ASSESSING THE SKILL OF THE SEASONAL RAINFALL PREDICTION OVER THE GREATER HORN OF AFRICA USING GLOBAL MODELS AS MULTIMODEL ENSEMBLE

GEORGE OTIENO
I56/73363/2012

DEPARTMENT OF METEOROLOGY
UNIVERSITY OF NAIROBI
BOX 30197-00100,
NAIROBI, KENYA.

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF MASTER OF SCIENCE IN METEOROLOGY

2013

DECLARATION

This dissertation is my original work and has not been presented for examination in any other University:

George Otieno

Signature……...

Date……

This dissertation has been submitted for examination with our approval as University Supervisors:

Prof. Laban Ogallo

Signature……

Date ……...

Dr. Joseph Mutemi

Signature……

Date……

Dr. Franklin Opijah

Signature……

Date……

DEDICATION

To my beloved wife Juliet and parents Mr. and Mrs. Oluoch for their wise counsel, advice, encouragement and spiritual inspiration they accorded me during the entire exercise.

ABSTRACT

The usefulness and limitations in seasonal forecasts are due to uncertainty inherent in the climate system. The reduction of errors in the forecasts systems increases the reliability of the forecasts. The improved seasonal rainfall prediction to reduce the climatic extreme events using dynamical models with fewer uncertainties is important to the socio-economic development of the Greater Horn of Africa (GHA).

In this study the overall objective of the study was to assess the skill and accuracy of the seasonal rainfall forecasting using global models as multi-model ensemble during October to December (OND) season over the study region. The data used in the study included the gridded rainfall data from Climate Research Unit, University of East Anglia (CRU) and hindcast data from eight Global Producing Center models (GPCs) for the period 1983 to 2001. The methodology employed included spatial analysis, correlation analysis, Model output Statistics (MoS), regression analysis, time series analysis, simple composite analyses, weighted average and categorical statistical skill score.

The spatial patterns of the individual models output from the models of Washington, Montreal, Melbourne and model from Centre for weather forecasting and climate studies (CPTEC) were closest to the observed rainfall patterns. The largest departure from observations in this season was observed in the northern and southern sectors of the GHA. The spatial distribution of rainfall anomalies of the observed and models output during extreme events showed that the ensemble models were able to simulate El Niño (1997) and La Niña (2000) years. The models were not able to capture the magnitude of the extreme events.

The skill of the ensemble model was higher than those of the individual member models in terms of its ability to capture the rainfall peaks during the El Niño Southern Oscillations phenomena (ENSO). The analysis for the correlation coefficients showed higher values for the ensemble model output than for the individual models over the Equatorial region (5°N to 5°S). Comparatively, the stations in the northern and southern sectors of the GHA had low skill. This is an indication that the models have better skill and accuracy over the Equatorial region.

In general, the skill of the models was relatively higher during the onset of the ENSO event and became low towards the decaying phase of ENSO period. Regarding the prediction of extreme low and high values, the models generally indicated the direction of the anomalies but such extremes were under-estimated or over- estimated in some cases.

ACKNOWLEDGMENTS

I take this opportunity to thank God for seeing me through to the completion of this work, His grace and mercies have been immeasurable. I would also like to express my greatest appreciation to the collaboration between ICPAC and University of Nairobi that granted me the opportunity to pursue a degree in Master of Science (Meteorology) through an academic scholarship.

Special thanks go to my Supervisors, Prof. Laban Ogallo, Director of ICPAC, Dr. Joseph Mutemi and Dr. Franklin Opijah of the University of Nairobi for their advice, direction and support they accorded to me during the study. I am grateful to the entire staff of the department of Meteorology for their support, counsel and direction since I joined the department in 2006 as an undergraduate student.

I acknowledge the warm relationship enjoyed during the entire Masters' program with my colleagues at ICPAC and fellow classmates at the Department of Meteorology for their academic and moral support which has been invaluable. I also acknowledge in a special way Joseph Karianjahi, Sabiiti Geoffrey and Jully Ouma for their guidance and motivation during the research.

Special thanks go to Prof. L. Ogallo, Director of ICPAC, for the academic mentorship he has provided during my entire time at ICPAC.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS AND THEIR ABBREVIATIONS

AGCM	Atmospheric General Circulation Model
ARPEGE	Action de Recherché Petite Echelle Grande Echelle
BCC	Beijing Climate Centre
CERFACS	Centre European de Recherche et de Formation Avancee en Calcul Scientifique
CFS	Climate Forecast System
CMIP5	Coupled Models Inter-comparison Project 5
COLA	Center for Ocean, Land and Atmospheric studies
CPC	Climate Prediction Center
CPI	Climate Prediction Index
CPTEC	Centre for weather forecasting and climate studies
CRU	Climate Research Unit
DEMETER	Development of European Multimodal Ensemble system for seasonal to inter-annual prediction
DMI	Dipole Mode Indices
EAC	East African Community
ECHAM	European Centre Hamburg Model
ECMWF	European Centre for Medium Range Weather Forecasts
EEMD	Ensemble Empirical Mode Decomposition technique
ENSE 2	Ensemble Two
ENSE1	Ensemble One
ENSO	El Niño Southern Oscillations
FAO	Food Agricultural Organisation
FAR	False Alarm Ratio
GCMs	Global Circulation Models
GDPFS	Global Data Processing and Forecasting Systems
GHA	Greater Horn of Africa
GHG	Green House Gases
GPCs	Global Producing Centers
HSS	Heidke Skill Score
ICPAC	Inter-governmental authority and development Climate Prediction and Application Centre
ICs	Initial Conditions
ICTP	International Center for Theoretical Physics
IGAD	Inter-Governmental Authority and Development
IOD	Indian Ocean Dipole
IPCC	Inter -Governmental Panel on Climate Change
IRI	International Research Institute for climate and society
ITCZ	Inter-Tropical Convergence Zone
JMA	Japan Meteorological Agency
KMA	Korean Meteorological Agency
KMD	Kenya Meteorological Department
LBCs	Lateral Boundary Conditions
LC- SVSLRF	Lead Centers for Standard Verification System of Long Range Forecast
LC-LRFMME	Lead Centre for Long Range Forecasts Multi-Modal Ensemble
LDCs	Least Developed Countries

LODYC	Laboratories d'Oceanographie Dynamique et de Climatologie
MAC	Modulated Annual Cycles
MAM	March, April and May
MJO	Madden Julian Oscillation
MME	Multi-Model Ensemble
MOM3	Modular Ocean Model version 3
MOS	Model Output Statistics
MOSES	Met Office Surface Exchange Scheme
MSL	Mean Sea Level
MSLP	Mean Sea Level Pressure
MSSS	Mean Square Skill Score
NARCCAP	North American Regional Climate Change Assessment Program
NEMO	Nucleus for European Modeling of the Ocean for ocean
NMHSs	National Meteorological and Hydrological Services
NWP	Numerical Weather Prediction
OGCM	Ocean General Circulation Models
OND	October, November and December
PoD	Probability of Detection
QBO	Quasi-Biannual Oscillation
RCM	Regional Climate Models
RCOFs	Regional Climate Outlook Forums
RegCM3	Regional Climate Model Version 3
RMIP	Regional climate Model Inter-comparison Project for Asia
RMSD	Root Mean Squared Deviation
RMSE	Root Mean Square Error
SAWS	South African Weather Services
SRES	Special Report on Emissions Scenarios
SST	Sea Surface Temperature
TCC	Temporal Correlation Coefficient
TDD	Traditional Dynamical Downscaling
UK-MET	United Kingdom-Meteological office
UN	United Nations
WCRP	World Climate Research Programme
WMO	World Meteorological Organisation
WRF	Weather Research and Forecasting

CHAPTER ONE

1.0 INTRODUCTION

1.1 Background Information

Most countries in the Greater Horn of Africa (GHA) belong to a category classified as the Least Developed Countries (LDCs) that are characterised by extreme poverty, where the average citizens live on less than one dollar a day. The region faces frequent catastrophic famines associated with recurrent hazards such as droughts, floods and other climatic extremes. These events often devastate most economic, social and environment systems in the region. Floods usually lead to the destruction of property, infrastructure and settlements, the loss of life, diseases among many other negative impacts. Drought which is the most common hazard on the other hand is associated with loss of society's livelihoods, negative national economic growth and many other socio-economic miseries. Some examples of adverse climatic impacts within the GHA are shown in Figure 1.

Figure 1: Pictures of impacts caused by climate extreme conditions within the GHA, (Source: UNISDR, 2012)

Floods and droughts affect the transport system, livestock, crops, buildings and investments. The Fourth Inter-governmental Panel on Climate Change report (IPCC, 2007) showed that climate change in the world including the GHA region is real. Climate change will lead to changes in the space-time patterns of climate extremes in the GHA with far reaching socio-economic implications. Availability of reliable and timely weather or climate information products and early warning advisories is absolutely crucial for addressing these challenges.

Over the years the IGAD Climate Prediction and Applications Centre (ICPAC), National Meteorological and Hydrological Services (NMHSs) of the region, regional universities, the World Meteorological Organisation (WMO) among others have been involved in various activities and research for monitoring, prediction and early warning of regional weather and climate. These initiatives have been reported by (IRI, 2005, ACMAD, 2010 and Ogallo, 2008) among others.

This research has not only led to the understanding of processes that are linked to regional weather and climate anomalies but also in the improvement of weather and climate forecasts at various time scales that has put a lot of global interest improving the predictability of regional climate especially at seasonal time scale. The most innovative approach has been through Regional Climate Outlook Forums (RCOFs), a process introduced by WMO for the development of consensus seasonal climate outlook to support regional efforts in reducing negative impacts of climate extremes (WMO, 2007, Ogallo, 2008 and Chavas, 2008).

Most of the probabilistic regional forecasts have many challenges and limitations. There are still limitations in the level of their skill; especially during periods when global sea surface temperature anomalies are neutral. This has posed a difficulty in application by most users of the seasonal forecasts. These limitations and challenges in the probabilistic forecasts over the GHA have been studied by (Timothy *et al.*, 2009).

Recently, the World Meteorological Organisation has designated several advanced climate centers worldwide as WMO Global Producing Centers (GPCs) for long range forecasts (Timothy *et al.*, 2009). The skill and accuracy of products from the Global Producing Centers (GPCs) over the GHA has not been determined using systematic and consistent methodologies, this formed the basis and focus of this study.

1.2 Problem Statement

Most of the economic activities within Africa are rain dependent. Rainfall is therefore the climatic factor of the greatest socio-economic significance for the GHA. The rainfall over the region is highly variable with extreme events resulting in droughts and floods with adverse effects to the communities. These extremes including intra-seasonal episodes like wet and dry spells are among the anomalous weather and climate regimes that pose risks in the GHA region. Droughts and floods are amongst the major hazards identified in the region that have the potential of causing negative impacts in the GHA region (ISDR, 2008).

Dynamical models have been identified as suitable tools for use in seasonal forecasts despite the errors that still pose uncertainties in the forecasts (Krishnamurti *et al.,* 2008). These errors are due to the complex nature of the climate processes and hence the models may not adequately simulate salient features of seasonal rainfall and its variability. However, model agreements with observations are still the only way to assign confidence in the models output. Getting relatively accurate forecasts and higher skill in predicting rainfall is an important accomplishment. The ensemble forecasting technique gives forecasts with less inherent uncertainties that are relatively more skillful.

1.3 Objectives of the Study

The overall objective of this study was to assess the skill and accuracy of forecasting seasonal rainfall using multi-model ensemble from Global Climate Producing centers during OND season. To achieve this main objective, several specific objectives were undertaken. These were to:

i. Determine the spatial and temporal distributions of observed and model output rainfall.
ii. Determine the skill and accuracy of individual climate models in predicting seasonal rainfall.
iii. Determine the accuracy and the skill of the multi-model ensemble.

1.4 Justification of the Study

Agriculture forms the backbone of the economy of the region that mainly depends on seasonal rainfall. However the high variability of seasonal rainfall associated with climate extremes affects agriculture. Therefore, knowledge about the quality of the seasonal forecasts would provide early warning information to farmers and other users of climate information for preparedness to alleviate

the devastating impacts associated with the climate extremes (Joliffe and Stephenson, 2003). This provision of early warning information has been made possible by use of dynamical models in prediction. The introduction of products from the GPCs in the region in the recent past has continued to encourage more detailed analyses of the African RCOFs using improved datasets and methodologies. Studies have proved that a single model may not adequately provide sufficient skill to the forecasts; hence the ensemble approaches have been confirmed to improve the skill and accuracy of the seasonal forecasts (Palmer *et al.,* 2004 and Wang *et al.,* 2008).

Evaluating the performance of the GPC models within the GHA will improve the RCOF processes by minimizing any bias in the forecast system, which should subsequently lead to improvements in the forecast methodology. Skillful forecasts would be useful for decision making in hydrology, agriculture, public health, and other sectors of the economy that are rain dependent. Application of a forecasting system that takes advantage of the best available multi-model forecasts and generates skillful forecasts over the GHA region would provide a useful contribution to the climate monitoring and early warning sector.

The research being undertaken by ICPAC and other regional institutions would accelerate the understanding of processes that are linked to regional weather and climate anomalies at various time scales. This would improve the predictability of regional climate. The use of GPC model output for regional development of consensus seasonal climate outlooks would strengthen the efforts being undertaken currently in the region in reducing negative impacts of climate extremes, hence improving the accuracy and reliability of the forecasts.

1.5 Domain of the Study

Figure 2 shows the domain of the study comprising the eleven countries of the Greater Horn of Africa (GHA) region. The region lies between latitudes 21°N to 12°S and longitudes 23.5°E to 52°E.

