Conditioning Tropical Rainfall Measuring Mission (TRMM) data using ground based rainfall data

by

SIMON MALOSE NGOEPE

Submitted in fulfilment of the academic requirements of

Master of Science in Engineering

School of Civil Engineering

College of Agriculture, Engineering and Science

Howard College

University of KwaZulu-Natal

Durban

South Africa

November 2016

PREFACE

The research contained in this dissertation/thesis was completed by the candidate while based in the Discipline of CIVIL ENGINEERING, School of Civil Engineering of the College of Agriculture, Engineering and Science, University of KwaZulu-Natal, Howard College Campus, Durban, South Africa.

The contents of this work have not been submitted in any form to another university and, except where the work of others is acknowledged in the text, the results reported are due to investigations by the candidate.

____

Signed: Prof. Geoff Pegram

Date: 14 November 2016

DECLARATION 1: PLAGIARISM

Note that two declaration sections are required if there are papers emanating from the dissertation/thesis. The first (obligatory) declaration concerns plagiarism and the second declaration specifies your role in the published papers.

I, Simon Malose Ngoepe, declare that:

(i) the research reported in this dissertation, except where otherwise indicated or acknowledged, is my original work;

(ii) this dissertation has not been submitted in full or in part for any degree or examination to any other university;

(iii) this dissertation does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons;

(iv) this dissertation does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

a) their words have been re-written but the general information attributed to them has been referenced;

b) where their exact words have been used, their writing has been placed inside quotation marks, and referenced;

(v) where I have used material for which publications followed, I have indicated in detail my role in the work;

(vi) this dissertation is primarily a collection of material, prepared by myself, published as journal articles or presented as a poster and oral presentations at conferences. In some cases, additional material has been included;

(vii) this dissertation does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the dissertation and in the References sections.

Signed: Simon Malose Ngoepe

Date: November 14, 2016

DECLARATION 2: PUBLICATIONS

My role in each paper and presentation is indicated. The * indicates the candidate.

Pegram GGS., Sinclair S., Ngoepe S*. Comparing interpolated daily gauge rainfall over RSA with TRMM to determine possible bias correction for hydrological applications. In: Proceedings of South African Society for Animal Science (SASAS) Conference 48th, September 2015.

Pegram, Geoff, András Bárdossy, Scott Sinclair, & Simon Ngoepe* (2015). Accounting for uncertainty in the repair of rain gauge records and their spatial interpolation, *European Geosciences Union General Assembly*, Vienna

Pegram GGS., Sinclair S., Bárdossy A., Ngoepe S*. 2016. New methods of infilling Southern African raingauge records enhanced by Annual, Monthly and Daily Precipitation estimates tagged with uncertainty. Water Research Commission Report No. K5 - 2241

Signed: Simon Malose Ngoepe

Date: November 14, 2016

ABSTRACT

Rainfall helps to structure society in a geographical sense, thus correct capturing of rainfall data and recording is very importantant in ensuring that water resources planners have information that can be used to make informed decisions concerning agriculture and water provision to people, the environment and other industries.With a loss in a number of old reliable rainfall gauges, implementation of new gauges and also missing data, there is a need to evaluate other options like Tropical Rainfall Measuring Mission (TRMM) data. The tropics play an important role in the global hydrological cycle, and tropical rainfall is the critical component of this role. TRMM provides systematic, multi-year, visible, infrared, and microwave estimates of rainfall in the tropics and subtropics as key inputs to weather and climate research. The TRMM satellite orbited around the Earth and it was not sun synchronous. The TRMM science team developed a range of gridded rainfall products; the product used for this research was 3B42RT which is a similar rainfall product to 3B42. Furthermore, TRMM data was selected at the same locations with intent to have the ground based gauge stations measurements compared with TRMM satellite derived precipitation pixel value at the same site. The data considered was from March 1st, 2000 to February 28th, 2010. In the 10 year period, the analysis was for the daily, pentads, monthly and annual data comparisons. The different methods applied for analysing and comparing TRMM and Block Averaged Gauge Data (BAGD) datasets were linear regression, standardization, cross validation and the introduction of quantile-quantile (Q-Q) transform methods. Considering the high variability in time and space of rainfall and that the gauges used to measure BADG are at times sparse, TRMM had a high potential to estimate precipitation relatively accurately over large areas. TRMM pixel values can be used to get information on an area that does not have gauges or is poorly gauged. The research findings indicate that it is likely that TRMM data will be useful for large-scale hydrology and agriculture, particularly at the monthly scale, in contrast with daily.

ACKNOWLEDGMENTS

To God be the glory, as He has provided me with an opportunity and guiding my Principles to make the research possible to complete.

To my supervisor, Professor Geoff Pegram of the School of Civil Engineering at University of KwaZulu Natal; you have kept me on my toes from day one. The door to Prof. Pegram's office was always open whenever I encounter a challenge that is difficult to overcome. I thank you and may the Lord keep your mind fresh like a 25 year old, we still need your wisdom.

I would like to acknowledge the help of Dr Scott Sinclair for lessons in data management and introducing me to Python - without that, this work of managing large data-sets meaningfully would not have been feasible.

Mama, my brothers, sisters and my extended family, I warmly thank and appreciate the support and courage you have provided in all aspects of my life.

May I express my gratitude and deepest appreciation to my girls, Oaratwa and Orefile and you boy, Boitshepo. Daddy is back!

My lovely wife, Molebogeng, your support, understanding, encouragement, and mostly your love especially through difficult at times; you are forever amazing. May God continue to give you the strength to continue to shine.

Amen

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Figure **Page**

LIST OF ACRONYMS

BAGD	Block Averaged Gauge Data
ccc	cross correlation coefficient
cdf	cumulative distribution function
CERES	Clouds and the Earth's Radiant Energy System
cfdf	cumulative frequency distribution function
CSAG	Climate System Analysis Group
DAAC	Goddard Distributed Active Archive Centre
DISC	Data and Information Services Center
ERB	Earth Radiation Budget
GES	Goddard Earth Sciences
GOES	Geostationary Operational Environmental Satellites
GSFC	Goddard Space Flight Centre
HQ	High Quality
IR	infra-red
JAXA	Japan Aerospace Exploration Agency
LIS	Lightning Imaging Sensor
MSc	Master of Science
NASA	National Aeronautics and Space Administration
NetCDF	Network Common Data Form
PMW	passive microwave
POES	Polar Operational Environmental Satellites
PR	Precipitation Radar
Q-Q	Quantile-Quantile
R^2	Coefficient of determination
SAST	South African Standard Time
SAWS	South African Weather Services
SCF	Science Computing Facility
TDRSS	Tracking and Data Relay Satellite System
TMI	TRMM Microwave Imager
TMPA	TRMM multi-satellite precipitation analysis
TRMM	Tropical Rainfall Measuring Mission
TSDIS	TRMM Science Data and Information System
U.S.	United State
UTC	Coordinated Universal Time
VAR	Variable rain-rate
VIRS	Visible Infrared Scanner

CHAPTER 1: INTRODUCTION

In Summary:

The introductory chapter looks at the rationale for the research, with rainfall as one of the most important inputs, rare and variable in water resourece studies. With a loss in a number of old reliable rainfall gauges, implementation of new gauges and also missing data, the need to evaluate other options like Tropical Rainfall Measuring Mission (TRMM) data is key.

1.1 Rationale for the research

South Africa is a water scarce country. Rainfall is a key parameter in hydrology at every scale from the local to the global. However, rainfall is highly variable, and has been more poorly monitored in many areas than is necessary for adequate assessment of amounts and their temporal and spatial variations, for appropriate management of water resources, and for hydrological and related modelling and prediction. Rainfall helps to structure society in a geographical sense. Water is an essential element for life, thus the more water that is relatively available in an environment, the more potential that environment has for sustenance of life.

Correct rainfall data capturing and recording is important to ensure that modellers have information that can be used to make informed decisions. South Africa relies heavily on ground based rainfall data. Ground based rainfall data is data that is observed through rain gauges and radars across the country. The most common method of rainfall data-capture in South Africa is the use of a standard non-recording rain gauge, but nowadays estimation by radar and satellite is practised as well. Nevertheless, the primary source of rainfall data is still provided by the daily non-recording rain gauges. This is because rain gauges are cheap and generally reliable and unbiased. Rain gauge data are also available for long time periods, which is advantageous in many respects.

With a loss in a number of old reliable rain gauges, low implementation of new gauges and unsatisfactory data sharing, there is a need to evaluate other options like the Tropical Rainfall Measuring Mission (TRMM) data to determine how effective this method is compared to ground base data.

TRMM was a joint United State (U.S.).-Japan satellite mission to monitor tropical and subtropical precipitation and to estimate its associated latent heating. The tropics play an important role in the global hydrological cycle, and tropical rainfall is the critical component of this role.

The paper describing the 3B42RT (real-time) and 3B42 (research quality) of the TRMM Multi-satellite Precipitation Analysis (TMPA) products is Huffman et al. (2007). The TRMM science team developed a range of gridded rainfall products, some of their products are 3B42, 3B42RT and 3B43. The 3B43 rainfall product is at a higher spatial and temporal resolution than 3B42 and is a merger of 3B42 and rain gauge data products to form a single rain product, with a 0.25^0 resolution. The 3B42 estimates (computed in monthly blocks) are considered to supersede the 3B42RT estimates (computed in a 30-day sliding window) as each month of the 3B42 data are computed during the following month. The 3B42 processing is designed to maximize data quality, so 3B42 is strongly recommended for any research work not specifically focused on real-time applications.

Satellite estimates provide an average precipitation over an area of the satellite pixel, while the gauge provides measurements at points. Block-averaged ground-based gauge station measurements are compared with the corresponding TRMM satellite derived pixel precipitation value at the same location.

The key question is to find out how whether TRMM data is comparable to ground based rainfall data. Can TRMM data be used to replace or provide a second option to ground based rainfall data and to meaningfully estimate rainfall where there are no gauges? In this research, an updated daily rainfall gauge database which was repaired or patched under a Water Research Commission project number K5/2241 (Pegram et al., 2016), in which the candidate was a partner, is used as the basis for comparison with TRMM rainfall estimates.

Conclusion:

The TRMM satellite monitors tropical and subtropical precipitation and it remains to be seen how well it compares with ground base rainfall gauges. The next chapter expands on what the TRMM mission was and also looks at the different products that the satellite produces and also which will be appropriate for our research.

CHAPTER 2: THE RATIONALE BEHIND THE TROPICAL RAINFALL MEASURING MISSION

In Summary:

This chapter focuses on the review done on the formation and the importance of evaluating TRMM by the two giant countries in weather satellite deployment, Japan and the United States. It looks at the variability of tropical rainfall which affects the lives and economies of more than half of the world's population. We learn how the TRMM satellite relates to the Earth being at a defined altitude. The chapter further introduces the variety of TRMM products, and also how the data was produced and stored. The TRMM data product that was considered in this research project is 3B42RT.

The international cooperation between the two countries Japan and the United States through the Japan Aerospace Exploration Agency (JAXA) and the National Aeronautics and Space Administration (NASA) respectively, allowed them to get together to share their visions. According to JAXA's document titled "JAXA Technology Challenge" published in March 31, 2009, JAXA's vision was to build a secure and prosperous society through the utilisation of aerospace technology while on www.nasa.gov/about/index.html (November 2016), NASA's vision is to reach for new heights and reveal the unknown so that what they do and learn will benefit all human kind. The two organizations combined their minds to bring the first ever joint mission to measure tropical and subtropical rainfall. They named their mission the Tropical Rainfall Measuring Mission (TRMM).

TRMM was launched on November 27, 1997 from the Tanegashima Space Center, on the Japanese H-II vehicle, with continuous science data collection that began in December 8, 1997. The data was archived and distributed by Goddard Distributed Active Archive Centre (DAAC), Science Computing Facility (SCF) and others. Mission operations were terminated in 2015 April, after which the spacecraft re-entered the Earth's atmosphere and burned up in 2015 June (science.nasa.gov).

According to NASA's website, science.nasa.gov, TRMM was a satellite that provided more information both to test and improve models. TRMM was devoted to find out where and how hard it is raining. Not all clouds cause rain, and when rain does fall, it falls through various heights in the atmosphere, sometimes not reaching the ground at all! Because rain is so variable, we cannot know beforehand just how much rain actually falls across the earth, but it is something we need to measure. TRMM accomplished this not just by providing rainfall data, but more importantly, by providing information on heat released into the atmosphere as part of the process that leads to rain generation.

In summary, TRMM was motivated by

- the important role that the tropics play in the global hydrological cycle, and tropical rainfall is the critical component of this role. Three quarters of the atmosphere's heat energy derives from the release of latent heat of condensation in the process of precipitation. Two-thirds of the global precipitation occurs in the tropics (NASA).
- the variability of tropical rainfall which affects the lives and economies of more than half of the world's population.

TRMM provided systematic, multi-year, visible, infrared and microwave measurements of rainfall in the tropics as key inputs to weather and climate research. The satellite observations were complemented by ground radar and rain gauge measurements to validate the satellite rain estimation techniques.

The key objectives on the science behind TRMM were:

- to obtain and study multi-year science data sets of tropical and subtropical rainfall measurements.
- to understand how interactions between the ocean, air, and land masses produce changes in global rainfall and climate.
- to improve modelling of tropical rainfall processes and their influence on global circulation in order to predict rainfall and its variability at various space and time scales.
- to test, evaluate, and improve satellite rainfall measurement techniques.

The NASA Goddard Space Flight Centre (GSFC), provides three rain measuring instruments with their data archived at the Goddard DAAC, and two other products with their data not archived at the Goddard DAAC while JAXA brings in one product. The TRMM multi-satellite precipitation analysis (3B42*/TMPA) product will continue to produce (https://pmm.nasa.gov/data-access/downloads/trmm) data for the 3-hourly (3b42), daily (3b42) and monthly (3b43) through early 2018.

The three NASA GSFC instruments used are specified and discussed in brief below:

1. TRMM Microwave Imager (TMI), that was designed to

    - provide quantitative rainfall information over a wide swath under the TRMM satellite,
    - consume little power,

so by combining these two outputs, the TMI is the "workhorse" of the rain measuring package on TRMM.

2. the Visible Infrared Scanner (VIRS),

- senses radiation coming up from the Earth in five spectral ranges,
- has the ability to delineate rainfall,
- serves as a transfer standard to other measurements that are made routinely using Polar Operational Environmental Satellites (POES) and Geostationary Operational Environmental Satellites (GOES) satellites. POES is the satellite system that offers the advantage of daily global coverage, by making nearly polar orbits 14 times per day approximately 520 miles above the surface of the Earth while GOES is a collection of Geostationary Operational Environmental Satellites circling the Earth in a geosynchronous orbit, allowing them to hover continuously over one position on the surface; the geosynchronous surface is about 35 888 km above the Earth, high enough to allow the satellites an almost full-disc view as Earth radius is 6 371 km.

3. the Tracking and Data Relay Satellite System (TDRSS),

- is a network of communications satellites and ground stations used by NASA for space communications,
- is designed to replace an existing network of ground stations that had supported all NASA's manned flight missions,
- determines the orbital position of the spacecraft,

The two instruments not having their data archived at DAAC, are

1. the Clouds and the Earth's Radiant Energy System (CERES), that

- was inherited from NASA's Earth Radiation Budget Experiment (ERB),
- measure the energy at the top of the atmosphere, as well as estimate energy levels in the atmosphere and at the Earth's surface,

2. the Lightning Imaging Sensor (LIS), with the aim of

- detecting and locating lighting over the tropical region of the globe,
- observing the distribution and variability of lightning over the Earth,
- providing information that could lead to future advanced geostationary lightning sensors capable of significantly improving weather "nowcasting".

The product supplied by JAXA was the Precipitation Radar (PR), which

- provided three-dimensional maps of storm structure,
- measured the same variables as the ground-based systems, i.e. the intensity and distribution of the rain, the rain type, the storm depth and the melting layer, i.e. the height at which snow melts into rain,
- was proven to be consistent with calibration accuracy.

TRMM orbited around the Earth and it was not sun synchronous. It was first placed at an altitude of 350 km with an inclination of 35 degrees to the Equator, and later, in August 2001 was moved to a higher orbit at an altitude of 402.5 km. The reason for moving the satellite was to conserve fuel through less drag and thereby extend its life.

The two characteristics of the satellite orbit that can be used to understand the behaviour of the satellite are the time it takes to orbit the earth once and also the longitudinal shift with each orbit. For the spacecraft to complete one orbit around the Earth, it required about 91 minutes, two orbits in three hours, eight orbits in 12 hours and 16 orbits each day. The longitudinal shift of TRMM's orbit, at an altitude of 350 km, was approximately -23.3^0. Below is Figure 1 showing an image of a sequential set of TRMM data swaths.

