

# Development of High Spatial Resolution Rainfall Climatology for Ghana

by

Aryee Jeffrey, B Sc (Meteorology and Climate Science)

A Thesis Submitted to the Department of Physics,  
Kwame Nkrumah University of Science and Technology in

partial fulfillment of the requirements for the degree

of



MASTER OF PHILOSOPHY

(Meteorology and Climate Science)

College of Science

Supervisor:-Dr. Leonard K. Amekudzi

c Department of Physics

July 2015

# Declaration

I hereby declare that this thesis is my own work towards the MPhil and that, to the best of my knowledge, it contains no material previously published by another person nor material which has been accepted for the award of any other degree of the University, except where due acknowledgement has been made in the text.

.....  
Student Name & ID                      Signature                      Date

Certified by:

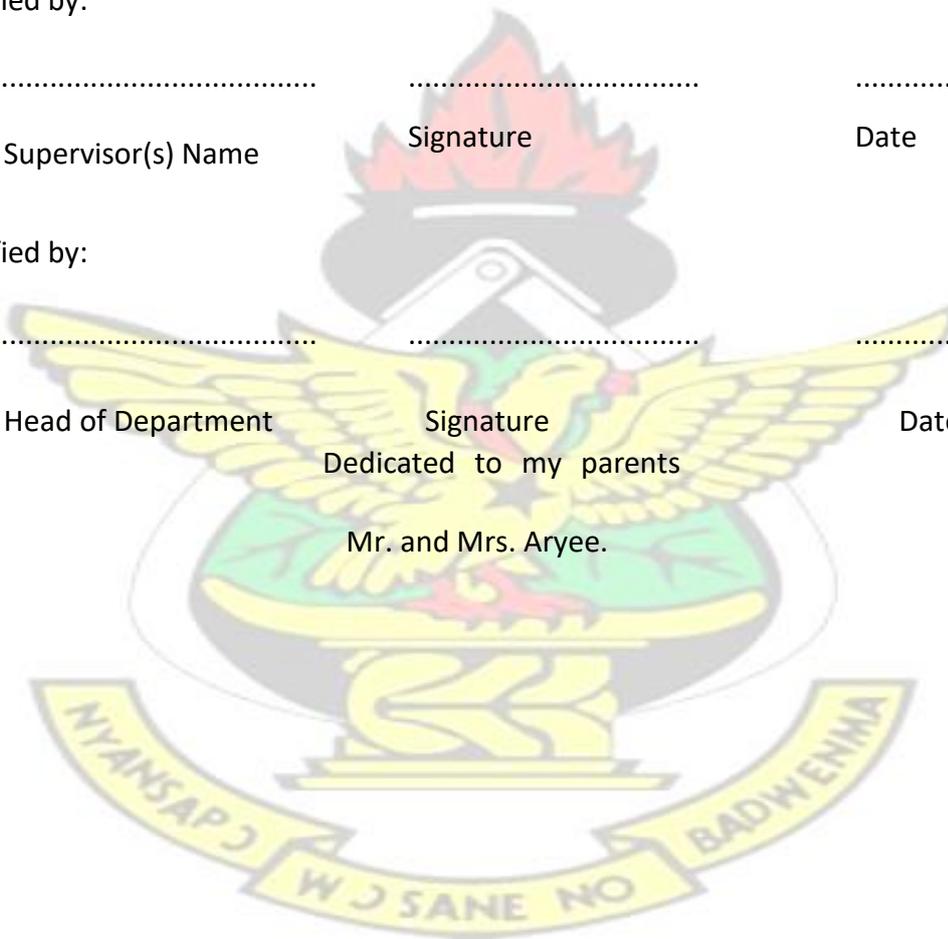
.....  
Supervisor(s) Name                      Signature                      Date

Certified by:

.....  
Head of Department                      Signature                      Date

Dedicated to my parents

Mr. and Mrs. Aryee.



# Abstract

Various sectors of the country's economy (health, energy, agriculture, planning and many others) depend on climate, and as such availability of quality climate data becomes essential for climate impact studies in these sectors. In this study, rainfall climatology database has been developed for Ghana using GMet station datasets distributed over the four agro-ecological zones and spanning a 33-year period (1980 – 2012). Datagaps within the rainfall time-series were filled by Regularized Expectation Maximization (RegEM) and homogenization of the time-series was performed by Quantile Matching Adjustments (QMadj). The homogenized datasets were then gridded at a high-spatial resolution ( $0.25^{\circ} \times 0.25^{\circ}$ ) using Minimum Surface Curvature (MSC) with tensioning parameter. Seasonal rainfall for the four agro-ecological zones have been derived based on the grids covering the entire country and this allowed a clear evidence of the migration of Inter-Tropical Discontinuity (ITD) from the South of the country to the North and back; thus, establishing a uni-modal rainfall regime over the Northern part of the country and a bi-modal rainfall regime over the Southern part of the country. Finally, Climatic Research Unit Time-Series 3.22 (CRU TS 3.22) monthly precipitation data was used to validate the gridded dataset, obtaining high Pearson's correlation co-efficients (0.5 – 0.9), low relative mean difference (0 – 0.3) and low relative root mean square error values (0 – 8). At present, a country-wide rainfall climatology has been developed from GMet rainfall time-series which will serve as a precursor for further climate impact study, in the aforementioned sectors, across the country.

# Contents

Declaration		i
Abstract		iv
Table of Content		v
List of Tables		viii
List of Figures		ix
1 Introduction	1	
1.1 Research Background	1	
1.2 Problem Statement	2	
1.3 Justification of the Study	3	
1.4 Main Objective	4	
1.4.1 Specific Objectives	4	
1.5 Organization of the thesis	4	
2 Literature review	6	
2.1 Rainfall - formation and measurement	6	
2.1.1 Rain formation	6	
2.1.2 Measurement of Rainfall	7	
2.2 Climate of Ghana	8	
2.3 Surface Climate Data	10	
2.3.1 Climate Data Reconstruction	11	
2.4 Data Homogenization	15	
3 Data and Methodology	20	
3.1 Study Area and Data Source	20	
3.1.1 Climate of the Study Area	20	
3.1.2 Data Source	21	
3.1.2.1 Observed Data	21	
3.1.2.2 Classification of the Various Meteorological Stations	23	

3.1.2.3	CRU TS Data .....	24
3.2	Methods .....	25
3.2.1	Regularized Expectation Maximization Algorithm for Missing Data Estimation .....	25
3.2.2	Homogenization .....	27
3.2.3	Gridding by Minimum Surface Curvature With a Tension Pa- rameter .....	28
3.2.4	Validation .....	29
3.2.4.1	Relative Mean Difference .....	29
3.2.4.2	Pearson’s Correlation Co-efficient .....	29
3.2.4.3	Relative Root Mean Square Error .....	29
3.3	Summary of Methodology .....	30
4	Results: analysis and discussion	32
4.1	Data Reconstruction .....	32
4.2	Rainfall Seasonality in the Agro-Ecological Zones .....	38
4.2.1	Northern Zone .....	38
4.2.2	Transition Zone .....	40
4.2.3	Forest Zone .....	42
4.2.4	Coastal Zone .....	44
4.3	Gridding .....	46
4.3.1	Monthly Rainfall Climatology .....	46
4.3.1.1	Summary of Monthly Rainfall Climatology .....	52
4.3.2	Seasonal Rainfall Climatology .....	54
4.3.3	Annual Rainfall Climatology .....	56
4.3.3.1	Summary of Seasonal & Annual Rainfall Climatology	57
4.4	Validation .....	59
5	Conclusion and recommendations	62
5.1	Conclusion .....	62
5.2	Recommendations .....	63

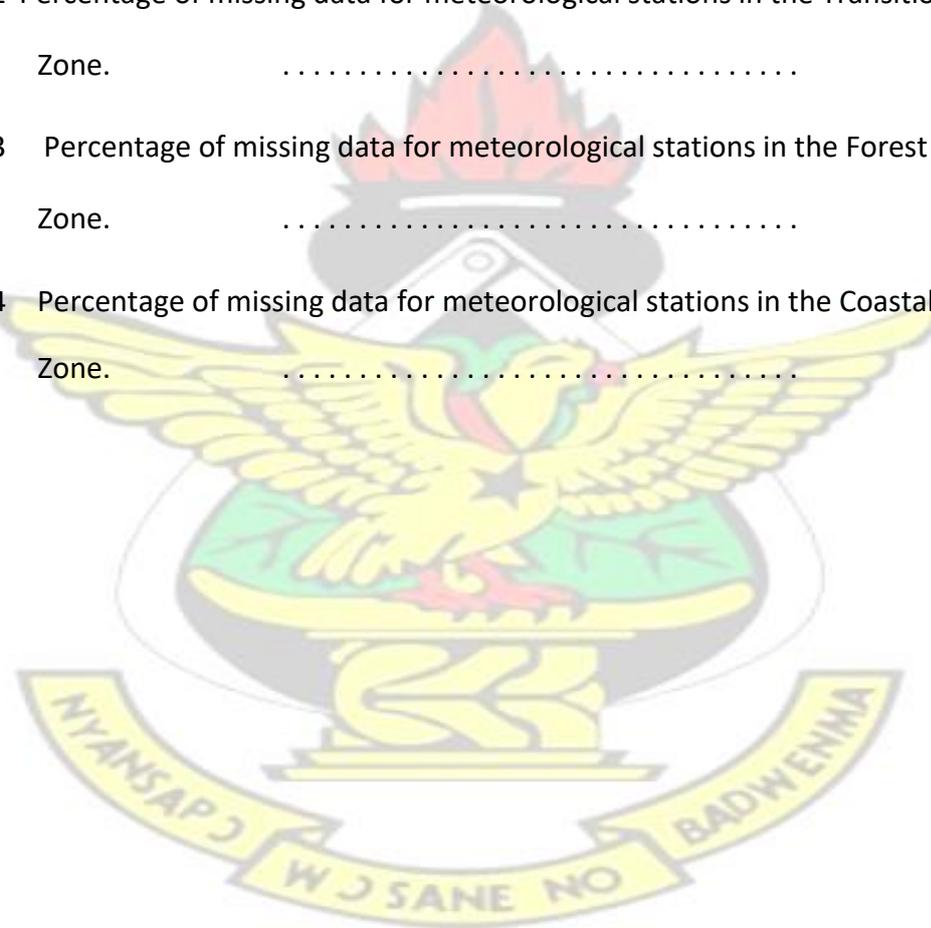
5.2.1 Recommendation for Policy .....	63	
5.2.2 Recommendation for Future Research .....	64	
Bibliography		66
A Appendix	73	
B Appendix	75	
C Appendix	88	

# KNUST



# List of Tables

4.1	Estimates of monthly rainfall total in the four agro-ecological zones. .	52
4.2	Estimates of seasonal & annual total in the four agro-ecological zones.	57
A.1	Percentage of missing data for meteorological stations in the Northern Zone. ....	73
A.2	Percentage of missing data for meteorological stations in the Transition Zone. ....	73
A.3	Percentage of missing data for meteorological stations in the Forest Zone. ....	74
A.4	Percentage of missing data for meteorological stations in the Coastal Zone. ....	74



# List of Figures

2.1	Annual Migration of ITD .....	9
3.1	Spatial distribution of meteorological stations employed in the study .	21
3.2	Percentage of missing data in each station .....	22
3.3	Percentage of missing data as a function of cumulative number of stations with data gaps. ....	23
3.4	Flowchart of datagap estimation, homogenization and gridding steps carried out in the study. ....	31
4.1	Homogenization by QMadj performed on Abetifi data .....	32
4.2	Homogenization by QMadj performed on Aburi data .....	33
4.3	Homogenization by QMadj performed on Asuansi data .....	34
4.4	Homogenization by QMadj performed on Atebubu data .....	35
4.5	Homogenization by QMadj performed on Damango data .....	36
4.6	Homogenization by QMadj performed on Winneba data .....	37
4.7	Seasonal Rainfall Patterns over selected stations in the Northern Zone.	38
4.8	Seasonal Rainfall Patterns over selected stations in the Transition Zone.	40
4.9	Seasonal Rainfall Patterns over selected stations in the Forest Zone. .	42
4.10	Seasonal Rainfall Patterns over selected stations in the Coastal Zone. .	44
4.11	Gridded Monthly Rainfall Total (Jan-Feb) .....	46
4.12	Gridded Monthly Rainfall Total (Mar-Apr) .....	47
4.13	Gridded Monthly Rainfall Total (May-Jun) .....	48
4.14	Gridded Monthly Rainfall Total (Jul-Aug) .....	49
4.15	Gridded Monthly Rainfall Total (Sep-Oct) .....	50

4.16 Gridded Monthly Rainfall Total (Nov-Dec) .....	51
4.17 Gridded Seasonal Rainfall Total (DJF - MAM) .....	54
4.18 Gridded Seasonal Rainfall Total (JJA - SON) .....	55
4.19 Annual Rainfall Total .....	56
4.20 Validation of GMet with CRU TS .....	59
4.21 RMD and Pearson's Correlation Statistical Tests .....	60
4.22 Relative Root Mean Square Error .....	61



# Acknowledgments

My utmost appreciation is to the Almighty God who made all things work together for good throughout the period of my study. I appreciate Him for the successful completion of this study.

To Dr. Leonard Kofitse Amekudzi, the supervisor of my research, I am most grateful for your awesome supervision and unrelenting support throughout the study period. You always challenged me to add something new to my knowledge-base. You introduced me to various programming languages (FORTRAN, gnuplot, SHELL SCRIPT and many more) which are now to my advantage in this field and I am so indebted to you.

This will be incomplete without acknowledging Dr. Kwasi Preko who always had an encouraging word when the going got tough. His high sense of parenthood and charisma was always an inspiration for the study. You were more as a second supervisor for this work. I am so grateful, "Doc".

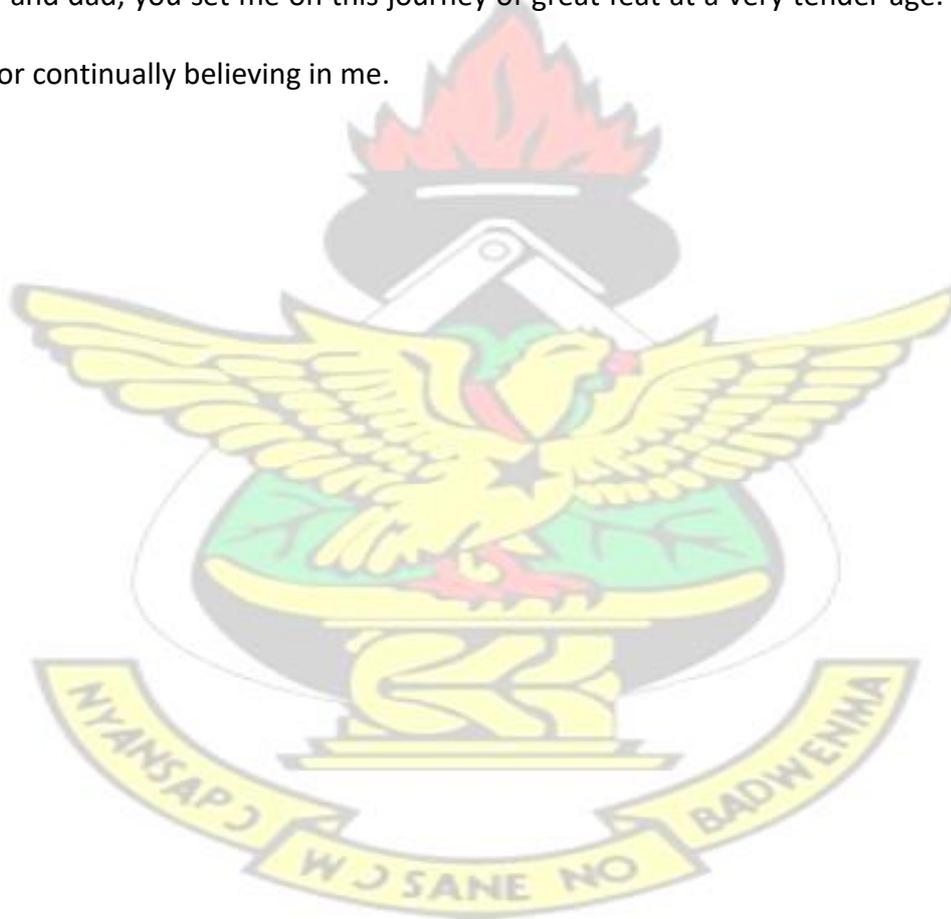
My sincere appreciation also goes to Mr. David Dotse Wemegah, Mr. Benjamin Boadi, Mr. Daniel Manu Marfo and Mr. Clement Obeng Manu who introduced me to various geophysical software packages. You all deposited volumes of knowledge in my knowledge-base.

Lecturers and fellow postgraduate students, especially those in the Meteorology and Climate Science Unit, you all challenged me through constructive criticisms and challenges throughout this study. I appreciate you all.

Also, I am so thankful to the Ghana Meteorological Agency (GMet) for allowing me use their data for this research, especially Mr. Charles Yorke and Mr. Christian Ayi.

A very big thank you to Prof. Mengistu Tsidu, G. for his support with necessary RegEM codes, Wang and Feng for the RHtestV3 package that was used for homogenization of rainfall time-series, and the Climatic Research Unit of the University of East Anglia for providing the CRU TS precipitation data.

Last but not least, for the immense contributions of my parents, I am so grateful. Mom and dad, you set me on this journey of great feat at a very tender age. Thank you for continually believing in me.



# CHAPTER 1

## Introduction

### 1.1 Research Background

Present increase in global warming and subsequent climate change has necessitated a more rigorous research geared towards finding and understanding their possible cause. According to Barnett et al. (1996), knowing both the spatial and temporal patterns of climate change over the past several centuries remains a key to assessing a possible anthropogenic impact on post-industrial climate. To this end, availability of climate data is very vital to climate change studies. Climate researches, over time, have been carried out using various climate variables. Precipitation is one climate variable, almost as important as atmospheric surface temperature, employed in providing rich evidence of climate change at different spatial scales (Mengistu Tsidu, 2012).

Such climate researches tend to rightly represent the phenomenon but a major limitation of them is the handling of gaps inherent in the dataset and homogeneity testing (Mengistu Tsidu, 2012). Manzanas et al. (2014) also makes the assertion that one major challenge of such climate researches is the lack of dense-instrumental network. Rainfall tends to be highly differential on a spatial scale across the sub-region unlike temperature. Due to this, employing a synoptic station's temperature measurements for neighbouring areas would not be too erroneous as in the case of rainfall.

Rainfall records are often incomplete because of missing rainfall data in the measured period, or insufficient rainfall stations in the study region (Feng-Wen and Chen-Wuing, 2012). It is much more expedient finding ways that will estimate rainfall data gaps to a higher degree of accuracy. A common means of resolving rainfall data gaps is by using interpolation techniques. Such interpolation techniques include Minimum Surface Curvature (Smith and Wessel, 1990), nearest neighbour (Yakowitz, 2008), global and local polynomial (Feng-Wen and Chen-Wuing, 2012), Thiessen polygons (Kopeck, 2010) and inverse distance weighting (Lu and Wong, 2008; Feng-Wen and Chen-Wuing, 2012) methods. Others include various forms of Kriging method (Price et al., 2000; Jeffrey et al., 2001; Li and Heap, 2008; Yeh et al., 2011; Mengistu Tsidu, 2012). Using appropriate interpolation techniques will aid in the development of a quality climate database that will allow for further climate-related impact studies. This work seeks to reconstruct quality and high-resolution, gridded rainfall database for Ghana.

## 1.2 Problem Statement

Climate-related study is still at its growing stage in Ghana. Paramount among other reasons for this limited study is inhomogeneity and inherent gaps within the climate data such as rainfall and temperature. These gaps within the data-time series yield difficulties in the analysis and sampling of the climate. Most rainfall studies carried out represent missing data as -99.9 or 0 mm (no rain) for convenience in analysis. This, although might not affect the result of the study holistically, but in exact terms, is not a true depiction of the variability of the rainfall parameter.

Also, very sparse meteorological station network exists countrywide and as such, not every location's rainfall can be quantified and further used to carry out quality impact studies.

Due to the insufficient manning and supervision of climatological and agrometeorological stations, data collected from these observatories may have poor quality as compared to data from synoptic stations.

Lastly, there is lack of a very organised countrywide rainfall climatological database to enhance quality climate-based impact studies.

### 1.3 Justification of the Study

With advances in recent climate studies, availability of quality climate data becomes very essential. Most sectors of the country's economy are climate change dependent, thus, climate impact studies in these sectors such as agriculture, health, energy among others will require quality climate data (Ofori-Sarpong, 2001; Singh and Ranade, 2010; Mathugama and Peiris, 2011). Reconstructing the rainfall dataset will be a very good resolve to eliminate data gaps, which are known to introduce inconsistencies in climate data-time series and as such, aid in quality climate impact studies. The study intends to develop a highly-resolved rainfall database that will aid in quality impact studies countrywide.

Reto et al. (2010) accounts that, with precipitation being the most challenging aspect of climate modeling, there is therefore a strong need for quality estimates of precipitation. The development of a quality rainfall dataset at a high resolution will

help in dealing with the challenge, since the ground-based rainfall measurements serve as input data for climate models and their simulations.

## 1.4 Main Objective

This study seeks to develop a high-resolution rainfall climatology for Ghana.

### 1.4.1 Specific Objectives

- To estimate missing values in the gauge data by regularized expectation maximization (hereafter referred to as RegEM).
- To homogenize rainfall dataset by Quantile Matching Adjustment (hereafter referred to as QMadj) Regression Fit.
- To grid homogenized datasets on spatial scales at a high resolution ( $0.25^{\circ} \times 0.25^{\circ}$ ) by Minimum Surface Curvature (hereafter referred to as MSC) Method.
- To validate generated datasets with climate data from Climate Research Unit Time Series 3.22 (CRUTS 3.22).

## 1.5 Organization of the thesis

The thesis is organized into five chapters as follows:

Chapter one presents a general overview of data reconstruction, homogenization and gridding, the motivation for the study, the significance of the study, the objectives and the thesis structure.

The first section of Chapter two focuses on the Climate of Ghana. A review of various studies conducted, with regards to climate data reconstruction, is carried out in the second section of Chapter two. The third section highlights the processes of homogenization and works carried out in that light.

Chapter three focuses on data organization and various methods employed in the study.

This Chapter holistically presents the RegEM algorithm for missing data estimation, Homogenization by QMadj Regression Fit, Gridding by Minimum Surface Curvature and a flowchart of the steps undertaken in the study.

The results are presented in the fourth chapter. Various panels showing the reconstructed and homogenized datasets have been presented and analysed in the first section. In the second section, rainfall seasonality is shown and discussed. The homogenized datasets have also been gridded on a high-resolution in the next section. Validation of generated dataset with CRU TS dataset is presented in the final section.

Chapter five comprises of conclusions based on the objectives of the study. Recommendations have also been made to individuals and institutions that the project will be beneficial to, for future research work.

## CHAPTER 2

## Literature review

### 2.1 Rainfall - formation and measurement

#### 2.1.1 Rain formation

Rainfall is condensation of water droplets from atmospheric water vapor, which becomes heavy enough to fall under gravity. It forms a dominant input for water within the hydrological cycle and affords a suitable condition for many ecosystem types, as well as water for hydroelectric power plants, crop irrigation and other domestic and industrial purposes. Rain is formed when a sharp contrast occurs between two distinct air masses which are of different physical properties such as, temperature, moisture and others. Precipitation falls from convective clouds if there is enough moisture and upward motion. Rainfall can be classified into three types.

##### Convective Rain

Convective rain occurs from convective clouds. Heating from the surface causes the air mass below to be warmer and less dense, thus rising. As it is lifted, the water vapor encounters cloud condensation nuclei (CCN) and thus condenses on them. They then accumulate and form heavy drops. Gravity acts on the heavy drops and so they fall back to the ground. As it falls, buoyancy acts on it and the force tends to split the huge droplet into smaller and finer ones. Also, the drops as they fall, hit against each other, thereby causing further disintegration. Convective rains fall as showers with rapidly changing intensity.

