

VALIDATION OF SATELLITE DERIVED RAINFALL ESTIMATES OVER KENYA.

BY

EZEKIEL MUGAI NJOROGI



A PROJECT SUBMITTED IN PARTIAL FULFILMENT FOR THE DEGREE OF MASTER OF SCIENCE IN METEOROLOGY AT THE UNIVERSITY OF NAIROBI.

AUGUST, 2010.

DECLARATION

This study is my original work and has not been presented for a degree in any other University.

Signed  24/08/2010


EZEKIEL MUIGAI NJOROGE

ID NO. I56/71203/2007


This project has been submitted for examination with our approval as University Supervisors.

Signed  24/08/10

Prof. N. J. MUTHAMA

Signed  24/08/2010

Dr. G. O. OUMA

Signed  24/08/2010

Mr. C. B. LUKORITO

DEPT. OF METEOROLOGY & CLIMATE
UNIVERSITY OF NAIROBI

DEPARTMENT OF METEOROLOGY

UNIVERSITY OF NAIROBI

P. O. BOX 30197

NAIROBI, KENYA

ABSTRACT

The aim of the study was to determine the validity of satellite derived rainfall estimates for use in Kenya. This will improve monitoring of rainfall in Kenya. The study analyzed eleven years of monthly rainfall estimates (1998-2008) produced by Tropical Rainfall Measuring Mission (TRMM)'s 3B-43 algorithm and compared them with gridded monthly rainfall totals from twenty six synoptic and Agrometeorological stations over Kenya for the same period.

The quality of rain gauge records was assessed before they were included in this study. These monthly rain gauge records were then gridded using Kriging method to a grid scale of 0.25° by 0.25° to match with the TRMM satellite's rainfall estimates. The two gridded data sets were then compared by plotting scatter diagrams for a dry season (January to February), a wet season (March to May) and for the whole period of study. From the scatter plots, both variables were compactly arranged displaying a strong joint behaviour (correlation) of these two data sets.

The results from simple correlation analysis revealed significant correlation between gridded rainfall and TRMM satellite's rainfall estimates. The largest value of correlation obtained was 0.9 for the dry season while the smallest value was 0.847 in wet season.

Principal component analysis was performed in both spatial and temporal modes to investigate the underlying physical processes which gave rise to the two data sets. The first principal components were presented on spatial maps for both Kriged rainfall and satellite rainfall estimates respectively. A pair of maps with similar patterns, one for each variable, was obtained for the wet season, dry season and for the whole period of study. Time series of the first and the second principal components of the Kriged rainfall and TRMM satellite's rainfall estimates revealed a common trend.

The mean absolute percentage errors ranged from zero to 99.9% for PCA's spatial mode against a target of 10% and threshold of 15%. The temporal mode

generated large errors ranging from 0.27% to 212%, however some of these high errors in this mode are outliers since the average was 44.4%. Mean errors (Bias) ranged from -0.024 to +0.03 for spatial mode while temporal mode had errors ranging from -0.093 to +0.09 against the target of ± 0.025 and threshold of ± 0.038 . Mean absolute errors (MAE) in spatial mode ranged from 0.009 to 0.061 while those in temporal mode which ranged from 0.033 to 0.1727. The targeted and threshold MAE values were 0.025 and 0.038 respectively. Root mean squared errors (RMSE) ranged from 0.039 to 0.22 and 0.012 to 0.06 for temporal and spatial modes respectively. The RMSE targeted and threshold values were 0.027 to 0.041. From these results, temporal mode generated larger errors than the spatial mode. In spatial mode more than half of the country had errors within the acceptable range. The TRMM's 3B43 algorithm tended to overestimate rainfall during the wet season.

Canonical correlation analysis was done to determine a linear combination of each of the two sets of variables such that the correlation between two functions is maximized. CCA which is equivalent to multiple regressions was also used to develop models for estimation of areal rainfall using satellite derived rainfall estimates. The results from CCA revealed high correlation coefficients between Kriged rainfall and TRMM rainfall estimates. For the overall period of study eight out of ten Eigen vectors analyzed had CCA coefficients greater than 0.5.

Based on these results it was concluded that satellite rainfall estimates can be modeled to represent areal rainfall in areas with inadequate ground based rainfall observations, especially over Northwestern, Northern, Northeastern and Southern Kenya. In this study an attempt was done to design these models.

DEDICATION

*This work is dedicated to my Parents, my wife Grace, my
daughter Sarah and my son Daniel,*

With great respect and love.

ACKNOWLEDGEMENT

I wish to thank all those who assisted me during the progress of this study. Special thanks go to Professor J. N. Muthama, Dr. G. O. Ouma and Mr. C. B. Lukorito all of Department of Meteorology, University of Nairobi for their personal interest and willingness to help me throughout this study. Without their valuable guidance, this work would never have come in this form.

Special thanks to Mary C. Greene and all the staff of Goddard Earth Sciences Data and Information Services Centre, National Aeronautics and Space Administration Centre, (NASA) for providing me with the satellite data. My appreciation goes to Kenya Meteorological Department for providing me with rain gauge data without which this work could have been impossible.

My profound thanks go to the Ministry of Transport, Kenya and Regional Universities Forum for Capacity Building in Agriculture, (RUFORUM) for the financial support for this study.

I am thankful to all my colleagues at the University of Nairobi who have contributed to this work through their helpful discussions and criticism.

My deep thanks to my mother and father who sacrificed whatever they could to let me get high education. Last but most importantly, I express my sincere gratitude to my wife Grace and my children, Sarah and Daniel for their permanent love and spiritual support during my study.

**EZEKIEL MUIGAI NJORGE,
UNIVERSITY OF NAIROBI
AUGUST, 2010.**

TABLE OF CONTENTS

Title	i
Declaration	ii
Abstract	iii
Dedication	v
Acknowledgement	vi
Table of contents	vii
List of Figures	xi
List of Tables	xiv
List of Acronyms	xv

CHAPTER ONE.....1

1	Introduction	1
1.1	Background	1
1.2	Objectives of the study	4
1.3	Justifications of the Study	5
1.4	Area of study	5

CHAPTER TWO 7

1	Literature review	7
2.1	Rainfall interpolation methods	7
2.2	Satellite based rainfall estimation methods and finding.....	9
2.2.1	Tropical Rainfall Measuring Mission (TRMM)	9
2.2.2	Radar based techniques	13
2.2.3	Visible / Infrared techniques	15

2.2.4	Passive microwave techniques	16
2.2.5	Data fusion procedures and findings	16
2.3	Comparison of rain gauge, precipitation radar and satellite based rainfall estimates	17
CHAPTER THREE		21
3	Data and Methodology	21
3.1.1	Rainfall data	21
3.1.2	TRMM rainfall data and Algorithm	22
3.2	Methodology	23
3.2.1	Data quality control	23
3.2.1.1	Estimation of missing rain gauge data and satellite derived rainfall estimates	23
3.2.1.2	Homogeneity test	24
3.2.2	Gridding	24
3.2.2.1	Derivation of Kriging equation	25
3.2.3	Principal component analysis	27
3.2.4	Scatter plots	29
3.2.5	Correlation analysis	30
3.2.5.1	Simple Correlation analysis	30
3.2.5.2	Canonical correlation analysis (CCA)	30
3.2.5.2.1	Computation of canonical correlation analysis	31
3.2.6	Error Analysis	33
3.2.6.1	Mean absolute percentage error method (MAPE)	34
3.2.6.2	Bias or mean Error	34
3.2.6.3	Mean Absolute error (MAE)	34
3.2.6.4	Root Mean Square Error (RMSE)	35

3.2.7	CCA models	36
CHAPTER FOUR		38
4	Results and Discussions	38
4.1	Quality control of the data	38
4.2	Validation	40
4.2.1	Kriging	40
4.2.2	Principal Component Analysis	43
4.2.2.1	PCA in S-mode method	43
4.2.2.1.1	PCA in S-mode method for the overall period	43
4.2.2.1.2	PCA in S-mode method for the dry season	47
4.2.2.1.3	PCA in S-mode method for a Wet season	50
4.2.2.2	PCA in T-mode method	52
4.2.3	Time series analysis	54
4.2.4	Canonical correlation	55
4.2.4.1	Canonical correlation for the overall period of study	55
4.2.4.2	Canonical correlation for seasons	57
4.2.5	Error Analysis	58
4.2.5.1	Mean absolute percentage error method (MAPE)	58
4.2.5.1.1	Mean absolute Percentage error in temporal mode	59
4.2.5.1.2	Mean absolute Percentage error in spatial mode	59
4.2.5.2	BIAS (Mean error)	62
4.2.5.2.1	BIAS (mean error) in temporal mode	62
4.2.5.2.2	BIAS (mean error) in spatial mode	63
4.2.5.3	Mean absolute error (MAE)	66
4.2.5.3.1	Men absolute error (MAE) in temporal mode	66

4.2.5.3.2	Mean absolute error (MAE) in spatial mode	67
4.2.5.4	The root mean square error (RMSE)	69
4.2.5.4.1	The root mean square error (RMSE) in temporal mode	69
4.2.5.4.2	Root Mean squared error (RMSE) in spatial mode method ...	70
4.3	Canonical correlation analysis (CCA) models	72
4.3.1	Canonical correlation analysis (CCA) spatial models	73
4.3.1.1	CCA spatial models for the overall period	73
4.3.1.2	CCA spatial models for dry season	76
4.3.1.3	CCA spatial models for wet season	79
4.3.2	CCA temporal models	80
 CHAPTER FIVE		82
 5	 Summary Conclusion and recommendations	 82
5.1	Summary	82
5.2	Conclusion	83
5.3	Recommendations	85
 CHAPTER SIX		86
 6	 References	 86

LIST OF FIGURES

Fig 1:	Map of Kenya (area of study) and rainfall stations	6
Fig 2:	TRMM space craft	10
Fig 3:	TRMM Microwave Imager	11
Fig 4:	East African Climatic zones for MAM season	29
Fig 5a:	Mass curve for Dagoretti station	38
Fig 5b:	Mass curve for Moyale station	39
Fig 5c:	Mass curve for Kisii station	39
Fig 6a:	scatter plot of long term monthly means kriged and satellite rainfall estimates	41
Fig 6b:	scatter plot for kriged rainfall against TRMM rainfall estimates for January to February	42
Fig 6c:	Scatter plot Kriged rainfall against Trmm rainfall estimates for March to May	42
Fig 7:	Scree plot for latent roots of the Kriged rainfall (PCA S-mode method)	45
Fig 8:	Scree plot for latent roots of the TRMM rainfall (PCA S-mode method)	45
Fig 9:	Pattern of the first Eigen vector (first principal component) of Kriged rainfall for the overall period	46
Fig 10:	Pattern of the first Eigen vector (principal component) of TRMM rainfall estimates for the overall period	47
Fig 11:	Pattern of the first Eigen vector of TRMM rainfall for January to February.....	49
Fig 12:	Pattern of the first Eigen vector of Kriged rainfall for January to February.....	49
Fig 13:	Pattern of the first Eigen vector of Kriged rainfall for March to May	51
Fig 14:	Pattern of the first Eigen vector of TRMM rainfall estimates for March to May	51
Fig 15:	Plot for latent roots of the Kriged rainfall (PCA T-mode)	53

Fig 16:	Scree plot for latent roots of the TRMM rainfall (PCA T-mode)	53
Fig 17:	Time series analysis for the first principal component of TRMM and Kriged rainfall from January 1998 to December 2008	54
Fig 18:	Time series analysis for the second principal component of TRMM and Kriged rainfall from January 1998 to December 2008.....	55
Fig 19:	The temporal representation of the mean absolute percentage error (MAPE)	59
Fig 20:	The overall spatial representation of the mean absolute percentage error (MAPE)	60
Fig 21:	The spatial representation of the mean absolute percentage error (MAPE) for a dry season (January to February)	61
Fig 22:	The spatial representation of the mean absolute percentage error (MAPE) for a wet season (MAM)	62
Fig 23:	The temporal representation of the mean error (BIAS)	63
Fig 24:	The overall spatial representation of the Mean error (BIAS) ...	64
Fig 25:	The spatial representation of the Mean error (BIAS) for a dry season (January to February)	64
Fig 26:	The spatial representation of the Mean error (BIAS) for a wet season, (March to May),.....	65
Fig 27:	The temporal representation of the mean absolute error (MAE)	66
Fig 28:	The overall spatial representation of the Mean absolute error (MAE)	67
Fig 29:	The spatial representation of the Mean absolute error (MAE) for a dry season (January to February)	68
Fig 30:	The spatial representation of the Mean absolute error (MAE) for a wet season (March to May)	68
Fig 31:	The temporal representation of root mean square error (RMSE)	69
Fig 32:	Overall spatial representation of the Root Mean Square Error (RMSE)	70

Fig 33:	The spatial representation of the Root mean square error (RMSE) for a dry season (January and February)	71
Fig 34:	The spatial representation of the Root mean square error (RMSE) for a wet season (March to May)	72
Fig 35:	TRMM rainfall predictor (a) and Areal rainfall predictand (b) CCA loadings for mode 1 for overall period.....	74
Fig 36:	TRMM rainfall predictor (a) and Areal rainfall predictand (b) CCA loadings for mode 2 for overall period	75
Fig 37:	TRMM rainfall predictor (a) and Areal rainfall predictand (b) CCA loadings for mode 3 for overall period	75
Fig 38:	TRMM rainfall predictor (a) and Areal rainfall predictand (b) CCA loadings for mode 1 for January to February	76
Fig 39:	TRMM rainfall predictor (a) and Areal rainfall predictand (b) CCA loadings for mode 2 for January to February	77
Fig 40:	TRMM rainfall predictor (a) and Areal rainfall predictand (b) CCA loadings for mode 1 for March, April and May	78
Fig 41:	TRMM rainfall predictor (a) and Areal rainfall predictand (b) CCA loadings for mode 2 for March, April and May	78
Fig 42:	TRMM rainfall predictor (a) and Areal rainfall predictand (b) CCA loadings for mode 3 for March, April and May	79
Fig 43:	Canonical component predictor (TRMM rain) and predictand (Kriged rain) for mode 1 in T-mode	80
Fig 44:	Canonical component predictor (TRMM rain) and predictand (Kriged rain) for mode 2 in T-mode	81
Fig 45:	Canonical component predictor (TRMM rain) and predictand (Kriged rain) for mode 3 in T-mode	81

LIST OF TABLES

Table 1:	List of rain gauge stations and their location	21
Table 2:	Target and threshold values for MAPE, BIAS, MAE and RMSE.....	35
Table 3:	Eigen vectors, Eigen values and the percentage of the variance for Kriged rainfall when subjected to PCA S-mode method	44
Table 4:	Eigen vectors, Eigen values and the percentage of the variance for Kriged rain and TRMM rainfall for January to February when subjected to PCA S-mode method	48
Table 5:	Eigen vectors, Eigen values and the percentage of the variance for Kriged and TRMM rainfall for March to May when subjected to PCA S-mode method	50
Table 6:	Eigen vectors, Eigen values and the percentage of the variance for Kriged and TRMM rainfall in PCA T-mode	52
Table 7:	Canonical correlation for S-mode for the overall period of study	56
Table 8:	Canonical correlation for T-mode for the overall period of study	56
Table 9:	CCA results for a dry season (January to February)	57
Table 10:	CCA results for a wet season (March to May)	58

LIST OF ACRONYMS

APMEL	Atmospheric Administration / Pacific Marine Environment Laboratory
AVHRR	Advanced Very High Resolution Radiometer
BMRC	Bureau of Meteorology Research Centre Australia
CAMS	Climate Assessment and Monitoring System
CCA	Canonical Correlation Analysis
CERES	Clouds and Earth's Radiant Energy System
DSD	Drop size distribution
EOC	Earth Observation Center
EOS	Earth Observing System
GDP	Gross Domestic Product
GES	Goddard Earth Science
GPCC	Global Precipitation Climatology Centre
IPGEG	International Precipitation Working Group
ITCZ	Inter-Tropical Convergence Zone
JAXA	Japan Aerospace Exploration Agency
KDP	Specific differential phase
KOUN	Norman, Oklahoma weather surveillance radar
KR	Kwajalien ground based radar
LIS	Lightning Imaging System
MAE	Mean Absolute Error
MAM	March, April and May
MAPE	Mean Absolute Percentage Error
NASA	National Aeronautical and Space Administration
NASDA	National Space Development Agency
NN	Neural Network
NOA	National Oceanic and Atmospheric Administration
OND	October, November and December
PCA	Principal Component Analysis
PR	Precipitation Radar
R	Rain rate

r	Correlation coefficients
RMSE	Root Mean Squared Error
S-mode	Spatial mode
SSM/I	Special Sensor Microwave Imager
TMI	TRMM Microwave Imager
T-mode	Temporal mode
TOVAS	TRMM online visualization and analysis system
TRMM	Tropical Rainfall Measuring Mission
VIRS	Visible and Infrared Radiometer System
Z	Radar reflectivity factor
Z_{dr}	Differential reflectivity
Z_h	Horizontal polarization

CHAPTER ONE

1 Introduction

1.1 Background

Precipitation is one of the major components of the earth's climate system. It is the result of upward movement of air, which causes cooling by expansion beyond the level of condensation. Precipitation takes various forms such as rain, snow and hail. In the tropics precipitation consists almost entirely of rainfall, Nieuwolt, (1977).

Many countries in the tropics depend on rainfall for the agricultural and hydrological activities which are dominant in their economies. For example, in Kenya, agriculture is the mainstay of the economy supporting the livelihoods of about eighty per cent of the rural population. The agricultural sector employs 70% of the national labor force through forward and backward industrial linkages, thus providing food and incomes to individuals and households. Agriculture's contribution to Kenya's GDP is twenty six percent, Omiti et al, (2009).

The process of precipitation provides an amount of latent heat release to the atmosphere. This heating is one of the strongest driving forces of the atmospheric general circulation in the tropics, Fisher, (2004).

Rainfall information is therefore a crucial aspect not only for sustainable social-economic development of many countries but also for study of atmospheric circulations, climate analysis and global energy balance. Hence it is important to use reliable and accurate rainfall data in any planning.

Rainfall data can be acquired through in situ measurements or estimation methods.

One of the oldest and most common methods of measuring precipitation amount is to collect and measure amount of rainfall over a set period of time using a rain gauge. Methods that use standard rain gauges involve the conversion of the point

rainfall values given by a network of rain gauges in to an estimate of the areal average for the catchment. These methods include the arithmetic mean, Thiessen's polygons, Isopercentile and Isohyetal, Ouma, (1988).

Most of rainfall data in Kenya is derived from rain gauge records. They also only accurately indicate rainfall in a localized area, Ouma, (1988). Although improvements in satellite observations of precipitation are sought, it should be noted that gauge measurements provide the only long-term direct measure of precipitation and should not be overlooked, Gruber and Levizzani, (2008). Recently, more accurate automatic weather stations have been introduced but their network is too sparse due to their high cost.

In many parts of the world especially African continent the rain gauge network is too sparse to produce reliable areal rainfall estimates. In Kenya, the number of rain gauge stations is limited and the distribution of rain gauges is very uneven, with most stations located near main towns. As a result, these gauges may not represent the rainfall over the rural areas where the information is needed most. Even where data exist, there are other limitations including short historical time series, missing data and reading errors.

Many researchers are therefore focusing upon getting better and alternative methods of collecting areal rainfall data. One of the modern methods being used to acquire rainfall data is remote sensing. Remote sensing is the science and art of obtaining information about an object, area or phenomenon through the analysis of data acquired by a device that is not in contact with the object, area or phenomenon under investigation, Lillesand, (1979). Remote sensing methods of rainfall estimation use information received from radar and weather satellite based sensors.

The radar based techniques for rainfall estimation involves the hypothesis of raindrop size distribution (DSD). This hypothesis states that the rain drops are as homogeneously distributed in space and time as randomness allows. This hypothesis forms the basis of the sampling theory of in situ rainfall observations

and can be considered one of the cornerstones of the physical theory of precipitation induced pulse-to-pulse echo fluctuations in weather radar observations, R. Uijlenhoet et al, (2009). The rain drop size distribution determines the microwave radiometric brightness temperature, the attenuation and optical extinction as well as liquid water content, Nzeukou A. et al, (2004). The drop size distribution (DSD) describes the distribution of rain water in the atmosphere and is used to derive relationships between bulk rain variables (total liquid water per unit of air volume, the attenuation and scattering of electromagnetic waves propagation through the rain and the rain erosivity) and the rain intensity. The rain drop distribution is expressed using distribution function that gives the mean number of drops per unit of air volume, Sempere Torres et al, (1994). The drop size distribution is characterized by three parameters namely diameter, concentration and shape of the drops. These parameters are used to develop algorithms based on radar observables (the radar reflectivity factor of rain (Z) and rain rate (R)) or polarimetric radar algorithms based on reflectivity at horizontal polarization, Z_h , differential reflectivity, Z_{dr} and specific differential phase k_{dp} , for rainfall estimation, Bringi et al, (2003).

The use weather radar is hampered by a few problems. The area of effectiveness of radar is relatively smaller than that of satellite. The intensity of the radar beam decreases with the distance from the radar and hence for coverage of an adequate area, many radar stations need to be set up. Serious difficulties arise in mountainous regions where the beam loses intensity due to blocking. Very heavy precipitation may also reduce the intensity of the beam.

The satellite techniques of rainfall estimation use sensors to collect radiation from clouds and rain drops. Algorithms are then used to estimate the surface rainfall intensity and amount based on physical and statistical relations between radiation and precipitation, Fisher, (2004).

Satellite based remote sensors have many advantages over rain gauge based measurements. These include the spatial coverage, especially over the oceans where rainfall measurements are sparse. Temporal sampling capabilities for large areas are also a major advantage. Only timely satellite estimates permit the

development of real time or near-real time applications. Satellite data are also more homogeneous and objective than gauge data, Tapiador et al, (2004).

The satellite-derived rainfall estimates need to be validated so that they can be used for many diverse meteorological, climatic, hydrological, agricultural and other applications Nicholson et al, (2003). Validation is a systematic process for reviewing data against a set of criteria to provide assurance that the data are adequate for their intended use.

This study aims at validating satellite-derived rainfall estimates retrieved from TRMM's monthly rainfall retrieval algorithm, (3B-43 algorithm), over Kenya. The details of the objectives are highlighted in the following section.

1.2 Objectives of the study

The overall objective of this study is to determine the validity satellite derived rainfall estimates over Kenya using gridded rainfall field. This will enhance the usage of these estimates in areas with inadequate rain gauges and therefore strengthen the monitoring of rainfall in this country.

The specific objectives are:

- Develop gridded rainfall field based on the rain gauge data.
- Perform a principal component analysis on data set to identify patterns in the data and to highlight their similarities and differences.
- Determine the validity of the TRMM rainfall estimates.
- Develop temporal and spatial models for estimating areal rainfall using TRMM satellite's estimates.

1.3 Justifications of the Study

Weather and climate are important aspects in social-economic planning, climate analysis study of atmospheric circulations and global energy balance. Despite its importance, precipitation is one of the most difficult atmospheric parameters to measure because of large variability in space and time, Kim (2004). It is especially difficult to measure precipitation over mountainous regions and oceans because surface-based observations such as radars and rain gauges are rare, Fisher (2004). We therefore need to find an alternative method of getting rainfall data. Estimation of precipitation from satellite data will therefore fill in this gap.

Sensors on board satellites observe the distribution of precipitation more homogeneously than rain gauge network. They monitor large areas and yield spatially continuous data.

If the methods of estimating rainfall using data retrieved from satellites are close to the rain gauge observations, then satellites can be used to estimate precipitation in areas where rain gauge network is sparse. Thus validation should be done to ensure confidence in satellite derived rainfall estimates for various meteorological and hydrological applications, Nicholson et al, (2003).

1.4 Area of study

Kenya lies astride the equator and is situated between longitudes 34°E to 42°E and latitudes 5.5°N and 5°S. Uganda borders it to the west, Tanzania to the south, Somali and the Indian Ocean to the east and Ethiopia and Sudan to the north. The total area of the country is about 582,646 sq. Km. Kenya has complex topography ranging from Coastal lowlands to snow-capped mountains among them Kenya (5199 m) and Elgon (4321 m), large lakes, and Great Rift Valley.

The main rainfall generating system over the study area is the convergence zone between the northeast monsoon and the southeast monsoon referred to as inter-tropical convergence zone (ITCZ), Asnani, (2005). The ITCZ moves northward

passing over East Africa during March to May (MAM), while the southward movement takes place during October to December (OND), Okoola,(1999). Local features such as topography and large water bodies generate mesoscale systems which also play role in modulating the weather and climate over the country leading to spatial variation of rainfall, Ininda, (2008).

Figure 1 gives the location of Kenya and the synoptic stations used for this study.

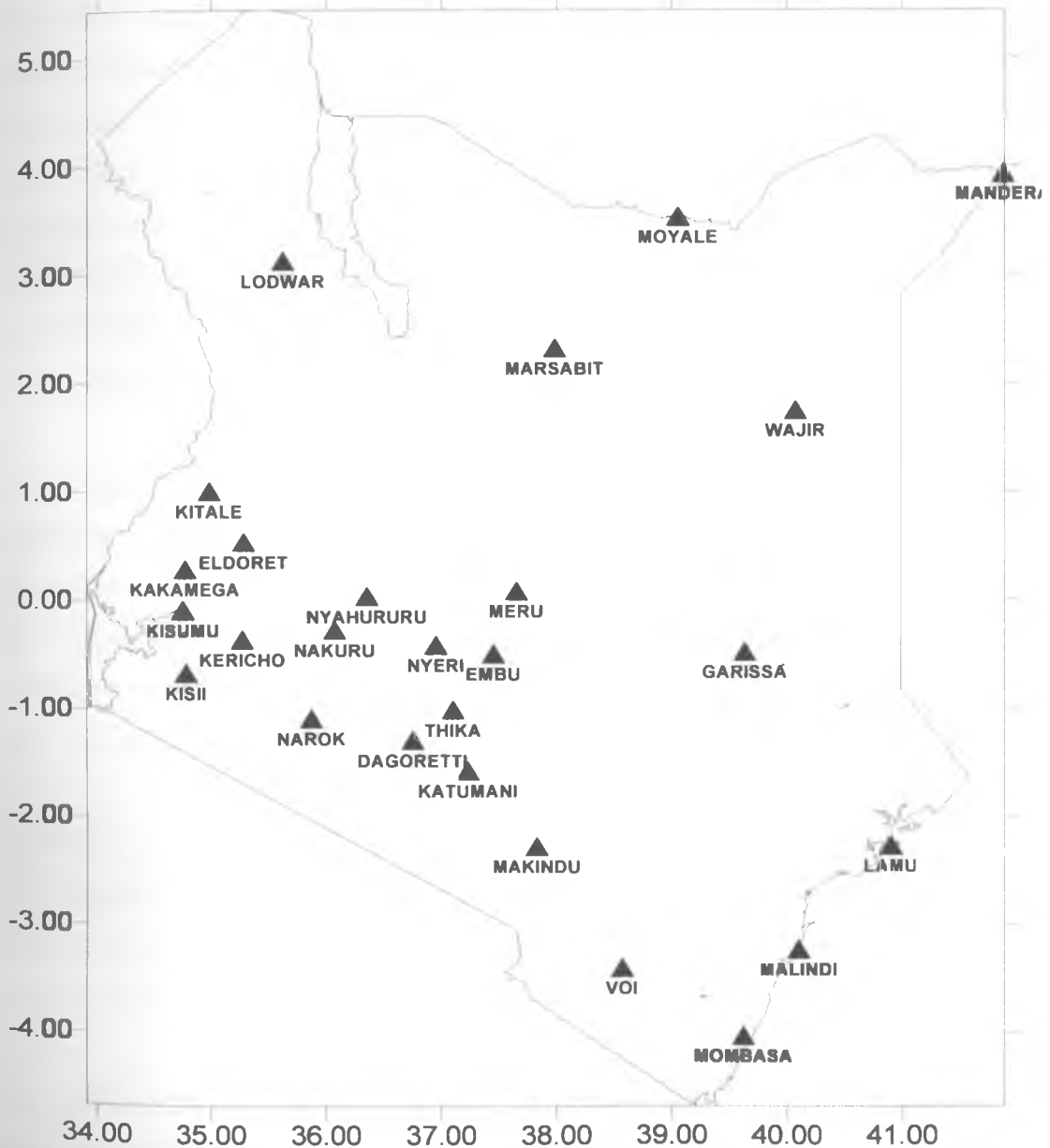


Fig 1: Map of Kenya (area of study) and rainfall stations

CHAPTER TWO

2 Literature review

2.1 Rainfall interpolation methods

Interpolation is the method of constructing new data points within the range of a discrete set of known data points. Data sets of spatially irregular meteorological observations are interpolated to a regular grid and are important for climate analyses, Naoum and Tsanis, (2004). Interpolated data sets allow best estimates of climate variables at locations away from observing stations, thereby allowing studies of local climate in data-sparse regions. Principal component analysis and canonical correlation analysis techniques prefer regularly spaced observations so as not to bias the eigenvalues to regions with a higher density of observations, Haylock et al, (2008).

