual temporal scales. Typically, these adopt only a single model to fit the semivariogram (Buytaert et al., 2006; Moral, 2010 and Cheng et al., 2017). The focus in this study is the interpolation of daily accumulations, therefore different theoretical models were experimented with, because of the high variability of rainfall on a daily scale. The semivariogram model which produced best Kriging results was chosen as appropriate for representing the rainfall structure on a particular day. The objective was to obtain a theoretical model that adequately captures the spatial structure of the experimental semivariogram and reduces the computational time.

3.5.2 Spatial Estimation

The Ordinary Kriging equation estimates a rainfall value at an ungauged location using a weighted linear combination of the observed neighbouring rainfall measurements using equation 2.8 in section 2.4.4.3.

To remove bias, the OK error $\left[\hat{Z}_{ok}(x_0) - Z(x_0)\right]$ is calculated subject to the weights, thus the error is given as:

$$r = \sum \lambda_i z(x_i) - Z(x_0) \tag{3.8}$$

where r, referred to as the bias, is the Kriging error at a specific location, $\sum \lambda_i z(x_i)$ is the estimated rainfall value and $Z(x_0)$ is the actual observed rainfall value. According to Wackernagel (2014), and Cressie and Wikle (2015) the resulting error when predicting unknown variables is considered as an outcome of a random variable. We assume that the expected error at any specific location in a rainfall field is zero. The expected error can be defined as:

$$E_r = E\left[\sum \lambda_i z(x_i) - Z(x_0)\right] \tag{3.9}$$

Hence, equating the expected error (E_r) to zero introduces a condition of unbiasedness as follows:

$$E_r = E\left[\sum \lambda_i z(x_i) - Z(x_0)\right] = 0 \tag{3.10}$$

$$E\sum \lambda_i.z(x_i) = EZ(x_0) \tag{3.11}$$

This can be simplified as:

$$\sum_{i=1}^{n} \lambda_i = 1 \tag{3.12}$$

with the variance of the estimation given by:

$$\sigma^2 = E\left[\{\hat{Z}_{ok}(x_0) - Z(x_0)\}^2\right] = 2\sum_{i=1}^{n} \lambda_i \gamma(x_i, x_0) - \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j \gamma(x_i, x_j) \tag{3.13}$$

where $\gamma(x_i, x_j)$ is the semivariance between rain gauge observations at points x_i and x_j, and $\gamma(x_i, x_0)$ represents semivariance between the estimation point x_0 and i^{th} rain gauge observation point. The rain gauge observations in this study are located in unique 1 km grid cells such that the rainfall recorded by each rain gauge at a point is used to represent the rain falling in that cell. Taking this into consideration, the Kriging variance was calculated for each grid cell.

3.5.2.1 Introducing the Lagrange multiplier

A Lagrange multiplier is introduced into the Kriging equation based on the constraint that the sum of Kriging weights is equal to 1 (Cressie and Wikle, 2015). The Lagrange multiplier minimizes the estimation variance. This is achieved by equating the weight constraint to zero

and then multiplying it by the Lagrange multiplier μ as follows:

$$\mu \left\{ \sum_{i=1}^{n} \lambda_i^{ok} - 1 \right\} = 0 \tag{3.14}$$

Equation 3.14 is then combined with the Kriging variance to be minimized to obtain an auxiliary function:

$$F(\lambda_i, \mu) = \sigma^2 - \sum_{i=1}^{n} \gamma(x_i, x_0) - 1 \tag{3.15}$$

This auxiliary function is partially differentiated in terms of the weights λ_i and the Lagrange multiplier μ:

$$\begin{aligned} \frac{\partial(\lambda_i, \mu)}{\partial \lambda_i} &= 0, \\ \frac{\partial(\lambda_i, \mu)}{\partial \mu} &= 0, \qquad for \quad i = 1, 2, \ldots, n \end{aligned} \tag{3.16}$$

Differentiating with respect to the Lagrange multiplier gives:

$$\sum_{i=1}^{n} \lambda_i^{ok} = 1 \tag{3.17}$$

The Kriging linear system then becomes:

$$\sum_{i=1}^{n} \lambda_i^{ok} \gamma(x_i, x_j) + \mu = \gamma(x_j, x_0) \tag{3.18}$$

Equation 3.17 and Equation 3.18 form the Ordinary Kriging system. Solving this system gives the input weights of Equation **??** and the variance is obtained as follows:

$$\sigma_{ok}^2(x_0) = \sum_{i=1}^{n} \lambda_i \gamma(x_i, x_0) + \mu(x_0) \tag{3.19}$$

Equation 3.19 measures the quality of the estimation and depends on the quality of the semivariogram. The system of OK equations can be combined in matrix form as:

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \mu(x_0) \end{bmatrix} = \begin{bmatrix} \gamma(x_1,x_1) & \gamma(x_1,x_2) & \ldots & \gamma(x_1,x_n) & 1 \\ \gamma(x_2,x_1) & \gamma(x_2,x_2) & \ldots & \gamma(x_2,x_n) & 1 \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ \gamma(x_n,x_1) & \gamma(x_n,x_2) & \ldots & \gamma(x_n,x_n) & 1 \\ 1 & 1 & \ldots & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} \gamma(x_1,x_0) \\ \gamma(x_2,x_0) \\ \vdots \\ \gamma(x_n,x_0) \\ 1 \end{bmatrix} \tag{3.20}$$

This augmented matrix can be simply denoted as:

$$W_{ok}.\lambda_{ok} = b_{ok} \tag{3.21}$$

where the first vector contains the semivariogram values between the observed gauge observations and the second vector contains the Kriging weights in which μ is the Lagrange multiplier.

The OK weights are therefore:

$$\lambda_{ok} = W_{ok}^{-1}.b_{ok} \tag{3.22}$$

The resulting Kriging weights account for the structural relationship between the observed data and the estimation point through the experimental semivariogram. In general the observed rainfall value that is close to the estimated point receives a larger weight. The factors affecting the Kriging weights are summarized below:

- spatial correlation between the available rain gauge observations.

- spatial positioning of the available rain gauge observations relative to the estimated point.

- spatial continuity and structure of rainfall in the study region, this is characterized by the semivariogram.

3.5.3   Kriging Performance Measures

Daily rainfall accumulations were used to asses the performance of the OK method. The daily interpolated rainfall field was verified using cross-validation. Cross-validation (CV) is widely used in the analysis of rainfall interpolation methods (McKee, 2015). It is mostly used in geostatistical studies where an independent rain gauge network does not exist to asses the rainfall estimates (Webster and Oliver, 2007). The CV procedure used in this context involves excluding one rain gauge observation for validation from the rainfall data used to construct the semivariogram and Kriging estimation. The rainfall data is divided into two subsets, a validation subset which contains the excluded rain gauge observation only and a training subset which contains the rest of the rainfall data. This type of CV is referred to as leave one out cross-validation. The test error which is based on a single observation is given by:

$$MSE_i = (y_i - \hat{y}_i)^2 \tag{3.23}$$

where y_i is the actual observed rainfall value and $\hat{y}_i$ is the predicted rainfall value at the validation point. The test error is calculated for each observed rainfall point i in the rainfall dataset and then averaged to yield a mean CV error defined by:

$$CV_{(n)} = \frac{1}{n}\sum_{i}^{n} MSE_i \tag{3.24}$$

where n is the number of observed rainfall data. During cross-validation an assumption is made that the excluded rain gauge measurement is the actual rainfall amount at that location.

3.6   Comparison statistics

Comparison statistics are typically used to measure the accuracy of the rainfall amount attained from different sensors. This section describes the statistics commonly used to validate radar and gauge derived rainfall products. The statistics were computed using 24-hr rainfall accumulations

from the rain gauges and the data observed by the MRL5 radar. A rainfall day was considered for assessment if there were more than 10 pairs of radar and rain gauge data available. Besides, only rain gauges which recorded values above 2 *mm* of rainfall were selected.

3.6.1 Mean Absolute Error (MAE) and Root Mean Square (RMSE)

Mean absolute error (MAE) gives a measure of the average error magnitude. It is given by:

$$MAE = \frac{\sum_{i=1}^{N} |R_i - G_i|}{N} \tag{3.25}$$

where R_i is the radar rainfall value at location i, G_i is kriged rainfall value at location i and N represents the number of rain gauges used. However, MAE tends to be less sensitive to large errors so it was used as the first coarse error statistic. RMSE also measures the average magnitude of the error but gives greater weight to large errors. It is defined as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (R_i - G_i)^2}{N}} \tag{3.26}$$

3.6.2 Correlation Coefficient (R)

The correlation coefficient is typically used to compare rainfall estimates measured by different sensors or methods. The correlation coefficient shows the strength and direction of a linear relationship between the observed gauge rainfall values and radar estimates. The most used correlation coefficient in radar-gauge comparison studies is the Pearson correlation coefficient which is defined as:

$$R = \frac{\sum_{i=1}^{n} (G_i - \overline{G}).(R_i - \overline{R})}{\sqrt{\sum_{i=1}^{n} (G_i - \overline{G})^2 . \sum_{i=1}^{n} (R_i - \overline{R})^2}} \tag{3.27}$$

where $\overline{G}$ is the average rain gauge rainfall accumulation and $\overline{R}$ represents the average radar rainfall accumulation in a particular image. The R values ranges between -1 and $+1$; an R value above zero shows that a positive linear correlation exists, an R value of 0 shows that no correlation exists and an R value below zero indicates that a negative linear correlation exists.

3.7 Summary

This chapter provided a description of the data processing of daily rainfall attained from both the radar and rain gauges and described in detail the spatial distribution of gauge rainfall measurements. The beginning of the chapter elaborated methods for preliminary analysis of rainfall data that assist in developing an interpolation model as well as proper interpretation of the modelling results. This analysis is based on classical descriptive statistics which include simple histogram and box plots. The section that followed described the Ordinary Kriging technique used to spatially distribute gauge measurements to match the radar rainfall field. This Kriging technique considers the spatial structure of a rain gauge network and assigns

relative weights to each individual gauge according to the position of the gauge in the network. Lastly three common statistics used in validation of precipitation studies (MAE, RMSE and the correlation coefficient R) were described in this chapter.

Chapter 4

Results and Discussion of Statistical Methods of Interpolating Gauge and Radar Rainfall Information

4.1   Introduction

This chapter presents the key findings of this study, putting into practice methodologies described above. It analyses the daily rainfall accumulations derived from the radar, rain gauges and a kriged rain gauge network. This analysis is focused on factors which have been shown to have an influence on spatial rainfall estimation as described in Chapter 2. These factors include rain gauge density, range dependency, storm type and seasonal variation. To achieve this, the following objectives are addressed:

(i)   analyse the rain gauge rainfall data using visual and quantitative methods,

(ii)   identify the similarities and differences between radar and rain gauge rainfall fields, and

(iii)   evaluate the effect of gauge density, storm type, range dependence and seasonal variations on the estimated rainfall accumulations.

4.2   Exploratory Data Analysis

This section analyses the rain gauge rainfall data with the aim of understanding and summarizing the dataset. The analysis is important because it facilitates the development of a suitable interpolation method and the interpretation of rainfall characteristics of the study region. The study area contains 18 gauge stations measuring the rainfall depth in mm at specified longitude and latitude positions. The gauge stations are distributed over a 7000 km^2 grid. The gauge stations are separated by 30 km on average, with the closest distance between gauge stations being 1 km. The rain gauge stations are distributed as shown in Figure 4.1. The gauges appear to be evenly spaced throughout the study region except towards the south-east part of the region.
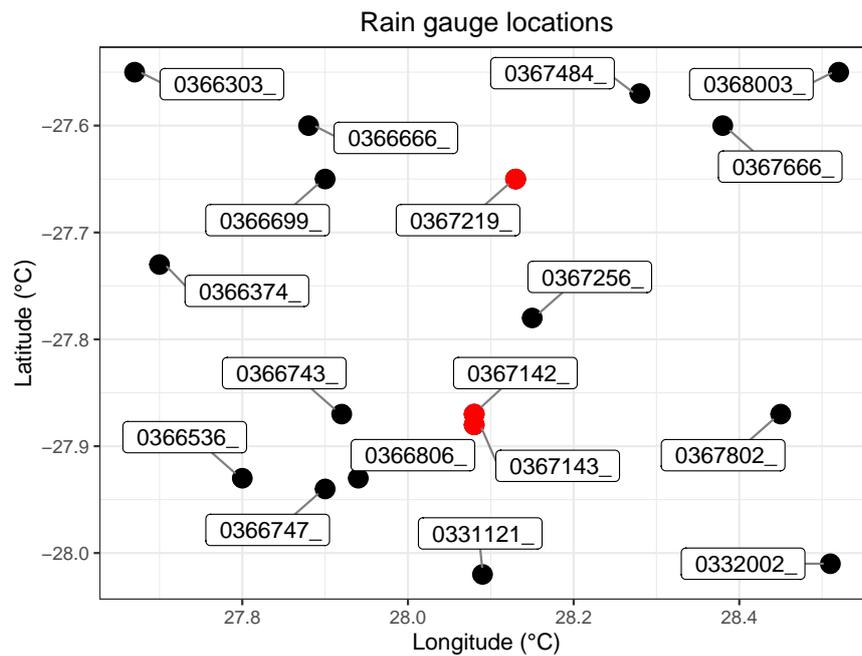
Figure. 4.1 Location of the rain gauge stations and their respective gauge IDs, the red coloured stations are discussed after Figure 4.5.

Table 4.1 Gauge ids with their corresponding station names

Gauge id	Staion name
0331121	LE LONG
0332002	VISGAT
0366303	EDENVILLE
0366374	DOORNKLOOF
0366536	VAN TONDERTST
0366666	EKEINHOFF
0366699	UITZICHT
0366743	LINDLEY MUN
0366747	ERFDEEL
0366806	BRAKDAM
0367142	RAFIDUM
0367143	NEWTON GRANGE
0367219	PETRUS STEYMUN
0367256	SLAAPPLAATS
0367484	ERFENIS
0367666	JACOBSDAL
0367802	TERTUIS
0368003	TWEEKING POL

The descriptive statistics for rainfall depths in the study region between 1994-2003 are shown in Table 4.2, which is comprised of descriptive statistical concepts (measures of central tendency, measures of central variability and measures of symmetry) introduced in Section 3.4.1. The descriptive statistics were calculated from each gauge station, identified by a station identifier

(gauge id) for all gauge readings above 3mm as described in Section 3.4.1. From the table it can be concluded that rainfall data from all the gauge stations are positively skewed, that is, their means are higher than the median values.

Table 4.2 Summary statistics of the 18 study rain gauges data

Gauge_id	Number of daily readings	Altitude (m)	Mean (mm)	Standard deviation	Min (mm)	25% (mm)	50% (mm)	75% (mm)	Max (mm)	Skew
0331121	523	1666	12.3	10.2	3.0	5.0	9.0	16.0	75.0	2.1
0332002	474	1691	15.2	12.4	3.0	6.5	11.0	20.0	86.0	2.0
0366303	400	1509	14.5	12.9	3.0	6.0	10.5	17.6	85.0	2.3
0366374	540	1515	12.8	10.6	3.0	5.0	10.0	16.0	85.0	2.4
0366536	527	1760	14.5	12.3	3.0	6.0	10.0	20.0	100.0	2.3
0366666	475	1539	13.5	11.2	3.0	6.0	10.0	17.5	90.0	2.4
0366699	487	1584	13.5	11.8	3.0	6.0	10.0	16.0	76.0	2.5
0366743	438	1540	15.1	13.1	3.0	6.0	11.0	20.0	84.4	2.1
0366747	492	1584	13.8	11.3	3.0	6.0	10.3	18.0	85.0	2.4
0366806	481	1580	13.7	11.6	3.0	6.0	10.0	18.0	98.0	2.6
0367142	363	1691	12.9	11.5	3.0	5.0	9.1	16.0	80.0	2.3
0367143	318	1710	14.1	13.0	3.0	5.6	9.0	19.4	85.0	2.3
0367219	330	1700	15.9	15.6	3.0	6.1	11.1	20.0	137.5	3.6
0367256	567	1656	12.6	11.2	3.0	5.5	9.0	17.0	105.0	2.9
0367484	513	1705	14.1	11.5	3.0	6.1	10.5	18.0	92.0	2.3
0367666	511	1650	13.6	11.9	3.0	5.7	9.8	17.5	74.0	2.3
0367802	522	1698	12.7	11.0	3.0	5.2	9.5	16.5	110.0	2.8
0368003	392	1424	12.9	10.9	3.0	5.9	10.0	16.0	80.0	2.6