## Orographic Rain

Orographic rain forms on the windward side of mountains and is caused by the rising air motion of a large-scale flow of moist air across the mountain, resulting in adiabatic cooling and condensation. A more moist climate usually prevails on the windward side of a mountain than on the leeward or downwind side. Moisture is removed by orographic lift, leaving drier air on the descending and generally warming, leeward side where a rain shadow is observed (Pidwirny, 2008).

## Frontal Rain

Stratiform and dynamic precipitation occur as a consequence of slow ascent of air in synoptic systems, such as in the vicinity of cold fronts and near and poleward of surface warm fronts. These frontal rains occur when two contrasting air masses (cold-warm air masses or cold-colder air masses) tend to displace each other, thereby causing an ascent with subsequent rainfall associated with the ascent.

### 2.1.2 Measurement of Rainfall

Rain is measured in units of depth per unit time (for example  $mm\ hr^{-1}$ ,  $inches\ hr^{-1}$ ), as a representation of the depth of rain water that would be accumulated on a flat, horizontal and impermeable surface during a given period (Cerveny and Balling, 1998).

The standard raingauge is used in measuring rainfall. The raingauge is kept in a 100 mm (4 in) plastic and 200-mm (8-in) metal varieties. The inner cylinder is filled by 25

mm (0.98 in) of rain, with overflow flowing into the outer cylinder. Other types of gauges include the tipping bucket rain gauge, and the weighing rain gauge.

For remote locations, in the advent of rigorous climatic studies, remote sensing provides a better alternative for rainfall data collection and analysis. The weather radar and rain-observing satellites are also used in order to assess the amount of precipitation over remote areas. These rainfall estimates also compliment surface station data which can be used for calibration. These remote sensing methods also better represent spatial distribution of rainfall than the former approach.

## 2.2 Climate of Ghana

Interaction of the West African Monsoon (WAM) and the Inter-Tropical Discontinuity (ITD) dominates the climate of Ghana (Stanturf et al., 2011). The ITD is a region of calm winds separating the north-easterly (NE) and south-easterly (SE) trade winds. This region is important for African agriculture, because the rising air and water vapor caused by the warmth of the sun at that region leads to formation of clouds and rainfall (Hastenrath, 1995). The ITD migrates all year round, reaching its northernmost apex during the northern hemisphere summer and its southernmost apex during the northern hemisphere winter. The migration of the ITD downwards (southwards) brings the dry, dust-laden Trade Winds (Harmattan) countrywide and its migration upwards (northwards), brings in moisture-laden Monsoon winds countrywide as shown in Figure 2.1.

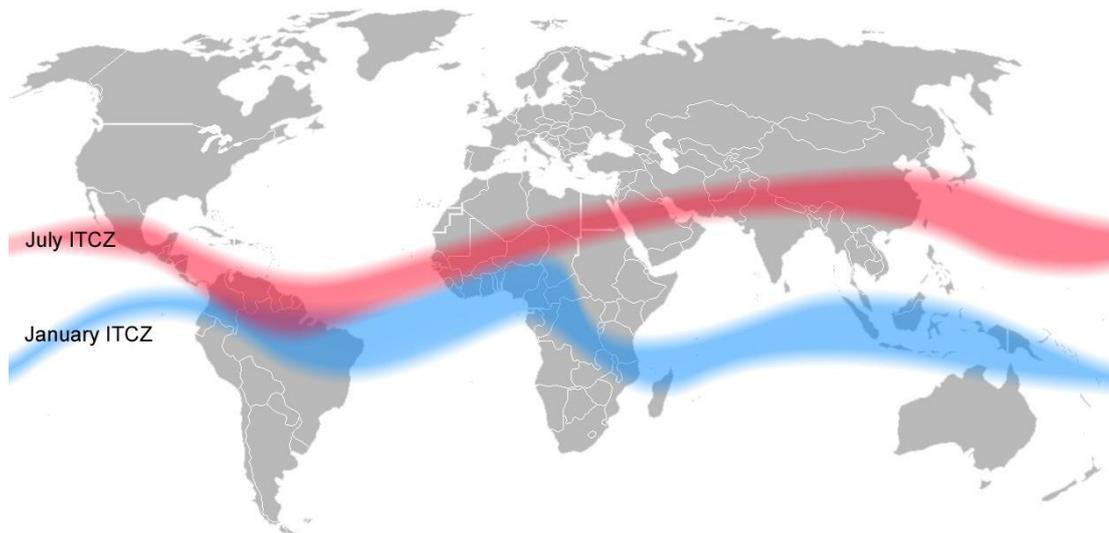


Figure 2.1: Annual migration of the Inter-Tropical Discontinuity (ITD). The red fill indicates the position of the ITD in July and the blue shows its position in January (accessed from [https://en.wikipedia.org/wiki/Intertropical\\_Convergence\\_Zone](https://en.wikipedia.org/wiki/Intertropical_Convergence_Zone)).

The movement of the ITD is very variable within the overall repeating pattern, thereby causing variations in rainfall patterns. There are two distinct wet seasons (bimodal rainfall pattern) in Southern Ghana, but Northern Ghana has only one (unimodal rainfall pattern). The ITD position does not move gradually over several months, but there are many processes that impact the ITD at various time-scales, from daily low-level jets (Flamant et al., 2009) to multi-day pulsations with cycles of about 5 days (Couvreur et al., 2010). These pulsations are associated with increased meridional low-level winds and bring additional moisture to the north. The diurnal cycle of the ITD is one of the atmospheric processes that play a crucial role in the WAM system. With the associated (mainly nocturnal) low-level winds, the diurnal variation of the ITD position has been recognized in several previous studies as a key factor in the northward transport of moisture (Lothon et al., 2008).

Studies have also shown that the climate of Ghana is getting warmer. In a recent study by Stanturf et al. (2011), the authors identified temperature data and patterns to be depicting a warming climate in Ghana with the drier northern area warming more rapidly than southern Ghana. Their study show a 1.0 °C rise in mean annual temperature since 1960 as well as, an increase in 'Hot' days and nights.

## 2.3 Surface Climate Data

Climate data provides great depth of information about the atmosphere that impacts almost all aspects of human life (Peterson et al., 1998). Global change research and impact studies are highly dependent on the description of the mean state and variability of recent climate (New et al., 1999). As such, continuous and quality climate datasets are precursors for excellent climate-impact studies. An example is the indication by Segond et al. (2007) that high spatial and temporal rainfall resolutions are required for urban drainage and urban flood modeling applications.

In another study, Brohan et al. (2006) indicated that impact models for determining the possible consequences of climate change often require continuous and evenly spaced data. Historically, surface climate data which describe variability in space and time (Easterling, 1997), have had incomplete spatiotemporal coverage. Until recent times, the instrumental record of such climate variables have been quite poor and full of gaps. This has necessitated several works of surface climate data reconstruction and gridding by various scientists with varying interpolation methods and schemes. Such interpolated datasets allow best estimates of climate variables at

locations away from observing stations, thereby allowing studies of local climate in data-sparse regions (Atkinson and Lloyd, 1998).

### 2.3.1 Climate Data Reconstruction

Various interpolation methods have been applied widely for data reconstruction in many disciplines. Jeffrey et al. (2001) constructed a comprehensive archive of Australian rainfall and climate data from ground-based observational data using spatial interpolation algorithms. In their study, ordinary kriging was used for interpolation of rainfall and thin plate smoothing spline for interpolation of the other climate variables. In another study, high-resolution gridded dataset for surface climate variables (precipitation, temperature) for the period 1950 - 2006 over Europe was developed by Haylock et al. (2008). The developed dataset improved on earlier products in spatial resolution and extent, time period, number of contributing stations, and attention to finding the most appropriate method for spatial interpolation of climate observations. To enable direct comparison with Regional Climate Models (RCMs), each dataset was designed to provide the best estimate of grid box averages rather than point values. Haylock et al. (2008) interpolated monthly precipitation totals and monthly mean temperature using three-dimensional thin-plate splines, then interpolated the daily anomalies using indicator and universal kriging, then combined the monthly and daily estimates. The authors quantified uncertainty of interpolation by providing daily standard errors for every grid square. The daily uncertainty averaged across the entire region was shown to be largely dependent on the season and number of contributing observations.

Feng-Wen and Chen-Wuing (2012) also utilized inverse distance weighting (IDW) method to estimate the rainfall distribution in the middle of Taiwan. The authors evaluated the relationship between interpolation accuracy and two critical parameters of IDW: power ( $\alpha$  value), and a radius of influence (search radius). To obtain optimal interpolation data of rainfall, the value of the radius of influence, and the control parameter- $\alpha$  were determined by root mean squared error. Per their study, the optimal parameters for IDW in interpolating rainfall data have a radius of influence up to 30 km in most cases.

Li and Heap (2011) also analysed the performance of some methods/sub-methods. The impacts of sample density, data variation and sampling design on the estimations the methods were quantified. The authors identified inverse distance weighting (IDW), ordinary kriging (OK), and ordinary co-kriging (OCK) as the most frequently used methods. According to Li and Heap (2011), data variation is a dominant impact factor and has significant effects on the performance of the methods. Data variation inversely relates the accuracy of all methods, while the magnitude of performance is method dependent. Irregular-spaced sampling design might improve the accuracy of estimation and the effect of sampling density on the performance of the methods is insignificant. Due to the deterministic nature of IDW, it remains a suitable method of spatial interpolation for estimation of probable rainfall at a reference station, provided the neighbouring stations have high quality rainfall dataset. Imputation by the IDW scheme, in the study area, will be insufficient due to lack of dense network of meteorological stations and also due to the fact that, most agrometeorological and climatological stations have sparse data.

Canonical correlation analysis (CCA) was also tested for paleoclimate field reconstructions of pseudoproxy experiments assembled from the millennial integration of the National Center for Atmospheric Research Community Climate System Model by Smerdon et al. (2010). The authors presented a method for selecting the order of the CCA model. Their results suggested that the method can resolve multiple (3-13) climatic patterns given the estimated proxy observational network and the observational uncertainty. Comparison of CCA reconstructions to derivations of regularized expectation maximization method using ridge regression regularization (RegEM–Ridge) was made. The authors found that CCA and RegEM–Ridge yielded similar skill patterns, characterized by high correlation regions collocated with dense pseudoproxy sampling areas. The two schemes generated reconstructed datasets characterized by spatially variable warm biases and variance losses, particularly at high pseudoproxy noise levels. RegEM–Ridge in particular is subject to significantly larger variance losses than CCA, even though the spatial correlation patterns of the two methods are comparable. Their results showed the importance of testing the performance of methods that target spatial climate patterns during the last several millennia and are indications that results of available climate field reconstructions should be carefully interpreted.

Bo et al. (2009) also reiterated the need to compare recent changes with past variability and as such, presents a study on comparison of the properties of some reconstruction methods. In the authors' study, a systematic study of the properties of reconstruction methods (both direct hemispheric-mean reconstructions and field reconstructions, including reconstructions based on canonical regression and RegEM algorithms) is presented. The study was based on temperature fields where the

target of the reconstructions is known, and with emphasis on how well the reconstructions reproduce low-frequency variability, biases, and trends. Bo et al. (2009) employed a climate simulation from an ocean-atmosphere general circulation model of the period 1500 – 1999, including both natural and anthropogenic forcings. The reconstructions, however, included large elements of stochasticity, and reconstructions of a large ensemble of realistic temperature fields were needed to draw robust statistical inferences. The authors developed a technique to generate surrogate fields with similar spatio-temporal characteristics as the original surface temperature field from the climate model. The authors found that all reconstruction methods contained large elements of stochasticity, and it was not possible to compare them and draw conclusions from a single or a few realizations. This, they further explained that, very different results can be obtained using the same reconstruction method on different surrogate fields and this could be a possible explanation for some of the recently published divergent results.

Schneider (2001) proposed that the RegEM scheme is applicable to climate datasets, in which the number of variables typically exceeds the sample size. According to Schneider (2001), estimation of the mean and covariance matrix of incomplete dataset and filling in missing values is a nonlinear problem, which demands an iterative solution. The author considers the expectation maximization (EM) algorithm for Gaussian data, as departure for the development of RegEM algorithm.

Schneider (2001) cites that the RegEM algorithm can estimate, and exploit for the imputation of missing values, both synchronic and diachronic covariance matrices, which may contain information on spatial covariability, stationary temporal covariability, or cyclostationary temporal covariability. A test of the RegEM algorithm

with simulated surface temperature data, by the author, showed that the scheme can be applied to typical sets of climate data and that it leads to more accurate estimates of the missing values than a conventional noniterative imputation technique.

Mengistu Tsidu (2012), in his recent study, also addressed the concern regarding the extent and quality of historical climate data for Ethiopia. The author employs rainfall records of selected gauge stations over Ethiopia for the 1978-2007 period in an analysis that involved homogenization, reconstruction, and gridding onto a regular  $0.5^\circ \times 0.5^\circ$  resolution grid. Inhomogeneity was detected and adjusted based on quantile matching. The RegEM and multichannel singular spectrum analysis (MSSA) algorithms were also used for missing value imputation. The latter was determined to have quite a slim advantage. Mengistu Tsidu (2012) further used ordinary kriging to create a set of gridded monthly rainfall data. An assessment of spatio-temporal coherence of the dataset was performed with harmonic analysis (HA), self-organizing maps (SOMs), and intercomparison with global datasets which included Global Precipitation Climatology Project (GPCP), Climate Prediction Center (CPC) Merged Analysis of Precipitation (CMAP), Climatic Research Unit time series version 3 (CRUTS3.0), Tropical Rainfall Measuring Mission (TRMM), and interim ECMWF Re-Analysis (ERA-Interim) rainfall. Correlation values of the datasets typically ranged from 0.52 to 0.95 over sparse to dense rain gauge regions.

In another study, Smith and Wessel (1990) used minimum curvature to grid geological datasets. The authors identified that minimum curvature surfaces may have large oscillations and extraneous inflection points which make them unsuitable for

gridding in many of the applications where they are commonly used. As a resolve, a tension is added to the elastic-plate flexure equation. Thus, solutions under tension require no more computational effort than minimum-curvature solutions, and any algorithm which can solve the minimum-curvature equations can solve the more general system. The authors further included this in the development of their Generic Mapping Tools (GMT) programming language.

## 2.4 Data Homogenization

The level of consistency in the trend of climate data time-series is often referred to as data homogeneity. Data homogenization is a two-phase technique. The first phase deals with the detection of the inhomogeneities in the data whereas the second phase has to do with the adjustment of the data in order to remove the inhomogeneities.

According to Blair (2012), the detection of inhomogeneities in a dataset is a wellexplored problem, both in climate science and in statistics. Costa and Soares (2009) and Mengistu Tsidu (2012) further classified data homogenization into two sets: absolute and relative homogenization. Relative homogenization deals with augmenting the inhomogeneities in the station using data from neighbouring stations whereas absolute homogenization employs statistical approaches to identify changepoints within the data from the station and augment based on the degree of variations within the detected trends. In further explaining some possible causes of inhomogeneity, Blair (2012) asserted that most climate instrumental networks change over time primarily due to the importance of a fixed network for monitoring

climate variability and change being less appreciated in the past than in more recent times. In addition, various socioeconomic considerations, such as changes in demographics and infrastructure, affect the ability to maintain a fixed station network over a long period of time.

Inhomogeneities, as discussed by Menne and Williams Jr. (2008) and Lei et al. (2014), can also be induced by an alteration in instrumentation or observation practice. The relocation, replacement, or recalibration of an instrument, for example, can lead to an abrupt shift in time-ordered observations that is unrelated to any real change in climate. Various homogenization studies have been carried out using climate data. In a study by Menne and Williams Jr. (2008), the authors described an automated homogenization algorithm based on the pairwise comparison of monthly temperature data. The algorithm works by developing pairwise difference series between monthly temperature values from a network of observing stations. Undocumented shifts and stations responsible for the shifts are then sampled by evaluating each difference series. The algorithm also makes use of available station metadata to improve the identification of artificial shifts in temperature data. In their study, when magnitude of a shift linked to a particular station was reliably estimated, an adjustment was made for the target series. Pairwise homogenization algorithm was seen to be robust and efficient at detecting undocumented step changes under varying scenarios with step- and trend-type inhomogeneities. The approach has been shown to yield lower false-alarm rate for undocumented changepoint detection relative to the more common approach (use of a reference series).

In another study, Lei et al. (2014) performed relative homogenization on temperature data of Huairou station in Beijing for a 49-year period by applying data of two nearby stations. The authors estimated and compared the linear trends of the original and adjusted temperature series and concluded that the significant discontinuities in their data series was linked to relocation of the station from downtown to suburb in 1996. The homogenized data was then used to study the change in climate, which further revealed that inhomogeneous data can lead to a significant overestimate of rising trends of climate variables, and this necessitates a careful evaluation and adjustment for urban biases before the data is used to analyse local and regional climate change.

Peterson et al. (1998) reviewed the methods and techniques developed for homogeneity adjustments and described many different approaches involved in adjusting in situ climate data. Some findings are made in the researchers' study and these include the effect of inhomogeneities on temperature being often opposite during winter and summer. Therefore, evaluating annual mean temperatures, if the series are intended for studies of seasonal trends, is not recommended. Evaluation can be done on single months, but the noise level is then increased. The authors further proposed the use of metadata alongside the annual mean homogeneities for seasonal trend studies of temperature. For precipitation, the authors suggested the use of both seasonal and annual test results in conjunction with metadata, to state if and when there is an inhomogeneity.

Noise levels are larger for seasonal testing, therefore, many inhomogeneities are most easily detected using annual values. The authors admonish that series tested and adjusted only on annual basis should not be used for seasonal trend studies. In

confirming the work of Hanssen-Bauer et al. (1997), the authors stated that the adjustment values for precipitation may vary throughout the year.

Time-consistent homogeneity, in high-resolution investigations, can rely on strong correlations among series, which must meet the quality standard in terms of completeness (Eccel et al., 2012). The authors pre-processed fifty-nine daily precipitation and temperature series of 50 years from Trentino, northern Italy, for climatic analysis. The authors first filled the data gaps using geostatistical correlations on both horizontal and vertical domains and then developed an algorithm to reduce inhomogeneity owing to the systematic snowfall underestimation of rain gauges. Finally, the processing protocol to factor any source of undocumented discontinuities in series, was described. Homogenization was performed by the F-test and T-test as designed by Wang et al. (2010). Pre-processing showed straightforward results; homogenization increases the strength of the climatic signal and reduces the scattering of time trends, assessed over a few decades, by a factor of 2.

Lanzante and Klein (2002) in another study, performed temporal homogenization on radiosonde data to determine long-term trends. The homogeneity methods were applied to a network of 87 stations using monthly temperature data, spanning a 50 year period and at mandatory pressure levels. The first phase dealt with identification of artificial inhomogeneities through visual examination of graphical and textual materials, including temperature time series, transformations of the temperature data, and independent indicators of climate variability, as well as station history metadata. Modification, in the form of data adjustment or data deletion, was then applied to correct each problem encountered. The authors performed various analyses, particularly trend, in a second part of the study using the homogenized

datasets. Application of the procedures to the station network revealed a number of systematic problems. Some stations show a tendency for episodic drops in temperature that produce spurious downward trends. Stations from Africa and neighboring regions were found to be the most problematic with the character of the interannual variability being unreliable. Temporal variations in observation time were also found to lead to inhomogeneities as serious as the worst instrument-related problems.

Costa and Soares (2009) also reviewed the characteristics of several widely used procedures and discusses the potential advantages of geostatistical techniques. The geostatistical simulation approach is applied to precipitation data from 66 stations located in the southern region of Portugal for a 22-year period. The results from this procedure were then compared with those from three well established statistical tests: the Standard normal homogeneity test (SNHT) for a single break, the Buishand range test, and the Pettit test. The case study provided promising results that open new research perspectives on the homogenization of climate time series.

The greater concern at the moment is that such studies have not been carried out on GMet datasets, hence this study becomes essential towards addressing the challenge of datagaps and inhomogenities within rainfall time-series from meteorological stations across the country, through the application of interpolation and homogenization techniques.

## CHAPTER 3

## Data and Methodology

### 3.1 Study Area and Data Source

#### 3.1.1 Climate of the Study Area

Two main seasons dominate the country due to its location in the tropics. These are the dry (Harmattan) and the wet (Monsoon) seasons. The Harmattan spans the months of November to March, with the remaining months being the rainy period. Rainfall in this region is mainly associated with mesoscale convective systems and controlled by the advection of moisture from the Gulf of Guinea in the low level atmosphere (Sultan and Janicot, 2003). This system is usually referred to as the West Africa Monsoon (WAM) and it is driven by energy and temperature gradient between the Gulf of Guinea and the Sahara. The maritime tropical air mass which originates from the Atlantic Ocean is moisture laden and converges with the dry north-east continental tropical air mass usually along the Inter-Tropical Discontinuity (ITD) [(Amekudzi et al. (2015) and references therein)]. The north and south movement of the ITD regulates the dynamics of the seasons in Ghana (Biederlack and Rivers (2009); Manzanas et al. (2014)). There is a change in the rainfall regime from the north to the south of the country. A uni-modal regime dominates the northern part and a bi-modal regime dominates the southern parts of the country [(Yamba (2010) and references therein)].

### 3.1.2 Data Source

#### 3.1.2.1 Observed Data

Monthly rainfall datasets for various meteorological (synoptic, agrometeorological and climatological) stations over Ghana were acquired from the Ghana Meteorological Agency (GMet) database. Almost all the datasets had inherent data-gaps. A criteria was set for sorting the datasets. RegEM is less well-suited to deal with datasets that exhibit relatively long continuous gaps (Mengistu Tsidu, 2012) so rainfall data with 10% or more gaps were discarded, as well as data with continuous gaps. A total of 113 stations that met this criteria were employed in this study. The sampled stations are spread over the four agro-ecological zones (Northern, Transition, Forest, Coastal) as shown in Figure 3.1, with a dense station-network in the Forest zone.

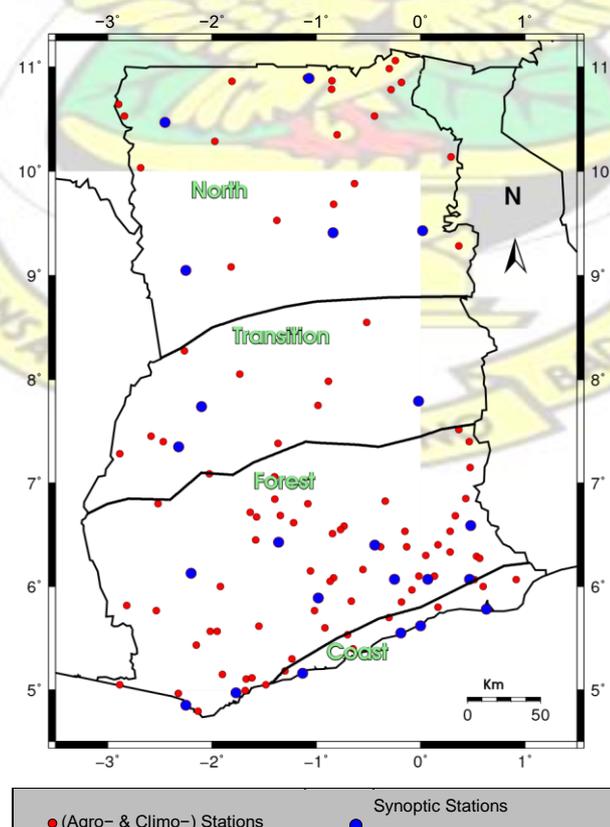


Figure 3.1: Spatial distribution of 113 meteorological (synoptic, agrometeorological and climatological) stations employed in the study. The percentage of data gaps in each station's data is illustrated in Figure 3.2 and Figure 3.3 shows the percentage of missing data as a function of cumulative number of stations with data gaps. Data were collected for a 33-year period, from 1980 to 2012.