Interpolation of observed data sets to regular or equal grids is referred to as gridding. Construction of a gridded data set where each grid value is a best estimate average of the grid square observations is the most appropriate data set for validation of the model or algorithm estimates, rather than comparison between the model/ algorithm estimates and point observations directly, Haylock et al, (2008).

There is a variety of spatial interpolation methods which range from simple intuitive predictions to more complex procedures. Among these techniques are Kriging, Spline, Inverse distance weighting, Thiessen polygons, Isohyetal, Trend surface and Polynomial, Naoum and Tsanis, (2004).

Taesombat and Nuchanart, (2009), compared Thin plate spline technique and two conventional techniques, the isohyetal and Thiessen polygon techniques to approximate areal rainfall over the upper Ping river basin in Thailand. They did this study for the purpose of improving the accuracy of runoff and flood estimation for mitigation of water related problems. Two data sets of maximum rainfall

registered in August 2001 and September 2003 at 68 non-automatic rainfall stations located in the basin and nearby areas were used in the analysis. The Thin plate spline technique proved to provide more accurate results of rainfall estimation than the other two techniques.

Sun and Ebert, (2003), combined rain gauge measurements with satellite data for the year 1997-1998 to estimate rainfall over Australia. The study was done in order to provide better and more effective ways of improving daily raining area delineation and spatial rainfall interpolations. They used rain gauge data from a network of one thousand and five hundred stations which collect data on a daily real time basis. Estimates were generated in two steps:

- The indicator Kriging technique was used first to delineate the raining areas.
- Then ordinary Kriging was used to determine the rainfall estimates in these areas.

To implement indicator Kriging technique, a binary variable was used to describe the rain gauge rainfall. The satellite rainfall probability of occurrence was calculated from its infrared temperature. They used CoKriging to merge the two spatial observational data sets. Using a two-year period of rain gauge and satellite data, the results of delineating the raining areas from indicator Kriging show an improvement over the results from using rain gauge data only and without indicator Kriging, especially in rain gauge data-sparse areas.

Naoum and Tsanis, (2004), ranked some of spatial interpolation techniques used in studying rainfall variability. These techniques are Spline (regularized and tension), inverse distance weighting, Kriging (linear, Gaussian, circular, spherical, exponential and universal 1 & 2), second order Polynomial and Thiessen polygons. Ordinary Kriging which is represented by spherical, circular, exponential, Gaussian and linear methods, was generally found to be the best interpolator and therefore adapted in this study.

2.2 Satellite based rainfall estimation methods and findings

2.2.1 Tropical Rainfall Measuring Mission (TRMM)

Earlier research on satellite rainfall estimation used infrared or visible channels data. These methods include the use of cloud top temperature, cloud brightness, cloud type and duration as parameter in quantifying the area rainfall, Ouma, (1988). However significant errors existed because these channels could not penetrate cloud layers. The rainfall intensity was frequently over estimated when high cirrus cloud or anvil clouds appeared in infrared and visible data.

Microwave channel data are less influenced by cloud layers and are therefore more suitable for rainfall intensity estimation. The importance of microwave remote sensing has been demonstrated by some new experiments such as Tropical Rainfall Measuring Mission (TRMM), Bowman, (2004).

The Tropical Rainfall Measuring Mission (TRMM) satellite was launched in November 1997 as a joint scientific initiative between the National Aeronautical and Space Administration (NASA) of USA and National Space Development Agency (NASDA) of Japan. The TRMM satellite collects precipitation information within a large sampling domain that extends from 40°N to 40°S, Fisher, (2004). The primary rainfall instruments on TRMM are the TRMM Microwave Imager (TMI), the precipitation radar (PR) and the Visible and Infrared Radiometer System (VIRS), (fig 2). Additionally, the TRMM satellite carries two related Earth Observing System (EOS) instruments in the Clouds and Earth's Radiant Energy System (CERES) and the Lightning Imaging System (LIS).

The TRMM Microwave Imager (TMI) is a Multi-channel dual-polarized passive microwave radiometer. It utilizes nine channels with frequencies of 10.65 GHz, 19.35 GHz, 21.3 GHz, 37 GHz and 85 GHz. The TMI instrument provides data related to rainfall rates over oceans, but less reliable data over land, where non-homogeneous surface emissions make interpretation difficult. The TMI data combined with the data from the PR and VIRS are also utilized for deriving

precipitation profiles, JAXA/EOC, (2006). TMI quantifies water vapour, the cloud water and the rainfall intensity in the atmosphere. Figure 3 gives main parts of TMI.

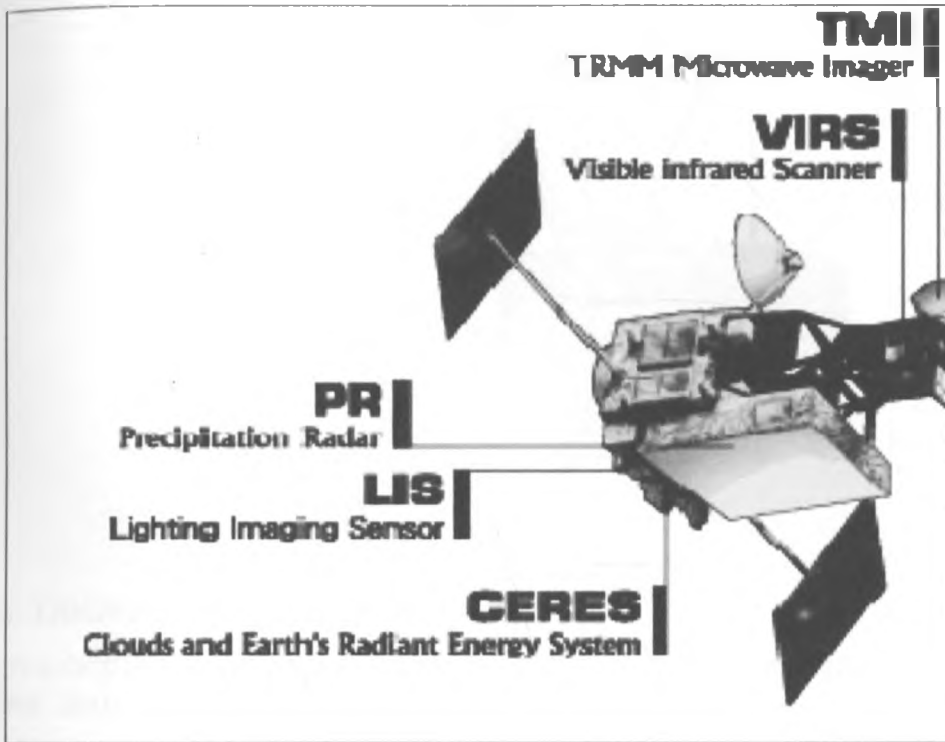


Fig 2: TRMM space craft

Source: http://disc.sci.gsfc.nasa.gov/precipitation/additional/instruments/tmm_instr.shtml, June, 2010.

The VIRS is a five-channel imaging spectro-radiometer with bands in the wavelength range from 0.6 to 12 μm . The VIRS data is used to observe precipitation using visible and IR techniques. It also provides a link between the derived precipitation and similar data both historical and contemporaneous, from geosynchronous and low earth orbiting sensors. Comparison of the microwave data with VIRS visible and infrared data provides the means whereby precipitation is estimated more accurately than by visible and infrared data alone. The VIRS instrument serves as a background imager and provides the cloud context within which the passive microwave and radar observations are made. Data from the VIRS instrument is used in rain estimation algorithms based primarily on the passive and active microwave sensors, JAXA/EOC, (2006).

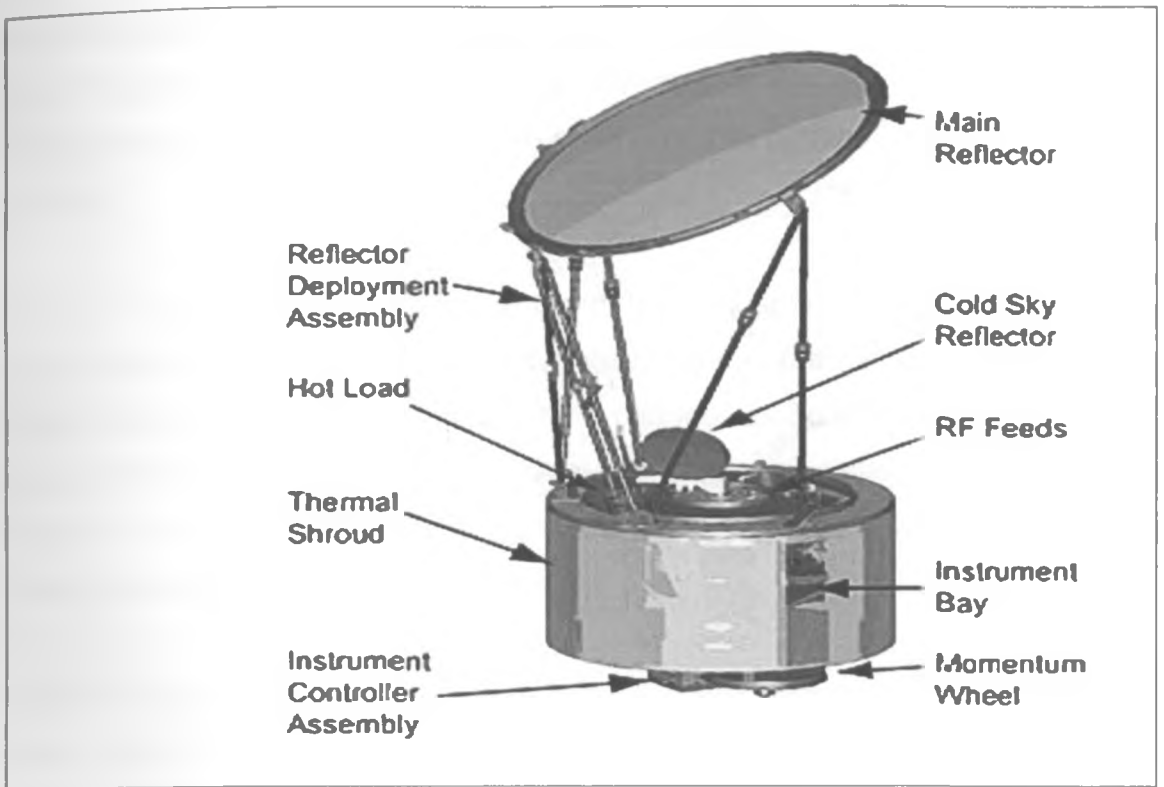


Fig3: TRMM Microwave Imager

Source: http://disc.sci.gsfc.nasa.gov/precipitation/additional/instruments/trmm_instr.shtml, June, 2010.

The Precipitation Radar (PR) is the primary instrument onboard TRMM. The most innovative of the five TRMM instruments, the PR is the first quantitative rain radar instrument to be flown in space, Kummerow et al, (1998). The major objectives of the PR instrument are as follows:

- Provides a 3-dimensional rainfall structure,
- Achieves quantitative measurements of the rain rates over both land and ocean,
- Improve the overall TRMM precipitation retrieval accuracy by combined use of active (PR) and passive (TMI and VIRS) sensor data.

When combined with TMI measurements, the PR data is instrumental in obtaining the height profile of the precipitation content, from which the profile of latent heat release from the Earth can be estimated. The rain rate is estimated from the radar

reflectivity factor when the rain rate is small by applying conventional algorithms used for ground-based radar. For large rain rates, a rain attenuation correction is made using the total-path attenuation of land or sea surface echoes, JAXA/EOC, (2006).

CERES measures the energy at the top of the atmosphere and estimates energy levels within the atmosphere and at the earth's surface. It helps reduce one of the major uncertainties in predicting long-term changes in the Earth's climate. CERES also provides information to determine the surface radiation budget, which is important in atmospheric energetics, studies of biological productivity and air-sea energy transfer, JAXA/EOC, (2006).

Lightning Imaging Sensor (LIS) is an optical staring telescope and filter imaging system that acquires and investigates the distribution and variability of both intra-cloud and cloud-to-ground lightning over the Earth. The LIS data also is used with PR, TMI and VIRS data to investigate the correlation of the global incidence of lightning with rainfall and other storm properties, JAXA/EOC, (2006).

One of the major goals of the TRMM is to produce quantitative and accurate mean monthly rain estimates with errors not exceeding 10%-15% of the true precipitation for $0.25^\circ \times 0.25^\circ$ grid spaces. There are three possible errors caused by instrument noise and statistical effects of thermal emission, retrieval and sampling methods.

However studies have been done to minimize these errors. The study by Fisher (2004) introduced an error model that empirically estimates sampling and retrieval errors for the TMI and PR. The methodology is fundamentally dependent on statistical rain estimates from gauges that have been sub sampled at the satellite overpass times.

2.2.2 Radar based techniques

Active and passive sensors on board satellites are used for precipitation estimation. Relative to passive sensors, important satellite radiation measurements occur for visible/ infrared (VIS/IR) spectrum and microwave (MW) spectrum.

Various techniques to estimate rainfall using these satellites based sensors have been developed and a lot of studies have been done on these techniques.

Precipitation measurement using active microwave from space or ground based sensors involves use of radars. The radar techniques of rainfall estimation involves hypothesis about raindrop size distribution. The rainfall size distribution (DSD) determines the microwave radiometric brightness temperature, the attenuation and optical extinction as well as liquid water content. The radar reflectivity factor of rain (Z) and rain rate (R) are linked by a relation of the form

$$Z = a R^b$$

Where, a , b are coefficients of raindrop size distribution (DSD).

The above approach was used by Nzeukou A. et al, (2004) to estimate precipitation over West Africa. They found out that the shape of the averaged raindrop size distribution (DSD) was very similar from one year to the next. From rain rates R higher than about 20mm/h, the slope of (DSD) was tending towards a constant value. The coefficients of the Z-R relation, between the radar reflectivity factor Z and R were different for convective and stratiform parts of squall lines. However the Z-R relations for convective rain intersect the relation for stratiform rain for high rates. Hence they used a single Z-R relation to correct representation of the whole Z and R range variation in West Africa. The coefficients of Z-R relation were found to be close to that of the stratiform and to that of observed in other West African areas. The conditional probability density function of rain rate, (R), was found to be very stable year after year. The coefficient of variation of R as found to be constant and close to 2.24 and was observed at many other sites. From $P(R)$, the linear coefficient $S(\tau)$ of the relation that links the area-averaged

rain rate to the fractional area where the rain rate was higher than the threshold (τ) was computed and was found to be very stable for values of τ close to MR, the mean climatic value of R (around 5-6 mm/h). Since the sub-Sahara West African sites have a similar MR, comparison showed that $S(\tau)$ was homogeneous over this area. The results suggested that $S(\tau)$ can be used with confidence for average rainfall estimation over a climatically homogeneous region.

Similar approach was applied by Gabremichael et al, (2005) to estimate precipitation over Mississippi River basin. They compared these rainfall estimates with global precipitation climatology project daily rainfall model. The authors found out that in comparison with radar-based estimates the model overestimated rainfall.

Studies have shown that there is improvement in rainfall estimates if polarization radar is used. In addition to conventional radar reflectivity factor Z, polarimetric radar is capable of measuring the differential reflectivity ZDR, specific differential phase KDP and the cross-correlation coefficient between two orthogonally polarized radar returns. This improves the quality of the data.

This approach was used by Ryzhkov et al, (2005) to test the capability of KOUN radar (Norman, Oklahoma weather surveillance radar) to estimate rainfall over Oklahoma. A dense gauge network was used to validate different polarimetric algorithms for rainfall estimation. One hour totals were estimated from the KOUN radar using conventional and polarimetric algorithms and were compared with hourly accumulations measured by gauges. Both point and areal rain estimates were examined. The use of the synthetic polarimetric algorithms resulted in significant reduction in root mean squared errors (RMSE) of hourly rain estimates when compared with the conventional non polarimetric relation.

2.2.3 Visible / infrared techniques

Passive sensors on board satellites measure radiation for visible/infrared spectrum and for the microwave spectrum.

Ouma, (1988) discussed various methods that use the visible and infrared radiation to estimate areal average rainfall. These methods are classified into seven groups as methods that use:-

- Visible and infrared characteristics of clouds separately and/ or collectively,
- Cloud cover and cloud types,
- Cloud top heights or cloud top temperatures,
- Convective cloud area,
- Fractional cloudiness and
- Cold cloud duration in estimating rainfall.

Infrared and visible based rainfall estimation methods offer unique advantages of extensive coverage at relatively high temporal sampling rates.

Hong et al, (2004), used a satellite based algorithm, which extracts local and regional cloud features from infrared ($10.7\mu\text{m}$) geostationary satellite imagery in estimating fine scale, ($0.04^\circ \times 0.04^\circ$ every 30min), rainfall distribution. The algorithm processes satellite clouds into pixel rain rates by:

Separating cloud images into distinctive cloud patches,

Extract cloud features including cloud temperature, geometry and texture,

Clusters cloud patches in to well-organized subgroups and

Calibrates cloud top and rainfall (Tb-R) relationship for the classified cloud groups using gauge-corrected radar hourly rainfall data.

Hong et al, (2004), used radar and gauge rainfall measurements to evaluate the algorithm's rainfall estimates at a range of temporal and spatial scales. The evaluation results showed correlation coefficients of 0.45 to 0.59 for temporal and 0.57 to 0.63 for spatial scales respectively.

2.2.4 Passive microwave techniques

Currently, research is ongoing on microwave channels. This has been enhanced by the Tropical Rainfall Measuring Mission (TRMM). The passive microwave observations from radiometers on board low earth-orbiting platforms have better physical connection to precipitation processes as compared to the visible/infrared sensors that can offer quasi-continuous coverage from space. The measurement by passive microwave radiometers is less sensitive to the presence of cirrus clouds, which is one of the major problems in infrared based rainfall estimation algorithms.

Klepp et al, (2003) used rainfall estimates derived from the Special Sensor Microwave Imager (SSM/I) on board National Oceanic and Atmospheric Administration (NOAA) National Environmental Satellites to study rainfall processes associated with frontal and cyclonic systems over North Atlantic. A multi satellite method was applied for complete coverage of North Atlantic twice a day. Different special sensor microwave imager precipitation algorithms were tested for individual cyclones and compared to the Global Precipitation Project datasets. An independent rainfall pattern and intensity validation method was presented using voluntary observing ship datasets and AVHRR images. The results showed that mesoscale backside rainfall events contributed up to 25% of the total amount of rainfall in North Atlantic cyclones.

2.2.5 Data fusion procedures and finding

Passive microwave based estimates suffer from poor temporal sampling and coarse spatial resolution but determine rainfall more directly. Infrared based sensors provide better temporal resolution and moderate spatial resolution but determine rainfall indirectly by inference from cloud top temperatures. Data fusion is therefore necessary.

Data fusion procedures aim to combine both datasets with precipitation radar datasets without their limitations while reinforcing their strength.

Tapiador et al, (2004) evaluated an operational procedure to produce half-hourly rainfall estimates at a 0.1 spatial resolution which combined rainfall estimated by a Neural Network (NN) approach utilizing passive microwave and infrared satellite measurements. Half hourly rain gauge data over Andalusia, Spain, were used for validation purposes. Results showed fused methodologies improved the performance of estimations.

Nesbitt and Zipser, (2003) combined the Tropical Rainfall Measuring Mission (TRMM) satellite measurements from the precipitation radar and microwave imager to yield a comprehensive 3-year database of precipitation features (PF) throughout the global Tropics ($\pm 36^\circ$ latitude).

Furuzawa and Nakamura, (2005) used Tropical Rainfall Measuring Mission precipitation radar (PR) and TRMM Microwave Imager, to investigate the performance of TMI rainfall estimates. The results showed that TMI underestimates rainfall with low cloud height and overestimate rainfall with high cloud height. Combination of microwave imager and precipitation radar improved rainfall estimates.

2.3 Comparison of rain gauge, precipitation radar and satellite based rainfall estimates

Satellite derived rainfall estimates can only meet acceptable or standard requirements if they are supported by correlative data of known quality and are continually challenged by reliable ground based observations and qualitative science. Hence remote sensing measurements should be validated and calibrated. These activities ensure the scientific value of the data.

Studies in to validating rainfall estimates have been carried out by several scientists. Bowman, (2005), Compared four years of precipitation retrieved from Tropical Rainfall Measuring Mission (TRMM) satellite with data from 25 surface rain gauges on the National Oceanic and Atmospheric Administration / Pacific

Marine Environment Laboratory (NOAA/APMEL) in Tropical Pacific. The author found out that:-

- When precipitation is correlated with itself both in space and/ or time, maximum correlation r were in range of 0.6 to 0.7.
- For large satellite averaging areas, correlation with gauges are smaller and the optimum gauge-averaging time is longer.
- Relative RMSE difference between satellite data and a gauge is in the range of 40% to 70%.
- The bias between gauges and satellite retrieves was estimated by correlating the long-term time-mean precipitation estimates across the set of gauges. TRMM microwave imager (TMI) gave r (squared) of 0.97 and a slope of 0.97 indicating very little bias with respect to gauges.

A similar study was done by Fisher, (2004). He analyzed 4 years of precipitation estimates (1998-2001) produced from data collected by TRMM microwave imager (TMI) and precipitation radar PR and compared them with corresponding estimates computed using 5-min rain accumulation from 66 rain gauges in Oklahoma Masonet. The methodology he applied estimated bulk climate-scale sampling and retrieval errors and bias for the (TMI) and PR at two resolutions: 1° by 1° and 2° by 5° . This approach generated two gauge-inferred gridded estimates of monthly precipitation over the study period:

- G_G , which was computed by performing a complete integration of monthly gauge time series and
- G_S , which consisted of gauge-inferred rain rates sub sampled to TRMM overpass at a gridded resolution of 1° by 1° .

The author found out that the overall random sampling and retrieval errors for PR exceeded the TMI errors for the period of study. PR showed a greater reduction in errors when the scale was increased to 2° by 5° . Annual coefficients of variation

were lower for the PR than the TMI at this scale. PR retrievals biases were positive over all 4 years and exceeded the TMI biases in every year of study.

Kim et al, (2004) validated maritime rainfall retrieved from TRMM microwave radiometer. They used Kwajalien ground based radar (KR) observations as the reference data. They found out that TRMM microwave radiometer overestimated surface rainfall with respect to KR by 16%.

An evaluation of version-5 precipitation radar (PR- algorithm 2A25) and TRMM microwave imager (TRMM-algorithm 2A12) was performed by Nesbitt et al (2004) across the tropics in two ways:-

By comparing long term TRMM rainfall products with Global Precipitation Climatology Centre (GPCC), global rain gauge analysis. This was done to evaluate the overall biases of the TMI and PR relative to 'ground truth' in order to examine regional difference in the estimates

By comparing the rainfall estimates from the PR and TMI on a rainfall feature-by-feature basis within the narrow swath of the PR using a 1 year database of classified precipitation features. This allowed a direct comparison of the estimates with the same sampling area and identified relative biases as a function of storm type.

The authors found out that TMI overestimated rainfall in most of the deep tropics and mid latitudes warm seasons over land with respect to both GPCC gauge analysis and the PR. The PR agreed well with GPCC gauges in deep tropics globally. The analysis by feature type revealed that TMI overestimated rainfall relative to the PR.

Dinku and Anagnostou, (2005) used precipitation radar (PR) rainfall estimates for calibrating overland Tropical Rainfall Measuring Mission (TRMM) microwave imager (TMI) rain algorithm in four geographic regions consisting of Central Africa,

the Amazon, the USA Southern plains and the Ganges-Brahmaputra- Meghna River basin in Southern Asia. The algorithm consisted of:-

- Multi channel based rain screening and convective/stratiform classification schemes,
- Non linear/ linear regressions for rain rate retrieval of stratiform/ convective rain regimes.

They also examined regional differences in the algorithm performance. Year 2000 and 2001 data was used for calibration while validation was done using year 2002 data. The algorithm was compared with the latest version [V6] TRMM product in terms of rain detection and rain rate retrieval error statistics on the basis of PR reference rainfall. The results showed reduction in random error by 24% to 165% for the four regions respectively.

The validation approach by Nesbitt et al, (2004), Bowman, (2005), Fisher, (2004) and Dinku and Anagnostou, (2005) were adapted in this study.

CHAPTER THREE

3 Data and Methodology

3.1.1 Rainfall Data

The rainfall data used in this study was obtained from Kenya meteorological department and consisted of monthly totals of rainfall from 26 synoptic and Agrometeorological stations all over the country for the thirty years (1998 to 2008). The list of rainfall stations used in this study is shown in table 1.

LONGITUDE	LATITUDE	STATION
36.75	-1.3	DAGORETTI
35.28	0.53	ELDORET
37.45	-0.5	EMBU
39.63	-0.48	GARISSA
34.77	0.28	KAKAMEGA
37.23	-1.58	KATUMANI
35.27	-0.37	KERICHO
34.78	-0.68	KISII
34.75	-0.1	KISUMU
34.98	1	KITALE
40.9	-2.27	LAMU
35.62	3.12	LODWAR
37.83	-2.28	MAKINDU
40.1	-3.23	MALINDI
41.87	3.93	MANDERA
37.98	2.32	MARSABIT
37.65	0.08	MERU
39.62	-4.03	MOMBASA
39.05	3.53	MOYALE
36.07	-0.28	NAKURU
35.87	-1.1	NAROK
36.35	0.03	NYAHURURU
36.95	-0.42	NYERI
37.1	-1.02	THIKA
38.57	-3.4	VOI
40.07	1.75	WAJIR

Table 1: List of rain gauge stations and their location

3.1.2 TRMM rainfall data and Algorithm

Monthly retrievals were obtained from the Tropical Rainfall Measuring Mission (TRMM) which is a joint scientific initiative between NASA and National Space Development Agency of Japan. The retrievals were done by TRMM's 3B-43 algorithm, which is executed once per calendar month to produce the single, best estimate monthly precipitation and Root Mean Squared precipitation-error estimate field (3B-43) by combining the 3-hourly merged high-quality/IR estimates with the monthly accumulated Climate Assessment and Monitoring System (CAMS) or Global Precipitation Climatology Centre (GPCC) rain gauge analysis (3A-45). The 3-hourly merged high quality/IR estimates are summed for the calendar month, and then the rain gauge data are used to apply a large-scale bias adjustment to the 3B-42 estimates over land. The monthly gauge-adjusted merged estimate is then combined directly with the rain gauge estimates using inverse error variance weighting.

This data is available to the public through Goddard Earth Science (GES) and viewed by TRMM online visualization and analysis system, (TOVAS). The data has a resolution of $0.25^\circ \times 0.25^\circ$ grid on a global scale over latitude band 50°N-S , <http://mirador.gsfc.nasa.gov/cgi-bin>

The data used in this study was for Kenya, (34°E to 42°E and 5.5°N to -5°S).

3.2 Methodology

Several methods were used in this study and their details are outlined in the following sections.

3.2.1 Data quality control

3.2.1.1 Estimation of missing rain gauge data and satellite derived rainfall estimates

Before undertaking any study it is necessary to estimate all the missing rainfall data. One of the methods used for this purpose is linear correlation method. In this method it is assumed that some relation exists between station (or grid point) **A** and an adjacent station (or a grid point) **B**, whose data are to be used. This requires the knowledge of a neighboring station (grid point) **B** which is best correlated to the station (grid point) **A** with missing data, Basalirwa, (1979) and Nyakwada, (1991). The first step is to identify a neighboring station which has the highest correlation (r) with the station with the missing data. The correlation coefficient quantifies the degree of correlation between pairs of variables.

The neighboring station with the highest correlation coefficient (r) and reliable record is used to estimate the missing record as shown in equation (1).

$$x_{A_j} = \frac{x_{B_j} \bar{x}_A}{\bar{x}_B} \dots \dots \dots (1)$$

x_{A_j} is the missing record of station (grid point) **A** in the j^{th} month, x_{B_j} is the record for station (grid point) with reliable records **B** in the month j and \bar{x}_A and

\bar{x}_B are long term averages for stations (grid points) **A** and **B** respectively based on the period of records available at **A**.