Figure 4.2 shows a histogram combined with a rug plot of the rainfall data above 3 mm per day during the period 1993-2004. The histogram describes the rainfall distribution pooled into 10 mm bins and the rug plot describes each daily rainfall reading displayed as marks along the x-axis. The histogram of the rainfall data is positively skewed as shown in Figure 4.2, since it has a long tail at the right with a few days receiving high rainfall amounts. Most of the daily readings record low rainfall amounts showing that there is a high frequency of low rainfall in Bethlehem, which is to be expected. The histogram in Figure 4.2 was decomposed into plots for each individual rain gauge stations to represent the distribution of rainfall received at each gauge station. The resulting plots are shown in Figure 4.3 and Figure 4.4. All the histograms show that the gauge stations have their highest frequency at the lowest daily rainfall amounts with gauge 0366536, 0331121, 0367256 and 0367802 having the highest number of daily readings.
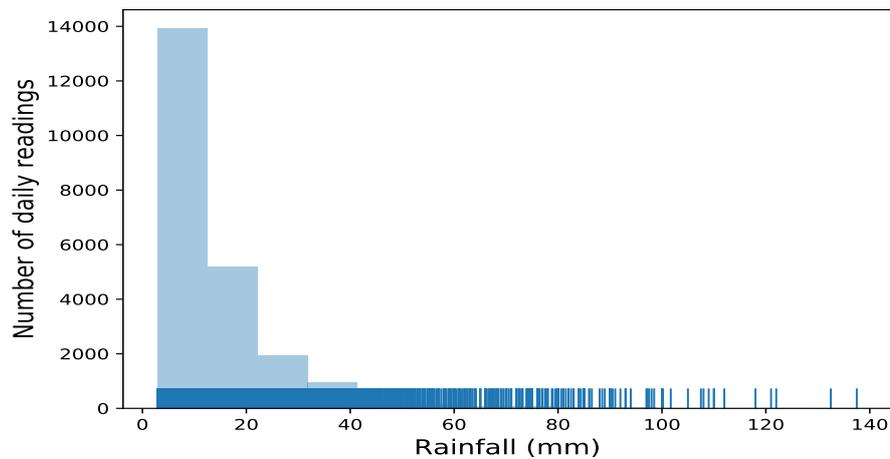
Figure. 4.2 Histogram and rug plot representing the daily rainfall distribution from 1994 to 2003. Only wet days above 3 mm depth are plotted here and in Figures 4.3 and 4.4.
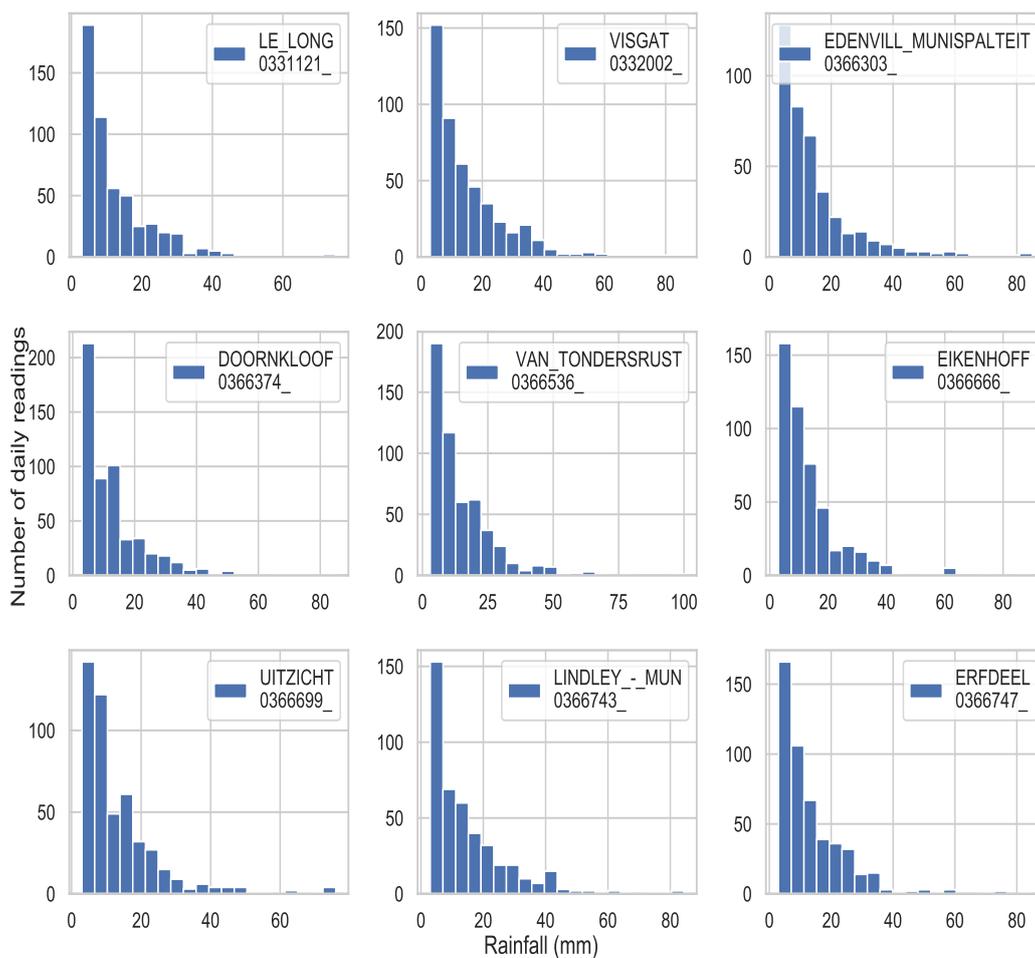
Figure. 4.3 Histogram plots for each individual rain gauge station showing distribution of rainfall during the study period.
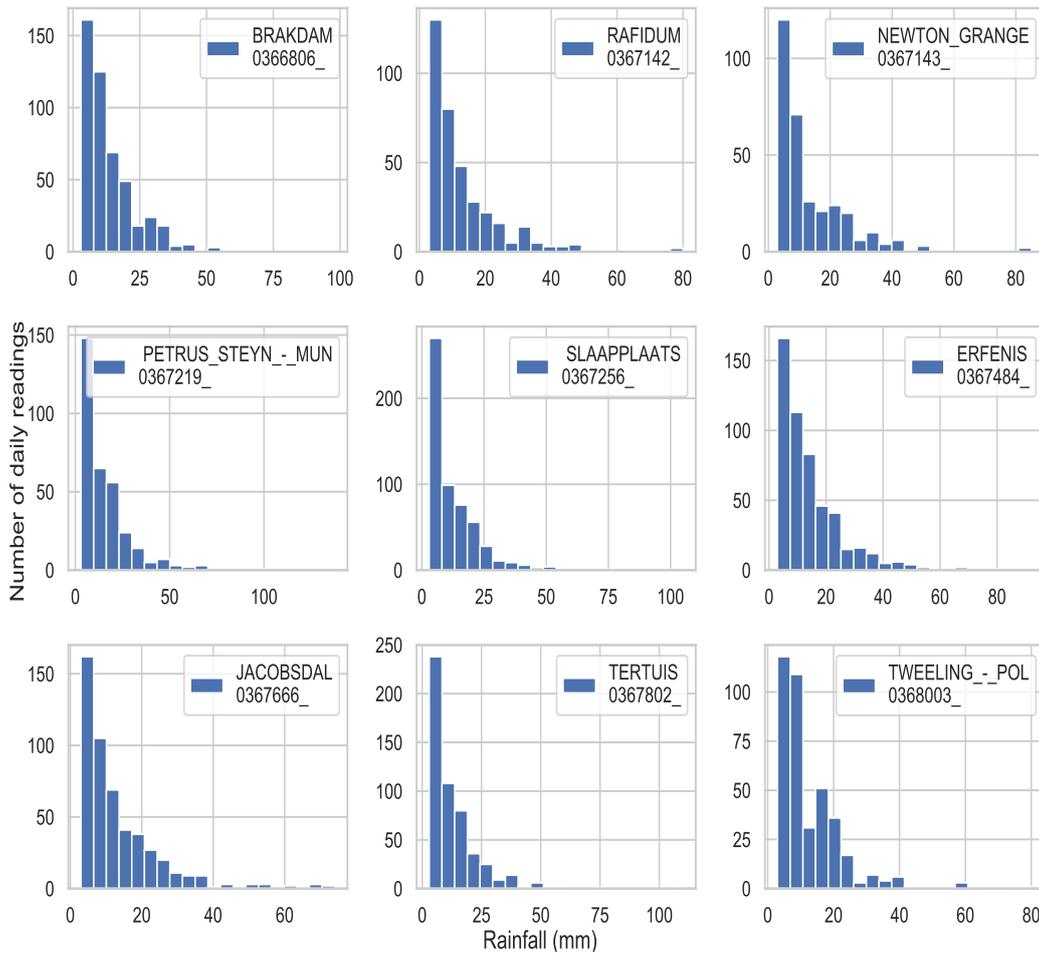
Figure. 4.4 Histogram plots for each individual rain gauge station showing distribution of rainfall during the study period.

At some stations (0367484, 0367142, 0367256 and 0366303) there is a smooth continuous decrease from low to high rainfall intensities. These histograms exhibit rainfall amounts in every rainfall bin over the range of data, this might correspond to gauge stations in humid locations (ŞEN and Eljadid, 1999). The rain gauge stations with the highest daily readings are located at relatively higher altitude compared to others. Figure 4.5 shows the variation of the number of daily readings with altitude, a relative increase in daily readings as the altitude increases is evident except for three stations (0367142, 0367219 and 0367143) which exhibit very low rainfall at higher altitudes whose positions are shown in red in Figure 4.1.
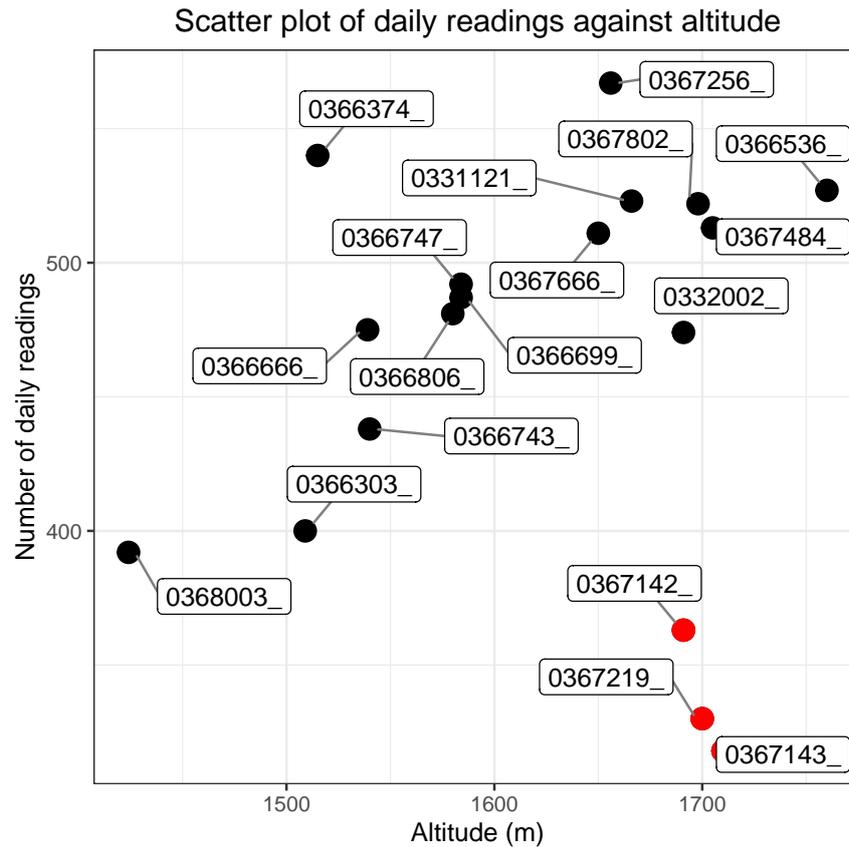
Figure. 4.5 The graph shows the variation of the number of daily readings with altitude. The red coloured stations were given in Figure 4.1 and are highlighted in Figure 4.6.

In Figure 4.6 the rain gauge stations are colour coded with respect to the total amount of rainfall per gauge station for the whole study period, to try and check if there is a clear distinction between areas of high and low rainfall intensities. It appears that there is a mixture of gauge stations with high and low rainfall totals throughout the study area. There is however a sparse group of gauge stations with which exhibits high rainfall totals at the east-side of the grid. For a 10-year period, the 1 day design rainfalls north of the MRL5 radar range from > 75 mm to < 100 mm.

Figure 4.7 and Figure 4.8 shows box plots describing the monthly rainfall distribution and variation for each individual gauge station throughout the whole study period (1994-2003). The results show that throughout the entire 7000 km^2 area, January and February receive the highest monthly average rainfall as well as the highest number of very heavy rainfall days.. The highest rainfall amount recorded within the study area by a single gauge station (0367219) was 137.5 mm on 11 February 1992. The threshold of heavy rainfall events depend on the geographical location. Extreme rainfall events are often defined using data from individual gauge stations.
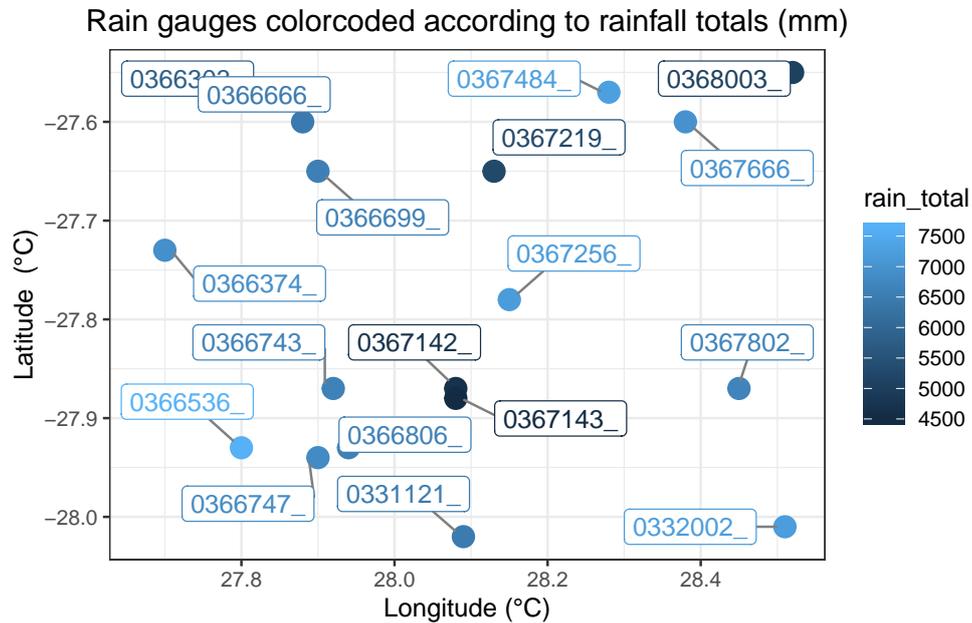
Figure. 4.6 Map showing the distribution of rainfall stations and their varying rainfall totals.

In Taiwan, Chen et al. (2007) define a heavy rainfall event if 1 or more rain gauge stations record more than 50 mm of rainfall and a very heavy rainfall when an excess of 150 mm of rain is received at a single station. However the, SAWS weather forecasters issue out warnings for heavy rainfall when rain above 50 mm is expected at any gauge station and a very heavy rainfall when any one gauge station records more than 115 mm (Rae, 2008). Throughout the study period, rainfall above 115 mm at any of the 18 gauge stations was received just once. The daily rainfall over Bethlehem is highly variable and exhibits a strong summer and winter seasonal trend. On average 75% of rainfall is received in the summer (October-March) and extreme rainfall events capable of generating high stream flows are likely to occur during this season. The summer season of 1995/1996 had the highest monthly average rainfall (134 mm) throughout the study period which corresponds to the findings of Dyson (2009) and the 2002/2003 season was the driest summer season with average monthly rainfall of 63 mm.
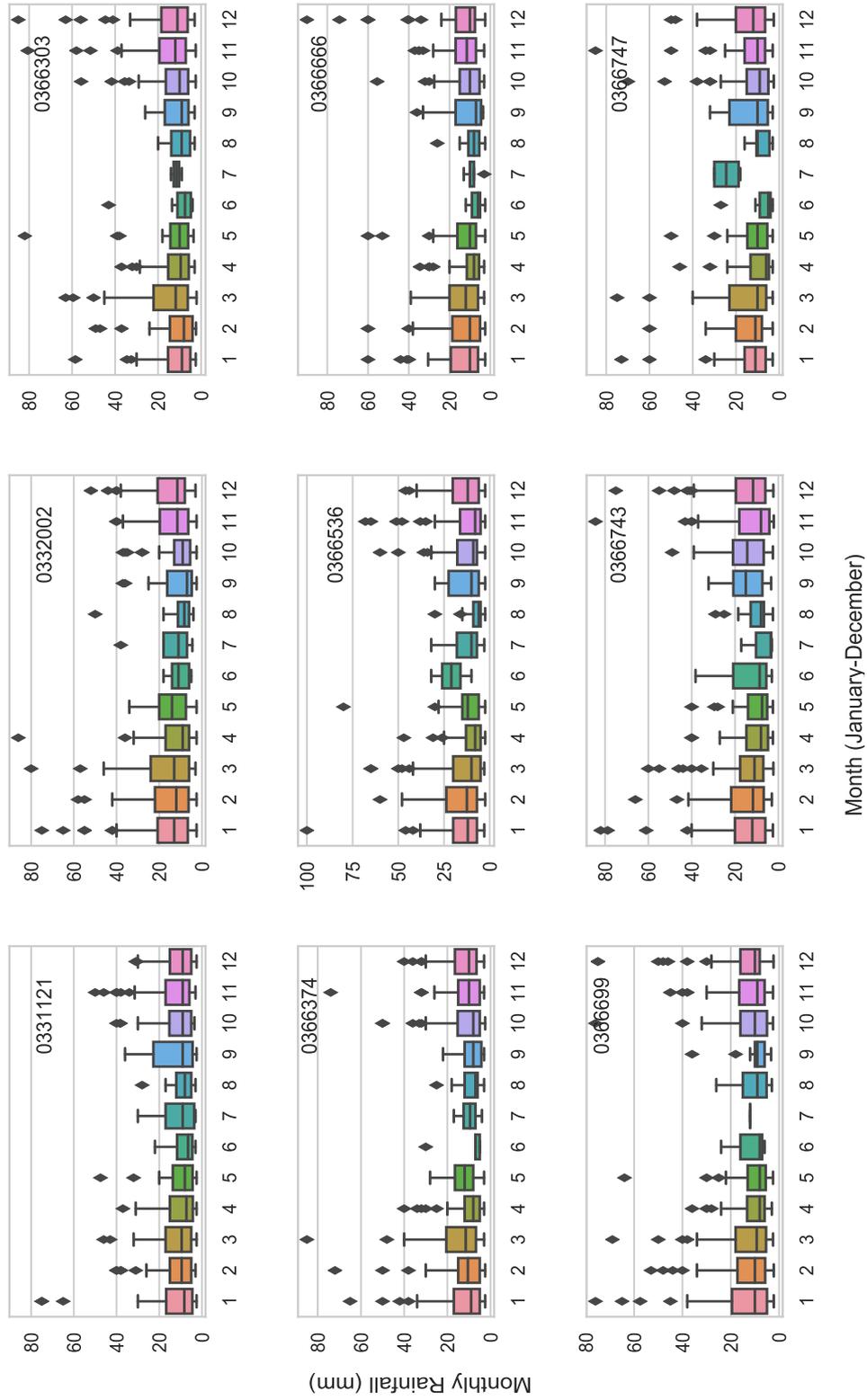
Figure. 4.7 Box plots for each individual rain gauge station showing the monthly distribution of rainfall above 3 mm per day based on information from Table 4.1. Note: Each box plot is color-coded according to month
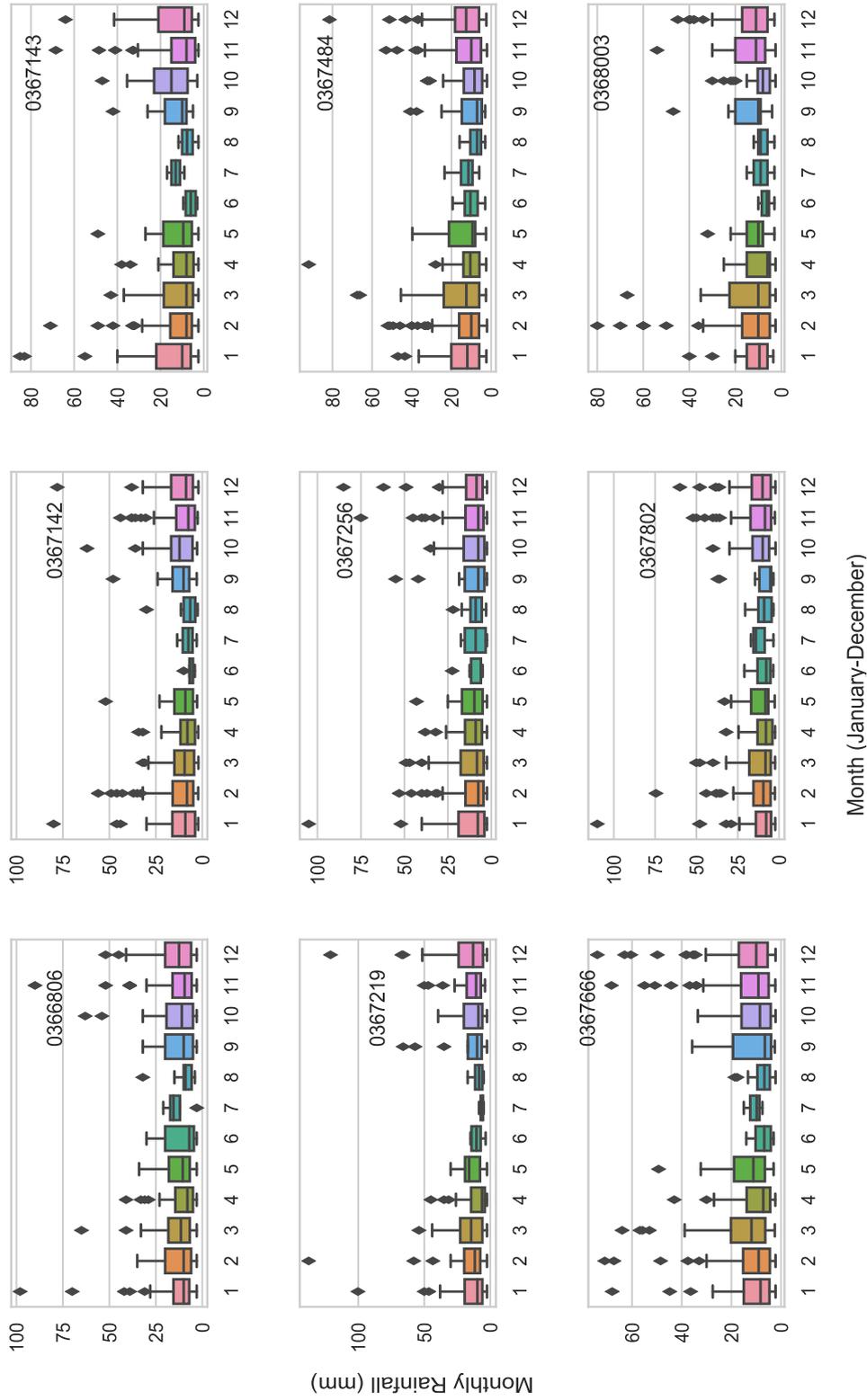
Figure. 4.8 Box plots for each individual rain gauge station showing the monthly distribution of rainfall above 3 mm per day based on information from Table 4.1. Note: Note: Each box plot is color-coded according to month