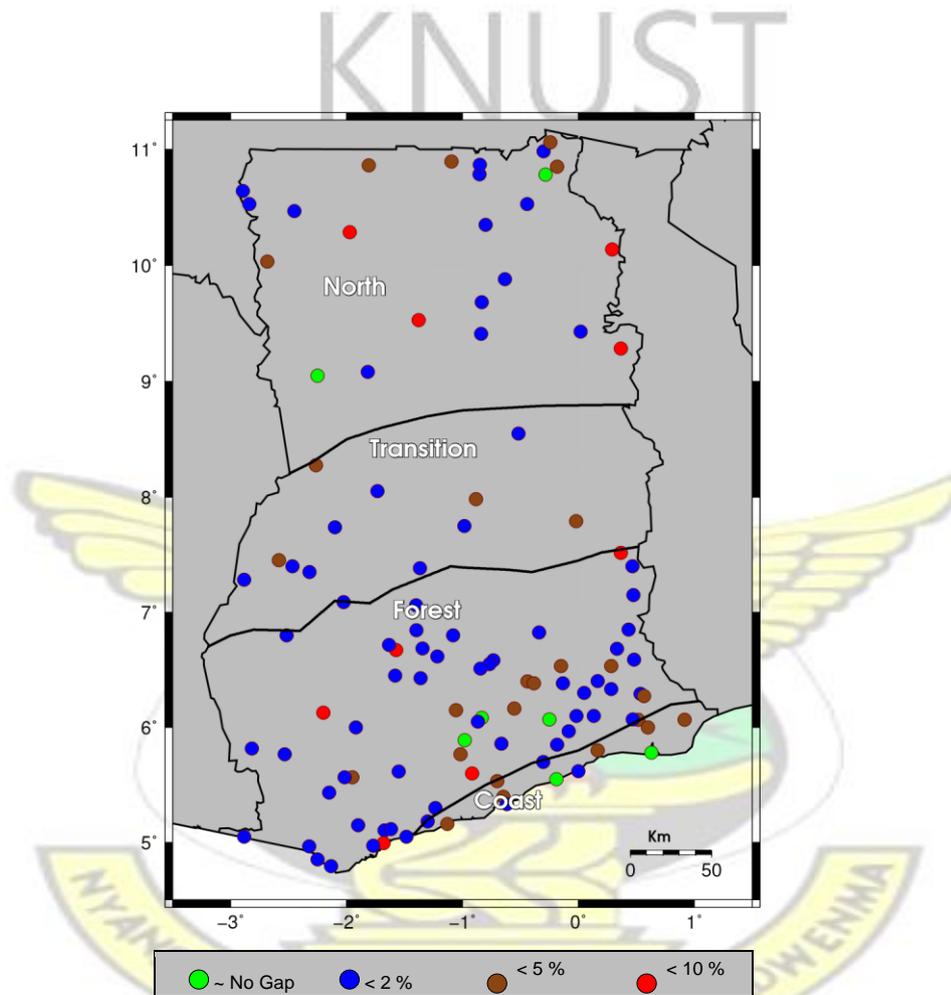


Figure 3.2: Percentage of missing data in each of the 113 station. Green circles show stations with approximately no data gap, blue circles show stations with less than 2% gap, brown circles show stations with less than 5 % gaps and the red circles show stations with less than 10 % gaps.

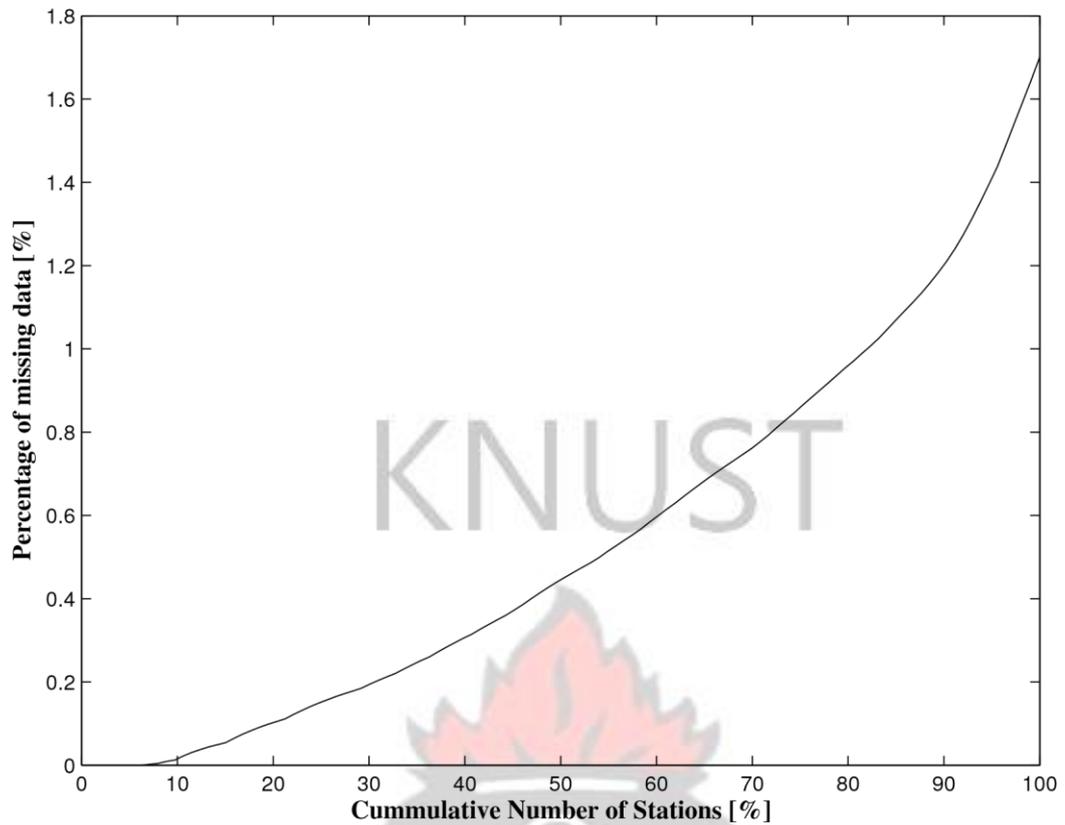


Figure 3.3: Percentage of missing data as a function of cummulative number of stations with data gaps.

### 3.1.2.2 Classification of the Various Meteorological Stations

#### 3.1.2.2a Synoptic Station

The synoptic stations are manned, per WMO standards, to observe majority of the climate variables. Observations at the synoptic stations can be classified into two: the minor and major observations. Minor observations are carried out at every 3 hour interval (i.e 0000, 0300, 0600, 0900, 1200, 1500, 1800, 2100 Greenwich Mean Time) and the major observations are carried out at every 6 hour interval (i.e 0000, 0600, 1200, 1800 Greenwich Mean Time). Rainfall, however, is recorded at 0900 GMT in the synoptic station.

#### 3.1.2.2b Climatological Station

These stations are primarily manned for meteorological data archiving. Observations are taken at every hour and they carry out lesser observations than the synoptic stations.

### 3.1.2.2c Agrometeorological Station

Agrometeorological stations are manned to carry out agricultural-relevant meteorological observations. Some of these observations include rainfall, evapotranspiration, soil matrix potential, and others. Lesser observations are taken in the agrometeorological station than the synoptic station.

### 3.1.2.3 CRU TS Data

CRU TS datasets are monthly variation in climate over the last century produced by the Climatic Research Unit (CRU) at the University of East Anglia. They are gridded datasets at a high-resolution of  $0.5^{\circ} \times 0.5^{\circ}$  with precipitation datasets provided by more than 4000 weather stations distributed around the world.

CRU also generates data of other climate variables such as diurnal temperature range, daily mean temperature, vapour pressure, wet day frequency, monthly average daily maximum temperature, wet day frequency and cloud cover, which allow for the study of climate variability. The latest data time series (CRU TS 3.22) are generated for the period 1901 – 2013. The CRU TS data are monthly gridded fields based on daily values -hence its precipitation storage files contain monthly total mean values.

## 3.2 Methods

### 3.2.1 Regularized Expectation Maximization Algorithm for Missing Data Estimation

Regularized Expectation Maximization (RegEM) algorithm described by Schneider (2001), is used in reconstructing the rainfall data-time series. RegEM works on the principles outlined in equations 3.1 - 3.7.

An unobservable population parameter that maximizes the log-likelihood function is extensively estimated using the maximum likelihood method described in equation

$$3.1 \quad L(\Theta; \mathcal{X}) = \sum_{i=1}^n \log P(\mathcal{X}_i | \Theta) \quad (3.1)$$

where, observations,  $X = \{x_i | i = 1, \dots, n\}$  are independently drawn from the distribution  $P(x)$  parameterized by  $\Theta$ . The Expectation-Maximization (EM) algorithm iteratively computes the maximum-likelihood estimates when the observations are an incomplete data with existence of additional but missing data  $Y = \{Y_i | i = 1, \dots, n\}$  corresponding to  $X$ . The EM algorithm maximizes the log-likelihood of the incomplete data by exploiting the relationship between the complete and the incomplete data.

Two processes (E-step and M-step) are involved in each iteration ( $t = 1, 2, 3, \dots, n$ ). In the E-step, the expectation of log-likelihood of the complete data, based on the incomplete data and the current parameter ( $\Theta_{(t)}$ ) is determined from equation 3.2

$$Q(\theta|\theta_{(t)}) = E(\log P(X,Y|\theta)|X,\theta_{(t)}) \quad (3.2)$$

The algorithm determines a new parameter by maximizing Q, in the M-step (see equa-

tion 3.3).

$$\theta_{(t+1)} = \underset{\theta}{\operatorname{argmax}} Q(\theta|\theta_{(t)}) \quad (3.3)$$

Each iteration is guaranteed to increase the likelihood, and finally the algorithm converges to a local maximum of the likelihood function. The missing data, Y, has strong effect on performance of the EM algorithm since the optimal parameter  $\theta^*$  is obtained by maximizing  $E(\log P(X,Y|\theta))$ . With the incomplete-data, X, there exists several different ways to define Y. The choice of a suitable Y to make the solution more plausible is an unaddressed question in EM algorithm because the likelihood function reflects no influence of the missing data. The information about one object contained in another object can be measured from equation 3.4

$$H(X|Y) = \sum_y P(y) H(X|Y=y) = - \sum_x \sum_y P(x,y) \log p(x|y) \quad (3.4)$$

The relationship between entropy (H) and information (I) in equation 3.5

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) \quad (3.5)$$

demonstrates that the mutual information measures the amount of information that one random variable contains about another one.

The conditional entropy  $H(Y|X)$  measures the uncertainty of Y on the average when X is known. In fact,  $H(Y|X) = 0$  if and only if Y is a function of X. Thus,  $H(Y|X)$  is expected to be small if the observations X and the missing data Y have a strong correlation.

To optimize the regularized likelihood, there is the need for a modification in the Mstep of the EM algorithm, which is given as

$$\Theta_{(t+1)} = \underset{\Theta}{\operatorname{argmax}} Q_e(\Theta | \Theta_{(t)}) \quad (3.6)$$

where

$$Q_e(\Theta | \Theta_{(t)}) = Q(\Theta | \Theta_{(t)}) - \gamma H(Y | X; \Theta) \quad (3.7)$$

This modified algorithm is the Regularized Expectation Maximization algorithm (see equation 3.6).

KNUST

### 3.2.2 Homogenization

There are two types of homogenization (absolute and relative) procedures. Absolute homogenization considers information in the particular station's climate data-time series alone whereas relative homogenization uses data from other neighbouring observatories to approximate the observatory of interest, and thus is more reliable, provided that neighbouring stations have high quality datasets (Mengistu Tsidu, 2012).

Due to the sparse datasets from neighbouring observatories, absolute homogenization was carried out in this study, using the RHtests V4 package described by Wang et al. (2010). The homogenization is based on a linear regression analysis of the reference series as shown in equation 3.8.

$$Y_i = X_i - \beta(t_i) \quad (3.8)$$

where  $Y_i$  is the homogenized dataset,  $X_i$  is the reference series observed at time ( $t_i$ ) and  $\beta(t_i)$  is the common linear trend component.

Homogenized datasets are based on quantile matching adjustments.

### 3.2.3 Gridding by Minimum Surface Curvature With a Tension

#### Parameter

Minimum Surface Curvature (MSC) is a surface interpolation method which is analogous to a thin, linearly-elastic plate moving through each of the data values with minimal amount of bending. The method interpolates the data with a surface having continuous second derivatives and minimal total squared curvature (Smith and Wessel, 1990). MSC generates smoothest possible surface but has been shown not to be an exact interpolator, implying that the data is not always honored.

A better resolve is the introduction of a tension parameter ( $T$ ) to the algorithm. The general algorithm is a numerical solution of modified biharmonic differential equation given in equation 3.9 as

$$(1 - T)\nabla^4 f(x,y) - (T)\nabla^2 f(x,y) = 0 \quad (3.9)$$

with three boundary conditions in equations 3.10, 3.11 and 3.12.

$$(1 - T)\frac{\partial^2 f}{\partial n^2} + T\frac{\partial f}{\partial n} = 0 \quad (3.10)$$

$$\frac{\partial(\nabla^2 f)}{\partial n} = 0 \quad \text{on the edges} \quad (3.11)$$

$$\frac{\partial^2 f}{\partial x \partial y} = 0 \quad \text{at the corners} \quad (3.12)$$

where  $T$  is the tension parameter for the boundary which varies from 0 to 1,  $\nabla^2$  is the Laplacian operator,  $\nabla^4$  is the biharmonic operator and  $n$  is the boundary normal.

### 3.2.4 Validation

Validation in this study helps in the quantification of the disparity or agreement between the gridded GMet data and CRU TS data. In the effort of such quantification, three statistical measures (Relative mean difference, Pearson's correlation coefficient and Relative root mean square error), described by Amekudzi et al. (2007) and Quansah et al. (2014) have been used. These show the level of agreement between the two datasets or the deviation of the reconstructed from the CRU TS data.

#### 3.2.4.1 Relative Mean Difference

$$\text{Mean Difference (M.D)} = \frac{1}{N^2} \sum_{i=1}^N |x_i - y_i| \quad (3.13)$$

$$R.M.D = \frac{M.D}{\text{arithmetic mean } (\mu_x)} \quad (3.14)$$

#### 3.2.4.2 Pearson's Correlation Co-efficient

$$r_{xy} = \frac{\sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y)}{(N - 1) S_x S_y} \quad (3.15)$$

#### 3.2.4.3 Relative Root Mean Square Error

$$RRMSE = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (x - y)^2}}{\mu_x} \quad (3.16)$$

where  $x$  is the reconstructed dataset,  $y$  is the satellite data,  $\mu$  is the arithmetic mean,  $S$  is the standard deviation,  $r_{xy}$  is the Pearson Correlation Coefficient ( $r \in -1,1$ ),  $RMD$  is the Relative Mean Difference and  $RRMSE$  is the Relative Root Mean Square Error.

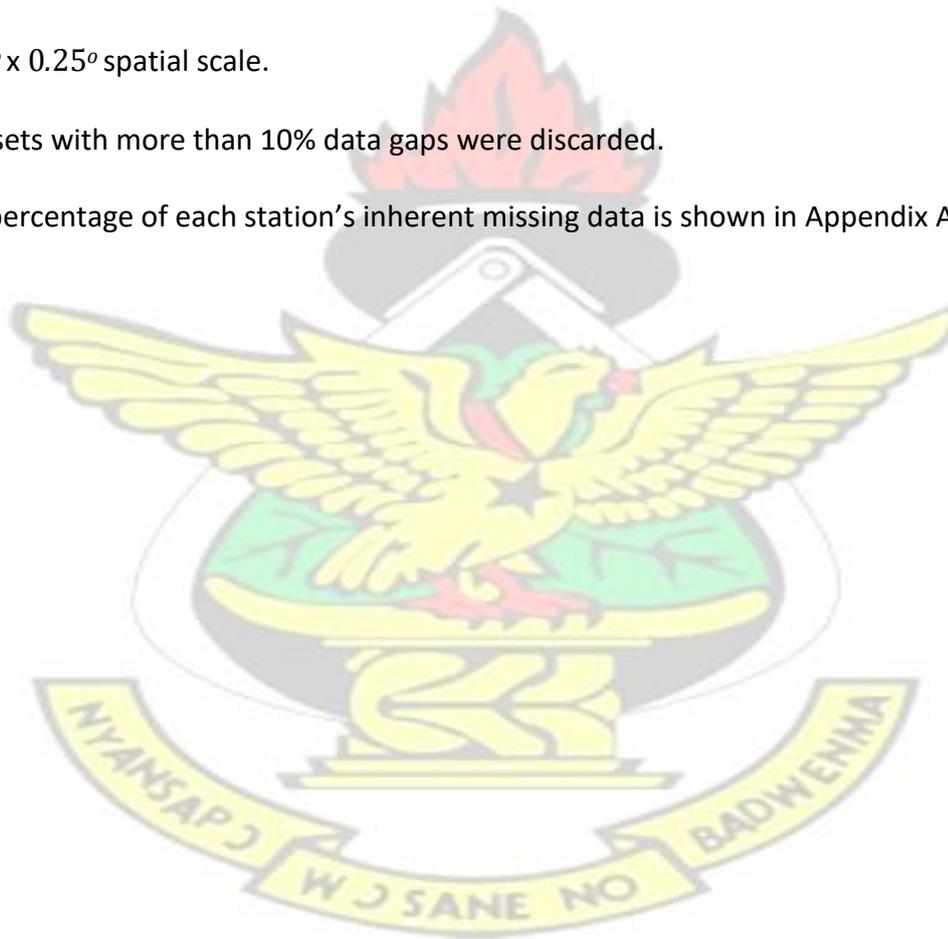
### 3.3 Summary of Methodology

The steps for reconstruction (missing data estimation, homogenization and gridding) of rainfall datasets have been summarised in a flowchart in Figure 3.4.

With less than 10% missing data, the gaps were estimated using RegEM and passed on for homogenization. With no data gap present, data was passed on for homogenization. If changepoints were detected during homogenization stage, datasets were augmented using QMadj. Homogenized data was then gridded on  $0.25^{\circ} \times 0.25^{\circ}$  spatial scale.

Datasets with more than 10% data gaps were discarded.

The percentage of each station's inherent missing data is shown in Appendix A.



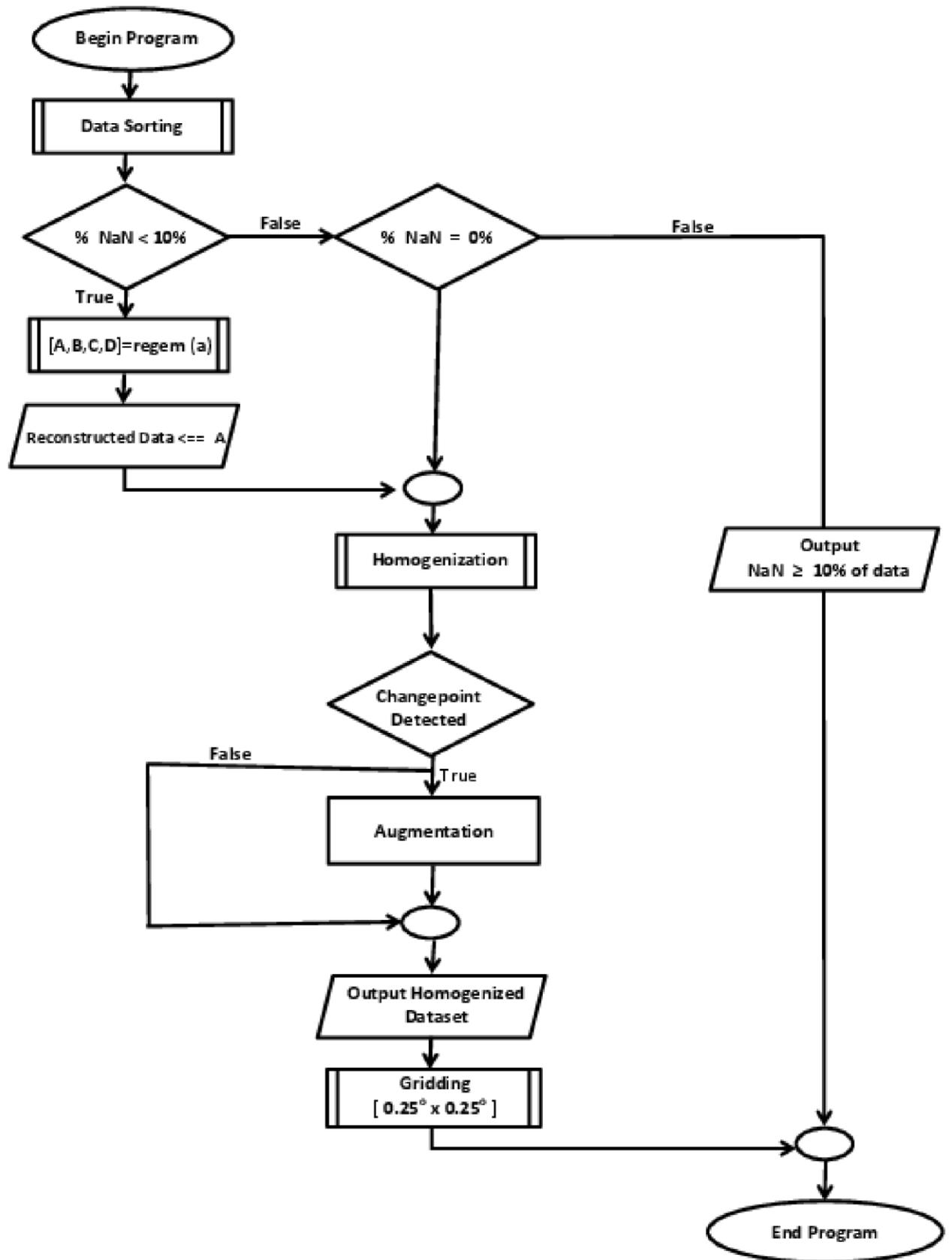


Figure 3.4: Flowchart of datagap estimation, homogenization and gridding steps carried out in the study.

## CHAPTER 4

### Results: analysis and discussion

#### 4.1 Data Reconstruction

RegEM was used in missing data estimation, and QMadj regression fit used for homogeneity testing. Some of the results are shown and discussed below in this section.

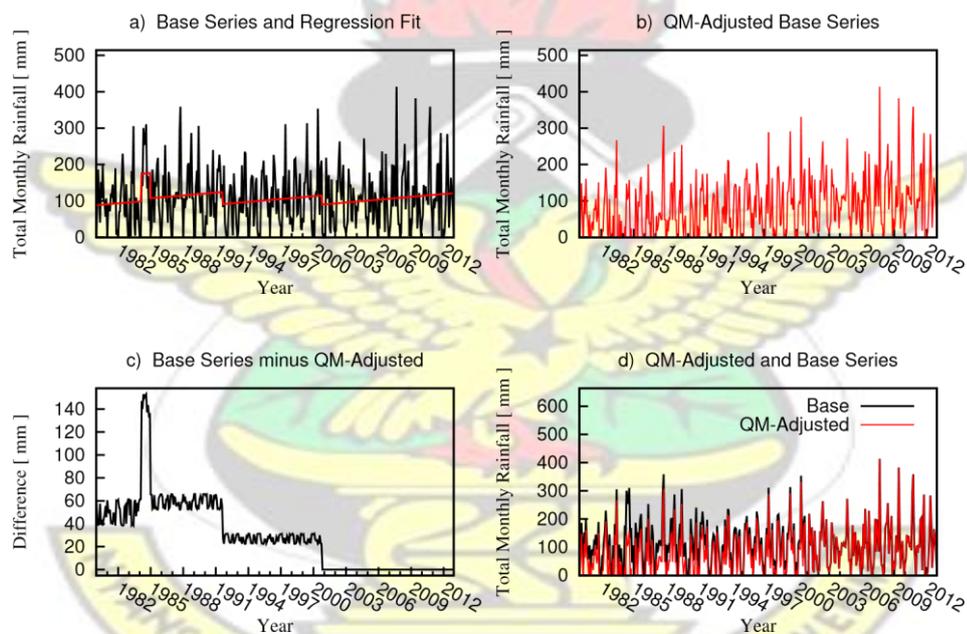


Figure 4.1: Homogenization by QMadj performed on Abetifi data. In Fig. 4.1a, the black line represents the detrended data (base series) and the red line is the regression fit (trend line) with any red vertical mark indicating a changepoint. Fig. 4.1b shows the homogenized dataset. The difference between the base series and the adjusted series is shown in Fig. 4.1c. An overplot of the two time-series is illustrated in Fig. 4.1d. Base series are the gap-filled datasets and QM-adjusted are the homogenized datasets.