Estimation of missing records ensured point to point comparison of satellite derived rainfall estimates with the gridded rainfall.

3.2.1.2 Homogeneity test

The data homogeneity test was done using mass curve analysis to investigate the behavior of the observed data and to determine whether there was a need for corrections to the data to account for changes in the data collection procedures or other local conditions.

In this technique, the observed values were compared with the preceding ones. Cumulative records of a station were plotted against time for each station used in this study. These plots were fitted with trend line. Any break in slope of the resulting curve would indicate inconsistency.

If the dataset was heterogeneous, a cumulative plot of the station's data against two or more neighboring homogeneous station was performed. The graph obtained (double mass curve) was used to adjust the heterogeneous records accordingly.

3.2.2 Gridding

Rainfall data sets of spatially irregular observations were interpolated to a regular grid for direct comparison with TRMM's 3B-43 algorithm's data. This was done by Kriging method.

Kriging is a stochastic technique that uses a linear combination of weights at known points to estimate the value at an unknown point. The most commonly applied form of Kriging uses a semi-variogram which is a measure of spatial correlation between pairs of points describing the variance over a distance or lag, Hartkamp et al, (1999).

The variogram model mathematically specifies the spatial variability of the data set and the resulting grid. The interpolation weights, which are applied to data points during the grid node calculations, are direct function of variogram model, Hartkamp et al, (1999). Kriging also provides a measure of error or uncertainty of the estimated surface.

In this study, Kriging was done by Surfer, a grid based contour program which has several gridding methods. Surfer uses grid features to generate contour maps or surface plot. The advantage of grid based approach is that the tasks of drawing contour lines, volumetric calculations of map modifications are much faster. With most data-sets the default gridding method, Kriging with a linear variogram is quite effective, Daly, (2006).

3.2.2.1 Derivation of Kriging equation

The derivation of Kriging equations, used in spatial rainfall estimation in this study, has been discussed by Creutin and Obbled, (1982).

For a random function, $Z(x)$, if the joint probability of Z_i ($i=1, n$) has normal (Gaussian) distribution, then its linear or conditional expectation is optimal for Kriging estimate, where

$Z(x)$ is rainfall at a spatial point.

We consider a location x_0 with an unsampled rainfall Z_0 and neighboring rainfall observations Z_i , so that its linear estimation can be expressed as

$$E^* \left[\frac{z_0}{z} \right] = E [Z_0] + \sum_{i=1}^n w_i (Z_i - E [Z_i]) \dots \dots \dots (2)$$

Where w_i is the Kriging weight and $E [Z_0]$ is the rainfall expectation at Z_0 and n is the gauge number. Under the hypothesis of local stationary of random functions Z ,

$$E [Z_0] = E [Z_i] = m$$

m is the stationary mean.

The weights w_i are determined by minimizing the unbiased Kriging variance

$$E[Z_0 - E^* [Z_0 | Z]]^2$$

This minimization results in a set of n linear equations,

$$\sum_{i=1}^n w_i c(x_i, x_j) + \mu = c(x_j, x_0), j=1, \dots, n \dots\dots\dots (3)$$

Where, $c(x_i, x_j)$ is spatial covariance at different observational locations and $c(x_j, x_0)$ is covariance between the observation and an estimated field point. μ is a Lagrange multiplier in ordinary Kriging that is linked to the unbiased condition,

$$\sum_{i=1}^n w_i = 1 \dots\dots\dots (4)$$

In equation (3), if $\mu=0$, we have simple Kriging, which is equivalent to the statistical interpolation method. The statistical interpolation method is not linked to the unbiased conditions but requires prior knowledge background field. In simple Kriging, the stationary mean, m , is regarded as background field. To estimate for the rainfall expectation, $E[Z_0]$ at, Z_0 we use the local mean of the observed gauges, m under stationary conditions. Where

$$m = \frac{1}{n} \sum_{i=1}^n R_i,$$

In Kriging, because the covariance matrix is positive definite, equation (3) will always have a solution. The Kriging variance equation can then be derived as,

$$\sigma_k^2 = C(0) + \mu - \sum_{i=1}^n w_i c(x_i, x_0) \dots\dots\dots (5)$$

This is a measure of the quality of the fit to the Kriging equation. Kriging provides optimum estimation, relative to other methods, in the sense that it minimizes the least square error for a covariance model with the unbiased condition. The weights, w_i , depend on the position of the observed and calculated points and the number of observations.

In this study, Kriging with a linear variogram method of gridding was used to interpolate rainfall data to a grid scale of 0.25° by 0.25° to match with the TRMM satellite rainfall estimates. It is effective in areas with poor distribution of data and produces reasonable presentation of the data, Sun and Ebert, (2003). The interpolated data was referred to as Kriged rain.

3.2.3 Principal component analysis

Mathematically, principal component analysis (PCA) can reduce a large number of correlated quantities in time and space into a small number of orthogonal functions that is linear combinations of the original observations and account for a large percentage of the total variance, Richman, (1986). The basic PCA model is:

$$Z=FA' \dots\dots\dots (6)$$

Where, z is the matrix of raw data ($n \times p$ matrix, n time points and p variables). F is the $n \times p$ PC score.

A- Is a $p \times p$ loading matrix whose columns a_k are the p principal component patterns that can be scaled in various ways.

The aim of PCA is to find a_k such that the variance of the first PCA series is maximized, the variance of the second series is maximized subject to orthogonality with the first series, the variance of the third series is maximized subject to orthogonality with the first two series and all subsequent modes were obtained in the same way. Using Equation (6), score matrix F is solved as follows:

$$ZA=FA' A \dots\dots\dots (7a)$$

$$F=ZA(A' A)^{-1} \dots\dots\dots (7b)$$

Where

$A' A$ - Is the diagonal matrix of the Eigen values.

The sum of Eigen values ($\sum_{k=1}^n \lambda_k$) equals to the total variance in the data, and each Eigen value λ_k gives the variance that the corresponding Eigen mode accounts for. The sequence of the function of the Eigen modes represents the importance of the mode in explaining the fraction of the total variance in the data, if the Eigen values are arranged in descending order of magnitude. The important Eigen modes (principal components) are usually retained for further analysis. The choice of truncation level is aided by use of scree graphs. When using the scree graph qualitatively, the goal is to locate a point separating a steeply sloping portion to the left and a more shallowly sloping portion to the right. The principal component number at which the separation occurs is then taken as the truncation level, Wilks, (1995). In this paper this method was therefore adapted

In this study, PCA was purposely done to reduce the size of Kriged rainfall and TRMM satellite rainfall estimates to a pair of data sets containing fewer variables, but that nevertheless represent a large fraction of variability contained in the original data, Wilks, (1995). Beyond mere data compression, however PCA was done so that the results could be used as an important tool for exploring the data sets. Here the PCA has the potential for yielding substantial insights into both the spatial and temporal variations exhibited by two fields being analyzed. PCA was also done as a useful preliminary step in canonical correlation analysis. This step is also referred to as pre-orthogonalization, Mutemi, (2003) and is useful step in pre-filtering the two fields of raw data before subjecting them to canonical correlation analysis, Wilks, (1995).

Studies by Mutua et al, (1999) and Richman, (1996) showed that that rainfall patterns in Kenya and East Africa in general is not homogeneous. Hence there was also a need to perform PCA in S-mode in order to come up with climatic zones, Fig 4

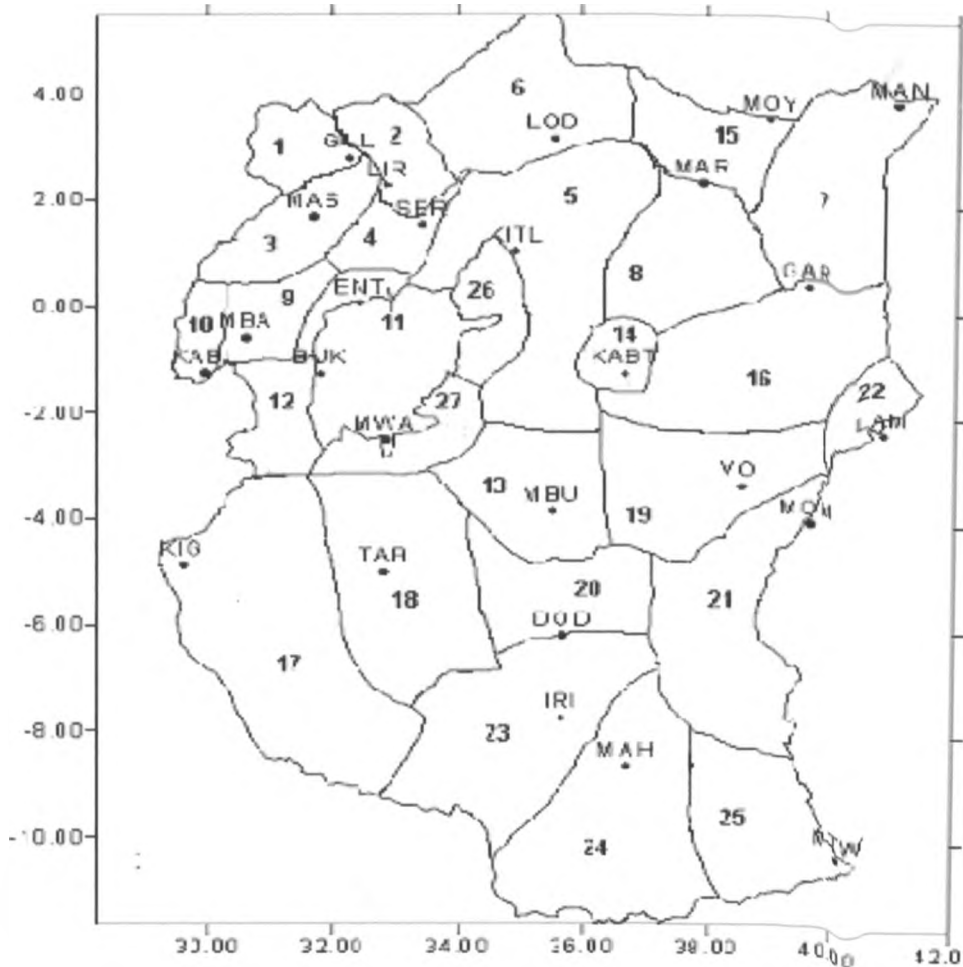


Figure 4: East African Climatic zones for MAM season, (Mutua et al, 1999)

3.2.4 Scatter plots

A scatter plot is a graphical extension of correlation matrix. A scatter plot matrix is an arrangement of individual scatter plots according to the same logic as the one which governs the placement of individual correlation coefficients in a correlation matrix. A scatter plot of gridded rain gauge data and satellite rainfall estimates was drawn against time. This plot was to show how rainfall estimates compared with the observed rain gauge data.

3.2.5 Correlation analysis

3.2.5.1 Simple Correlation analysis

Correlation analysis was done to find out whether there was a degree of relationships between satellite-derived rainfall estimates and the observed rain gauge data. A simple correlation coefficient r between satellite-derived rainfall estimates (E_i) and the corresponding rain gauge data (O_i) was calculated as follows:-

$$r = \frac{\sum_{i=1}^n (E_i - \bar{E})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^n (E_i - \bar{E})^2} \sqrt{\sum_{i=1}^n (O_i - \bar{O})^2}} \dots\dots\dots (8a)$$

The level of significance was tested using student t-test as given by equation (8b) at 95% significance levels for values of n , where n is number of observations.

$$t = r \sqrt{\frac{n-2}{1-r^2}} \dots\dots\dots (8b)$$

3.2.5.2 Canonical correlation analysis (CCA)

Canonical correlation analysis, (CCA), is a statistical technique that identifies a sequence of pairs of patterns in two multivariate data sets, and constructs sets of transformed variables by projecting the original data on to these patterns. In CCA, the patterns are chosen such that the new variables defined by projection of the two data sets on to these patterns exhibit maximum correlation, but are uncorrelated with the projections of the data on to any of the other identified patterns, Wilks, (1995). In other words, CCA identifies new variables that maximize the interrelationships between two data sets, in contrast to the patterns

describing the internal variability within a single data set identified in PCA, (section 3.2.3).

CCA can also be viewed as a linear multiple regressions applied to a multivariate (pattern) predictand. Thus, multi-component predictors are linearly related to multi-component predictand such that the sum of the squared errors is minimized, Barnson and He, (1996). A set of weights for the predictor components (elements) is related to an analogous set of weights for predictand components. These weighting sets, called loading patterns illustrate the associated predictor-predictand patterns and thus, can be used as guidance to underlying physical processes as well for descriptive purposes, Barnston and Ropelewski, (1992).

3.2.5.2.1 Computation of canonical correlation analysis

If we take variables x and y which in this study represent TRMM satellite rainfall estimates and Kriged rainfall respectively, then the information drawn upon by CCA is contained in the joint variance-covariance matrix of these variables. The two data vectors are concatenated into a single vector,

$$C^T = [x^T, y^T].$$

This vector contains $I+J$ elements, the first I of which are the elements of x and the last J of which are the elements of y . The $(I+J \times I+J)$ variance-covariance matrix of C , $[S_c]$ is then partitioned in to four blocks. Thus,

$$[S_c] = \frac{1}{n-1} [C']^T [C'] = \begin{bmatrix} [S_{xx}] & [S_{xy}] \\ [S_{yx}] & [S_{yy}] \end{bmatrix} \dots\dots\dots (9)$$

Each of the n rows of the $(n \times I + J)$ matrix $[C]$ contains one observation of the vector x' and one observation of the vector y' the primes indicate centering of the data by subtraction of each of the respective sample means.

The $(I \times I)$ matrix $[S_{xx}]$ is the variance-covariance matrix of I variable in x .

The ($J \times J$) matrix $[S_{yy}]$ is the variance-covariance matrix of J variables in y .

The matrices $[S_{xy}]$ and $[S_{yx}]$ contain the covariance between each of the elements of x and each element of y and they are related according to

$$[S_{xy}] = [S_{yx}]^T$$

The canonical correlations r_c are given by the square roots of the non-zero Eigen values of the matrices

$$[M_x] = [S_{xx}]^{-1} [S_{xy}] [S_{yy}]^{-1} [S_{yx}] \dots \dots \dots (10a)$$

And

$$[M_y] = [S_{yy}]^{-1} [S_{yx}] [S_{xx}]^{-1} [S_{xy}] \dots \dots \dots (10b)$$

The matrix $[M_x]$ is dimensioned ($I \times I$) and the matrix $[M_y]$ is dimensioned ($J \times J$). The first $M = \min(I, J)$ Eigen values of these two matrices will be identical and if $I \neq J$, the remaining Eigen values of the larger matrix will all be zero. The canonical vectors a_m and b_m are the respective Eigen vectors of these matrices, satisfying

$$[M_x] a_m = r_{cm}^2 a_m, \quad m = 1, 2, \dots, M \dots \dots \dots (11a)$$

And

$$[M_y] b_m = r_{cm}^2 b_m, \quad m = 1, 2, \dots, M \dots \dots \dots (11b)$$

Here r_{cm}^2 is the m^{th} canonical correlation.

As is the case with the use of PCA with spatial data, it is often informative to plot maps of the canonical vectors by associating the magnitude of their elements and the geographic locations to which they pertain. In this context the canonical vectors are sometimes called canonical patterns since the resulting maps show spatial patterns of the way in which the original variables contribute to the canonical variables. Examining the pairs of maps formed by corresponding vectors a_m and b_m may be informative about the nature of the relationship

between variations in the data over the two domains encoded in x and y , respectively, Wilks, (1995).

In this study, prior to canonical correlation analysis the two fields of raw data were gridded and then separately orthogonalized using principal component analysis (empirical orthogonal function). That is, separate principal components were computed for each of the two fields. The correlation based empirical orthogonal function analysis reduces large number of original variables to much fewer presumably essential variables, filtering incoherent variability, Barnson and He, (1996). After PCA, the most important pairs of Eigen vectors were retained for CCA analysis.

In this study, CCA was carried out as a tool for exploratory and diagnostic analysis of Kriged rainfall and TRMM satellite rainfall estimates.

3.2.6 Error Analysis

Other statistic methods used in this study to validate and compare rain gauge analysis (Kriged rainfall) and satellite rainfall estimates followed the guidelines of International Precipitation Working Group (IPWG), Ebert et.al, (2007). The statistics used in this study were the mean absolute percentage error (MAPE), the BIAS (mean error), mean absolute error (MAE) and the root mean square error (RMSE). After analysis, two sets of results were obtained, one from spatial (S-mode) method and the other from temporal (T-mode) method. For spatial mode method, errors were determined for overall period and as well as for, a dry season (January to February) and for a wet season (March to May). The skill scores used in this study are discussed below.

3.2.6.1 Mean absolute percentage error method (MAPE)

The mean absolute percentage error was used to measure the accuracy or error of satellite rainfall estimates / gridded rainfall. The mean absolute percentage error was calculated using the following relationship.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{O_i - E_i}{O_i} \right| \times 100 \% \dots\dots\dots (12)$$

Where O_i and E_i are Kriged rainfall and TRMM satellite rainfall estimates respectively.

3.2.6.2 Bias or mean Error

The Bias or mean error of n observations:

Is given by:-

$$BIAS = \frac{1}{n} \left(\sum_{i=1}^n (E_i - O_i) \right) \dots\dots\dots (13)$$

The mean error in this study is simply the difference between the average satellite rainfall estimates and average gridded rainfall and therefore expresses the bias of the satellite estimates. Satellite rainfall estimates that are on average too high will exhibit mean errors greater than zero and those that are on average too low will exhibit mean errors less than zero.

3.2.6.3 Mean Absolute error (MAE)

Is given by:-

$$MAE = \frac{1}{n} \left(\sum_{i=1}^n |E_i - O_i| \right) \dots\dots\dots (14)$$

In this study, Mean absolute error (MAE) is the arithmetic average of the absolute values of the difference between satellite rainfall estimates and the gridded

rainfall. It is a typical magnitude for errors between these two data sets. The MAE is zero if satellite rainfall estimates are equal to gridded rainfall and increases as the discrepancies between the two data sets become larger.

3.2.6.4 Root Mean Square Error (RMSE)

Is given by the following equation:-

$$RMSE = \frac{1}{n} \left(\sum_{i=1}^n (E_i - O_i)^2 \right)^{\frac{1}{2}} \dots\dots\dots(15)$$

In this study RMSE is the root of the average squared difference between satellite rainfall estimates and the gridded rainfall. It is more sensitive to larger errors than MAE. Squaring the errors produces positive terms so that RMSE increases from zero through large positive values as the discrepancies between satellite rainfall estimates and gridded rainfall become increasingly large.

Target mean absolute percentage errors of 10% and threshold of 15% are adapted, Fisher (2004) and Nicholson et al, (2003). The corresponding target and threshold values for Bias MAE and RMSE are shown in table 2.

Skill score	MAPE	BIAS	MAE	RMSE
Target	10%	±0.025	0.025	0.027
Threshold	15%	±0.038	0.038	0.041

Table 2: Target and threshold values for MAPE, BIAS, MAE and RMSE

3.2.7 CCA models

Another reason for using CCA was to develop models for estimating rainfall using TRMM estimates in areas where there is no rain gauge station. In the development of model, CCA identifies critical sequences of predictor patterns that tend to evolve in to sequent patterns that can be used to form a forecast (an estimate), Barnston and Ropelewski, (1992).

When CCA is used as part of predicting (estimating) procedure, it is necessary operationally to reconstruct estimated data values predicting (estimating) the y field (Kriged rainfall) using some number of the canonical vectors b_m . A synthesis equation can be derived for CCA by manipulating the matrix analysis equation, equation (16).

$$\begin{matrix} [W] & = & [Y][B] & \dots\dots\dots & (16) \\ (n \times M^*) & & (n \times J)(J \times M^*) & & \end{matrix}$$

Matrix $[Y]$ is the full matrix of predictand values. Each of the n rows of this matrix contains one data record y^T . The matrix $[B]$ has as its columns some number M^* of the canonical vectors b pertaining to y . $[W]$ is the matrix of the corresponding M^* canonical scores. Equation (16) can be solved by multiplying on the right by $[B]^T$ and then multiplying on the right again by $([B][B]^T)^{-1}$:

$$[W][B]^T = [Y][B][B]^T \dots\dots\dots (17)$$

$$[Y][B][B]^T ([B][B]^T)^{-1} = [W][B]^T ([B][B]^T)^{-1}$$

$$[Y][I] = [W][B]^T ([B][B]^T)^{-1}$$

$$[Y] = [W][B]^T ([B][B]^T)^{-1} \dots\dots\dots (18)$$

The relationship between $[v]$, $[X]$ and $[A]$ is analogous.

In this study, zero lagged data was used because our aim was to estimate areal rainfall using satellite rainfall estimates. Canonical correlation cross-data correlation matrix was constructed and the matrix was made to be symmetric, resulting in CCA Eigen vectors and their associated Eigen values for predictors and predictands. A CCA truncation mode was determined depending on the CCA Eigen value curve. The retained modes were used to develop prediction (estimation) models using equation 18.

Genstat Discovery Edition 3 application program was used in this study to compute Canonical correlations and Principal components, (CCA and PCA).

CHAPTER FOUR

4 RESULTS AND DISCUSSIONS

4.1 Quality control of the data

Monthly rainfall totals from twenty six synoptic and Agrometeorological stations were used. Both the satellite derived rainfall estimates and the rain gauge observations had no missing data. Data quality control was done by plotting mass curve and trend line for each of the twenty six stations. There was no break in slope of the resulting plots and we therefore conclude that there was no inconsistency of rainfall data for the stations used in this study. Mass curves for selected stations are presented below, (Figures 5a-c).

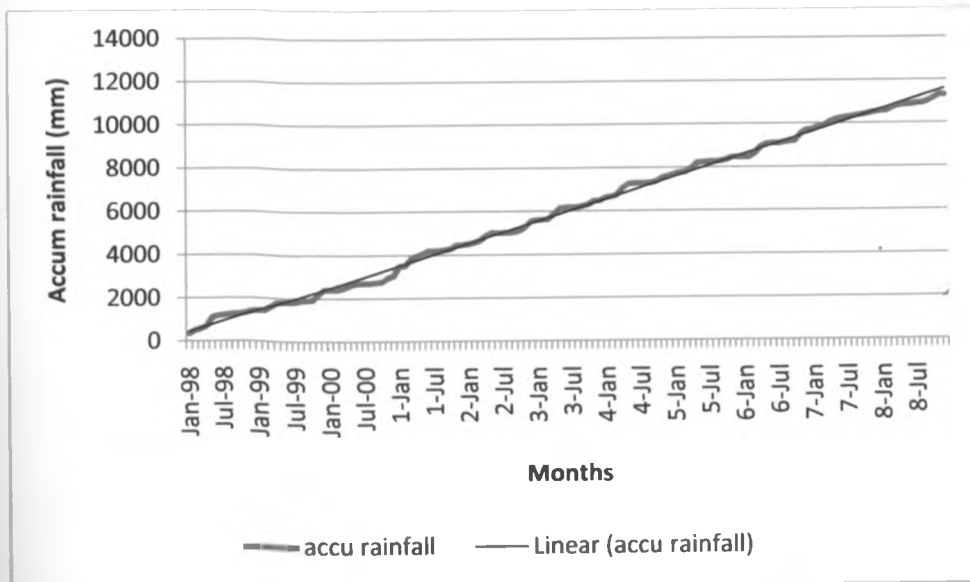


Fig 5a: Mass curve for Dagoretti station

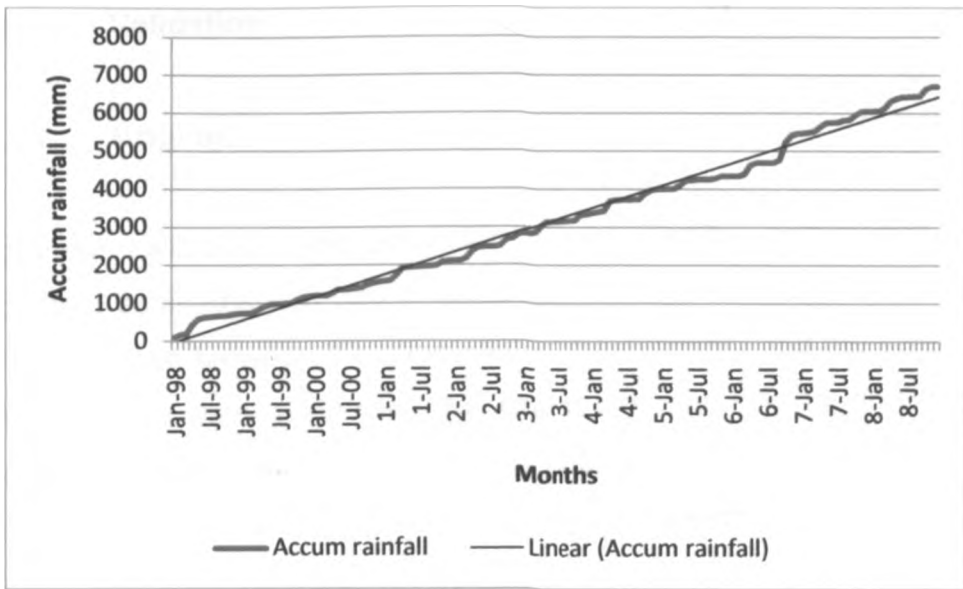


Fig 5b: Mass curve for Moyale station

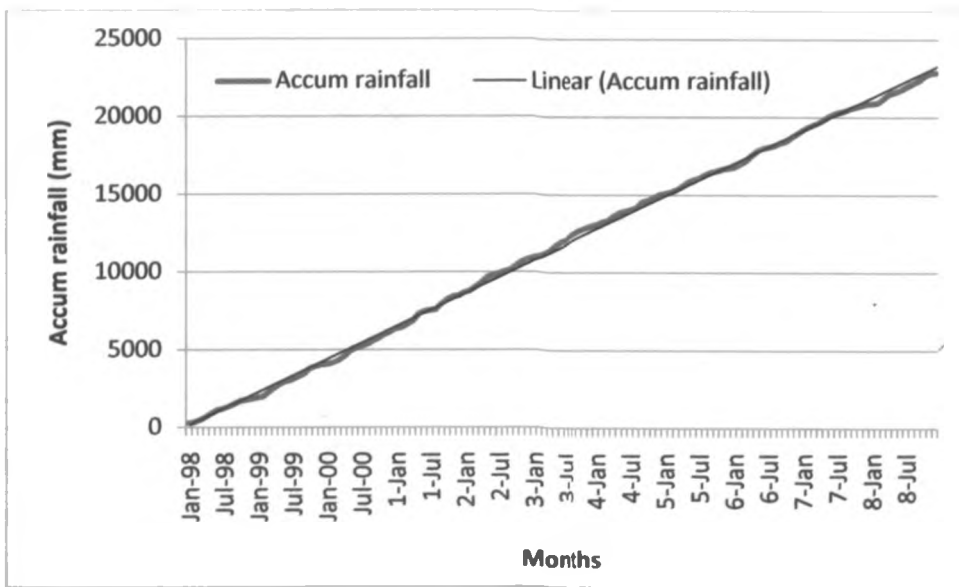


Fig 5c: Mass curve for Kisii station

4.2 Validation

4.2.1 Kriging

Algorithm calibration and validation require careful comparison between the algorithm estimates and gauge observations. One problem arises, the mismatch in spatial scale between the gauge data (observations) and the satellite estimates. This problem was addressed by using ordinary Kriging scheme with linear variogram at a spatial resolution of 0.25 degrees in order for the rain gauge data to give areal estimates corresponding to satellite pixels.

The scatter plot (Fig 6a) compared the long term monthly means Kriged rainfall and the monthly mean TRMM rainfall estimates at a resolution 0.25 by 0.25 degrees. The two variables were compactly arranged and the correlation coefficient for these fields was 0.871.

Scatter plots were also plotted for dry and wet seasons. For dry season, January to February was considered. The correlation coefficients between Kriged rainfall and TRMM rainfall estimates were 0.9, (fig 6b).

For wet a season, March to May (MAM), the long rain season was analyzed. The correlation coefficients for MAM were 0.847, (fig 6c). The correlation coefficients for dry season were higher than those of wet season. The two seasons show a symmetrical distribution of about one to one line. In all three cases the two variables were compactly arranged displaying a strong joint behaviour between the two data sets.

However, the utility of ordinary Kriging for rainfall estimation is limited by particular assumptions. First, rainfall stations used in this study were insufficient for the purpose of high resolution. Ideally, the preferred station density for high density resolution (25Km) gridding would be at least one station per 25 by 25 Km, Haylock et al, (2008). As a result errors are likely to larger than the targeted, (table 2).