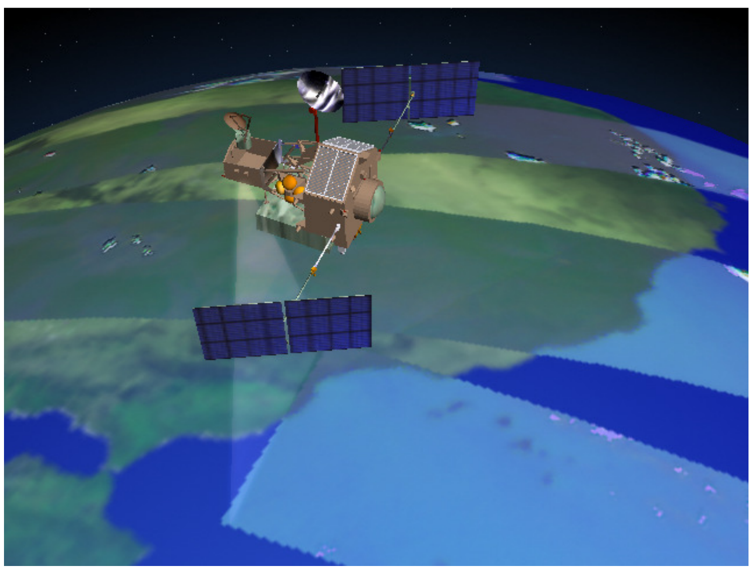


Figure 1: An image showing a sequential set of TRMM data swaths on the globe collected by the satellite

At the equator, the change in local time from orbit to orbit for TRMM, was about -0.033 hours (h), thus the local time at each equatorial crossing was 0.033 h early. For the satellite to re-cross the equator at the same location, 732 orbits (46.4 days) at local time were required. The local separation time between the descending and ascending nodes was about 12 hours. It took about 23 days for the ascending and descending branches of the orbit together to cover a diurnal cycle near the equator. In the later period of TRMM, when the satellite was moved to an altitude of 402.5 km, it took 47.5 days instead of 46.4 days to revisit a location.

These orbits provide extensive coverage in the tropics and allowed each location to be covered at a different local time each day, as in Figure 2 that displays a Map of TRMM's low Earth orbit swaths. The data for each orbit were stored on board and transmitted to the ground via the TDRSS, which was processed by TRMM Science Data and Information System (TSDIS) into a suite of standard products and the processed data transferred to Goddard DAAC for archival and distribution.

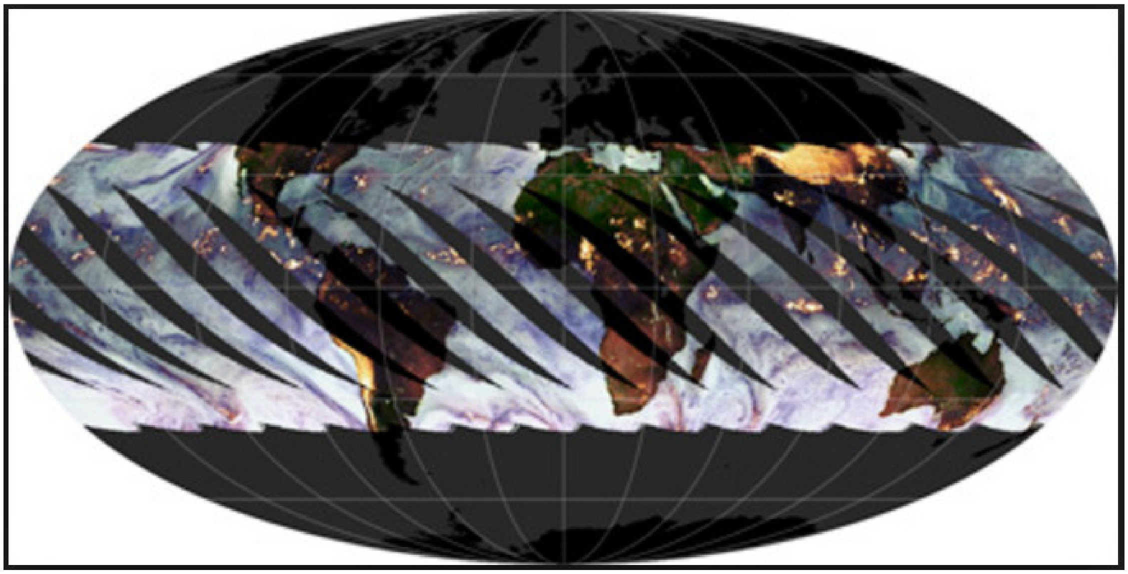


Figure 2: Map showing TRMM's low Earth orbit at an inclination of 35 degrees

TRMM Multi-satellite Precipitation Analysis (TMPA; Huffman **et al.** 2007) is the generic name for a suite of TRMM precipitation estimates also known as 3B4X and 3B4XRT. The main product (3B42) is a scaled version of the real-time product (known as 3B42RT) which is a combination of the passive microwave (PMW) High Quality (HQ) product (also known as 3B40) and the PMW-calibrated geostationary infra-red (IR) Variable rain-rate (VAR) product (also known as 3B41).

TRMM data products are archived at and distributed from the NASA Goddard Earth Sciences (GES) Data and Information Services Center (DISC). The GES DISC provides free, quasi-global archived and near-real-time precipitation products and services for research and applications, and it has collected and provided the services since 1997. The data products are organized as the following three categories:

- orbital products (also known as swath products);
- gridded products; and
- other TRMM-related products, consisting of TRMM ancillary products, ground-based instrument products, TRMM and ground observation subsets, and field experiment products.

TRMM science team developed a range of gridded rainfall products, some of their products are 3B42, 3B42RT and 3B43, as mentioned above. The 3B42 rainfall product is a Multi-satellite precipitation dataset calibrated by TRMM PR/TMI (Combined), 3-hours, 0.25^0 resolution, while 3B42RT is a similar rainfall product to 3B42 that is near real time that merges multi-satellite radiometer and radiometer adjusted IR precipitation data, every 3-hours with a 0.25^0 resolution. The 3B43 rainfall product is at a higher spatial and temporal resolution than 3B42 and is a merger of 3B42 and rain gauge data products to form a single rain product, with a 0.25^0 resolution.

The data product that is considered in this research project is 3B42RT as it is being used in several applications within South Africa, so is readily available.

On October 09, 2014, the TRMM satellite was reported by NASA's website that it had begun to descend. TRMM PR orbit number 96230's, the last data production available to the public, was in October 7, 2014. It was also stated that the TMI data would continue to be gathered and made available to the public during the descent of the spacecraft until it reached its decommissioning altitude of 335 km. By February 12, 2015, TRMM had descended to an altitude of around 350 km with a TRMM PR orbit number 96231's data being distributed for experimental operation period of 40 days. TRMM spacecraft re-entered the Earth's atmosphere on June 15, 2015, over the South Indian Ocean.

In this research project the results of a comparison between rainfall derived from the TRMM 3B42RT product and block averaged rainfall determined from daily rain gauges are presented. It is shown that while there are some similarities in the gross behaviour of the two rainfall estimates, there are also some significant differences which require a careful treatment of TRMM data before using it (Pegram et al. 2016).

Conclusion:

TRMM provides systematic, multi-year, visible, infrared, and microwave estimates of rainfall in the tropics and subtropics as key inputs to weather and climate research. The TRMM satellite orbited around the Earth and it was not sun synchronous. The data product that was considered in this research project is 3B42RT as it is freely available and used in several applications within South Africa. The next chapter introduces the four selected sites within South Africa where a comparison is made between the data from both TRMM and the ground based datasets.

CHAPTER 3: SITES AND DATA SETS

In Summary:

Four study areas were identified with different geographical characteristics which also have dense groups of rainfall gauges within South Africa. The areas were in the Gauteng, KwaZulu-Natal, Limpopo and Western Cape Provinces. The Climate System Analysis Group (CSAG) rainfall database was used to extract ground based rainfall data for the four identified sites. The period considered for the research was from March 2000 until February 2010 to maximise the overlap of the data-sets. In each of the four selected areas, sets of nine 0.25^0 blocks were identified that had the most usable data in terms of data reliability, missing gaps and number of stations located in each block. It is also noted how the two datasets behave, more especially the poor performance of TRMM in the coastal or mountainous regions and the sparseness of rain gauges.

3.1 Introduction

In South Africa, rainfall varies across the country, with the Western Cape having winter and some summer rainfall, while the rest of the country experiences mostly summer rainfall. Historically, the highest rainfall measured in one year was at Jonkershoek, Western Cape Province in 1950 at 3874 mm, the highest monthly rainfall record measured in Matiwa, Limpopo Province in 1958 at 1510 mm, while the highest 24 hours rainfall record was measured in St Lucia, KwaZulu-Natal Province on January 31, 1984 at 597 mm (www.weathersa.co.za), due to tropical cyclone Domoina. The Gauteng Province has the densest population of rain gauges as compared to the rest of the provinces.

In this research, four study areas were identified that have relatively dense groups of rain gauges. The areas are in Gauteng, KwaZulu-Natal, Limpopo and Western Cape Provinces. Both KwaZulu-Natal and Western Cape Provinces are coastal areas, see Figure 3.   A multiquadric interpolation technique developed by Pegram and Pegram (1993) was used to identify active rain gauges and their associated weights to be used in calculating average depth weight over rainfall block. This technique, written in FORTRAN programming language code, was wrapped in a Python package interface to make it more convenient to use in conjunction with the Python based workflow. The configuration of gauges on a day within a given block may change from day to day (depending on missing records). Since the active gauge configuration defines the weights, it was necessary to separately compute the weights for each wetted TRMM block on each day of the 10 year analysis period over RSA. A rainfall database archived by the Climate System Analysis Group (CSAG) based at the University of Cape Town, South Africa was used to extract ground based rainfall data for the four identified sites. The multiquadric interpolation technique provides a map with the number of stations and weighted daily average per block.

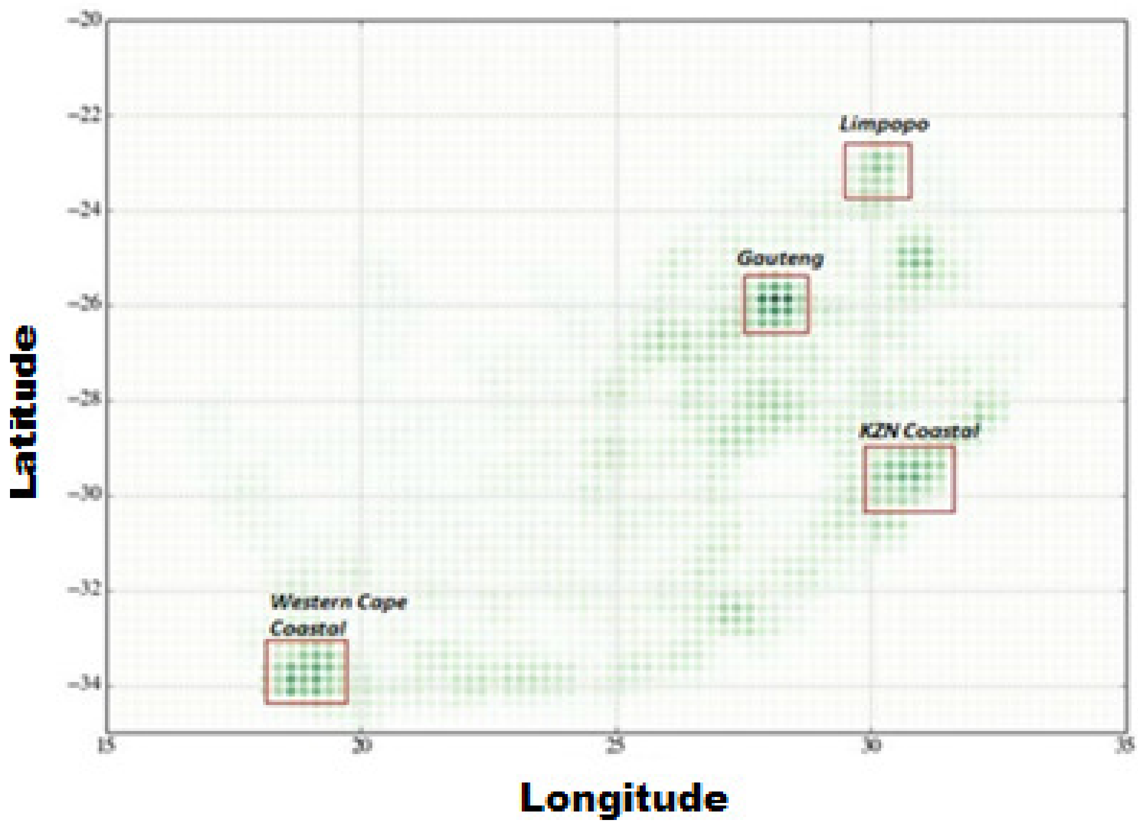


Figure 3: Layout of the 4 study sites in South Africa within red square boxes: Limpopo, Gauteng, KZN Coastal and Western Cape Coastal

In each area, nine ¼ degree blocks were identified that have the most usable data in terms of data reliability, missing gaps and number of stations located in each block. Figure 3 reflects the number of stations that were available to match TRMM data, from March 2000 until February 2010, where the darkness of the circles centred in the ¼ of degree blocks indicates the relative number of gauges available to be averaged.

3.1.1 Site 1: Gauteng

Of the nine blocks selected for Site 1 in Gauteng, one of the blocks, with coordinates 27.75° to 28.50° E by -25.50° to -26.25° N, had the highest number of gauges in the country at 73. Table 1 below presents the central position in degrees of each of the 9 blocks used in Site 1, with Figure 4 showing the number of gauges in each block.

Table 1: Gauteng selected blocks' centres in Longitude and Latitude

Block	Long	Lat
1	27.88	-26.13
2	28.13	-26.13
3	28.38	-26.13
4	27.88	-25.88
5	28.13	-25.88
6	28.38	-25.88
7	27.88	-25.63
8	28.13	-25.63
9	28.38	-25.63

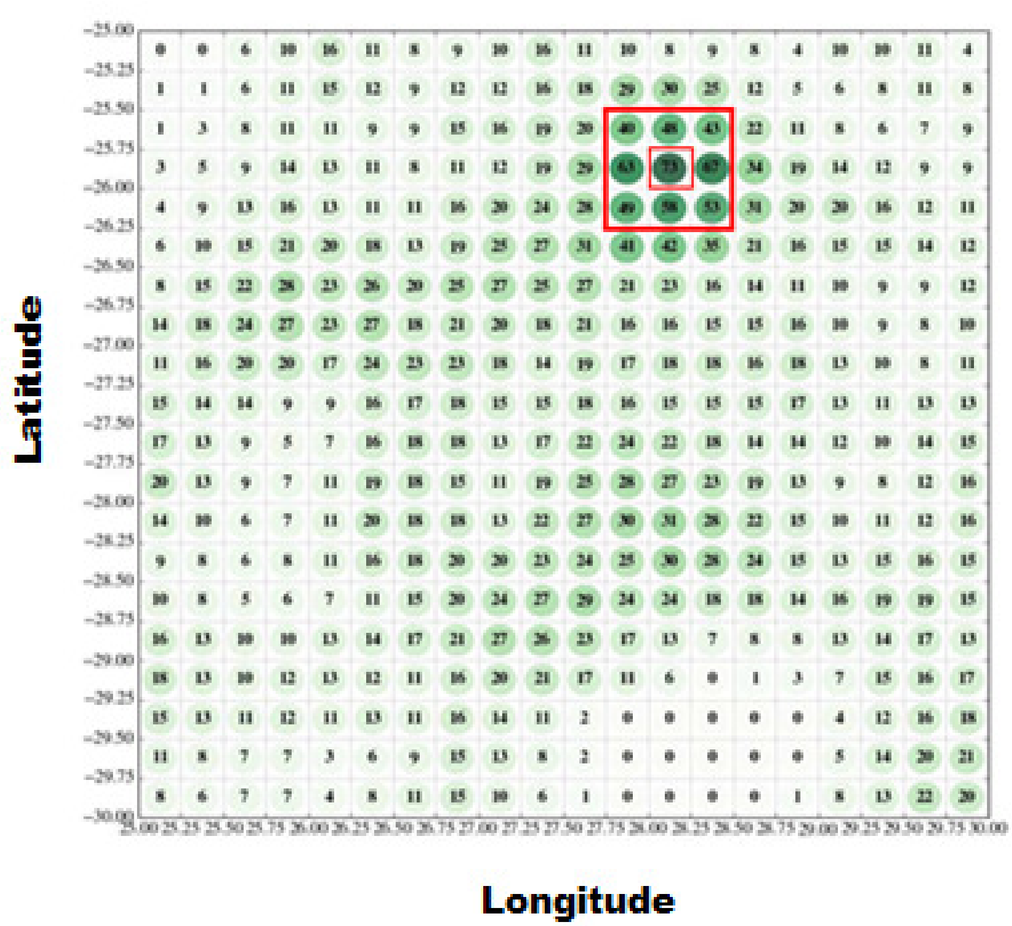


Figure 4: Layout of the Gauteng study site with 9 blocks highlighted in the outer red square box. The red box in the middle reflects the highest number of gauges at 73

3.1.2 Site 2: KwaZulu-Natal Coast

The nine blocks, selected for the most active gauges in the KwaZulu-Natal coastal area, had their central position located as per the Table 2 below:

Table 2: KwaZulu-Natal selected blocks' centres in Longitude and Latitude

Block	Long	Lat
1	31.13	-29.63
2	30.38	-29.38
3	30.88	-29.88
4	32.38	-28.38
5	30.38	-29.88
6	30.13	-28.13
7	32.13	-28.38
8	30.63	-29.63
9	31.88	-28.38

The highest number in one block for the area in Figure 5 is 11.