4.2.1 Probability and Clustering Analysis

The normality of the rainfall data in this study was assessed using a normal Q-Q plot. The Q-Q plot allows us to check if the data follows a normal distribution, if not, how the assumption of normality is violated and which of the data contribute to the violation. The normal Q-Q plots in Figure 4.9 were created by plotting the actual quantiles of the rainfall data from all 18 stations against theoretical quantiles (z-scores). If both sets of quantiles came from a normal distribution, the plotted points should form a line that is roughly straight (Ford, 2015). A straight line was drawn in both Q-Q plots (left panel and right panel) to check the extent to which the points deviate from normality. The results from Figure 4.9 (left panel) show that the rain gauge data follows a non-normal distribution. The rainfall data flies off at both ends of the straight reference line because it is skewed. The data was log transformed to remove skewness and to meet the normality requirements for Kriging estimation. Figure 4.9 (right panel) shows a Q-Q plot of the transformed data, there still exists a deviation at the lower end due to the dominance of small rainfall values in the rainfall dataset.
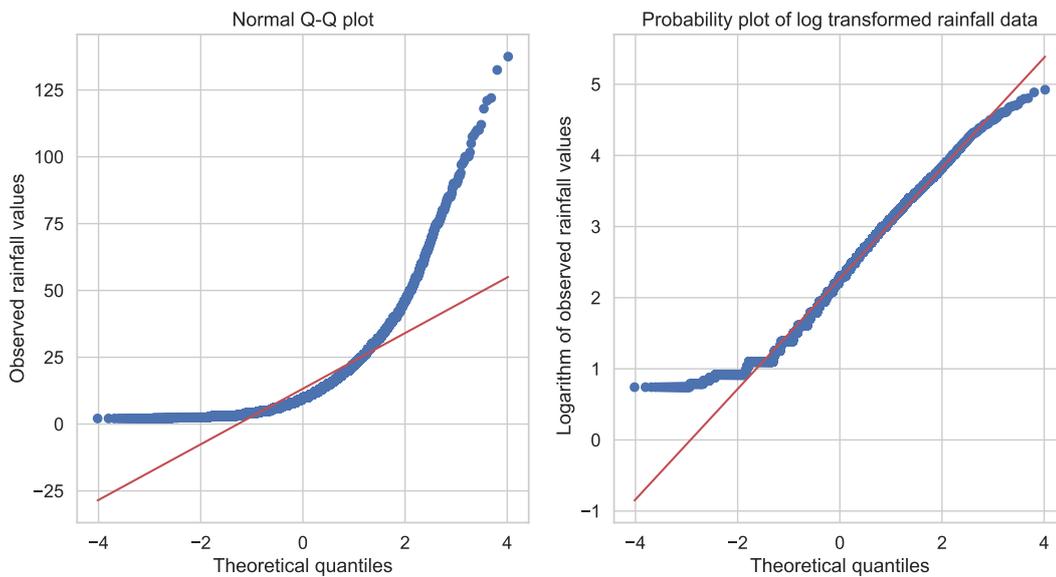
Figure. 4.9 Quantile-quantile plots of skewed and log-transformed rainfall data for all rain gauge stations.

Log-normal Kriging was carried out since the original rainfall data was log-transformed to assume normality. This means that the log-transformed rainfall values were used for the Kriging estimation, hence, the resulting estimated rainfall values were in logarithms. However, most hydrological applications require rainfall estimates expressed in original units (mm), therefore, the logarithmic results were back-transformed using the following equation:

$$\hat{Z}_{ok}(x_0) = exp\{\hat{Y}_{ok}(x_0) + \sigma_{ok}^2(x_0)/2 - \mu(x_0)\} \qquad (4.1)$$

where $\hat{Y}_{ok}$ is the kriged estimate of the natural logarithm at location (x_0), σ_{ok}^2 is the corresponding Kriging variance at (x_0) and μ is the Lagrange multiplier.

One of the most common exploratory data analysis methods is clustering. Clustering provides an understanding about the statistical relationships within a dataset. Clustering can be simply defined as a method of identifying and grouping similar data points such that data points in the same group (cluster) are very similar, while data points in different groups are very different according to a chosen set of criteria. Clustering analysis was carried out between two features (mean and standard deviation) to investigate if there exists a population of statistical oversampling, hence, determining if the Kriging estimation was to be applied within each cluster or not. Chen et al. (2018) show that a clustering analysis of gauge stations during the application of interpolation methods in mountainous areas enhances the accuracy of the resulting rainfall estimates. K-means clustering described in Section 3.4.2 was used to analyze and group rain gauge stations in this study according to their average mean and standard deviation. Figure 4.10 (left panel) indicates the results of the elbow method used to select the initial number of clusters. The results show that the gauge stations can be divided into two initial clusters , which subsequently matched the final number of clusters in which the gauge stations can be grouped. The clustering results in Figure 4.10 (right panel) show that there is no and a wide variation in the behaviour of the two Cluster except for one station in Cluster 2 which has both a mean and standard deviation above 15 mm. Most of the rain gauge stations in Cluster 2 are located at higher altitudes compared to those in Cluster 1, which is the reason for them having higher means and standard deviations because of the orographic effects.
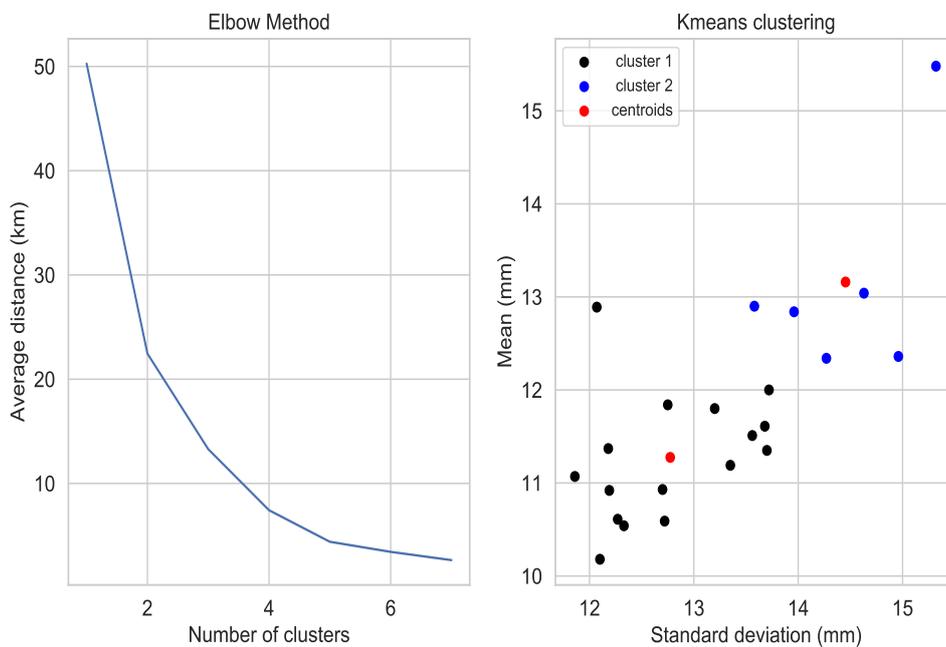
Figure. 4.10 The diagram shows the elbow method for selecting initial clusters of the k-means algorithm (left panel) and the final clusters of rain gauge stations obtained (right panel).

4.2.2   Summary of the Exploratory Data Analysis

The probability distribution of the rainfall in the Free State is non-normal. As expected, the rainfall data are positively skewed suggesting that there is a large number of data falling into small rainfall ranges. However no conclusions were made on the physical processes responsible for the rainfall distribution since it is not relevant for this study. The scatter plots of the rain gauge stations suggest that there is a mixture of low and high rain gauge observations throughout the study area. The rain gauges with the highest observation counts seem to be located at higher altitudes. A seasonal trend is also highlighted by the rainfall dataset with most of the rain falling in the summer season. The probability plots support the idea that the data do not follow a normal distribution and log-transformation of the dataset removes much of the skewness. In addition, the clustering analysis shows that there is not a wide variation in the behaviour of the clusters presented by the gauge rainfall data.

4.3   Study Period

The main aim of this study is to assess the difference between gauge, Kriged gauge and radar rainfall products. The 10 year period from 1994-2003 was selected for the assessment of the three rainfall products mentioned above. The reasonably homogeneous and large rainfall dataset will assist in selecting rainfall days with quality radar rainfall products to compare with the gauge products.

To put the study period in perspective, monthly averaged rainfall totals of all the 18 gauges were obtained from the 30-year (1981 to 2010) Climate Systems Analysis Group (CSAG) dataset including zeros. The monthly averaged rainfall totals from the 30-year CSAG dataset were compared with the monthly averaged totals of each of the 18 gauge stations during the 10-year study period. The monthly averaged data of each of the 18 gauge stations used in this study are listed in Table 4.3, and it can be seen that two gauges (0367142 and 0367143) record very low rainfall totals compared to the rest of the 16 gauges during the study period and these are at higher altitudes compared to the rest of the gauge stations (refer to Figure 4.5 (a) in Section 4.2.1). The lower rainfall totals challenge the assumption of increasing rainfall with altitude in the Free State, summer high veld region. This finding is supported by previous research studies elsewhere, that conclude altitude is not the only factor responsible for influencing rainfall in mountainous regions (prudhomme1998relationships,johansson2003influence,nel2005short). Prudhomme and Reed (1998) suggest complex topography and the sparsity of rainfall information available makes it difficult to find the relationship between rainfall and topography in the mountainous regions.

Table 4.3 Rain gauge stations listing the average total rainfall per month during the study period (1994-2003) compared to monthly averaged totals from the whole 30-year CSAG dataset.

Gauge id	Month												
	January	February	March	April	May	June	July	August	September	October	November	December	Annual total
0331121	101	81	97	41	23	7	7	9	16	20	65	88	645
0332002	107	99	104	53	33	11	7	14	24	56	83	126	720
0366003	75	67	99	43	36	8	2	13	14	58	71	93	505
0366374	116	81	102	50	34	5	4	15	12	68	76	121	690
0366536	115	110	115	45	39	8	21	13	26	80	84	104	763
0366666	91	84	93	46	37	6	4	9	17	55	76	121	643
0366699	96	87	87	42	33	6	2	12	13	66	83	126	656
0366743	103	85	105	40	25	8	2	16	17	71	75	112	660
0366747	97	91	97	43	31	6	9	9	19	68	83	125	679
0366806	93	84	91	40	31	7	5	12	18	64	78	129	658
0367142	73	66	46	36	26	3	2	6	15	53	60	83	467
0367143	75	65	57	29	24	2	2	3	13	51	55	71	448
0367219	81	84	65	37	28	7	1	11	24	38	50	92	523
0367256	114	93	104	54	36	9	3	15	24	62	88	122	717
0367484	110	111	103	51	33	6	5	13	20	61	84	130	725
0367666	92	95	104	46	34	3	3	17	23	59	93	133	697
0367802	85	96	86	29	34	7	3	15	20	64	98	128	665
0368003	57	93	61	25	20	5	1	10	19	41	78	92	505
CSAG average	114	92	78	49	25	10	11	17	31	73	92	101	671

4.3.1   Selection of Rainfall Events

The study period was limited between to 1994 and 2003 due to the limited availability of MRL5 radar data. Rainfall events were selected based on their magnitude and spatial distribution over the study area. A time-series and heat map of the rainfall data were used to asses the magnitude and distribution of the rainfall. The time-series and heat-map were produced using gauge data only because the MRL5 radar had periods of missing data during the span of the study. The time-series was used to identify the nature of rainfall presented by the gauge measurements and to identify trends in extreme rainfall events. Figure 4.11 shows the time-series plots of average weekly gauge rainfall data of all the gauges, including zeros, over the study area for each year from 1994 to 2003. The results in Figure 4.11 indicate high variability of rainfall as we move from January to December each year, with heavy weekly average rainfalls evident during the summer season (October-March). Figure 4.11 illustrates that little rain was received in 1994 and 1998, this corresponds to the droughts experienced during those years in South Africa. The droughts are suggested to be due to the El Nino cyclone experienced in Asia and Australia which leads to below average rainfall being received in many parts of Southern Africa (Masih et al., 2014; Monyela, 2017). Usually droughts happen in Southern Africa during the October-March summer season after the beginning of an El Nino event (Thomson et al., 2003).

To further investigate the time-series data, the time-series was decomposed to identify trends and seasonality.
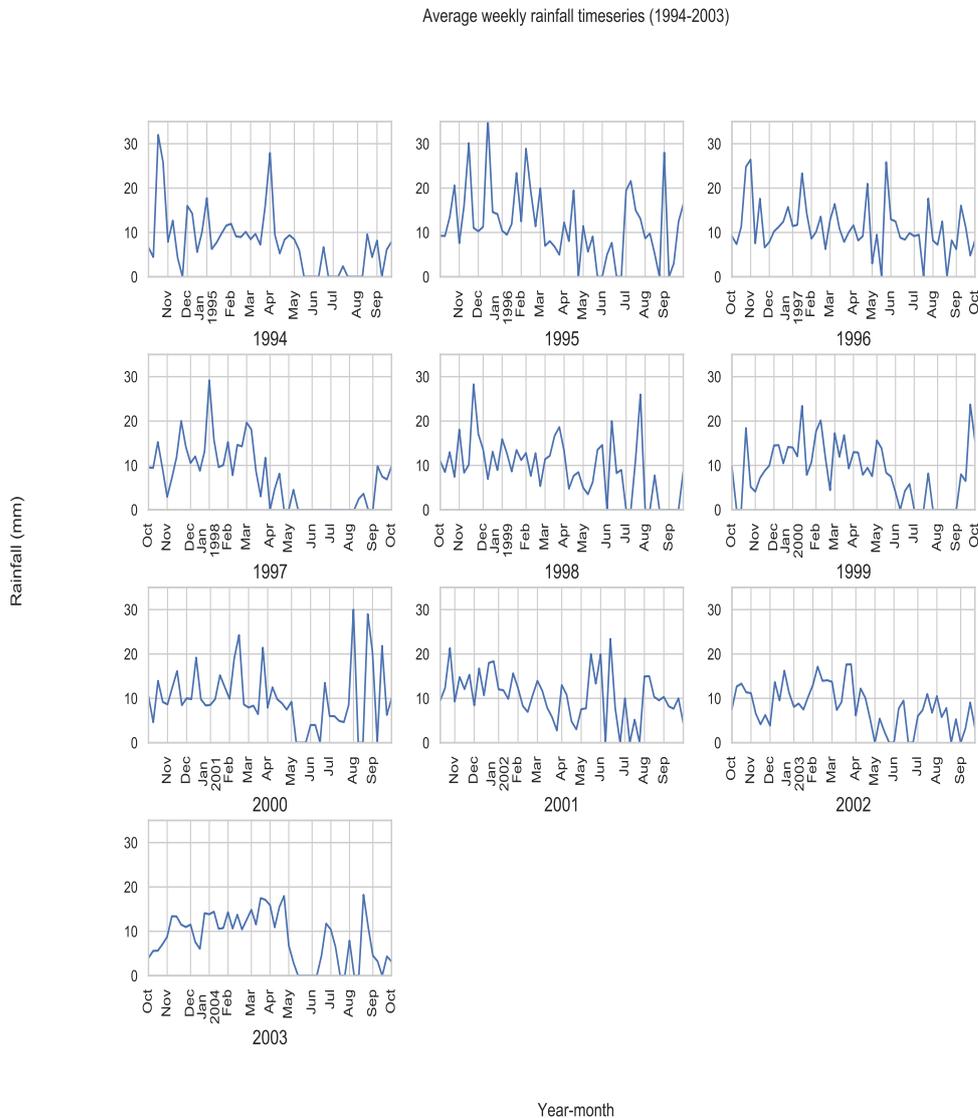
Figure. 4.11 Average weekly rainfall time-series of all the 18 gauges for each year between 1994 and 2003.