Secondly, Kriging involves solving a set of linear equations to minimize the variance of the observations around the interpolating surface. This least square problem therefore assumes that station data being interpolated are homogeneous in space. This is not the case when we have stations across the country from several climatic zones. Regions with higher precipitation will therefore have higher interpolation errors.

TRMM's 3B43 algorithm also overestimates deep stratiform rain associated with wet season with a widespread anvil around strong, deep convective rain in the mature stage and the decay phase because of the effects of an anvil rain with a little or no surface rain, evaporation near the surface or tilting of the storm system.

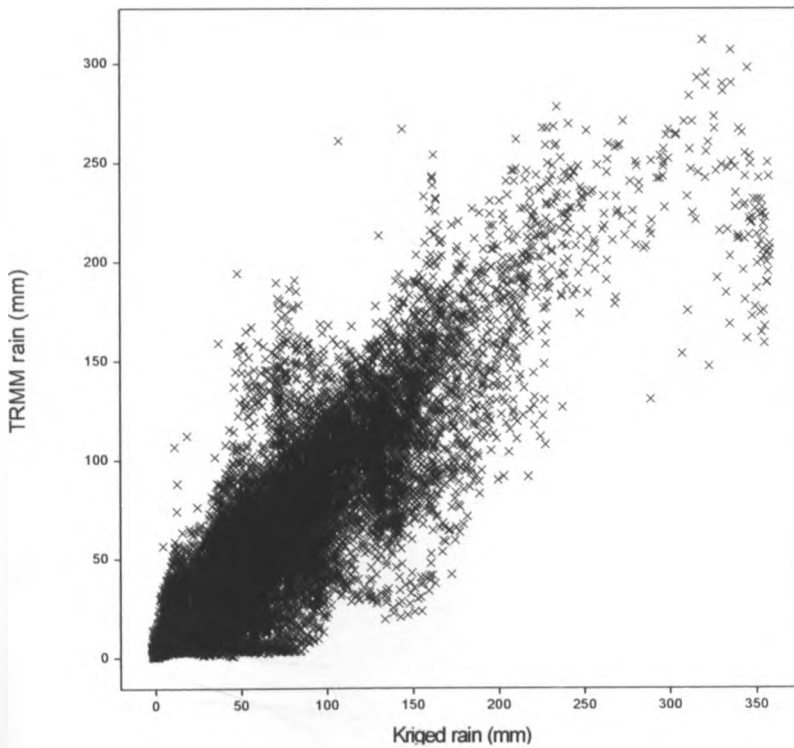


Fig 6a: scatter plot of long term monthly means kriged and satellite rainfall estimates

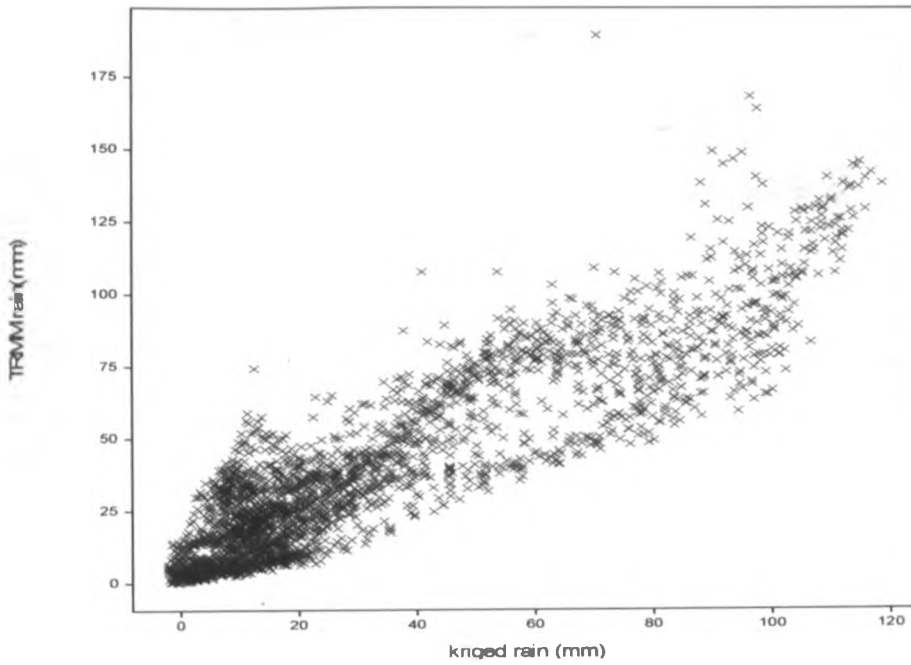


Fig 6b: scatter plot for kriged rainfall against TRMM rainfall estimates for January to February

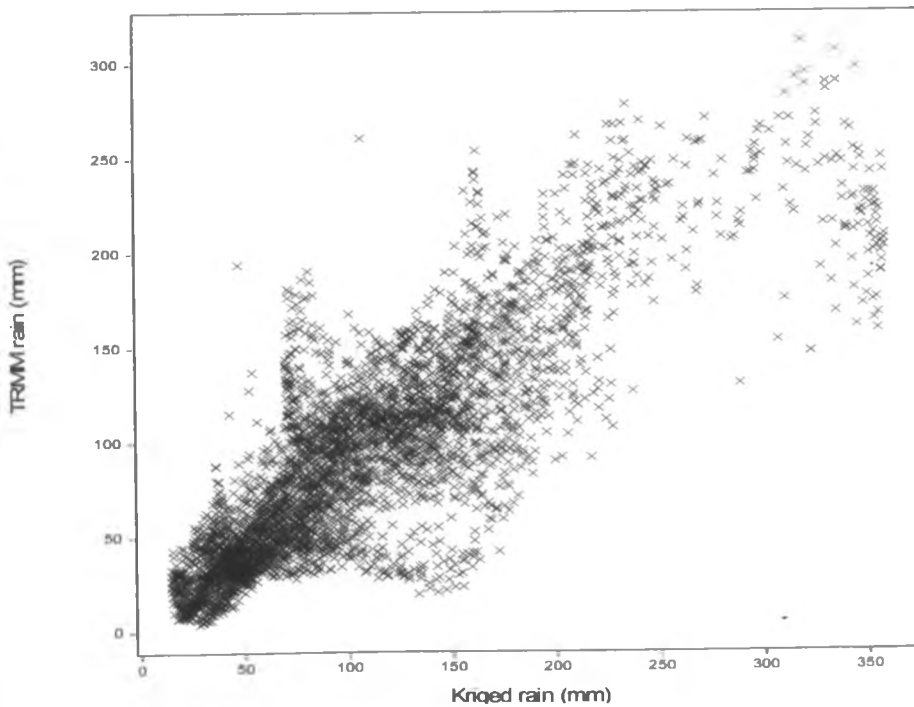


Fig 6c: Scatter plot Kriged rainfall against Trmm rainfall estimates for March to May.

4.2.2. Principal Component Analysis

In this study, PCA was also applied in both modes, (temporal and spatial) for both Kriged (gridded) rainfall and TRMM rainfall estimates to explore the pair of the data sets and to reduce the size of the data. These analysis results are discussed in sections 4.2.2.1-2.

4.2.2.1 PCA in S-mode method

4.2.2.1.1 PCA in S-mode method for the overall period

Table 3 gives the results of S-mode on Kriged rainfall and TRMM rainfall estimates. Twenty six and sixty latent roots (Eigen values) with their associated Eigen vectors were obtained for Kriged rainfall and TRMM rainfall estimates respectively.

The first ten principal components were retained for validation both for Kriged rainfall and TRMM rainfall estimates according to the Scree plots (Fig 7 and 8). Also threshold of the Eigen value was set at ten and the Principal component whose Eigen values are above this threshold was retained.

The Scree plots for both Kriged rainfall and TRMM rainfall estimates showed some similarities. The first Eigen value for Kriged rainfall was 739.89 while that of TRMM rainfall estimates was 694.28.

Eigen vector	Kriged rain- Eigen value	TRMM rain- Eigen value	Percentage of Var-Kriged rain	Percentage of Var- TRMM rain
1	739.89	694.28	54.69	51.31
2	176.84	152.18	13.07	11.25
3	123.53	108.65	9.13	8.03
4	116.56	91.39	8.61	6.75
5	51.69	39.23	3.82	2.90
6	30.04	26.95	2.22	1.99
7	25.13	23.96	1.86	1.77
8	20.11	17.12	1.49	1.27
9	15.01	15.26	1.11	1.13
10	10.34	13.80	0.76	1.02
11	9.54	9.90	0.70	0.73
12	6.52	9.47	0.48	0.70
13	5.02	8.01	0.37	0.59
14	4.13	7.16	0.30	0.53
15	3.49	6.68	0.26	0.49
16	3.05	6.67	0.23	0.49
17	2.61	6.51	0.19	0.48
18	2.51	6.21	0.19	0.46
19	1.90	5.48	0.14	0.41
20	1.40	4.97	0.10	0.37
21	1.00	4.67	0.07	0.35
22	0.79	4.48	0.06	0.33
23	0.63	3.98	0.05	0.29
24	0.59	3.84	0.04	0.28
25	0.37	3.41	0.03	0.25
26	0.31	3.35	0.02	0.25
27	0	3.16	0.0	0.23
28	0	2.86	0.0	0.21
29	0	2.77	0.0	0.20
30	0	2.69	0.0	0.20

Table 3: Eigen vectors, Eigen values and the percentage of the variance for Kriged rainfall when subjected to PCA S-mode method

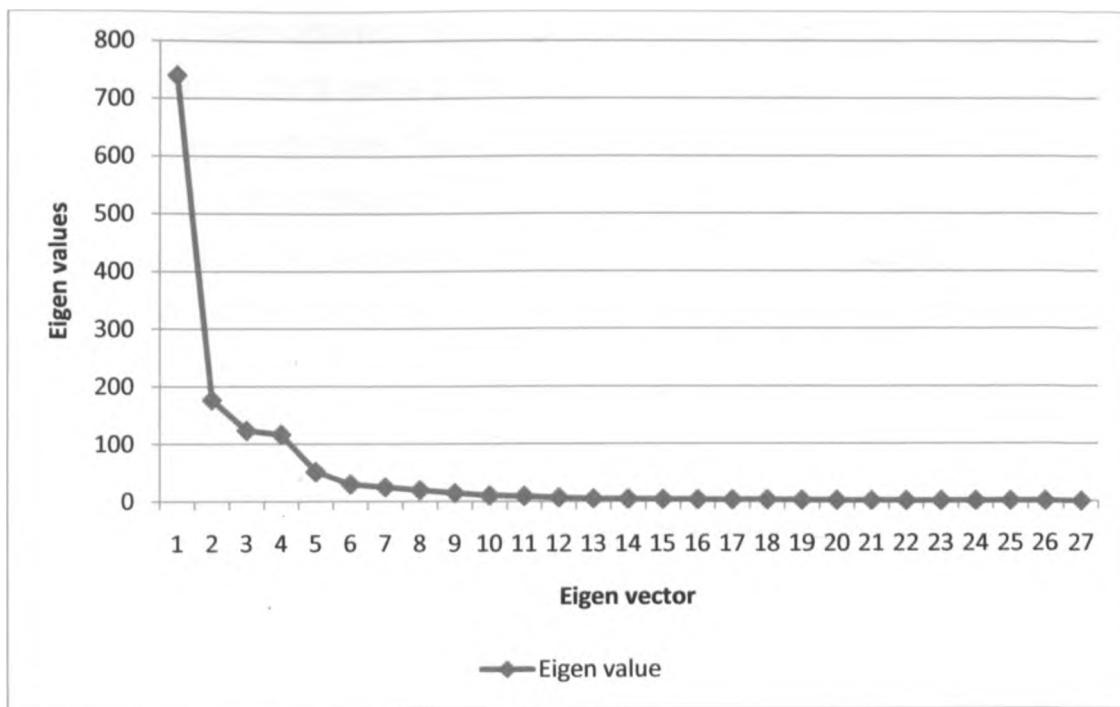


Fig 7: Scree plot for latent roots of the Kriged rainfall (PCA S-mode method)

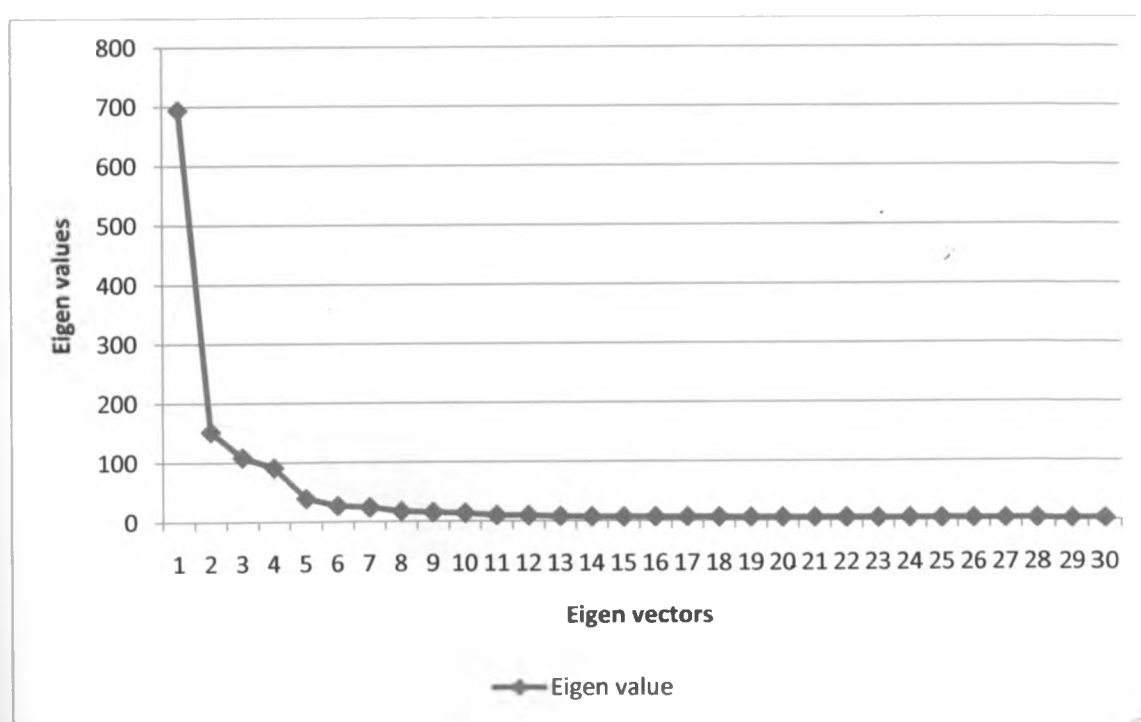


Fig 8: Scree plot for latent roots of the TRMM rainfall (PCA S-mode method)

Four regions were obtained from the plot of the first principal component for the Kriged rainfall for thirty years period (Fig. 9). The first Eigen vector for was plotted because it explained more than 50% of the total variance. These regions are the Western sector extending up to the Northwestern Kenya, a small region over the Southern part of the country near Kilimanjaro area, the Coastal strip and the Northern Kenya extending to Central and some parts of Southern Kenya. This plot compared well with that of the first principal component of TRMM rainfall estimates for most parts of the country.

Fig. 10 shows the three zones obtained from the corresponding plot for TRMM rainfall estimates. The first zone (zone 1) was Western sector of the country. The second zone, (zone 2) was the Kenyan Coast. The third zone, (zone 3) was composed of Northern, North Eastern, Central and Southern parts of Kenya.

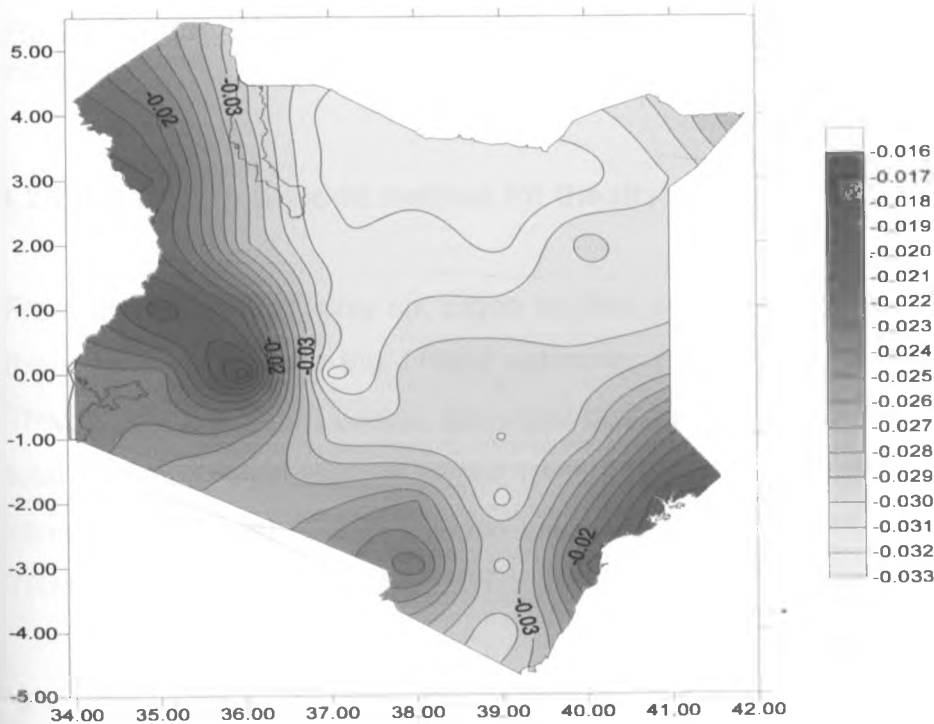


Fig 9: Pattern of the first Eigen vector (first principal component) of Kriged rainfall for the overall period (y-axis is longitude and x-axis is latitude)

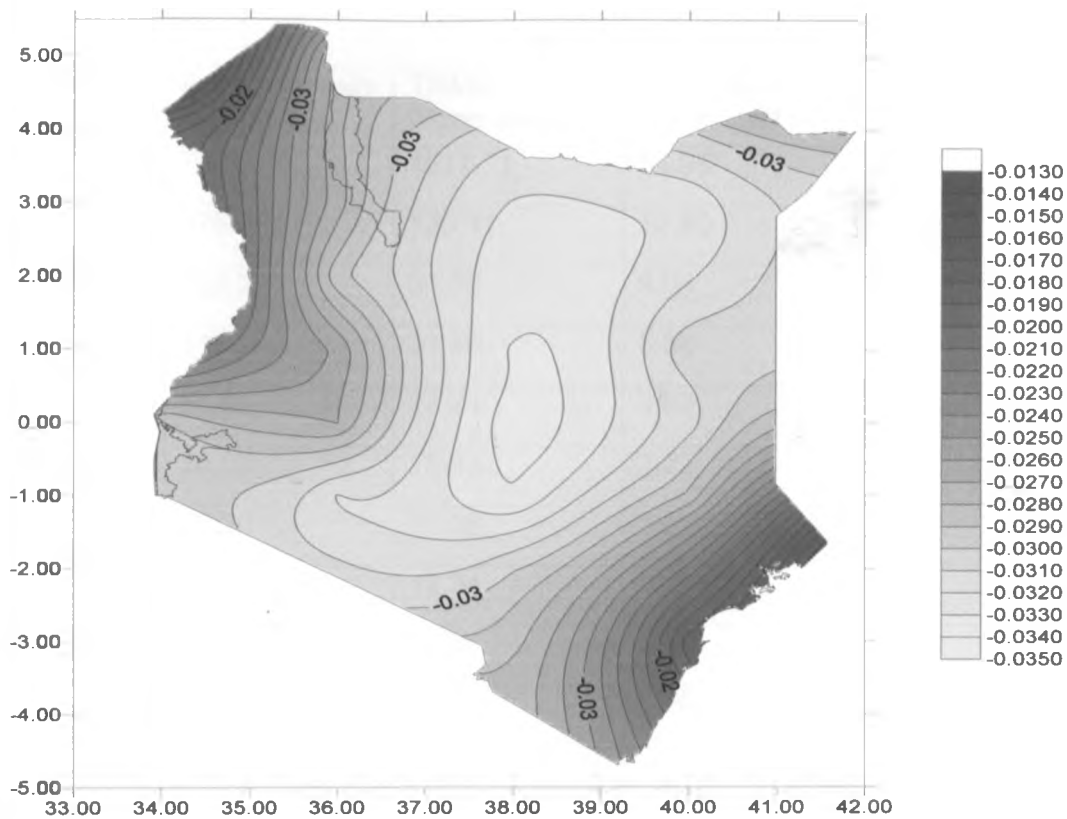


Fig 10: Pattern of the first Eigen vector (principal component) of TRMM rainfall estimates for the overall period (y-axis is longitude and x-axis is latitude)

4.2.2.1.2 PCA in S-mode method for the dry season, (January to February)

From the Scree plot, only six Eigen vectors were retained for validation for both the Kriged rainfall and the TRMM estimates for January and February data. For TRMM, the first Eigen vector, (principal component) explained 75 per cent of the total variance while that of Kriged rainfall explained 54.99 per cent of the total variance, (Table 4). The first Eigen value for Kriged rainfall was 744 while that of TRMM rainfall estimates was 1017.

Eigen vector	Kriged rain- Eigen value	TRMM rain- Eigen value	Percentage of Var-Kriged rain	Percentage of Var-TRMM rain
1	744	1017.11	54.99	75.17
2	175.38	139.11	12.96	10.28
3	122.29	51.39	9.04	3.80
4	115.76	31.40	8.56	2.32
5	51.83	22.24	3.83	1.64
6	30.06	18.83	2.22	1.39
7	25.20	13.08	1.86	0.97
8	20.14	12.34	1.49	0.91
9	14.90	8.78	1.10	0.65
10	10.09	7.01	0.75	0.52

Table 4: Eigen vectors, Eigen values and the percentage of the variance for Kriged rain and TRMM rainfall for January to February when subjected to PCA S-mode method (y-axis is longitude and x-axis is latitude).

Fig. 11 shows four regions namely, Western Kenya, North Western, Central extending to Southern, Coastal and Northern Kenya and finally a region over North Eastern Kenya-, derived from the first PC for the TRMM estimates. The corresponding plot for the Kriged rainfall over thirty year period resulted to four regions namely Western and Northwestern Kenya, Northern Kenya extending to Central and Southern Kenya, the Coastal strip and Southern Kenya, (Fig 12). The results of Kriged rainfall over thirty year period and that of TRMM had similar pattern over Northwestern, Western, Central and some parts of the Northern sector of the country for the dry period.

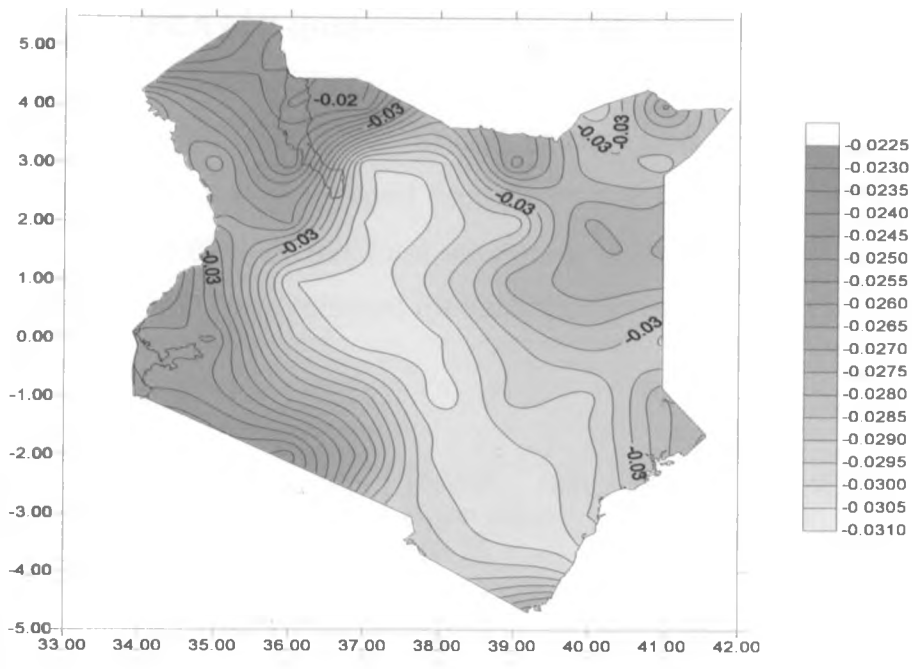


Fig 11: Pattern of the first Eigen vector of TRMM rainfall for January to February (y-axis is longitude and x-axis is latitude).

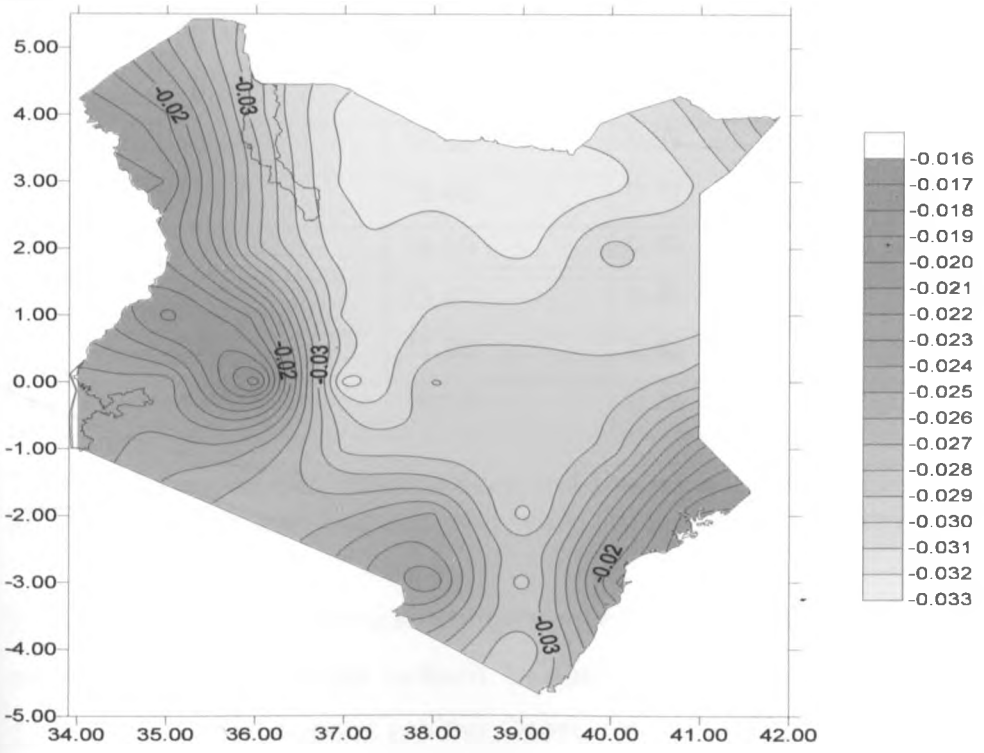


Fig 12: Pattern of the first Eigen vector of Kriged rainfall for January to February (y-axis is longitude and x-axis is latitude).

4.2.2.1.3 PCA in S-mode method for a Wet season (March to May)

The first Eigen value for Kriged rainfall was 654.54 and the total variance explained by the associated Eigen vector was 48.67%, while that of TRMM was 553.25, (40.89% of the total variance). With the aid of scree plot, the first ten Eigen vectors were retained for validation in each case, (Table 5).

Eigen vector	Kriged rain-Eigen value	TRMM rain-Eigen value	Percentage of Var-Kriged rain	Percentage of Var-TRMM rain
1	658.54	553.25	48.67	40.89
2	240.30	208.91	17.76	15.44
3	145.38	156.27	10.74	11.55
4	92.27	103.33	6.82	7.64
5	43.28	42.15	3.20	3.12
6	35.86	34.22	2.65	2.53
7	33.21	33.27	2.45	2.46
8	20.31	22.83	1.50	1.69
9	16.06	22.37	1.19	1.65
10	13.10	17.83	0.97	1.32
11	10.47	16.46	0.77	1.22
12	8.15	14.39	0.60	1.06
13	6.18	13.47	0.46	1.00
14	5.86	11.98	0.43	0.89
15	4.78	10.13	0.35	0.75

Table 5: Eigen vectors, Eigen values and the percentage of the variance for Kriged and TRMM rainfall for March to May when subjected to PCA S-mode method

Map plot for the first principal component of both Kriged rainfall and TRMM rainfall estimates showed similar pattern. In each case three regions were obtained, (fig 13 and 14). These regions are the Coast, North Western Kenya and the rest of the country.