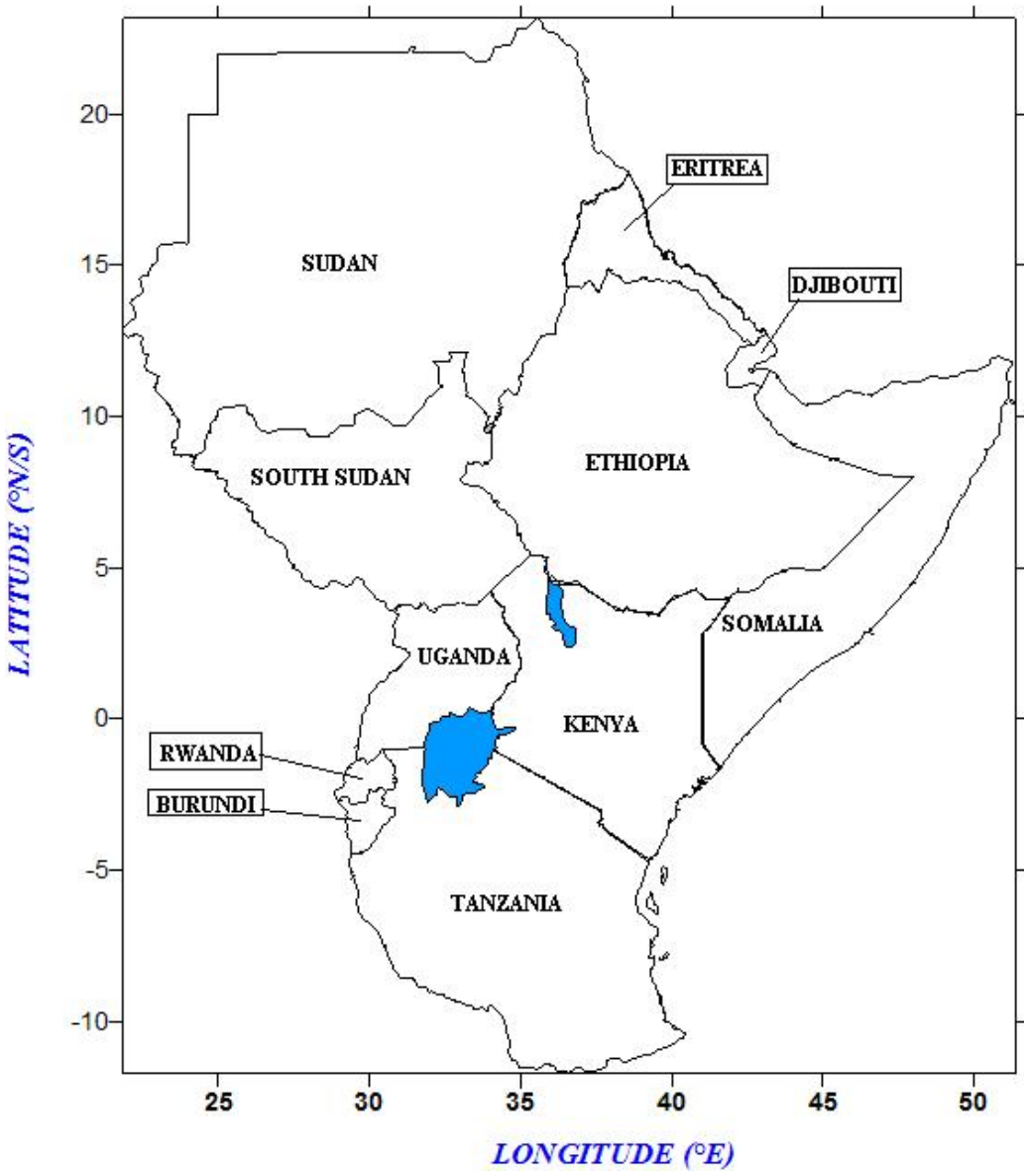

Figure 2: Domain of the study showing the countries within the GHA region

The subsections 1.5.1, 1.5.2 and 1.5.3 discuss briefly the physical features of the study domain, rainfall climatology and systems that influence rainfall over the study region.

1.5.1 Physical Features of the Study Domain

The GHA region has complex topographical features, which include the Ethiopian highlands to the Northeast and East African highlands to the southwest. The East African highlands include high mountains like Kenya (5199 metres), Kilimanjaro (5895 metres), Elgon (4321 metres), Aberdare Ranges (3999 metres) and the Mau escarpment (3098 metres). Some of these mountains like Kenya and Kilimanjaro have permanent glaciers at their top throughout the year which makes

them very special as potential indicators of regional large-scale, together with long-term climate fluctuations. The complex mountains are also the source of some major rivers of the region. The unique physical characteristics of the region are the water masses, for example, Lake Victoria and the Indian Ocean and the Western and Eastern Highlands which decline to a plain towards the Indian Ocean.

1.5.2 Rainfall Climatology of the Study Domain

Figure 3 is the topographic map of the study region depicting physical features of the Greater Horn of Africa. The GHA region is characterized by widely diverse climates ranging from desert to forest over relatively small areas. The presence of the water bodies generates land/sea and land/lake breezes, as a result of the water and land temperature contrasts, due to differential solar heating and radioactive cooling. Lake Victoria, for instance, has a strong circulation of its own with a semi-permanent trough, which migrates from land to lake and lake to land during the night and day respectively.

The influence of the lakes over the region, together with the nearby topography (Figure 3), enhances convection associated with thunderstorms in the region, the interaction with large-scale systems and the processes through which these local features modify the weather are not well understood. These pose unique numerical modeling challenges that can be addressed within the region.

The northern and southern part of GHA region experiences rainfall during June-July-August and September-October-November-December seasons. The processes and mechanisms believed to influence rainfall during these seasons are global Sea Surface Temperatures over the Indian and Atlantic Oceans. The processes are modulated by regional and local scale features including large inland lakes and the complex topographical patterns. The ENSO conditions over the tropical Pacific and the associated monsoonal winds are believed to influence rain over the North and Southern sectors of the GHA.

Rainfall pattern over the GHA is also modulated by the Great Rift Valley which runs from Ethiopia down to Mozambique. Most countries around the Great Rift Valley experiences cool highland climate due to increased elevation. For example the highland areas of Ethiopia are cool to cold. This causes abundant precipitation in these areas. Several regions receive rainfall throughout much

of the year, while in other areas precipitation is only seasonal. The climate characterizes cool, temperate, and hot conditions. The cool zone consists of the Western, Eastern, and central portions of the Northwestern plateau of Ethiopia. During the day, conditions are extremely hot and arid and average temperatures range from 27 degrees Celsius to 50 degrees Celsius. There is a wide variation in precipitation throughout the country due in large part to the differing elevations and seasonal changes in surrounding atmospheric pressure.

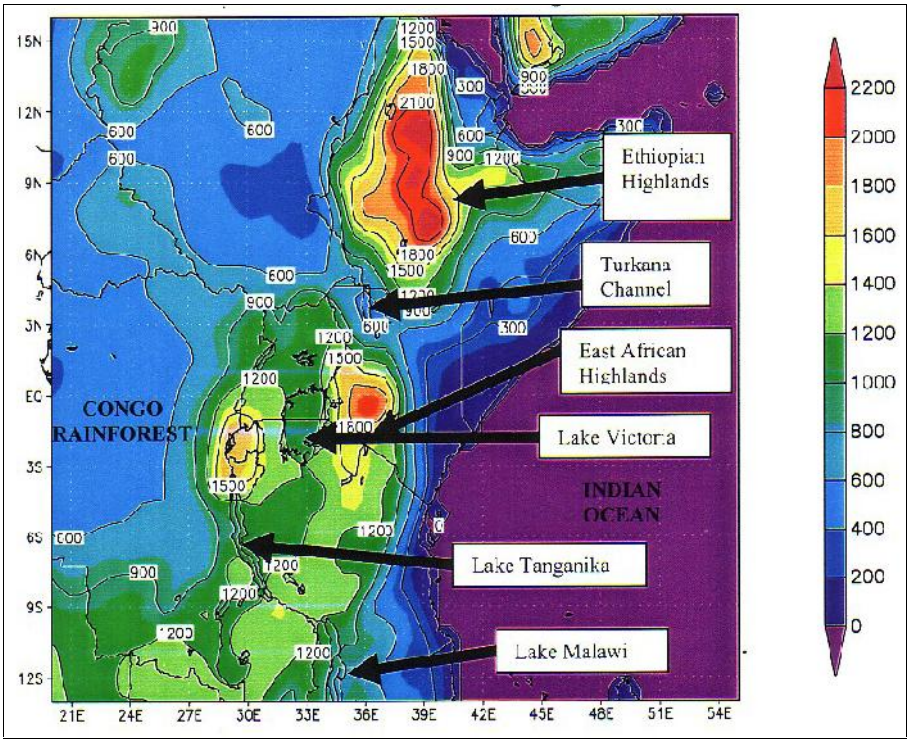

Figure 3: Topographic map depicting physical features of the Greater Horn of Africa. Elevation is in meters (Source: Bowden 2004)

The rainfall patterns over the Equatorial parts of the GHA are quite complex. Close to the large water bodies, substantial rainfall is received throughout the year. Over much of the sector, however, the major rainfall periods are concentrated within two peak seasons of March-May and October-December. Parts of the western and coastal regions also receive significant rainfall during the months of July-August.

1.5.3 Systems Influencing Rainfall Distribution over the Region

The ITCZ is the main synoptic system controlling the climate over the East Africa region (Okoola, 1998). The ITCZ is a key component of the global circulation system. Weather stations in the Equatorial region record precipitation up to 200 days each year, making the Equatorial and ITCZ zones the wettest on the planet (Nicholson, 2013).

The location of the ITCZ varies throughout the year and while it remains near the equator. The ITCZ over land ventures farther north or south than the ITCZ over the oceans due to the variation in land temperatures. The location of the ITCZ can vary by as much as 40° to 45° of latitude north or south of the equator based on the pattern of land and ocean. In Africa, the ITCZ is located just south of the Sahel at about 10°, dumping rain on the region to the south of the desert. There is a diurnal cycle to the precipitation in the ITCZ (Gitau, 2011). Clouds form in the late morning and early afternoon hours and then by 3 to 4 p.m., the hottest time of the day, convectional thunderstorms form and precipitation begins. These storms are generally short in duration.

Tropical storms, easterly waves, jet streams, the continental low level trough, extra-tropical weather systems, interactions between mesoscale flows and the large-scale monsoonal flows, teleconnection with global-scale climatic anomalies like those associated with SST, the Quasi-biennial Oscillation in the Equatorial lower stratospheric zonal wind (QBO), solar and lunar forcing and inter-seasonal 30-60 day Madden-Julian waves are some of the other factors that influence rainfall over the region. These climatic factors that modulate climate over the GHA have been studied by many researchers including (Indeje and Semazzi, 2000, Omeny *et al.,* 2008, Nyakwada *et al.,* 2009 and Nicholson, 2013)

1.5.4 Teleconnections

Apart from the ITCZ and other systems discussed in sections 1.5.3 believed to influence climate over GHA, teleconnections also play a key role in influencing climate over the GHA. By definition teleconnections are linkage between climate changes occurring at global scales to regional scales that are widely separated. Examples of the teleconnections systems are SSTs, QBO, ENSO, IOD and MJO amongst others. The Teleconnection with global-scale systems like the El Niño/Southern Oscillation (ENSO) and Indian Ocean Dipole and regional systems have been found to play a key role in modulating climate pattern over the GHA.

Rainfall variability over GHA is known to resonate with the coupled ocean-atmosphere phenomena of El Nino Southern Oscillation (ENSO) and the Indian Ocean Dipole (IOD). The rainfall variability is believed to show strong link to ENSO and IOD conditions (Nicholson, 2013). ENSO is a leading mode of tropical climate variability at inter-annual timescales and is characterized by sea surface temperature (SST) and surface pressure anomalies across the Pacific Ocean. It has a positive phase when El Niño phase occurs. This occurs when SSTs are warm in the eastern tropical Pacific Ocean contrasted by its negative La Niña phase when the SSTs cool. ENSO impacts on GHA are believed to occur in part via tropical Atlantic and Indian Ocean teleconnection.

The IOD is another coupled ocean-atmosphere mode with a positive (negative) phase characterized by warm (cool) SSTs over the western Indian Ocean and cool (warm) SSTs in the eastern Indian Ocean. Indian Ocean Dipole is reportedly rather a zonal mode instead, with the apparent dipole oscillation linked mostly to eastern Indian Ocean SSTs (Williams and Hanan, 2011).

The OND season have been found to have a positive correlation ENSO with high predictive skill based on sea surface temperature including the oceans Pacific, Indian, and Atlantic Ocean than one based on ENSO alone (Mutai and Neil, 2000).

CHAPTER TWO

2.0 LITERATURE REVIEW

This section contains literature on some of the past studies relevant to this study. Discussion on the gaps still lacking on the field of study have also been highlighted in the sections.

2.1 Introduction

Knowledge of the processes and/or systems that control East African rainfall is essential for the development of seasonal forecasting systems, which may mitigate the effects of flood and drought. Seasonal forecasting is giving most probable climate outlook for the forthcoming seasons based on the past and present evolution of climate systems.

The sections below discusses briefly the variability in the state of the atmosphere, processes and systems that influence rainfall over the GHA, Dynamical and climate modeling over the GHA, challenges and difficulties associated with the dynamical models and the applications of the regional and global climate models over the GHA.

2.2 The Variability in the state of the Atmosphere

The changes and variability in the atmospheric circulation greatly influences the climate of a given location. The atmospheric states have had a larger temporal variability than can be accounted for by the unpredictable, chaotic component of the flow. This variability could be as a result of external forcing mechanisms such as the sensible and latent heat fluxes from the oceans, the continents or the cryosphere. Model validation studies have been used to address these challenges. Taylor *et al.*, (2001) and Boer and Lambert (2001) characterized model performance based on various methods including correlation analysis. Murphy *et al.,* (2004) introduced a Climate Prediction Index (CPI), which measures the reliability of a model based on the composite mean square errors of a broad range of climate variables. Min and Hense (2006) introduced a Bayesian approach into model evaluation, where skill is measured in terms of a likelihood ratio of a model forecast with respect to some reference or climatology.