The time-series of the average weekly rainfall in Figure 4.11 indicates peaks in rainfall but are difficult to use for identifying the exact days with extreme rainfall. Therefore a heat-map was used to identify days with heavy daily rainfall magnitudes from all the 18 gauges for each calender year. Recently, heat-maps have been a common tool for representing complex statistical data. A heat-map can be defined as a data analysis software that uses colour as a data visualization tool. A heat-map was used to identify days with the highest daily rainfall totals in a visual way during the study period. Figure 4.12 shows the heat-map produced from the sum of daily rainfall totals from all the gauges over the study area. The magnitude of rainfall is represented by a dark (heavy rainfall) to light (low rainfall) color spectrum.
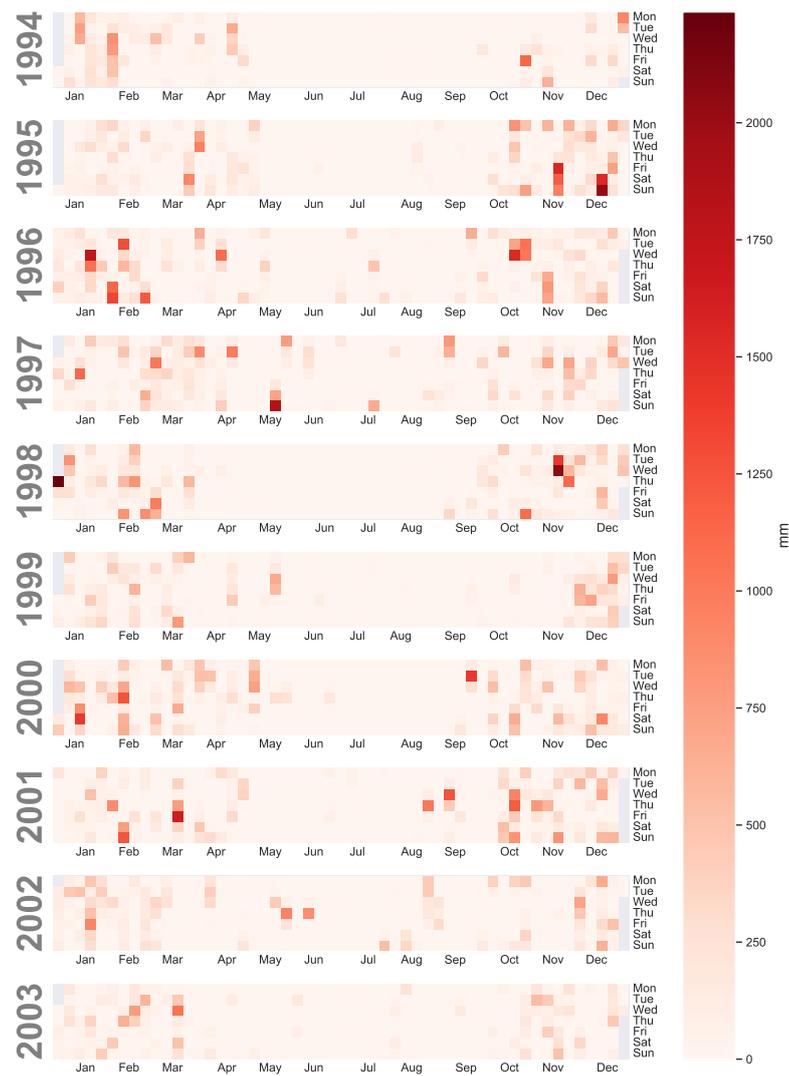
Figure. 4.12 Heat-map showing the sum of daily rainfall totals for the 10-year study period.

Only rainfall events that had more than 50% of the 18 gauge stations reporting rainfall in excess of 5 mm were selected for analysis since the purpose of the rainfall estimates in this study is targeted at flood forecasting applications. Traces of rainfall (less than 3 mm) have no detectable effect on stream-flows and furthermore, OK is sensitive to low values because it is difficult to estimate a robust semivariogram for low rainfall events (Soenario et al., 2010). However, the selected rainfall events were further limited because of missing MRL5 daily accumulations on some of the rainfall days which met the wet day criteria discussed above. Table 4.4 shows the 20 selected rainfall events for analysis. The table shows the date of the rainfall event, the number of gauge stations which received rainfall on that day or during that period and the sum of the total rainfall received by each of the gauge stations.

Table 4.4 Characteristics of selected rainfall events

Date	Number of gauge observations	Average gauged rainfall (mm)
17 December 1995	18	63.44
27 July 1997	15	16.73
16 February 1998	18	15.50
26 March 1998	15	16.87
24-26 October 1998	31	6.51
17-18 November 1998	40	36.78
15 January 2000	17	34.29
10 February 2000	18	24.88
4 April 2000	15	14.67
3 May 2000	20	23.55
19 September 2000	18	37.44
10 July 2001	9	11.67
12 September 2001	18	32.00
25 October 2001	18	33.17
30 May 2002	18	22.94
13 June 2002	16	26.50
4 August 2002	17	15.76

4.4 Kriging Estimation and Gauge Sensitiviy Analysis

4.4.1 Semivariogram Analysis

Three semivariogram models namely: Gaussian, exponential and spherical models were generated from the rainfall data for each month of the 10-year gauge dataset. The MRL5 radar data were not used to generate the semivariograms since the OK interpolation in this study is based on rain gauge data only. The generated semivariogram models varied significantly from month to month. Figure 4.13 shows the fitted semivariogram models for the summer months (October-March). The semivariograms were modelled at a 10 *km* lag which is the average separation distance between the nearest rain gauge stations. The use of large lag distances increases the available data pairs for estimation but results in averaged detail in the semivariogram which does not accurately represent the actual rainfall data.

The results in Figure 4.13 display a clear increase in the semivariance as the separation distance increased, indicating that gauge stations close to each other have more similar data and hence their squared difference is less significant than those far apart as expected. The parameters of theoretical models fitted to the semivariograms for all months are described in Table 4.5. The models were fitted using an automated weighted least squares criterion described in Section 3.5.1. The models fitted for January, March and December have the largest sills, showing that there is a large variance of rainfall values including large total rainfall amounts for the these months. The months of June and August had the smallest sills.

(a) Semivariogram: January rainfall

(b) Semivariogram: February rainfall

(c) Semivariogram: March rainfall

(d) Semivariogram: October rainfall

(e) Semivariogram: November rainfall

(f) Semivariogram: December rainfall
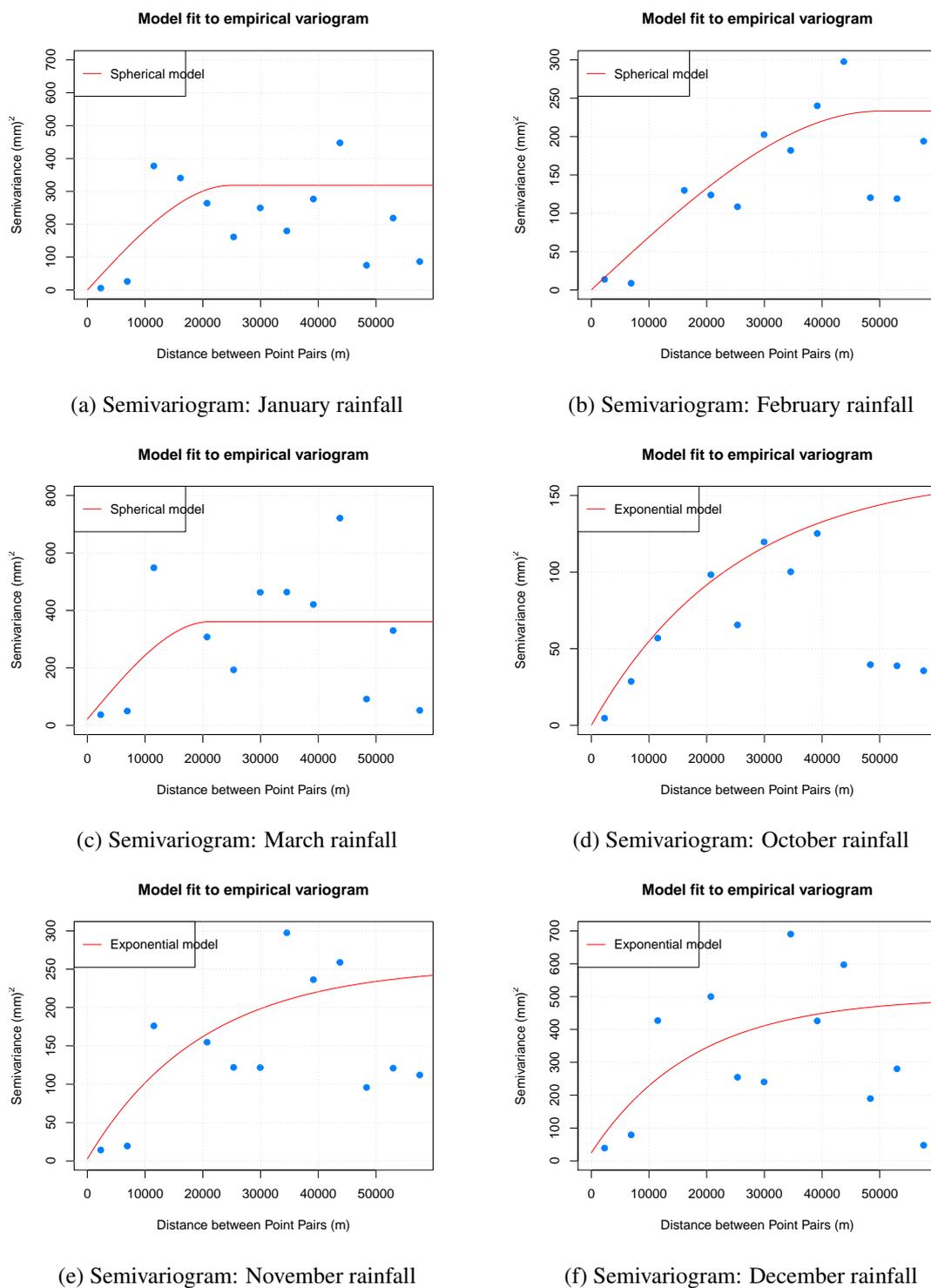
Figure. 4.13 The empirical (dotted) and theoretical (line) semivariograms of average monthly rainfall during the 10-year study period.

Overall the Spherical models was the most frequently best fitted to the monthly empirical semivariogram, followed by the Exponential model. The Gaussian model exhibited a good fit for the month of June when few rain gauges received rainfall. The best fitted semivariograms

for December and March have high nugget values as it appears that there is high variability in the average monthly rainfall data. These high nugget values can lead to a smoothing effect on the Kriging interpolation results (Wackernagel, 2014).

Table 4.5 Parameters of the theoretical models fitted to the semivariogram

Month	Semivariogram model	Nugget	Sill	Range (km)	AIC
January	Spherical	0.0	318.5	25.0	149.0
February	Spherical	0.0	233.2	45.0	136.0
March	Spherical	21.7	339.1	21.2	153.0
April	Spherical	9.8	60.5	16.5	126.6
May	Spherical	0.0	34.5	30.0	109.7
June	Gaussian	0.2	5.0	7.6	80.0
July	Exponential	0.0	50.3	50.0	104.7
August	Spherical	2.2	10.5	13.7	98.0
September	Spherical	0.6	28.2	40.0	101.0
October	Exponential	0.0	166.3	25.0	131.0
November	Exponential	2.7	251.9	44.9	141.0
December	Exponential	24.6	474.2	17.7	154.4

It is concluded that the monthly rainfall over the study area is best fitted by the Spherical model as it has a best fit frequency of 50% or more during the summer and winter season as shown in Figure 4.14, which shows the frequency of the best fitted semivariogram models with respect to the summer and winter seasons.

Figure. 4.14 Count of the different models chosen for the summer (October-March and winter (April-September) months.

4.4.2   Impact of Gauge Density

The characterization of rainfall patterns requires a dense rain gauge network because of the high variability of rainfall. The Kriging accuracy depends on the rain gauge density used for the estimation. The analysis in this section explores the effect of varying gauge density on Ordinary Kriging estimation. The effect of gauge density on Kriging is useful to determine the minimum network density required to obtain a given level of accuracy. Ideally the rain gauges must be sited in a pattern that closely provides a uniform spatial coverage of the study area. However, the distribution of the rain gauges within the study area was not uniform. Gauges are located based on catchment characteristics and practical considerations. The effect of removing rain gauges within the gauge network depends on the position of the removed rain gauge. It is not surprising that the removal of certain gauges will have a significant impact on the Kriging results. Taking this into account, rain gauges were removed using an approach which least disturbed the spatial structure of the gauge network. This was achieved by removing rain gauges according to their euclidean distances where one rain gauge is removed at each instance from the closest pair. Table 4.6 shows the rain gauges used for each gauge density and Figure 4.15 shows the spatial distribution of the highest (421 km^2) and lowest (889 km^2) gauge densities used, as the mean area gauge, within the study area.

Table 4.6 Gauges used for each gauge density

Number of gauges	Gauge density (gauge per km^2)	LE_LONG (0331121)	VISGAT (0332002)	EDENVILLE_MNC (0366003)	DOORNKLOOF (0366374)	VAN_TONDERTRST (0366536)	EIKENHOFF (0366666)	UITZICHT (0366699)	LINDLEY_MUN (0366743)	ERFDEEL (0366747)	BRAKDAM (0366806)	RAFIDUM (0367142)	NEWTON_GRANGE (0367143)	PETRUS_STEYNMUN (0367219)	SLAAPPLAATS (0367256_)	ERFENIS (0367484)	JACOBSDAL (0367666)	TERTUIS (0367802)	TWEELING_POL (0368003)
18	444	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
16	500	✓	✓	✓	✓	✓		✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓
13	615	✓	✓	✓	✓	✓		✓			✓	✓	✓	✓		✓		✓	✓
9	889		✓	✓	✓	✓		✓				✓		✓		✓			✓

(a) Highest rain gauge density used (18 gauges).   (b) Lowest rain gauge density used (9 gauges).

Figure. 4.15 Spatial spread of the highest and lowest gauge network density used.

Table 4.7 presents the cross-validation results of the OK estimation for all stations in each gauge density examined. The table shows the MAE, RMSE and Pearson correlation coefficient (R^2), respectively at the four gauge densities for the selected rainfall events using cross-validation of the observed versus the estimated values.

Table 4.7 Summary of Cross-validation results

Gauge density (gauge per km^2)	MAE (mm)	RMSE (mm)	$R^2 * 100$
444	3.14	4.08	76.89
500	3.28	4.35	68.00
615	4.27	4.62	64.75
889	4.49	5.13	29.57

Figure 4.16 displays the variation of cross-validation errors for the four gauge densities. The cross-validation error is the difference between the rainfall observed from the gauges and predicted rainfall after Leave One Out Cross-Validation (LOOCV). The errors were averaged at each gauge density for selected study rainfall events with 100% rain observations for the entire gauge network. As anticipated, the MAE and RMSE increase as the rain gauge density decreases. The MAE increased from 3.14 to 4.49 mm, while decreasing rain gauge density from 444 km^2 to 889 mm^2 as given in Table 4.6 as expected.

Decreasing the gauge density from 444 to 500 km^2 results in a low increase in the MAE but a gauge density reduction to 615 km^2 results in a large increase in the MAE, indicating that a certain number of gauges should be removed to cause a significant effect on the Kriging
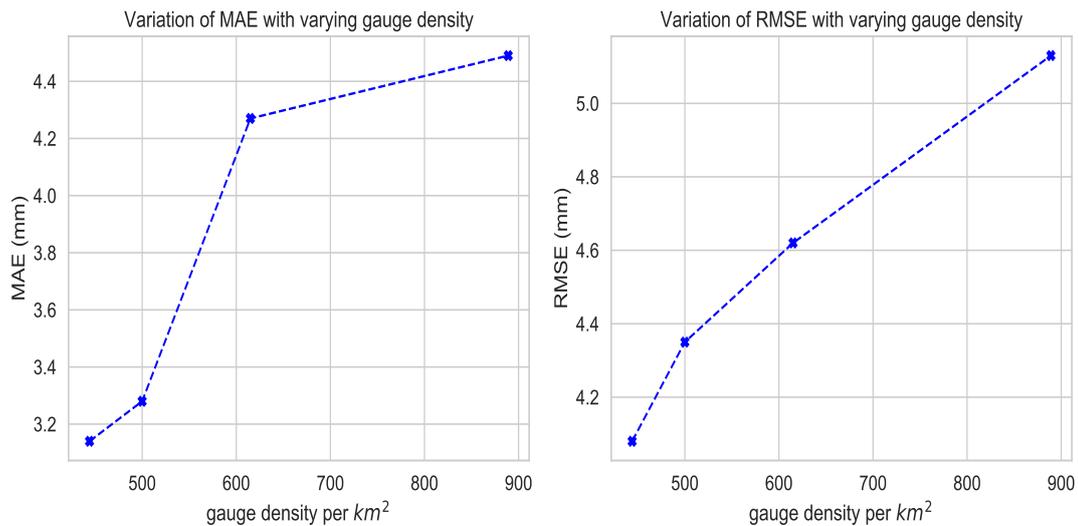
Figure. 4.16 MAE and RMSE for each rain gauge density examined for rainfall events were all gauges received rainfall.

accuracy. A change of the gauge density from 615 km^2 to 889 km^2 resulted in a slight change in the MAE which is in agreement to the findings of Biggs and Atkinson (2011) that a change in the gauge density affects the interpolation accuracy only up to a certain point. In this case, an RMSE of 4.08 mm at the highest gauge density (444 km^2) and a corresponding Pearson correlation coefficient of approximately 77% shows that the OK estimator performs reasonably well for interpolating gauge observations. Ly et al. (2011) investigated the performance of four interpolation methods (IDW, TP, OK, UK), and they report that OK produced the lowest RMSE value of about 2.42 mm on average. Furthermore lower gauge densities lead to mis-characterization of rainfall surfaces as expected. Studies by Otieno et al. (2014), Li and Heap (2011), Wagner et al. (2012) and McKee (2015) illustrated using different interpolation methods how few rain gauges mis-represent rainfall surfaces and conclude that a dense gauge network improves the explanation of rainfall variation. Figure 4.17 shows interpolated rainfall fields for a rainfall day derived from four different gauge densities. The Figure shows an increase in the smoothing effect on the rainfall fields as the number of rain gauges used decrease. It is important to note that the rainfall surface derived from gauge data only shows smoothed averages between the lowest and highest recorded rainfall values.
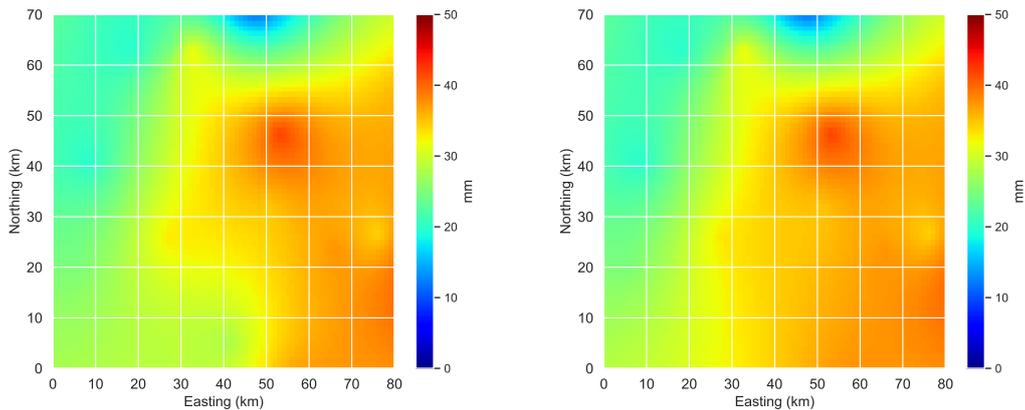
(a) Kriged rainfall field derived using 18 gauges.

(b) Kriged rainfall field derived using 16 gauges.
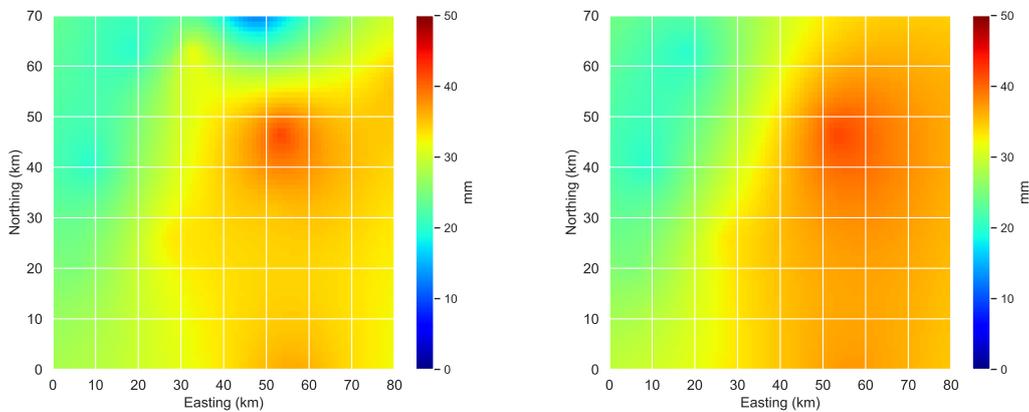
(c) Kriged rainfall field derived using 13 gauges.