Four different changepoints are identified within the datasets (see Figure 4.1a). These changepoints are associated with non-climatic shifts in the trend of the data timeseries. Such shifts are likely attributed to a total shift of the station or rearrangement of instruments in the observatory and other factors as described in section 2.4. Figure 4.1b shows the homogenized datasets. The detrended datasets are augmented based on the identified changepoints. Figure 4.1c also shows the relative changepoints (i.e the difference between the homogenized datasets or QM-adjusted series and the base series), ranging from 0 to 150 mm. An overplot of the base (black line) and QM-adjusted (red line) datasets is shown in Figure 4.1d. A clear indication of the magnitude of deviation of the homogenized datasets from the base series is clearly seen.

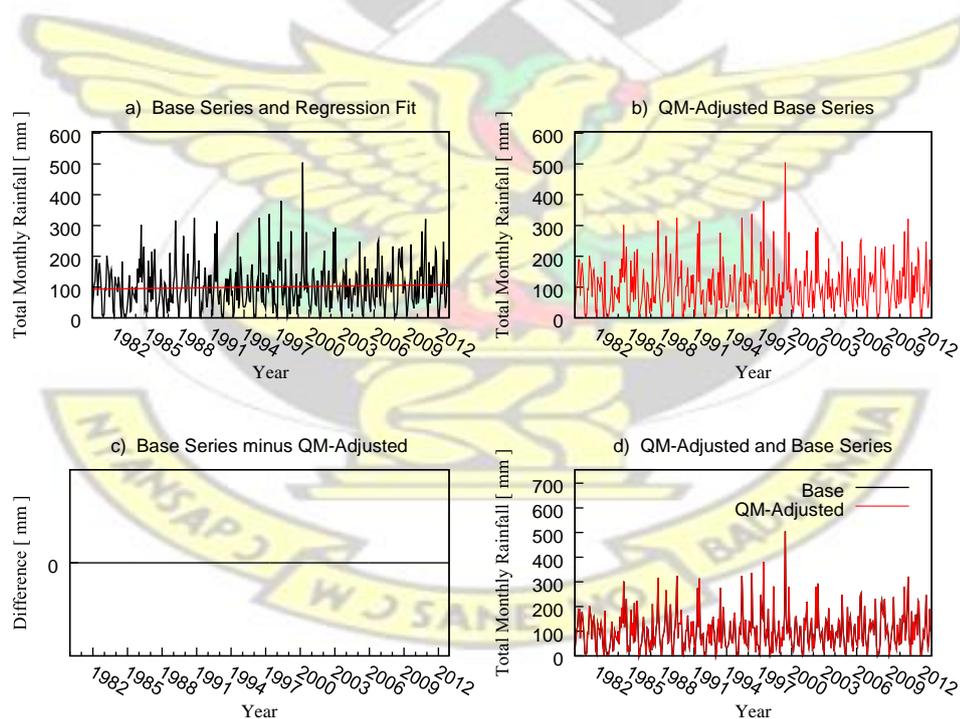


Figure 4.2: Homogenization by QMadj performed on Aburi data. In Fig. 4.2a, the black line represents the detrended data (base series) and the red line is the regression fit (trend line) with any red vertical mark indicating a changepoint. Fig. 4.2b shows the homogenized dataset. The difference between the base series and the adjusted series is shown in Fig. 4.2c. An overplot of the two time-series is illustrated in Fig.

4.2d. Base series are the gap-filled datasets and QM-adjusted are the homogenized datasets.

No changepoint is detected in the rainfall data-time-series for Aburi-Parks.

This is clearly seen by the 0 mm difference between the 2 datasets, as shown in Figure 4.2c. This indicates that variability of rainfall over the station is solely driven by climatic factors as discussed in Section 2.4

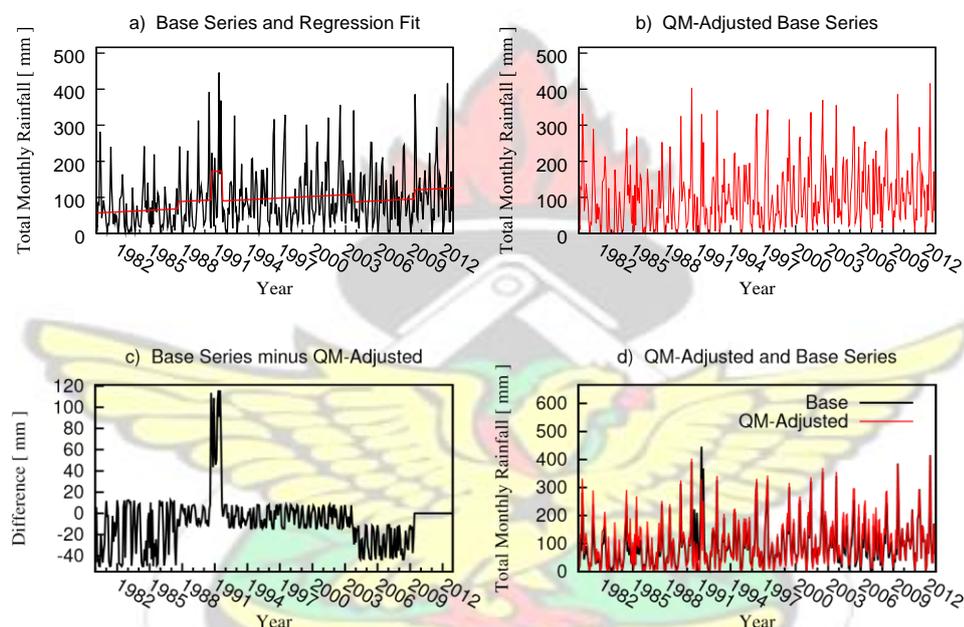


Figure 4.3: Homogenization by QMadj performed on Asuansi data. In Fig. 4.3a, the black line represents the detrended data (base series) and the red line is the regression fit (trend line) with any red vertical mark indicating a changepoint. Fig. 4.3b shows the homogenized dataset. The difference between the base series and the adjusted series is shown in Fig. 4.3c. An overplot of the two time-series is illustrated in Fig. 4.3d. Base series are the gap-filled datasets and QM-adjusted are the homogenized datasets.

Five changepoints (non-climatic shifts) are detected in the rainfall data-time-series for Asuansi in 4.3a. Relative difference between the two datasets (typically between -50 to 120) is shown in Figure 4.3c. An overplot of the two datasets in 4.3d, reveals the deviation between the base series and QM-adjusted series.

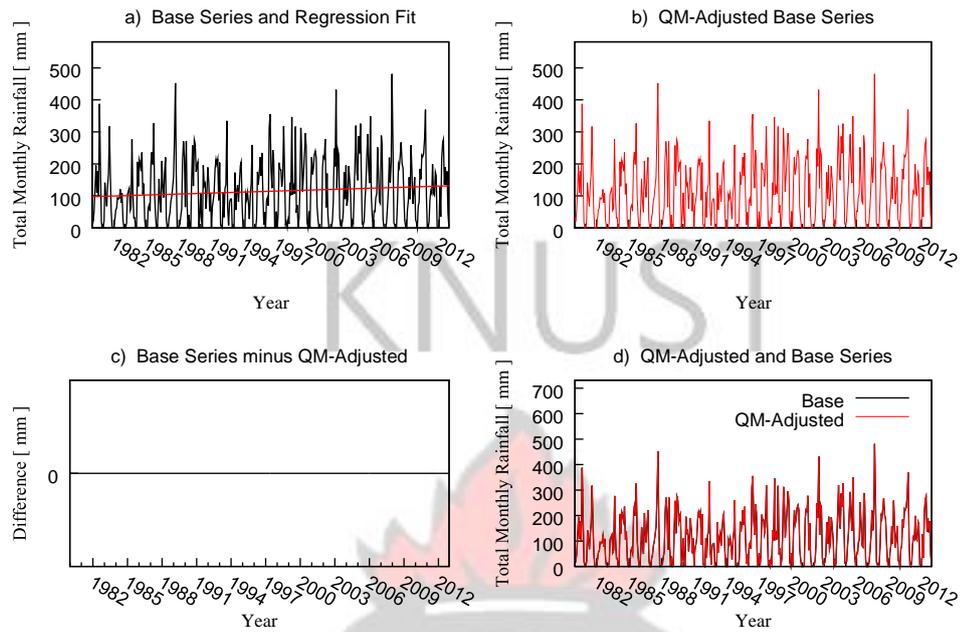
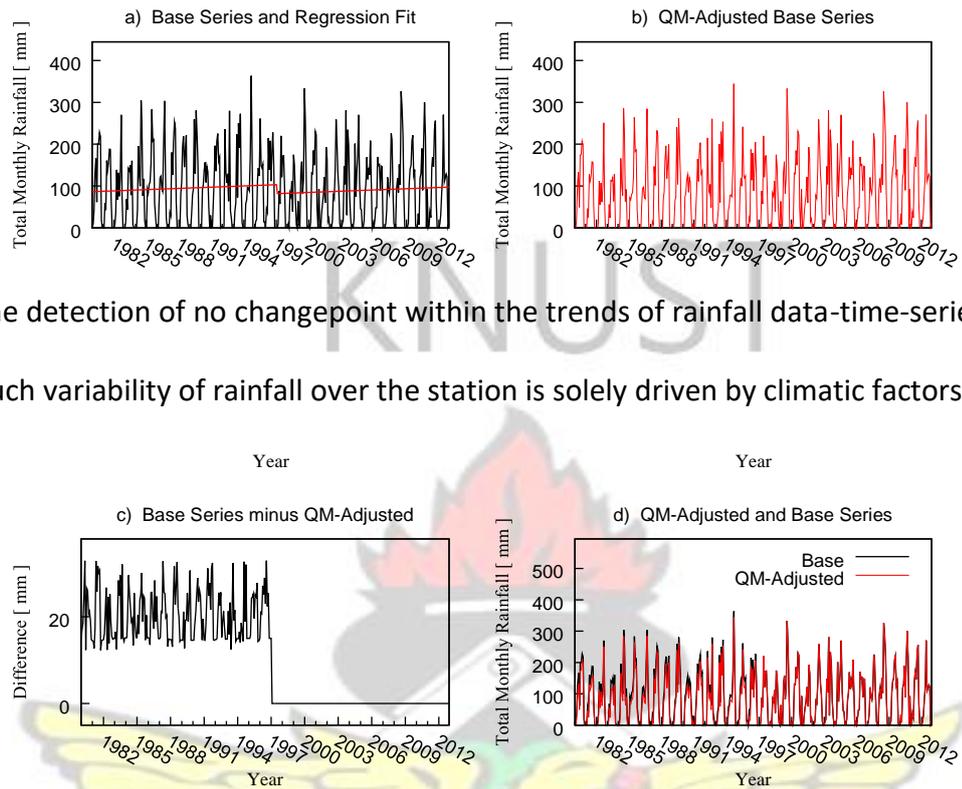


Figure 4.4: Homogenization by QMadj performed on Atebubu data. In Fig. 4.4a, the black line represents the detrended data (base series) and the red line is the regression fit (trend line) with any red vertical mark indicating a changepoint. Fig. 4.4b shows the homogenized dataset. The difference between the base series and the adjusted series is shown in Fig. 4.4c. An overplot of the two time-series is illustrated in Fig. 4.4d. Base series are the gap-filled datasets and QM-adjusted are the homogenized datasets.

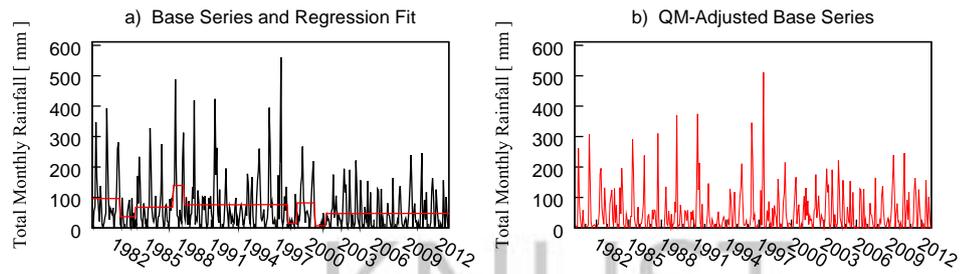
No changepoint is detected in the rainfall data-time-series for Atebubu. The difference of 0 mm between the 2 datasets, as shown in Figure 4.4c, is an indication



of the detection of no changepoint within the trends of rainfall data-time-series and as such variability of rainfall over the station is solely driven by climatic factors.

Figure 4.5: Homogenization by QMadj performed on Damango data. In Fig. 4.5a, the black line represents the detrended data (base series) and the red line is the regression fit (trend line) with any red vertical mark indicating a changepoint. Fig. 4.5b shows the homogenized dataset. The difference between the base series and the adjusted series is shown in Fig. 4.5c. An overplot of the two time-series is illustrated in Fig. 4.5d. Base series are the gap-filled datasets and QM-adjusted are the homogenized datasets.

One changepoint (non-climatic shift) is detected in the rainfall data-time-series for



Damango in Figure 4.5a. Relative difference between the two datasets is shown in Figure 4.5c. An overplot of the two datasets in Figure 4.5d, reveals the deviation between the base series and QM-adjusted series.

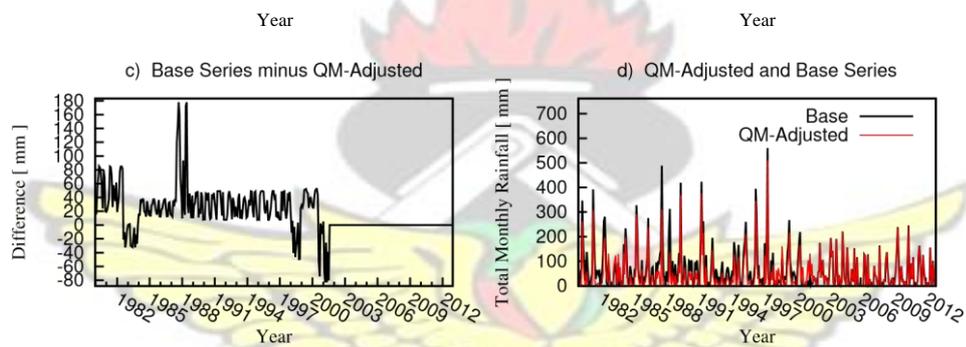
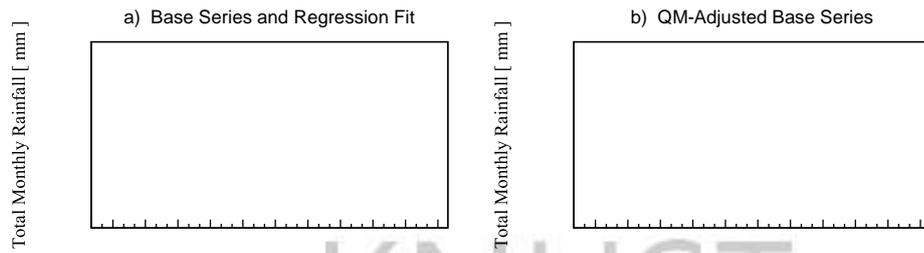
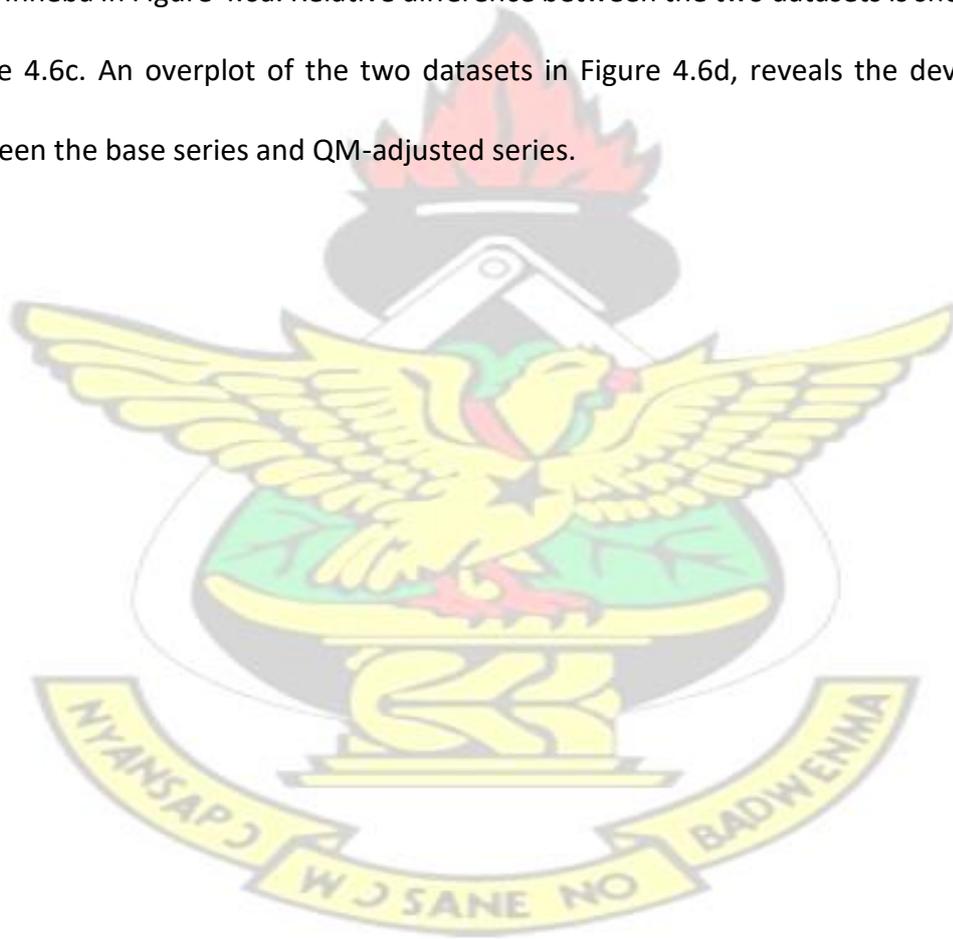


Figure 4.6: Homogenization by QMadj performed on Winneba data. In Fig. 4.6a, the black line represents the detrended data (base series) and the red line is the regression fit (trend line) with any red vertical mark indicating a changepoint. Fig. 4.6b shows the homogenized dataset. The difference between the base series and the adjusted series is shown in Fig. 4.6c. An overplot of the two time-series is illustrated in Fig. 4.6d. Base series are the gap-filled datasets and QM-adjusted are the homogenized datasets.



Eight changepoints (non-climatic shifts) are detected in the rainfall data-time-series for Winneba in Figure 4.6a. Relative difference between the two datasets is shown in Figure 4.6c. An overplot of the two datasets in Figure 4.6d, reveals the deviation between the base series and QM-adjusted series.



## 4.2 Rainfall Seasonality in the Agro-Ecological Zones

### 4.2.1 Northern Zone

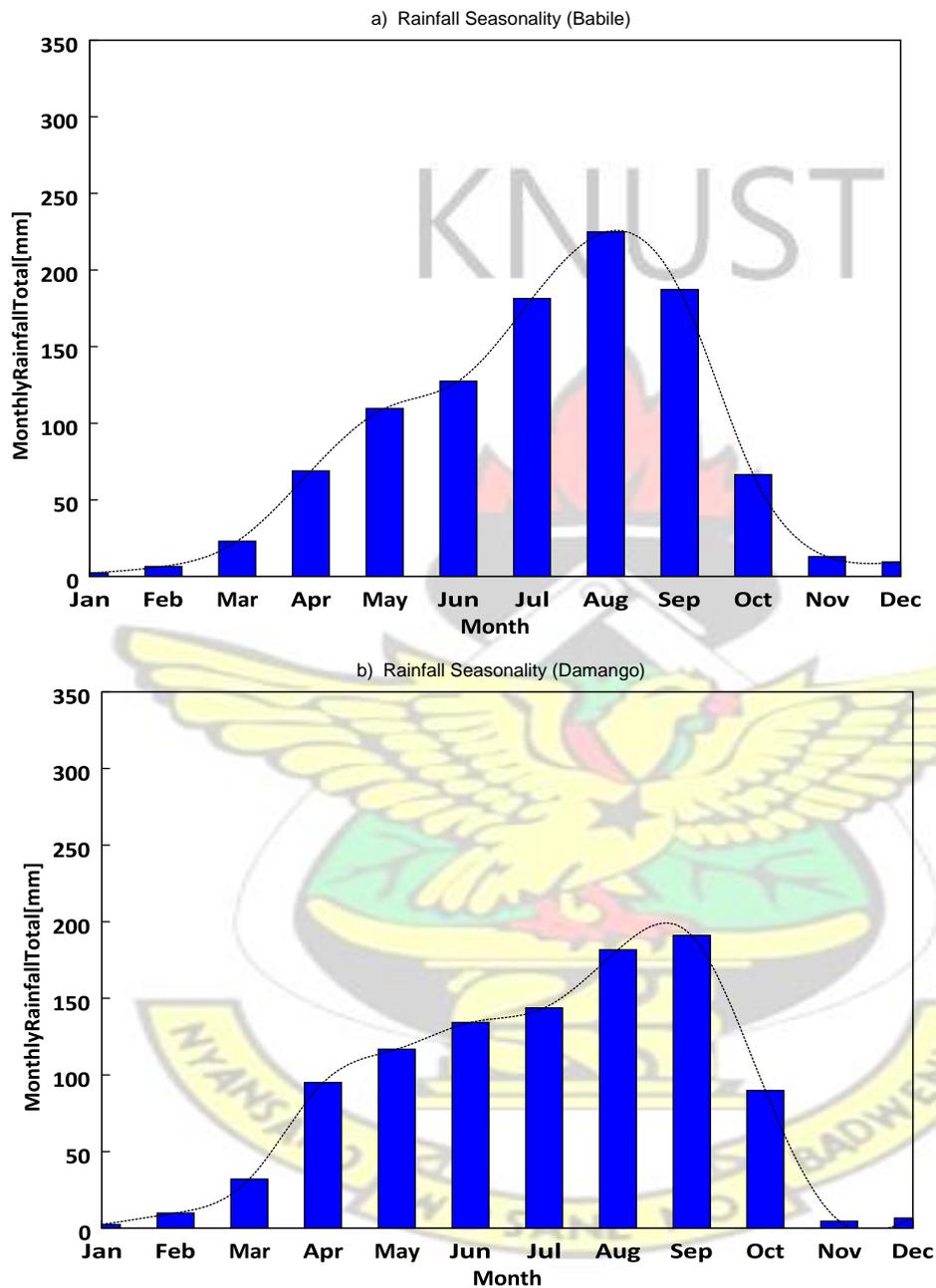


Figure 4.7: Seasonal Rainfall Patterns over selected stations in the Northern Zone for (a) Babile and (b) Damango.

A unimodal pattern of rainfall is observed in the Northern Zone (see Figure 4.7), with the onset of rains in April and its cessation in October. The unimodal pattern of rainfall over the area is associated with the once-in-a-year migration of the ITD

northwards. August marks the month of maximum rainfall in Babile (see Figure 4.7a) whereas August and September are the peak months of rainfall in Damango (see Figure 4.7b). Within these periods, the ITD is directly over the area. Damango, due to its location closer to the Transition zone, tends to have a rainfall pattern that shows some similarities to that of the transition. November to March are the dry months in the northern zone. This dryness is attributed to the southward location of the ITD below the area and as such, dry, dust-laden Trade Winds dominate the area within this period.