2.3 Processes and Systems Influencing Rainfall over the Study Region

The general circulation patterns of East African Rainfall have been studied by using various climate models at different extreme conditions, Giorgi *et al.,* (2000). Several studies have pointed out that global climate models have allowed for a better scientific understanding of various factors that cause global climate change and this has brought commensurate developments in mitigation strategies. These studies include the works of (Hulme *et al.,* 2002, Arnell *et al.,* 2003 and IPCC, 2007). However, at the regional scale, there remains an urgent need for relevant, targeted projections of regional climate change. Study by Shongwe *et al.,* (2011) has given an indication that Africa is the least studied region in terms of ecosystem dynamics and climate variability. This in the recent past has posed a challenge in understanding the primary mechanisms and synoptic systems associated with coupled climate human-ecosystem changes in the region.

Studies by (Owiti *et al.,* 2008 and Nyakwada *et al.,* 2009) and on the predictability of seasonal rainfall using Atlantic Indian Ocean Dipole indicated a strong possible interaction between ENSO and IOD. Their results documented a mode together with the associated gradient that can be used to represent the combined influence of the Indian and Atlantic oceans on the rainfall over the region, and improve the monitoring and Prediction of seasonal rainfall over the region. Omeny *et al.*, (2008) studied the Variability of East African Rainfall using MJO and found high association between rainfall and MJO and high skill of predicting seasonal rainfall over the western part of the region.

Similar studies on the variability of East Africa rainfall and their predictability using SSTs have been done (Omondi *et al.*, 2009). The study provided some evidence of decadal variability in the inter-annual patterns of East Africa rainfall. The MAM and OND seasonal rainfall showed 20 years cycles of wet and dry phases with OND season having high predictability due to dominant synoptic features.

Okoola *et al.*, (2008) studied the Extreme wet patterns that affect the East Africa Coast during the MAM season with respect to 1997 October wet spell event. Their study revealed a high spatial coherence in the rainfall over the EAC. These findings were also consistent with the studies about homogeneity in rainfall variability along the EAC North. The coherence of the rainfall over a larger area observed suggested high association with a large-scale system (Okoola *et al.*, 2009).

2.4 Climate Forcings

Solar variations and Volcanoes have been recurring themes historically in discussions of seasonal prediction. The subsections 2.4.1 and 2.4.2 presents briefly the influence of these forcings on climate variability over the region.

2.4.1 Solar Forcing

The variations in solar forcing are, however, generally comparatively small and tend to operate on long timescales with the most notable being the 11-year solar cycle. Van Loon *et al.,* (2008) reviewed some aspects of solar forcing and found out that the effects were not strong on seasonal timescales. While annual cycle of solar radiation is the dominant external forcing on the climate system, other external forcings like the 11-year solar cycle have been suggested though not very strong on seasonal time scales.

2.4.2 Volcanoes

On the internal climate forcing mechanisms, volcanoes for example, have been found to affect climate and have the potential to affect the skill of certain seasonal forecasts made after large eruptions (Robock, 2000 and Stenchikov *et al.,* 2006).

The effects from GHGs and aerosols have often been neglected in seasonal prediction with the presumption that the effects are small compared to natural variability and that the global warming signal largely is incorporated into the forecast in the initial and boundary conditions. However, Boer (2009) suggested that specification of anthropogenic forcing influence seasonal forecasts. The time-variation of greenhouse gas and aerosol forcing are now increasingly being introduced into seasonal prediction schemes. Doblas (2012) found out that most seasonal forecasts do not explicitly include the effects of anthropogenic forcing but assumes that the effect is small compared to that of the natural variability. However, a warming trend due to the GHG forcing has been identified as a main source of skill in temperature forecasts as confirmed by Doblas-Reyes *et al.,* (2006).

2.5 Dynamical and Climate Modeling

Climate modeling and seasonal forecasting is widely being used as a tool to predict and understand the extremes of the climate conditions. To address challenges common with GCMs the uses of climate modeling and downscaling of climate information to users have been employed.

Downscaling is done through the use of climate models using statistical, dynamical or combination of both (Krishnamurti *et al.*, 2008).

In an attempt to solve the problem of uncertainty and sensitivity to parameterization and initial conditions that are characterized in most GCMs, the concept of multimodal ensemble prediction has been developed and tested for forecasting purposes, Palmer *et al.,* (2000). The use of an ensemble prediction from one model systematically provides better results than the standard deterministic forecasting with only one run which improves the accuracy and forecasts reliability to the consumers. This in the past has been studied by researchers including (Krishnamurti *et al.,* 2000 and Hagedorn *et al.*, 2005).

CMIP5 models have been studied in the recent past by Anyah (2012). The study used these model output to investigate seasonal climatic conditions over the GHA. The models were able to capture the main features of seasonal mean rainfall distribution and its annual cycle. The significant deviations were observed on individual models which depended on the region and season considered. The study revealed a signal of predictably of seasonal rainfall using the GCMs despite the several dynamics and variability of synoptic systems within the region.

Anyah *et al.*,(2006) studied the multiyear simulation of East Africa rainfall during the OND season, using the International Center for Theoretical Physics (ICTP) regional climate model version 3 (ICTP-RegCM3). Observed rainfall variability over distinct homogeneous climate sub regions was fairly reproduced by the model. The spatial correlation between the simulated seasonal rainfall and some of the global teleconnection (DMI and Nino3.4 indices) showed that the regional model conserves some of the observed regional hot spots where rainfall-ENSO/DMI associations are strong.

2.6 The Global Producing centre Models

The key tools in this research were the Global Producing Centers (GPCs) model output. An attempt is made in this section to provide detailed information on the GPCs. The global producing centres have 12 models with each centre giving outputs from their models monthly and seasonally. The GPCs models being used worldwide for seasonal predictions are Glosea4, ECMWF, Meteo-France, Melbourne, Montreal, Moscow, Toulouse, Tokyo, Beijing, Seoul, CPTEC and Washington models.

In this study only 8 models were used for analyses due to challenge in data acquisition for all the models.

Over the last 15 years, a number of international climate centers have developed operational capabilities for global long-range prediction, typically for 3-months-mean climate anomalies and to 6 months ahead, using ensemble integrations of dynamical models. Working through Expert Teams, WMO has fostered coordination between these centers, leading to the establishment of new infrastructure within the Global Data Processing and Forecasting System (GDPFS) that improves both access to the forecast information and the usefulness of this information for generating climate services. New nodes within the GDPFS include 12 WMO-designated Global Producing Centers (GPCs) for seasonal forecasts, which adhere to the criteria for long-range forecast output and verification developed by the Expert Teams. The section 2.6.1 briefly describes the physics and the parametization schemes for each of the 12 GPC models.

2.6.1 Physics and Parametization Schemes of the GPC Models

The European Centre for Medium-Range Weather Forecasts (ECMWF) forecasting system has an ocean analysis model Nucleus for European Modeling of the Ocean (NEMO) for the ocean to estimate the initial state of the ocean, a global coupled ocean-atmosphere general circulation model to calculate the evolution of the ocean and atmosphere. To minimize the uncertainty in the ocean state, a 5-member ensemble analysis is created using perturbed versions of the wind forcing. Prior to starting coupled model forecasts, the ocean analyses are further perturbed by adding estimates of the uncertainty in the sea surface temperature to the ocean initial conditions. Thus all 51 members of the ensemble forecast have different ocean initial conditions.

The atmospheric component of the coupled model is the Integrated Forecast System. It has 91 levels in the vertical with 0.01hPa as the pressure at the top of the model. The model physical parameterization which includes clouds, rain and the land surface are calculated on a reduced 0.7 degrees spacing. The parametization are ice data for resolved lakes, non-orographic gravity wave drag and stratospheric volcanic aerosol within the forecast system. The atmospheric initial conditions come from ERA Interim with seasonal forecasts consisting of a 51 member ensemble. The products from the model are precipitation, temperature, Humidity, Pressure levels and wind components.

The GPC model from Tokyo is run and maintained at Japan Meteorological Agency (JMA). The prediction model system used is an atmospheric general circulation model (AGCM) of resolution TL159. The model is coupled ocean-atmosphere general circulation model (CGCM), which consists of the AGCM and the ocean general circulation model (OGCM). The model has vertical layers of 40 with pressure at top models being 0.4hpa. The initial conditions for atmosphere are JMA Climate Data Assimilation System, for ocean it is Ocean Data Assimilation and for land surface it is climatology with ensemble size of 51 members. The physics of the model is Arakawa –Schubert cumulus parametization, middle level convection of mass flux and large scale condensation. The parametization schemes are radiation, direct effects of aerosols, cloud, gravity wave drag, planetary boundary layer and land surface.

The Meteo-France is another GPC model maintained and run at Toulouse centre. It has ARPEGE as Atmospheric General Circulation Model (AGCM). The coupled model is composed of ARPEGE as an atmospheric GCM and oceanic GCM with coupled with spectral AGCM using a linear T63 truncation. It has Vertical representation using finite-difference with hybrid sigma-pressure coordinate. The Vertical resolution has 19 levels, surface pressure of 1000 hPa, 5 levels are below 800 hPa (Déqué, 2001). The physics of the GCM is hydrostatic shallow-atmosphere approximation, spectral horizontal representation of major variables (vorticity and divergence), temperature and the logarithm of surface pressure. It has Pressure-based hybrid vertical coordinate. The parametization of the ARPEGE take into account the effects of gravity wave drag, diffusion, chemicals and aerosols, convective and cloud formation processes, precipitation, planetary boundary layer, snow cover, and surface and soil characteristics and surface fluxes. ARPEGE has 41 Ensemble members, with temporal resolution of T63, and vertical levels of L31.

The GloSea4 model is the fourth version of the Met Office seasonal ensemble prediction system run at Met office Hardly Centre. It consists of the Met office Unified model (UM) for atmosphere, NEMO for the ocean, Los Alamos sea ice model (CICE) for sea ice and Met Office Surface Exchange Scheme (MOSES) for land surface components. The spatial resolution is N96L85 for atmosphere, which is approximately 135 km in the horizontal with 85 vertical levels, and tri-polar ORCA1L75 for ocean. The horizontal grid distance is 1 degree with 1/3 of a degree between 20 degrees south and 20 degrees North with 75 vertical levels from the sea surface to the bottom. The ocean has 75levels with atmospheric initial conditions from meteorological global analyses. It has

42 ensemble members. Model uncertainties are represented through the use of stochastic physics schemes (Arribas *et al.,* 2011). Global ENSO teleconnections and Madden–Julian oscillation anomalies are well represented in GloSea4. The model is very skillful over the regions that are strongly linked to ENSO conditions.

The Climate Forecast System (CFS) is a fully coupled ocean–land–atmosphere dynamical seasonal prediction system from Washington GPC centre. The atmospheric model has a spectral triangular truncation of 126 waves (T126) in horizontal (100 Km grid resolution) and a finite differencing in the vertical with 64 sigma-pressure hybrid layers. The ocean component of the CFS model is GFDL Modular Ocean Model version 3 (MOM3). The parametization of the model is gravity wave drag, deep cumulus convection, cloud convections, advanced cloud radiation to address unresolved variability in cloud layer and surface runoff parameters (Saha *et al.,* 2012).

The Seoul model is coupled model being run and maintained at Korean Meteorological Agency (KMA). It is a 2-tier global spectral model with hydrostatic primitive equations having hybrid sigma-as pressure coordinates. It has a resolution of T106L21 with ensemble size of 20 members. Some of the parametization schemes of the model physics are cloud convection, land surface, Planetary Boundary Layer, radiation and large scale condensation. The spatial resolution is 2.5°×2.5°with global as spatial coverage. The products of the model are 850hPa temperature anomaly, precipitation anomaly and SST anomaly both of resolution 2.5°×2.5° (Kristler *et al.,* 2001). It has a hindcast period of 1979-2010 with observed climatology as land surface initial conditions and NCEP/NCAR reanalysis as initial conditions for atmosphere.

The Predictive Ocean Atmosphere Model for Australia (POAMA) is a state-of-the-art seasonal to inter-annual seasonal forecast system based on a coupled ocean/atmosphere model and ocean-atmosphere-land observation assimilation systems from Melbourne GPC centre. A multi-model approach of three different configurations of the model has been developed using 33-member ensemble which is made up of 11 member ensembles from each model configuration. The model is useful in simulating El Niño and La Niña condition (Cottrill *et al*., 2013) and has atmospheric resolution of T63.

Montreal Canadian GPC model is a coupled model with atmospheric component of Coupled General Circulation Model version 4 and Ocean General Circulation Model version 4. The OGCM4

uses a Z -level vertical coordinate, with horizontal differencing formulated on an Arakawa B-Grid. There are 40 vertical levels with spacing ranging from 10m near the surface to nearly 400m in the deep ocean. Horizontal coordinates are spherical with grid spacing approximately 1.41 degrees in longitude and 0.94 degrees in latitude. The spectral representation currently used corresponds to a higher horizontal resolution of a 63 wave triangularly truncated (T63) spherical harmonic expansion. The vertical domain of Atmospheric General Circulation Model version 4 extends from the surface to the stratopause region. The physical parameterizations are radiative transfer schemes which accounts for the direct and indirect radiative effects of aerosols. The land surface processes are soil layers, snow layer where applicable and a vegetative canopy treatment. Soil surface properties such as surface roughness heights for heat and momentum , and surface albedos are taken to be functions of the soil and vegetation types and soil moisture conditions within a given grid volume.

The Moscow GPC model is maintained at Russia .It has horizontal resolution of 1.13 degrees latitude and 1.4 degrees longitude and a vertical level of 28 with 28 sigma levels in the verticals. The Physics of the model is moist processes where the precipitation is produced by the large-scale and deep convective condensation processes. Other schemes are radiation, planetary boundary layer and land surface processes. The model has ensemble size of 10 members. The vertical turbulent transport of momentum, heat and moisture in the surface layer is described using Monin − Obukhov theory for different stratification type. Above in the PBL the K − theory is used. The modified Richardson number is applied in this procedure.