(d) Kriged rainfall field derived using 9 gauges.

Figure. 4.17 The diagram shows the kriged rainfall fields from four different gauge densities on 12 September 2001.

This section evaluated the performance of the OK method under varying gauge densities. The use of inadequate rain gauge stations for rainfall estimation leads to unreliable results, therefore hindering design and development of flood forecasting schemes. The MAE and RMSE results confirm that OK performs better with a dense rain gauge network. In addition, the section illustrates that cross-validation and gauge density are useful for assessing interpolation methods, as expected from the literature referenced in Chapter 2.

4.5 Effect of Radar Range Dependency

One of the main and universal problems in radar-rainfall estimation is range-degradation which results in underestimation of rainfall accumulation at greater distance from the radar station (Kitchen and Blackwell, 1992; Meischner, 2005 and Villarini and Krajewski, 2010). The range-related biases that influence radar rainfall measurements are mainly due to vertical variability of reflectivity and radar sampling geometry. The interest in considering range-related effects

influencing radar estimates arises from the observation that radar reflectivity scans at high elevations ($> 2km$) are frequently used for rainfall estimation to avoid beam interaction with the terrain. An analysis of range-related errors provides useful information for the development of radar-range correction procedures. In this section rain gauges were grouped based on the distance from the radar station to examine the radar range effect as suggested by Goudenhoofdt and Delobbe (2009). The groupings were done based on 20 *km* radius intervals extending outward from the radar station. The examination was carried out in terms of seasonal rainfall, radar pixels and their corresponding rain gauge stations for the 20 selected rainfall events in Table 4.4 were split to fit into summer (October-March) and winter (April-September) seasons.
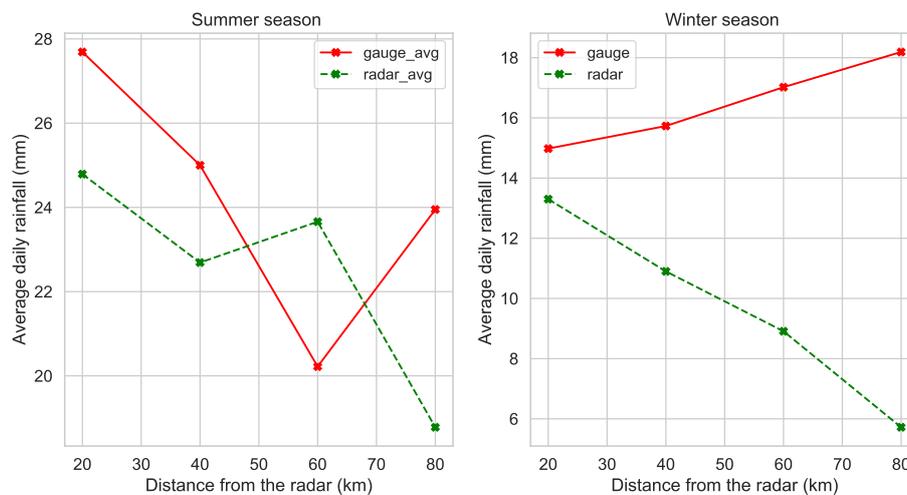
Figure. 4.18 Variation of average daily rainfall accumulations from both gauge and radar measurements with distance from the radar location for summer (October-March) and winter (April-September) rainfall events. Note that the winter rain rates are much lower than in summer.

Figure 4.18 shows the variation of average daily gauge rainfall accumulations with distance from the radar location presented for both summer and winter events. The results in Figure 4.18 show a definite seasonal bias, with the radar pixels underestimating gauge rainfall by as much as 85% during winter. At short ranges 0-40 *km* the average radar and gauge estimates line up well with an average difference of 2.5 and 4.5 mm between them for both summer and winter seasons respectively. As the distance increases, the radar underestimates the high rainfall from gauge stations, a peak in gauge accumulation is evident around 80 *km* during summer and beyond 60 *km* during winter, which are almost missed by the radar estimates.

The radar underestimation of gauge rainfall is due to partial beam filling and overshooting of low precipitation systems at far range. The radar beam sampling geometry is responsible for the partial beam filling problem. At far distances, the radar sample volume increases and therefore, small but intense features of the precipitation system are averaged out, resulting in a bias in the measured reflectivity. Overshooting of precipitation systems by the radar beam is due to increase in elevation of the beam at greater distances and is more significant during the winter season, when shallow and low rainfall systems are dominate. This is the case since

the MRL5 beam samples the atmosphere at too high an altitude (2 *km* above the ground) to observe the low level precipitation. Studies by Kitchen and Blackall (1992), Smith et al. (1996), Ciach et al. (2007), and Villarini and Krajewski (2010) have illustrated strong range degradation due to beam overshooting of precipitation systems at far range. Ciach et al. (2007) highlight that radar tend to overestimate the total rainfall in summer seasons and to underestimate it at far range in the winter season in Oklahoma, USA. This is evident in Figure 4.18 for the winter season where the radar underestimates the average gauge rainfall by 70%. However, during the summer season, mean radar rainfall at 60 *km* overestimates gauge rainfall and decreases sharply as expected beyond that point. This is probably due to the beam interception with the melting layer that presents enhanced reflectivity (bright band) compared to the precipitation below. Tabary (2007); Zhang et al. (2016) mention that reflectivity can be overestimated by a factor of 2-5 when the radar beam intercepts the bright band.

For the same distance the significant difference between average daily gauge and radar estimates might be due to the error in location caused by the non-vertical fall of precipitation. The non-vertical fall of precipitation is caused by wind drift especially in convective events (Lack and Fox, 2005). Wind drift causes rain drops to fall in a non-vertical line, so that the raindrops fall according to a parabolic trajectory under constant wind drift (Gunn and Marshall, 1955). It is a common assumption in radar rainfall estimation that precipitation observed aloft hits the ground directly below the volume sampled. However, rain may be observed above one area by the radar but fall in another in extreme cases of wind drift (Lack and Fox, 2007). The results in Figure 4.18 show a peak of radar rainfall at a distance of 60km from the radar station for the summer season which is not evident for the winter season. This anomaly is possibly due to wind drift in convective events which usually dominate the summer season. Lack and Fox (2007) investigated the effects of wind drift on stratiform and convective rainfall events who found that during convective events the radar overestimated gauge rainfall by as much as 30 mm in 3-hr accumulation periods and for stratiform events it overestimated gauge rainfall by an average of 8 mm due to wind drift. The differences for these events are from 3-hour accumulation periods, therefore, if the events last longer it could exceed these amounts for a given location.

Figure 4.19 shows a Height Time Indicator (HTI) for a storm event from the study of Berne et al. (2004), where precipitation cells initiating at the melting layer are made visible by a clear bright band at an altitude of 3.7 *km*. The HTI offers detail in terms of the vertical structure of the storm event, an interesting feature for the purpose of understanding the wind drift influence. Within the initial 500 *s* , 20 *dBZ* of reflectivity is observed just above the ground whilst more than 40 *dBZ* is observed 2.5 *km* above the ground. This results in a 67% difference in the rainfall observed between the gauge and radar estimates for the given time period. It is recommended to lower the Constant Altitude Plan Position Indicators (CAPPI) height to reduce the range over which wind-drift can influence rainfall but of course the down-side is that ground clutter becomes a problem (Berne et al., 2004).
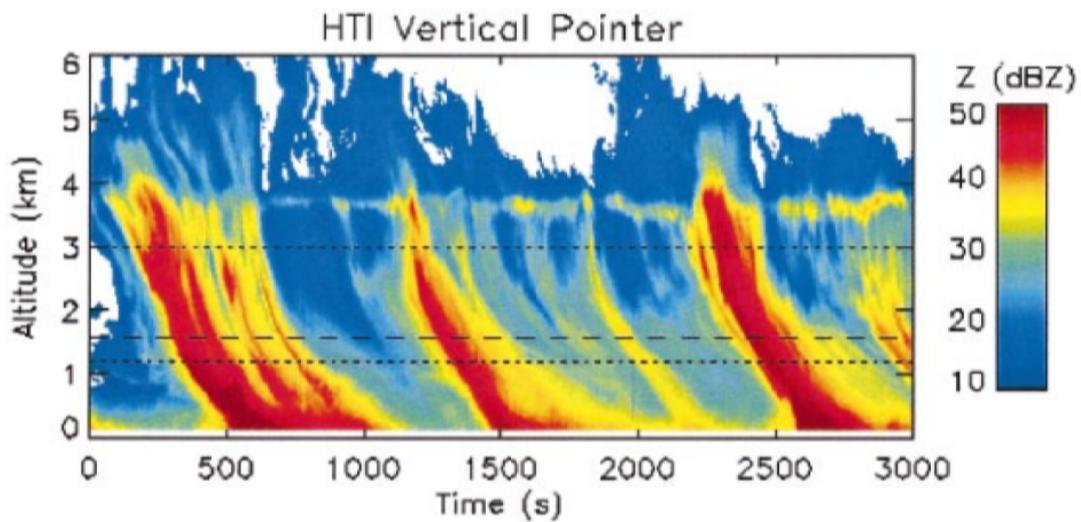
Figure. 4.19 Height time indicator (HTI) from an X-band radar vertically pointing radar for a 50 minute period (Berne et al., 2004).

Figure 4.20 shows rainfall fields for the MRL5 radar for three days with the least rainfall (a-c) and most rainfall (d-f) observed by both gauge and radar. The white areas shown in Figure 4.20 (a-c) display high reflectivity during dry days which are eliminated during the quality control procedure. Rings are visible in Figure 4.20 (d-f) which start from the radar station going outwards, these rings are due to the change in CAPPI levels of the MRL5 radar during the reflectivity volume scan. The rings represent different CAPPI levels with range of each elevation during a radar scan. The lowest CAPPI levels will observe the least reflectivity volume.

(a) Radar 24-hr accumulation on 04-08-2002.

(b) Radar 24-hr accumulation on 12-09-2001.

(c) Radar 24-hr accumulation on 26-08-2002.

(d) Radar 24-hr accumulation on 10-02-2000.

(e) Radar 24-hr accumulation on 18-11-1998.

(f) Radar 24-hr accumulation on 16-02-1998.

Figure. 4.20 The figure shows the MRL5 rainfall field with clutter fields for the three selected days with least (a-c) and most (d-f) rainfall received for the selected rainfall events. The rings are the result of the CAPPI treatment introduced by Terblanche (1996)

Generally, the quality of radar estimates decreases as the distance from the radar increases. The results in Figure 4.21 show that the mean RMSE value does not exceed 15 mm within 50 *km* from the radar, while the correlation becomes less after 40 *km*. The correlation diminishes as the distance from the radar increases because of beam broadening and overshooting of weather systems. In Figure 4.21, the highest mean RMSE value is observed between 40-70 *km*, these results support the findings of McKee (2015), in which the highest RMSE mean value was attained in the same distance range for an C-band radar and three radar-gauge merging methods. Krajewski et al. (2010) developed a model that describes range dependent errors in radar rainfall estimates. Their proposed model represents the reflectivity vertical profile structure in relation to the radar beam height. Using a 4-year data-set of radar rainfall, the range dependent error model illustrated a good match with differences observed between radar and gauge estimates.
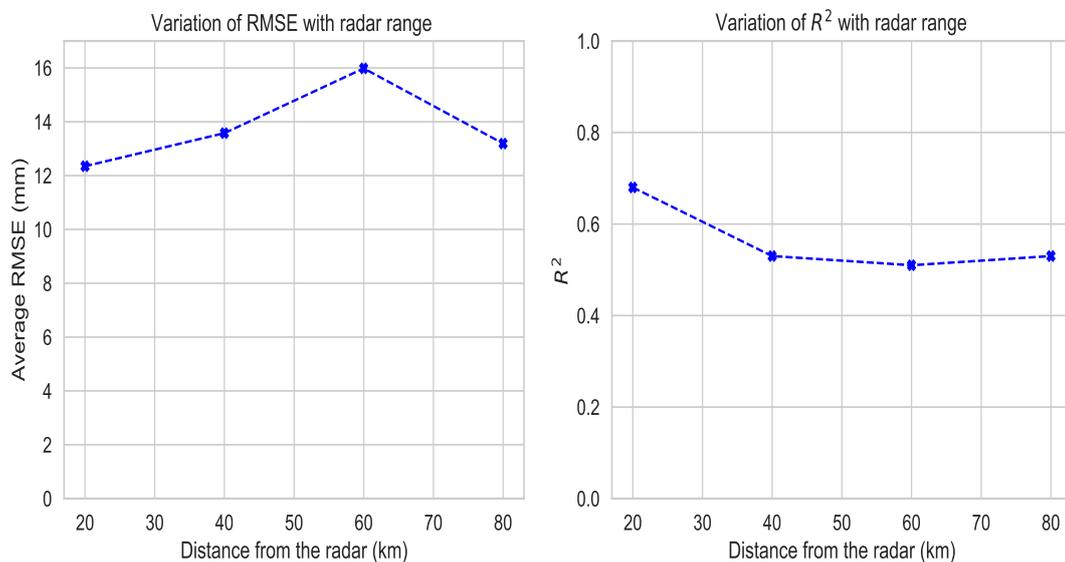
Figure. 4.21 Variation of the RMSE and the Pearson correlation coefficients with distance from radar.

This section aimed at providing an overview of how range degradation influences radar rainfall estimation. The section highlights limitations of radar rainfall estimates with varying radar range for summer and winter seasons. It is evident that the radar rainfall estimates are reliable within a limited distance around the radar location as demonstrated by the RMSE and correlation coefficients. The comparison of daily rainfall averages derived from both the radar and gauges confirm that there can be significant differences between the estimates due to range dependency, ground clutter and wind drift. Generally, the MRL5 radar rainfall underestimates gauge rainfall as we move away from the radar station because of beam broadening and blocking as well as the land curvature effect on the radar beams.

4.6   Effect of Seasonal Changes and Storm Type on Radar Rainfall Estimates

This analysis documents the difference between radar, gauge and interpolated rainfall data. The interpolated rainfall data are derived from rain gauge observations, and the resulting estimates are used across the sub-catchment regardless of the shape and topography. Previous studies indicate that radar rainfall estimates underestimate gauge-derived data by 5- 10% (Morin and Gabella, 2007 and Price et al., 2014) and other studies report a great difference ($\sim 50\%$) between the two sensors (Villarini, 2008 and Espinosa et al., 2015). The following analysis is employed on a daily scale with the objective to determine if there exist significant seasonal biases in the radar and gauge-derived rainfall products. The daily MRL5 radar rainfall was compared with interpolated rainfall estimates and point gauge measurements. The interpolated rainfall estimates were obtained by distributing gauge accumulations using OK while the MRL5 accumulations were obtained by summing up 5 minute radar scans into 24-hr accumulations (08:00-08:00). Four rain gauge stations were left out during the Kriging estimation as check gauges and only radar pixels and grid cells directly above the reference gauges were used for the analysis. Figure 4.22 shows the location of the radar and the selected reference gauge stations. The four reference gauges were selected randomly by picking one gauge from each quadrant of the study area.
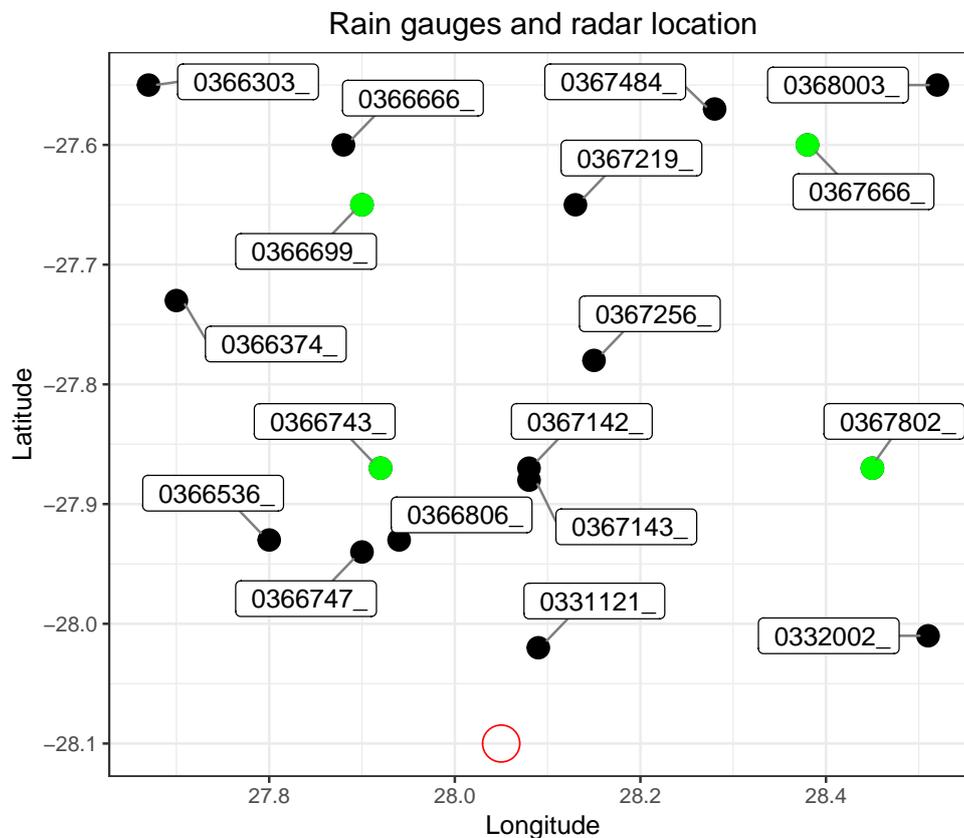
Figure. 4.22 The image shows the location of the radar (red circle) and the reference gauge stations marked in green across the area.