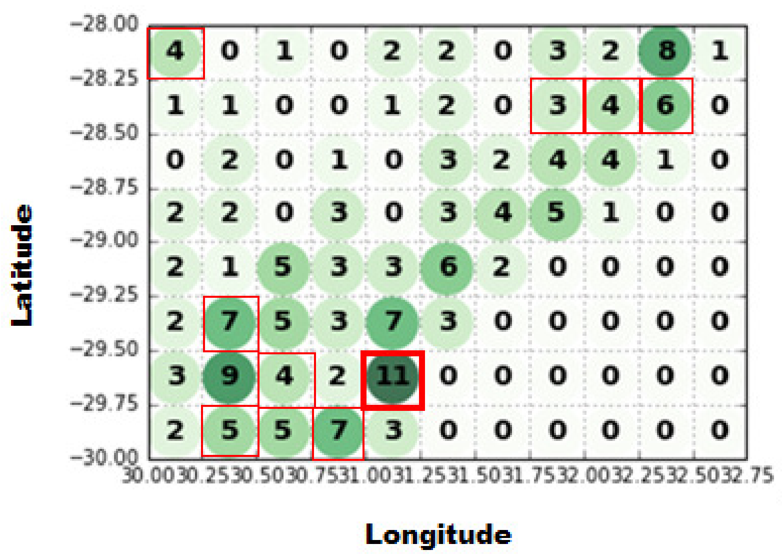


Figure 5: Layout of the KwaZulu-Natal study site with the 9 selected blocks highlighted in red squares, with the highest count in the heavier red square

3.1.3 Site 3: Western Cape Coastal

The selected blocks' central positions (longitude and latitude) for the Western Cape Coastal are present in a Table 3 and Figure 6. The block with 16 gauges in one block is the highest number in any block. The block was not used as it failed the selection criteria because of the number of missing daily data.

Table 3: Western Cape selected blocks' centres in Longitude and Latitude

Block	Long	Lat
1	18.38	-34.13
2	19.13	-34.13
3	18.63	-33.38
4	18.88	-32.13
5	18.88	-32.38
6	18.88	-32.88
7	19.88	-33.88
8	18.38	-32.13
9	18.63	-32.13

The blocks in Table 3 above are highlighted in red squares in Figure 4 below

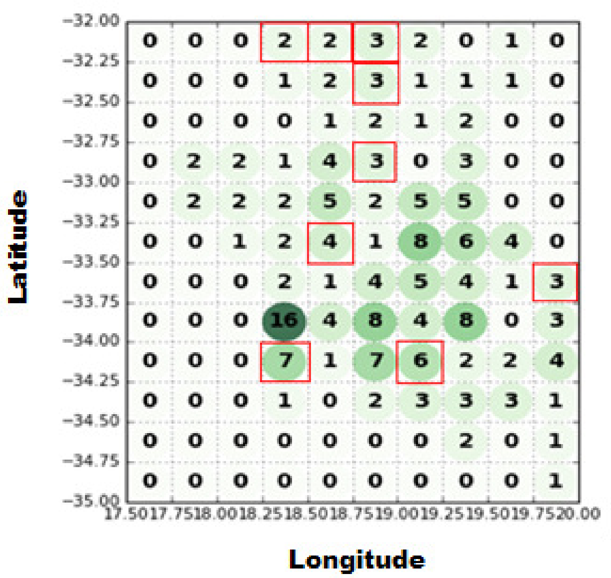


Figure 6: Layout of the Western Cape study site with the 9 selected blocks highlighted in red squares

3.1.4 Site 4: Limpopo

The set of nine blocks selected in Limpopo have a block with the highest number of 6 6 gauges. The centre positions of the blocks and the image are presented below in Table 4 and Figure 7, respectively. The block with 10 gauges was ignored because many of them had too many missing days.

Table 4: Limpopo selected blocks' centres in Longitude and Latitude

Block	Long	Lat
1	30.13	-23.63
2	29.88	-23.38
3	27.63	-23.63
4	31.38	-23.63
5	28.63	-22.88
6	29.63	-23.13
7	29.38	-23.88
8	30.63	-23.88
9	29.88	-23.13

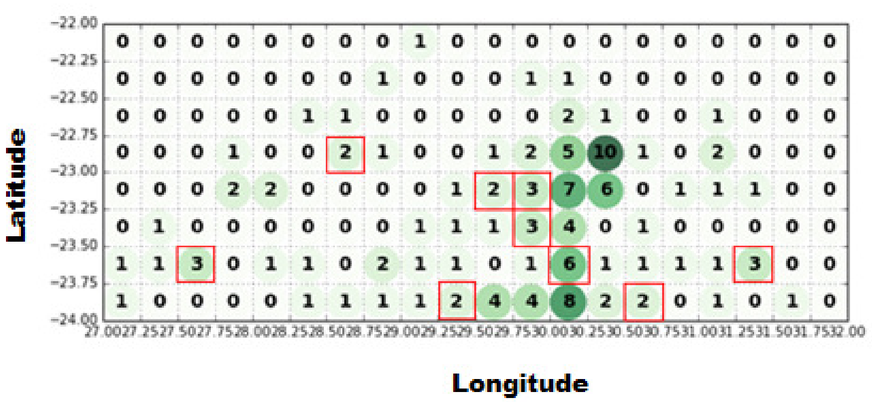


Figure 7: Layout of the Limpopo study site with the 9 selected blocks highlighted in red squares

3.2 Data Sources

3.2.1 Introduction

The Observation times for TRMM 3B42 3-hourly data are 00:00, 03:00, 06:00, 09:00, 12:00, 15:00, 18:00 and 21:00 UTC. The time denotes the mid of 3-hourly interval, i.e., the time 00:00 UTC denotes that the rainfall is averaged over the period 22:30 UTC of the previous day to 01:30 UTC of the next day.

The product of the Multiquadric analysis was a NetCDF file containing a three dimensional array of block averaged daily rainfall totals for each TRMM block on all 3682 days in the analysis period running from 2000-03-01 until 2010-03-31. This overlap period was chosen because TRMM dataset runs from 2000-03-01 until April 2015, while the ground-based rain-gauge available dataset spans the period 1850-01-01 until 2010-03-31.

3.2.2 Data Processing

A difficulty is encountered when comparing data from ground based gauges with those from satellite products in that they provide two different kinds of information. Satellite estimates provide an average precipitation over an area of the satellite pixel, while the gauge provides measurements at points. Hence, one of them should be transformed to the format of the other to make them comparable. The approach that was used follows.

Block-averaged ground-based gauge station measurements are compared with the corresponding TRMM satellite derived pixel precipitation value at the same location. Once all gauge weights were calculated for each block, they were arranged in a matrix relating gauge weights in each block with individual gauges. This matrix ensures that the correct information is collected in each column of block averages.

A similar dataset of daily rainfall accumulations was developed for TRMM data, being careful to match the accumulation times of TRMM in UTC to those of the gauge reporting periods in South African Standard Time (SAST) (a 2 hour shift). It was important to ensure that TRMM accumulations represented the 24 hour accumulation reported at 08:00 SAST.

3.2.2.1 Ground based gauges

A list of all gauging stations located within a pixel or a block was created. Records of each of the gauges were studied in order to choose those that have good reliable data over the 10 year period. The first criterion was to retain the stations with less than 10% missing data and discard others. Daily values were measured by SAWS from 08:00 on a given day to 08:00 on the next. The nearest neighbour patch method was applied to the remainder of the eligible stations where there were gaps. After patching the gauges in each block, the technique that was used for computing weighted average rainfall was based on integration of multiquadric interpolated surfaces over target areas as described by Pegram and Pegram (1993). This programming work was initiated by Dr Sinclair. After scaling, the weights associated with each gauge can be calculated. These weights are then used to determine the average depth over the horizontal surface of the target area for rainfall occurrence, be it daily or over another interval (Pegram et al., 2013). A value of rainfall in the pixel is determined using the weights of all the active gauges in the pixel.

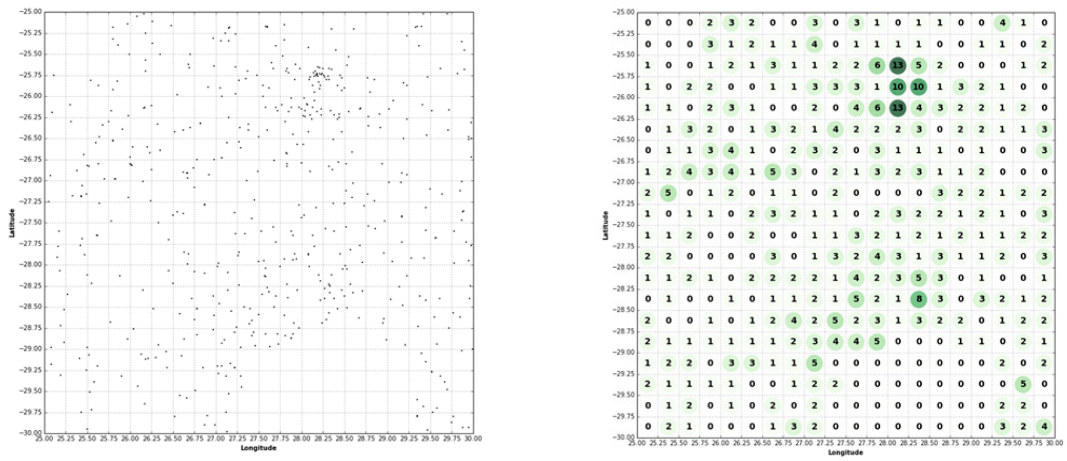
Once the daily average data were calculated per block, with 3652 values for the ten year period under consideration, two of the values on February 29th from the two leap years, 2004 and 2008 were discarded, to simplify the analysis. Thus the total number of daily values matching TRMM and gauge data to be considered per block was 3650, which yields 730 pentad data values, 120 monthly values and 12 yearly values.

3.2.2.2 TRMM data

The TRMM dataset that was considered for the research is version 7 of the real time TRMM Multi-satellite Precipitation Analysis (TMPA-RT). This version of data was made available for use from March 1st, 2000 for research purpose, and the 3B42 product is applicable to South Africa. A whole calendar month is used for IR calibration in 3B42, whereas in 3B42 RT the trailing window of approximately 30 days is used, hence this choice.

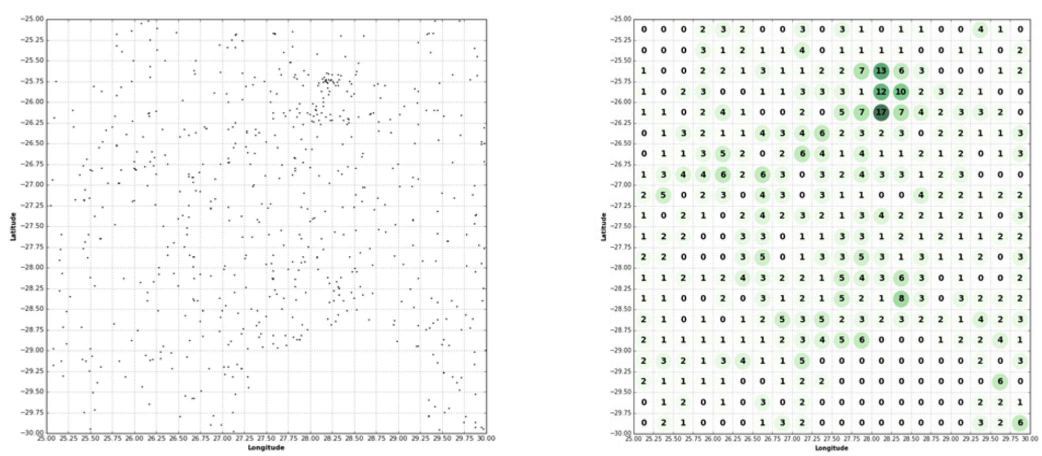
TRMM data files were downloaded in TRMM HDF file format. Once the data was downloaded, the satellite TRMM average values were converted to daily ground based gauge data format to 08:00 to 08:00. To do this, a Python script for each of the four sites was programmed, the results exported into an excel and jpeg file format. By taking the previous day's TRMM snapshot value (mm/h) at 09:00 multiplied by 0.5h, then the current day's 09:00 value multiplied by 2.5h, then each intervening value at 12:00, 15:00, 18:00, 21:00, 00:00, 03:00 and 06:00 multiplied by 3h, the sum of these products gave a total daily value in mm, matching the timings of the gauges.

Figures 8.1a and 8.1b display, (i) the point locations of the rainfall gauges (ii) the TRMM blocks with the number of stations in each block on 3rd March 2000 at the beginning of the ten year period in the Gauteng region first shown in Figure 4. The gauge locations and the population of the TRMM blocks the end of the period 2000-03-01 to 2010-03-31 are shown in Figure 8.2 (a) and (b), respectively. Note the lower gauge counts in the dense cluster in the upper right corner of Figure 8.1b when compared to Figure 8.2b. The layout of active gauges is not constant throughout the period and this had to be accounted for in the analysis, by recalculating the weights, in each gauge-active block, on each day.



a) Ground base gauges in 03 - 2000 **b) TRMM grid in 03 - 2000**

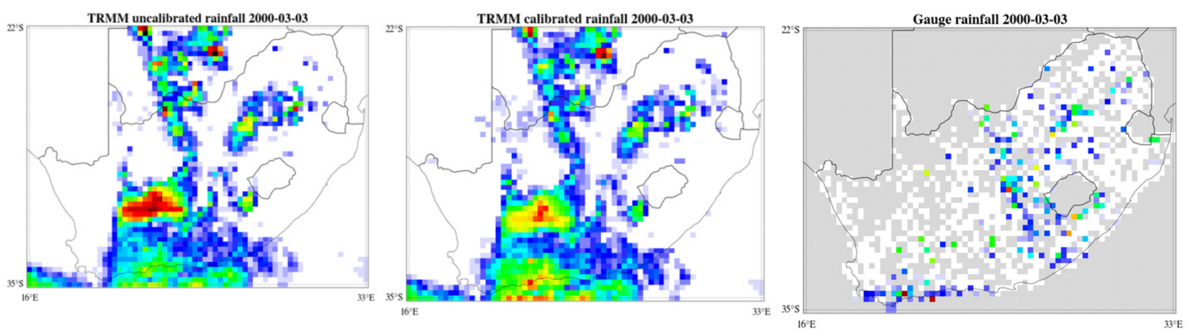
Figure 8.1: The 5° square subregion of South Africa showing gauges active on the first day of the overlapping data-sets: day (2000-03-01) and overlaid by the 0.25° TRMM grid (left panel).



a) Ground base gauges for 10 years **b) TRMM grid for 10 years**

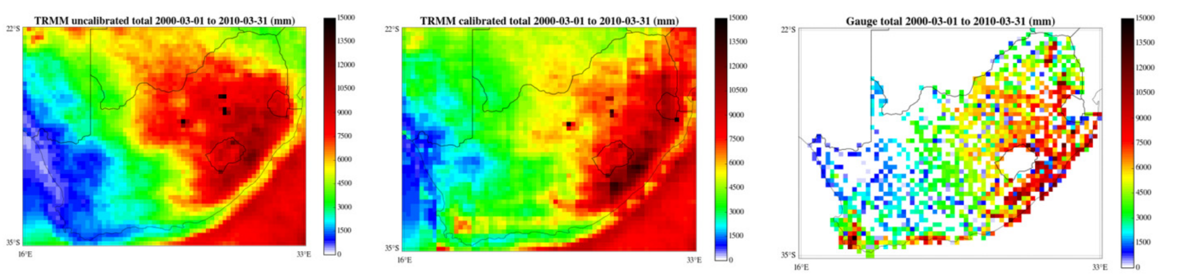
Figure 8.2: The 5° square subregion of South Africa indicated in Figure 8c, illustrating the layout of rain gauges active within the period 2000-03-01 to 2010-03-31 and overlaid by the 0.25° TRMM grid (right panel)

In order to reflect how the two datasets compare, uncalibrated and calibrated TRMM images in Figure 9.1 (a & c) below and BAGD rainfall in Figure 9.2 (a & b) respectively show the rainfall amount estimated for 3rd March 2000 and the totals in the full period 2000-03-01 to 2010-03-31. An 'uncalibrated' TRMM image (both left panels (a) of TRMM images) refers to the rainfall estimates that are made using only satellite data and retrieval algorithms. The grey coloured blocks in Figure 9.1b mean that there was no data to report in the 10 year period.



a) TRMM Uncalibrated vs Calibrated **b) BAGD rainfall**

Figure 9.1. A comparison of daily totals from gauges and TRMM on 3 March 2000



a) TRMM Uncalibrated vs Calibrated **b) BAGD rainfall**

Figure 9.2. A comparison of daily totals from gauges and TRMM for a 10 year analysis period

Note the poor agreement of the daily rainfall totals of Figure 9.1a and 9.1b, with far more zeros in the BAGD estimates in Figure 9.1b when compared to the two pairs of TRMM Figures 9.1a. Figures 9.2 (a & b) show good agreement between the 10-year totals in general, with the BAGD values showing considerable noisy variation. Note the difference in the South West corner (Western Cape) in Figure 9.2, between the Uncalibrated TRMM and the BAGD rainfall accumulations. Although there is general agreement between TRMM and BAGD over the rest of the domain, it is only the Calibrated TRMM which comes close to the BAGD in the problem area. Unfortunately, the Calibrated TRMM product is only available 3 months after data capture, so cannot be used for near real-time comparison with gauges.