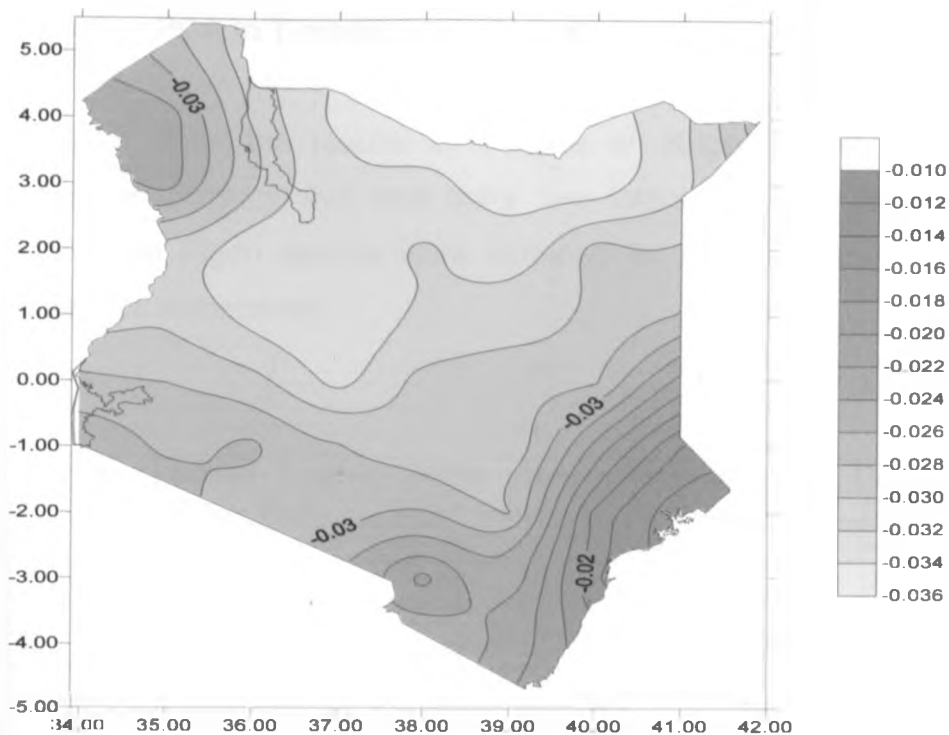


Fig 13: Pattern of the first Eigen vector of Kriged rainfall for March to May (y-axis is longitude and x-axis is latitude).

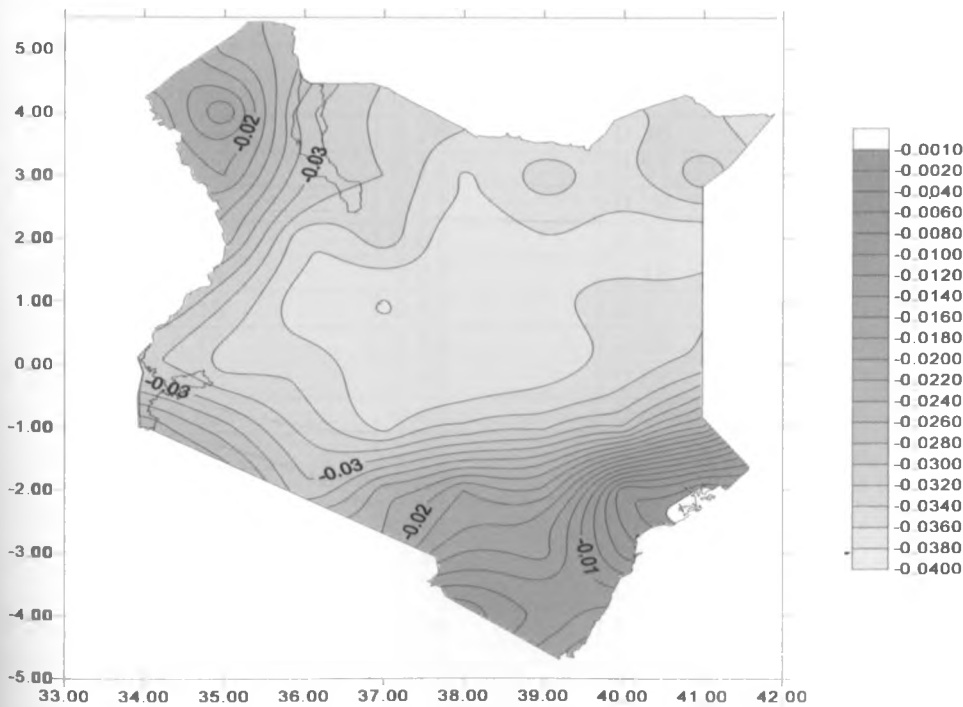


Fig 14: Pattern of the first Eigen vector of TRMM rainfall estimates for March to May (y-axis is longitude and x-axis is latitude).

4.2.2.2 PCA in T-mode method

Table 6 gives the results of T-mode on Kriged rainfall and TRMM rainfall estimates. Twenty five and thirty two latent roots (Eigen values) with their associated Eigen vectors were obtained for Kriged rainfall and TRMM rainfall estimates respectively.

Eigen vector	Kriged rain- Eigen value	TRMM rain- Eigen value	Percentage of Var-Kriged rain	Percentage of TRMM rain	Var-
1	58.64	45.63	44.42	34.57	
2	28.23	24.68	21.39	18.70	
3	12.80	12.22	9.69	9.26	
4	12.00	10.57	9.09	8.01	
5	5.18	3.95	3.92	2.99	
6	3.47	3.65	2.63	2.76	
7	2.47	3.51	1.87	2.66	
8	1.96	1.95	1.49	1.48	
9	1.19	1.74	0.90	1.32	
10	0.99	1.37	0.75	1.04	
11	0.89	1.19	0.67	0.90	
12	0.71	1.00	0.53	0.75	
13	0.58	0.95	0.44	0.72	
14	0.54	0.87	0.41	0.66	
15	0.48	0.76	0.36	0.57	
16	0.41	0.74	0.31	0.56	
17	0.37	0.71	0.28	0.54	
18	0.26	0.67	0.20	0.51	
19	0.24	0.61	0.18	0.46	
20	0.15	0.57	0.11	0.43	
21	0.12	0.53	0.09	0.40	
22	0.10	0.48	0.08	0.36	
23	0.09	0.46	0.07	0.35	
24	0.08	0.45	0.06	0.34	
25	0.05	0.43	0.04	0.33	
26	0.00	0.41	0.00	0.31	
27	0.00	0.37	0.00	0.28	
28	0.00	0.36	0.00	0.27	
29	0.00	0.34	0.00	0.26	
30	0.00	0.33	0.00	0.25	

Table 6: Eigen vectors, Eigen values and the percentage of the variance for Kriged and TRMM rainfall in PCA T-mode

The first ten principal components were retained for validation both for Kriged rainfall and TRMM rainfall estimates according to the Scree plots (Fig 15 and 16). The two plots suggested that Kriged rainfall and TRMM rainfall estimates had some common characteristics.

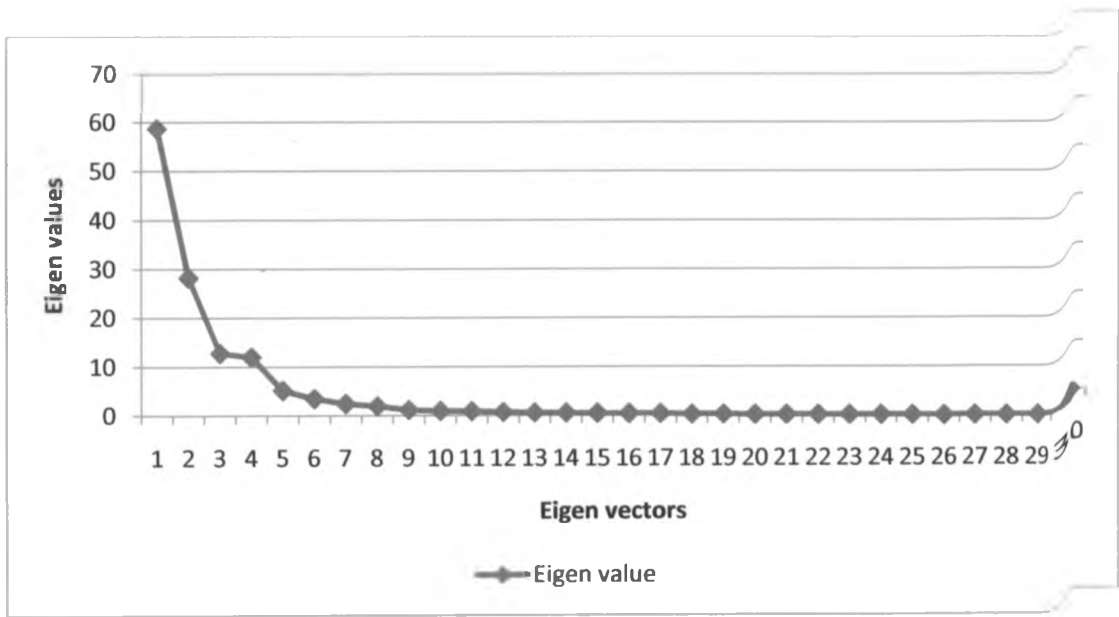


Fig 15: plot for latent roots of the Kriged rainfall (PCA T-mode)

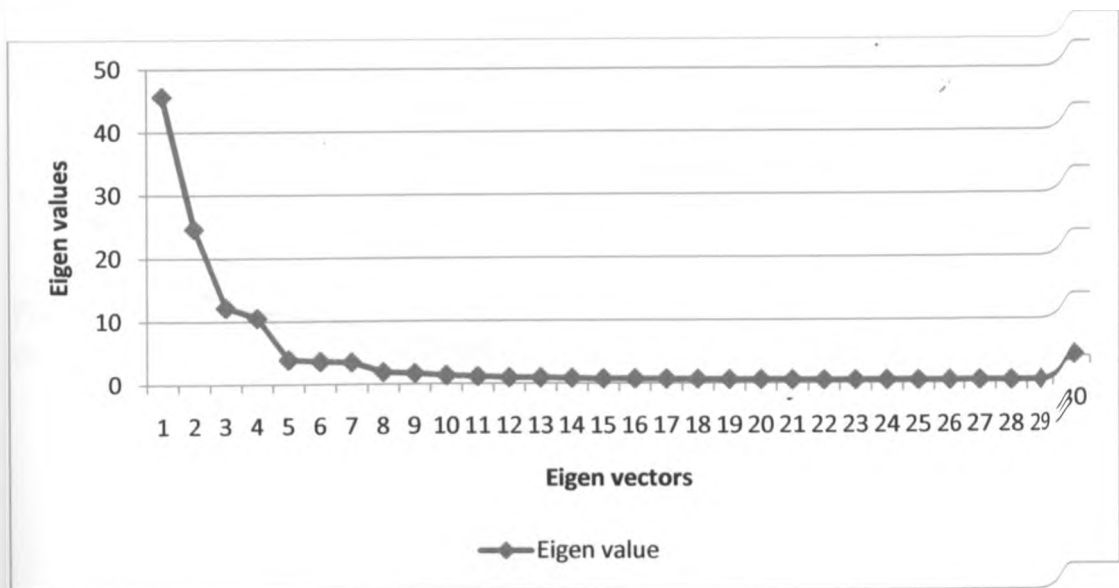


Fig 16: Scree plot for latent roots of the TRMM rainfall (PCA T-mode)

4.2.3 Time series analysis

Figures 17 and 18 give the results of time series analysis for the first and second Eigen of the Kriged rainfall and TRMM rainfall estimates. Figure 17 indicates that Kriged rainfall's loadings are larger than those of TRMM rainfall estimates. However, in both cases loadings are negative. Figure 18 indicates that extremes of TRMM rainfall estimates are either larger or smaller than those of Kriged rainfall. The two variables are either underestimating or overestimating one another. However, from these results, it is possible to estimate areal rainfall using satellite rainfall estimates.

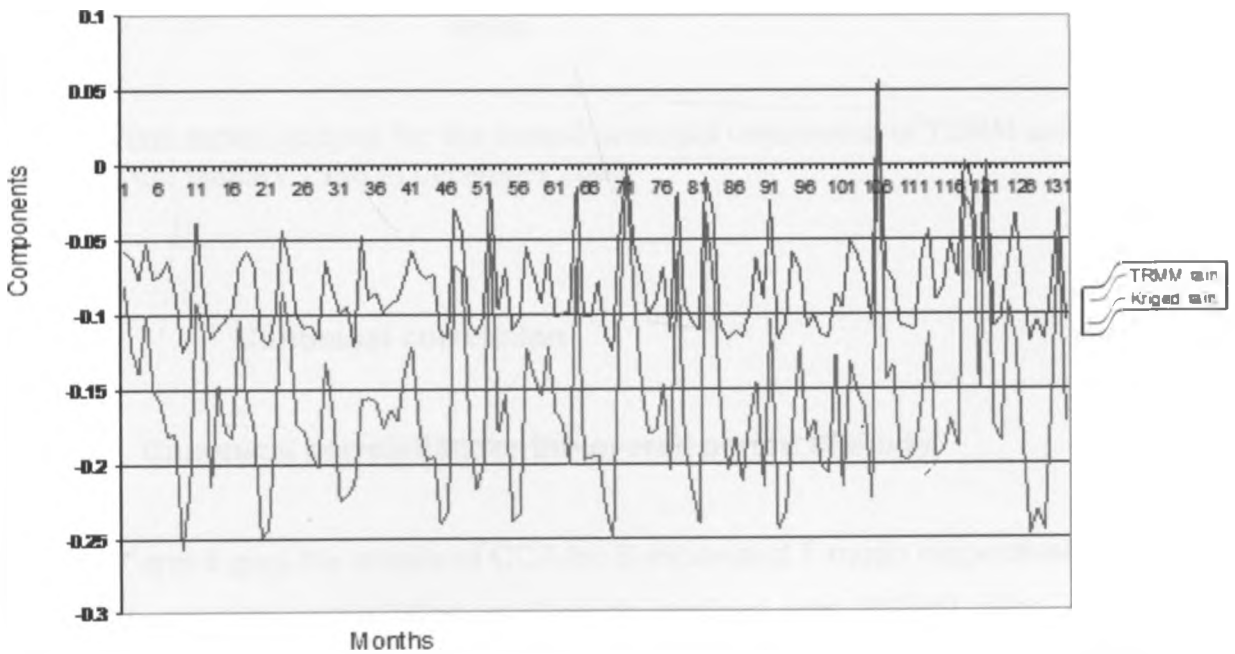


Fig 17: Time series analysis for the first principal component of TRMM and Kriged rainfall from January 1998 to December 2008

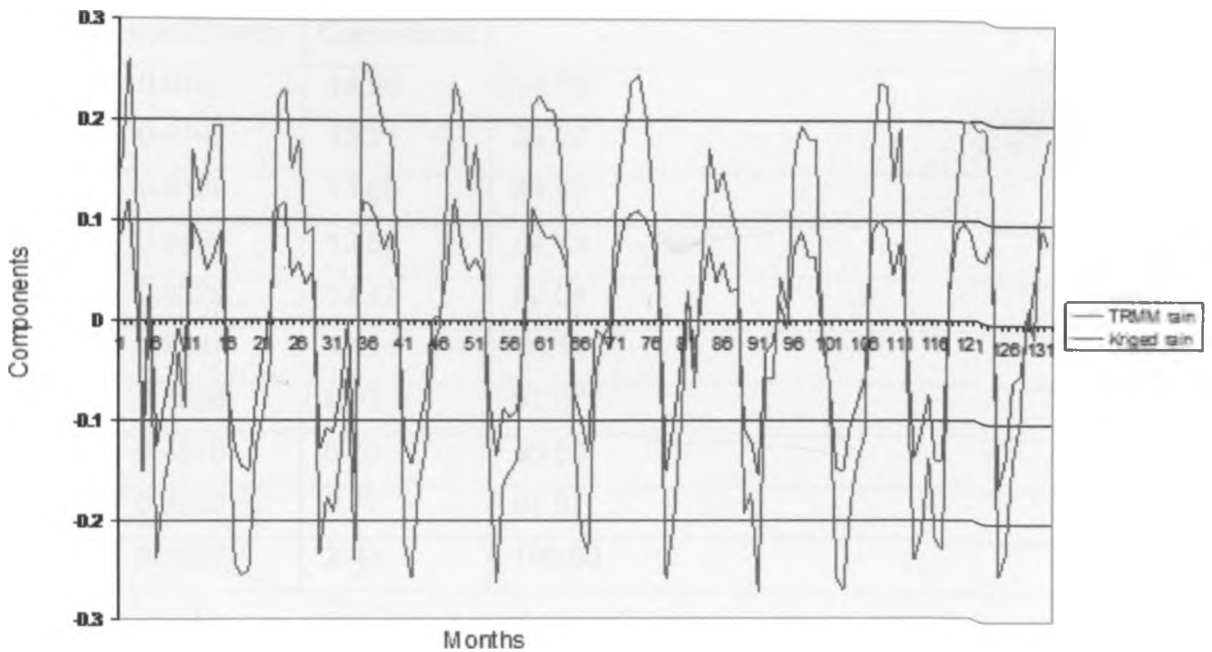


Fig 18: Time series analysis for the second principal component of TRMM and Kriged rainfall from January 1998 to December 2008

4.2.4 Canonical correlation

4.2.4.1 Canonical correlation for the overall period of study

Tables 7 and 8 give the results of CCA for S-mode and T-mode respectively. The first canonical correlation in S-mode method was 0.9767. For T-mode, correlation between Kriged rainfall and TRMM was 0.9859. These were the largest possible correlation between pairs of the linear combinations of these datasets. However, they represent only 14.56 and 12.60 per cent of the total correlations respectively. For S-mode method, seven out of ten Eigen vectors had canonical correlation coefficients of more than 0.5, which is 87.07% of the cumulative correlation coefficients. T-mode method had eight out of ten Eigen vectors with canonical correlation coefficients of more than 0.5.

Eigen vector	Canonical correlations coefficients	% of the total Correlations	Cumulative % Correlations
1	0.9767	14.56	14.56
2	0.9197	13.71	28.27
3	0.9156	13.65	41.92
4	0.8430	12.57	54.49
5	0.8329	12.42	66.91
6	0.7537	11.24	78.14
7	0.5988	8.93	87.07
8	0.4516	6.73	93.80
9	0.2530	3.77	97.57
10	0.1627	2.43	100.00

Table 7: Canonical correlation for S-mode for the overall period of study

Eigen vector	Canonical correlation coefficients	% of the total Correlations	Cumulative % Correlations
1	0.9859	12.60	12.60
2	0.9770	12.49	25.08
3	0.9677	12.37	37.45
4	0.9334	11.93	49.38
5	0.9214	11.77	61.15
6	0.8730	11.16	72.31
7	0.8094	10.34	82.65
8	0.7937	10.14	92.79
9	0.4356	5.57	98.36
10	0.1283	1.64	100.00

Table 8: Canonical correlation for T-mode for the overall period of study

4.2.4.2 Canonical correlation for seasons

Tables 9 and 10 give CCA results for Kriged rainfall and TRMM satellite for a dry season (January to February) and a wet season (March to May) respectively.

For dry season the highest correlation coefficient was 0.926, however this was only 24.9 percent of the total correlations, (table 9). In this season four out of six Eigen vectors had correlation coefficients greater than 0.5 which was 79.73% of the total correlation.

For March to May season seven out of ten Eigen vectors had canonical correlation coefficients greater than 0.5. This was 87.72% of the total correlations. The highest correlation coefficient obtained in this season was 0.967, (table 10).

Eigen vector	Canonical correlation coefficients	% of the total Correlations	Cumulative Correlations %
1	0.926	24.90	24.90
2	0.807	21.69	46.58
3	0.666	17.91	64.49
4	0.567	15.24	79.73
5	0.415	11.16	90.90
6	0.339	9.10	100

Table 9: CCA results for a dry season (January to February)

Eigen vector	Canonical correlation coefficients	% of the total Correlations	Cumulative %Correlations
1	0.967	14.66	14.66
2	0.951	14.42	29.08
3	0.944	14.32	43.40
4	0.865	13.11	56.51
5	0.796	12.07	68.58
6	0.720	10.91	79.49
7	0.543	8.23	87.72
8	0.479	7.26	94.98
9	0.261	3.95	98.94
10	0.070	1.06	100

Table 10: CCA results for a wet season (March to May)

The results of the canonical correlations for the overall period of study and the two seasons indicated that the TRMM satellite rainfall estimates and the Kriged rainfall were highly correlated. It was concluded that TRMM satellite rainfall estimate could be used to estimate areal rainfall. The results of canonical correlation analysis were therefore used to develop models for estimating areal rainfall.

4.2.5 Error Analysis

The error analysis results are presented below.

4.2.5.1 Mean absolute percentage error method (MAPE)

The mean absolute percentage error analysis was performed to assess the accuracy of TRMM satellite rainfall estimates. The (MAPE) error gives an indication of the accuracy of the satellite based estimates with respect to ground

based observations. The percentage error was determined using spatial and temporal methods.

4.2.5.1.1 Mean absolute Percentage error (MAPE) in temporal mode

Figure 19 represents the mean absolute percentage error analysis in temporal mode. These errors ranged from 0.27% to 212.1%. The average percentage absolute error in this mode was 44.4%. The target and the threshold values were 10% and 15% respectively. This indicated that the difference between the satellite-derived rainfall estimates and the interpolated rainfall was quite large in this mode.

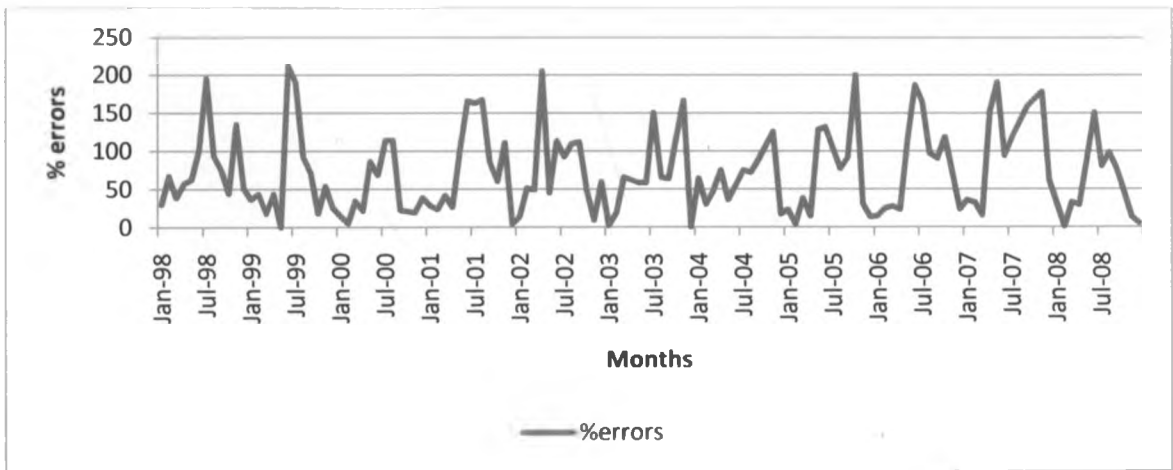


Fig 19: The temporal representation of the mean absolute percentage error (MAPE)

4.2.5.1.2 Mean absolute Percentage error (MAPE) in spatial mode

Mean absolute percentage error analysis in spatial mode method was performed to determine areas where satellite rainfall estimates agree with the Kriged rainfall. This analysis was performed for the overall period of study and also for a dry and a wet season.

(a) Mean absolute percentage error for the overall period of study

Figure 20 represents the spatial representation of the mean absolute percentage error for the overall period of study. These errors ranged from 0.1% to 99%. However, the average percentage error was found to be 28% against a target of 10% and threshold of 15%. Central and Northwestern Kenya recorded the largest errors. Over these areas the TRMM satellite rainfall estimates and the Kriged rainfall are not in good agreement. The Western and Eastern sectors of the country recorded smallest percentage errors. In these areas the differences between the satellite-derived rainfall estimates and the Kriged rainfall are relatively small.

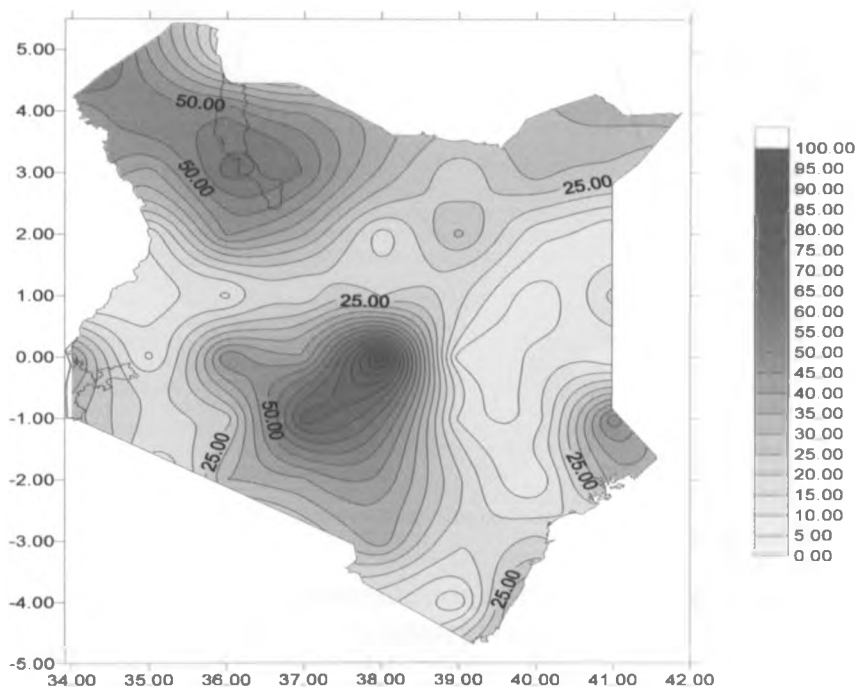


Fig 20: The overall spatial representation of the mean absolute percentage error (MAPE) (y-axis is longitude and x-axis is latitude).

(b) Mean absolute percentage error for a dry season (January to February)

Figure 21 shows the spatial representation of mean absolute percentage error for a dry season (January to February). The range of absolute percentage errors was between 0% and 86.4% with an average of 14.9%, which is less than the

threshold value of 15%. Generally the plot show that apart from the Western, the Coast and some few parts of Northwestern Kenya, most parts of the country have mean absolute percentage errors less than 20%. This indicates that the differences between the satellite-derived rainfall estimates and the interpolated rainfall are relatively small, hence this verifies that satellite estimations may be used to estimate areal rainfall where ground based observations are non existence.

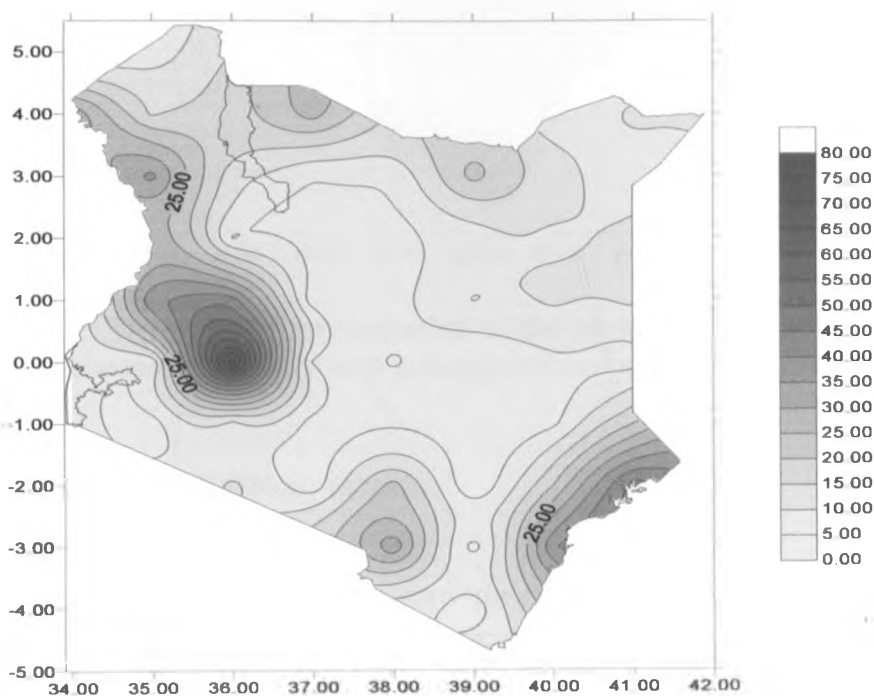


Fig 21: The spatial representation of the mean absolute percentage error (MAPE) for a dry season (January to February) (y-axis is longitude and x-axis is latitude).