It is important to consider the daily statistics of the wet days (defined in Section 4.3.1) over the study area before doing any comparison, since the examined days were selected from the wet days. To complement this, Figure 4.23 shows the total number of wet days (> 3 mm) during the 10-year study period on the study area as measured by the network of 18 gauge stations. Figure 4.23 again clearly indicates the seasonal trend in the rainfall experienced over the study area (refer to Section 4.3), with December having the most number of rainy days which exceed the 3mm threshold amount used and July having the least rainy days.
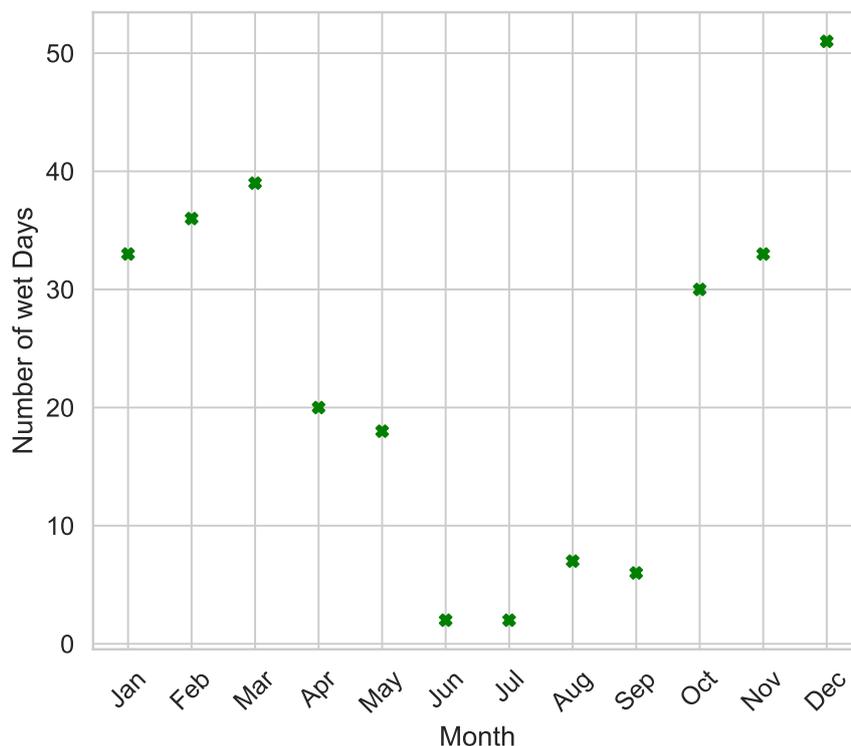
Figure. 4.23 The image shows the total number of wet days observed in each month during the 10-year period (1994-2003).

Figure 4.24 shows the means and standard deviations of gauge, OK and MRL5 radar rainfall estimates for days with the heaviest rainfall in each month across the selected rainfall events. The results reflect a definite seasonal bias with MRL5 radar underestimating both the gauge and OK rainfall values in April by as much as 95% and 125% respectively. However, during the summer season (October-March) the agreement between gauge and OK rainfall is much better. The MRL5 radar overestimates both gauge and OK rainfall estimates from November to February, with almost equal rainfall in March. Throughout the summer events the total radar and gauge rainfall are within $\sim 30\%$ of each other, whilst the gauge and OK estimates have a difference of $\sim 4\%$. Similarly, during winter rainfall events (May to August) the radar and gauge rainfall are within $\sim 31\%$, whilst the gauge and OK estimates have a difference of $\sim 22\%$.
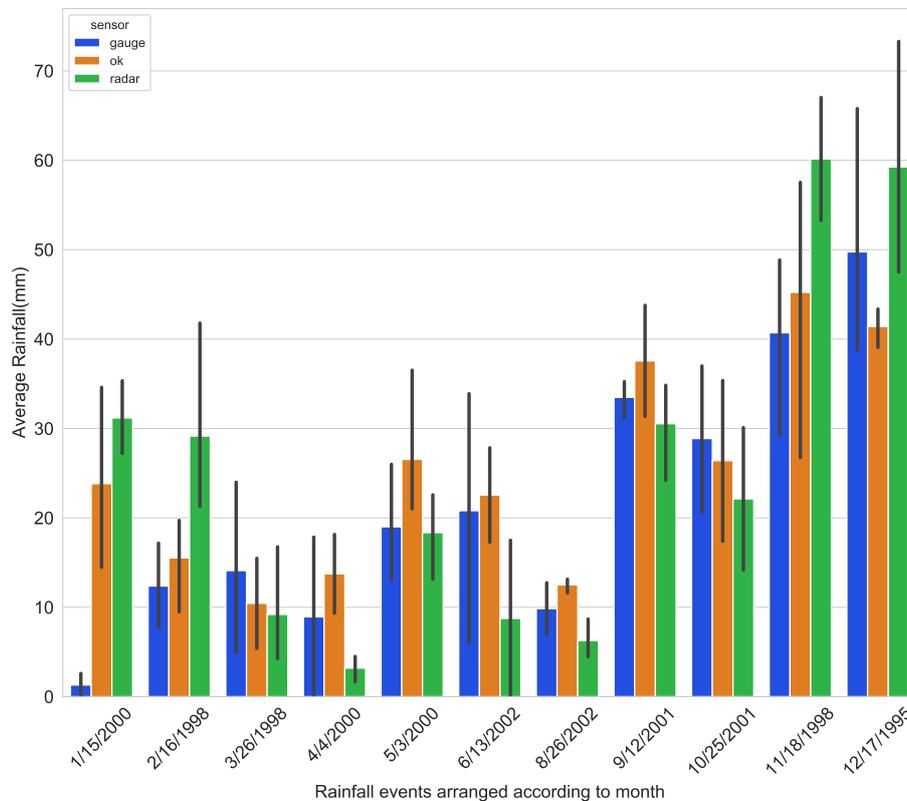
Figure. 4.24 Mean rainfall (mm) by month for selected rainfall events with standard deviations for gauge, OK and radar estimates for days with the heaviest rainfall in each month over the 10 years, with standard deviations illustrated by black bars.

The percentage differences between gauge and OK estimates are probably lower than radar-gauge differences for both seasons because the Kriging estimation is a weighted linear sum of the neighbouring gauge observations and more weight is assigned to the closest neighbor. Therefore, if nearby gauge values resemble the actual amount of rainfall received at an unobserved location then the OK estimate is more likely to be close to the actual observed value. The decrease in the difference between OK estimates and gauge observations during the winter season is most probably due to the low temporal variability of rainfall totals during rare stratiform events.

Generally, the MRL5 tends to overestimate rainfall during summer and underestimates rainfall during winter across the selected rainfall events. Studies have reported several factors that cause underestimation of gauge rainfall by radar which include errors due to Z-R relationship, anomalous propagation, beam blocking, bright band and beam filling (Uijlenhoet, 2001; Kurri and Huuskonen, 2008; Villarini and Krajewski, 2010). It is difficult to identify which factors have the greatest effect on daily rainfall estimates (Stellman et al., 2001). However, the MRL5 data were subjected to quality control for anomalous propagation, ground clutter and bright band before being published. Therefore in this case, during convective rainfall events, beam filling could be the source of inaccuracy due to the strong reflectivity gradients associated

with convective events. This would not explain the MRL5 radar underestimation during the winter rainfall events. The radar underestimation of rainfall in winter is due to radar beams overshooting the precipitation cells, probably because of stratiform rather than convective rainfall, as described in the previous section. Standard deviations in Figure 4.24 also indicate a seasonal bias. The standard deviations for OK estimates are relatively greater during the summer season compared to the winter season. This is due to the enhanced variability in rainfall totals during convective storms which are mostly received in summer in the Free State. However, the MRL5 radar does not reflect the variability in standard deviations during summer as it captures the highly variable rainfall events better than gauges.

Figure 4.25 presents the trend lines between OK and radar data against Gauge data at the sites of the reference gauges. The correlation coefficients (R^2) values show that OK gives a good estimation of the gauged data than the radar. The slightly high correlation coefficients attest to OK capability to represent the rainfall distribution.
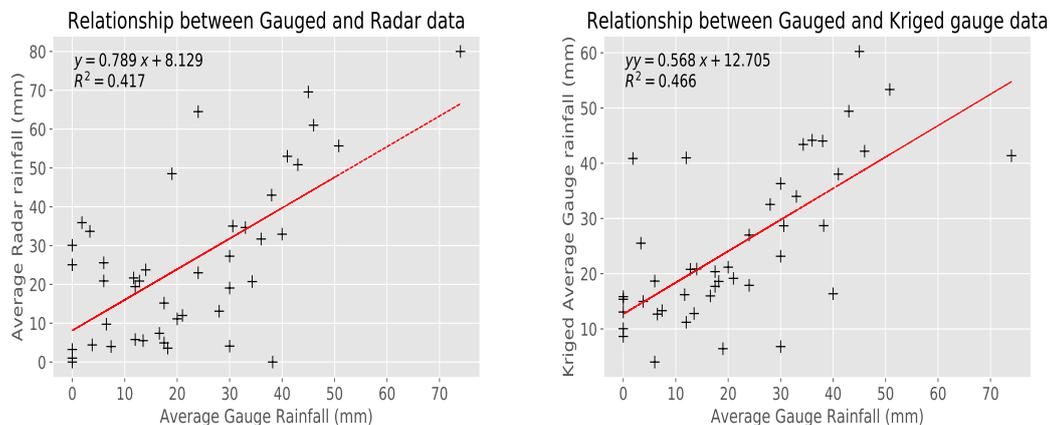
(a) Radar against Gauge rainfall plot with trend line and correlation coefficient.

(b) Kriged gauge against Gauge rainfall plot with trend line and correlation coefficient

Figure. 4.25 Plots of Kriged gauge (OK)and Radar against Average Gauge Rainfall

This section documented the differences between radar, gauge and Kriged gauge rainfall products during the summer and winter seasons. Radar and gauge rainfall fields were computed from 24-hour rainfall accumulations. The results indicate that the radar underestimates rainfall in winter and overestimates the rainfall during summer compared to gauge products. The OK estimates generally overestimate rainfall events due to their mis-characterization of rainfall. Furthermore, the radar rainfall exhibits higher correlation coefficients with gauge rainfall during the summer season than the winter season whilst OK estimates have high correlation coefficients with gauge rainfall in few instances. The conclusion is that the contrasting differences between convective and stratiform rainfall results in seasonal biases mentioned above.

Chapter 5

Summary, Discussion and Conclusion

5.1 Introduction

This dissertation compares the MRL5 weather radar, gauge and Kriged gauge rainfall measurements in the Free State (Bethlehem) using data covering 10 rainfall seasons. The aim of the study was to present a general overview of opportunities that result from an integrated analysis of gauge and weather radar rainfall products, and assess the operational and environmental factors influencing the accuracy of spatial rainfall estimates in the area of interest. The thesis highlights factors influencing the accuracy of spatial rainfall estimation so that it can motivate more thorough research in flood prone regions of South Africa. This chapter provides a summary and discussion of the conclusions drawn from this study, which are divided into two categories:

(i)   Summary from a comprehensive review to indicate the derivation, application and limitations of using gauge, radar and a Kriged gauge network for rainfall estimation; and

(ii) Results from an analysis of daily rainfall characteristics in the Free State;

(iii) Results of generated spatial rainfall fields from rain gauges and weather radar; and

(iv) Results from an analysis of operational and enviromental factors influencing rainfall estimation using radar.

The last section of this chapter outlines recommendations for areas of future research.

5.2 Rainfall Estimation

Hydrological models are an important tool used for forecasting floods. The development and application of hydrological models requires accurate rainfall data as input to produce results with a high degree of confidence. Rainfall estimates at high spatial and temporal scales are preferable for flood forecasting purposes. Recently, radar combined with rain gauges are the most widely used sensors in rainfall estimation for near real-time applications (Zhang et al., 2018). Rain gauges measure the rainfall amount at a given point and, therefore, present accurate rainfall depth readings. However, rainfall fields can exhibit high variability in both space and time. Since gauges have the ability to only measure rainfall at a given point, they fail to capture the spatial and temporal variability exhibited by rainfall fields. Consequently, variability in rainfall fields can influence stream flow responses and hence the accuracy of hydrological models that rely on gauge data as input can be reduced. The spatial distribution of gauge rainfall to obtain a better representation of rainfall fields can be achieved using geostatistical interpolation methods.

Following a comprehensive review (Chapter 2) on geostatistical interpolation of rainfall, it is clear that there is a possibility for near real-time application of Kriging methods which merge gauge and radar data. The radar-gauge merging methods, however, require high computational efficiency and quality secondary data. They also smooth out the interpolated rainfall field and cannot recover intensities outside their collective range. This has hindered their widespread application in an operational context.

The radar provides high resolution precipitation data that is favourable in forecasting applications although it still lacks quantitative precision due to several sources of error (Villarini and Krajewski, 2010 and Berne and Krajewski, 2013). Weather radar is a remote sensing tool which measures rainfall indirectly through reflectivity off hydrometers in the atmosphere and converting the reflectivity to rainfall intensity using a mathematical relationship suggested by Marshall and Palmer (1948). Van Heerden and Steyn (1999), Pegram (2003), Sinclair and Pegram (2005), Clothier (2011) and Becker (2016) investigated the performance and application of radar QPEs in South African hydrology. These studies indicate that there is a substantial gap between radar and gauge rainfall products. The errors in radar rainfall products can be intensified when radar rainfall estimates are used as input in flood forecasting applications, thus leading to large prediction errors (Zhu et al., 2013).

The application of radar, gauge and interpolated rainfall data is affected by several operational and environmental factors. These factors tend to influence the performance and quality of rainfall sensors in an operational context. The use of radar in rainfall estimation is often influenced by the type of precipitation (storm type), proximity to the radar station (radar range) and topography, while rain gauge networks are often affected by catchment characteristics and density of the gauge network. The estimation of rainfall using radar, gauge and kriged data has been studied in South Africa but few studies have compared these rainfall products. How and to what extent each factor influences rainfall estimation techniques needs to be taken into consideration. Therefore, there is a need to investigate the performance of each rainfall estimation technique in attaining daily rainfall under varying operational and environmental conditions.

5.3 Daily Rainfall Accumulations near Bethlehem, Free State

5.3.1 What are the Characteristics of Daily Rainfall near Bethlehem?

A 10-year rainfall dataset has been analysed in a comprehensive manner in Chapter 4. The daily rainfall over a region North of Bethlehem is highly variable and exhibits a strong summer and winter seasonal trend. On average 75% of the rainfall is received in summer (October-March) and extreme rainfall events are more likely to occur during this season. Most of the summer rainfall is due to convective activity and the low winter rainfalls are due to stratiform storms.

Through-out the entire study period the highest average rainfalls occurred in the late summer (January and February), the year 2000 having the highest average rainfall record. The peak in average rainfall during the late summer of 2000 was most likely associated with the Tropical Cyclone Eline in February. The Exploratory data analysis (EDA) carried out in Chapter 4 shows that there is a high frequency of low daily rainfall in Bethlehem, with heavy rainfall being received at higher altitudes probably due to orographic effects. The result is that the rainfall distribution in Bethlehem is non-normal.

5.3.2   How to Obtain Spatial Rainfall Estimates from Rain Gauges and Weather Radar?

Spatial rainfall fields in this study were generated from rain gauges using the Ordinary Kriging technique based on 20 carefully selected rainfall events between the period 1994-2003. 24-hour gauge accumulations were used to distribute daily rainfall amounts throughout the study region and the rainfall days considered for Kriging estimation include the heaviest daily rainfall events during the study period. The Ordinary Kriging method (OK) estimates rainfall values at ungauged locations using a weighted linear combination of the observed neighbouring rainfall measurements. The spatial structure of the rainfall observations was determined using the semivariogram. The spherical semivariogram model provided the best fit based on the weighted least squares criterion for representing the rainfall structure in most rainfall events. The resulting Kriged rainfall fields examined in Chapter 4 perform differently depending on the amount of daily rainfall received and the number of rain gauge observations used for the estimation. The performance of OK was assessed by applying leave-one-out cross validation on the interpolation results. According to results drawn from the analysis of Mean absolute error (MAE), Root mean square error (RMSE) and correlation (R^2) of daily kriged rainfall fields, OK consistently performs better under a dense rain gauge network density. When looking at the radar image output, OK well represents both high and low rainfall. However, a smoothing effect is evident in the Kriged rainfall surfaces when a less dense gauge network is used. This smoothing effect leads to poor estimation of rainfall fields.

The selection of rainfall events with varying rainfall caharacteristics was used to investigate the quality of the MRL5 radar product. The Python NETCDF4 package was used to generate MRL5 radar rainfall surfaces for the selected 20 rainfall events investigated in Chapter 4, yielding a product at $1\ km \times 1km$ grid spacing. Significant improvements in the spatial representation of rainfall is provided by the MRL5 radar when compared to Kriged gauge estimation, up to a radius of 100 km. The MRL5 radar captured the variability of precipitation systems better during heavy rainfall events, which are mostly a product of convective activity during the summer season. The benefits are due to an enhanced spatial and temporal resolution provided by the radar. The MRL5 completed one reflectivity volume scan in 5-minutes with a spatial resolution of 1 km and it was operated in S-band mode, thus, making it less prone to attenuation errors at long range. During low rainfall events which are associated with stratiform precipitation, patches of clutter dominate the MRL5 radar rainfall field. Most MRL5 radar artefacts which

appear on the rainfall surface map are reduced by eliminating non-meteorological returns and the bright band.

5.3.3 What are the Fffects of Operational and Environmental Factors on Daily Rainfall Estimates?

Making use of carefully selected daily rainfall events the effect of gauge density, range dependency, storm type and seasonal changes on radar, gauge and Kriged gauge rainfall products were investigated. The rain gauge observations were used as the ground truth rainfall values. A conclusion on the findings is made in this section.

Rain gauges were removed systematically, from the set of 18, to asses the effect of gauge network density on the accuracy of Kriged gauge rainfall estimates. As expected, a dense rain gauge network provides a better accuracy of the Kriging estimations compared to a less dense network. OK estimates indicated a gradual increase in MAE and RMSE as the gauge density decreased from one gauge station per 444 km^2 to one gauge station per 889 km^2. This confirms the obvious that the use of inadequate rain gauge stations for rainfall estimation leads to unreliable results, thereby hindering the design of flood forecasting schemes. The OK method, however, did not display a substantial increase in error beyond a gauge density of one gauge per 615 km^2. This suggests that a reduction in gauge density affects the Kriging estimation accuracy up to a certain extent beyond which it remains almost constant.

The differences between the MRL5 radar and gauge rainfall, and correlations between the two variables with respect to the radar range were investigated. To assess the effect of radar range on the published rainfall estimates, rain gauge stations were grouped into 20 km intervals moving away from the radar tower and each radar pixel was compared to the gauge station directly above it. It is clear that the radar estimates are reliable within a distance of 70 km from the radar. This was indicated by the RMSE values and correlation (R^2) between the gauge and radar obtained as a function of the distance. Generally the MRL5 radar rainfall likely underestimates rain gauge measurements. This underestimation increases gradually as the distance from the radar station increases due to land curvature effects on radar beams. Since the topography of the study area is relatively flat, the difference between the radar and gauge estimates tend to increase in a uniform manner. The RMSE values confirm that interaction of radar beams with the terrain and range degradation are capable of causing significant underestimation of rainfall by the MRL5 radar.

Results of the seasonal analysis of radar rainfall estimates compared with gauge and Kriged gauge estimates showed that a seasonal bias is present in the radar rainfall product. The summer rainfall events (October-March) results show that the MRL5 radar overestimates gauge measurements and the two rainfall products tend to match towards the end of the summer events. Conversely, OK estimates have almost similar rainfall amounts with the gauge rainfall throughout the summer rainfall events. This close match between OK estimates and gauge observations is probably due to the neigbouring rainfall observations used for the Kriging estimation resembling the true rainfall amounts at the estimated points. During winter rainfall events (April-September)

the MRL5 radar underestimates gauge rainfall, while OK overestimates gauge rainfall. The sparse and low winter rainfall events are greatly underestimated by the MRL5 radar. The low precipitation systems in winter which are likely overshot by radar beams are responsible for the MRL5 radar rainfall deficit during winter events. Correlations between radar and gauge were observed to be high during the high rainfall events in summer, while the OK method has a poor correlation with gauges throughout the selected rainfall events. The conclusion is that the contrasting differences between convective and stratiform rainfall are responsible for the seasonal biases mentioned above.