Beijing climate centre (BCC) produces its products monthly and seasonally. The Global Ocean Data assimilation model is based on the observation of Global Ocean (GODA) system runs at the early beginning of each month to produce the assimilated data such as sea temperature, precipitation and other climate variables. The model is a coupled atmosphere- ocean- land system. The physics of the model are atmospheric chemistry, dynamic vegetation, ice sheets. The parametization schemes are clouds, convection, boundary layer, short and long wave radiation. The last GPC centre is Pretoria which hosts a model called ECHAM4.5 for its Ensemble Prediction System. The ECHAM4 model has 19 vertical levels, ECHAM4/L19, and with 39 vertical levels, ECHAM4/L39DLR. The uppermost full level at 10 hPa, with the L39 version has better resolution mainly around the tropopause.

Attempts have been made to test the skill of some the GPC models over Africa but no effort has been put over GHA or East Africa. The models that have been tested over Africa are Glosea4, ECMWF, Meteo-France, CPTEC and ECHAM 4.5. Studies by Marengo *et al.,* (2003) using GLosea4, ECMWF and Meteo-france models from EUROSIP multi-model ensemble over Africa showed high skill in simulating rainfall over African continent during OND season than the statistical models.

Ininda (2008) studied and analyzed the circulation of rainfall pattern over Tanzania simulated by the ECHAM4 model forced with observed SSTs. The annual cycle and inter-annual rainfall variability were well reproduced by the model especially during the OND season. Similar studies using ECHAM5 version model generated by MoS technique of downscaling over Tanzania significantly improved the prediction skill especially during the OND season.

Marengo *et al.,* (2004) did a study on the annual and inter-annual variability of regional rainfall produced by the Center for Weather Forecasts and Climate Studies/Center for Ocean, Land and Atmospheric Studies (CPTEC/COLA) for atmospheric global climate model. An evaluation was made of a 9-member ensemble of the model forced by observed global sea surface temperature (SST) anomalies for a 10-year period (1982–1991). The annual cycle of precipitation was well simulated by the model for several continental and oceanic regions in the tropics and mid latitudes.

Table 1 summarizes the main characteristics of the 12 Global Producing Center models, including the resolution of the models and hindcasts period of each Center model as well as the ensemble size of the forecasts model. The GPC model centres include Beijing, Exeter, Melbourne, Montreal, Seoul, Tokyo, Toulouse, Washington, ECMWF, Pretoria, CPTEC and Moscow. These have been designated as Lead Centers for Standard Verification System of long range forecasts. Seoul-Washington designated as a Lead Center for LRF Multi-Model Ensemble (LC-LRFMME).

Table 1: The WMO Global Producing Centers (GPCs)

GPC name	Center	Ensemble size	Resolution (Atmospheric component)	Hindcast period
Beijing	BCC	Coupled (48)	T63/L16	1983–2004
ECMWF	ECMWF	Coupled (51)	T159/L62	1981–2005
Exeter	Met Office Hadley Center	Coupled (42)	1.25° × 1.85°/L38	1989–2002
Melbourne	Australian Bureau of Meteorology	Coupled (33)	T47/L17	1980–2006
Montreal	Meteorological Service of Canada	2-tier (40)	T32/T63/T95/2.0° × 2.0° (4- model combination)	1969–2004
Seoul	KMA	2-tier (20)	T106/L21	1979–2007
Tokyo	JMA	Coupled (51)	TL159/L40	1979–2008
CPTEC	CPTEC	2-Tier(15)	T62/L28	1979-2001
Toulouse	Meteor-France	Coupled (41)	T63/L91	1979–2007
Washington	NCEP	Coupled (41)	T62/L64	1981-2004
Moscow	Hydromet Center of Russia	2-tier (10)	1.1° × 1.4°/L28	1979–2003
Pretoria	SAWS	2-tier (6)	T42/L19	1983–2001

2.6.2 Delineated Homogenous Rainfall Zones over GHA

ICPAC has grouped the rainfall records of the region into homogenous groups for the individual season. Figure 4 shows the climatological homogeneous rainfall zones of the GHA based on the Principal Component Analysis that was adopted for the study. The delineation reduces the number of the stations used in the study and only retains one station amongst others which have the characteristics of other stations within a given zone.

This study focused on OND season that is one of the key rainfall seasons. The season was preferred since the large scale systems that influence the systems are easily simulated well by the GCMs during the OND season.

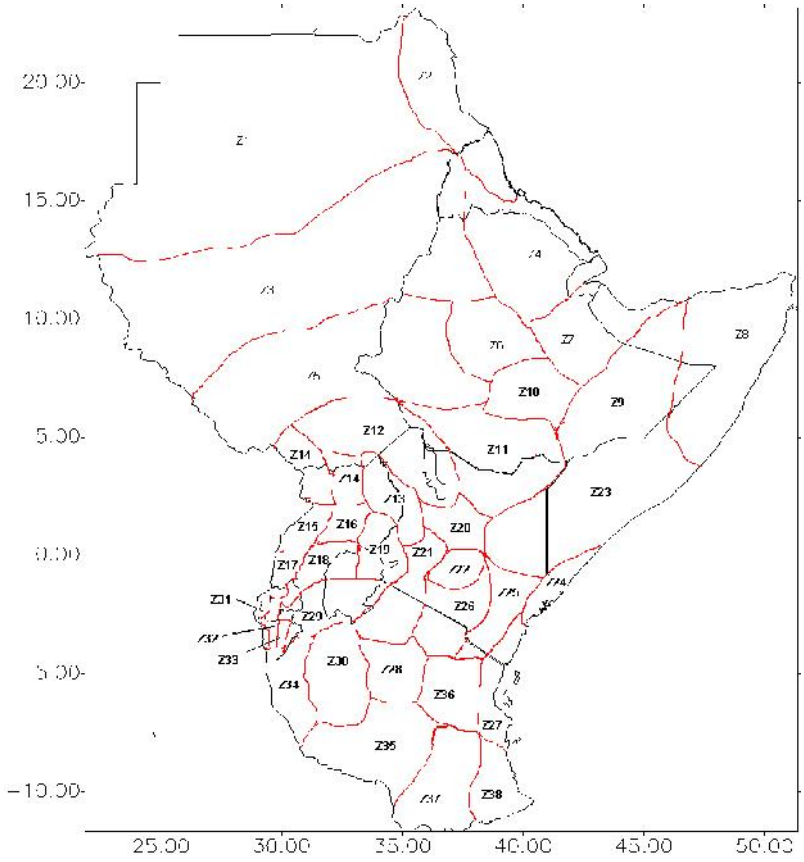

Figure 4 : Climatological zones within the GHA adopted for the study (Source: Ogallo 1989)

2.7 Challenges Associated with Dynamical Models in Seasonal Forecasting

Kalnay (2003) reported that models that consider the dynamics of the atmosphere and oceans should be able to predict better than the statistical models that do not consider the physical parameterization and initial boundary conditions of the atmospheres like. However the predictions and forecasting of future climate scenarios have received challenges due to the errors in the models that cause unreliable forecasting information.

The errors in the models and uncertainties in the forecasts have led to the investigations and validations of the model skill. Experiments have shown that dynamical seasonal forecast models are over realistic and their spread is too narrow to match the range of observed outcomes. There is often little relationship between ensemble spread and the error in the forecast. The reason for this is believed to be in model error. This has been studied by Jones (2005) and Achutarao (2006).

Multi-model approaches, where ensembles from different state-of-the-art models are combined have been found to reduce model errors and produce more skillful forecasts than the single models. Palmer *et al.,* (2004) and Wang *et al.,* (2008) did a study on the multi model ensemble approaches and found out that ensemble technique significantly improves the model performance. Mutemi (2003) used ECHAM4.5 to study the variability of East African climate. The model reproduced the climatological mean pattern such as the bimodal seasonality of rainfall associated with the north–south migration of the ITCZ and monsoonal flow. The study however did not get the correct amplitudes of the inter-annual variability linked to extreme El Nino episodes such as the 1982 and 1997. The skill of the model was found to be higher during ENSO when large SST values are found over many parts of the Equatorial tropics.

2.8 Applications of Regional Climate Models and Global Climate Models

The experience of modeling groups and WCRP on Seasonal Prediction (Kirtman and Pirani, 2009) shows that dynamical models offer a long term potential in providing seasonal predictions in terms of climate and its variability despite the many problems with dynamic models. Seasonal forecasting models still require very accurate modeling of basic features of the climate system such as atmospheric boundary layers, convection, land surface processes and ocean mixing. These parameters amongst others require adequate treatment of a range of other processes to improve on the forecast skill for dynamic models (Latif *et al.,* 2001).

It is vital that seasonal forecasts be accompanied by a measure of expected forecast skill, based on the past operation of the system, in order to be properly utilized. Plisnier *et al.,* (2000) research suggests that warming SSTs in the southwest Indian Ocean, in addition to inter-annual climate variability (El Niño/Southern Oscillation (ENSO) may play a key role in precipitations over East African. Warm SSTs are thought to be responsible for the recent droughts in Equatorial and subtropical Eastern Africa during the 1980s to the 2000s (Funk *et al.,* 2005).

Despite the uncertainties in the forecasts, the skill and value of seasonal forecasts promise to increase as investments are continued in the various components of the observation-analysis-prediction-application system. Forecast skill is usually highest when considering multi-model ensemble (MME) forecasts. This technique of MME has been studied by (Palmer *et al.,* 2004 and Jin *et al.,* 2008) and revealed an improvement in forecasts using MME.

The use of seasonal Ensemble prediction or several models systematically provides better results than the standard deterministic forecasting. Statistical analyses based on output from GCMs, the Model Output Statistics approach correct systematic errors in the GCM fields such as spatial shifts and down-scales the climatic information when simulations and observations are not at the same spatial resolution. This concept has been studied by Krishnamurti *et al., (*2000), Landman and Goddard, (2002), Hagedorn *et al.,* (2005) and Wills (2006).

High resolution regional climate models (RCMs) simulation (dynamical downscaling) or statistical methods (statistical downscaling) are the common approaches to obtain fine spatial-scale information from long-term global circulation model (GCM) simulations. RCMs are formulated in terms of physical principles and therefore have the potential for capturing fine spatial-scale nonlinear effect which increases confidence in their abilities to downscale future climate. RCMs have been widely used to provide a better projection of future climate at the regional scale and proven to be a powerful tool in the dynamical downscaling of regional climate since 1980s (Wang *et al.,* 2004, Lo *et al.,* 2008 and Heikkila *et al.,* 2010).

Although there have been studies of Global Climate Models (GCM)-simulated climate change for several regions in Africa, the downscaling of GCM output to finer spatial and temporal scales has received relatively little attention in Africa and GHA. It is from the challenges with dynamical model and uncertainties in the seasonal forecasts reviewed in the literature above that forms the motivation and drive for this study. The study was devoted in assessing the skill and accuracy of global models in simulating regional rainfall scale within the GHA.

CHAPTER THREE

3.0 DATA AND METHODOLOGY

This chapter describes the datasets, model applied and various methods that were used to achieve the overall and specific objectives of the study.

3.1 Data

The data used in this study included gridded observed rainfall data from Climate Research Unit (CRU), University of East Anglia from 1983 to 2001 and GPCs model hindcasts data from the Individual GPC website.

3.1.1 Climate Research Unit Data

The gridded observed rainfall data used in the study was from the Climate Research Unit (CRU) obtained from the University of East Anglia for the period 1983-2001. The CRU datasets have been interpolated at different regular spacing. For example the datasets are available at regular spacing of 0.5, 1.5, 2.0 and 2.5. For this study, the datasets at 2.5 regular spacing was used to match with the resolution of the model output. The datasets contain five climatic parameters precipitation, surface temperature, diurnal temperature range (DTR), cloud cover and vapor pressure.

In this study only the precipitation parameter was used for model evaluation. Monthly gridded rainfall from CRU was used. The CRU monthly datasets are derived from satellite data, model estimates and rain gauge data from ground stations. The final merged product is generated by combining the satellite and reanalysis data. It contains precipitation distributions with full global coverage and improved quality compared to the individual data sources. Comparisons of the CRU with other data sources revealed remarkable agreements over the global land areas and over tropical and subtropical oceanic areas, with differences observed over extra tropical oceanic areas.

Studies within the GHA and East Africa using CRU as the observed have revealed remarkable agreement and high skill in simulation of climate variables. Omondi (2010) studied teleconnections between decadal rainfall variability and global sea surface temperatures over East Africa using CRU as observed. He found high relationships between the teleconnections and rainfall variability. Similar studies on the variability of East Africa rainfall and their predictability using SSTs have been done (Omondi *et al.,* 2009). In both cases CRU was used as the observed. Their studies found

some evidence of decadal variability in the inter-annual patterns of East Africa rainfall during the OND season. Yang *et al*., (2013) studied the decadal variability of the East African precipitation during MAM season and the performance of a series of models in simulating the observed features using CRU as the observed. Observational results show that the drying trend of the long rains is associated with decadal natural variability associated with sea surface temperature.

3.1.2 Model Data from Global Producing Centres

In this study only eight models were used for the analysis due to challenges encountered in data acquisition for all the 12 models. The model hindcast datasets for a set of 19 years from 1983 to 2001 all months and years inclusive was obtained from individual GPC centers. The grid resolutions of the models datasets and hindcast period of datasets available for each model are shown (Table 1).

Table 2 shows a list of rainfall stations that were used for the study. The stations were selected based on homogenous zones and correlations between inter-stations. In each one of the 38 homogeneous zones, only one representative station was selected. The total number of stations selected based on the homogenous zones were then correlated with each other. From the inter-stations correlations only one representative station was picked form those that were strongly correlated, hence the 25 rainfall stations used for the analysis.