Conclusion:

The discussion on the precipitation estimates at the four sites indicates the variety of different rainfall patterns that the country has and gives a view of how well TRMM does at the sites. It is also evident that TRMM does misinterpret rainfall signals in some of the key areas, like in the Western Cape's mountainous ranges, which experience Winter rather than Summer rainfall. BAGD shows noisy variation in some areas, which can be attributed to the variability in available record lengths which strongly affects the total. The TRMM data in the Western Cape underestimates annual rainfall when compared to BAGD. The next chapter focuses on different methods used to compare and analyse the TRMM and BAGD datasets.

CHAPTER 4: DATA ANALYSIS AND RESULTS

In Summary:

In this Chapter, the intent is to compare for, validation, the TRMM precipitation estimates with BAGD rainfall, the ground-based measurements. It presents a comparison carried out for the different daily, pentad, monthly and yearly data from March 2000 until February 2010 at all four sites. The methods used were linear regression, with and without standardisation and cross validation.

4.1 Introduction

In this Chapter, the intent is to compare for validation, TRMM precipitation estimates with Block Averaged Gauge Data (BAGD), the ground-based measurements. It presents a comparison carried out for the different daily, pentad, monthly and yearly data from March 2000 until February 2010 at all four sites: Gauteng, KwaZulu-Natal, Limpopo, and Western Cape. All the matching data for TRMM and BAGD accumulated over the ten year period, that started in March 2000 until February 2010 for both TRMM and BAGD datasets on each of the four sites, were plotted to determine the correlation between the two sets of data at the four locations.

4.2 Data Analysis

In this analysis, the focus was on the rainfall data for daily, pentad, monthly and annual datasets. The 3-hourly TRMM data were converted to daily data to correspond with the timings of BADG daily data, as described in Chapter 3. The data for both TRMM and BAGD that were used had a period from March 2000 until February 2010. Before using the daily rainfall data for analysis, continuity and consistency had to be checked, with the first criterion been to retain the stations with less than 10% missing data and discard the others.

As mentioned in Chapter 3, for the 10 year period, the datasets had 3650 daily values by omitting the two 29 February dates. The remaining daily values were grouped into pentad, monthly and annual values.

The scatter plots for each of the 9 blocks at each site were prepared with TRMM the dependent and BAGD the independent variable.

4.3 Comparing TRMM and BAGD Datasets

Linear regressions of each set for each block in all 4 regions were calculated. These are summarised in this section in tables, with selected scatter-plots of the individual comparisons.

4.3.1 Comparing TRMM and BAGD daily Precipitation Data

The data used to create the tables in this subsection was the raw daily data for the period from March 1st, 2000 until February 28th, 2010. Of the 3 650 data points, only 7 of the TRMM data points needed to be patched because some hours in those days had missing data.

Table 5a and 5b reflect the number of zero values for BAGD and TRMM, respectively, out of the 3650 days.

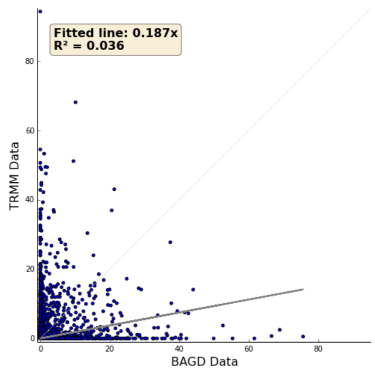
Table 5a: Number of dry days in BAGD Precipitation Daily Data

Block Number	Daily: Number of zero (P0) values			
	Site 1: Gauteng	Site 2: KwaZulu-Natal	Site 3: Western Cape	Site 4: Limpopo
Block 1	2515	2173	2277	2226
Block 2	2102	2046	2380	2924
Block 3	2589	1961	2690	2981
Block 4	2796	2243	3212	3150
Block 5	2275	2157	3057	3201
Block 6	2511	2432	2988	3143
Block 7	2704	2343	2933	2985
Block 8	2401	2357	3049	3069
Block 9	2677	2812	3069	2644

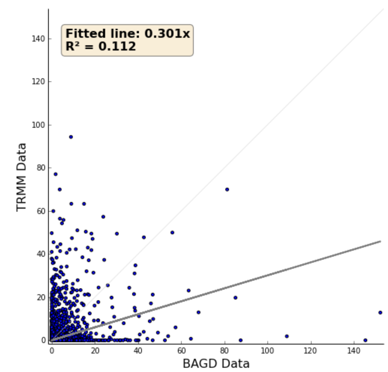
Table 5b: Number of dry days in TRMM Precipitation Daily Data

Block Number	Daily: Number of zero (P0) values			
	Site 1: Gauteng	Site 2: KwaZulu-Natal	Site 3: Western Cape	Site 4: Limpopo
Block 1	855	2720	2910	2897
Block 2	842	2688	2847	2847
Block 3	2189	2811	3166	2704
Block 4	2415	2844	3309	2997
Block 5	1963	2573	3294	2929
Block 6	2628	2492	3082	2771
Block 7	2442	2874	3236	2386
Block 8	1901	2682	3315	2893
Block 9	2583	2874	3353	2887

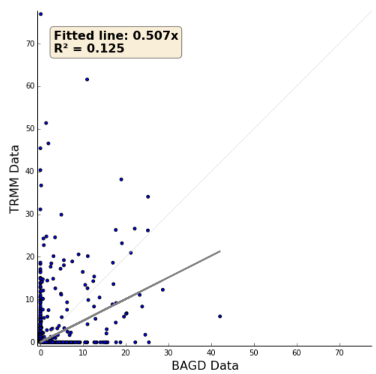
An example of scatterplots comparing TRMM against BAGD data in selected blocks are given below in Figures 10a and Figure 10b. In these, Figure 10a shows the trend-line of the lowest correlation scatter plots of the blocks at each site, while Figure 10b shows the trend-line of the highest correlation scatter plots, from the daily datasets of BAGD and TRMM. The R^2 values of the trend-lines in Figure 10a are 0.036, 0.112, 0.125 and 0.058 for the sites in Gauteng, KwaZulu-Natal, Western Cape and Limpopo, respectively in the selected block. In Figure 10b, the highest R^2 for these sites were 0.114, 0.283, 0.343, and 0.475, respectively. The figures contain the fitted regression equations together with the coefficient of determination R^2, obtained from the graph when the line is constrained to pass through the origin.



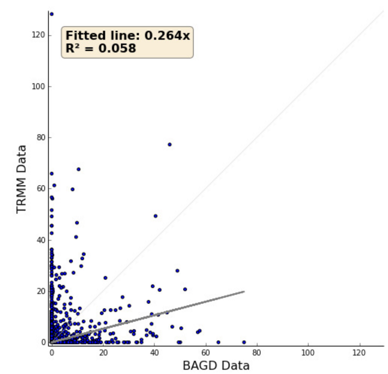
Gauteng



KwaZulu-Natal

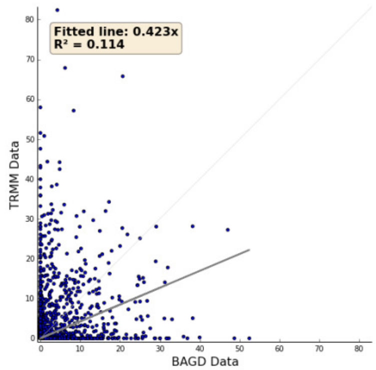


Western Cape

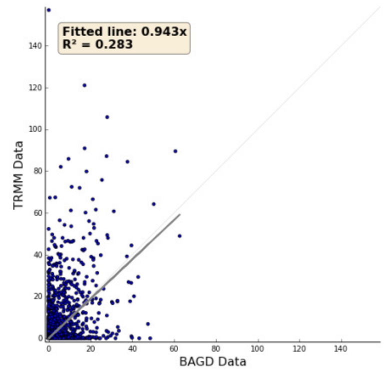


Limpopo

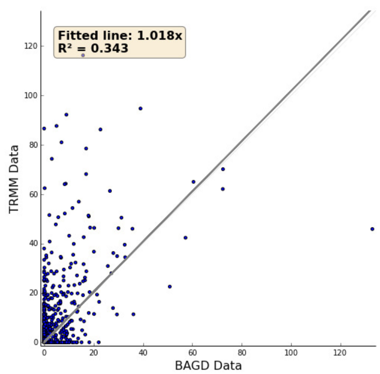
Figure 10a: Lowest correlated Daily Scatter Plot at each site



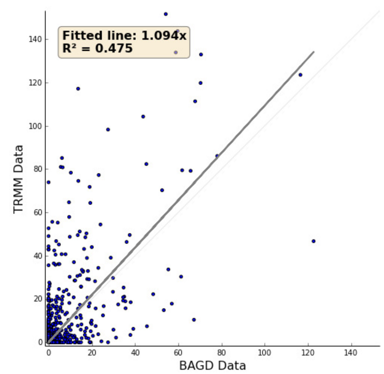
Gauteng



KwaZulu-Natal



Western Cape



Limpopo

Figure 10b: Highest correlated Daily Scatter Plot at each site

Below is Table 6 with the outcome of the comparison of the nine blocks in each of the four sites

Table 6: R^2 values for TRMM and BAGD Precipitation Daily Data

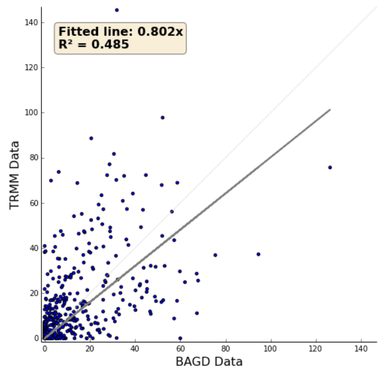
Block Number	Daily: Coefficient of Determination Results (R^2)			
	Site 1: Gauteng	**Site 2: KwaZulu-Natal**	**Site 3: Western Cape**	**Site 4: Limpopo**
Block 1	0.064	0.112	0.262	0.322
Block 2	0.091	0.117	0.293	0.337
Block 3	0.067	0.202	0.250	0.320
Block 4	0.048	0.229	0.167	0.475
Block 5	0.077	0.248	0.172	0.058
Block 6	0.036	0.283	0.241	0.315
Block 7	0.114	0.227	0.342	0.335
Block 8	0.071	0.252	0.125	0.312
Block 9	0.041	0.266	0.194	0.318

TRMM and BAGD daily datasets for all the blocks show mixed results of poor and weak correlation between the datasets.

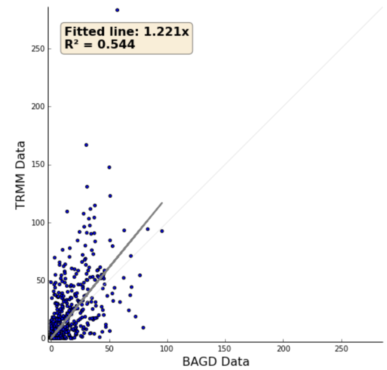
4.3.2 Comparing TRMM and BAGD pentads Precipitation Data

A Pentad is by definition a group or series or sum of five. The pentad dataset consists of 73 pentads of days per year, a total of 730 pentads in a ten year period.

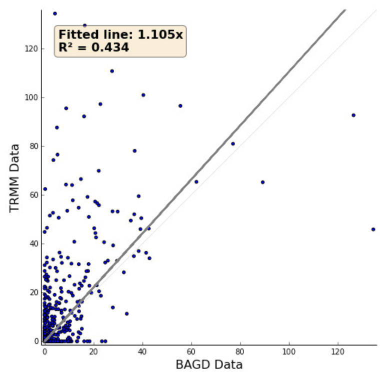
Figure 11 below shows the scatter plots with the strongest correlation between the pentad datasets of the four selected sites. The Limpopo plot reflects good correlation between the two datasets with a coefficient of determination of R^2 = 0.628 as compared to 0.485, 0.544, and 0.434 values for Gauteng, KwaZulu-Natal and Western Cape, respectively. The coefficients of determination were determined from the graph when the line passes through the origin.



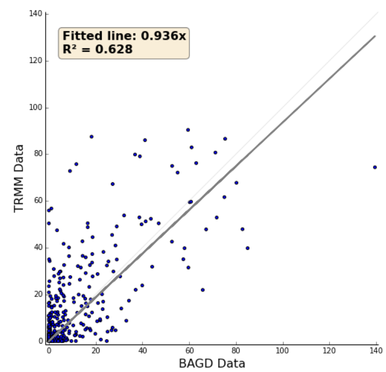
Gauteng



KwaZulu-Natal



Western Cape



Limpopo

Figure 11: Normal Pentads Scatter Plot with the strongest correlation at a site

Table 7 lists the outcome of the pentad comparison of the blocks in each region with the strongest correlation, where it is seen that there is relatively good consistency between the correlations in each site, although they vary between sites.

Table 7: R^2 values for TRMM and BAGD Precipitation Pentads Data

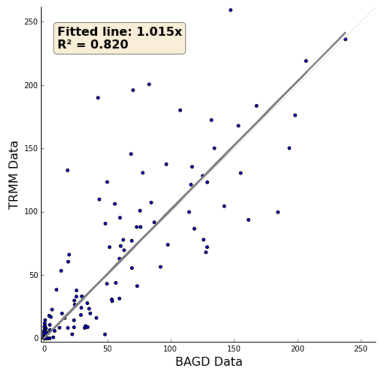
Block Number	Pentads: Coefficient of Determination Results (R^2)			
	Site 1: Gauteng	Site 2: KwaZulu-Natal	Site 3: Western Cape	Site 4: Limpopo
Block 1	0.456	0.270	0.412	0.589
Block 2	0.473	0.340	0.395	0.569
Block 3	0.454	0.383	0.376	0.557
Block 4	0.446	0.392	0.243	0.607
Block 5	0.480	0.497	0.286	0.222
Block 6	0.416	0.544	0.355	0.628
Block 7	0.485	0.346	0.434	0.510
Block 8	0.462	0.444	0.156	0.559
Block 9	0.468	0.417	0.267	0.544

TRMM and BAGD datasets for all the blocks show a reasonable correlation (0.6 to 0.8 based on R^2 values of 0.4 to 0.6) between the datasets with an exception of few blocks in between that reflect fair correlations. It is noted at this stage that the aggregation of daily to pentad intervals provides an improved correlation between the two data sources.