5.4   Recommendations for Future Research

5.4.1   Availability of rainfall data

A good spatial coverage is required to produce robust rainfall estimates over the study area. While Bethlehem has a relatively good dense rain gauge network coverage, the effective range of the MRL5 radar is limited. Therefore, the new dual pol S-band radar installed in Bethlehem in March 2010 (Becker, 2016) is likely to produce reliable rainfall estimates for the region. Missing data in our rainfall datasets affect our results depending on the proportion. The lack of availability of radar data limited the selection of rainfall events for analysis. Extending the number of rainfall events for analysis will reduce the effect of errors from outlying events for each estimation method.

5.4.2   Quality of Rainfall Data

The reliability of our statistical analysis relies on the quality of rainfall data. The uncertainty associated with results in Chapter 4 are likely related to the relatively small amount of gauge and radar data with good quality. One can argue that corrupted data are unlikely to come from the rain gauges and MRL5 radar due to the quality control procedures undertaken before they are published. However, the effect of calibration errors on both gauge and radar measurements must not be ignored. Therefore, it would be interesting for further studies to examine the validity of the quality control procedures. Goudenhoofdt and Delobbe (2009) suggests several decades of rainfall data are needed to validate results associated with extreme rainfall phenomena.

5.4.3   Spatial Rainfall Estimation

Although considerable efforts have been made to generate the best spatial rainfall estimates in this study, there are still ways to improve the estimation. Sinclair and Pegram (2005), Goudenhoofdt and Delobbe (2009), , Ly et al. (2011), Wang et al. (2013), McKee (2015) and Gabriele et al. (2017) show that the radar-gauge merging methods described in Chapter 2 improve the accuracy of rainfall accumulation estimates. Specifically, extreme rainfall estimation could benefit from combining radar and gauge measurements (Boodoo et al., 2015). Evaluating the performance of

each radar-gauge merging method for specific rainfall events, as in McKee and Binns (2016) should be done as well .

To conclude, rainfall exhibits high temporal and spatial variability. For flood forecasting applications, the space-time dynamic of rainfall is an important requirement. The spatial rainfall fields generated in this study have a temporal and spatial resolution of 24-hours and 1 *km* respectively, which is not sufficiently detailed for most river basins. Flood forecasting applications require high resolution rainfall data because of the small scale nature of the processes. For a particular spatial scale, the required temporal resolution required is associated with the characteristics of the rainfall field and the response time of flow rates to the rainfall event. A possible solution to obtain rainfall estimates with a higher resolution is to combine radar measurements at 5-minute and 1 *km* with 5-minute tipping bucket rain gauge measurements from Automatic weather stations (AWS) operated by the SAWS, possibly using the conditional merging method of Sinclair and Pegram (2005).

References

Adhikary, S. (2017), Optimal Design of a Rain Gauge Network to Improve Streamflow Forecasting, PhD thesis, Victoria University.

Adhikary, S., Yilmaz, A. and Muttil, N. (2015), 'Improved spatial interpolation of rainfall using genetic programming', *http://www. mssanz. org. au/modsim2015/authorsA-B. html* pp. 2214–2220.

Alexander, W. J. R. (1990), *Flood hydrology for southern Africa*, South African National Committee on Large Dams.

Anselin, L., Syabri, I. and Kho, Y. (2006), 'Geoda: an introduction to spatial data analysis', *Geographical analysis* **38**(1), 5–22.

Atlas, D., Ulbrich, C. W., Marks Jr, F. D., Black, R. A., Amitai, E., Willis, P. T. and Samsury, C. E. (2000), 'Partitioning tropical oceanic convective and stratiform rains by draft strength', *Journal of Geophysical Research: Atmospheres* **105**(D2), 2259–2267.

Austin, P. M. (1987), 'Relation between measured radar reflectivity and surface rainfall', *Monthly Weather Review* **115**(5), 1053–1070.

Barge, B., Humphries, R., Mah, S. and Kuhnke, W. (1979), 'Rainfall measurements by weather radar: Applications to hydrology', *Water Resources Research* **15**(6), 1380–1386.

Battan, L. J. (1973), 'Radar observation of the atmosphere'.

Bech, J., Codina, B., Lorente, J. and Bebbington, D. (2003), 'The sensitivity of single polarization weather radar beam blockage correction to variability in the vertical refractivity gradient', *Journal of Atmospheric and Oceanic Technology* **20**(6), 845–855.

Becker, E. H. (2016), Application of a quantitative precipitation estimation algorithm for the S-band radar at Irene, South Africa., Master's thesis, University of Kwazulu-Natal.

Bell, V. and Moore, R. (1998), 'A grid-based distributed flood forecasting model for use with weather radar data: Part 2. case studies', *Hydrology and Earth System Sciences Discussions* **2**(2/3), 283–298.

Berndt, C., Rabiei, E. and Haberlandt, U. (2014), 'Geostatistical merging of rain gauge and radar data for high temporal resolutions and various station density scenarios', *Journal of Hydrology* **508**, 88–101.

Berne, A., Delrieu, G., Andrieu, H. and Creutin, J.-D. (2004), 'Influence of the vertical profile of reflectivity on radar-estimated rain rates at short time steps', *Journal of Hydrometeorology* **5**(2), 296–310.

Berne, A. and Krajewski, W. F. (2013), 'Radar for hydrology: Unfulfilled promise or unrecognized potential?', *Advances in Water Resources* **51**, 357–366.

Beven, K. J. (2011), *Rainfall-runoff modelling: the primer*, John Wiley & Sons.

Biggs, E. M. and Atkinson, P. M. (2011), 'A comparison of gauge and radar precipitation data for simulating an extreme hydrological event in the severn uplands, uk', *Hydrological Processes* **25**(5), 795–810.

Boodoo, S., Hudak, D., Ryzhkov, A., Zhang, P., Donaldson, N., Sills, D. and Reid, J. (2015), 'Quantitative precipitation estimation from a c-band dual-polarized radar for the 8 july 2013 flood in toronto, canada', *Journal of Hydrometeorology* **16**(5), 2027–2044.

Borga, M. (2002), 'Accuracy of radar rainfall estimates for streamflow simulation', *Journal of Hydrology* **267**(1-2), 26–39.

Borga, M., Anagnostou, E. N. and Frank, E. (2000), 'On the use of real-time radar rainfall estimates for flood prediction in mountainous basins', *Journal of Geophysical Research: Atmospheres* **105**(D2), 2269–2280.

Burcea, S., Cheval, S., Dumitrescu, A., Antonescu, B., Bell, A. and Breza, T. (2012), 'Comparison between radar estimated and rain gauge measured precipitation in the moldavian plateau.', *Environmental Engineering & Management Journal (EEMJ)* **11**(4).

Chaudhary, S., Dhanya, C. and Vinnarasi, R. (2017), 'Dry and wet spell variability during monsoon in gauge-based gridded daily precipitation datasets over india', *Journal of Hydrology* **546**, 204–218.

Chen, C.-S., Chen, Y.-L., Liu, C.-L., Lin, P.-L. and Chen, W.-C. (2007), 'Statistics of heavy rainfall occurrences in taiwan', *Weather and Forecasting* **22**(5), 981–1002.

Chen, Y.-C., Wei, C. and Yeh, H.-C. (2008), 'Rainfall network design using kriging and entropy', *Hydrological Processes: An International Journal* **22**(3), 340–346.

Chen, Z.-M., Yeh, Y.-L. and Chen, T.-C. (2018), 'The clustering analysis and spatial interpolation of intense rainfall data', *International Journal of Emergency Management* **14**(2), 122–136.
URL: *https://www.inderscienceonline.com/doi/abs/10.1504/IJEM.2018.090882*

Chumchean, S., Seed, A. and Sharma, A. (2006), 'Correcting of real-time radar rainfall bias using a kalman filtering approach', *Journal of Hydrology* **317**(1-2), 123–137.

Ciach, G. J., Krajewski, W. F. and Villarini, G. (2007), 'Product-error-driven uncertainty model for probabilistic quantitative precipitation estimation with nexrad data', *Journal of Hydrometeorology* **8**(6), 1325–1347.

Clothier, A. N. (2011), High resolution space-time modelling of rainfall: the string of beads model., PhD thesis, University of Kwazulu-Natal.

Cole, S. J. and Moore, R. J. (2008), 'Hydrological modelling using raingauge-and radar-based estimators of areal rainfall', *Journal of Hydrology* **358**(3-4), 159–181.

Collier, C. G. (1989), *Applications of weather radar systems*, Wiley.

Cressie, N. and Hawkins, D. M. (1980), 'Robust estimation of the variogram: I', *Journal of the International Association for Mathematical Geology* **12**(2), 115–125.

Cressie, N. and Wikle, C. K. (2015), *Statistics for spatio-temporal data*, John Wiley & Sons.

Davis, C. L. and Vincent, K. (2017), 'Climate risk and vulnerability: A handbook for southern africa'.

De Coning, E. and Poolman, E. (2011), 'South african weather service operational satellite based precipitation estimation technique: applications and improvements', *Hydrology and Earth System Sciences* **15**(4), 1131–1145.

Delhomme, J. (1978), 'Kriging in the hydrosciences', *Advances in water resources* **1**, 251–266.

Deyzel, I., Pegram, G., Visser, P. and Dicks, D. (2004), Spatial interpolation and mapping of rainfall (simar). volume 2: Radar and satellite products, Technical report, WRC report No 1152/1/04, ISBN 1-77005-160-0.

Dingman, S. L. (2015), *Physical hydrology*, Waveland press.

Doviak, R. J. et al. (2006), *Doppler radar and weather observations*, Courier Corporation.

Durre, I., Menne, M. J., Gleason, B. E., Houston, T. G. and Vose, R. S. (2010), 'Comprehensive automated quality assurance of daily surface observations', *Journal of Applied Meteorology and Climatology* **49**(8), 1615–1633.

Durre, I., Menne, M. J. and Vose, R. S. (2008), 'Strategies for evaluating quality assurance procedures', *Journal of Applied Meteorology and Climatology* **47**(6), 1785–1791.

Dyson, L. L. (2009), 'Heavy daily-rainfall characteristics over the gauteng province', *Water Sa* **35**(5).

Einfalt, T., Arnbjerg-Nielsen, K., Golz, C., Jensen, N.-E., Quirmbach, M., Vaes, G. and Vieux, B. (2004), 'Towards a roadmap for use of radar rainfall data in urban drainage', *Journal of Hydrology* **299**(3-4), 186–202.

Eric, C. (2011), ' South Africa: Floods kill 120 and destroy crops'.
    URL: *https://www.pri.org/stories/2011-01-28/*

Falah, A. N., Abdullah, A. S., Parmikanti, K. and Ruchjana, B. N. (2017), 'Prediction of cadmium pollutant with ordinary point kriging method using gstat-r', *AIP Conference Proceedings* **1827**(1), 020019.
    URL: *https://aip.scitation.org/doi/abs/10.1063/1.4979435*

Faurès, J.-M., Goodrich, D., Woolhiser, D. A. and Sorooshian, S. (1995), 'Impact of small-scale spatial rainfall variability on runoff modeling', *Journal of hydrology* **173**(1-4), 309–326.

Ford, C. (2015), Understanding q-q plots.

Frei, C. and Schär, C. (1998), 'A precipitation climatology of the alps from high-resolution rain-gauge observations', *International Journal of Climatology: A Journal of the Royal Meteorological Society* **18**(8), 873–900.

Friedrich, K., Hagen, M. and Einfalt, T. (2006), 'A quality control concept for radar reflectivity, polarimetric parameters, and doppler velocity', *Journal of Atmospheric and Oceanic Technology* **23**(7), 865–887.

Gabriele, S., Chiaravalloti, F. and Procopio, A. (2017), 'Radar–rain-gauge rainfall estimation for hydrological applications in small catchments', *Advances in Geosciences* **44**, 61–66.

Germann, U., Galli, G., Boscacci, M. and Bolliger, M. (2006), 'Radar precipitation measurement in a mountainous region', *Quarterly Journal of the Royal Meteorological Society* **132**(618), 1669–1692.

Golding, B. (2009), 'Uncertainty propagation in a london flood simulation', *Journal of Flood Risk Management* **2**(1), 2–15.

Goovaerts, P. (2000), 'Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall', *Journal of hydrology* **228**(1-2), 113–129.

Goudenhoofdt, E. and Delobbe, L. (2009), 'Evaluation of radar-gauge merging methods for quantitative precipitation estimates', *Hydrology and Earth System Sciences* **13**(2), 195–203.

Gourley, J. J., Maddox, R. A., Howard, K. W. and Burgess, D. W. (2002), 'An exploratory multisensor technique for quantitative estimation of stratiform rainfall', *Journal of Hydrometeorology* **3**(2), 166–180.

Gunn, K. L. S. and Marshall, J. S. (1955), 'The effect of wind shear on falling precipitation', *Journal of Meteorology* **12**(4), 339–349.

Habib, E. H., Meselhe, E. A. and Aduvala, A. V. (2008), 'Effect of local errors of tipping-bucket rain gauges on rainfall-runoff simulations', *Journal of Hydrologic Engineering* **13**(6), 488–496.

Hall, F. (1952), 'Status of possibilities of artificial precipitation', *Eos, Transactions American Geophysical Union* **33**(6), 866–870.

Hamidi, A., Devineni, N., Booth, J. F., Hosten, A., Ferraro, R. R. and Khanbilvardi, R. (2017), 'Classifying urban rainfall extremes using weather radar data: An application to the greater new york area', *Journal of Hydrometeorology* **18**(3), 611–623.

Harrison, D., Driscoll, S. and Kitchen, M. (2000), 'Improving precipitation estimates from weather radar using quality control and correction techniques', *Meteorological Applications* **7**(2), 135–144.

Hertz, H. and Kelvin, W. T. B. (1893), *Electric waves*, Macmillan London.

Hitschfeld, W. and Bordan, J. (1954), 'Errors inherent in the radar measurement of rainfall at attenuating wavelengths', *Journal of Meteorology* **11**(1), 58–67.

Houlding, S. (2000), *Practical geostatistics: modeling and spatial analysis. Manual*, Vol. 1, Springer Science & Business Media.

Jager, E. D. (2017), 'Climate data'.
URL: *http://www.weathersa.co.za./climate center/*

Jakubicka, T., Vos, F., Phalkey, R., Guha-Sapir, D. and Marx, M. (2010), Health impacts of floods in europe: data gaps and information needs from a spatial perspective, Technical report, Centre for Research on the Epidemiology of Disasters (CRED).

Jayakrishnan, R., Srinivasan, R. and Arnold, J. (2004), 'Comparison of raingage and wsr-88d stage iii precipitation data over the texas-gulf basin', *Journal of Hydrology* **292**(1-4), 135–152.

Jewell, S. A. and Gaussiat, N. (2015), 'An assessment of kriging-based rain-gauge–radar merging techniques', *Quarterly Journal of the Royal Meteorological Society* **141**(691), 2300–2313.

Kitanidis, P. K. (1997), *Introduction to geostatistics: applications in hydrogeology*, Cambridge University Press.

Kitchen, M. and Blackall, R. (1992), 'Representativeness errors in comparisons between radar and gauge measurements of rainfall', *Journal of Hydrology* **134**(1-4), 13–33.

Kneis, D., Abon, C., Bronstert, A. and Heistermann, M. (2017), 'Verification of short-term runoff forecasts for a small philippine basin (marikina)', *Hydrological sciences journal* **62**(2), 205–216.

Kouwen, N. (1988), 'Watflood: a micro-computer based flood forecasting system based on real-time weather radar', *Canadian Water Resources Journal* **13**(1), 62–77.

Krajewski, W. F., Villarini, G. and Smith, J. A. (2010), 'Radar-rainfall uncertainties: Where are we after thirty years of effort?', *Bulletin of the American Meteorological Society* **91**(1), 87–94.

Krige, D. G. (1951), 'A statistical approach to some basic mine valuation problems on the witwatersrand', *Journal of the Southern African Institute of Mining and Metallurgy* **52**(6), 119–139.

Kurri, M. and Huuskonen, A. (2008), 'Measurements of the transmission loss of a radome at different rain intensities', *Journal of Atmospheric and Oceanic Technology* **25**(9), 1590–1599.

Lack, S. A. and Fox, N. I. (2005), 'Errors in surface rainfall rates retrieved from radar due to wind-drift', *Atmospheric Science Letters* **6**(1), 71–77.

Lack, S. A. and Fox, N. I. (2007), 'An examination of the effect of wind-drift on radar-derived surface rainfall estimations', *Atmospheric research* **85**(2), 217–229.

Lark, R. (2000), 'A comparison of some robust estimators of the variogram for use in soil survey', *European Journal of soil science* **51**(1), 137–157.

Lee, G. and Zawadzki, I. (2006), 'Radar calibration by gage, disdrometer, and polarimetry: Theoretical limit caused by the variability of drop size distribution and application to fast scanning operational radar data', *Journal of Hydrology* **328**(1-2), 83–97.

Li, J. and Heap, A. (2008), 'Spatial interpolation methods: a review for environmental scientists', *Geoscience Australia, Record. Geoscience Australia, Canberra* .

Li, J. and Heap, A. D. (2011), 'A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors', *Ecological Informatics* **6**(3-4), 228–241.

Lin, X., Kidd, C., Tao, J. and Barros, A. (2016), 'Comparisons of rain estimates from ground radar and satellite over mountainous regions'.

Liptser, R. S. and Shiryaev, A. N. (2013), *Statistics of random processes: I. General theory*, Vol. 5, Springer Science & Business Media.

Lloyd, C. (2005), 'Assessing the effect of integrating elevation data into the estimation of monthly precipitation in great britain', *Journal of Hydrology* **308**(1-4), 128–150.

Ly, S., Charles, C. and Degre, A. (2011), 'Geostatistical interpolation of daily rainfall at catchment scale: the use of several variogram models in the Ourthe and Ambleve catchments, belgium', *Hydrology and Earth System Sciences* **15**(7), 2259–2274.

Lynch, S. (2004), *Development of a raster database of annual, monthly and daily rainfall for Southern Africa: Report to the water research commission*, Water Research Commission.

MacQueen, J. et al. (1967), Some methods for classification and analysis of multivariate observations, *in* 'Proceedings of the fifth Berkeley symposium on mathematical statistics and probability', Vol. 1, Oakland, CA, USA, pp. 281–297.

Mair, A. and Fares, A. (2010), 'Comparison of rainfall interpolation methods in a mountainous region of a tropical island', *Journal of Hydrologic Engineering* **16**(4), 371–383.

Marshall, J. S. and Palmer, W. M. K. (1948), 'The distribution of raindrops with size', *Journal of meteorology* **5**(4), 165–166.

Masih, I., Maskey, S., Mussá, F. E. F. and Trambauer, P. (2014), 'A review of droughts on the african continent: a geospatial and long-term perspective', *Hydrology and Earth System Sciences* **18**(9), 3635–3649.
URL: *https://www.hydrol-earth-syst-sci.net/18/3635/2014/*

Matheron, G. (1965), Les variables régionalisées et leur estimation, PhD thesis.

Matheron, G. (1971), 'The theory of regionalised variables and its applications', *Les Cahiers du Centre de Morphologie Mathématique* **5**, 212.

McClure, M. and Howell, T. (2013), 'Forecast failure: How flood warnings came too late for southern albertans', *Calgary Herald* .