(c) Mean absolute percentage error for a wet season (March to May)

Figure 22 gives the mean absolute percentage errors for March to May. In this season the mean absolute percentage error ranged from 0% to 99.9% with an average of 27.9%. Generally most parts of the country recorded errors which were less than 20%. These were within the acceptable limit. However, Coastal strip reported errors larger than 20%.

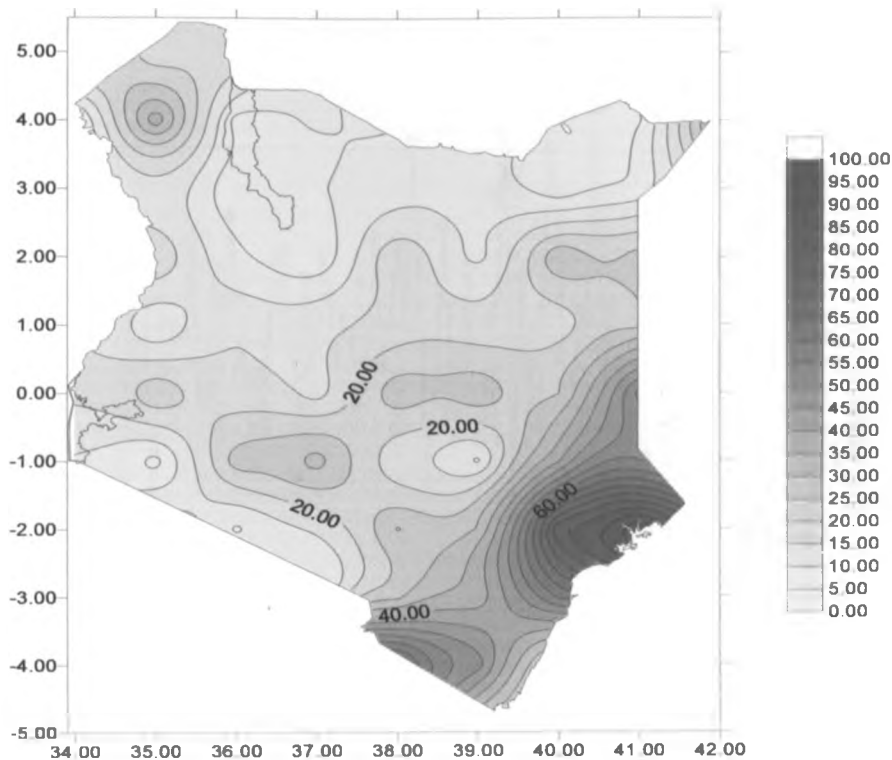


Fig 22: The spatial representation of the mean absolute percentage error (MAPE) for a wet season (MAM) (y-axis is longitude and x-axis is latitude).

4.2.5.2 BIAS (Mean error)

The mean errors (BIAS) were determined using two methods, the spatial (S-mode) and the temporal (T-mode). In spatial method, BIASES were determined for the overall period of study, a dry season (January to February) as well as for a wet season (March to May).

4.2.5.2.1 BIAS (mean error) in temporal mode

Figure 23 gives the overall BIAS results in temporal mode. These mean errors ranged from -0.093 to +0.09. The lowest bias value was recorded in January 1998 while the highest bias was recorded in April 2007 respectively. These maximum and minimum BIASES were treated as outliers and it was found that in most cases the BIASES were within the acceptable limit of ± 0.038 .

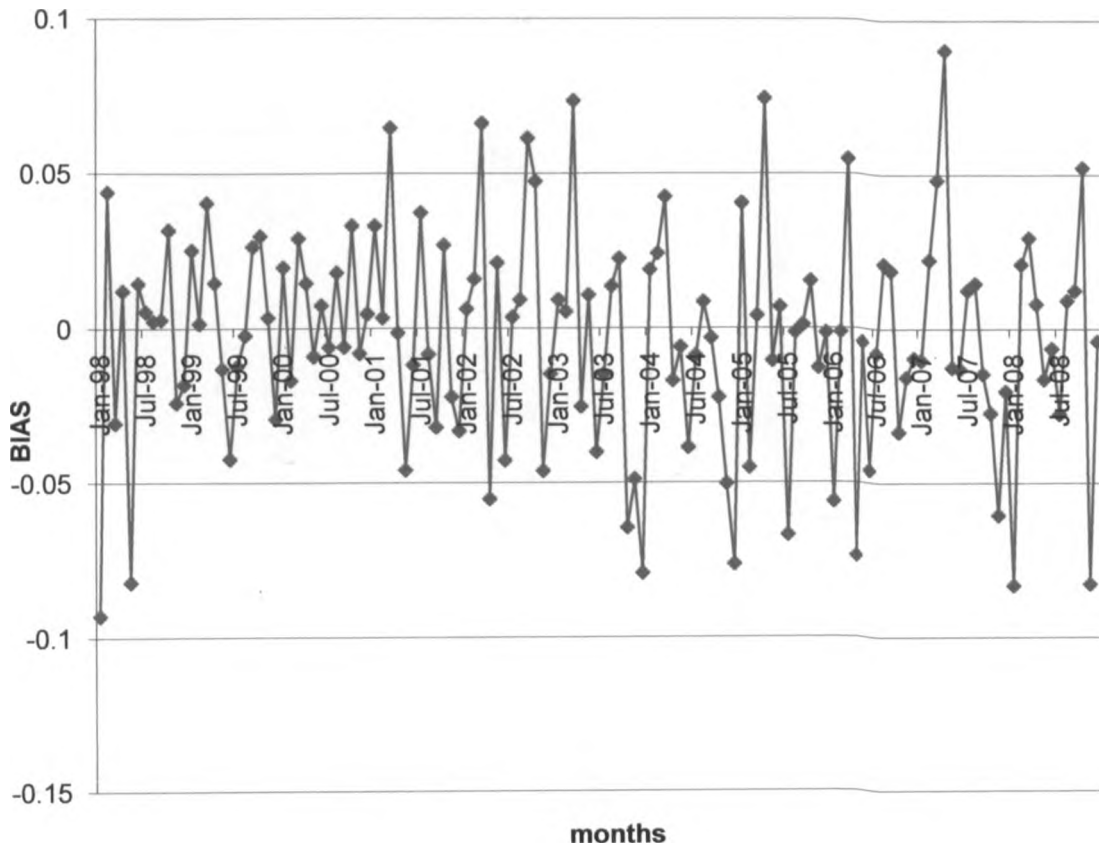


Fig 23: The temporal representation of the mean error (BIAS)

4.2.5.2.2 BIAS (mean error) in spatial mode

(a) Bias for the overall period of study

Figure 24 gives results of the Bias in S-mode method for the overall period of study. These errors ranged from -0.026 to +0.03, which was within the acceptable limit of ± 0.038 . When these Bias values were plot on a map, results showed that the Northern, Central and the Kenyan coast had the highest values of the BIAS. These positive bias values indicate that in these areas satellite rainfall estimates exceed the observed values on average. Western and Eastern Kenya had negative BIAS. In these areas satellite underestimate rainfall.

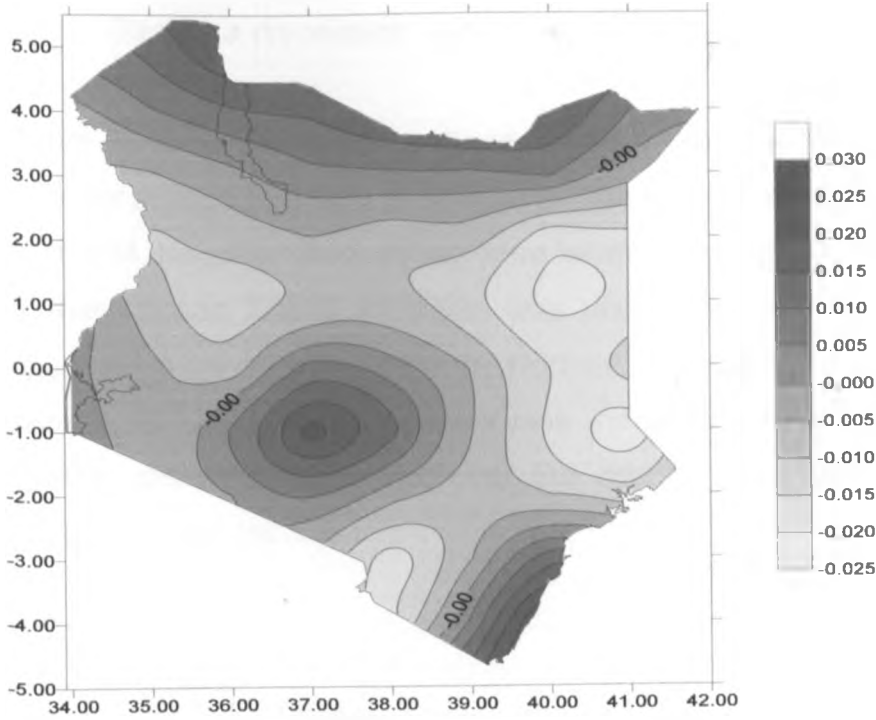


Fig 24: The overall spatial representation of the Mean error (BIAS) (y-axis is longitude and x-axis is latitude).

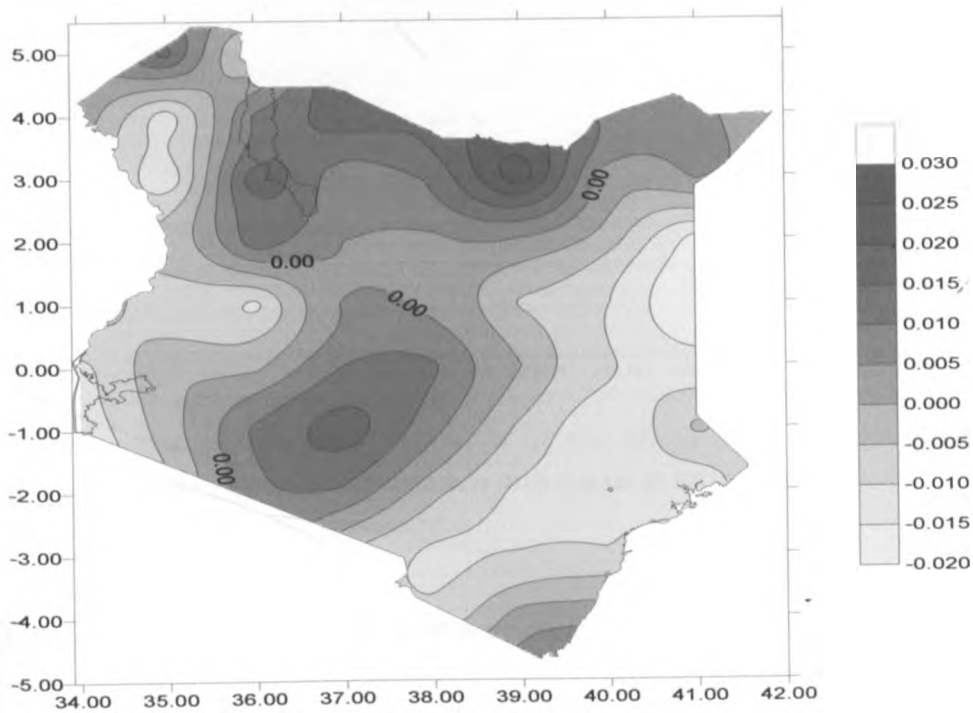


Fig 25: The spatial representation of the Mean error (BIAS) for a dry season (January to February), (y-axis is longitude and x-axis is latitude).

(b) Bias for a dry season (January to February)

Figure 25 gives the BIAS results of Kriged rainfall and TRMM rainfall estimates for a dry season (January to February). The range of the BIAS were from -0.038 to +0.034. Negative BIAS values were located over Western and Eastern Kenya. In these areas TRMM algorithm was underestimating rainfall. Positive BIAS values were concentrated over the Northern, Central and southern sectors of the country. In these areas positive bias indicates that the satellites estimates exceed observed values, all over the country the BIASES were within the acceptable limit (± 0.038).

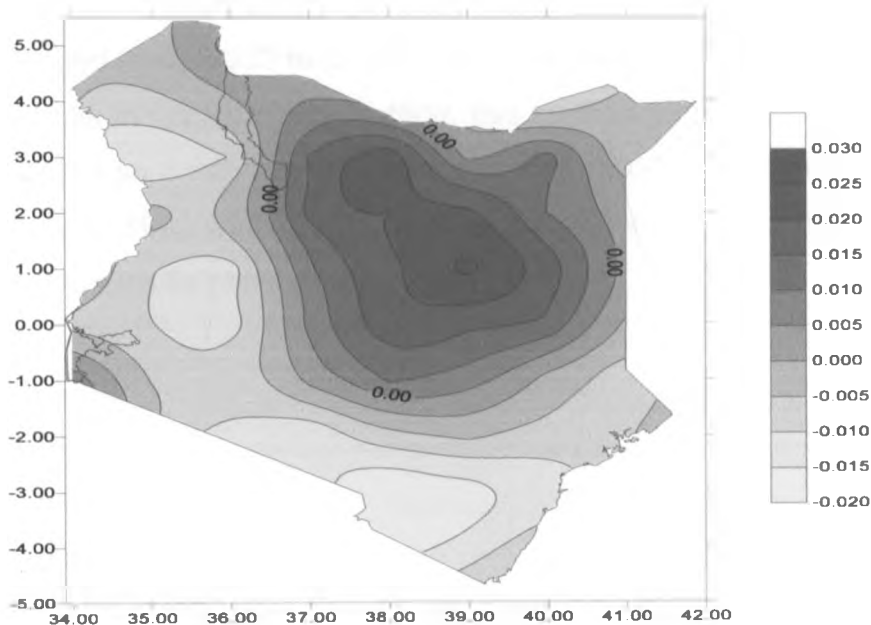


Fig 26: The spatial representation of the Mean error (BIAS) for a wet season, (March to May), (y-axis is longitude and x-axis is latitude).

(c) BIAS (Mean error) for a wet season (March to May)

Figure 26 represents the spatial mean error (BAIS) for a wet season (March to May). The BAIS values for this period ranged from -0.024 to +0.03. These values were within the threshold value of ± 0.038 , although the targeted values were ± 0.025 . Northern and some parts of Central Kenya reported positive BIAS values

with the highest values being concentrated around Marsabit. Western, Northwestern, Southern and the Coastal strip had negative BAIS values.

4.2.5.3 Mean absolute error (MAE)

The mean absolute errors were determined using both temporal and spatial method. The results are presented below.

4.2.5.3.1 Mean absolute error (MAE) in temporal mode

Figure 27 gives the results of mean absolute errors in temporal mode. The errors ranged from 0.033 to 0.1727, which is considered as an outlier. Compared to the results in spatial mode method, this method generated larger errors. The targeted MAE was 0.025 and the threshold was 0.038. The lowest value was recorded in August 1999 while the highest was recorded in January 2008. However, most of the values lie between 0.04 and 0.12.

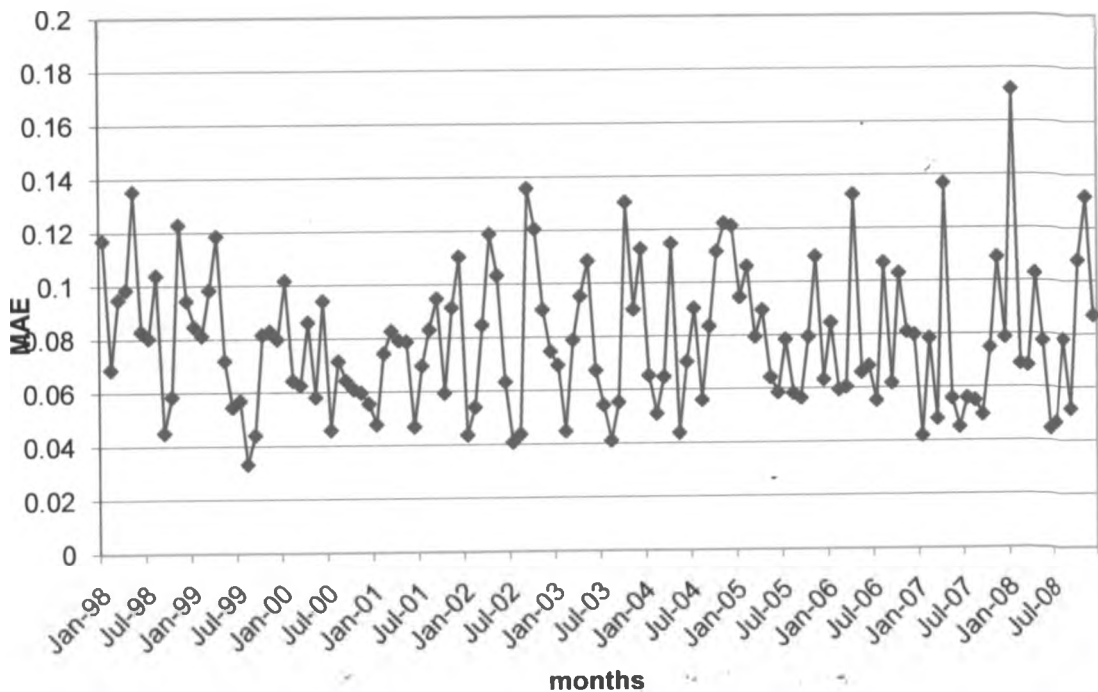


Fig 27: The temporal representation of the mean absolute error (MAE)

4.2.5.3.2 Mean absolute error (MAE) in spatial mode

(a) Overall Mean absolute error (MAE)

Figure 28 gives the results for the overall mean absolute error. The MAE ranged from 0.011 to 0.046, however more than half of the country had errors below the threshold value, (0.038). Western, Central, Southern Kenya and the Coast reported highest mean absolute errors. Northern Kenya reported lowest MAE values. The errors obtained in this mode are smaller than those in temporal mode method.

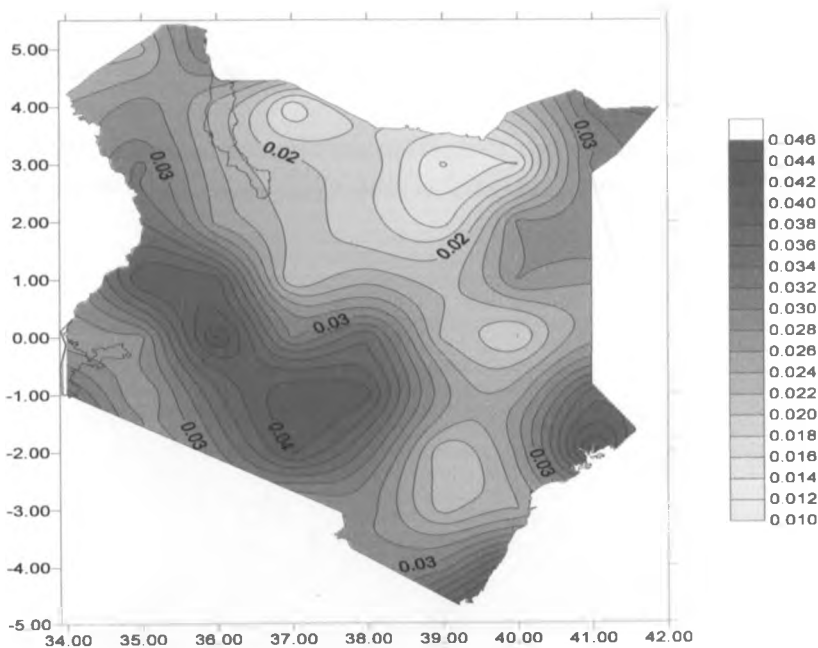


Fig 28: The overall spatial representation of the Mean absolute error (MAE), (y-axis is longitude and x-axis is latitude).

(b) Mean absolute error for a dry season (January to February),

Figure 29 gives the results MAE for January to February (dry season). This season had mean absolute errors ranging from 0.009 to 0.048. The largest mean absolute errors values were reported over Western and Northern Kenya, however more than half of the country had errors below the threshold value, (0.038). In both cases TRMM algorithm seemed to overestimate rainfall over the Western and Northern parts of the country.

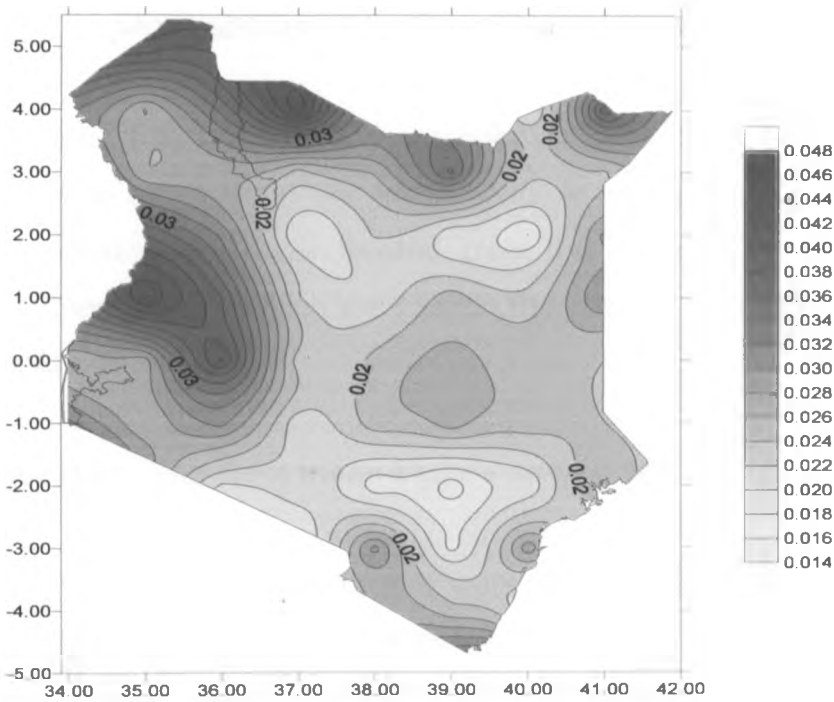


Fig 29: The spatial representation of the Mean absolute error (MAE) for a dry season (January to February), (y-axis is longitude and x-axis is latitude).

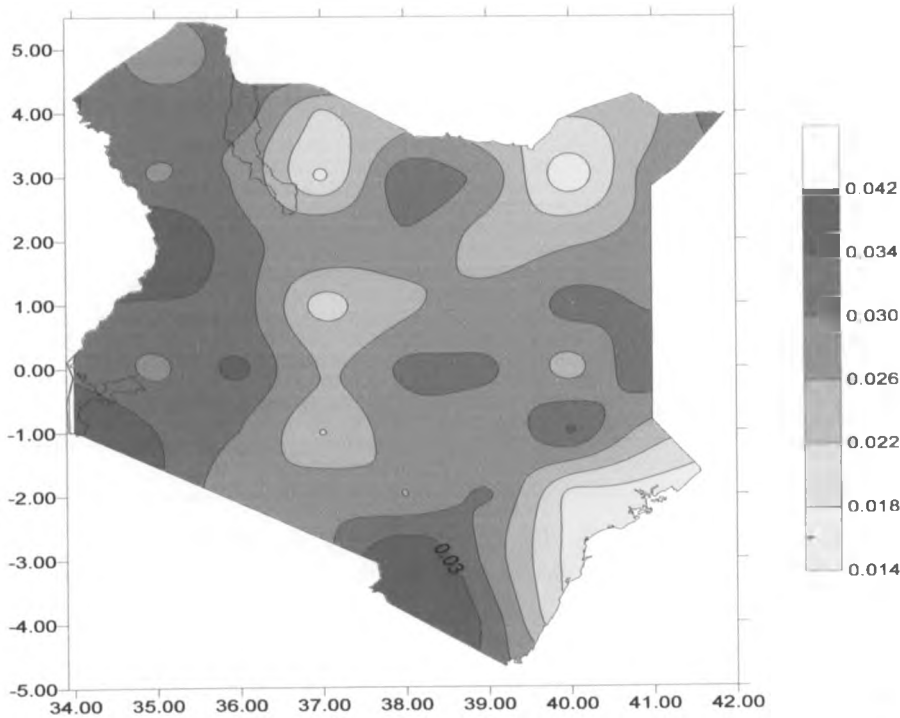


Fig 30: The spatial representation of the Mean absolute error (MAE) for a wet season (March to May), (y-axis is longitude and x-axis is latitude).

(c) Mean absolute error for a wet season (March to May)

Figure 30 is the spatial representation of MAE for long rain season, (March to May). In this season the mean absolute errors ranged from 0.01 to 0.043. The smallest errors were located over the Coast, Central and some parts over northern Kenya, which were within the acceptable limit of 0.025 and the threshold value of 0.038.

4.2.5.4 The Root mean square error (RMSE)

The root mean square errors were determined using both temporal and spatial method. The results are presented below.

4.2.5.4.1 The root mean square error (RMSE) in temporal mode method

Figure 31 is the temporal representation of root mean square error. The RMSE ranged from 0.039 to 0.22. These errors were above the targeted and the threshold values, (0.027 and 0.041). The lowest RMSE value was recorded on August 1999 while the highest value was recorded on January 2008.

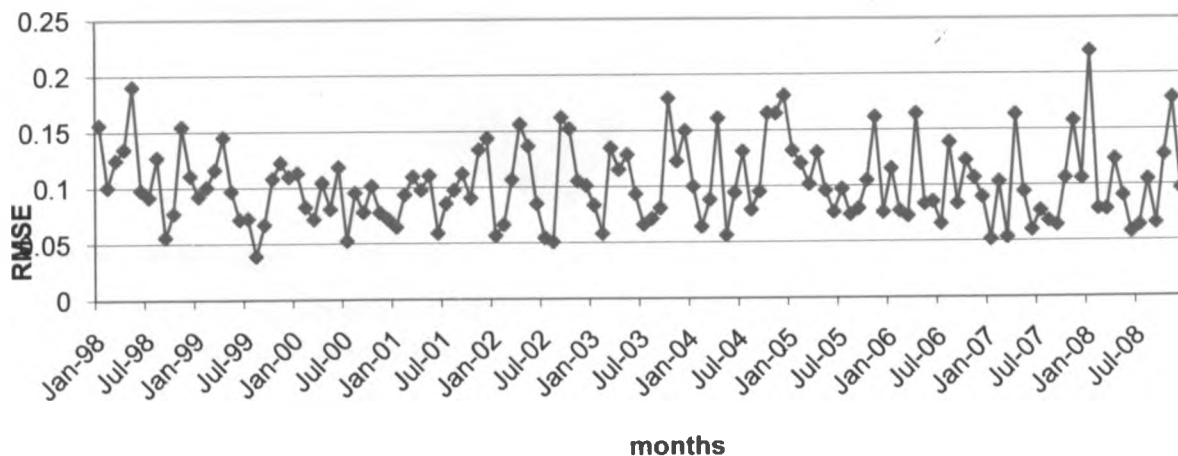


Fig 31: The temporal representation of root mean square error (RMSE)

4.2.5.4.2 Root Mean squared error in spatial mode method

(a) Overall Root mean squared error, (RMSE).

Figure 32 gives the results of the overall RMSE in spatial mode. These errors ranged from 0.018 to 0.06. More than half of the country reported RMSE values less than the threshold value, (0.041). This range was smaller than that obtained from temporal method. Western, Central and the Northern Coast recorded the largest root mean squared errors. The northern and Southern Kenya recorded small errors.

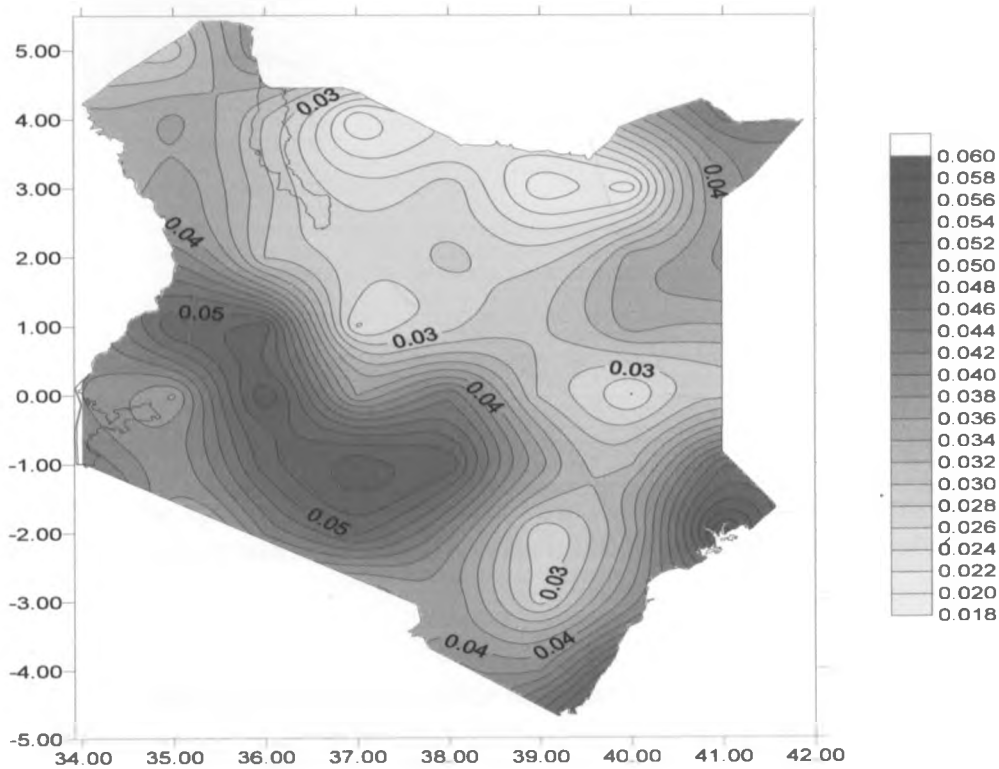


Fig 32: Overall spatial representation of the Root Mean Square Error (RMSE), (y-axis is longitude and x-axis is latitude).