Table 2: Locations of the Rainfall Stations used in the Study

Stations No.	Zones	Representative Stations	Country	Longitude	Latitude	Altitude (m) Above MSL
1	1	Khartoum	Sudan	32.55	15.6	394
2	1	Edduim	Sudan	32.33	14.0	418
3	2	Asmara	Eritrea	56.5	15.19	2350
4	3	Abuhamad	Sudan	33.32	19.53	317
5	4	Djibouti	Djibouti	43.15	11.55	1010
5	5	Wau	South Sudan	28.13	7.7	1470
6	5	Juba	South Sudan	31.6	4.87	469
7	11	Combolcha	Ethiopia	39.72	11.08	1860
8	11	Wajir	Kenya	31.08	1.44	267
9	12	Lodwar	Kenya	35.36	3.06	510
10	14	Gulu	Uganda	32.17	2.46	1100
11	16	Entebbe	Uganda	32.27	0.03	1200
12	19	Kabale	Uganda	29.59	1.15	1870
13	21	Kericho	Kenya	35.35	-0.37	3190
14	21	Dagoretti	Kenya	36.45	1.18	1820
15	21	Kisumu	Kenya	34.46	0.05	1200
16	22	Narok	Kenya	35.52	1.55	915
17	25	Lamu	Kenya	40.54	2.16	29
18	26	Makindu	Kenya	37.83	-2.28	1000
19	30	Mwanza	Tanzania	32.92	-2.47	1139
20	31	Kigali	Rwanda	30.12	-1.97	1460
21	32	Bujumbura	Burundi	29.32	-3.37	795
22	23	Mogadishu	Somalia	45.25	2.02	52
24	37	Mtwara	Tanzania	40.18	-10.27	113
25	34	Kigoma	Tanzania	29.63	-4.88	999

3.2 Methodology

This section presents the various methods employed in the study. These include methods used to assess the skill of individual models in representing the spatial and temporal distribution of rainfall, and determination of the ensemble models that could provide improved skill over GHA region.

3.2.1 Standardization of Rainfall Records

Before the analysis all the datasets were standardized for easier comparison using Equation 1. Standardization involves converting the individual rainfall records into comparable indices. Various indices have been developed and used in the region. The most common method of standardization is to express the observation as a percentage of long term average using same baseline period to compute the averages. The standardized anomaly z was adopted in this study for fair comparison between the models simulated and observed records.

$$Z_i = \frac{y_i - \mu}{\sigma_y} \quad \text{.....1}$$

In Equation 1, Z_i is the i^{th} standardized observed or predicted rainfall value, y_i is the i^{th} observed or predicted value, μ is the mean of observed or predicted value, σ_y is the standard deviation of observed or predicted rainfall, and i =1,2,3,4…..N and N is the total number of forecasts or observations. All observed and predicted values were expressed as anomalies for easier comparison.

3.2.2 Assessment of the Spatial and Temporal Distribution of Observed and Predicted Rainfall Anomalies for Selected Extreme Events

Analysis was done to select years which experienced extreme events for the period 1983-2001. The seasonal long term mean was calculated for each of the 19 years in some few selected stations. Any year which recorded above 125% of seasonal long term mean was regarded as enhanced, while any year which recorded less than 75% of seasonal long term mean was regarded as a depressed year.

Both observed and predicted rainfall anomalies were then compared through visual analysis between trends of observed and model output for extreme events. This visual comparison was done in order to assess the departure or agreement between the model output and observed rainfall in capturing the annual cycle, mean seasonal rainfall patterns and inter-annual variability. The methods applied were time series and graphical displays.

3.2.3 Correlation Analysis

Correlation analysis was done between the observed rainfall data to determine stations that were strongly correlated using Equation 2. In each of the 38 homogeneous zones, only one representative station was selected from the stations that were strongly correlated with each other giving a total of 25 rainfall stations.

The validation of CRU datasets with the observed station datasets was also done using correlation analysis. The correlation analysis was also applied between observed and model output to compare models output and observed data using Equation 2. The correlation coefficient (r) between a model output variable (f_i) and the corresponding observation (***O****i*) is given by:

$$r = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(f_i - \bar{f}\right)\left(o_i - \bar{o}\right)}{\left[\frac{1}{N}\sum_{i=1}^{N}\left(f_i - \bar{f}\right)^2 \cdot \frac{1}{N}\sum_{i=1}^{N}\left(o_i - \bar{o}\right)^2\right]^{1/2}} \quad \text{…………. ………….……………………} (2)$$

In Equation 2, N is the total number of years used for analysis, $\bar{f}$ is the mean of the model output and $\bar{O}$ is the mean observation of the observed variable. The correlation ranges from -1 to 1 where a value of 1 denotes perfect linear relationship and -1 denotes a perfect inverse linear relationship.

3.2.4 Testing for the Significance of the Correlation Coefficient

The computed value of correlation coefficient between observed and model output was tested using the student T-test. The test significance level considered was 95% level of interval. If the computed value of t is greater than the tabulated value, then the correlation coefficient is significant. The t-test for correlation coefficient is given by:

$$t_{n-2} = r\sqrt{\frac{(n-2)}{1-r^2}} \quad \text{…………………………………………………………………..………} (3)$$

In Equation 3, *n* is the total number of years used in the study; $n-2$ is the degrees of freedom, t_{n-2} is the value of the confidence level computed from the correlation coefficient and r is the correlation coefficient. If the computed value of t is greater than the tabulated value of t_{n-2}, then the correlation coefficient is significant.

3.2.5 Multiple Linear Regression

Once the correlation between variables has been established, it is usually important to determine the nature of the relationship between the correlating variables. Regression analysis helps determine linear relationships between variables. In this study regression analysis was done using Equation 4.

$$y_i = a_o + \sum_{i=1}^{n} b_i x_i + e_i \quad \text{..... (4a)}$$

In Equation 4a, a_o and b_i are the intercept and regression coefficients for the predictors x_i applied in the Equation. In this study, the predictors were the model output. The variance of the error term e_i , in this case is given by Equation 4b.

$$S^2 = \frac{SSE}{n-(k+1)} \quad \text{..... (4b)}$$

In Equation 4b, SSE is the sum of the squared of errors, n is the period of time considered for the study.

The test of the adequacy of the model is done by computing R^2 (the multiple coefficient of determination) given by

$$R^2 = 1 - \frac{SSE}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \quad \text{..... (4c)}$$

In Equation 4c, y_i is the model output and $\bar{y}$ is the mean of the model output. For a perfect model the value of R^2 should be 100%. For $R^2 = 0$, it implies lack of fit, while $R^2 = 1$ implies perfect fit. Regression analysis in this study was used to determine the linear relationship between the model output and OND seasonal rainfall anomalies for the 25 rainfall stations used in the study. The stepwise regression technique was used as a means of picking the best individual predictor into the regression model Equation.

3.3 Categorical Statistics

Categorical statistics was used to analyze the relationship of model output and the observed rainfall values. A 3 by 3 contingency Table was used to display the data.

Table 3 gives the basic structure and entries from categorical analysis from which some skill score were evaluated. The letters in the Table were used to calculate the various score as discussed in the

next section below. Below Normal (BN), Normal (N) and Above Normal (AN) are various categories where the rainfall was observed and predicted. Letters A-I denotes the values obtained at different categories for the predicted and observed events. Letters J-O shows the totals of the events observed at different categories, and letter T is the total number of events carried out.

Table 3: A 3 by 3 contingency Table

	FORECAST				TOTAL
OBSERVED		BELOW NORMAL	NORMAL	ABOVE NORMAL	
	BELOW NORMAL	A	B	C	M
	NORMAL	D	E	F	N
	ABOVE NORMAL	G	H	I	O
TOTAL		J	K	L	T

3.3.1 Bias Score

The bias score measures the ratio of the frequency of forecast rainfall events to the frequency of observed rainfall events. It indicates whether the forecast system has a tendency to under forecast (Bias<1) or over forecast (Bias>1) rainfall events. It ranges from 0 to ∞ the perfect score is 1 (100%). The Bias score was calculated using Equation 5 generated from Table 3 and the letters have their meanings as defined in subsection 3.4.

$$Bias = \begin{Bmatrix} \frac{J}{M} & BN \\ \frac{L}{O} & AN \\ \frac{K}{N} & N \end{Bmatrix} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (5)$$

3.3.2 Probability of Detection (PoD)

The Probability of Detection (PoD) gives a simple measure of proportion of rainfall events successfully forecast by the model. It is calculated by dividing the total number of correct forecasts by total number of events observed. PoD ranges from 0 to 1 where a perfect score is 1 (100%). Equation 6, gives the formula for computing the PoD for below normal, normal and above normal categories with the letters have their meanings as defined in subsections 3.4.

$$\text{PoD}=\begin{cases}100*\dfrac{A}{M} & BN\\[2ex] 100*\dfrac{I}{O} & AN\\[2ex] 100*\dfrac{E}{N} & N\end{cases}\quad\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ (6)$$

3.3.3 False Alarm Ratio (FAR)

The FAR gives a simple proportional measure of the model's tendency to forecast rain where no rain was observed. The score ranges from 0 to 1 (100%), the perfect score is 0. The FAR for the below and above normal categories was given by Equation 7 with the letters having their meanings defined in subsections 3.4.

$$\text{FAR}=\begin{cases}100-100*\dfrac{A}{J} & BN\\[2ex] 100-100*\dfrac{I}{L} & AN\end{cases}\quad\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ (7)$$

3.3.4 Heidke Skill Score (HSS)

The Heidke Skill Score (HSS) measures the fraction of correct forecasts after eliminating those forecasts which would be correct due purely to random chance. The numerator is the number of correct forecasts, and the reference forecast in this case is the rainfall events experienced by a given geographical location (Climatology). The score ranges from -∞ to 1, the perfect score is 1 (100%). Any score less than zero means the forecasts model is worse off than climatology. The HSS for this study was computed using Equation 8. The meanings of the letters in Equation 8 are defined in subsection 3.4.

$$\text{HSS}=\frac{A+E+I-\dfrac{JM+KN+LO}{T}}{T-\dfrac{JM+KN+LO}{T}}\quad\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ (8)$$

3.4 Determination of the Ensemble Model for Forecasting Rainfall

The results from the analysis was compared and used to determine best ensemble model that is useful for rainfall forecasting over GHA region based on the regression weights or performance of individual models. The down-scaling technique used was model output statistics (MOS). The Model Output Statistics (MOS) is an objective weather forecasting technique which consists of determining a statistical relationship between a predictand and variables forecast by a numerical model at some projection time(s). This technique, together with regression, is applied to the prediction of the parameter.

The regression analysis for the individual models yielded the R-square and P-values which were treated as the weights. For a perfect model R-square value is 100%, hence the higher the R-square value the better the skill of the model. This formed the basis of the determination of the ensemble model. The two methods applied to generate the ensemble models are discussed in the next section below.

3.4.1 Simple Composite Analyses for the First Ensemble Model (ENSE 1)

In this case, the model rainfall output was averaged for all the 8 models. This technique, also called simple composite gives each model equal weight. Thus the simple composite for the first ensemble model forecast (ENSE 1).

$$ENSE1 = \frac{1}{n}\sum_{i=1}^{n} mi \quad \text{..... (9)}$$

In Equation 9, *mi* is the rainfall output for the i^{th} model and *n* is the total number of years used for the prediction.

3.4.2 Weighted Multi-model Ensemble by Linear Regression

The second Ensemble (ENSE 2) model was developed as an improvement to the first ensemble from the four models Washington, Montreal, Melbourne and Moscow that had high skill using Equation 10.

$$\bar{y} = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_k x_k \quad \text{..... ..(10)}$$

In Equation 10, b_k is the k^{th} regression weight for the model, x_k is the number of k^{th} models used to generate the ensemble model and $\overline{y_i}$ is the predictand (rainfall). The regressions coefficients and the R-square explained by the individual models were multiplied and then divided by their totals to generate the weights. The best most skillful model got higher weight and the less skillful model amongst the four got the least weight. The method involved getting sum totals between the product of the regression coefficients and R-square which were used as the regression weights for all the four models in the ENSE 2 model.

CHAPTER FOUR

4.0 RESULTS AND DISCUSSION

This chapter discusses the results that were obtained in the study using different methodologies.

4.1 Introduction

The results obtained included seasonal mean spatial distribution of the observed rainfall, comparison of the spatial distribution of the observed anomalies versus model output anomalies during extreme rainfall events, correlation analysis, regression analysis, Model output Statistics (MoS), simple composite analyses, weighted averages and the categorical statistics.

4.2 Distribution of Observed Rainfall and Model Output for 1983-2001

Figure 5 show the spatial distribution of seasonal mean rainfall for observed rainfall, Washington and Beijing models output. The plotted means were calculated for the period 1983-2001. The general distribution pattern shows that high amounts of rainfall are received on the western sector of the GHA as shown in Figures 5 (a) and 5 (b). Beijing model had poor distribution of rainfall as it overestimated and underestimated the rainfall in the lower part of central Tanzania and the coastal strip of Kenya over the study region (Figure 5).

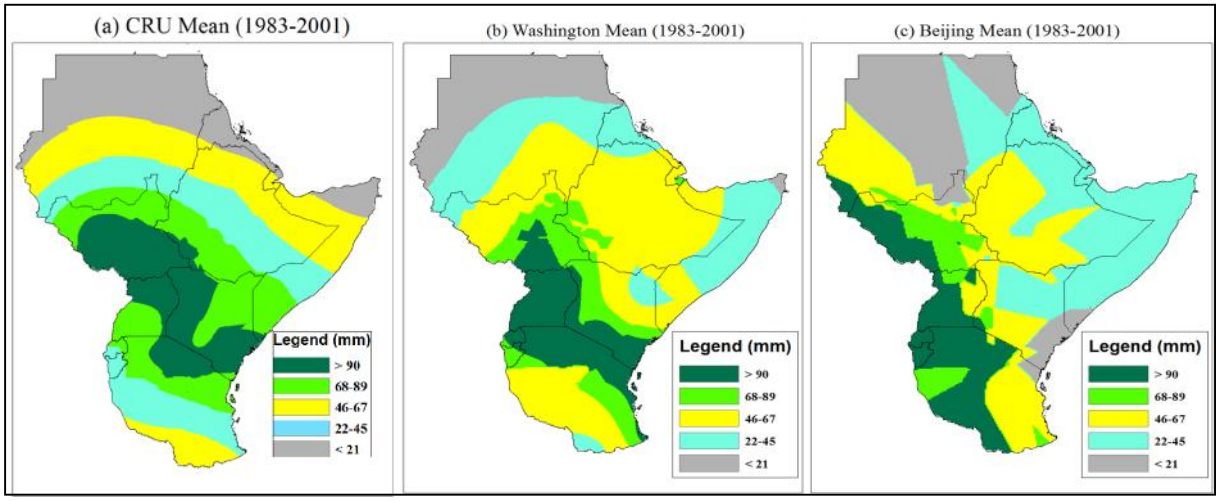

Figure 5: Spatial distribution of seasonal mean of observed rainfall, Washington and Beijing models output for 1983-2001. Only the best and poor models are shown.