## 4.2.2 Transition Zone

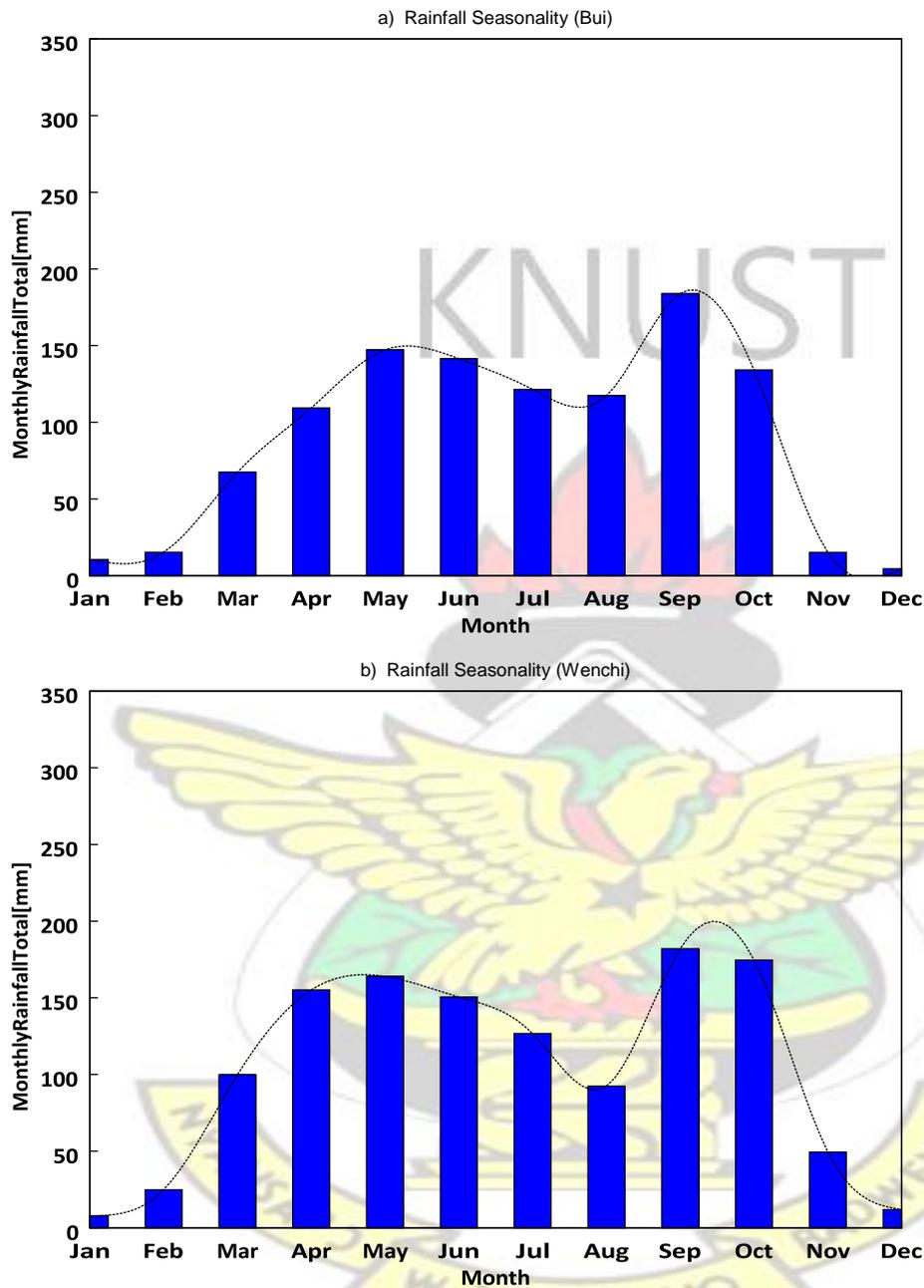


Figure 4.8: Seasonal Rainfall Patterns over selected stations in the Transition Zone for (a) Bui and (b) Wenchi.

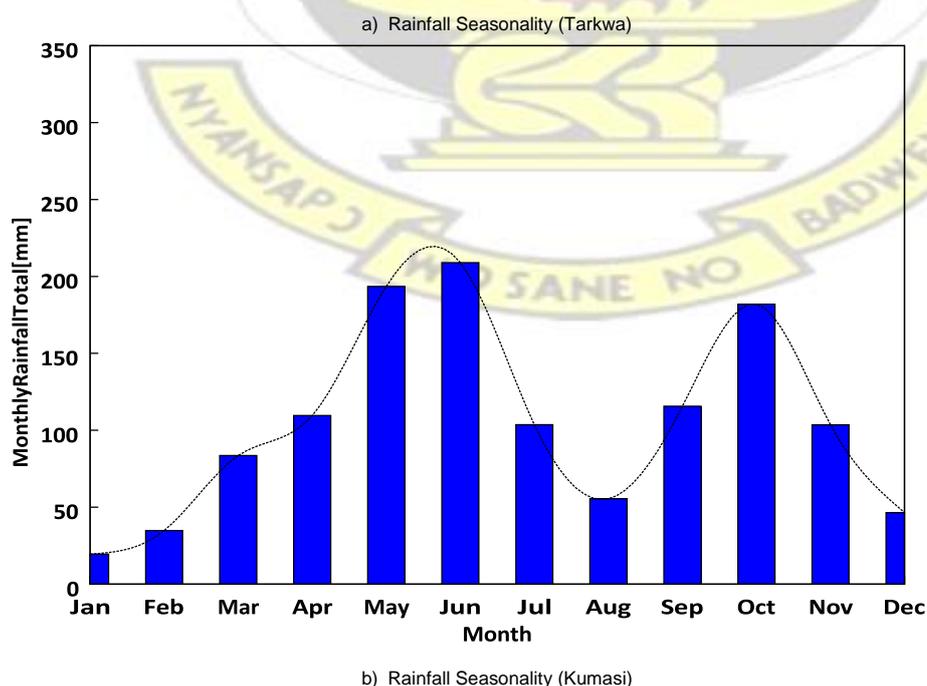
The transition zone is characterized by a bi-modal rainfall pattern with maximum rainfall recorded in the minor rainy season (see Figure 4.8). The transition zone has a longer growing season and thus, serves as the food hub of the country. Its very slim

temporal cessation is as a result of the very sharp return of the ITD downwards after it has moved northwards.

The transition zone has its rainy season spanning the months of April to October. Rainfall peaks over areas in the upper transition zone such as Bui are recorded in September (see Figure 4.8a) while rainfall peaks over areas in the lower transition zone such as Wenchi are recorded in September-October (see Figure 4.8b).

A temporal cessation – associated with the location of the ITD above the zone – is observed in August and this separates the two rainfall regimes. The temporal cessation is however pronounced in the lower transition zone and this creates a rainfall pattern with some similarities to that of the Forest Zone. The Harmattan season dominates the months of November to March in the transition zone. It is attributed to the location of the ITD below the Zone, and as such the dry, dust-laden Trade Winds dominate the zone within that period.

### 4.2.3 Forest Zone



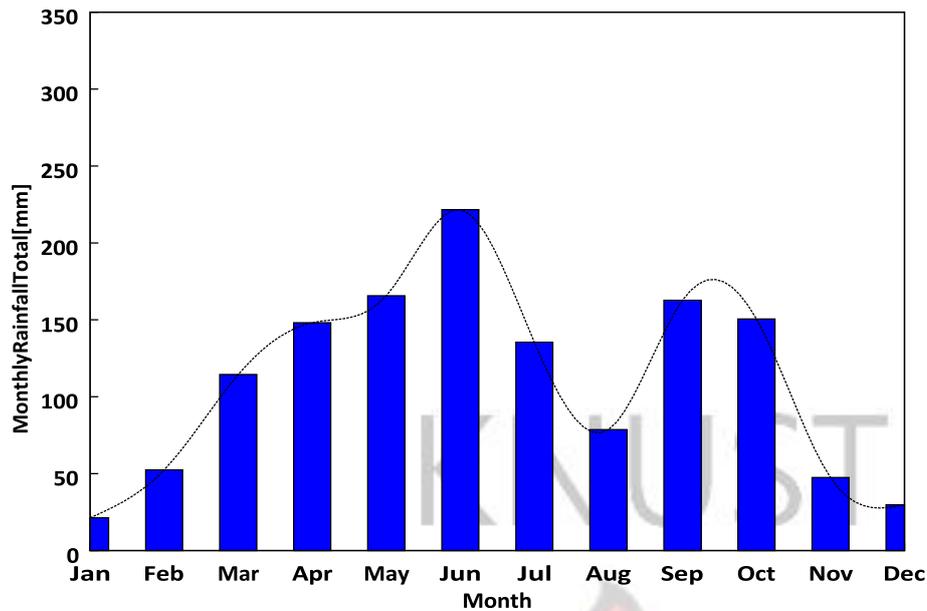


Figure 4.9: Seasonal Rainfall Patterns over selected stations in the Forest Zone for (a) Tarkwa and (b) Kumasi.

The forest zone is characterized by a bi-modal (major and minor) rainfall regime (see Figure 4.9). Rainfall onset and cessation are in April and October respectively in the upper forest zone (see Figure 4.9a) whereas the lower Western forest zone has onset and cessation are in March and November respectively (see Figure 4.9b). The major rainy season has rainfall peaks in the months of May and June. September marks the rainfall peaks of the minor rainfall season in the upper forest zone whereas October marks the rainfall peaks of the minor rainfall season in the lower forest zone. The major (minor) rainy season is associated with ITD migration northwards(southwards) over the forest zone. The recess of the ITD is observed to be sharper than its northward advancement, resulting in a shorter minor season and a bit longer major season.

In the month of August, the ITD migrates northwards beyond the forest zone, thereby establishing a rainfall cessation over the zone. The periods of November to the March mark the dry (Harmattan) season which is attributed to ITD location below the zone.

Dry, dust-laden Trade Winds dominate the zone within this period.

#### 4.2.4 Coastal Zone

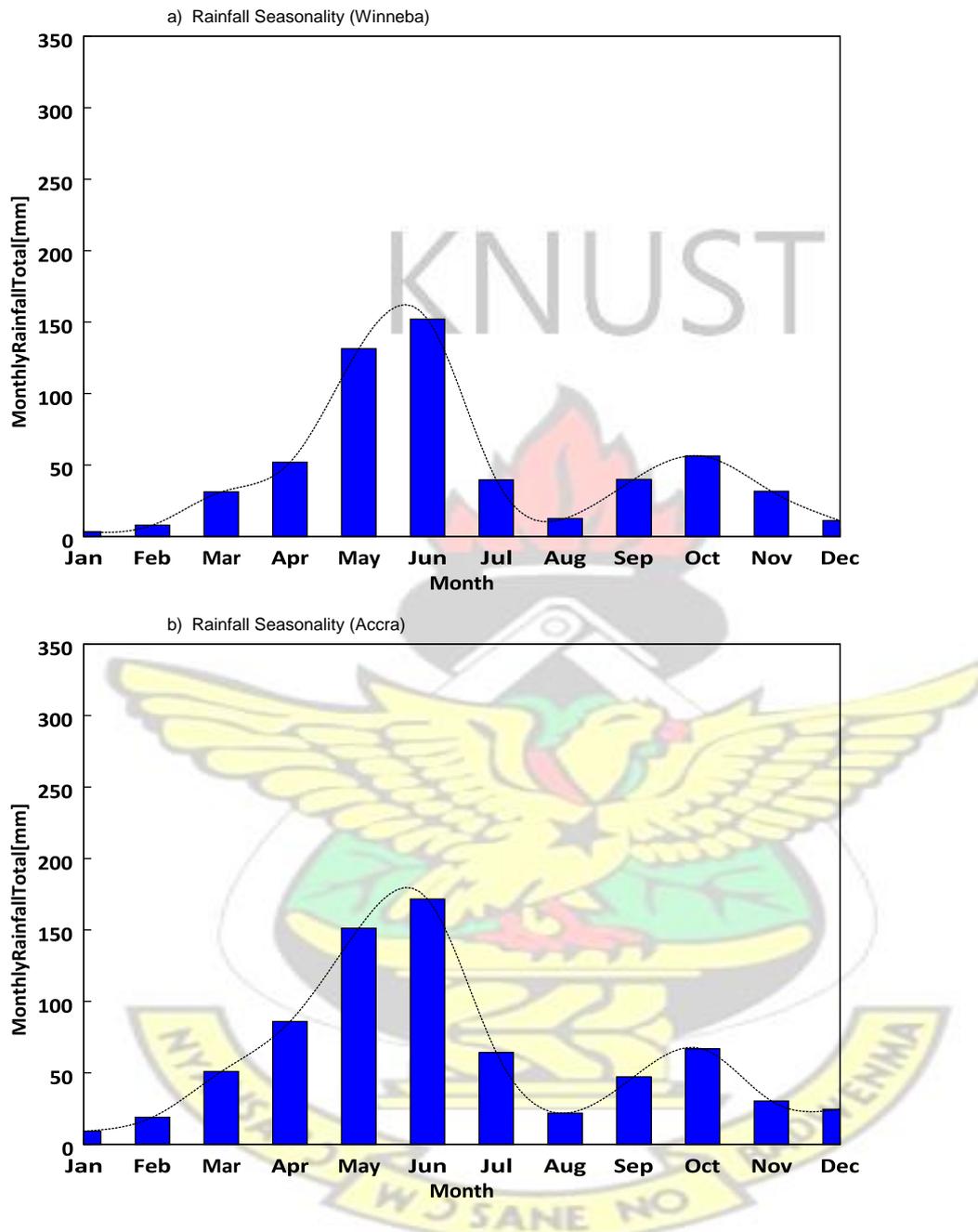


Figure 4.10: Seasonal Rainfall Patterns over selected stations in the Coastal Zone for (a) Winneba and (b) Accra.

The coastal zone is also characterized by a bi-modal (major and minor) rainfall pattern

[see Figure 4.10]. There is a similarity in the rainfall pattern of the coast and the forest

zone, however, rainfall amounts are lower in the coastal zone than the forest zone.

Rainfall onset and cessation are in April and October respectively in the Coastal zone.

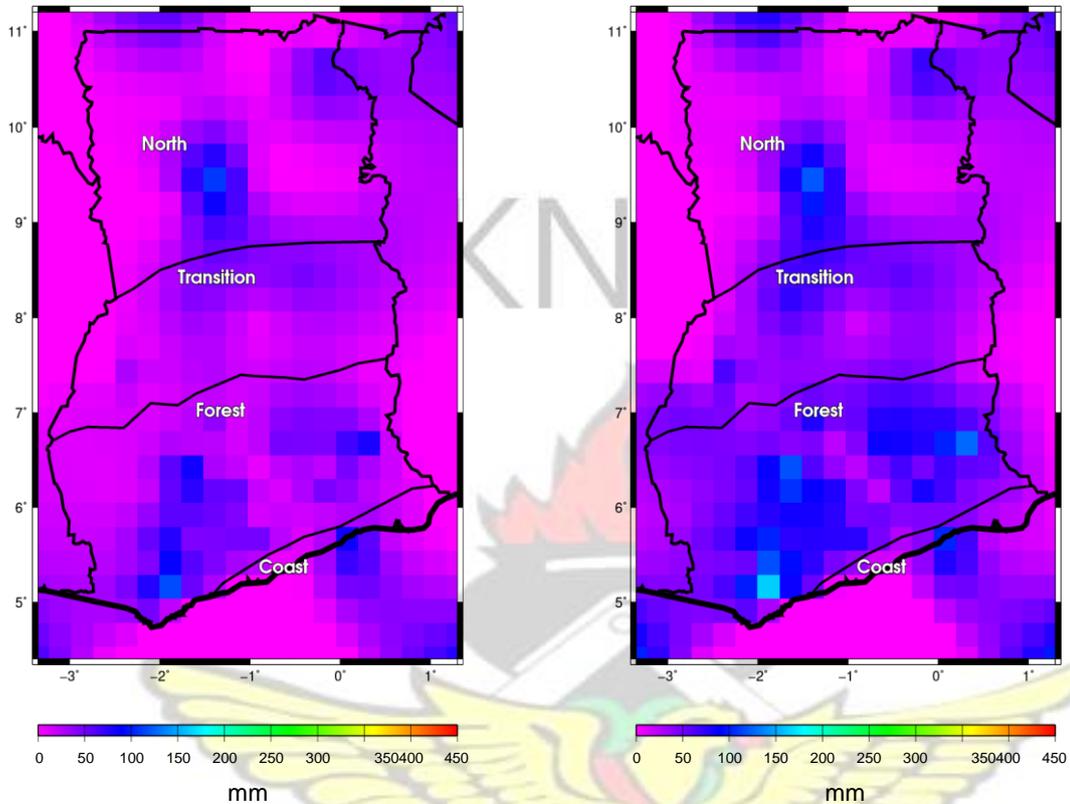
The major rainy season has rainfall peaks in the months of May and June whereas October marks the rainfall peaks of the minor rainfall season. The recess of the ITD is again observed to be sharper than its northward advancement, resulting in a shorter minor season and a longer major season.

In the month of August, the ITD migrates northwards beyond the forest zone, thereby establishing a temporal cessation over the zone. Also, the months of November to March are the dry (Harmattan) season in the Coastal Zone which is associated with the location of the ITD below the zone.



## 4.3 Gridding

### 4.3.1 Monthly Rainfall Climatology



(a) January

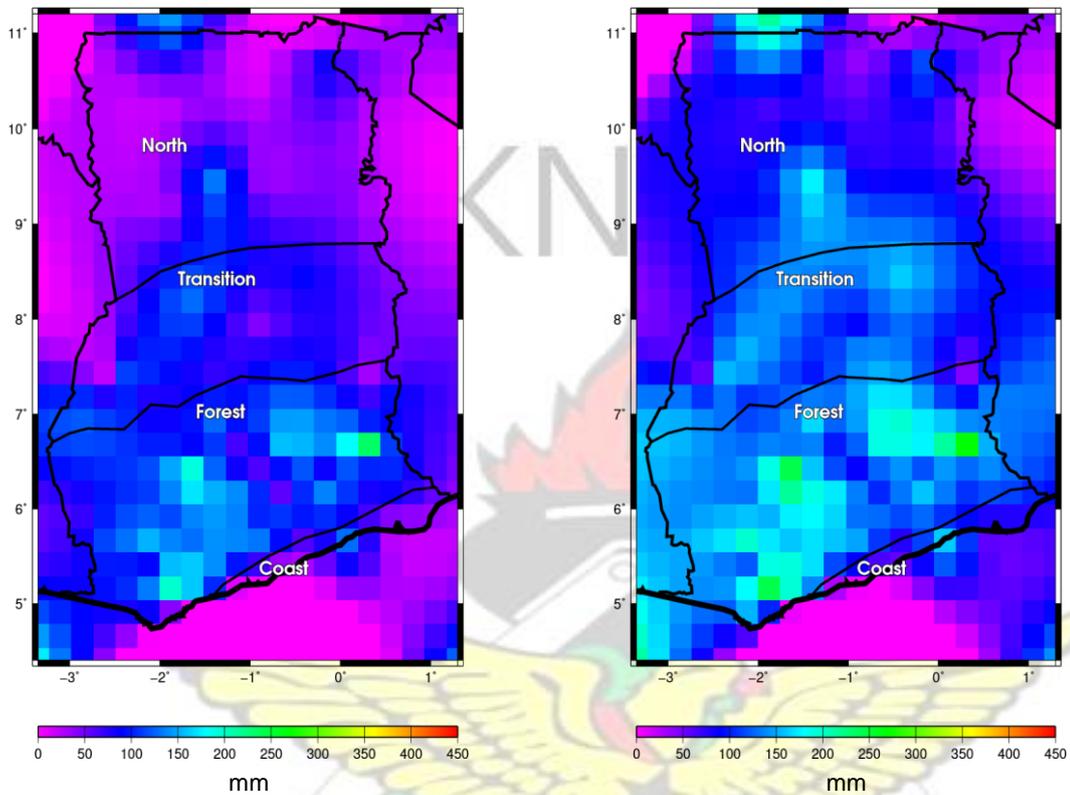
(b) February

Figure 4.11: Gridded Monthly Rainfall Total. (a) represents January and (b) represents February.

January is a dry month country-wide. The ITD is located at its southward apex beyond the country and as such, the dry, dust-laden Trade Winds are expected to dominate over the entire country. The gridded data shows monthly rainfall totals to be ranging up to a maximum of ~ 120 mm (see figure 4.11a) over the four agro-ecological zones of Ghana.

February is also a dry month over the entire country, with similarities to the January rainfall trend. The ITD, although has begun its migration northwards, is still located southwards beyond the country and as such, the dust-laden Trade Winds are

expected to continue dominating the country. Rainfall amounts over the entire country for February, as shown in figure 4.11b, ranges up to a maximum of approximately 150 mm.



(a) March

(b) April

Figure 4.12: Gridded Monthly Rainfall Total. (a) represents March and (b) represents April.

The month of March happens to be the rainfall onset month for some grids within the forest zone. This is associated with the continuous and gradual, northward migration of the ITD. Rainfall within these months range up to a maximum of approximately 200 mm as shown in figure 4.12a.

Owing to the gradual northward migration of the ITD, all forest and coastal grids fully experience the onset of the rainy season latest by April as shown in figure 4.12b, with monthly rainfall totals ranging up to 250 mm. An inflexion point is observed at grid “9.5N 1.0W” in the Northern Zone and grids directly beneath it (see Figure 4.12b).

These grids are the point of confluence for the Black and White Volta. The high inflexions are likely associated to local hydroclimatic factors that tend to impact on the climatology of the area.

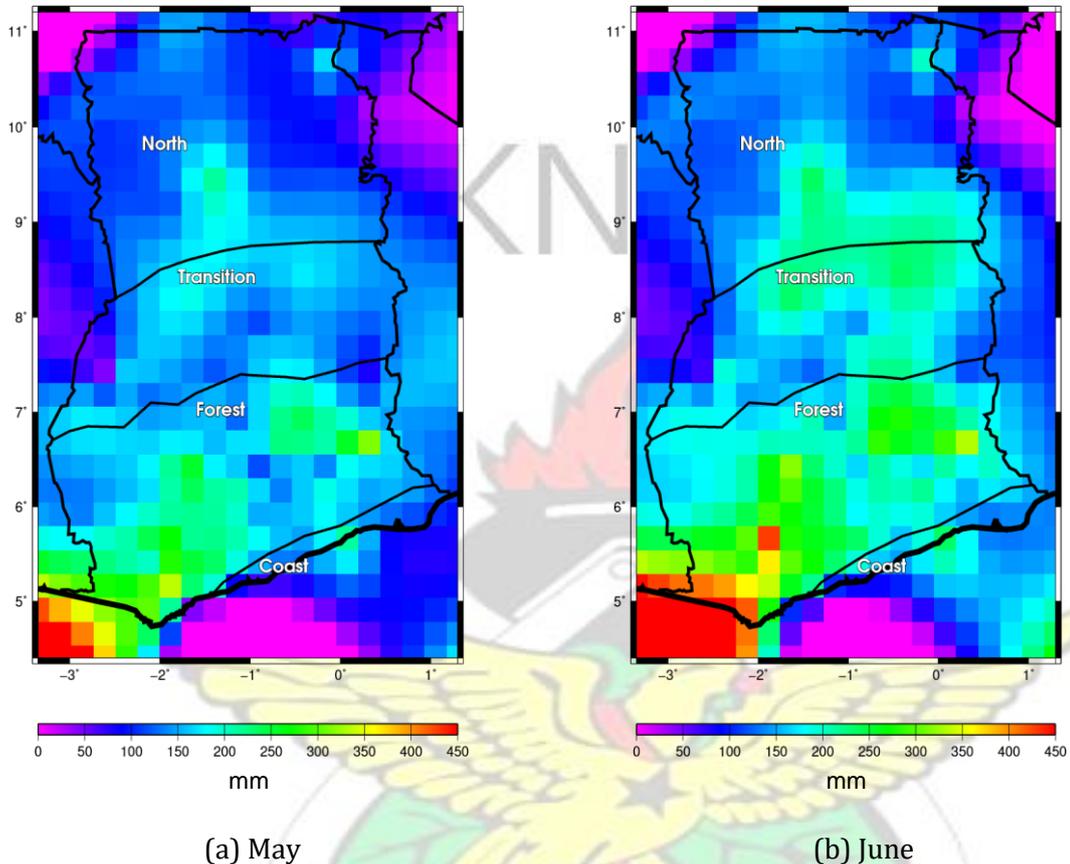
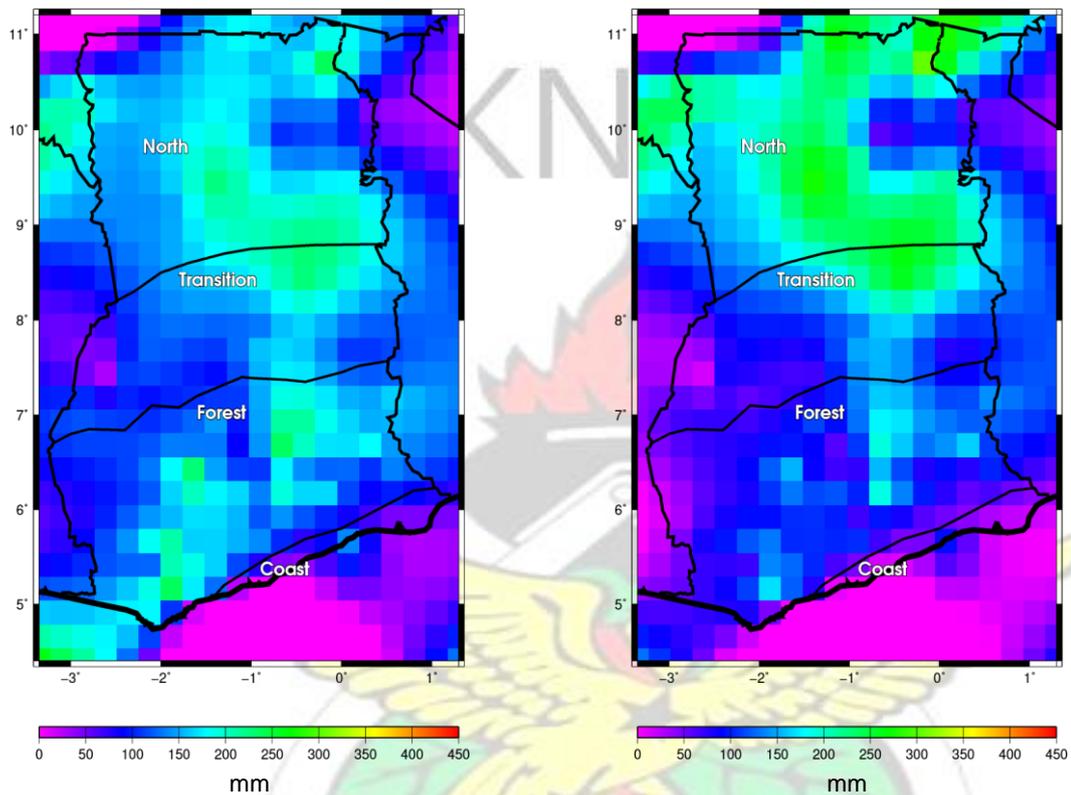


Figure 4.13: Gridded Monthly Rainfall Total. (a) represents May and (b) represents June.