(b) Root mean square error for a dry season (January to February)

Figure 33 is the spatial representation of RMSE for the dry season (January to February). These errors range from 0.012 to 0.06. Apart from Western and Northwestern Kenya, most parts of the country recorded smaller root mean

squared error than the threshold value. The Northern part of the country reported RMSE values which were less than the target, (0.027).

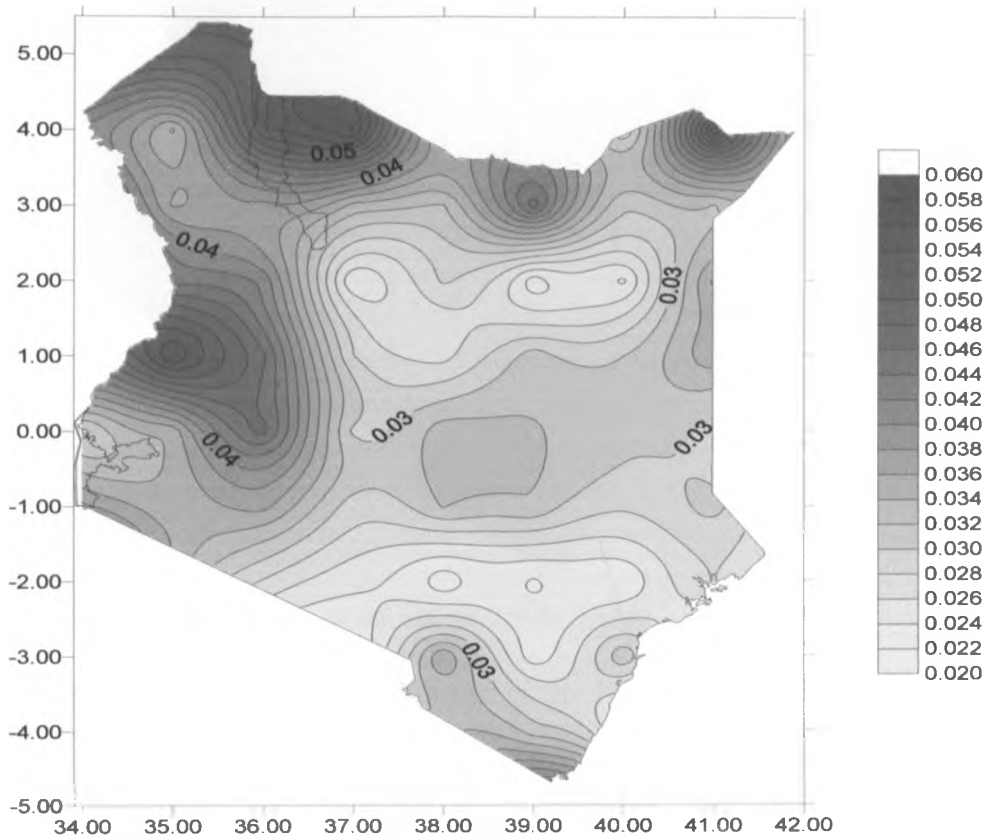


Fig 33: The spatial representation of the Root Mean Square Error (RMSE) for a dry season (January and February), (y-axis is longitude and x-axis is latitude).

(c) Root mean square error for a wet season (March to May)

Figure 34 gives the results of RMSE for March to May in spatial mode. Over this period, the RMSE ranged from 0.012 to 0.060. Western, Northwestern and Southern sectors of the country recorded the larger than the target errors while Coast recorded smaller than the targeted root mean squared errors, (0.027). In both seasons Northwestern and the western sectors of the country seemed to have larger errors than the rest of the country;

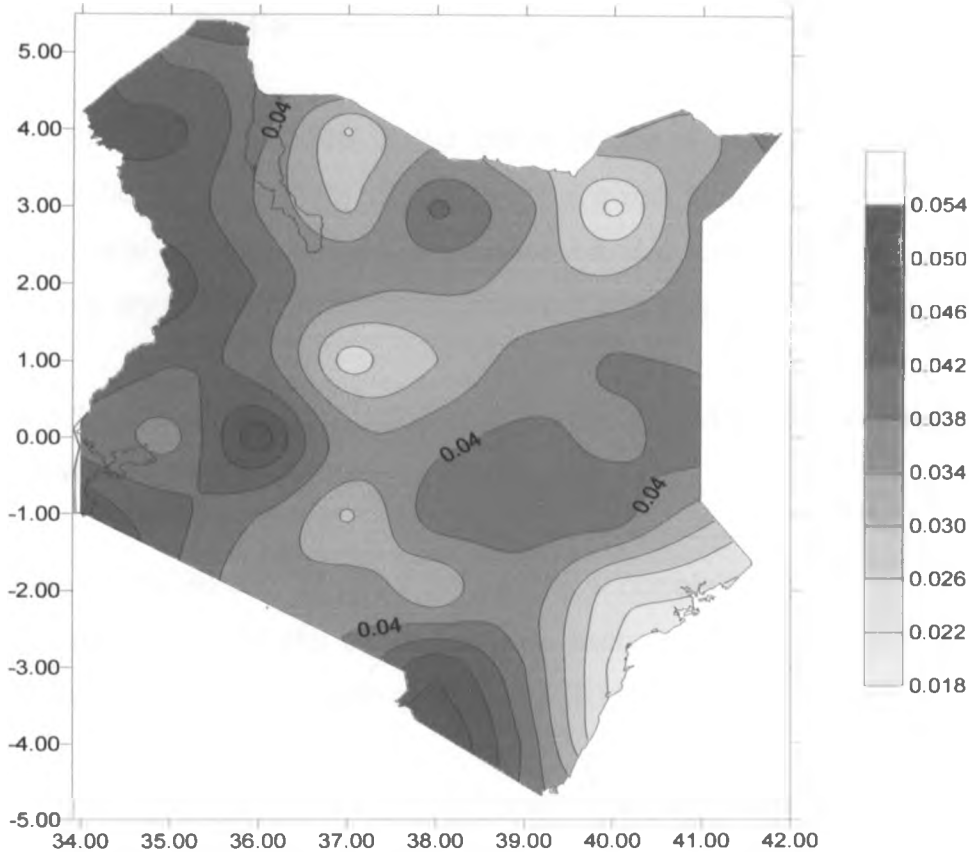


Fig 34: The spatial representation of the Root Mean Square Error (RMSE) for a wet season (March to May), (y-axis is longitude and x-axis is latitude).

4.3 Canonical correlation analysis (CCA) models

A set of weights for predictor components (elements) is related to an analogous set of weights for the predictand (estimated) components. These weighting sets called loading patterns illustrate the associated predictor-predictand patterns and thus can be used as guidance to the underlying physical processes as well for descriptive purposes. The magnitudes of the loadings represent the relative importance of predictor and the predictand to prediction (estimation) skill provided by CCA models in this study, Barnston and He, (1996). In this study, these loadings are presented in two forms, the spatial and the temporal modes. Hence we have spatial and temporal CCA models.

4.3.1 Canonical correlation analysis (CCA) spatial models

To examine the nature of the contributions of the predictor field to the skill of predictions (estimation) of rainfall of a given location, the spatial distribution of canonical loading patterns are examined. We forecast first on the overall period, then a dry season (January to February) and finally a wet season (March to May).

4.3.1.1 Canonical correlation analysis (CCA) spatial models for the overall period

Figures 35, 36 and 37 show the canonical correlation patterns for the predictor (fig a) and predictand (fig b) for the first, second and third modes respectively. The CCA canonical correlation coefficients (R) for these modes are also shown.

From figures 35a and 35b, we found that positive loadings are found over the Southern part of the country extending to the Coast, Central Kenya and the Lake Victoria region. Negative loadings are found over the Western, the Northern and Northeastern Kenya. The canonical correlation coefficient between TRMM rainfall and the areal rainfall was 0.9767. This pattern accounted for 14.56% of the total variance of TRMM rainfall or Kriged rainfall in the data set analyzed.

It should be noted there that prediction (estimation) was constructed using the pattern vector shown in fig 35 (a) to predict (estimate) the pattern in 35 (b).

From fig 36 (a) and 36 (b), the positive loadings were found over the Western sector while the negative loadings were recorded over the Eastern sector of the country. The canonical correlation coefficient between the two fields in this mode was 0.9197. This pattern accounted for 13.71% of the total variance of TRMM rainfall or Kriged rainfall in the data set analyzed.

Figures 37 (a) and 37 (b) showed that the positive canonical loadings were dominant over the Western, Northwestern, Kenya and the Coastal strip. The negative canonical loadings were found over Northeastern and Northern Kenya extending all the way to Central, Southern and Southeastern Kenya. The

canonical correlation between areal rainfall and TRMM rainfall was 0.9156. This pattern accounted for 13.71% of the total variance of TRMM rainfall or Kriged rainfall in the data set analyzed.

Positive loadings indicated their strong predictive value while weak ones indicated their lesser predictive value. Mode 1 (fig 35) would be therefore suitable for estimating areal rainfall over the Southern sector (half) of the country. Areal rainfall over the Western sector would be estimated by mode 2 (fig 36) while Coastal rainfall would be estimated using the third model (fig 37).

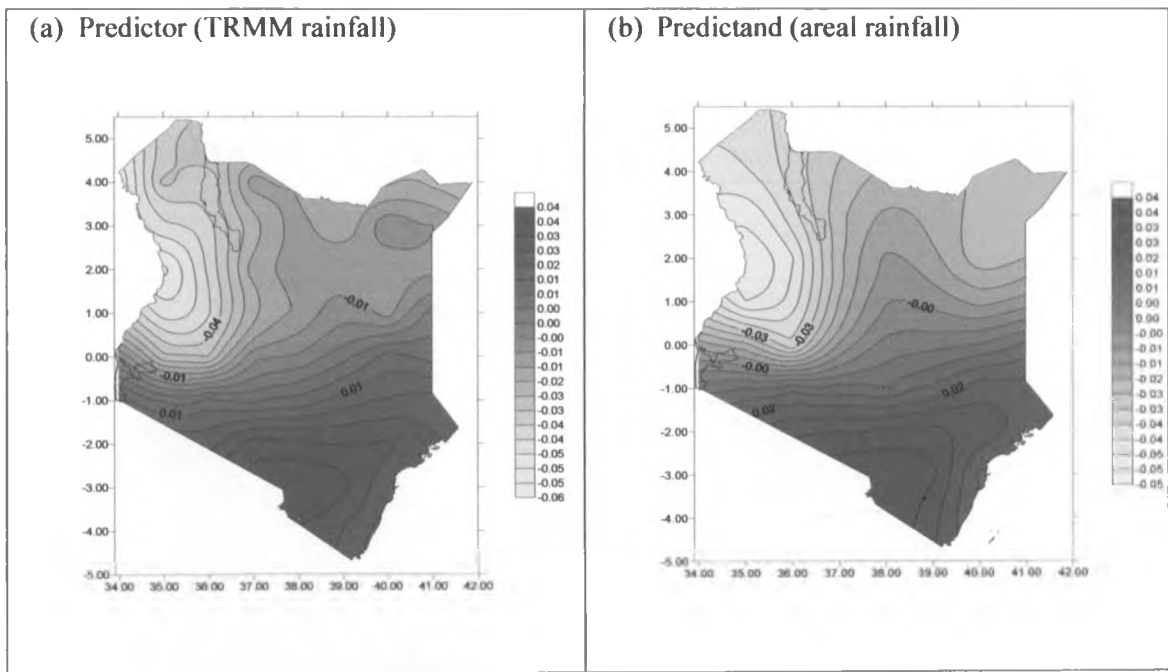


Fig 35 TRMM rainfall predictor (a) and Areal rainfall predictand (b) CCA loadings for the overall period for mode 1, (y-axis is longitude and x-axis is latitude), $R=0.9767$

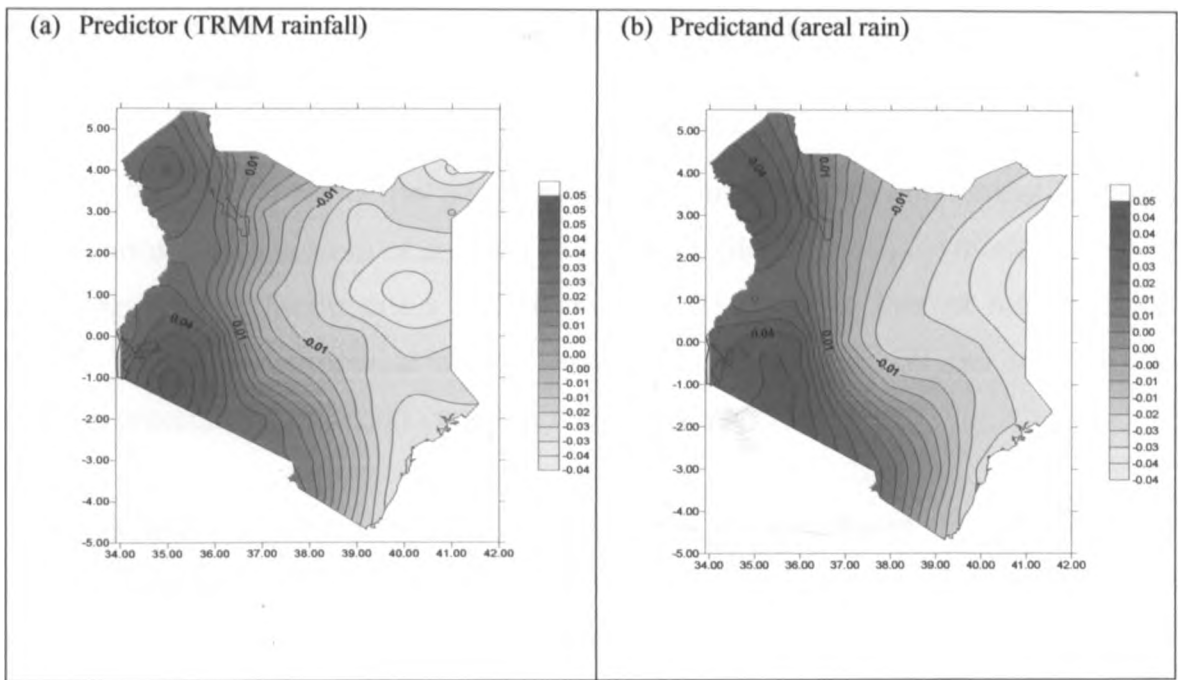


Fig 36 TRMM rainfall predictor (a) and Areal rainfall predictand (b) CCA loadings the overall period for mode 2, (y-axis is longitude and x-axis is latitude), $R=0.9197$

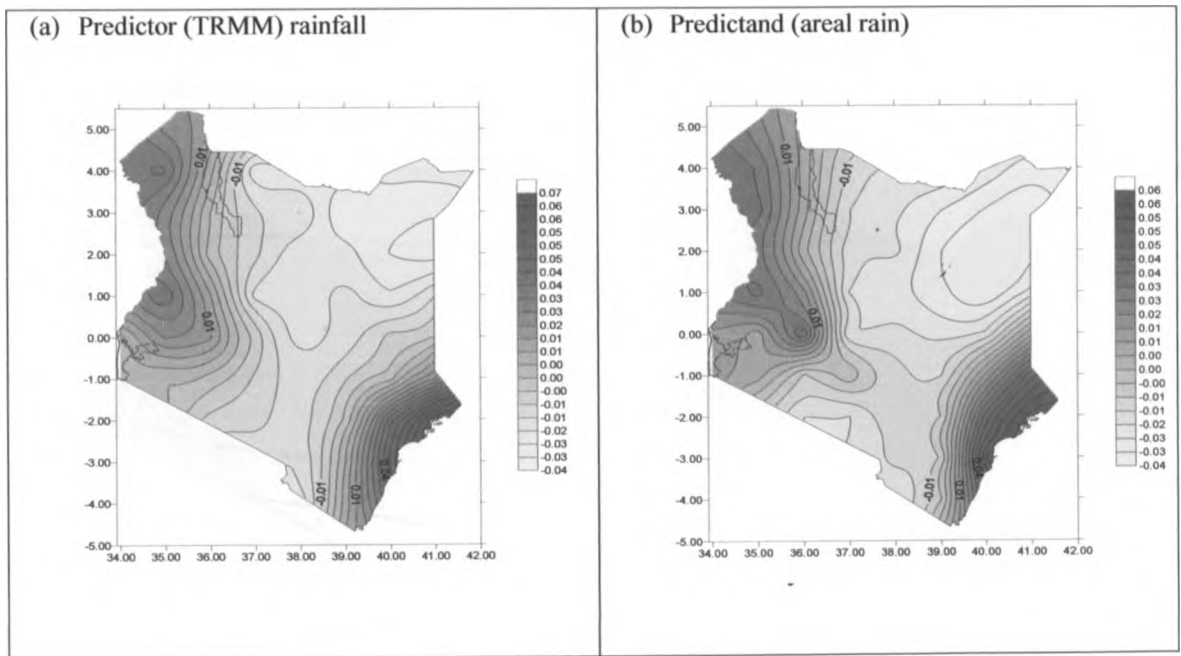


Fig 37 TRMM rainfall predictor (a) and Areal rainfall predictand (b) CCA loadings the overall period for mode 3, (y-axis is longitude and x-axis is latitude), $R=0.9156$

4.3.1.2 Canonical correlation analysis (CCA) spatial models for dry season (January to February)

Figures 38 and 39 show the geographical distribution of the canonical loadings and canonical coefficients of January to February (dry season) for modes one and two respectively. Mode one's coefficient was 0.926 while that of mode 2 was 0.807. These two canonical modes explained 24.9 and 21.69 per cent of the original predictor/ predictand variance respectively.

The principal predictor and predictand loadings for mode one (fig 38) indicated positive values over Northwestern, Western and Southern Kenya. Negative loading values were recorded over Northeastern parts of the country while the rest of the country recorded near zero values. This mode would be suitable for estimation of rainfall over the Northwestern and Western sector of the country.

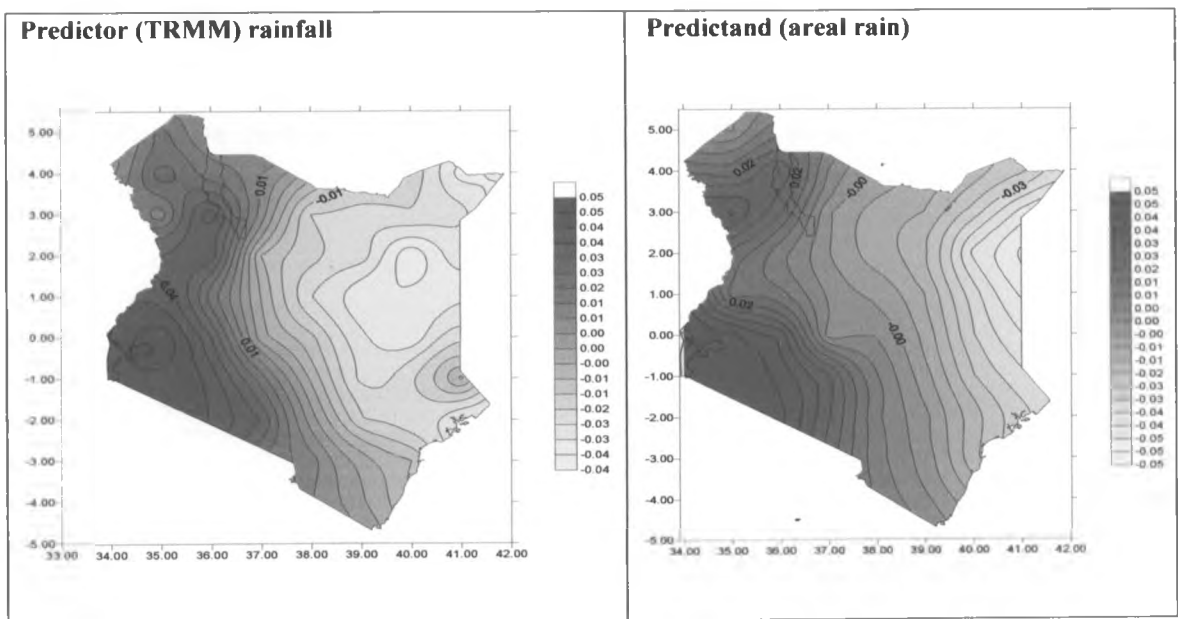


Fig 38 TRMM rainfall predictor (a) and Areal rainfall predictand (b) CCA loadings for mode 1 for January to February, (y-axis is longitude and x-axis is latitude), $R=0.926$

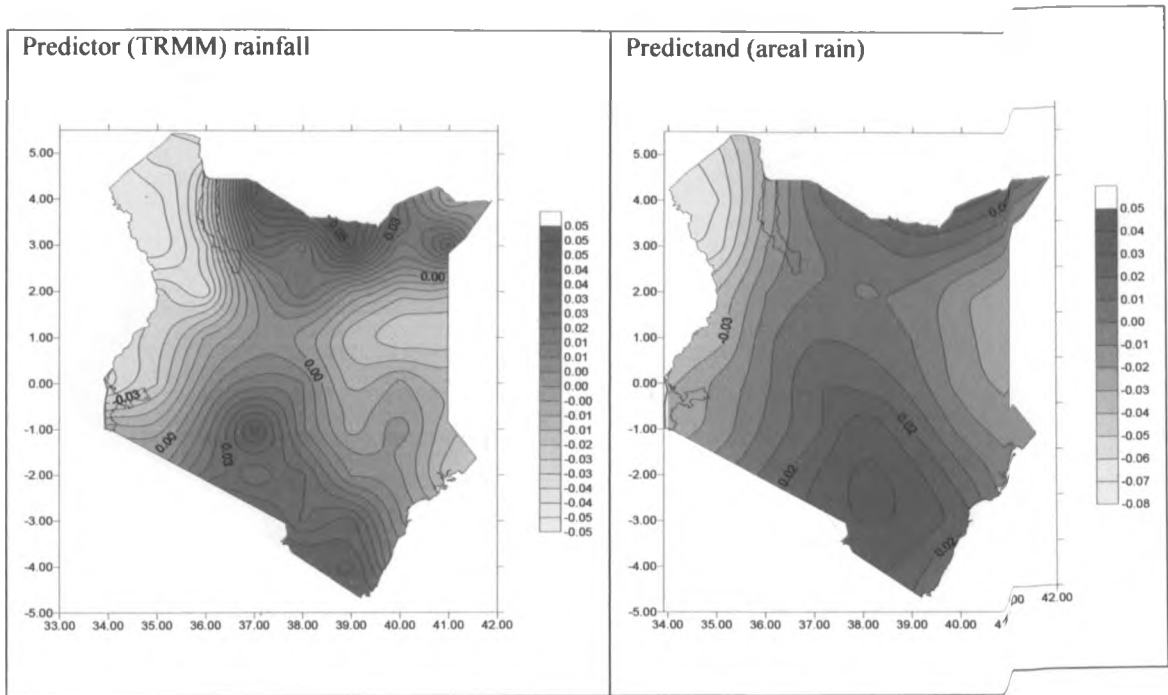


Fig 39 TRMM rainfall predictor (a) and Areal rainfall predictand (b) CCA loadings for mode 2 for January to February, ($R=0.807$, y-axis is longitude and x-axis is latitude)

Figures 39 (a) and (b) indicated that the positive loadings were found over the extreme Northeastern, Northern, Central, Southern Kenya and the Coast. Northwestern, Western and extreme Eastern Kenya had negative values. This mode would be suitable for areal rainfall estimation in areas with positive loadings named above.