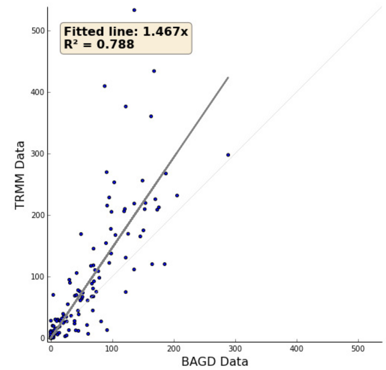
4.3.3 Comparing TRMM and BAGD monthly Precipitation Data

The data used in this subsection was monthly data for the 10 year period that started in March 2000 until February 2010 for both TRMM and BAGD datasets. The total number of data points was 120 in each block.

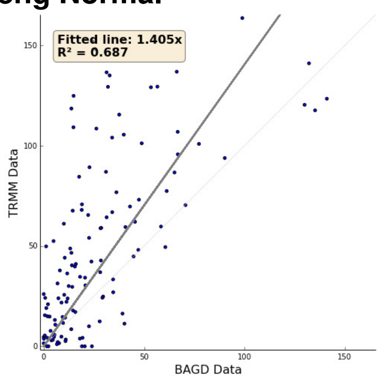
Figure 12 below shows the scatter plots of the blocks with highest R^2 of the months in Gauteng, KwaZulu-Natal, Western Cape and Limpopo., Figure 12 confirms the very good correlation between the two datasets for all four sites, Gauteng having an R^2 of 0.82, KwaZulu-Natal at R^2 = 0.788, and Western Cape and Limpopo claiming a coefficient of correlations of 0.687 and 0.850, respectively.



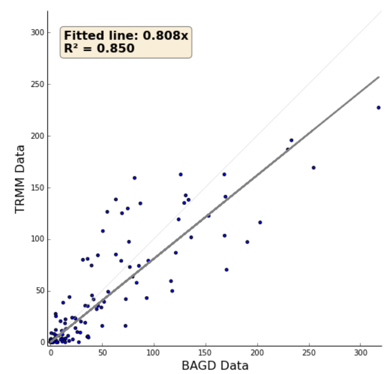
Gauteng Normal



KwaZulu-Natal



Western Cape



Limpopo

Figure 12: Normal Monthly Scatter Plot

Table 8 displays the outcome of the monthly comparison of all nine blocks for each of the four sites

Table 8: TRMM and BAGD Precipitation Monthly Data

Block Number	Monthly: Coefficient of Determination Results (R^2)			
	Site 1: Gauteng	**Site 2: KwaZulu-Natal**	**Site 3: Western Cape**	**Site 4: Limpopo**
Block 1	0.767	0.631	0.669	0.840
Block 2	0.779	0.691	0.670	0.816
Block 3	0.816	0.692	0.683	0.785
Block 4	0.801	0.659	0.561	0.792
Block 5	0.820	0.757	0.643	0.523
Block 6	0.808	0.788	0.665	0.862
Block 7	0.794	0.667	0.688	0.809
Block 8	0.785	0.754	0.399	0.767
Block 9	0.771	0.709	0.525	0.850

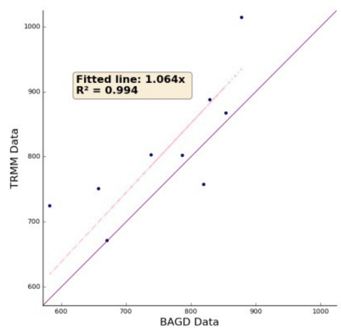
TRMM and BAGD monthly datasets for all the blocks show a very good R between the datasets, when compared to daily and then pentad, as expected. These range between 0.88 to 0.90, 0.79 to 0.88, 0.63 to 0.82 and 0.72 to 0.92 for Gauteng, KwaZulu-Natal, Western Cape and Limpopo, respectively. TRMM and BAGD datasets for all blocks show a very good correlation with correlation coefficients (R) between of 0.723 and 0.928 except the Western Cape Block 8 with R = 0.632, which generally shows a good correlation between the datasets at the monthly time scale.

Block 4 of site 1 in Gauteng, with only 1 station in the block displays an even better correlation between the datasets compared to most of the blocks. The coefficient of determination is R^2 = 0.801, which is useful when compared to daily and pentad data.

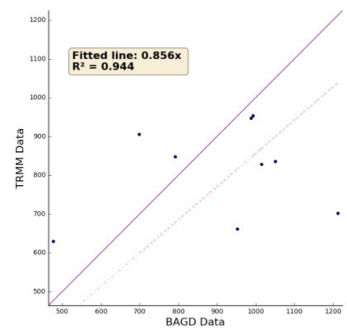
4.3.4 Comparing TRMM and BAGD annual Precipitation Data

The data used in this subsection was annual data for a 10 year period that started in March 2000 until February 2010 for both TRMM and BAGD datasets. The beginning of the each annual year was March 1st and ending in February 28th or 29th if it was a leap year. Below is a selection of scatterplots with the outcome of the comparison.

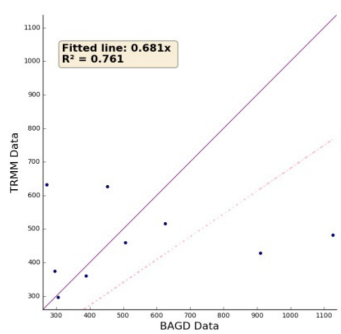
Figure 13 shows the scatter plot of the annual datasets of BAGD and TRMM with the highest coefficient of determination of each of the four sites. Gauteng has R^2 value of 0.994 while KwaZulu-Natal, Western Cape and Limpopo have their values at 0.944, 0.761, and 0.961, respectively. These high numbers are the result of the values clustering in a patch far from the origin, so are not useful comparisons as they stand. In Section 4.4 the forcing of the origin is relaxed; the data are standardised and replotted with modified results.



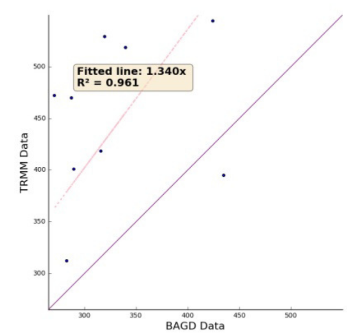
Gauteng



KwaZulu-Natal



Western Cape



Limpopo

Figure 13: Normal Annual Scatter Plot

Table 9 displays the outcome of the annual comparison for the four sites

Table 9: TRMM and BAGD Precipitation Annual Data

Block Number	Annual: Coefficient of Determination Results (R^2)			
	Site 1: Gauteng	Site 2: KwaZulu-Natal	Site 3: Western Cape	Site 4: Limpopo
Block 1	0.937	0.944	0.761	0.929
Block 2	0.952	0.909	0.749	0.961
Block 3	0.939	0.916	0.709	0.940
Block 4	0.916	0.883	0.634	0.863
Block 5	0.953	0.834	0.578	0.906
Block 6	0.943	0.799	0.682	0.932
Block 7	0.956	0.906	0.718	0.900
Block 8	0.984	0.896	0.674	0.870
Block 9	0.994	0.870	0.748	0.921

TRMM and BAGD datasets for all the blocks show results of good correlations between the datasets. The lowest R^2 value is 0.916, 0.799, 0.578 and 0.863 for Gauteng, KwaZulu-Natal, Western Cape and Limpopo, consecutively. The reason for the high values of correlation is that the data cluster far from the origin, so to obtain a fairer comparison, the data are standardised and a full regression line is plotted through each pair of data-set. The outcome of this exercise is reported in the next subsection.

4.4 Standardization of BAGD and TRMM data

Standardization was used to rescale the variables to have a mean of zero and a standard deviation of one. This method is widely used for normalization in many data related algorithms (e.g., support vector machines, logistic regression, and neural networks). Variables are standardized for a variety of reasons, for example, to make sure all variables contribute evenly to a scale when items are added together, or to make it easier to interpret results of a regression or other analysis. If we standardise a variable **x**, obtain a variable **x***, as:

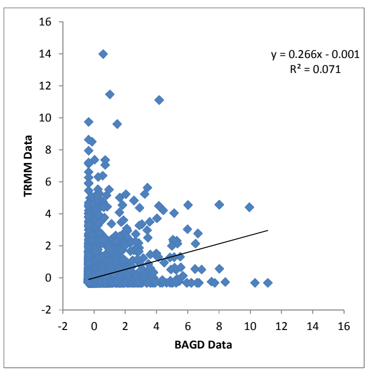
$x^* = (x-m)/sd$

Where **m** is the mean of **x**, and **sd** is the standard deviation of **x**.

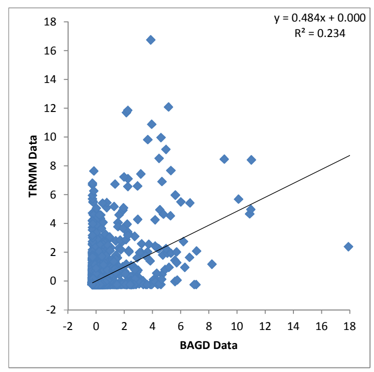
4.4.1 BAGD and TRMM daily correlation of Standardised data

The data used below was the daily data for the period from March 1st, 2000 until February 28th, 2010. The total number of data points were 3 650 per dataset used for comparison.

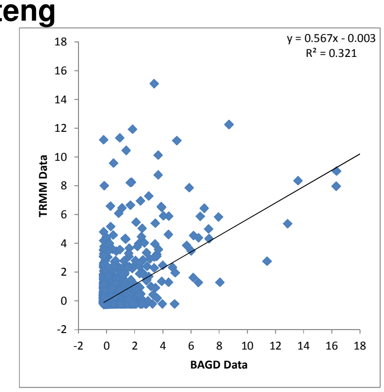
Figure 14 below displays the scatter plots with the highest value of coefficients of determinations for Gauteng, KwaZulu-Natal, Western Cape and Limpopo selected sites. The highest R^2 values for Gauteng, KwaZulu-Natal, Western Cape and Limpopo obtained were 0.071, 0.235, 0.321 and 0.459, respectively, as indicated in the figures. Their fitted regression equations have offsets which are zero, or very near zero, as expected.



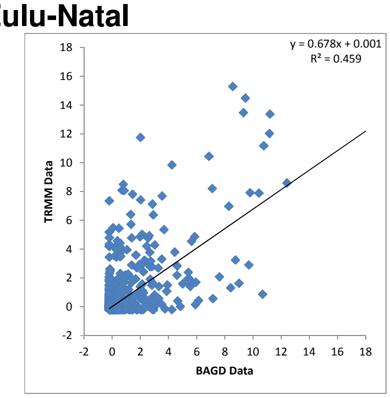
Gauteng



KwaZulu-Natal



Western Cape



Limpopo

Figure 14: Standardised Daily Scatter Plot

Below is Table 10 with the outcome of the comparison of all blocks in each region, showing low correlations across the country, slightly weaker than those in Table 5 .

Table 10: BAGD and TRMM daily correlation for Standardised data

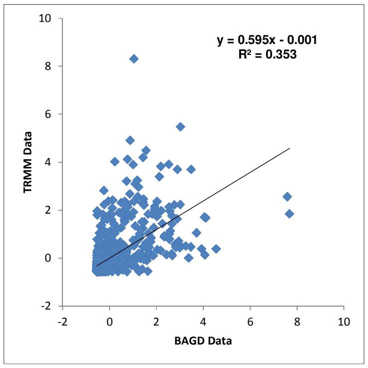
Block Number	Standardised Daily: Coefficient of Determination Results (R^2)			
	Site 1: Gauteng	Site 2: KwaZulu-Natal	Site 3: Western Cape	Site 4: Limpopo
Block 1	0.013	0.078	0.230	0.289
Block 2	0.034	0.067	0.262	0.307
Block 3	0.031	0.158	0.226	0.291
Block 4	0.020	0.191	0.147	0.459
Block 5	0.037	0.193	0.150	0.041
Block 6	0.011	0.227	0.213	0.291
Block 7	0.071	0.194	0.321	0.302
Block 8	0.035	0.206	0.108	0.286
Block 9	0.016	0.235	0.172	0.282

TRMM and BAGD datasets for all the blocks show a poor correlation between most of the datasets for the coastal regions, Western Cape and KwaZulu-Natal and also Gauteng but fair for most of correlations in Limpopo. The cross correlation R ranges between 0.105 to 0.267, 0.259 to 0.485, 0.329 to 0.567, and 0.203 to 0.678 for the 9 blocks in each of the four regions respectively.

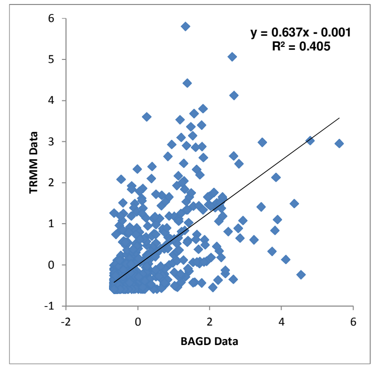
4.4.2 BAGD and TRMM pentad correlation of Standardised data

The data used below was the pentad data for the period from March 1st, 2000 until February 28th, 2010. The total number of data points was 730 per dataset used for comparison.

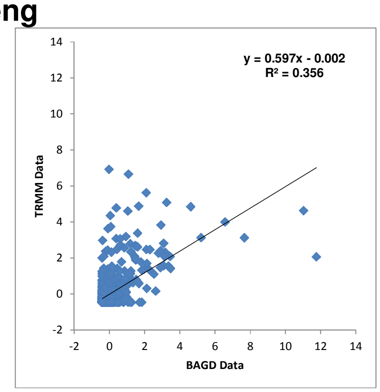
Figure 15 below displays the scatter plots with the highest pentad value of coefficients of determination for Gauteng, KwaZulu-Natal, Western Cape and Limpopo selected sites, which were 0.353, 0.406, 0.356 and 0.574, respectively.



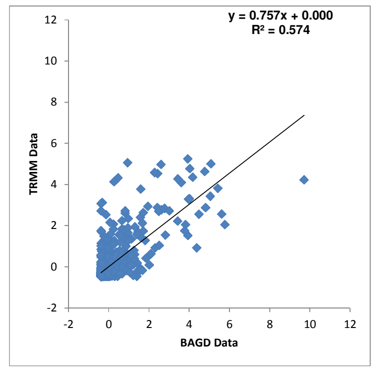
Gauteng



KwaZulu-Natal



Western Cape



Limpopo

Figure 15: Standardised Pentads Scatter Plot

Below is Table 11 with the outcome of the comparison of all regressions between TRMM and BAGD pentads by blocks in each region

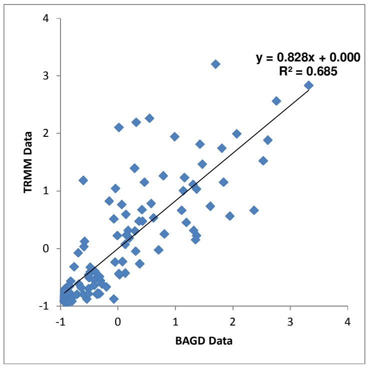
Table 11: BAGD and TRMM pentads correlation for Standardised data

Block Number	Standardised Pentads: Coefficient of Determination Results (R^2)			
	Site 1: Gauteng	Site 2: KwaZulu-Natal	Site 3: Western Cape	Site 4: Limpopo
Block 1	0.261	0.131	0.299	0.521
Block 2	0.233	0.164	0.283	0.499
Block 3	0.285	0.234	0.290	0.488
Block 4	0.316	0.268	0.174	0.561
Block 5	0.337	0.330	0.206	0.150
Block 6	0.302	0.406	0.264	0.574
Block 7	0.353	0.239	0.355	0.413
Block 8	0.328	0.294	0.095	0.501
Block 9	0.341	0.324	0.190	0.467

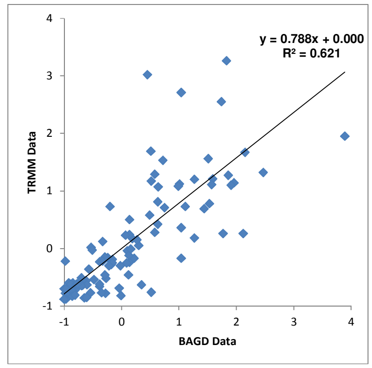
TRMM and BAGD datasets for all the blocks show a weak but fair correlation between the datasets except the site in KwaZulu-Natal with one block showing poor correlation. The correlation coefficient R ranges between 0.482 to 0.594, 0.362 to 0.637, 0.308 to 0.598, and 0.387 to 0.758 for the 9 blocks in each of the four regions.