4.2.1 Evaluation of CRU datasets against the Stations Observation

The choice of CRU datasets for this study was based on the fact that it strongly correlates with the observed rainfall station data during the OND season over the GHA. Figure 6 shows the inter-annual variability of CRU and observed rainfall datasets for some few stations on the western sector of the GHA, i.e Dagoreti, Makindu and Gulu stations. From Figures 6 (a), 6 (b) and 6 (c) it can be noticed that CRU datasets picked most of the station extremes with significant correlations of 0.65, 0.51 and 0.59 respectively. However the highest correlation explained only 42% and less in accuracy of the datasets over these stations. Therefore all subsequent results in the study used CRU for the analyses.

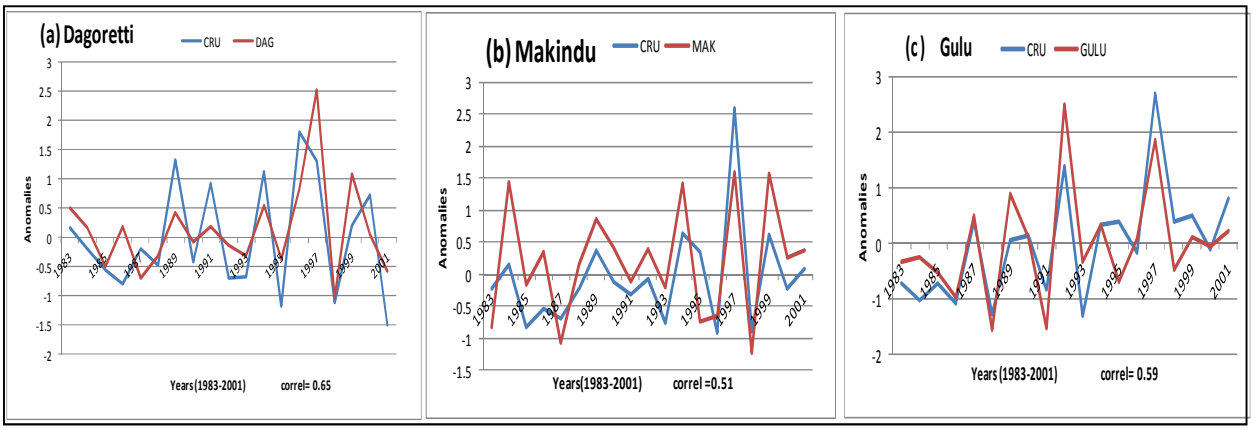

Figure 6: Inter-annual variability of rainfall output of CRU and station observations over (a) Dagoreti (b) Makindu (c) Gulu. Blue and Red shadings represents CRU and Stations respectively.

4.2.2 Distribution of Observed Rainfall and Model output during 1997 El Niño Episode

Figure 7 shows the results of the spatial analysis of observed rainfall anomalies Figure 7 (a), and models output from eight model centers during the OND season for the year 1997. Also included are Beijing Figure 7(b), CPTEC Figure (7c), Melbourne Figure (7d), Montreal Figure (7e), Moscow Figure (7f), Seoul Figure (7g), Tokyo Figure (7h) and Washington Figure (7i).

In Figure 7 (a), high rainfall was majorly distributed around the Equatorial sector and low rainfall in the Northern and Southern sectors of the study region. The stations that had high distribution of rainfall around the Equatorial sector were from zones 14, 16, 21, 25 and 32. These zones represented the western part of the GHA and lower part of Kenyan coastal strip. This distribution pattern of observed rainfall from CRU was closest to the observations made from the spatial pattern

from four models; Washington, Moscow, Montreal and CPTEC (Figure 7). The product from Beijing, Tokyo, Melbourne and Seoul centers showed greater departure from the observed rainfall pattern.

The high rainfall around the Equatorial sector and low rainfall around the Northern and Southern sectors is because during the OND season high rainfall is received on the Equatorial sector of the GHA and low rainfall is experienced in the Northern and southern sectors of the study region.

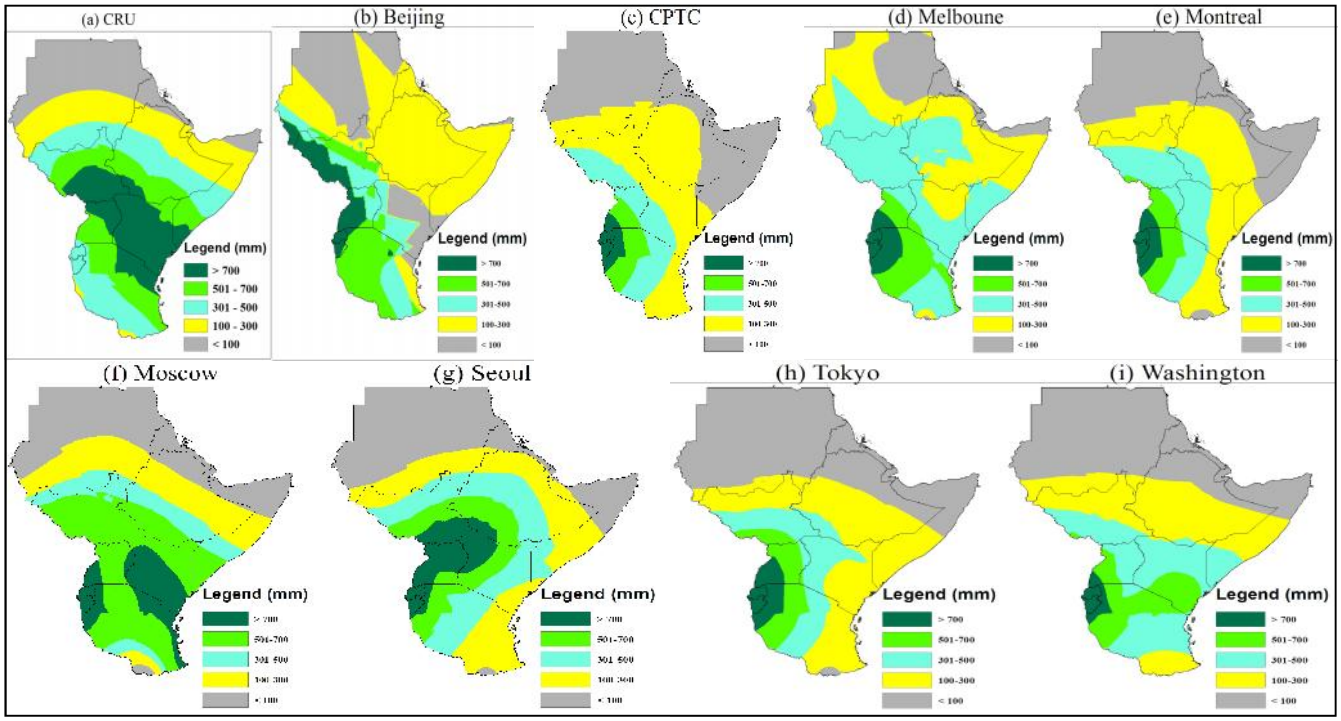

Figure 7: Spatial distribution of observed and 8 model rainfall output for El Niño Episode (1997)

During the El Niño year (1997/1998), model output from Moscow, Seoul, Montreal and Washington centers simulated the observed rainfall distribution for the region well. High amount of rainfall was distributed on the Equatorial sector of the study region. The rest of the models missed some stations in central part of Kenya and lower part of southern Sudan. The Beijing model captured the pattern of observed rainfall in some stations at the southern part of Tanzania. Most stations over the Equatorial region recorded rainfall above 700 mm.

4.2.3 Distribution of Observed Rainfall and Model Output during 2000 La Niña Episode

Figure 8 shows the results of the spatial analysis of the observed rainfall and model output anomalies from the eight models in the OND season for the year 2000.

Rainfall was observed to be mainly concentrated around the Equatorial sector. The distribution of rainfall was observed to be low in Figures 8 (c), 8 (e), 8 (f), 8 (h) and 8(i) in the Northern and Southern sectors of the study region. The stations that depicted low distribution of rainfall around the Equatorial sector were from zones 14, 16, 21, 25 and 32. These zones represent the western and the South East part of the GHA.

The distribution pattern of observed rainfall was closest to the observations made from the spatial pattern of the products from 5 models; Washington, Moscow, Montreal, CPTEC and Seoul (Figure 8). The products from Beijing, Tokyo, Melbourne and Seoul showed enhanced distribution of rainfall hence departure from the observed rainfall pattern.

During the La Niña year 2000, the distribution pattern of rainfall observed in the southern and central part of Tanzania was well simulated by most of the models output except for the products from Beijing and Washington centers that showed opposite signal. The regions in the Northern and Southern of the GHA, lower part of central Kenya and southern part of Tanzania that showed departure from the observations seemed to have an opposite signal to the rest of the GHA region when ENSO events are considered (Indeje *et al.*,2000).

The rainfall pattern for the observed rainfall showed that most stations on the western parts and Equatorial sector of the study region recorded rainfall above 500 mm as observed in Washington, Moscow, Montreal, Melbourne and Beijing models. The observed pattern is because during the OND season high amounts of rainfall is received on the western sector of the GHA. The low distribution of rainfall over Equatorial sector is because of the depressed rainfall experienced over the Equatorial sector during the 2000 La Niña episode.

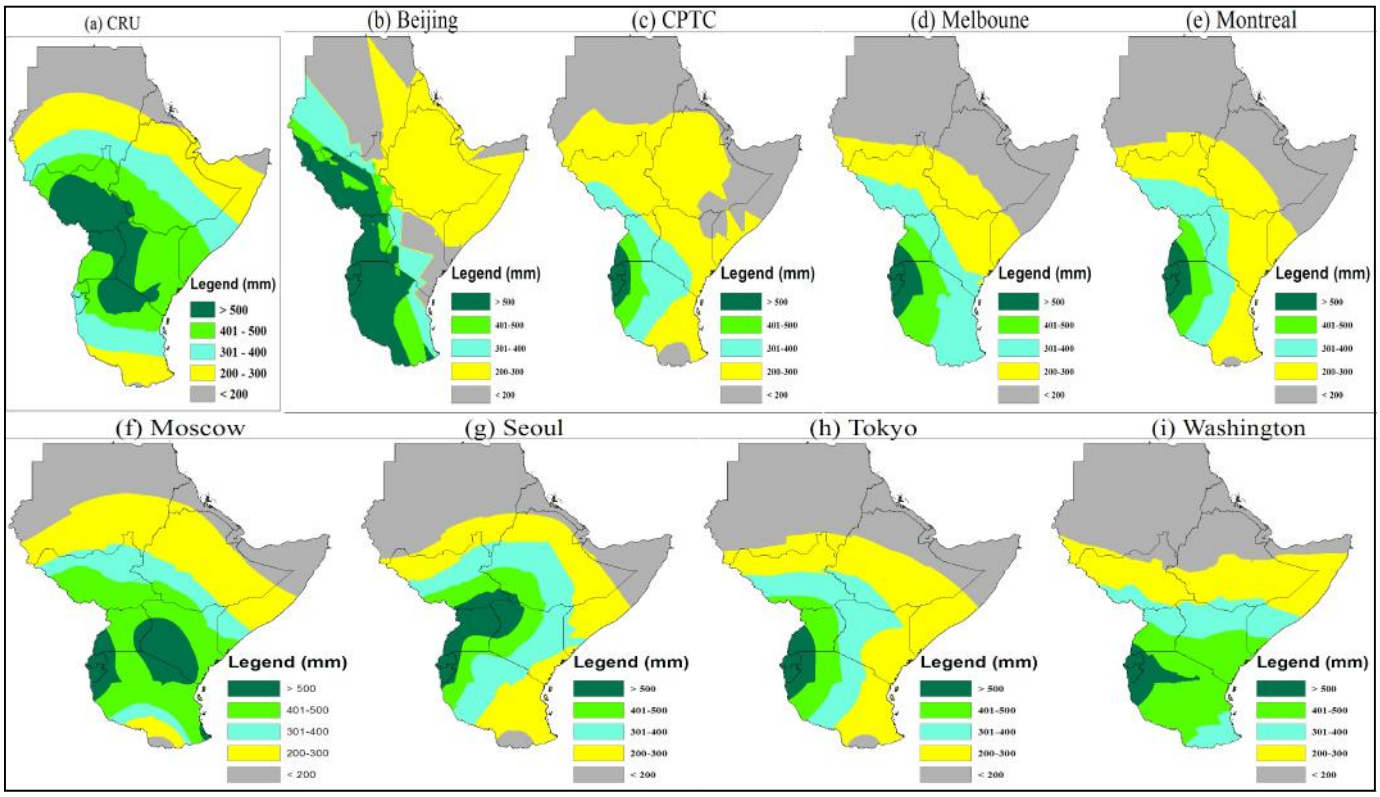

Figure 8 : Spatial distribution of rainfall between the observed and Model output for La Niña episode (2000).