By May, more rains feed in over the Coastal and Forest Zones, with rainfall amounts ranging up to 300 mm. Sections of the Northern and Transition Zones now experience the onset of rains within this month. Rainfall amounts over the Transition Zone ranges up to about 200 mm. The Northern Zone also has rainfall amounts ranging to 150 mm. Again, the inflexion point is observed at the locations of confluence for the Black and White Volta (see Figure 4.13a).

June is the wettest month for the Forest and Coastal Zone, with rainfall amounts ranging to an approximate of 450 mm and 350 mm respectively. Within this period the ITD is

directly overhead the Forest Zone and as such, accounts for the annual rainfall maxima over the zone. Also, rainfall amounts in the Transition Zone ranges up to 320 mm and the North ranges up to 200 mm. Similarly, portions of the Transition zone and lower portions of the Northern zone record high rains as shown in Figure 4.13b.



(a) July

(b) August

Figure 4.14: Gridded Monthly Rainfall Total. (a) represents July and (b) represents August.

By July, the ITD is expected to be traversing the Upper Transition and the Northern Zone, and continuing its northward migration (see Figure 4.14a). Rainfall, as such is expected to be quite high over portions ranging from the mid Forest Zone upwards.

Rainfall values are expected to be up to a maximum of approximately 270 mm. The

Coastal Zone, within this period is nearing a temporal cessation. Rainfall amounts are expected to be reducing within this period, with values less than 150 mm.

The ITD reaches its northward apex in the month of August. Due to its location at the upper Northern Zone, the zone tends to receive its all year-round maximum rainfall

within this month as shown in Figure 4.14b. Rainfall over the North ranges up to approximately 350 mm. The Upper Transition Zone also receives quite substantial rainfall amounts (less or equal to 350 mm). On the contrary, the Forest and Coastal Zones are in a temporal cessation period with rainfall totals ranging up to 180 mm.

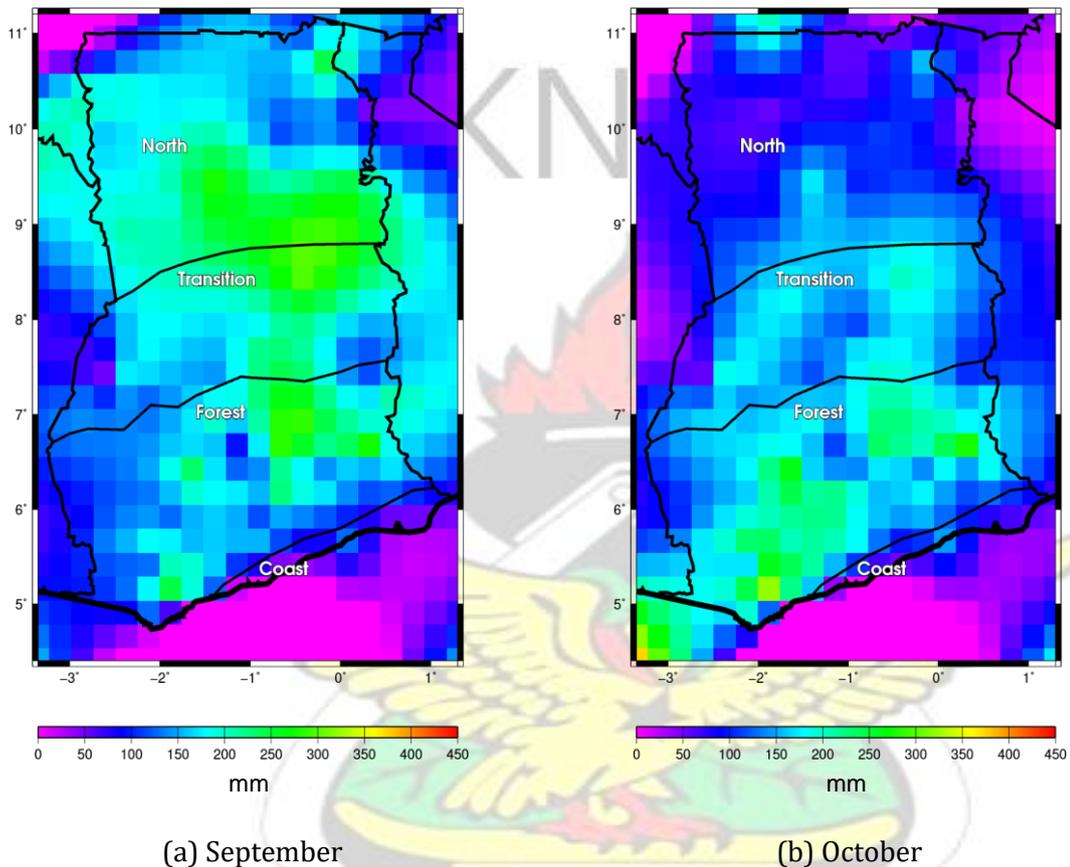


Figure 4.15: Gridded Monthly Rainfall Total. (a) represents September and (b) represents October.

Upon reaching its northwards apex, the ITD returns downwards. Its return is sharper than its advancement. The very sudden return of the ITD causes the Transition zone and some lower parts of the Northern zone to experience maximum monthly rainfall within September (in the minor rainy season). Rainfall amounts over the transition Zone ranges up to about 350 mm. Upper North, Forest and Coastal Zones record rainfall values up to about 200 mm.

October happens to be the peak months for rainfall in the minor rainy season over the Forest and Coast. Rainfall amounts typically range up to 300 mm and 200 mm respectively. As earlier stated, these peaks are associated with the Southward migration of the ITD. Within this month, the ITD is expected to be over the Forest Zone. Within this month, rainfall over the Northern Zone is expected to be low with amounts typically of 100 - 150 mm.

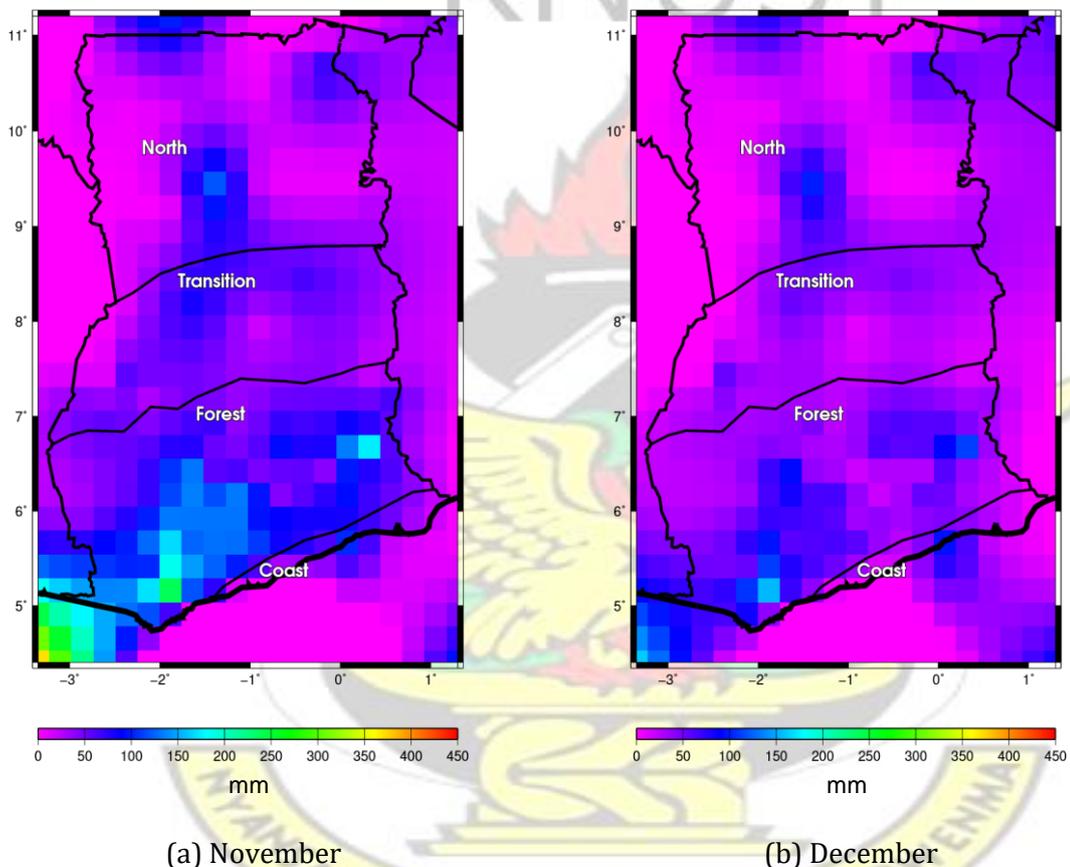


Figure 4.16: Gridded Monthly Rainfall Total. (a) represents November and (b) represents December.

Due to the southward ITD migration, the Trade Winds start to dominate over the Northern and upper Transition Zone in the month of November. This is the onset of the Harmattan periods for the country. The Forest Zone records some few rains, especially around the south-west shorelines, within the month of November before the Harmattan sets in. Rainfall amounts over the entire country are less than 250 mm as shown in Figure 4.16a.

The entire country is dominated by the dry, dust-laden Trade Winds (Harmattan) in December due to the southward ITD location, approaching its southwards apex. Rainfall amounts over the entire country, within the month of December, are less than 150 mm as shown in Figure 4.16b.

#### 4.3.1.1 Summary of Monthly Rainfall Climatology

Table 4.1: Estimates of monthly rainfall total in the four agro-ecological zones.

Month	Rainfall ( mm)			
	North	Transition	Forest	Coast
January	< 100	< 100	< 100	< 100
February	< 100	< 100	< 150	< 100
March	< 150	< 150	< 250	< 150
April	50 - 200	50 - 200	100 - 250	50 - 200
May	50 - 200	50 - 200	100 - 350	100 - 250
June	100 - 300	100 - 300	150 - 450	100 - 300
July	100 - 300	100 - 300	100 - 250	< 200
August	100 - 300	100 - 300	50 - 200	< 150
September	50 - 300	50 - 300	50 - 300	< 150
October	50 - 150	50 - 200	50 - 300	< 200
November	< 100	< 100	< 200	< 100
December	< 100	< 100	< 100	< 100

Monthly rainfall totals of the four agro-ecological zones are presented in Table 4.1, with each entry representing the probable range of monthly rainfall total for any grid that lies within a particular zone.

The north records monthly rainfall peaks of about 300 mm in any of the months between June and September. The forest is the zone with the overall highest cumulative rainfall amount (450 mm) amongst the four agro-ecological zones, whereas the coast is the zone with the least cumulative rainfall amount. The coastal

zone has rainfall pattern similar to that of the forest zone, however, rainfall amounts over the coast are lesser. All the zones tend to record their maximum monthly rainfall within the period of June to August, and their minimum between November and May.

# KNUST



### 4.3.2 Seasonal Rainfall Climatology

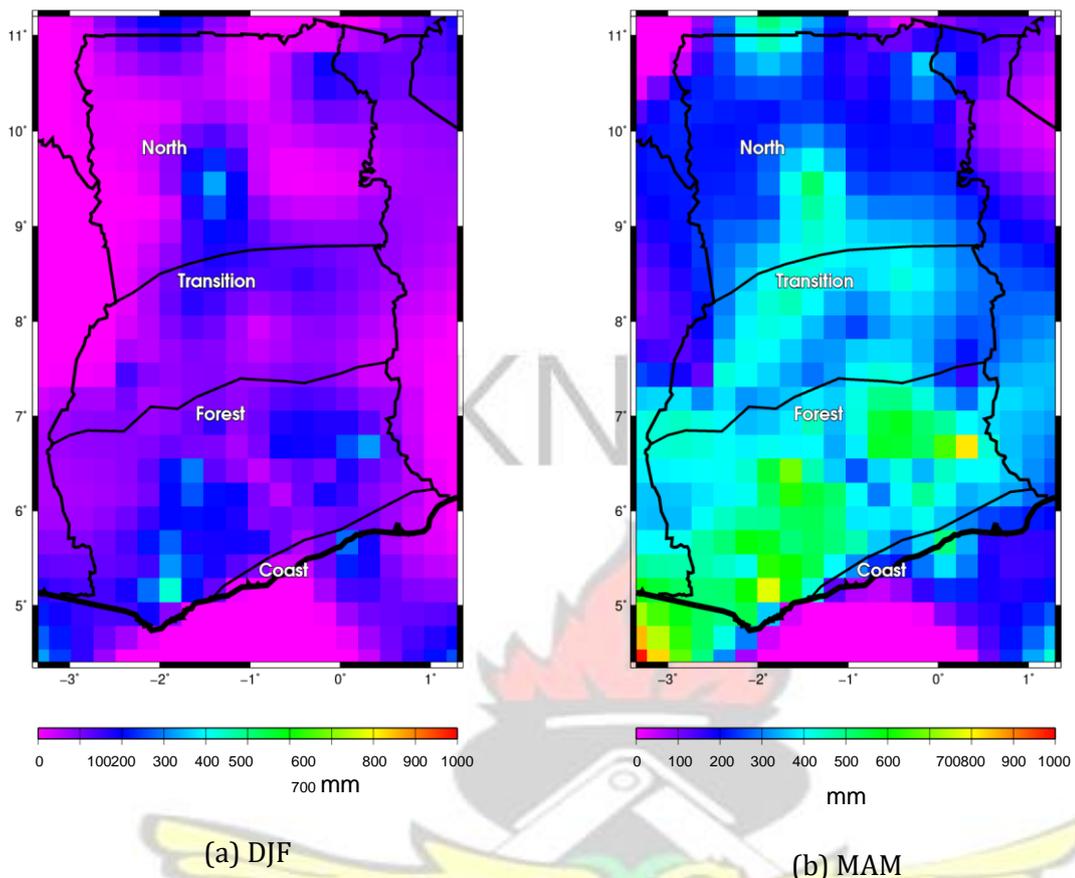
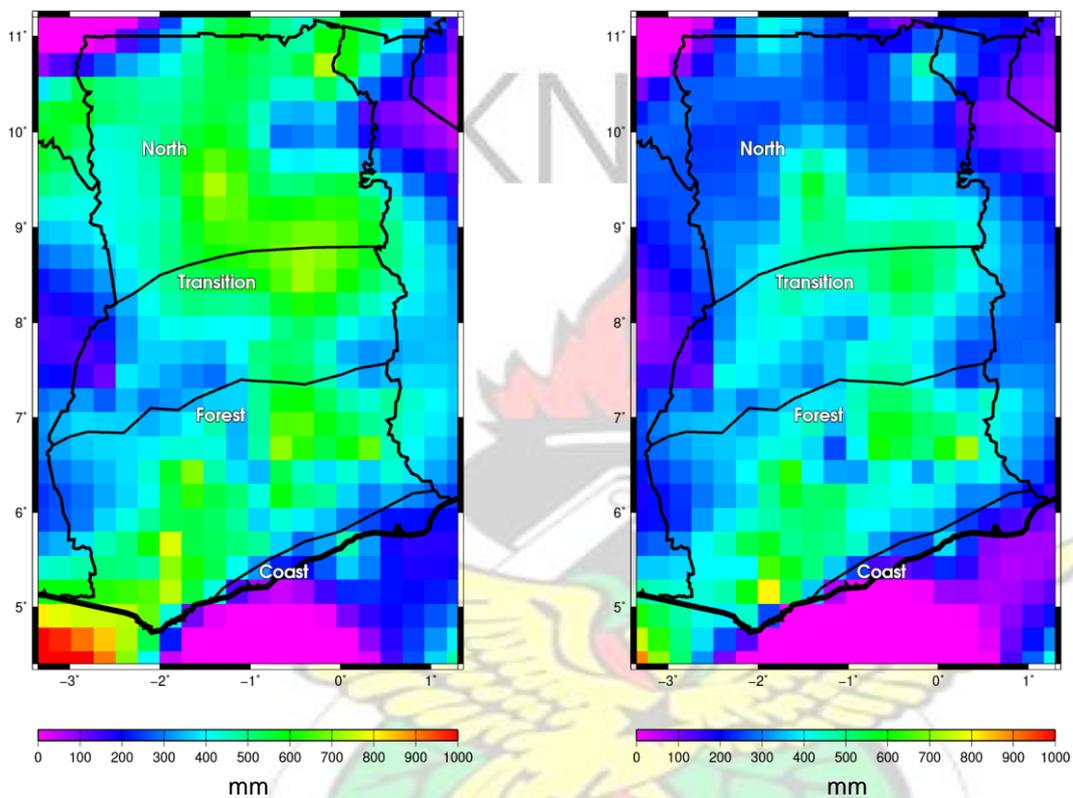


Figure 4.17: Gridded Seasonal Rainfall Total. (a) represents December-January-February(DJF) and (b) represents March-April-May (MAM).

The very first trimester (DJF) are the driest periods for the entire country. The whole country, within these periods, receive very minimal rains. These total to about 300 mm or less for the entire three months period over the entire country (see Figure 4.17a). Within these trimestry, ITD is located southwards beyond the country and as such the prevailing winds are the Trade Winds which are dry and dust-laden, thereby establishing the Harmattan periods over the country within these months.

The second trimester (MAM) happens to be the rainfall onset periods for the various zones (see Figure 4.17b). Within these periods, rainfall amounts are expected to be higher than the first trimester. By this trimester, the ITD has started migrating northwards, bringing in rains. These periods, being onsets of rains, are ideally the best months for farmers to start

planting. Since the onset for the Forest and Coastal Zone precedes that of the Transition and North in that respective order, rainfall amounts over the former zones will be higher than the latter, with rainfall amounts up to 700 mm over the three months duration. The transition and Northern Zones are expected to record up to 500 mm over the period.



(a) JJA (b) SON  
 Figure 4.18: Gridded Seasonal Rainfall Total. (a) represents June-July-August(JJA) and (b) represents September-October-November (SON).

The third trimester (JJA) is the very wet period of the year countrywide. Within these periods, the ITD is migrating over the country and thus, feeds in more rains over the country. Total rainfall within this period can peak to as high as 800 mm (see Figure 4.18a). In June, more rains feed in to the Coast and Forest Zones and in August, more rains feed in to the North and Transition Zones.

The fourth trimester (SON) is the minor rainy period for the Transition, Forest and Coastal Zones. Rainfall amounts within this period ranges up to about 700 mm in the

aforementioned zones and about 400 mm in the upper part of the Northern Zone. This minor season is attributed to the sharp return of the ITD southwards. A near-dry condition is observed at the upper Northern Zone (see Figure 4.18b).

### 4.3.3 Annual Rainfall Climatology

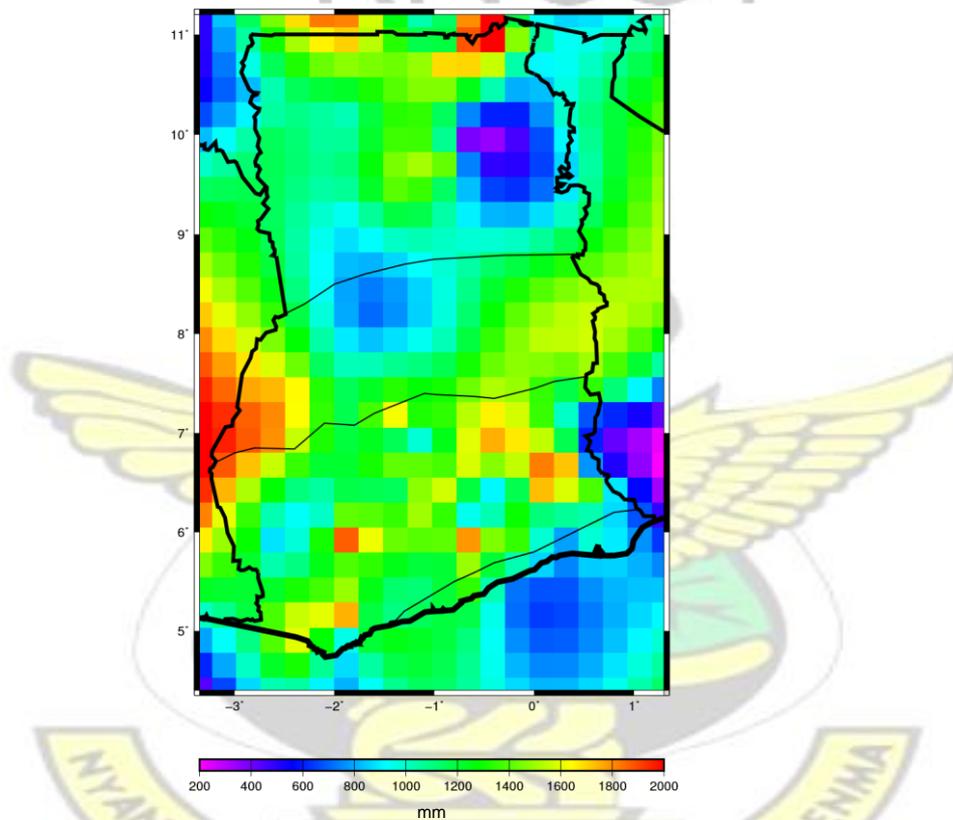


Figure 4.19: Annual rainfall total over the four agro-ecological zones. Figure 4.19 illustrates the annual rainfall climatology over the four agro-ecological zones. It is observed that the entire country receives high cumulative rainfall amounts per year, with the Forest zone receiving the most of rains (ranging up to approximately 2000 mm). The North and Transition zones also record substantial annual rainfall amounts (with majority of the grids recording rains up to approximately 1500 mm), however, annual rainfall totals in the coast are the least of

the four zones. Nonetheless, the west-Coastal zone recorded a relatively high rainfall amount (up to 1500 mm) than the east-Coastal zone (less than 1000 mm).

#### 4.3.3.1 Summary of Seasonal & Annual Rainfall Climatology

Table 4.2: Estimates of seasonal & annual total in the four agro-ecological zones.