4.4.3 BAGD and TRMM monthly correlation for Standardised data

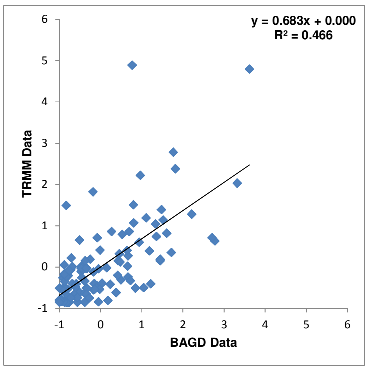
The data used below was the monthly data for the period from March 1st, 2000 until February 28th, 2010. The total number of data points was 120 per dataset used for comparison. Figure 16 below displays the scatter plots with the highest monthly value of coefficients of determinations for Gauteng, KwaZulu-Natal, Western Cape and Limpopo selected sites. The highest monthly R^2 values for Gauteng, KwaZulu-Natal, Western Cape and Limpopo obtained were 0.685, 0.621, 0.466 and 0.784, respectively.



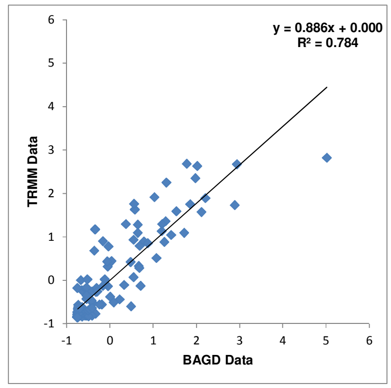
Gauteng



KwaZulu-Natal



Western Cape



Limpopo

Figure 16: Standardised Monthly Scatter Plot

Below is Table 12 with the outcome of the comparison of all blocks' monthly regressions in each region

Table 12: BAGD and TRMM monthly correlation for Standardised data

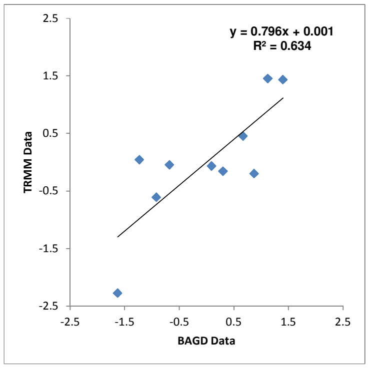
Block Number	Standardised Monthly: Coefficient of Determination Results (R^2)			
	Site 1: Gauteng	Site 2: KwaZulu-Natal	Site 3: Western Cape	Site 4: Limpopo
Block 1	0.549	0.283	0.362	0.743
Block 2	0.554	0.391	0.376	0.704
Block 3	0.665	0.407	0.466	0.649
Block 4	0.661	0.335	0.332	0.695
Block 5	0.686	0.511	0.436	0.317
Block 6	0.666	0.621	0.426	0.784
Block 7	0.653	0.382	0.422	0.670
Block 8	0.627	0.531	0.172	0.654
Block 9	0.616	0.490	0.268	0.756

TRMM and BAGD datasets for all the blocks show a good correlation between the datasets. The R ranges between 0.741 to 0.828, 0.532 to 0.788, 0.415 to 0.683, and 0.563 to 0.885 for the 9 blocks in each of the four regions. These values are useful, even though they are much lower than the values in Table 8, which reported R^2 values for regressions constrained to go through the origin.

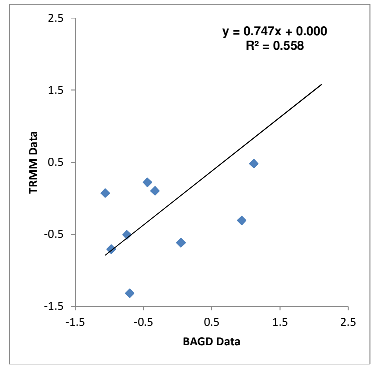
4.4.4 BAGD and TRMM yearly correlation for Standardised data

The data used below was the yearly data for the period from March 1st, 2000 until February 28th, 2010. The total number of data points was 10 per dataset used for comparison.

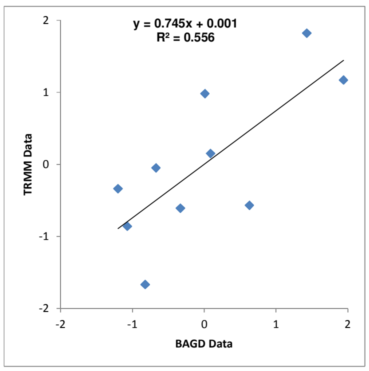
Figure 17 below displays the scatter plots with the highest annual value of coefficients of determinations for Gauteng, KwaZulu-Natal, Western Cape and Limpopo selected sites. The highest monthly R^2 values for Gauteng, KwaZulu-Natal, Western Cape and Limpopo obtained were 0.634, 0.558, 0.556 and 0.804, respectively.



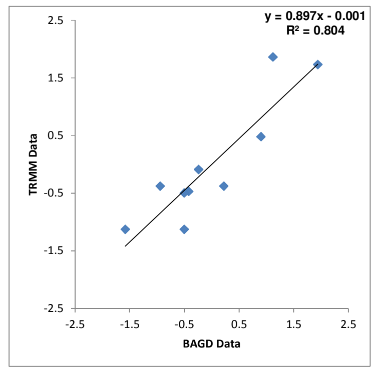
Gauteng



KwaZulu-Natal



Western Cape



Limpopo

Figure 17: Standardised Annual Scatter Plot

Below is Table 13 with the outcome of the comparison of all the annual results per blocks in each region.

Table 13: BAGD and TRMM yearly correlation for Standardised data

Block Number	Annual: Coefficient of Determination Results (R^2)			
	Site 1: Gauteng	Site 2: KwaZulu-Natal	Site 3: Western Cape	Site 4: Limpopo
Block 1	0.019	0.052	0.371	0.804
Block 2	0.098	0.005	0.317	0.471
Block 3	0.271	0.020	0.225	0.340
Block 4	0.584	0.558	0.176	0.344
Block 5	0.467	0.000	0.323	0.173
Block 6	0.294	0.487	0.276	0.771
Block 7	0.623	0.552	0.461	0.292
Block 8	0.579	0.067	0.556	0.399
Block 9	0.634	0.183	0.464	0.472

Taking the square root of the values in Table 13 shows that TRMM and BAGD standardized annual datasets for all blocks with correlation coefficients (R) range between of 0.000 (KwaZulu-Natal Block 5) and 0.796 (Limpopo at 0.796).

In summary, the daily totals show a very low correlation amongst datasets for regression lines forced through the origin. The daily R^2 has an average of 0.205, and once it was accumulated into pentads, monthly and yearly, the R^2 values improved to an average of 0.430, 0.723 and 0.959, respectively. These results were misleading, hence the need for the data to be standardised. For standardised data, the average R^2 values for daily are very low at (0.030 to 0.283) to monthly value averaging (0.363 to 0.664), which is not consistent in some areas.

Below is Table 14 with the outcome of the average values for normal (plots through the origin) and standardised datasets of all blocks in each region. It will be seen that the Normal method of line fitting used in section 4.3 produces a spurious coefficient of variation compared to the more realistic Standardised regression used in this section 4.4, a lesson which had to be learned by trial.

Table 14: BAGD and TRMM datasets average coefficient of determination

Time		Average Coefficient of Determination Results (R^2)			
		Site 1: Gauteng	Site 2: KwaZulu-Natal	Site 3: Western Cape	Site 4: Limpopo
Daily	Normal	0,068	0,215	0,227	0,310
	Standardised	0,030	0,172	0,203	0,283
Pentad	Normal	0,460	0,404	0,325	0,532
	Standardised	0,307	0,266	0,240	0,464
Monthly	Normal	0,793	0,705	0,611	0,783
	Standardised	0,630	0,439	0,362	0,664
Annual	Normal	0,952	0,884	0,695	0,914
	Standardised	0,398	0,214	0,355	0,452

The interim conclusion from sections 4.3 and 4.4 is that there is a poor information linkage between daily and annual data, but a reasonable one between pentads and a good one between monthly data. This is almost surely due to the relatively strong seasonality signal compared with noise found in the monthly data, but in these short records, this cannot be helped. An attempt was made to deseasonalise the data by filtering and fitting Fourier series, but this approach was abandoned because of the poor results, mainly due to the strong effect of dry periods causing the fitted functions to go negative.

The next subsection explores the usefulness of using regression to estimate monthly values of gauge data from TRMM and is validated by the method of cross-validation.

4.5 Cross validation of TRMM and BAGD Data

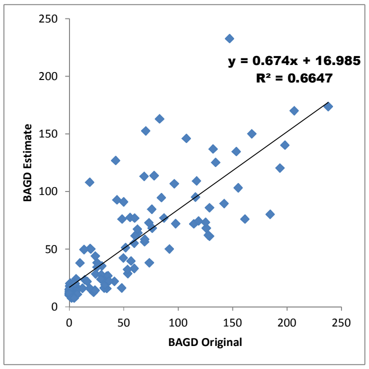
Suppose we have a block with no rainfall gauges and we need to estimate the BAGD dataset for the block from the TRMM dataset. How will cross validation assist in cases where we have no rain gauges? The answer is 'not directly', but if the regression equations linking TRMM and BAGD data (at blocks which have both sets) can be shown to be useful and similar over the set of blocks in the site, then we can use these regression equations to infill the empty blocks with confidence. Cross validation is a technique that can be used in blocks that have both data sources, to determine how well the TRMM dataset can predict or estimate the existing BAGD dataset, by leaving out some of the observations and comparing those with the estimated values. In order to use the cross validation technique on monthly datasets for both TRMM and BAGD, the method described in the following paragraphs will assist in getting the desired results. Monthly data were chosen because of their relatively high R^2 values obtained in Chapter 3.

Choose a block at one of the 4 sites, then from the 10 year period of BAGD and TRMM datasets, a year is removed, one at a time and the remaining 9 years of pairs of data will be used in turn to determine the regression equations linking them. The process of removal will be repeated 10 times for the ten years of the block's dataset. A set of 9 equations per year, one for each block at each site, will be obtained. Table 15 below displays the regression equations for the nine blocks at each of the four sites having had the first year omitted from the regression. There will be 9 more similar tables for the other missing years. The regressions are done with the TRMM and BAGD data as dependent and independent variables respectively. From the slope parameters, it is noted that the prediction of BAGD from TRMM has a ratio generally lower than unity and that there is a mostly a positive offset, because many small TRMM values correspond to zero BAGD values.

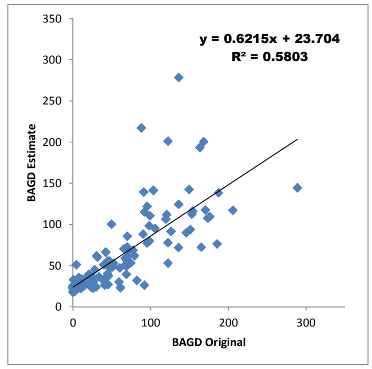
Table 15: Regression equations obtained by linking TRMM and BAGD monthly data at the four sites with only the first year omitted in each trial

Block Number	Site 1: Gauteng	Site 2: KwaZulu-Natal	Site 3: Western Cape	Site 4: Limpopo
Block 1	0.709*X -4.634	0.536X + 36.634	1.350X + 25.810	0.956X + 11.337
Block 2	0.695*X-2.933	0.818X + 47.270	1.653X + 36.396	0.720X + 8.825
Block 3	0.674*X+10.433	0.535X + 40.261	0.752X + 14.390	0.795X + 3.688
Block 4	0.881*X+11.286	0.492X + 60.357	0.587X + 5.572	0.562X + 1.806
Block 5	0.716*X+9.403	0.525X + 32.750	1.019X + 14.220	0.538X + 13.750
Block 6	0.822*X+9.542	0.441X + 19.968	0.530X + 10.971	0.765X + 1.611
Block 7	0.663*X+9.81	0.497X + 39.166	0.428X + 5.395	0.505X + 3.111
Block 8	0.786*X+10.681	0.424X + 28.619	0.351X + 10.487	0.602X + 4.038
Block 9	0.754*X+11.533	0.463X + 17.924	0.537X + 11.785	1.039X + 4.504

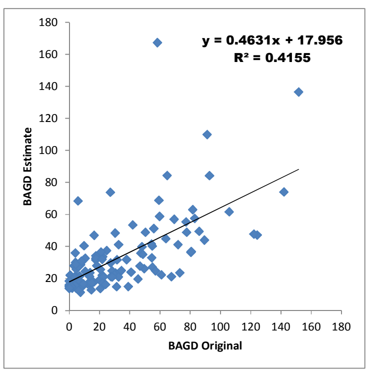
Once the set of equations is completed, they are then used to estimate the BAGD values, with the dependent variable X in the equation set as the TRMM observed value for that block. The product of this step will be BAGD Estimates. Figure 18 displays the scatter plots of BAGD Estimated vs BAGD original values, with only the highest coefficients of determinations plots chosen for display, for the Gauteng, KwaZulu-Natal, Western Cape and Limpopo selected sites. The monthly R^2 values obtained were 0.665, 0.580, 0.415 and 0.629, respectively. Note the slight positive offset of the estimates of the lower values.



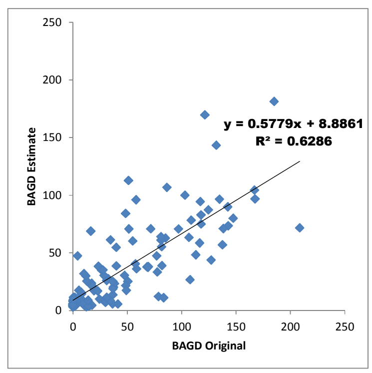
Gauteng



KwaZulu-Natal



Western Cape



Limpopo

Figure 18: BAGD original vs BAGD Estimate Monthly Scatter Plots

Below is Table 16 with the outcome of the R^2 comparison of all blocks for BAGD original and BAGD Estimates using cross-validation for each of the four sites, with the highest coefficient of determination in each column highlighted.

Table 16: BAGD Original and BAGD Estimated monthly R^2 results

Block Number	Coefficient of Determination Results (R^2)			
	Site 1: Gauteng	Site 2: KwaZulu-Natal	Site 3: Western Cape	Site 4: Limpopo
Block 1	0.528	0.229	0.337	0.544
Block 2	0.442	0.352	0.351	0.388
Block 3	0.651	0.358	0.415	0.629
Block 4	0.652	0.303	0.298	0.204
Block 5	0.665	0.456	0.375	0.523
Block 6	0.640	0.580	0.393	0.399
Block 7	0.635	0.327	0.393	0.454
Block 8	0.615	0.466	0.142	0.466
Block 9	0.575	0.409	0.236	0.437
Average	**0.600**	**0.387**	**0.327**	**0.449**

TRMM and BAGD datasets for all the blocks show a good correlation between the datasets except the KwaZulu-Natal and the Western Cape blocks, with the highest R^2 values of 0.580 and 0.415 respectively. The averages for both Gauteng and Limpopo are higher at 0.577 and 0.449, respectively as compared to the coastal sites' lower averages of 0.387 and 0.327 for KwaZulu-Natal and Western Cape, which nevertheless translate to useful correlations of from 0.5 to 0.8.

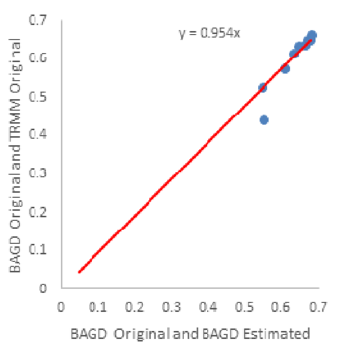
Table 17 below shows the R^2 results for the original BAGD versus TRMM datasets.

Table 17: BAGD Original and TRMM Original monthly R^2 results

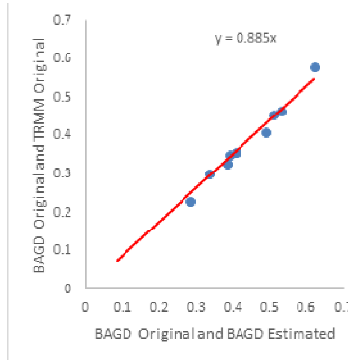
Block Number	Coefficient of Determination Results (R^2)			
	Site 1: Gauteng	Site 2: KwaZulu-Natal	Site 3: Western Cape	Site 4: Limpopo
Block 1	0.546	0.283	0.362	0.593
Block 2	0.550	0.391	0.376	0.413
Block 3	0.671	0.407	0.466	0.636
Block 4	0.676	0.335	0.332	0.192
Block 5	0.680	0.511	0.436	0.534
Block 6	0.664	0.621	0.426	0.424
Block 7	0.647	0.382	0.422	0.471
Block 8	0.633	0.531	0.172	0.453
Block 9	0.610	0.490	0.268	0.446
Average	**0.631**	**0.439**	**0.362**	**0.463**

The results clearly show a good comparison with the Table 16 results, in that the averages of the table with BAGD Original vs BAGD Estimate are 0.600, 0,387, 0.327 and 0,449 while the table with BAGD Original vs TRMM Original have an average of 0.631, 0,439, 0.362 and 0,463 for Gauteng, KwaZulu-Natal, Western Cape and Limpopo, respectively. The average value compares fairly well across all sites.