4.3 Inter-Annual Variability of Observed and Model Output Rainfall Anomalies for Some Selected Stations

Figure 9 shows the results of the inter-annual variability of the models output and observed rainfall anomalies across the 25 rainfall stations. Nine stations mostly from Equatorial sector i.e. Abuhamad, Bujumbura, Combolcha, Dagoretti, Djibouti, Entebbe, Gulu, Juba, and Kabale were displayed as having high correlations with the observed rainfall. This analysis was done to establish the ability of the model to simulate the year to year inter-annual variability in the seasonal rainfall pattern with particular reference to the El Niño and La Niña extreme years. A regression analysis was done at 95% level of confidence across the rainfall stations to determine which model(s) output had skill useful across each station. Only those models output that explained high R-square value and low p values were picked for each of the stations.

From the results in Figure 9, Washington model had high skill in four stations i.e. Combolcha,

Dagoretti, Djibouti and Entebbe. Montreal model had good skill across 3 stations i.e. Abuhamad, Gulu, and Juba. Moscow and Melbourne models had good skill in 2 stations, Beijing and Tokyo models had good skill in only one station each. The rest of the models i.e. CPTEC and Seoul were not skillful in any of the 9 stations.

The models output from Washington, Moscow, Melbourne and Montreal simulated well the peaks for the El Niño and La Niña years. The stations from the Equatorial sector, i.e. Figures 9 (b), 9(d), 9 (f), 9 (g) and 9 (i) tried to reproduce the pattern of the observed rainfall. The stations in the Northern part of the study region did not simulate the observed rainfall pattern well, as shown in Figures 9 (a) and 9 (c). The model output from Moscow, Montreal and Washington depicted high variability during the years 1997 and 2000.

The peaks during OND season associated with ENSO events were fairly replicated by some of the models output like Washington, Montreal, Moscow and Melbourne. These peaks could be associated with some global teleconnection like ENSO phenomenon which is the main driver during the OND season (Muhati *et al.,* 2007). The reason why some models missed the observed rainfall pattern in certain sectors of the GHA could be due to weak linkage with ENSO conditions in these regions.

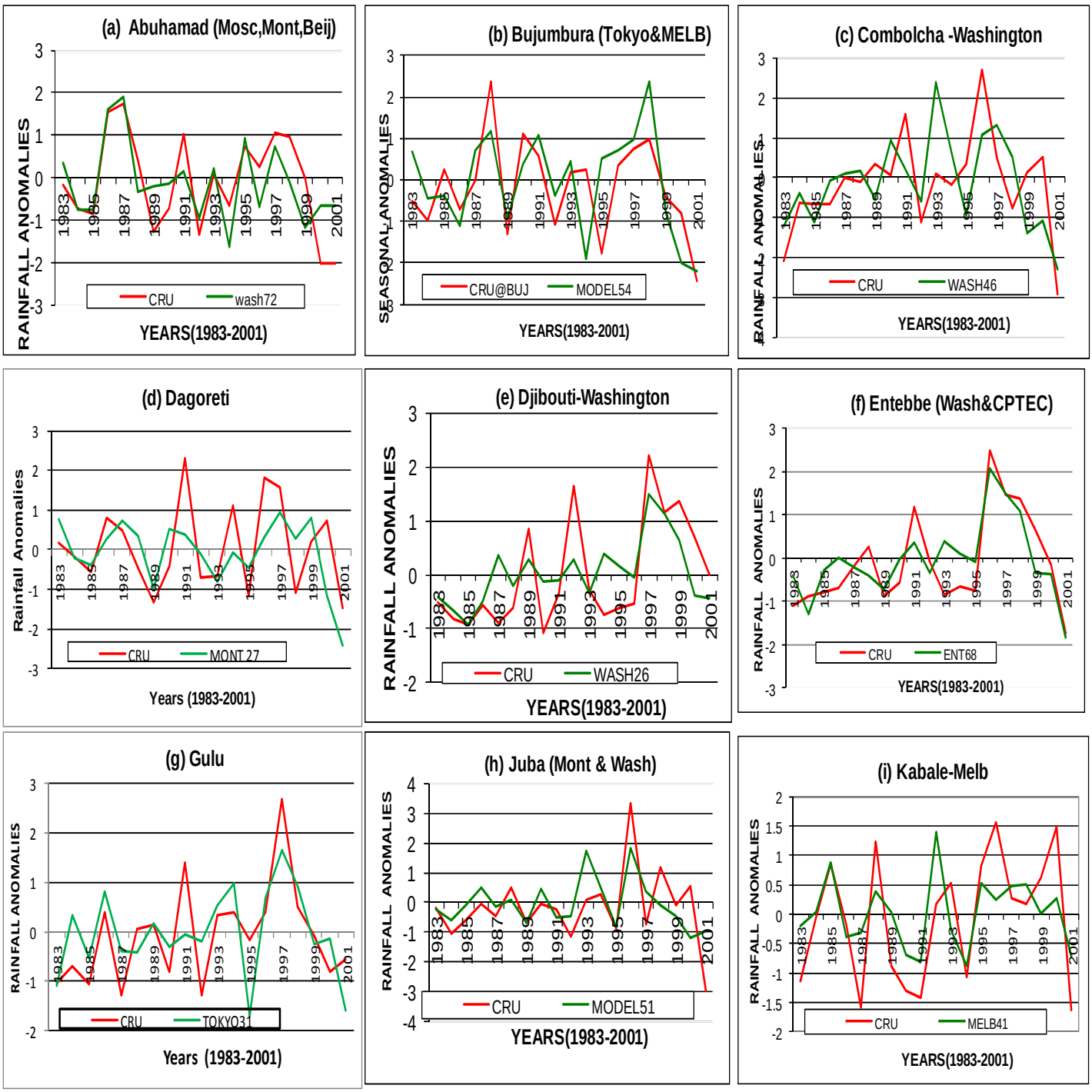

Figure 9 : Inter-annual variability of observed (red colour) and model output (green colour) rainfall anomalies for 1983 – 2001. Only the models output with the highest correlation are shown.

4.4 Correlation and Regression Analyses

Table 4 shows the correlation values obtained between the model output and the observed rainfall. The analysis was done and the significance of each correlation coefficient tested using the student T-test at 95% confidence level. Any correlation value equal and greater than 0.43 was statistically significant. This however explained less 7% of the total variance.

Table 4: Correlation Coefficients between Model output and Observed Rainfall Anomalies at various Stations in the GHA. Green and yellow shading indicates positive and negative significant correlations respectively, while unshaded values indicate insignificant coefficients at 95% confidence level

CRU	Beijing	CPTEC	Melbourne	Montreal	Moscow	Seoul	Tokyo	Washington
Abuhamad	-0.38	0.07	-0.38	-0.39	-0.09	0.16	-0.23	-0.05
Asmara	-0.35	-0.35	0.43	0.21	0.43	-0.09	0.29	0.03
Bujumbura	0	-0.15	-0.39	0.17	0	-0.03	-0.20	0.35
Combolcha	0	0.11	-0.37	0.12	0.12	-0.17	0.29	0.37
Dagoreti	0.05	0.43	0.10	-0.18	-0.14	-0.14	0.44	0.66
Djibouti	0.15	-0.26	0.35	0.58	0.50	0.37	0.38	0.64
Entebbe	0.04	-0.14	0.18	0.19	0	0.23	0.21	0.75
Gulu	0.35	-0.20	-0.18	0.07	-0.02	0.08	0.49	0.43
Juba	0.12	-0.23	0.12	-0.18	-0.14	0.13	0.33	0.51
Kabale	0.38	0.21	-0.56	0.36	-0.45	-0.01	0.18	0.27
Kericho	0.13	0.11	-0.04	0.08	-0.61	0.06	0.39	0.22
Khatoum	0.07	-0.16	-0.06	-0.08	-0.11	0.02	0.28	0.55
Lamu	0.12	-0.22	0.15	-0.16	-0.15	0.14	0.34	0.53
Lodwar	0.12	-0.22	0.15	-0.16	-0.15	0.14	0.34	0.53
Makindu	0.05	-0.57	0.01	-0.22	0.03	0.08	0.16	0.48
Mtwara	-0.06	-0.43	-0.19	0.70	0.09	0.06	0.76	0.66
Mwanza	0.03	-0.23	-0.11	-0.28	-0.01	0.19	0.32	0.37
Narok	0.49	-0.17	0.03	-0.23	0.03	0.11	0.40	0.43
Wajir	0.07	-0.16	-0.23	-0.06	-0.43	0.07	0.08	-0.08
Wau	-0.10	0.20	0.35	0.13	-0.21	0.07	0.26	0.36

From Table 4, it can be noted that the model output from the Washington GPC had the highest number of stations with significant correlation coefficients. Most significant correlations were found across the stations within the Equatorial sector of the GHA and central part of Tanzania of

the study domain. This can be attributed to large scale systems like ITCZ that influence OND rainfall over the Equatorial region.

Muhati *et al*., 2007 did an analysis on the correlation of station to station around the Equatorial sector. Their study found that most rainfall stations in the Equatorial sector correlate well with model rainfall output during the OND season. The distributions of these stations were found to be around the Equatorial sector of the GHA. The stations around the Equatorial sector of the study region were found to have high and significant correlations (Table 4). These findings were consistent with previous studies that found out that correlation is high during the OND season due to the influence of ENSO phenomenon (Muhati et al., 2007).

4.4.1 Distribution of Correlation Coefficients of Model Output and Observed Rainfall

Figure 10 shows the correlations of the model output with significant values plotted to show the spatial distribution of the coefficients during the OND season. High correlations were distributed on the Equatorial sector of the study domain representing Zones 14, 16, 21, 25, 32 and 34. Low coefficients were observed on the north eastern part of Somalia and southern part of Tanzania of the study domain as shown in Figures 10 (a), 10 (b) and 10 (c). This pattern of distribution is consistent with the previous studies that found high correlations between the stations over the Equatorial sector and low correlation values between the stations in the North and southern part of study domain during the OND season. The results show the ability of the models to simulate the climate features around the Equatorial sector of the GHA, and low skill over the northern and southern part of the study domain.

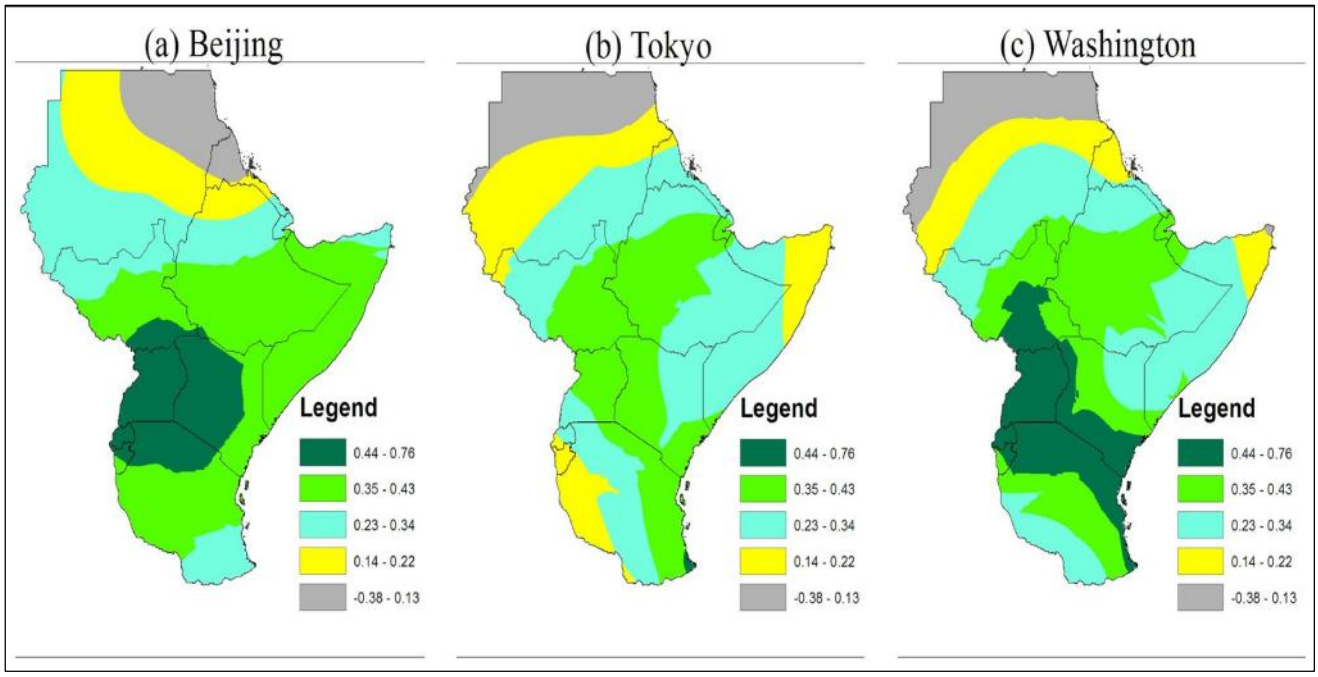

Figure 10 : Spatial distribution of Correlation Coefficients for the Model output (a) Beijing, (b) Tokyo and (c) Washington for the years 1983-2001. Only the best and poor models are shown.

Table 5 shows the performance of the individual models output based on their R-square and P-values from the regression analysis. The regression Equations developed for each station show the models output that were picked at 95% confidence level. High R-square values and low P-values is an indication of good performance of the model output across the stations. Most stations around Equatorial region and lower part of South Sudan had values of R-square above 45% and P-values below 0.1. These results could be an indication that the models have skill around these regions.

Regression analysis helped to determine linear relationships between model output and observed rainfall. The stepwise regression analysis was used as a mean of picking the best individual predictors (model output) into the regression model Equation. The stations around the Equatorial region showed high skill than those from the northern and southern sectors (Table 5).

Table 5: Regression Model Equations developed for the Models output at different Stations with R-square value above 25% and their P-values are shown.