Month	Rainfall ( mm)			
	North	Transition	Forest	Coast
Dec - Jan - Feb (DJF)	< 300	< 300	< 400	< 400
Mar - Apr - May (MAM)	100 - 500	100 - 500	100 - 900	100 - 700
Jun - Jul - Aug (JJA)	200 - 800	200 - 800	300 - 800	200 - 600
Sep - Oct - Nov (SON)	200 - 600	200 - 600	200 - 800	100 - 500
Annual	600 - 1800	800 - 2000	900 - 2000	600 - 1400

Table 4.2 shows estimates of seasonal rainfall totals for the four agro-ecological zones, with each column representing the range of rainfall total for any grid lying within a particular zone. The forest zone has an all-year maximum seasonal rainfall, with the coastal zone recording the least seasonal rainfall. In all zones, the second trimester (MAM) was found to be the rainfall onset over the country and thus, is the best season for farmers to engage in planting. The third trimester (JJA), is identified as the season of probable extreme events over the entire country. The very first trimester, however, is the driest period over the entire country. Within this period, the dry and dust-laden, North-East Trade Winds dominate the whole country, thereby establishing dry conditions over all zones. This is associated with the location of the ITD southwards, beyond the country.

On an annual basis, the coast tends to record the least of rains of the four zones while the Forest zone and portions of the Transitional zone record the highest rains.

## 4.4 Validation

CRU TS precipitation datasets are on a  $0.5^\circ \times 0.5^\circ$  spatial resolution, hence GMet point rainfall datasets were re-gridded on a  $0.5^\circ \times 0.5^\circ$  spatial resolution in order for the various statistical analyses to be performed to assess the agreement between the two datasets.

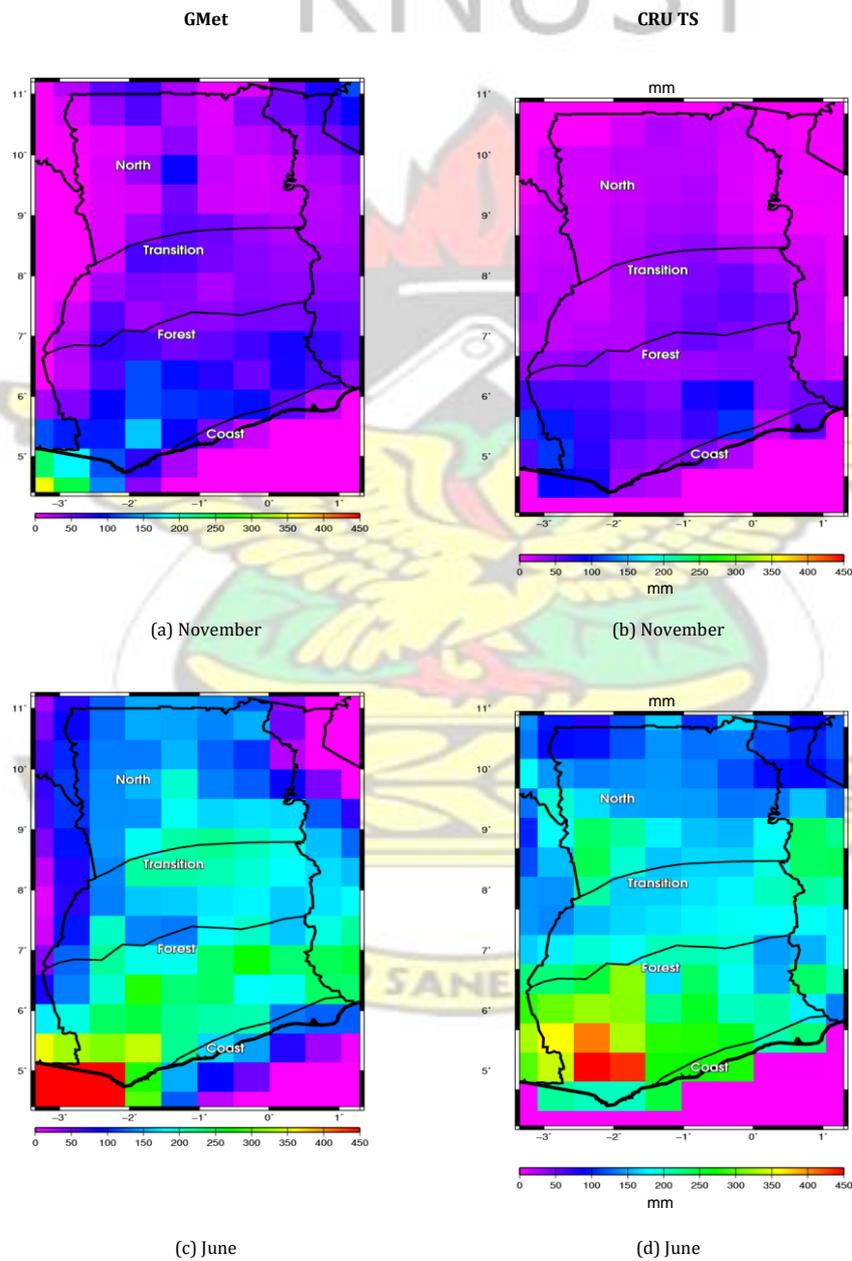


Figure 4.20: Validation of GMet dataset with CRU TS. (a) represents November rainfall from GMet dataset, (b) represents November from CRUTS, (c) represents that of June from GMet data while (d) represents June from CRUTS.

A good agreement is observed between the observed rainfall grids and CRU TS gridded precipitation, with many of the grids showing just little disparity. In the dry season, CRU TS tends to underestimate rainfall amount in all four zones (see Figure 4.20 a and 4.20 b ), however, in the rainy period, CRU TS tends to over-estimate rainfall values within the South-West forest zone, with the exception of the very last grid, where an under-estimation is observed (see Figure 4.20 c and 4.20 d ).

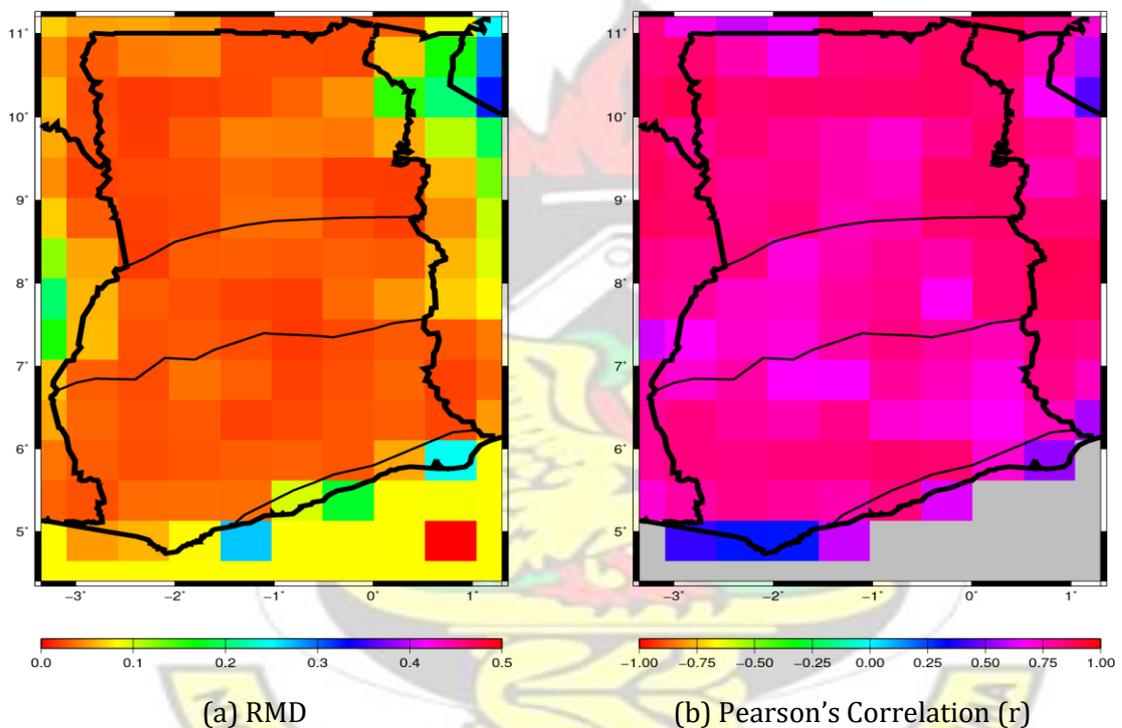


Figure 4.21: RMD and RMSE Statistical Tests. (a) represents RMD and (b) represents Pearson's Correlation.

RMD was used to quantify the magnitude of the variation between the CRU TS and the observed datasets. Low RMD values of about 0 to 0.3 was observed over the country between the two datasets (see Figure 4.21a): an indication of minimal variation between the observed and the CRU TS datasets. Also, some high inflection points were observed over some of the coastal grids.

The country-wide low RMD is corroborated by high positive Pearson's correlation between the two datasets, with correlation co-efficients ranging from 0.5 to 0.9 as shown in Figure 4.21b. This confirms a good agreement between the reconstructed grids and CRU TS precipitation data.

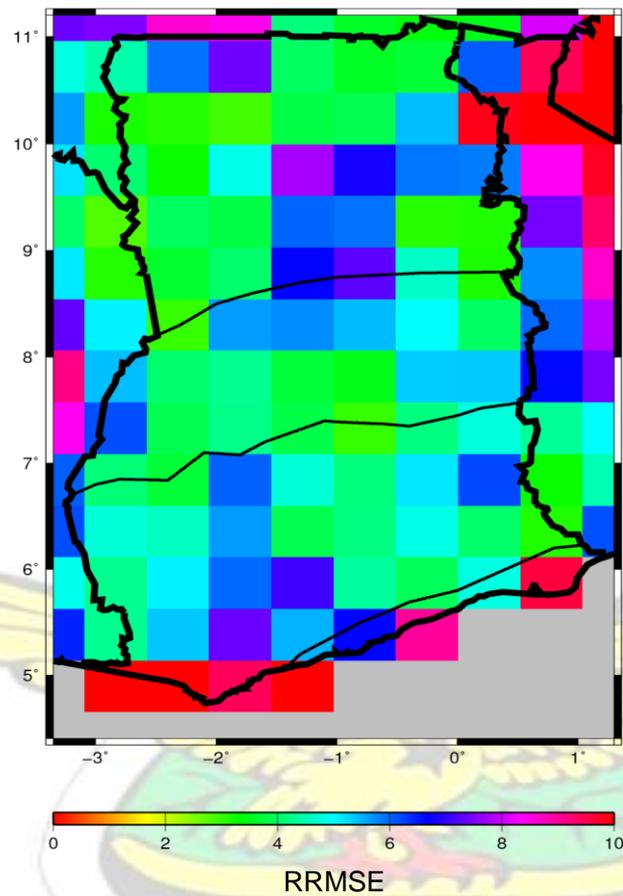


Figure 4.22: Relative Root Mean Square Error

Finally, RRMSE was computed to quantify the magnitude of variation between the two datasets at each grid. RRMSE values attained over the country ranged between 0 and 8, with inflection points located over some grids along the country's shoreline (see Figure 4.22). Low RRMSE values were observed over the interior parts of the country.

## CHAPTER 5

## Conclusion and recommendations

### 5.1 Conclusion

Monthly rainfall datasets from various meteorological (climatological, agrometeorological and synoptic) stations distributed over the four agro-ecological zones of Ghana, spanning a 33-year period (1980-2012), with inherent missing data have been reconstructed using various interpolation schemes. Regularized Expectation Maximization has been used in this study to estimate the missing data. RegEM was found to perform well for datasets with less and non-continuous data gaps. As such, a tolerance of 10 % datagap and non-continuous gaps were allowed in the missing data estimation. Absolute homogenization was performed to detect the various changepoints within observed data. These changepoints are supposed to be non-climate related discontinuities within the dataset. In the absence of station metadata, no reference series were used. The inhomogenized datasets were then detrended, per the magnitude of each changepoint within the time-series, using Quantile Matching Adjustments.

The reconstructed rainfall datasets were gridded at a high ( $0.25^{\circ} \times 0.25^{\circ}$ ) spatial resolution, both on monthly and seasonal timescales using the MSC algorithm with tensioning parameter. The grids revealed the all-year north-south migrations of the ITD which is clearly identified based on the latitudinal movement of the rains. From the gridded datasets, the Transition Zone was noted to have a longer growing season and as such, is the food hub of Ghana.

The gridded monthly and seasonal rainfall datasets were validated with Climatic Research Unit Time Series 3.22 (CRU TS 3.22) satellite data using three statistical tests (Pearson's Correlation, Relative Mean Difference and Root Mean Square Error). The satellite and reconstructed data showed a high agreement between them. Pearson's correlation coefficients were high (between the range of 0.5 to 0.9). Also, relative mean difference was between 0 to 0.3 which is low and thus, is an indication of minimal variation between the two datasets. The relative root mean square error also proved same with low values ranging from 0 to 8.

Although local factors influence daily rainfall variations, the overall seasonal variability of rainfall has been identified in this work to be linked with the migration of the ITD.

Finally, a highly-resolved rainfall climatological database has been developed from the GMet dataset over the entire country which will now serve as precursor for quality countrywide climate impact studies.

## 5.2 Recommendations

### 5.2.1 Recommendation for Policy

With regards to the high correlation levels of the reconstructed data and the CRU TS gridded data, the latter can be used as a substitute for the former when conducting climate-impact studies in the absence of observed data. By the statistical analysis,

CRU TS satellite data, if used for climate-impact studies will give a representation close to that of the gauge measurements.

Since the second trimester (MAM) is the rainfall onset period, it behoves on farmers to start their planting within this period since that is the set-in phase of the rains. Planting within this trimester, will afford crops quality time of sufficient rains to produce better yields.

Moreover, the third trimester (JJA) was found to be extremely wet and that is a possible period of climate extremes, such as flood. Precautionary measures (both on domestic and national scales) should be put in place to ameliorate the impact of these extremes that are likely to occur within this trimester.

Furthermore, the Transition Zone is characterized by a longer growing season and as such is a zone that could really boost agricultural yields. Since agriculture is the main propeller of the Ghanaian economy, it is recommended that much more efforts be principally focused within the Transition Zone since that is the country's food hub.

### 5.2.2 Recommendation for Future Research

Deterministic approaches such as IDW can be used to estimate missing data at a particular station, provided the reference (neighbouring) stations have quality data. A suitable approach will be to use the RegEM or other non-deterministic methods to reconstruct the reference station's data and then merging it with the deterministic approach to estimate data over remote areas.

Also, I recommend that further studies be carried out using other climate variables (eg. temperature) so as to develop quality countrywide climate database to aid in quality impact studies.

Finally, I recommend that further hydroclimatic studies be carried out to ascertain the impact of the confluence of the Black and White Volta on the climatology of its surrounding grids.

# KNUST



# Bibliography

- Amekudzi, L. K., Bracher, A., Bramstedt, K., Rozanov, A., Bovensmann, H., and Burrows, J. P. (2007). Towards validation of SCHIAMACHY lunar occultation NO<sub>2</sub> vertical profiles. *Advances in Space Research*, 41(2008):1921–1932.
- Amekudzi, L. K., Yamba, E., Preko, K., Asare, E. O., Aryee, J., Baidu, M., and Codjoe, N. A. S. (2015). Variabilities in rainfall onset, cessation and length of rainy season for the various agro-ecological zones of Ghana.
- Atkinson, P. M. and Lloyd, C. D. (1998). Mapping precipitation in Switzerland with ordinary and indicator kriging. *J. Geogr. Inf. Decis. Anal.*, 2:65–76.
- Barnet, T. P., Santer, B., Jones, P. D., Bradley, R. S., and Briffa, K. R. (1996). Estimates of low frequency natural variability in near-surface air temperature. *Holocene*, 6:255–263.
- Biederlack, L. and Rivers, J. (2009). Comprehensive food security and vulnerability analysis (CFSVA): Ghana. Technical report, United Nations World Food Programme.
- Blair, T. (2012). Techniques involved in developing the Australian climate observations reference network - surface air temperature (ACORN-SAT) dataset. Technical Report 049, Centre for Australian Weather and Climate Research.
- Bo, C., Schmith, T., and Thejll, P. (2009). A surrogate ensemble study of climate reconstruction methods: stochasticity and robustness. *J. Climate*, 22:951–976.

- Brohan, P., Kennedy, J. J., Harris, I., Tett, S. F. B., and Jones, P. D. (2006). Uncertainty estimates in regional and global observed temperature changes: A new dataset from 1850. *J. Geophys. Res.*, 111.
- Cervený, R. S. and Balling, R. C. (1998). Weekly cycles of air pollutants, precipitation and tropical cyclones in the coastal NW Atlantic region. *Nature*, 394(6693):561–563.
- Costa, A. C. and Soares, A. (2009). Homogenization of climate data: Review and new perspectives using geostatistics. *Math Geosci*, 41:291–305.
- Couvreux, F., Guichard, F., Bock, O., Lafore, J. P., and Redelsperger, J. L. (2010). Synoptic variability of the monsoon flux over West Africa prior to the onset. *Q. J. R. Meteorol. Soc.*, 136:160–174.
- Easterling, D. R. (1997). Maximum and minimum temperature trends for the globe. *Science*, 277:364–367.
- Eccel, E., Cau, P., and Ranzi, R. (2012). Data reconstruction and homogenization for reducing uncertainties in high-resolution climate analysis in Alpine regions. *Theor Appl Climatol*.
- Feng-Wen, C. and Chen-Wuing, L. (2012). Estimation of spatial rainfall distribution using inverse-distance weighting (idw) in the middle Taiwan. *Paddy Water Environ*, 10:209–222.
- Flamant, C., Knippertz, P., Parker, D. J., Chaboureaud, J. P., Lavaysse, C., AgustiPanareda, A., and Kergoat, L. (2009). The impact of a mesoscale convective

system cold-pool on the northward propagation of the inter-tropical discontinuity over West Africa. *Q. J. R. Meteorol. Soc.*, 135:139–159.

Hanssen-Bauer, I., Fjörland, E. J., Tveito, O. E., and Nordli, P. Å. (1997). Estimating regional precipitation trends – comparison of two methods. *Nordic Hydrol.*, 28:21–36.

Hastenrath, S. (1995). *Climate dynamics of the tropics*. Kluwer Acad.

Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D., and New, M. (2008). A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *Journal of Geophysical Research*, 113(20).

Jeffrey, S. J., Carter, J. O., Moodie, K. B., and Beswick, A. R. (2001). Using spatial interpolation to construct a comprehensive archive of Australian climate data. *Environ Model Softw*, 16(4):309–330.

Kopec, J. R. (2010). An alternative method for the construction of Thiessen polygons. *The Professional Geographer*, 15(5):24–26.

Lanzante, J. R. and Klein, S. A. (2002). Temporal homogenization of monthly radiosonde temperature data. *Journal of Climate*, 16:224–240.

Lei, Z., Guo-Yu, R., Yu-Yu, R., Ai-Ying, Z., Zi-Ying, C., and Ya-Qing, Z. (2014). Effect of data homogenization on estimate of temperature trend: a case of Huairou station in Beijing Municipality. *Theor Appl Climatol*, 115:365–373.

Li, J. and Heap, A. D. (2008). Spatial interpolation methods: a review of environmental scientists. *Geoscience Australia*.

- Li, J. and Heap, A. D. (2011). A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. *Ecological Informatics*, 6(3-4):228–241.
- Lothon, M., Said, F., Lohou, F., and Campistron, B. (2008). Observation of the diurnal cycle in the low troposphere over West Africa. *Mon. Weather Rev.*, 136:3477–3500.
- Lu, G. Y. and Wong, D. W. (2008). An adaptive inverse-distance weighting spatial interpolation technique. *Comput Geosci*, 34(9):1044–1055.
- Manzanas, R., Amekudzi, L. K., Preko, K., Herrera, S., and Gutierrez, J. M. (2014). Precipitation variability and trends in Ghana: An intercomparison of observational and reanalysis products. *Climate Change*, 124:805–819.
- Mathugama, S. C. and Peiris, T. S. G. (2011). Critical evaluation of dry spell research. *International Journal of Basic and Applied Sciences*, 11(6):153–160.
- Mengistu Tsidu, G. (2012). High resolution monthly rainfall database for Ethiopia: Homogenization, Reconstruction and Gridding. *Journal of Climate*, 25(24):8422–8443.
- Menne, J. M. and Williams Jr., C. N. (2008). Homogenization of temperature series via pairwise comparisons. *Journal of Climate*, 22:1700–1717.
- New, M., Parker, D. E., Martin, S., and Rigor, I. G. (1999). Surface air temperature and its changes over the past 150 years. *Rev. Geophys.*, 37:173–199.
- Ofori-Sarpong, E. (2001). Impact of climate change on agriculture and farmers coping strategies in the Upper East region of Ghana. *West Afr J Appl Ecol*, 2:21–35.

- Peterson, T. C., Easterling, D. R., Karl, T. R., Groisman, P., Nicholls, N., Plummer, N., Torok, S., Auer, I., Boehm, R., Gullett, D., Vincent, L., Heinof, R., Tuomenvirta, H., Mestre, O., Szentimrey, T., Salinger, J., FÅrland, E. J., Hanssen-Bauer, I., Alexandersson, H., Jones, P., and Parker, D. (1998). Homogeneity adjustments of in situ atmospheric climate data: A review. *International Journal of Climatology*, 18:1493–1517.
- Pidwirny, M. (2008). *Introduction to the Hydrosphere (e). Cloud formation Processes*, chapter 8.
- Price, D. T., McKenney, D. W., Nalder, I. A., Hutchison, M. F., and Kesteven, J. L. (2000). A comparison of two statistical methods for interpolation of Canadian monthly mean climate. *Agric Meteorol*, 101:81–94.
- Quansah, E., Amekudzi, L. K., Preko, K., Aryee, J., Boakye, R. O., Boli, D., and Salifu, M. R. (2014). Empirical models for estimating global solar radiation over the Ashanti Region of Ghana. *Journal of Solar Energy*, 2014.
- Reto, K., Reinhard, F., Claudia, T., Jan, C., and Gerald, A. M. (2010). Challenges in combining projections from multiple climate models. *J. Climate*, 23:2739–2758.
- Schneider, T. (2001). Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values. *J. Climate*, 14:853–871.
- Segond, M. L., Neokleous, N., Makropoulos, C., Onof, C., and Maksimovic, C. (2007). Simulation and spatio-temporal disaggregation of multi-site rainfall data for urban drainage applications. *Hydrol. Sci. J.*, 52(5):917–935.

- Singh, N. and Ranade, A. (2010). The wet and dry spells across india during 19512007. *Hydrometeor*, 11.
- Smerdon, J. E., Kaplan, A., Chang, D., and Evans, M. N. (2010). A pseudoproxy evaluation of the CCA and RegEM methods for reconstructing climate fields of the last millenium. *J. Climate*, 23(18):4856–4880.
- Smith, W. H. F. and Wessel, P. (1990). Gridding with continuous curvature splines in tension. *Geophysics*, 55(3):293–305.
- Stanturf, J. A., Warren Jr., M. L., Charnley, S., Polasky, S. C., Goodrick, S. L., Armah, F., and Nyako, Y. A. (2011). Climate change vulnerability and adaptation assessment. Technical report, United States Agency for International Development (USAID).
- Sultan, B. and Janicot, S. (2003). The West African Monsoon Dynamics part ii: Preonset and onset of the summer monsoon. *Journal of Climate*, 16:3407–3427.
- Wang, X., Chen, H., Wu, Y., Feng, Y., and Pu, Q. (2010). New techniques for detection and adjustment of shifts in daily precipitation data series. *J. Appl. Meteor. Climatol.*, 49:2416–2436.
- Yakowitz, S. (2008). Nearest-Neighbour methods for time series analysis. *Journal of Time Series Analysis*, 8(2):235–247.
- Yamba, E. (2010). Investigation of seasonal and annual variability in rainfall pattern over the transition and northern savanna belts of Ghana. Master’s thesis, Kwame Nrumah University of Science and Technology (KNUST).
- Yeh, H. C., Chen, Y. C., Wei, C., and Chen, R. H. (2011). Entropy and kriging approach to rainfall network design. *Paddy Water Environ*, 9:343–355.

## Appendix A

### Appendix

Table A.1: Percentage of missing data for meteorological stations in the Northern Zone.