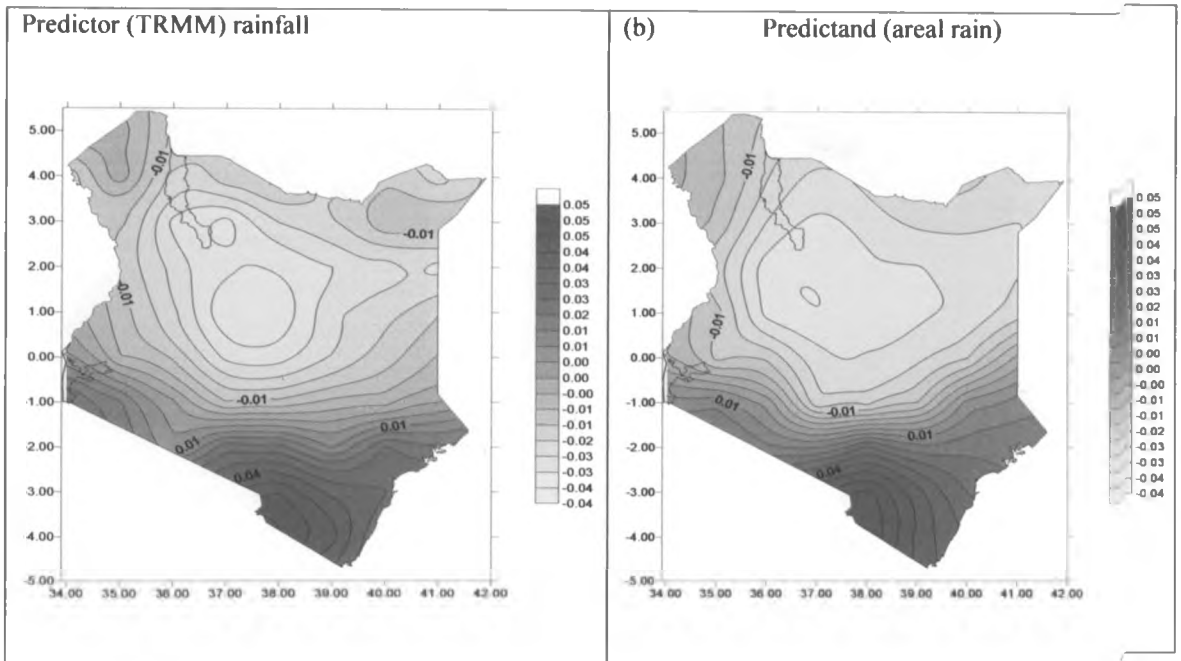


Fig 40 TRMM rainfall predictor (a) and Areal rainfall predictand (b) CCA loadings for mode 1 for March, April and May, (y-axis is longitude and x-axis is latitude), $R=0.967$

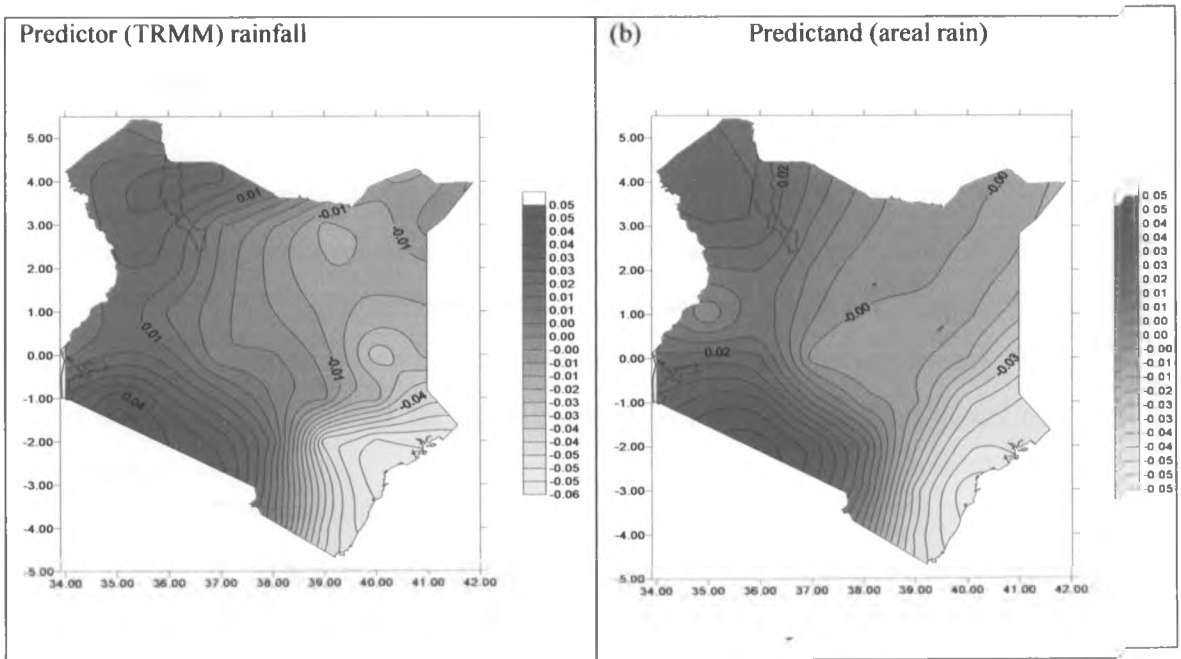


Fig 41 TRMM rainfall predictor (a) and Areal rainfall predictand (b) CCA loadings for mode 2 for March, April and May, (y-axis is longitude and x-axis is latitude), $R=0.951$

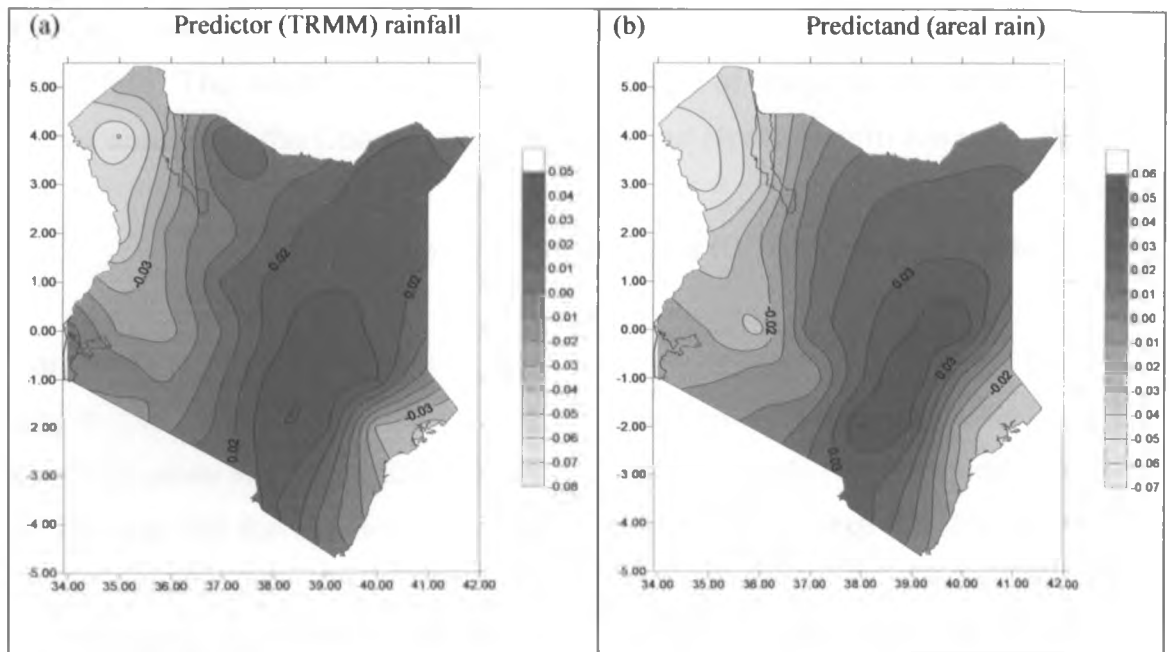


Fig 42 TRMM rainfall predictor (a) and Areal rainfall predictand (b) CCA loadings for mode 3 for March, April and May, (y-axis is longitude and x-axis is latitude), $R=0.944$

4.3.1.3 Canonical correlation analysis (CCA) spatial models for wet season (March to May)

The first and important three modes were examined. The first mode (fig 40) had canonical correlation coefficient of 0.967. The Southern part of the country including the Coast had positive canonical loadings while the rest of the country had negative canonical loadings indicating lesser predictive value. This mode would be therefore suitable for areal rainfall estimation over the Southern part of the country.

Figures 41(a) and 41(b) indicated that the Coast had negative canonical loadings while Northeastern Kenya had near zero values. The rest of the country had positive loadings. The canonical correlation coefficient in this model was 0.951. The mode could be used to estimate areal rainfall all over the country apart from the Coast.

Mode 3, figures 42(a) and 42(b) had negative canonical loadings over Northwestern, Western Kenya and the Coastal strip. The rest of the country

recorded positive loadings. The canonical correlation coefficient (R) for this model was 0.944. The mode is suitable for estimation of areal rainfall over the rest of Kenya apart from the Coastal strip, Western and Northwestern Kenya.

4.3.2 Canonical correlation analysis (CCA) temporal models

Figures (43, 44 and 45) show the canonical correlation results of predictand versus the observed. In this case TRMM rainfall estimates were taken as observed while Kriged rainfall field was being predicted. The canonical correlation coefficients (R) for the first, second and third modes were 0.9859, 0.9770 and 0.9677 respectively. These modes however represented 37.45 per cent of the total variance. It is evident from these graphs TRMM rainfall estimates provide reliable estimation information over the entire period (1998-2008), with skills of more than 0.96. The forecasted rainfall captured most significant fluctuations, including those associated with major extremes such as floods and droughts.

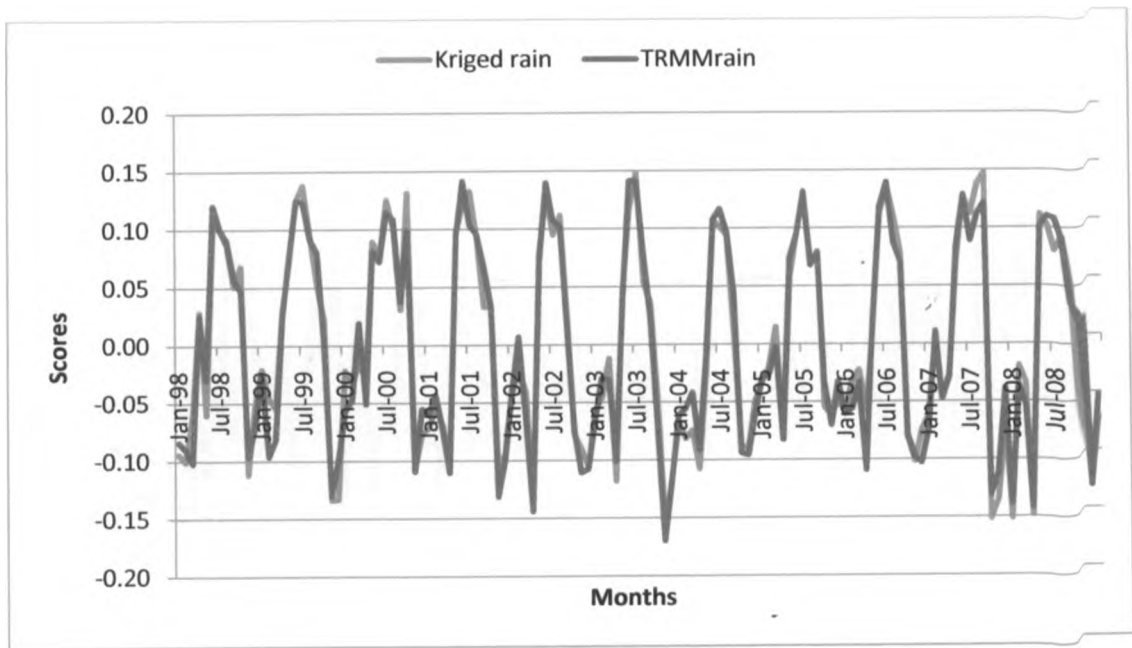


Fig 43: Canonical component predictor (TRMM rain) and predictand (Kriged rain) for mode 1, R=0.9859

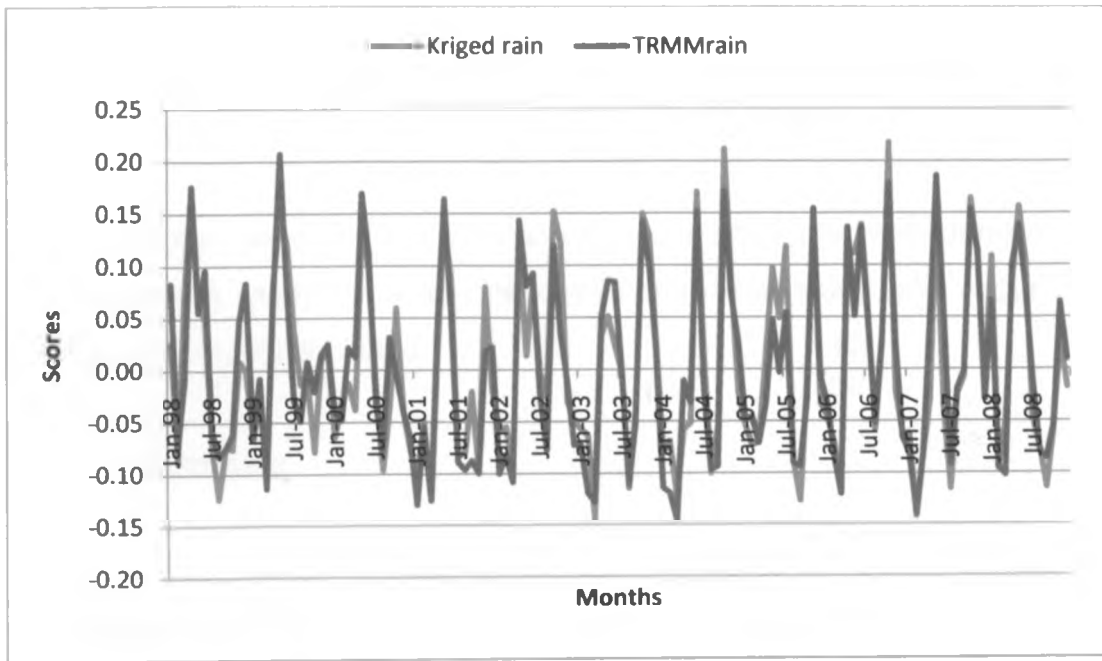


Fig 44: Canonical component predictor (TRMM rain) and predictand (Kriged rain) for mode 2, $R=0.9770$

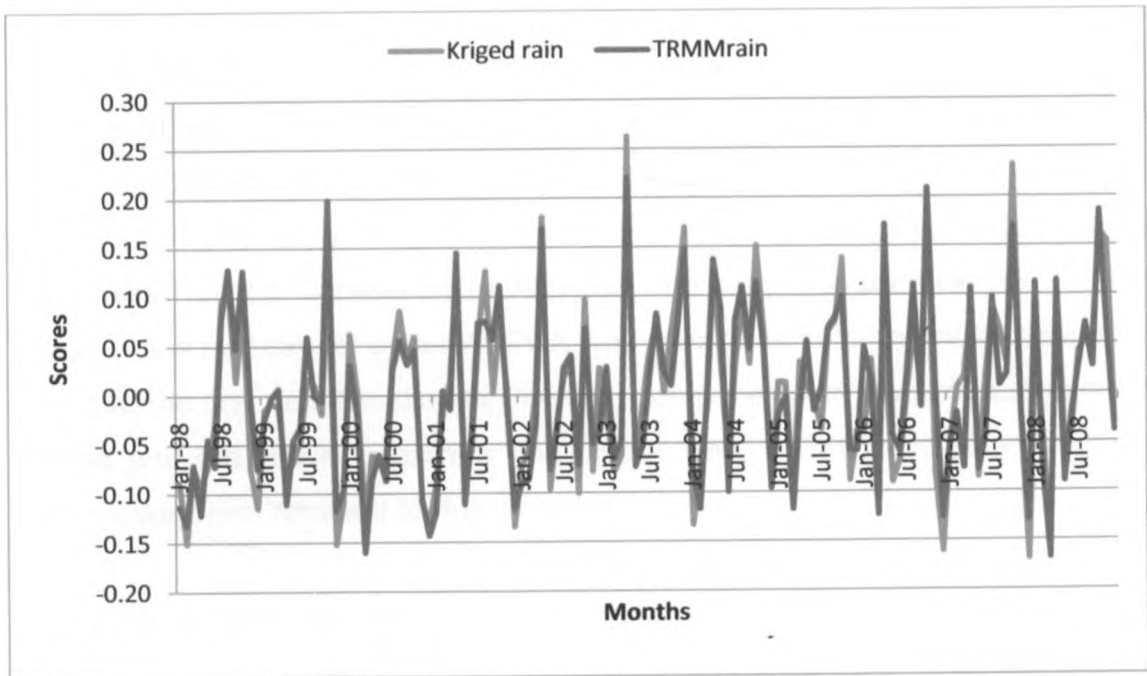


Fig 45: Canonical component predictor (TRMM rain) and predictand (Kriged rain) for mode 3, $R=0.9677$

CHAPTER FIVE

5.0 Summary, Conclusions and Recommendations

In this chapter, an attempt was made to summarize the work done and findings of the study. Conclusions drawn from this study and suggestions to improve it were also included.

5.1 Summary

This study was divided into four broad sections based on specific objectives. The first section dealt with the development of gridded rainfall field based on the rain gauge data over Kenya using Kriging method. This approach allowed an effective use of rain gauge-based observations to compare and validate satellite-based rainfall estimates.

In the second section the two data sets were analyzed using Principal Component Analysis to highlight their similarities and the differences. PCA was done in spatial mode for the two data sets for the overall period of study, wet season (March to May) and also for dry season (January to February). Spatial maps of the loadings of the first Principal Component were drawn in order to compare the two variables. These maps had similar patterns. PCA was also done in temporal mode and time series of the first and the second Principal components of the two data sets were plotted and the resulting graphs had similar trend. Significant Eigen vectors in both modes were retained for further analysis.

In the third part the validity of the satellite derived rainfall estimates was determined using gridded rainfall field as the reference data. This was achieved by use of scatter plots, simple and canonical correlations and also error analysis. Results from scatter plots showed that the two variables were compactly arranged with few outliers. Results from simple and canonical correlations showed large correlation coefficients. The statistical scores used in error analysis were mean absolute percentage error (MAPE), the BIAS (mean error), mean absolute error

(MAE) and the root mean square error (RMSE). The results obtained from these statistical scores indicated that the temporal mode method generated larger errors than the spatial mode method, while wet regions had larger errors than dry regions. Also dry season had smaller errors than the wet season.

Finally spatial and temporal Canonical Correlation (CCA) models were developed using the significant Principal components (PCA) to estimate rainfall in areas with sparse rain gauge network over Kenya. The study found no spatial mode was similar to the other. July was found to have largest CCA loadings while January had the smallest.

5.2 Conclusions

Gridding rain gauge data over Kenya using Kriging technique was found to be a good tool of solving the problem of mismatch between the satellites derived rainfall estimates and the rain gauge data (observations). Kriging method however assumes that the data being interpolated are homogeneous in space, which is not the case in Kenya. This probably resulted to generation of larger errors than expected especially over wet regions and wet season.

Conclusions drawn from Principal Component analysis was that this analysis provided large-scale information about the joint behaviour between TRMM's satellite derived rainfall estimates and the gridded rainfall. Of particular interest were the spatial distribution and the time series of the loadings of the Principal components (Eigen vectors). Spatial maps generated by the loadings of the first principal components of TRMM's satellite derived rainfall estimates and the gridded (Kriged) rainfall for the overall period of study, wet and dry seasons had identical patterns. The time series of the first and the second Principal components of the two variables had the same trend. These results show how the two data sets are closely related to each other and it is concluded that satellite derived rainfall estimates may be used to represent areal rainfall in regions with sparse gauge network.

The compactly arrangements of the two variables in scatter plots and large correlation coefficients from simple and canonical correlation analysis for the overall period of study, dry and wet seasons are indicators of a strong relationship between TRMM satellite derived rainfall estimates and the gridded rainfall data over Kenya. Based on these on these results, it may be suggested that the satellite derived rainfall estimates may be used to estimate rainfall in areas without rain gauge observatories.

Error analysis results from the temporal mode method showed that this method generated larger errors than the spatial method. The loadings of principal components (Eigen vectors) of the two data sets were used to determine errors. In principal component analysis, the temporal mode method disregards geographical locations while the spatial mode method data of each grid point is time averaged. This explains different error analysis results between the two methods.

Both the spatial mode and temporal mode methods generated larger errors than expected. There are three possible explanations for these findings, first the passive microwave sensors are carried on board polar orbiting TRMM satellite, with sensors providing only two to three overpasses at any given point each day. These observations from satellites are very few and may have lead to temporal sampling errors.

Another possible source of errors is inadequate validation rain gauge data. This data was generated through interpolation techniques which introduced errors.

The dry season had smaller errors than the wet season and also the dry parts of Kenya had smaller errors than the wet areas. These results suggest that TRMM algorithm tends to overestimates rainfall in wet season or in wet areas. This overestimation may have been caused by the presence of many different types of clouds, in wet season and regions. These clouds tend to decrease the brightness temperature observed in micro-wave channels.

CHAPTER SIX

6.0 REFERENCES

- Asnani G. C., (2005): Tropical meteorology vol. 1 to vol. 3, second edition, India Institute of Tropical Meteorology, Pashan.
- Barnston G. A. and He, (1996): Skill of canonical correlation analysis forecasts of 3-month mean surface climate in Hawaii and Alaska, J. climate, vol 9, pp 2579-2605.
- Barnston G. A. and Ropelewski F. C, (1992): Prediction of ENSO Episodes using canonical correlation analysis, J climate, vol 5, pp 1316-1345.
- Basalirwa, C.P.K (1979): Estimation of areal rainfall in some catchments of upper Tana River. MSc, Thesis. Dept. of meteorology, University of Nairobi, Kenya.
- Bowman K. P., (2005): Comparison of TRMM Precipitation Retrievals with Rain Gauge Data from Ocean Buoys, J. Climate, vol 18, pp 178-190.
- Bringi V. N., Chandrasekar V., Hubbert J., Randue L. W., & Schoenhuber M., (2003): Raindrop size distribution in different climatic regimes from disdrometer and dual-polarized radar analysis, J. Atmos. Sci., vol 60, pp 354-365.
- C. Daly, (2006): Guidelines for assessing the suitability of spatial climate data sets, Int. J. Climatology. 26 707-721.
- Creutin, J. D. and Obbled, C., (1982): Objective analysis and mapping techniques for rainfall field: an objective comparison. Water resour. Res. 18, 413-431.
- Daniel S. Wlks, (1995), Statistical methods in the atmospheric sciences, Academic press, volume 59, International Geophysics series.

From CCA temporal-mode models, July has high predictability while January has the smallest.

5.3 Recommendations

It is recommended that gridded rainfall field based on rain gauge data should be improved in Kenya through testing and application of various interpolation techniques. This will minimize errors generated by interpolation techniques.

Principal Component Analysis reduces large number of correlated data sets in time and space into small number of orthogonal functions that are linear combinations of the original observations and account for a large percentage of the total variance. In doing so some important information may have be discarded when truncating the PCA. It is therefore recommended that validation for satellite derived estimates over Kenya be done separately for each homogeneous zone. This will reduce the size of the two data sets.

In this study the satellite derived rainfall estimate used were obtained from combination of sensors (infra-red, passive microwave and precipitation radar). It is recommended that validation of rainfall estimates derived from each sensor be validated separately over Kenya. It is further recommended that rainfall estimates derived from different algorithms and satellites should also be validated over Kenya.

It is recommended that further studies should be carried out in order to find out why July has high predictability while January has the smallest. It is further recommended that models for estimating areal rainfall over Kenya using satellite derived rainfall estimates using other techniques apart from CCA should also be developed and tested.

CHAPTER SIX

6.0 REFERENCES

Asnani G. C., (2005): Tropical meteorology vol. 1 to vol. 3, second edition, India Institute of Tropical Meteorology, Pashan.

Barnston G. A. and He, (1996): Skill of canonical correlation analysis forecasts of 3-month mean surface climate in Hawaii and Alaska, *J. climate*, vol 9, pp 2579-2605.

Barnston G. A. and Ropelewski F. C, (1992): Prediction of ENSO Episodes using canonical correlation analysis, *J climate*, vol 5, pp 1316-1345.

Basalirwa, C.P.K (1979): Estimation of areal rainfall in some catchments of upper Tana River. MSc, Thesis. Dept. of meteorology, University of Nairobi, Kenya.

Bowman K. P., (2005): Comparison of TRMM Precipitation Retrievals with Rain Gauge Data from Ocean Buoys, *J. Climate*, vol 18, pp 178-190.

Bringi V. N., Chandrasekar V., Hubbert J., Randue L. W., & Schoenhuber M., (2003): Raindrop size distribution in different climatic regimes from disdrometer and dual-polarized radar analysis, *J. Atmos. Sci.*, vol 60, pp 354-365.

C. Daly, (2006): Guidelines for assessing the suitability of spatial climate data sets, *Int. J. Climatology*. 26 707-721.

Creutin, J. D. and Obbled, C., (1982): Objective analysis and mapping techniques for rainfall field: an objective comparison. *Water resour. Res.* 18, 413-431.

Daniel S. Wilks, (1995), *Statistical methods in the atmospheric sciences*, Academic press, volume 59, International Geophysics series.

Dinku T. and Anagnostou E. N., (2005): Regional Differences in overland Rainfall Estimates from PR-calibrated TMI Algorithm. *J. Appl meteor*, vol 44, pp 189-205.

Ebert E. E., J. Janowiak, C. Kidd, (2006), Comparison of near real precipitation estimates from satellite observations and numerical models, *Bull. Amer. Meteor. Soc*, 88, 47-64.

Fisher L. B., (2004) : Climatological Validation of TRMM TMI and PR Monthly Rain Products over Oklahoma, *J. Appl meteor*, vol 43, pp 519-535.

Furuzawa F. A. and Nakamura k., (2005): Differences of Rainfall Estimates over Land by Tropical Rainfall Measuring Mission (TRMM) Precipitation Radar (PR) and TRMM Microwave Imager (TMI)—Dependence on Storm Height. *J. Appl meteor*, vol 44, pp 367-383.

Gebremichael M, Krajewsk F. W., Morrissey L. M., Huffman J. G. and Adler F. R., (2005): A Detailed Evaluation of GPCP 1° Daily Rainfall Estimates over the Mississippi River Basin. *J. Appl meteor*, vol 44, pp 665-681.

Gruber A. and Levizzani V., (2008): Assessment of global precipitation products, World climate research programme, (WCRP-128), WMO/TD-NO. 1430.

Hartkamp A. D et al, (1999): Interpolation techniques for climate variables, ISSN: 1405-7484, NRG, Geographic information systems series 99-01.

Haylock M. R et al (2008): A European daily high-resolution gridded data set of surface temperature and precipitation for 1950-2006. *J. Geophys. Res.* Vol. 13, D20119.

Hong Y., Hsu Kuo-Lin, Sorooshian S. G. X. (2004): Precipitation Estimation from Remotely Sensed Imagery Using an Artificial Neural Network, Cloud Classification System. *J. Appl meteor*, vol 43, pp 1834-1853.

Ininda J. M, Athumani C, Mutemi J. N, (2008): Towards improvement of seasonal rainfall forecasting through model output statistics (MOS) downscaling of Echar forecasts over Tanzania, J. Kenya Meteor soc, Vol 2, No. 2, pp 99-108.

JAXA/EOC, (2006): TRMM Data Users Handbook, 3rd Edition, Earth Observation Center JAXA 1401, Numanoue, Hash, Hatoyama-machi Hiki-gun Saitama-Ke Japan.

Kim M. J., Weinman J. A, Houze R. A., (2004): Validation of Maritime Rainfall Retrievals from the TRMM Microwave Radiometer. J. Appl meteor, vol 43, pp 847-859.

Klepp C., P., Bakan S., Hartmut G., (2003): Improvements of Satellite-Derived Cyclonic Rainfall over the North Atlantic. J. climate, vol Improvements of Satellite-Derived Cyclonic Rainfall over the North Atlantic 16, pp 657-669.

Kummerow, C. D., W. Barnes, T. Kozu, J. Shiue, and J. Simpson, (1998): Tropical Rainfall Measuring Mission (TRMM) sensor package. J. Atmos. Oceanic Technol., 15, 809-817.

Lillesand K. (1979): Remote sensing and image interpretation, third edition, John Wiley & Sons, Inc.

Ministry of information & communication, (2006): 2006 Kenya ICT strategy, collaboration and outsourcing for economic growth.

Mutemi J. N. (2003): Climate anomalies over Eastern Africa associated with various ENSO evolution phases. PhD., Thesis, University of Nairobi.

Mutua, F. M, Mutemi J. M and Oludhe, C, (1999): Homogeneous Climatic zones, Drought Monitoring Center (DMC) lecture notes, Chapter 3, p29-43.

Naoum and Tsanis, (2004): Ranking spatial interpolation techniques using a GIS-based DSS. *Global Nest, Int. J. Vol 6, No 1*, pp 1-20.

Nesbit S., W. and Zipser E., J., Kummerow C. D. (2004): An Examination of Version-5 Rainfall Estimates from the TRMM Microwave Imager, Precipitation Radar, and Rain Gauges on Global, Regional, and Storm Scales. *J. Appl meteor*, vol 43, pp 1016-1036.

Nesbitt S. W. and Zipser E. W., (2003): The Diurnal Cycle of Rainfall and Convective Intensity according to Three Years of TRMM Measurements. *Climate journal*. vol 16, pp 1456-1475.

Nicholson S. E. and Coauthors (2003): Validation of TRMM and Other Rainfall Estimates with a High-Density Gauge Dataset for West Africa. Part II: Validation of TRMM Rainfall Products, *J. Appl meteor*, vol 42, pp 1355-1368.

Nieuwolt, (1977), *Tropical meteorology*, John Wiley & Sons

Nyakwada, W (1991): Relationships between satellite-derived outgoing long wave radiation and some important meteorological parameters. Master of Science Thesis, Department of meteorology, University of Nairobi, Kenya.

Nzeukou A., Souvageot H., Ochou A. D. , Kebe C. M. F. , (2004): Raindrop Size Distribution and Radar Parameters at Cape Verde. *J. Appl meteor*, vol 43, pp 90-105.

Okoola, R. E, (1999): A diagnostic study of the East African monsoon circulation during the northern spring. *Inter. J. Climatol*. 19, 143-168

Omiti J., Otieno D., Nyanamba T. and Cullough E., (2009): Factors influencing the intensity of market participation by smallholder farmers: A case study of rural and peri-urban areas of Kenya. *African Journal of Agricultural and Resource Economics*, (Afjare), Vol 3, No 1, pp 57-82.

Ouma G. O., (1988): Estimation of areal rainfall from satellite data. Thesis. Master of Science. University of Nairobi, Kenya.

R. Uijlenhoet, Porra J. M., Sempere Torres and Creutin J.D., (2009): Edge effect causes apparent fractal correlation dimension of uniform spatial raindrop distribution. *Nonlin. Processes Geophys.*, vol 16, pp 287–297.

Richman, M. B. (1986): Rotation of principal components. Review article. *J. climate*, 6, pp 293-335.

Ryznkhov A. V., Giangrande S. E., Schuur T. J. , (2005): Rainfall Estimation with a Polarimetric Prototype of WSR-88D. *J. Appl meteor*, vol 44, pp 502-515.

Sempere Torres D, Porra J. M. & Creutin J. D., (1994): A general formulation for raindrop size distribution. *J. Appl meteor*, vol 33, pp 1494-1502.

Sun X., M. J. Manton & E. E. Ebert, (2003): Regional rainfall estimation using double Kriging of rainfall and satellite observations, BMRC Research report No. 94.

Taesombat W. and Nutchant S., (2009): Areal rainfall estimation using spatial interpolation techniques. *Science Asia*, vol 35, pp268–275.

Tapiador F. J., Kidd C., Levizzani V., Marzano F. S., (2004): A Neural Networks–Based Fusion Technique to Estimate Half-Hourly Rainfall Estimates at 0.18 Resolutions from Satellite Passive Microwave and Infrared Data. *J. Appl meteor*, vol 43, pp 576-594.