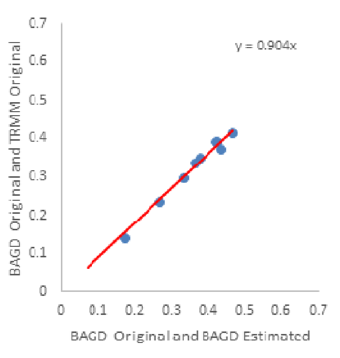
To support this observation, in Figure 19 the R^2 values of regressions of TRMM and BAGD original datasets are compared with the BAGD Original vs BAGD Estimate ones, derived from Tables 15 and 16. There is a small loss of information, but enough corroboration to justify using the TRMM-BAGD equations directly. It is worth noting that the Gauteng site 1 gauge data have at least 40 gauges per block, a coverage not repeated over the rest of the country. The result is that the site experiences relatively high correlations between TRMM and both the full and censored data sets. This observation is developed in Chapter 5, where the effect of gauge density is explored.



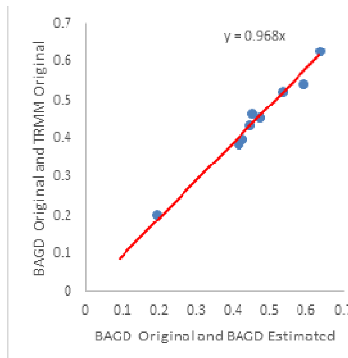
Gauteng



KwaZulu-Natal



Western Cape



Limpopo

Figure 19: BAGD observed and TRMM observed vs BAGD observed and BAGD Estimated R^2 Monthly Scatter Plots

Conclusion:

The results of cross-validation in Limpopo and Gauteng show some consistency, as compared to the coastal results like KwaZulu-Natal and Western Cape, which is evident with the monthly standardised average values of 0.439 and 0.362, respectively. It is clear that the relationship between TRMM and BAGD datasets depends on the site and factors like topography in the Western Cape and the number of gauges in a block. With that in mind, there is a need to look at each site with great care, taking into account its location. It turned out that all direct comparisons of daily and pentad data between the TRMM and BAGD series were not useful, because of the poor timing of TRMM compared to ground-based data. The monthly data were fairly well correlated, but not the annual data because of the very few years. Nevertheless, the message to carry forward is that monthly TRMM values can be used in the summer rainfall regions directly in agrohydrological applications in South Africa, based on the good results obtained from the well populated TRMM blocks in Gauteng. The sparsely populated blocks need further treatment.

The next chapter focuses on the introduction of a novel idea for performing a quantile-quantile (Q-Q) transform of TRMM to Block averaged gauge daily rainfall and highlights the problem of gauge sparseness.

CHAPTER 5: THE QUANTILE-QUANTILE (Q-Q) TRANSFORM OF TRMM TO BAGD

In Summary:

This chapter uses the ideas introduced in the previous chapter and focuses on quantile-quantile (Q-Q) transform methods and the probability of zero rainfall. The intent here is to determine if there is any useful link between TRMM and BAGD such that, given an area with poor or no gauges, TRMM data can be used to estimate the actual rainfall using the parameters suited for the area in question.

Having completed the cross correlation coefficient methodology of both original and standardised data of TRMM and BAGD datasets and found that the monthly data 'communicate' quite well, it still remains to be seen whether there exists a useful link between the two datasets at the daily time-scale, which is usually the one of interest. The useful link will assist in knowing how to use a TRMM pixel value to get information on an area that does not have gauges or is poorly gauged, i.e. how does one use TRMM to make a good estimate of the actual rainfall as if it were gauged and averaged over a TRMM pixel? A novel idea was used. The idea is to perform a quantile-quantile (Q-Q) transform of TRMM to Block averaged gauge daily rainfall, where there are ground-based records, and then interpolate the parameters describing the probability distributions to ungauged locations, as described in Pegram et al. (2016), which will not be repeated here. This Q-Q transformation was done by using the appropriate and easy-to-manipulate Weibull probability distribution fitted to gauge data, where available, and to all TRMM data. This chapter draws strongly from the WRC report of Pegram et al. (2016), but concentrates on the candidate's contribution and its extension.

The Weibull probability distribution has parameters that allow the function to take other distributions' characteristics: http://reliawiki.org/index.php/The_Weibull_Distribution .

The extended Weibull distribution function is given by:

(a) $F(x) = p = 1 - (1 - p_0)\ exp(-x/\alpha)^{\beta}$ (1)

With the inverse of the function above being

(b) $x = \alpha\ [ln\{(1-p_0)/(1-p)\}]^{1/\beta}$ (2)

where:

F(x) = p is the extended cumulative distribution function,

x is the rainfall value,

p_0 is the probability of zero rainfall,

α and β are scale and shape parameters.

The above equations are such that if we know the parameters of *F(x)* determined from TRMM data at a certain rainfall block, we can determine the transformed rainfall value from the inverse function.

The Q-Q procedure is easier to apply from location to location if sufficient rainfall data is available and the challenge comes when one has to bias-correct TRMM where there are no gauges in a given block or location.

Figure 20, from Pegram et. al. (2016), explains the Q-Q transform which uses the BAGD and TRMM daily values from Block 9 in Gauteng. The blue line represents the Weibull distribution fitted to TRMM daily data, the red line is the Weibull distribution fitted to the BAGD data in the same block. The parameter values for the distribution fitted to TRMM were p_0 = 0.709, α = 4.66 and β = 0.619 while those fitted to BAGD are p_0 = 0.733, α = 5.31 and β = 0.656. One sees that there are similarities between the two lines (red & blue). The simple explanation to bias correct the TRMM value is first to look at the green line with the arrow facing up with an x-axis value of about 1.2. Secondly, once the line meets or intercepts the blue pdf (TRMM model), then a line is drawn perpendicular to the line at 1.2 that meets the red line (BAGD model) and lastly, draw a line parallel and opposite to the first line towards the x-axis starting from the red line to get the transformed value of 0.907. Essentially the above equation (Eq. 2) was used to determine the BAGD estimate from the TRMM observation.

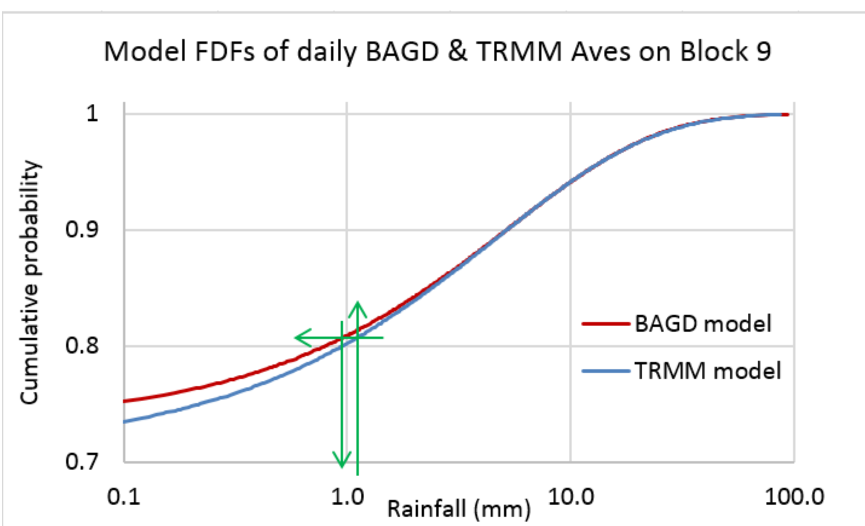


Figure 20: Sequence of calculations to perform a Q-Q transform of TRMM rainfall to Gauge. Blue curve: Weibull model fitted to TRMM; Red curve: Weibull distribution fitted to the BAGD data.

In Gauteng, a 4 x 4 matrix was selected and highlighted in the blue square shown in Figure 21 below. The selected block is within the Gauteng study area, chosen with consideration of the number of stations in each block. Table 18 a) below shows the number of stations in each block in Figure 20, with Table 18 b) containing the values of the P_0 of TRMM data for each block in the 4 x 4 matrix.

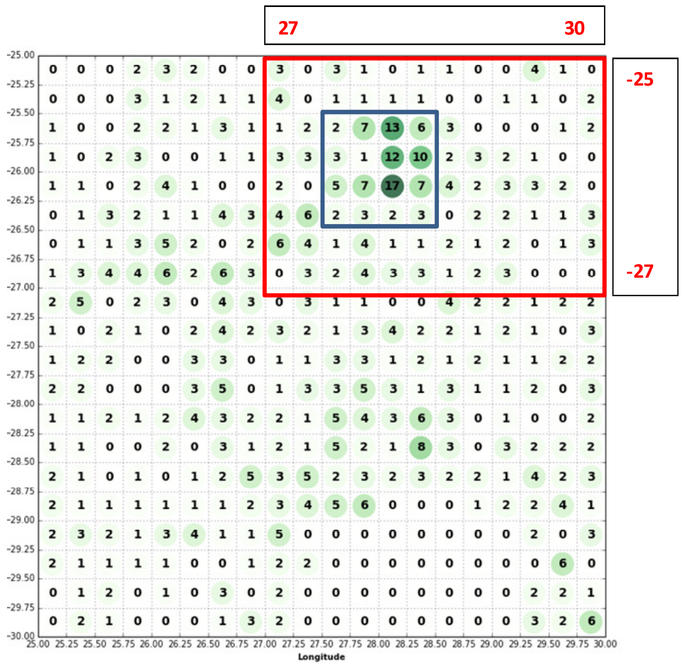


Figure 21: Gauteng area with the selected blocks highlighted in blue - the red numbers are degrees East and North

Table 18: Gauteng a) Number of stations in each block and b) observed gauge P0 values. The colours are to identify blocks with ties

a)

2	7	13	6
3	1	12	10
5	7	17	7
2	3	2	3

b)

0.8932	0.7533	0.6004	0.7347
0.8530	0.9100	0.5876	0.6666
0.7906	0.6655	0.5403	0.6905
0.8667	0.7803	0.8985	0.7644

The next step is to find a relationship between the number of gauges per block and its p_0 value. This can be described by a truncated exponential function (Equation 3).

The parameters values of this function to be fitted to these data are *a* and *b*, with the variable x is the number of gauges per populated block.

In Table 18 above, some of the blocks have the same number of stations, i.e. three sets of three blocks have 2, 3, and 7 stations respectively and are highlighted in the table. The blocks with the same number of stations in them, were given an average value of the p_0. Once the average values for the blocks with the same number of stations is calculated and consolidated, the 'Measured p_0 values' were obtained, shown in the middle column of Table 19 below. To obtain a relationship between p_0 and block station count, x, the following exponential formula was used:

$$p_0[x] = 1 - a[1 - \exp(-bx)] \qquad (3)$$

minimising the sum of squares of the differences between p_0 and $p_0[x]$, yielding an asymptotic value of p_0 for a heavily populated block equal to 1-a.

Table 19: Gauteng's p_0 consolidated values and Exponential distribution function (3) calculation

Number of Stations	Measured p_0 Value	Exponential function estimate
1	0.9100	0.9400
2	0.8861	0.8871
3	0.7992	0.8405
5	0.7907	0.7632
6	0.7347	0.7313
7	0.7031	0.7032
10	0.6666	0.6373
12	0.5876	0.6054
13	0.6004	0.5922
17	0.5403	0.5533

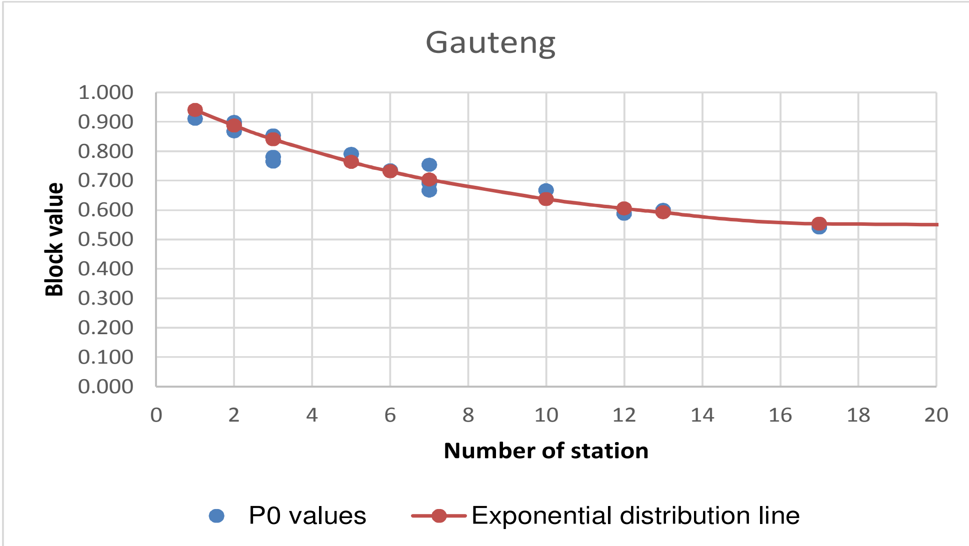


Figure 22: Gauteng Exponential function values

Figure 22 for Gauteng compares the p_0 values of the selected blocks above listed in Table 18, with the values plotted as blue dots. The red fitted line represents the Exponential function (3) estimates. The Exponential function fits the data very well and has an asymptotic p_0 value of 1 - a = 0.494, with a = 0.5058, b = 0.1262.

The Exponential distribution functions for Limpopo and KwaZulu-Natal sites both fitted very well to their selected data. The results are shown in Figure 23 and Figure 24 for Limpopo and KwaZulu-Natal, respectively, with asymptotic p_0 values of 0.451 and 0.628.

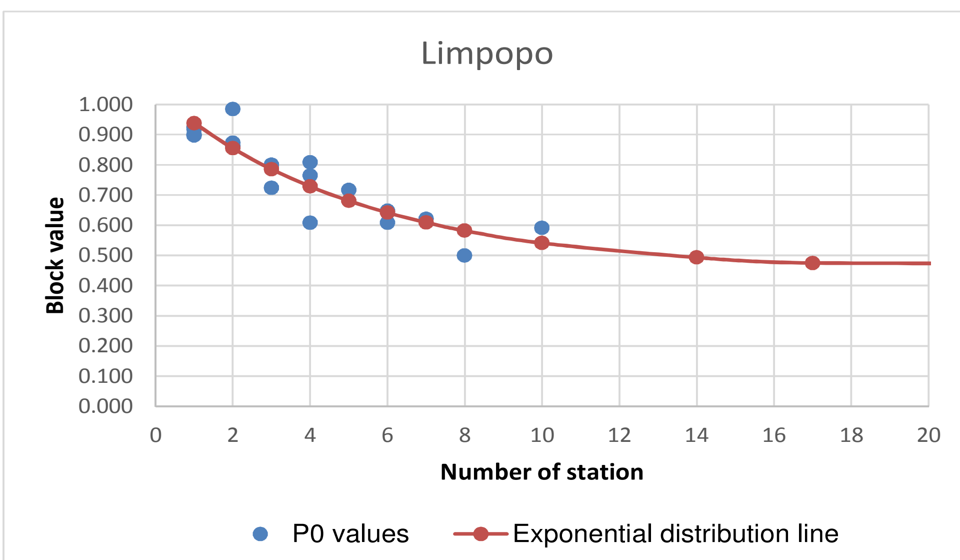


Figure 23: Limpopo Exponential function values

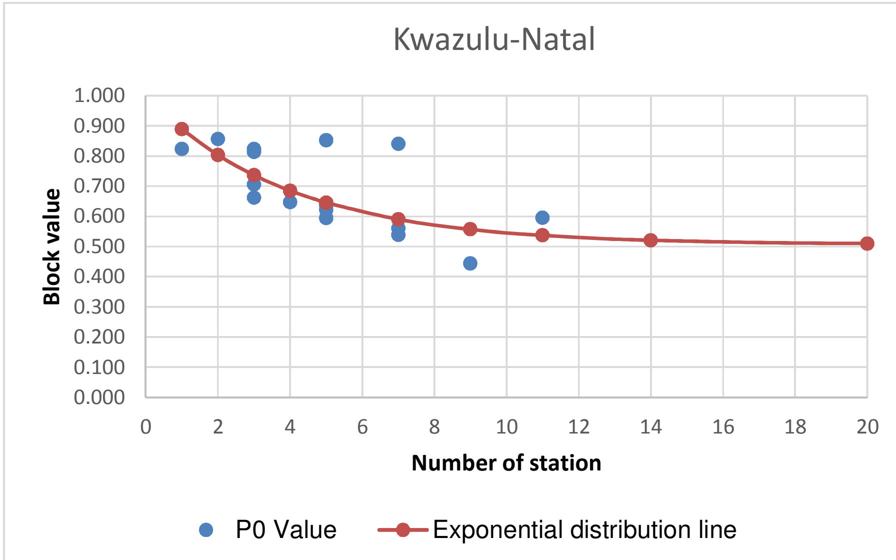


Figure 24: KwaZulu-Natal Exponential function values

However, the Western Cape selected data-set's fitted distribution function shows a flat red line that doesn't fit well to the data, see Figure 25 below, caused by a single station with a relatively low block p_0 value of 0.56, compared to the other block with a p_0 of 0.80, pulling the equation flat. The challenge with the Western Cape is that there is weak correlation between the gauge's readings because of the mountains and the prevailing wind, resulting in spatially variable rainfall.