Station	Missing Data [%]	Station	Missing Data [%]
Babile	1.77	Binduri	1.77
Bolgatanga-Agromet	0.63	Bole	0.00
Cherponi	6.99	Chira	2.53
Daboya	5.80	Damango	0.63
Dorimon	3.16	Lawra	1.77
Gambaga	7.34	Garu	0.88
Manga-Bawku	1.14	Kugri	3.03
Navrongo	0.38	Pong-Tamale	2.53
Salaga	1.52	Sampa	2.40
Walewale	1.39	Worawora	7.43
Ve	1.14	Funsi	6.33

Table A.2: Percentage of missing data for meteorological stations in the Transition Zone.

Station	Missing Data [%]	Station	Missing Data [%]
Bui	2.53	Kintampo	0.38
Sunyani	0.25	Forifori	1.77
Goaso	2.27	Nsoatre	5.69
Dormaa-Ahenkro	1.82	Bechem	1.39
Prang	1.39	Atebubu	0.76
Ejura	0.63	Berekum	2.15

Table A.3: Percentage of missing data for meteorological stations in the Forest Zone.

Station	Missing Data [%]	Station	Missing Data [%]
KNUST	6.06	Aburi-Parks	0.51
Agogo	0.76	Asuansi	0.25
Akropong-Akwapim	0.25	Akropong-Wassaw	1.01
Asamankese	0.51	Begoro	2.02
Kibi	3.26	Effiduasi	0.38
Akuse	0.13	Akosombo	1.64
Jasikan	5.17	Kade	1.39
Bekwai-Ashanti	1.14	Mampong	1.77
Enchi	1.81	Bobiri	1.77
Pankese	0.13	Nkawkaw	2.27
Asesewa	0.88	Konongo	0.40
Adidome	2.46	Akokoaso	4.80

Amedzofe	1.14	Anyinasi	3.16
Atieku	1.14	Avedotoe	1.04
Asantekrom-Dodi	3.90	Benso	0.76
Bogoso	0.76	Boso	0.63
Bunso	2.40	Dunkwa	0.51
Hohoe	1.64	Inchaban	1.64
Half-Assini	0.76	Komenda	0.13
Kpedzeglo-Mafi	2.78	Kpeve	1.14
Kusi	0.13	Kwanyako	2.53
Mpraeso	4.42	Nkroful	1.26
Prestea	2.27	Princes-Town	1.77
Somanya	2.65	Tarkwa	1.14
Tsito	3.79	Twifo-Praso	1.14

Table A.4: Percentage of missing data for meteorological stations in the Coastal Zone.

Station	Missing Data [%]	Station	Missing Data [%]
Pokuase	0.13	Pomadze	0.38
Winneba	0.38	Agona-Swedru	3.16
Breman-Asikuma	0.38	Brimso	0.88
Afienea	2.65	Afife-Weta	3.54
Nsawam	0.13	Sogakope	7.95

## Appendix B

### Appendix

```

RegEM MatLab Code by Schneider (2001) function [X, M, C, Xerr] =
regem(X, options) error(nargchk(1, 2, nargin)) % check number of
input arguments if ndims(X) > 2, error('X must be vector or 2-D
array.');
```

end % if X is a vector, make sure it is a column vector (a  
single variable) if length(X)==prod(size(X)) X = X(:); end

```

[n,p] = size(X);

% number of degrees of freedom for estimation of covariance matrix dofC
= n - 1; % use degrees of freedom correction

% ===== process options =====
```

```

if nargin ==1 | isempty(options)

fopts = []; else fopts =

fieldnames(options); end

% initialize options structure for regression modules

optreg = [];

regpar_given = 0; if strcmp('regress', fopts)

regress = lower(options.regress); switch regress

case 'mridge', 'iridge' if strcmp('regpar', fopts)

regpar_given = 1; if ischar(options.regpar)

error('Regularization parameter must be a

number') else optreg.regpar = options.regpar;

regress = 'mridge'; end end

case 'ttls' if isempty(strcmp('regpar',

fopts)) trunc_criterion = 'ne08'; elseif

ischar(options.regpar) trunc_criterion

= lower(options.regpar); else

regpar_given = 1; trunc =

min([options.regpar, n-1, p]); end if

strcmp('neigs', fopts) neigs =

options.neigs;

else

neigs = min(n - 1,p); end otherwise

error(['Unknown regression method ',

regress]) end

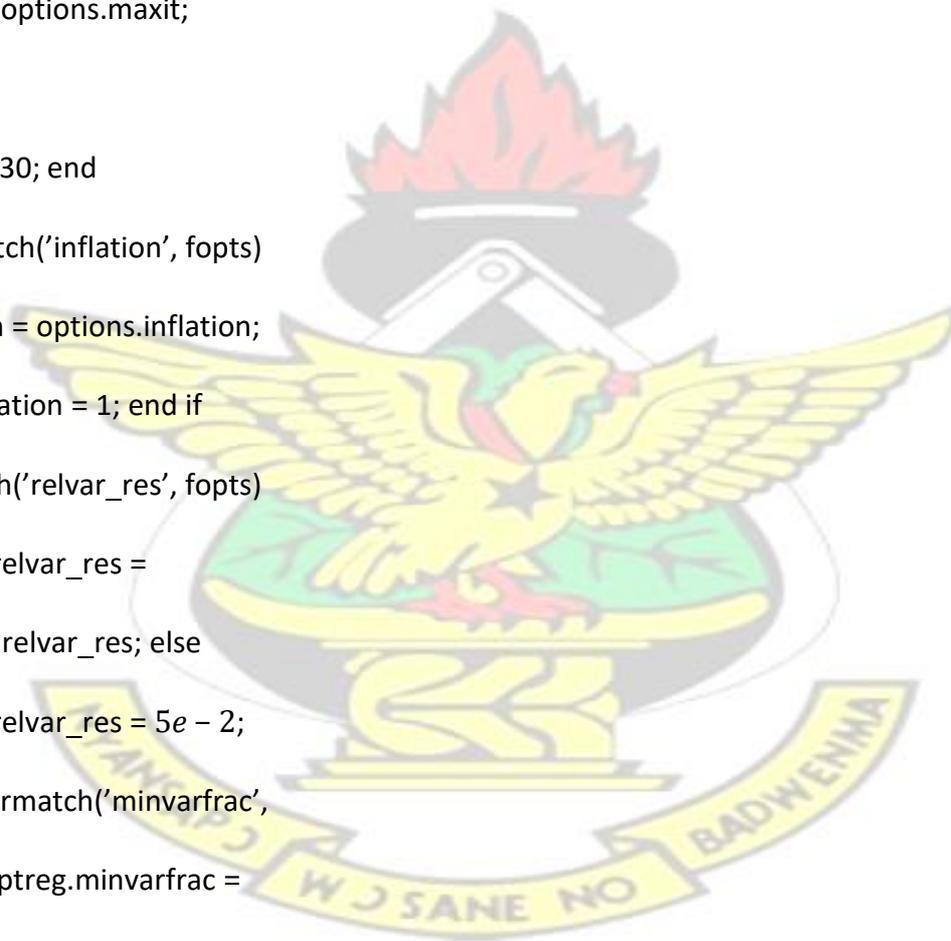
```

```

else regress =
'mridge'; end
if strmatch('stagtol', fopts)
stagtol = options.stagtol;
else
stagtol = 5e - 3; end if
strmatch('maxit', fopts)
maxit = options.maxit;
else
maxit = 30; end
if strmatch('inflation', fopts)
inflation = options.inflation;
else inflation = 1; end if
strmatch('relvar_res', fopts)
optreg.relvar_res =
options.relvar_res; else
optreg.relvar_res = 5e - 2;
end if strmatch('minvarfrac',
fopts) optreg.minvarfrac =
options.minvarfrac; else
optreg.minvarfrac = 0; end if
strmatch('disp', fopts);
dispon = options.disp; else

```

# KNUST



```

dispon = 1; end if

strmatch('Xmis0', fopts);

Xmis0_given= 1; Xmis0 = options.Xmis0; if

any(size(Xmis0) ~= [n,p]) error('OPTIONS.Xmis0 must

have the same size as X.') end

else

Xmis0_given= 0; end

if strmatch('C0', fopts);

C0_given = 1;

C0 = options.C0;

if any(size(C0) ~= [p, p]) error('OPTIONS.C0 has

size incompatible with X.') end

else

C0_given = 0; end if

strmatch('Xcmp', fopts);

Xcmp_given = 1; Xcmp = options.Xcmp; if

any(size(Xcmp) ~= [n,p]) error('OPTIONS.Xcmp must

have the same size as X.') end sXcmp = std(Xcmp);

else

Xcmp_given = 0; end

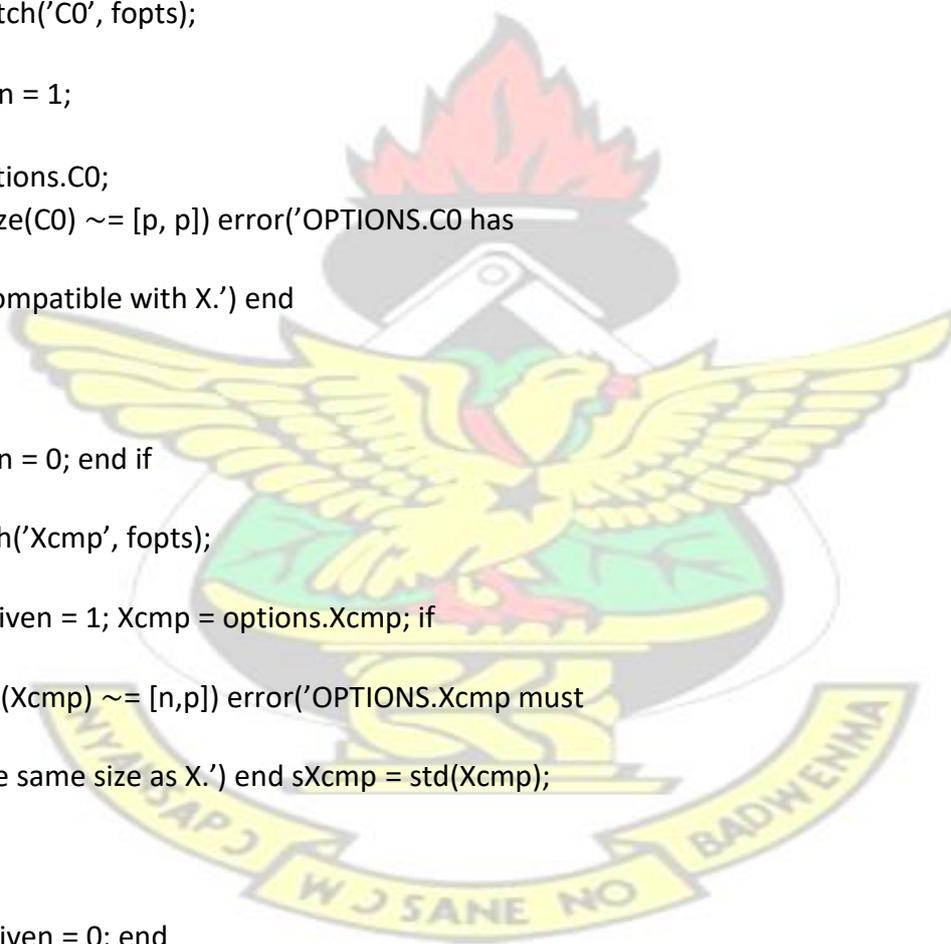
% get indices of missing values and initialize matrix of imputed values

indmis = find(isnan(X)); nmis = length(indmis);

if nmis == 0

```

KNUST



```

warning('No missing value flags found.')
```

return % no missing values end

```

[jmis,kmis] = ind2sub([n,p],indmis);
```

Xmis = sparse(jmis, kmis, NaN, n, p); % matrix of imputed values Xerr = sparse(jmis, kmis, Inf, n, p); % standard error imputed vals.

% for each row of X, assemble the column indices of the available  
% values and of the missing values kavlr = cell(n,1); kmisr = cell(n,1);  
for j=1:n kavlrj = find( isnan(X(j,:))); kmisrj = find(isnan(X(j,:))); end if  
dispon disp(sprintf('nREGEM:')) disp(sprintf('Percentage of values  
missing: %5.2f', nmis/(n\*p)\*100)) disp(sprintf('Stagnation  
tolerance: %9.2e', stagtol)) disp(sprintf('Maximum number of  
iterations: %3i', maxit)) if (inflation ~ = 1) disp(sprintf('Residual (co-  
)variance inflation: %6.3f ', inflation)) end if Xmis0\_given &  
CO\_given  
disp(sprintf(['Initialization with given imputed values and' ...  
' covariance matrix.'])) elseif CO\_given  
disp(sprintf(['Initialization with given covariance' ...  
' matrix.'])) elseif Xmis0\_given disp(sprintf(['Initialization  
with given imputed values.'])) else  
disp(sprintf('Initialization of missing values by mean  
substitution.)) end switch regress case 'mridge'  
disp(sprintf('One multiple ridge regression per record:'))  
disp(sprintf('==> one regularization parameter per  
record.)) case 'iridge' disp(sprintf('One individual ridge

```

regression per missing value:')) disp(sprintf('==> one
regularization parameter per missing value.')) case 'ttls'
disp(sprintf('One total least squares regression per
record.)) end switch regress case 'mridge', 'iridge' if
regpar_given disp(sprintf('Fixed regularization
parameter: %9.2e', optreg.regpar)) end
case 'ttls' if regpar_given disp(sprintf('Fixed truncation parameter:
%5i', trunc)) else disp(sprintf(['Truncation choice criterion: ',
upper(trunc_criterion)])) end end

if Xcmp_given disp(sprintf(['Iter mean(peff)2 |X-
Xcmp|/std(Xcmp)' ...
'|D(Xmis)|/|Xmis|']) else
disp(sprintf(['Iter mean(peff)2 |D(Xmis)| '
...
'|D(Xmis)|/|Xmis|'])
) end end

% initial estimates of missing values if Xmis0_given
% substitute given guesses for missing values
X(indmis) = Xmis0(indmis);

[X,M] = center(X); % center data to mean zero

else

[X,M] = center(X); % center data to mean zero

X(indmis) = zeros(nmis, 1); % fill missing entries with zeros

end if CO_given

```

```

C = C0;

else

C = X'*X / dofC; % initial estimate of covariance matrix

end it = 0; rdXmis = Inf; while (it < maxit & rdXmis >
stagtol) it = it + 1;

% initialize for this iteration ...

CovRes = zeros(p,p); % ... residual covariance matrix peff_ave
= 0; % ... average effective number of variables

% scale variables to unit variance D = sqrt(diag(C));
const = (abs(D) < eps); % test for constant variables
nconst = const; if sum(const) ~= 0 % do not scale
constant variables D = D .* nconst + 1*const; end

X = X . repmat(D', n, 1);

% correlation matrix

C = C . repmat(D', p, 1) . repmat(D, 1, p);

if strcmp(regress, 'ttls')

% compute eigendecomposition of correlation matrix

[V,d] = peigs(C,neigs);

% compute truncation selection criteria if
needed if ~regpar_given trunc_pars = [0:

length(d)-1];

[mdl,ne08,aic,aicc] = pca_truncation_criteria(d,p,trunc_pars,n); end

end for j=1:n % cycle over records pm = length(kmisrj); % number
of missing values in this record

```

```

if pm > 0
pa = p - pm; % number of available values in this record

% regression of missing variables on available variables

switch regress case 'mridge'

% one multiple ridge regression per record [B,S,h,peff] =
mridge(C(kavlrj,kavlrj),...
C(kmisrj,kmisrj), ...
C(kavlrj,kmisrj), n-1, optreg); peff_ave = peff_ave + peff*pmnmis; %
add up eff. number of variables dofS = dofC - peff; % residual degrees
of freedom

% inflation of residual covariance matrix
S = inflation * S;

% bias-corrected estimate of standard error in imputed
values Xerr(j, kmisrj) = dofC/dofS * sqrt(diag(S))'; case
'iridge'

% one individual ridge regression per missing value in this record [B,S,h,peff] =
iridge(C(kavlrj,kavlrj),...
C(kmisrj,kmisrj), ...
C(kavlrj,kmisrj), n-1, optreg); peff_ave = peff_ave + sum(peff)nmis; %
add up eff. number of variables dofS = dofC - peff; % residual degrees
of freedom

% inflation of residual covariance matrix
S = inflation * S;

% bias-corrected estimate of standard error in imputed values

```

```

Xerr(j, kmisrj) = ( dofC * sqrt(diag(S)) . dofS)';
case 'ttls'

% truncated total least squares if ~regpar_given imax = max(find(trunc_pars
<= min(pa, length(d)-1))); [dum,imin] = eval(['min(0,trunc_criterion,0 (1 :
imax))']); trunc = trunc_pars(imin); end peff_ave = peff_ave + trunc *
pmnmis; % add up eff. number of variables
[B,S] = pttls(V,d,kavlrj,kmisrj,trunc); dofS = dofC -
trunc; % residual degrees of freedom

% inflation of residual covariance matrix

S = inflation * S;

% bias-corrected estimate of standard error in imputed
values Xerr(j, kmisrj) = dofCdofS * sqrt(diag(S))'; end

% missing value estimates

Xmis(j, kmisrj) = X(j, kavlrj) * B;

% add up contribution from residual covariance
matrices CovRes(kmisrj, kmisrj) = CovRes(kmisrj,
kmisrj) + S; end end % loop over records

% rescale variables to original scaling

X = X .* repmat(D', n, 1);

Xerr = Xerr .* repmat(D', n, 1);

Xmis = Xmis .* repmat(D', n, 1);
C = C .* repmat(D', p, 1) .* repmat(D, 1, p);

CovRes = CovRes .* repmat(D', p, 1) .* repmat(D, 1, p);

% rms change of missing values dXmis =
norm(Xmis(indmis) - X(indmis)) sqrt(nmis);

```

```

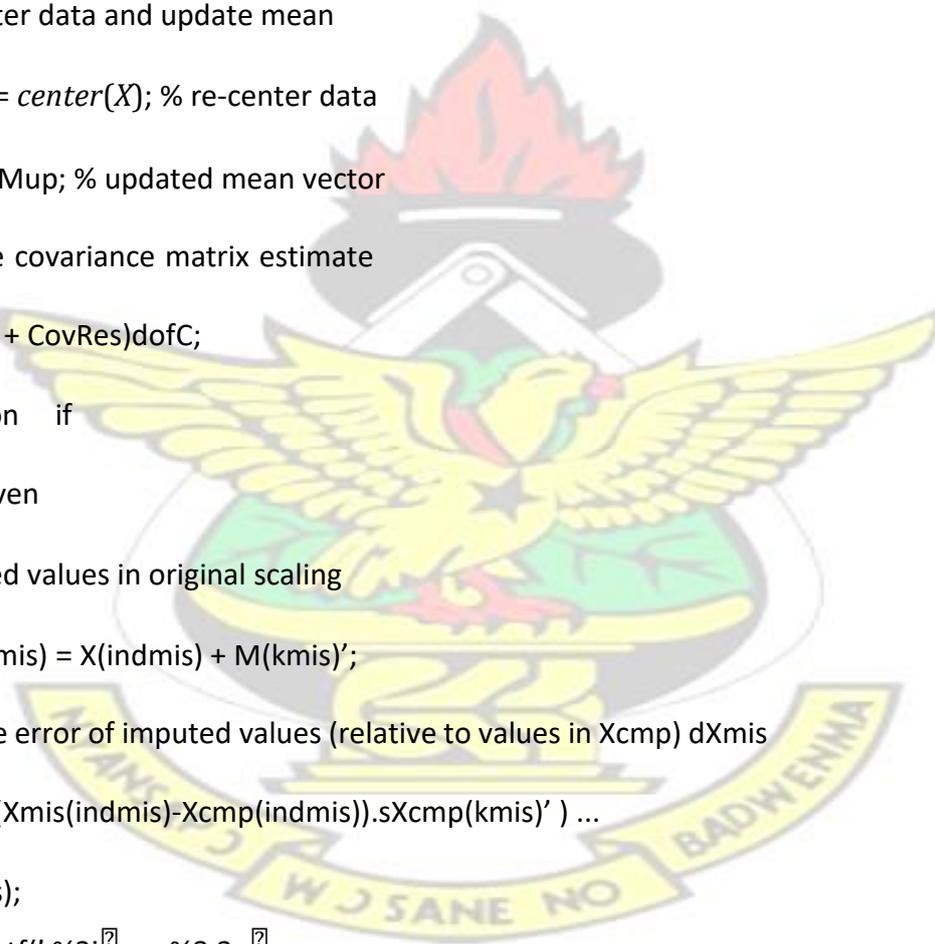
% relative change of missing values nXmis_pre =
norm(X(indmis) + M(kmis)') sqrt(nmis); if nXmis_pre <
eps rdXmis = Inf;
else
rdXmis = dXmis nXmis_pre; end

% update data matrix X
X(indmis) = Xmis(indmis);

% re-center data and update mean
[X,Mup] = center(X); % re-center data
M = M + Mup; % updated mean vector
% update covariance matrix estimate
C = (X'*X + CovRes)/dofC;
if dispon if
Xcmp_given
% imputed values in original scaling
Xmis(indmis) = X(indmis) + M(kmis)';
% relative error of imputed values (relative to values in Xcmp) dXmis
= norm( (Xmis(indmis)-Xcmp(indmis)).sXcmp(kmis)' ) ...
sqrt(nmis);
disp(sprintf(' %3i      %8.2e
%10.3e      %10.3e', ... it, peff_ave,
dXmis, rdXmis)) else disp(sprintf(' %3i
%8.2e      %9.3e      %10.3e', ...

```

KNUST



```
it, peff_ave, dXmis, rdXmis)) end end %  
display of diagnostics end % EM iteration  
% add mean to centered data matrix  
X = X + repmat(M, n, 1);
```

# KNUST



## Appendix C

### Appendix

Generic Mapping Tools (GMT) Code for Gridding

```
#!/bin/bash

# Define the names of the input and output files
name= <name>; folder= New_Plot2 out=
$folder/$name.eps; #seis_data= jan.dat topo=
Grid/trial.grd ; ext= extract_data.xyz pdf=
$folder/$name.pdf; png= $folder/$name.png
mask= mask.dat if [ -s $out ] || [ -s $pdf ] ; then
rm $out ; rm $pdf
fi

# Define your area -R-3.25/1.75/3.75/12.25 north= 11.20; south= 4.40;
east= 1.3; west= -3.35 tick= '-Ba1f1/a1f1.: '$name':WSen'; proj= '-JM6i';
palette= '-Cjeff1.cpt' GMT psbasemap -R$west/$east/$south/$north $proj
$tick -Y5.0 -P -K > $out awk 'if (NR > 0) print $2, $3, $4' Stat_Loc.dat > dat1
awk 'print $1, $2, $3 ' dat1 > $ext

GMT blockmean -R$west/$east/$south/$north -I0.25 -V $ext > 1_$ext

GMT surface 1_$ext -R$west/$east/$south/$north -I0.25 -G$topo -V
GMT psscale -D3i/-0.9i/6i/0.15ih $palette -B:mm: -O -K » $out

GMT grdimage $topo -R -J -O -K $palette » $out

GMT pscoast -R$west/$east/$south/$north $proj -O $tick -A12000 -Di -N11/3p,black
-W5p,black -P -K » $out

GMT psxy -R -JM -W10.0,black -O -K « EOF» $out
```

-2.5 8.2 ; -2.3 8.3 ; -2.0 8.5

-1.7 8.6 ; -1.3 8.7 ; -1.0 8.75

-0.3 8.79 ; 0.45 8.8

EOF

GMT psxy -R -JM -W10.0,black -O -K « EOF» \$out

-3.2 6.7 ; -3.0 6.8 ; -2.8 6.85

-2.4 6.84 ; -2.1 7.1 ; -1.8 7.08

-1.6 7.2 ; -1.1 7.4 ; -1.0 7.39

-0.8 7.38 ; -0.6 7.37 ; -0.4 7.35

0.0 7.45 ; 0.2 7.52 ; 0.52 7.57

EOF

GMT psxy -R -JM -W10.0,black -O « EOF» \$out

-1.42 5.07 ; -1.3 5.2 ; -0.8 5.5

-0.6 5.59 ; -0.4 5.69 ; 0.0 5.8

0.8 6.2 ; 1.05 6.23

EOF

convert \$out \$pdf ; convert \$out \$png gv

\$out &

KNUST