McKee, J. L. (2015), 'Evaluation of gauge-radar merging methods for quantitative precipitation estimation in hydrology: a case study in the upper thames river basin'.

McKee, J. L. and Binns, A. D. (2016), 'A review of gauge–radar merging methods for quantitative precipitation estimation in hydrology', *Canadian Water Resources Journal/Revue canadienne des ressources hydriques* **41**(1-2), 186–203.

McMillan, H., Jackson, B., Clark, M., Kavetski, D. and Woods, R. (2011), 'Rainfall uncertainty in hydrological modelling: An evaluation of multiplicative error models', *Journal of Hydrology* **400**(1-2), 83–94.

Meischner, P. (2005), *Weather radar: principles and advanced applications*, Springer Science & Business Media.

Michelson, D. and Koistinen, J. (2000), 'Gauge-radar network adjustment for the baltic sea experiment', *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere* **25**(10-12), 915–920.

Milrad, S. M., Gyakum, J. R. and Atallah, E. H. (2015), 'A meteorological analysis of the 2013 alberta flood: Antecedent large-scale flow pattern and synoptic–dynamic characteristics', *Monthly Weather Review* **143**(7), 2817–2841.

Mittermaier, M. P. and Terblanche, D. E. (1997), 'Converting weather radar data to cartesian space: A new approach using displace averaging', *WATER SA-PRETORIA-* **23**, 45–50.

Monyela, B. M. (2017), A two-year long drought in summer 2014/2015 and 2015/2016 over South Africa, PhD thesis, University of Cape Town.

Moral, F. J. (2010), 'Comparison of different geostatistical approaches to map climate variables: application to precipitation', *International Journal of Climatology: A Journal of the Royal Meteorological Society* **30**(4), 620–631.

Mukhopadhaya, S. (2016), 'Rainfall mapping using ordinary kriging technique: case study: Tunisia', *J. of Basic and Applied Engineering Res* **3**(1), 1–5.

Neary, V., Habib, E. and Fleming, M. (2004), 'Hydrologic modeling with nexrad precipitation in middle tennessee', *Journal of Hydrologic Engineering* **9**(5), 339–349.

Ngoepe, S. M. (2016), Conditioning Tropical Rainfall Measuring Mission (TRMM) data using ground based rainfall data., PhD thesis.

Niu, S., Jia, X., Sang, J., Liu, X., Lu, C. and Liu, Y. (2010), 'Distributions of raindrop sizes and fall velocities in a semiarid plateau climate: Convective versus stratiform rains', *Journal of Applied Meteorology and Climatology* **49**(4), 632–645.

Oliver, M. and Webster, R. (2014), 'A tutorial guide to geostatistics: Computing and modelling variograms and kriging', *Catena* **113**, 56–69.

Otieno, H., Yang, J., Liu, W. and Han, D. (2014), 'Influence of rain gauge density on interpolation method selection', *Journal of Hydrologic Engineering* **19**(11), 04014024.

Otto, T. (2012), A short course on radar meteorology.

Park, N.-W., Kyriakidis, P. C. and Hong, S. (2017), 'Geostatistical integration of coarse resolution satellite precipitation products and rain gauge data to map precipitation at fine spatial resolutions', *Remote Sensing* **9**(3), 255.

Pegram, G. G. S. (2003), *Spatial Interpolation and Mapping of Rainfall (SIMAR).: Data Merging for Rainfall Map Production*, Water Research Commission (WRC) Report No 1153/1/04 ISBN 1-77005-159-7.

Pegram, G. G. S., Sinclair, S. and Wesson, S. (2006), *Daily rainfall mapping over South Africa: Modelling*, Water Research Commission (WRC) Report No K5/1425.

Peleg, N., Marra, F., Fatichi, S., Paschalis, A., Molnar, P. and Burlando, P. (2018), 'Spatial variability of extreme rainfall at radar subpixel scale', *Journal of Hydrology* **556**, 922–933.

Pettazzi, A. and Salsón, S. (2012), Combining radar and rain gauges rainfall estimates using conditional merging: a case study, *in* 'The Seventh European Conference on Radar in Meteorology and Hydrology', MeteoGalicia, Galician Weather Service Santiago de Compostela, Spain, p. 5.

Prudhomme, C. and Reed, D. W. (1998), 'Relationships between extreme daily precipitation and topography in a mountainous region: a case study in scotland', *International Journal of Climatology* **18**(13), 1439–1453.

Rabiei, E. and Haberlandt, U. (2015), 'Applying bias correction for merging rain gauge and radar data', *Journal of Hydrology* **522**, 544–557.

Rae, K. (2008), 'Personal communication, chief forecaster, national forecast centre'.

Rinehart, R. E. (1991), *Radar for meteorologists*, University of North Dakota, Office of the President.

Ryde, J. (1941), 'Echo intensities and attenuation due to clouds, rain, hail, sand and dust storms at centimetre wavelengths', *Report* **7831**, 22–24.

Sanchez-Moreno, J. F., Mannaerts, C. M. and Jetten, V. (2014), 'Influence of topography on rainfall variability in santiago island, cape verde', *International journal of climatology* **34**(4), 1081–1097.

Schilling, W. and Fuchs, L. (1986), 'Errors in stormwater modeling—a quantitative assessment', *Journal of Hydraulic Engineering* **112**(2), 111–123.

ŞEN, Z. and Eljadid, A. G. (1999), 'Rainfall distribution function for libya and rainfall prediction', *Hydrological sciences journal* **44**(5), 665–680.

Seo, B.-C., Dolan, B., Krajewski, W. F., Rutledge, S. A. and Petersen, W. (2015), 'Comparison of single-and dual-polarization–based rainfall estimates using nexrad data for the nasa iowa flood studies project', *Journal of Hydrometeorology* **16**(4), 1658–1675.

Seo, B.-C., Krajewski, W. F., Kruger, A., Domaszczynski, P., Smith, J. A. and Steiner, M. (2011), 'Radar-rainfall estimation algorithms of hydro-nexrad', *Journal of Hydroinformatics* **13**(2), 277–291.

Seo, B.-C., Krajewski, W. F. and Mishra, K. V. (2015), 'Using the new dual-polarimetric capability of wsr-88d to eliminate anomalous propagation and wind turbine effects in radar-rainfall', *Atmospheric Research* **153**, 296–309.

Sinclair, S. and Pegram, G. (2005), 'Combining radar and rain gauge rainfall estimates using conditional merging', *Atmospheric Science Letters* **6**(1), 19–22.

Smith, J. A., Baeck, M. L., Meierdiercks, K. L., Miller, A. J. and Krajewski, W. F. (2007), 'Radar rainfall estimation for flash flood forecasting in small urban watersheds', *Advances in Water Resources* **30**(10), 2087–2097.

Smith, J. A., Hui, E., Steiner, M., Baeck, M. L., Krajewski, W. F. and Ntelekos, A. A. (2009), 'Variability of rainfall rate and raindrop size distributions in heavy rain', *Water Resources Research* **45**(4).

Smith, J. A., Seo, D. J., Baeck, M. L. and Hudlow, M. D. (1996), 'An intercomparison study of nexrad precipitation estimates', *Water Resources Research* **32**(7), 2035–2045.

Soenario, I., Plieger, M. and Sluiter, R. (2010), *Optimization of rainfall interpolation*, Royal Dutch Meteorological Institute Dutch.

Steiner, M. and Smith, J. A. (2002), 'Use of three-dimensional reflectivity structure for automated detection and removal of nonprecipitating echoes in radar data', *Journal of Atmospheric and Oceanic Technology* **19**(5), 673–686.

Stellman, K. M., Fuelberg, H. E., Garza, R. and Mullusky, M. (2001), 'An examination of radar and rain gauge–derived mean areal precipitation over georgia watersheds', *Weather and Forecasting* **16**(1), 133–144.

Strangeways, I. (2004), 'Improving precipitation measurement', *International Journal of Climatology: A Journal of the Royal Meteorological Society* **24**(11), 1443–1460.

Strangeways, I. (2006), *Precipitation: theory, measurement and distribution*, Cambridge University Press.

Suzana, R. and Wardah, T. (2011), Radar hydrology: New z/r relationships for klang river basin, malaysia, *in* 'International Conference on Environment Science and Engineering. IPCBEE', Vol. 8.

Tabary, P. (2007), 'The new french operational radar rainfall product. part i: Methodology', *Weather and forecasting* **22**(3), 393–408.

Takeuchi, K. (2001), 'Increasing vulnerability to extreme floods and societal needs of hydrological forecasting', *Hydrological Sciences Journal* **46**(6), 869–881.

Terblanche, D. E. (1996), 'A simple digital signal processing method to simulate linear and quadratic responses from a radar's logarithmic receiver', *Journal of Atmospheric and Oceanic Technology* **13**(2), 533–538.

Terblanche, D., Pegram, G. and Mittermaier, M. (2001), 'The development of weather radar as a research and operational tool for hydrology in south africa', *Journal of Hydrology* **241**(1-2), 3–25.

Thomson, M., Abayomi, K., Barnston, A. G., Levy, M. and Dilley, M. (2003), 'El niño and drought in southern africa', *The Lancet* **361**(9355), 437–438.

Uijlenhoet, R. (2001), 'Raindrop size distributions and radar reflectivity–rain rate relationships for radar hydrology', *Hydrology and Earth System Sciences* **5**(4), 615–628.

Uijlenhoet, R. and Berne, A. (2008), 'Stochastic simulation experiment to assess radar rainfall retrieval uncertainties associated with attenuation and its correction', *Hydrology and Earth System Sciences* **12**(2), 587–601.

Ulbrich, C. W. and Atlas, D. (2002), 'On the separation of tropical convective and stratiform rains', *Journal of Applied Meteorology* **41**(2), 188–195.

Van de Beek, C., Leijnse, H., Stricker, J., Uijlenhoet, R. and Russchenberg, H. (2010), 'Performance of high-resolution x-band radar for rainfall measurement in the netherlands', *Hydrology and Earth System Sciences* **14**(2), 205–221.

Van Heerden, J. and Steyn, P. (1999), *Weather radar measurement of rainfall for hydrological and other purposes*, Water Research Commission.

Vehvilainen, B., Cauwenberghs, M., Cheze, J., Jurczyk, A., Moore, R., Olsson, J., Salek, M. and Szturc, J. (2004), 'Evaluation of operational flow forecasting systems that use weather radar', *COST717, Working group* **1**.

Villarini, G. and Krajewski, W. F. (2010), 'Review of the different sources of uncertainty in single polarization radar-based estimates of rainfall', *Surveys in Geophysics* **31**(1), 107–129.

Volkmann, T. H., Lyon, S. W., Gupta, H. V. and Troch, P. A. (2010), 'Multicriteria design of rain gauge networks for flash flood prediction in semiarid catchments with complex terrain', *Water resources research* **46**(11).

Wackernagel, H. (2014), 'Geostatistics', *Wiley StatsRef: Statistics Reference Online* .
  URL: *http://doi.org/10.1002/9781118445112.stat03434*

Wagner, P. D., Fiener, P., Wilken, F., Kumar, S. and Schneider, K. (2012), 'Comparison and evaluation of spatial interpolation schemes for daily rainfall in data scarce regions', *Journal of Hydrology* **464**, 388–400.

Wang, L.-P., Ochoa-Rodríguez, S., Simões, N. E., Onof, C. and Maksimović, Č. (2013), 'Radar–raingauge data combination techniques: a revision and analysis of their suitability for urban hydrology', *Water science and technology* **68**(4), 737–747.

Webster, R. and Oliver, M. A. (2007), *Geostatistics for environmental scientists*, John Wiley & Sons.

Wilson, J. W. and Brandes, E. A. (1979), 'Radar measurement of rainfall—a summary', *Bulletin of the American Meteorological Society* **60**(9), 1048–1060.

WMO (2008), 'Guide to hydrological practices. volume i: Hydrology–from measurement to hydrological information'.

Xie, H., Zhou, X., Hendrickx, J. M., Vivoni, E. R., Guan, H., Tian, Y. Q. and Small, E. E. (2006), 'Evaluation of nexrad stage iii precipitation data over a semiarid region 1', *JAWRA Journal of the American Water Resources Association* **42**(1), 237–256.

Xu, H., Xu, C.-Y., Chen, H., Zhang, Z. and Li, L. (2013), 'Assessing the influence of rain gauge density and distribution on hydrological model performance in a humid region of china', *Journal of Hydrology* **505**, 1–12.

Xu, W., Zou, Y., Zhang, G. and Linderman, M. (2015), 'A comparison among spatial interpolation techniques for daily rainfall data in sichuan province, china', *International Journal of Climatology* **35**(10), 2898–2907.

Yuter, S. E. and Houze Jr, R. A. (1997), 'Measurements of raindrop size distributions over the pacific warm pool and implications for z–r relations', *Journal of Applied Meteorology* **36**(7), 847–867.

Zawadzki, I. (2006), Sense and nonsense in radar qpe, *in* 'Proc. Fourth European Conf. on Radar in Meteorology and Hydrology', pp. 121–124.

Zhang, J., Howard, K., Langston, C., Kaney, B., Qi, Y., Tang, L., Grams, H., Wang, Y., Cocks, S., Martinaitis, S. et al. (2016), 'Multi-radar multi-sensor (mrms) quantitative precipitation estimation: Initial operating capabilities', *Bulletin of the American Meteorological Society* **97**(4), 621–638.

Zhang, M., Leon, C. d. and Migliaccio, K. (2018), 'Evaluation and comparison of interpolated gauge rainfall data and gridded rainfall data in florida, usa', *Hydrological Sciences Journal* **63**(4), 561–582.

Zhu, D., Peng, D. and Cluckie, I. (2013), 'Statistical analysis of error propagation from radar rainfall to hydrological models', *Hydrology and Earth System Sciences* **17**(4), 1445–1453.

Ziervogel, G., New, M., Archer van Garderen, E., Midgley, G., Taylor, A., Hamann, R., Stuart-Hill, S., Myers, J. and Warburton, M. (2014), 'Climate change impacts and adaptation in south africa', *Wiley Interdisciplinary Reviews: Climate Change* **5**(5), 605–620.

Chapter A

National Weather Radar Network

Table A.1 Characteristics of the Weather radars operated by the National Weather Radar Network (NWRN) (Jager, 2017).

Radar	Type	Latitude (deg)	Longitude (deg)	Altitude (m)	Wave-length (cm)	Beam width (degrees)	Doppler	Dual Polarized
Bethlehem	METEOR 600S	-28.09836	28.16323	1722	10	1	Yes	Yes
Bloemfontein	METEOR 600S	-29.16627	26.05105	1566	10	1	Yes	No
Cape Town	EEC WSR 81	-34.05405	18.38532	905	5	1	Yes	No
De Aar	EEC WSR 76	-30.66475	23.99267	1284	5	1.5	No	No
Durban	METEOR 600S	-29.70723	31.08154	137	10	1	Yes	No
East London	METEOR 600S	-32.75566	27.66160	603	10	1	Yes	No
Ermelo	METEOR 600S	-26.49803	29.98406	1773	10	1	Yes	No
George	METEOR 600S	-34.21950	21.78264	236	10	1	Yes	No
Irene	METEOR 600S	-25.91193	28.21071	1532	10	1	Yes	No
Ottosdal	METEOR 600S	-26.73519	26.08766	1514	10	1	Yes	No
Polokwane	METEOR 600S	-23.89357	29.50568	1396	10	1	Yes	No
Port Elizabeth	EEC WSR 88D	-33.98466	25.61074	75	5	1	No	No
Skukuza	EEC WSR 76	-24.97395	31.60064	299	10	2	No	No
Umtata	METEOR 600S	-31.53714	28.76446	857	10	1	Yes	No
Cape Town	METEOR 50DX	-33.97152	18.60213	62	3	Yes	Yes	Yes
OR Tambo	METEOR 50DX	-26.13672	28.24111	1 680	3	Yes	Yes	Yes

Chapter B

Python programming codes for OK algorithm and Netcdf4 radar data extraction

Ordinary Kriging algorithm

```python
# -*- coding: utf-8 -*-
"""
Created on Thu Apr 11 12:29:12 2019

@author: Tadianashe .S. Gutsa
"""

from pylab import *
from scipy.spatial.distance import pdist, squareform
import matplotlib.pyplot as plt
import seaborn as sns; sns.set()  # for plot styling
import pandas as pd
import datetime as dt
from scipy.stats import norm
import scipy.stats as stats
from pykrige.ok import OrdinaryKriging
import numpy as np
import pykrige.kriging_tools as kt

#### read in gauge and rainfall data
gauges=pd.read_csv("input/gauges_clean.csv")
gauges=gauges[["gauge_id","easting", "northing"]]
rainfall_complete=pd.read_csv("input/rainfall_data.csv")
rainfall_complete.gauge_id = [c.strip() for c in rainfall_complete.gauge_id]
gauges.gauge_id = [c.strip() for c in gauges.gauge_id]
rainfall_df=pd.merge(left=rainfall_complete, right= gauges)
rainfall_df.to_csv("input/rainfall_df.csv", index=False)
rainfall_df=pd.read_csv("input/rainfall_df.csv")
rainfall_df['date'] = pd.to_datetime(rainfall_df['date'])

#set date of investigation
```

```python
day=rainfall_df[rainfall_df.date=="    "]


#Posting of data to visualize the distribution of observations
Post= plt.figure(1, figsize=(6, 4))
post=plt.scatter(day.easting, day.northing, c =day.variable, cmap = "viridis")
plt.ylabel("")
plt.xlabel("")
plt.title("")
Post.savefig("")# saves the plot


#Perfom Ordinary Kriging  on 100km by 80 km grid
data = np.array(day[["easting","northing","rain"]])


gridx = np.arange(__,__,__ )
gridy = np.arange(__,__,__ )
OK = OrdinaryKriging(data[:, 0], data[:, 1], data[:, 2], variogram_model='  ',
                     verbose=True, enable_plotting=True ,weight = True, nlags = 7)


OK.display_variogram_model()# enables the visualization of the variogram models


#display the kriged rainfall grid as numpy arrays
z, ss = OK.execute('grid', gridx, gridy,backend="loop")


#Save the kriging results on an asc file
kt.write_asc_grid(gridx, gridy, z, filename="kriged.asc")
pd.read_table("kriged.asc", sep= "/t", skiprows = 7)


#plot the rainfall surface of the kriging
kriged_map= plt.figure(1, figsize=(, ))
map_data=pd.read_table("kriged.asc", delim_whitespace= True, skiprows = _)
x = plt.imshow(z, aspect ="auto")
plt.colorbar(x).set_label("mm")
kriged_map.savefig(" ")# saves the resulting map
```

Netcdf data extraction

```python
# -*- coding: utf-8 -*-
"""
Created on Thu Apr 11 14:54:25 2019


@author: Tadiwanashe .S. Gutsa
"""
```

```python
import matplotlib.pyplot as plt
import numpy as np
import netCDF4 as nc4


#import and explore the radar.nc files
dataset = nc4.Dataset("  ")
print(dataset)
dataset.variables.keys()
precipitation= dataset.variables["rain"]
print(precipitation)
precipitation.dimensions
precipitation.shape

#select area of study from radar field
study_area=precp[--:--,--:--]

#plot the region of radar coverage
image = plt.figure(1,figsize=(,))
x = plt.imshow(study_area, aspect ="auto")
image.colorbar(x).set_label("mm")
image.savefig("  ")#
dataset.close()
```