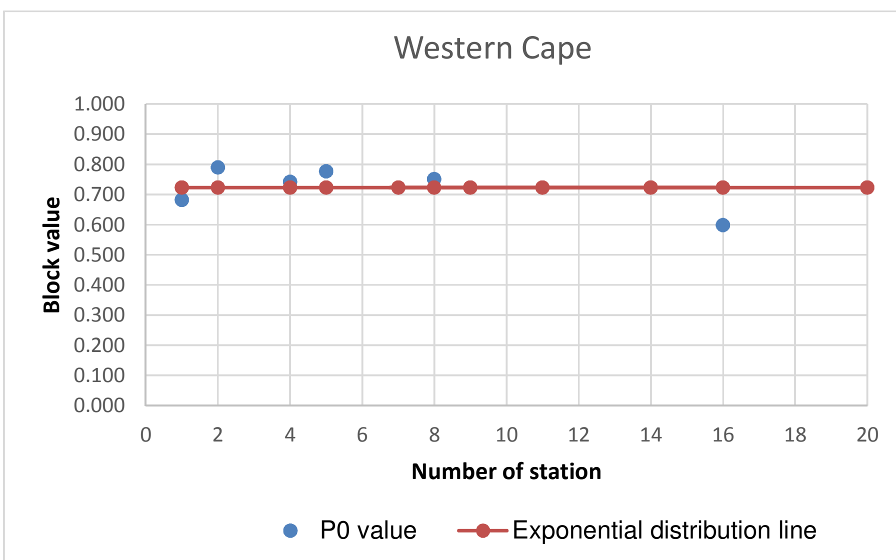


Figure 25: Western Cape Exponential function values

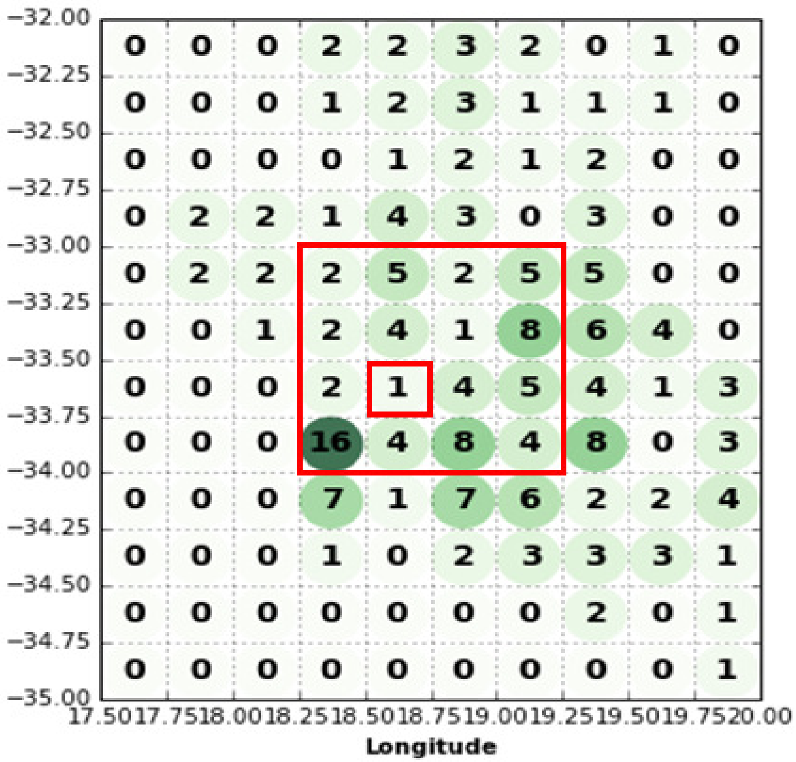


Figure 26: Western Cape area with the 12 selected blocks in the outer red square. The red block inside has a p_0 value of 0.5655 as highlighted in yellow in the table below

Table 20: The Western Cape p_0 values consolidated and Exponential function calculation

0.83451	0.81034	0.80082	0.76563
0.78116	0.73690	0.79777	0.79151
0.74016	0.56550	0.75096	0.75210
0.59812	0.72420	0.70948	0.75384

The observed TRMM p_0 values of the Western Cape are displayed above the Table 20; the highlighted block in yellow shows a p_0 value of 0.5655 that seems to be an outlier from the set of p_0 values - it belongs to a very wet gauge near Cape Town. By removing this value from the set, and fitting the Exponential function to the remaining data as in Figure 27 below; one can see the red curve fitting sensibly to the datasets. It is not a great fit but with the data we have available it is reasonable. It is another indication that TRMM finds it almost impossible to get a good rainfall estimate in the Western Cape region.

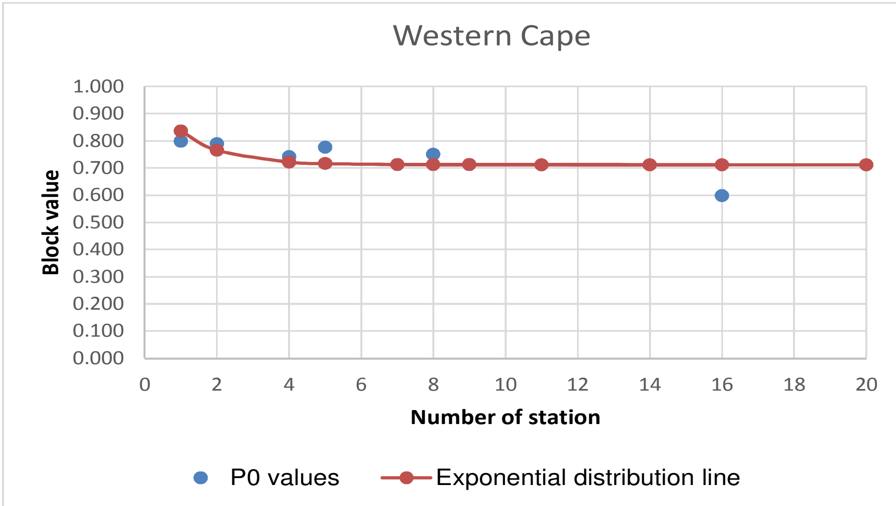


Figure 27: Western Cape Alternative Exponential function values

The p_0 values are one of the important variables in determining the probability of rainfall on a TRMM pixel. What we have discovered here is that as the number of gauges increase in a TRMM block, their collective p_0 value increases to closely agree to the TRMM estimate, something that was only discovered late in this investigation, while working on the WRC report (Pegram et al., 2016). The problem has not been resolved satisfactorily but will be dealt with in follow-up studies.

Conclusion:

The Q-Q transform is a good method for bias correction; it ensured correctness of the probability of rainfall at the four selected sites. It is evident that Q-Q transform does have a weakness in certain areas due to amongst others, topography. What this section has done is to highlight the need for care when using BAGD data to obtain TRMM pixel size probability distribution estimates of rainfall. When there are many gauges in a pixel, the p_0 *value is lower that when there are few. This point needs to be followed up and exploited.*

CHAPTER 6: CONCLUSIONS

The main objective of the research was to find out whether TRMM data is comparable to ground based rainfall data and furthermore to determine whether TRMM data can be used to replace or provide a second option to ground based rainfall data. Point measurements at ground based gauge stations were compared with satellite derived precipitation pixel values at the same locations/areas. The TRMM 3B42 product and BAGD datasets were compared and the comparison was completed on the daily, pentad, monthly and annual data.

Technically, TRMM data provides systematic, multi-year, visible, infrared, and microwave estimates of rainfall in the tropics and subtropics as key inputs to weather and climate research. The TRMM satellite orbited around the Earth and it was not sun synchronous. The data product that was considered in this research project was 3B42RT as it is still being used in several applications within South Africa.

Four study areas were identified with different geographical characteristics which also have dense groups of rainfall gauges within South Africa. The areas were in the Gauteng, KwaZulu-Natal, Limpopo and Western Cape Provinces. The Climate System Analysis Group (CSAG) rainfall database was used to extract ground based rainfall data for the four identified sites. The period considered for the research was from March 2000 until February 2010. In each of the four selected areas, sets of nine 0.25^0 blocks were identified that had the most usable data in terms of reliability, missing gaps and number of stations located in each block, in order to obtain sensible comparisons.

Analysis of the precipitation estimates at the four sites indicates the variety of different rainfall patterns that the country has and gave a view of how well TRMM does at the sites. It is also evident that TRMM does miss some of the key spots, like in the Western Cape's mountainous ranges, which experience Winter rather than Summer rainfall. BAGD shows noisy variation in some areas, which can be attributed to the variability within available record lengths which strongly affects the total.

The results of TRMM/BAGD comparisons in Limpopo and Gauteng show some consistency, as compared to the coastal results like KwaZulu-Natal and Western Cape. It was found that the relationship between TRMM and BAGD datasets depends on the site and factors like topography and the number of gauges in a block. With that in mind, there is a need to look at each site with great care, taking into account its location. It turned out that all direct comparisons of daily and pentad data between the TRMM and BAGD series were not useful, because of the poor timing of TRMM compared to ground-based data. The monthly data were fairly well correlated, but not the annual data because of the very few years of available data. Nevertheless, the message to carry forward is that monthly TRMM values can be used in the summer rainfall regions directly in agrohydrological applications in South

Africa, based on the good results obtained from the well populated TRMM blocks in Gauteng. The sparsely populated blocks need further treatment.

A difficulty was encountered when comparing data from gauges with those from TRMM in that they provide two different kinds of information. TRMM estimates provide an average precipitation over an area of the 0.25° pixel while a gauge provides measurements at a point. Hence, one of them was transformed to the format of the other to make them comparable. Therefore the final chapter focused on the introduction of a novel idea for performing a quantile-quantile (Q-Q) transform of TRMM to Block averaged gauge daily rainfall and highlighted the problem of gauge sparseness. The ideas introduced in previous chapters were used and the technique of Q-Q transform methods and the variability of the probability of zero rainfall with instrument was introduced. The intent here was to determine if there is any useful link between TRMM and BAGD such that given an area with poor or no gauges, TRMM data can be used to estimate the actual rainfall using the parameters suited for the area in question.

The Q-Q transform was found to be a good method for bias correction; it ensured correctness at the four selected sites. It is evident that the Q-Q transform does have a weakness in certain areas due to amongst others, topography. What the last section has done is to highlight the need for care when using BAGD data to obtain TRMM pixel size probability distribution estimates of rainfall; among other points, when there are many gauges in a pixel, the p_0 value is lower that when there are few.

In summary, the research findings indicate that it is likely that TRMM data will be useful for large-scale hydrology and agriculture, particularly at the monthly scale, in contrast with daily. Thus crop monitoring and reservoir storage calculations will benefit, but not flash floods, i.e. TRMM is useful in a coarse way, but poor in detail.

REFERENCES

Journal articles

Huffman J. George, Adler RF., Bolvin DT., Guojun Gu, Nelkin EJ., Bowman KP., Hong Y., Stocker EF., Wolf DB. , (2007) TRMM multisatellite precipitation analysis (TMPA): Quasi-Global, multiyear, combined-sensor precipitation estimates at fine scale. Journal of Hydrometeorology 8: 38 – 55.

Kummerow C.,Barnes W., Kozu T., Shiue J., Simpson J. , (1998) The tropical rainfall measuring mission sensor package. Journal of Atmospheric and Oceanic Technology 15: 810 – 817

Pegram G., Bárdossy A. (2013) Downscaling regional circulation model rainfall to gauge sites using recorrelation and circulation pattern dependent quantile – quantile transforms for quantifying climate change. Journal of Hydrology 504: 142 – 159.

Pegram G.C. and, Pegram G.G.S. (1993) Integration of rainfall via multiquadric surfaces over polygons. Journal of Hydraulic Engineering 119(2): 151 – 163.

Tian Y., Pieters-Lidard CD., Choudhury BJ., Garcia M. (2007) Multitemporal analysis of TRMM-Based satellite Precipitation products for land data assimilation applications. Journal of Hydrometeorology 8: 1165 – 1183.

Books

Langtangen HP. (2011). A Primer on Scientific Programming with Python (2nd edn). Springer Heidelberg Dordrecht London New York.

McKinney W. (2012). Python for Data Analysis (1st edn). O'Reilly Media.

Sheppard K. (2014). Introduction to Python for Econometrics, Statistics and Data Analysis. University of Oxford.

Chapter in a book

Huffman, G J, Adler, RF, Bolvin, DT, and Nelkin, E J (2010). Satellite Rainfall Applications for Surface Hydrology. Gebremichael, Mekonnen, Hossain, Faisal (Eds.). – TRMM Multi-Satellite Precipitation Analysis (TMPA). Pp 3 - 22

Published research reports

Pegram G., Sinclair S., Bárdossy A. (2013). Modelling daily rain-gauge network measurement responses under changing climate scenarios. Water Research Commission Report No. K5 - 1964

Pegram GGS, Sinclair DS, Bárdossy AB, Ngoepe S* (2016). New Methods of Infilling Southern African Raingauge Records Enhanced by Annual, Monthly and Daily Precipitation Estimates Tagged with Uncertainty, WRC Report No. 2241/1/15 ISBN 978-1-4312-0758-9

Government legislation

Republic of South Africa Government Gazette. 1998. National water act. 1998. Government Gazette 398. Republic of South Africa. Act 36 of 1998. Government Gazette, South Africa 398 (19182).

Conference/workshop proceedings

Barrett EC. (1997) Satellite rainfall monitoring: recent progress and remaining problems. In: Proceedings of Robat Symposium S3, April 1997 IAHS Publication, pp 141 – 148.

Goodrich D.C., Amitai E., Crow W.T., Hsu K., Wang J., AghaKouchak A., Bras R., Gochis D., Sorooshian S., Zeng X., Unkrich C., Habib E., Thill B., (2012) In-situ Verification of TRMM Radar Rainfall Intensity Estimates and a Strategy for GPM Ground Validations in the Southwest, 5th International Workshop for GPM Ground Validation, 10-12 July 12, 2012, Toronto, Canada.

Pegram GGS., Sinclair S., Ngoepe S*. (2015) Comparing interpolated daily gauge rainfall over RSA with TRMM to determine possible bias correction for hydrological applications. In: Proceedings of South African Society for Atmospheric Sciences (SASAS) Conference 48th, September

Electronic literature

About NASA. Retrieved from https://www.nasa.gov/about/index.html

An overview of TRMM by Goddard Earth Sciences Data and Information Services Center (GES DISC). Available at

http://disc.sci.gsfc.nasa.gov/precipitation/trmm_intro.shtml

Precipitation measuring mission instruments. 1997. From:

http://pmm.nasa.gov/TRMM/trmm-instruments

Huffman GJ., Pendergrass A., National Centre for Atmospheric Research Staff (Eds). (2015). The climate data guide: TRMM: Tropical Measuring Mission. “Retrieved from

https://climatedataguide.ucar.edu/climate-data/trmm-tropical-rainfall-measuring-mission#sthash.nGKorOjd.dpuf

Pegram G. (2008). The math behind the model, least squares: from Gauss to Kalman. University of KwaZulu-Natal, South Africa

Toshiyuki Iwamiya, Director, Program Management and Integration Department of Aerospace Research and Development Directorate, Director, Program Management and Integration Department of Aviation Program Group. 2009. JAXA Technology Challenge. JAXA

https://eeepitnl.tksc.jaxa.jp/mews/en/20th/data/1_13.pdf

TRMM project mission. Retrieved from: http://science.nasa.gov/missions/trmm/ .