STATIONS	REGRESSION EQUATION	R-SQUARE	P-VALUE
Abuhamad	ABUHAMAD72=0.304*BEIJING +0.596*MONTREAL+0.705*MOSCOW	72	0.012
Bujumbura	BUJU54=-0.983*TOKYO-0.43*MELBOURNE	54	0.024
Djibouti	DJIBOUTI26=0.564*WASHINGTON	26	0.289
Asmara	ASMARA58=-0.234*BEIJING-0.467*CPTEC	58	0.025
Combolcha	COMBOLCHA46=0.989*WASHINGTON	46	0.021
Lodwar	LODWAR37=0.53*WASHINGTON	37	0.049
Kericho	KERICHO36=-0.674*MOSCOW+0.369*MONTREAL	36	0.205
Wajir	WAJIR44=0.0254*MOSCOW	44	0.05
Lamu	LAMU53=0.021*WASHINGTON+0.111*MONTREAL	53	0.012
Dagoreti	DAGORETI27=-0.561*MONTREAL	27	0.123
Mtwara	MTWARA53=-0.281*MONTREAL+0.744*WASHNGTON	53	0.13
Wau	WAU37=0.59*MONTREAL	37	0.064
Juba	JUBA51=-0.308*MONTREAL+0.75*WASHINGTON	51	0.117
Khartoum	KHATOUM48=0.833*WASHINGTON	48	0.032
Mwanza	MWANZA44=-0.455*MELBOURNE	44	0.026
Entebbe	ENTEBBE68=-0.184*CPTEC+0.777*WASHINGTON	68	0.09
Gulu	GULU31=0.716*TOKYO	31	0.075
Kabale	KABALE41=-0.581*MELBOURNE	41	0.035

Table 6 shows models ranked based on their mean R-square, p-value and their frequency. The frequency is the total number of stations out of the 25 stations; each model output had better skill. The model output from the Washington GPC had high skill with mean R-square of 47% and regression coefficient of 0.744 in 8 rainfall stations. The second model output in terms of their skill was Montreal, followed by Melbourne, Beijing, Moscow, CPTEC and Tokyo respectively. The model output from the Seoul GPC had zero dominance and hence low skill. It was noted that at 95% confidence interval, four models i.e. Washington, Montreal, Melbourne and Moscow showed better skill across the rainfall stations within the GHA (Table 6).

Table 6: Models Ranked basing on their Mean R-square (%), Regression coefficients and their Frequency. The Total number of Stations indicates Models dominance within the Study domain

Model output	R-Square	Coefficients	Frequency
Washington	47	0.744	8
Montreal	41	0.634	7
Melbourne	17	1.29	4
Moscow	16	1.00	3
Beijing	16	1.17	2
CPTEC	16	-1.08	2
Tokyo	11	-0.2	2
Seoul	0	0	0
Total		1.041	28

4.5 Categorical Statistics

Table 7 shows the various score that were used to assess the skill and accuracy of the individual models output at various stations. A regression analysis was done to establish the best individual model output at 95% confidence interval level. Only the model output in the regression equation that had high R-square and low P-values were picked at each station. The results for Percent correct, Probability of Detection, Heidke Skill Score, the False Alarm Ratio and bias were calculated from the 3 by 3 contingency Table for the model output picked at each station.

From the analysis of the Percent correct in Table 7, all the stations recorded over 30% correct forecasts. In the Equatorial sector over 47% correct forecasts was achieved with highest score being 68%. These results show that the skill of the individual model is better over the Equatorial sector than other sectors of the study region.

From the analysis of the Heidke skill score (HSS); none of the model output had values close to a perfect score of 100%. The score were especially higher for the stations in the Equatorial sector (e.g., Dagoretti, Kericho, Bujumbura and Lamu).

The analyses from the Bias score show that the perfect score of 100% was achieved in seven instances for the model output presented. The cases of forecasting nearing almost perfect were

achieved in ten instances. The cases of over forecasting were more than those under-forecasted especially for most stations in the northern and southern sectors of the GHA (Table 7).

From the analysis of the Probability of Detection, for the normal category, 13 instances predicted above 50%; for the above normal category, 10 instances predicted above 50% and 8 instances predicted above 50% for below normal category. PoD gives the proportion of rainfall events successfully forecasted by the model. For a good forecast the PoD is 100%. Most of the stations around the Equatorial sector had score above 50%, while stations in the Northern and Southern regions recorded score less than 50% (Table 7). This indicates that the model successfully forecasts more than half of the rainfall events in the Equatorial region.

From the analysis of FAR score for the below and above normal categories, stations around Equatorial region recorded score less than 50% while those in the northern and southern sectors of the region recorded score of more than 50%. For a good forecast FAR score is 0%. From Table 7 the score was above 50% in some instances, indicating that when the model forecasted above normal or below normal rainfall the observation was not in the same category. Most models output across the stations recorded more than 50% in most instances except for a few stations to the north of the GHA.

Table 7: Percent correct (%), Probability of detection (POD) (%), False Alarm (FAR) (%), BIAS (%) and Heidke Skill Score (HSS) (%) for the various model output picked across all the Stations at different Categories, Below Normal (BN), Normal (N) and Above Normal (AN)

Stations	Models Output	Percent Correct	POD			FAR		BIAS			HSS
			BN	N	AN	BN	AN	BN	N	AN	
Abuhamad	Beijing, Montreal and Moscow	53	33	57	66	60	0	83.3	142	66	27
Bujumbura	Tokyo and Melbourne	53	50	16.7	86	50	40	100	50	140	58
Asmara	Beijing and CPTEC	47		57	83	0	50	0	129	167	22
Combolcha	Washington	42	50	28	50	50	62	100	70	130	14
Dagoretti	Montreal	42	50	29	50	25	67	70	86	150	46
Djibouti	Washington	68	33	86	83	0	37	33	130	133	52
Entebbe	Washington and CPTEC	52	33	57	67	0	50	33	133	133	30
Gulu	Tokyo	31	33	71	67	33	20	50	150	80	36
Juba	Montreal and Washington	68	67	71	67	20	43	80	100	100	50
Kabale	Melbourne	42	57	86	33	33	35	0	200	50	30
Kericho	Moscow and Montreal	63	67	71	50	20	40	80	130	80	40
Khartoum	Washington	37	50	14	50	50	57	100	90	100	50
Lamu	Washington and Montreal	63	83	43	67	29	43	120	70	120	50
Lodwar	Washington	47	33	43	67	33	50	50	110	130	20
Mtwara	Montreal and Melbourne	63	33	100	50	0	0	30	200	50	40
Mwanza	Melbourne	42	33	57	33	60	33	80	160	50	10
Wajir	Moscow	32	17	57	17	50	83	30	160	100	50
Wau	Montreal	53	50	57	50	50	25	100	130	70	30

4.6 Ensemble Model Output

The four individual models Washington, Montreal, Melbourne and Montreal were subjected to further analysis since they had better skill (Table 6). The ensemble models was developed according to Equations 9 and 10 by assigning the model which had high skill the largest weight and the model which had low skill the smallest weight possible. The R-square and the regression coefficients from regression analysis were used as the weights for each of the four models output.

The weight for each of the models was determined by multiplying its R-square value with regression coefficients and the product divided by the sum total of the coefficients. Washington model had higher R-square value and smallest regression coefficients. It was assigned a weight of 0.36, Montreal was assigned a weight of 0.27, Melbourne 0.20 and Moscow model that had relatively low R-square and high regression coefficient was assigned a weight of 0.17.

4.6.1 Ensemble Model Output and Observed Rainfall for 1983-2001

Figure 11 shows the results of the spatial analysis of the first model Ensemble (ENSE1), second model ensemble (ENSE 2) and observed rainfall distributions for the years 1983-2001. The spatial distribution pattern of rainfall for ENSE 1 and ENSE 2 show that most rainfall was over the Equatorial Sector (Figures 11 a, 11 b and 11 c). The distribution pattern of ENSE 2 model output was very close with the observed rainfall pattern (Figure 11 c).

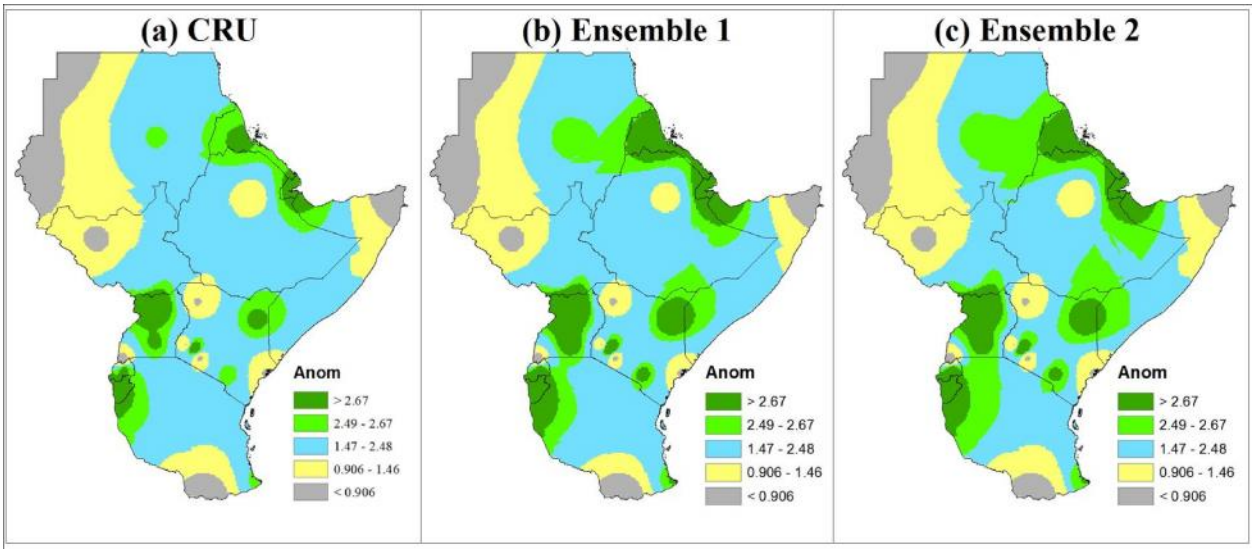

Figure 11: Spatial Distribution of (a) CRU, (b) ENSE 1 (c) ENSE 2 for 1983-2001.

4.6.2 Ensemble Model Output and Observed Rainfall for El Niño year (1997)

Figure 12 shows the spatial distribution of the observed rainfall, ENSE 1 and ENSE 2 model rainfall output over the study domain. The rainfall distribution pattern is concentrated on the western sector of the GHA. ENSE 2 depicted high simulation of observed rainfall pattern than ENSE 1 as shown in Figures 12 (a) , 12 (b) and 12 (c). It can be observed that during El Niño year for the OND season, the distribution of rainfall pattern was on the Equatorial sector of the study domain.

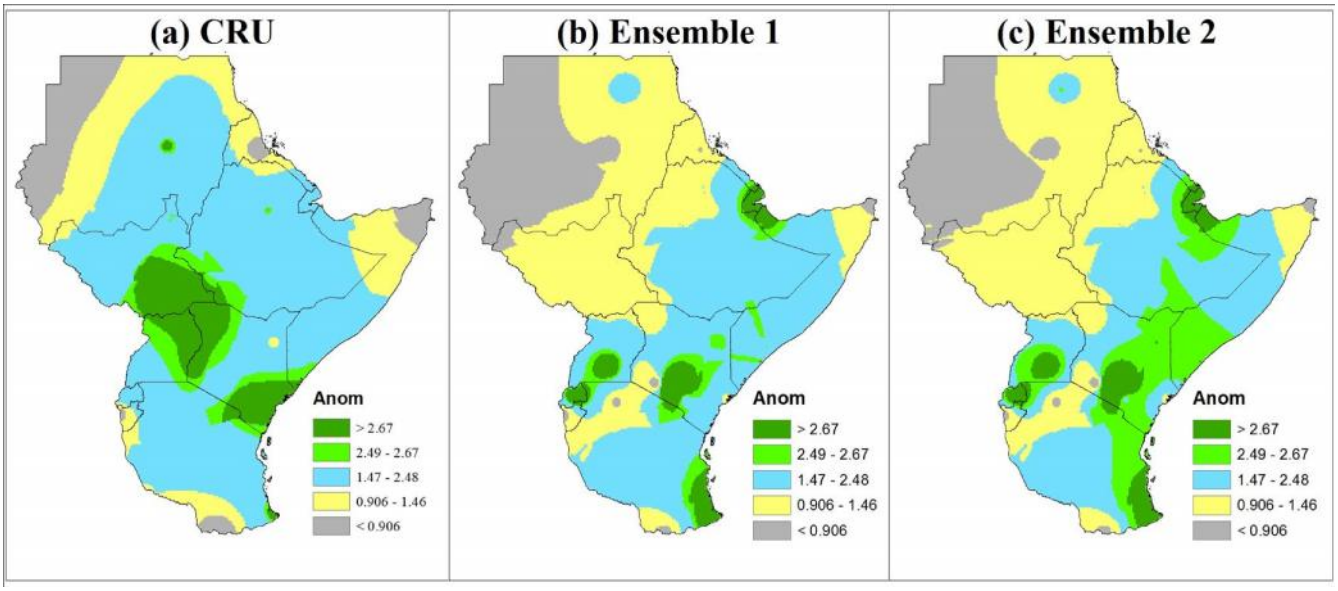

Figure 12: Spatial distribution of (a) CRU (b) ENSE 1, (c) ENSE 2 models output for 1997.

4.6.3 Ensemble Model Output and Observed Rainfall for La Niña Year (2000)

Figure 13 shows the spatial distribution for the observed rainfall, ENSE 1 and ENSE 2 models output over the study domain. The rainfall distribution pattern was concentrated on the western sector of the GHA during the year 2000 which is La Niña year. The distribution pattern of rainfall was better in ENSE 2 than ENSE 1 models output as shown in Figures 13 (b) and 13 (c). It can be observed that during La Niña year for the OND season, the distribution of rainfall pattern was on the Equatorial sector of the study domain.

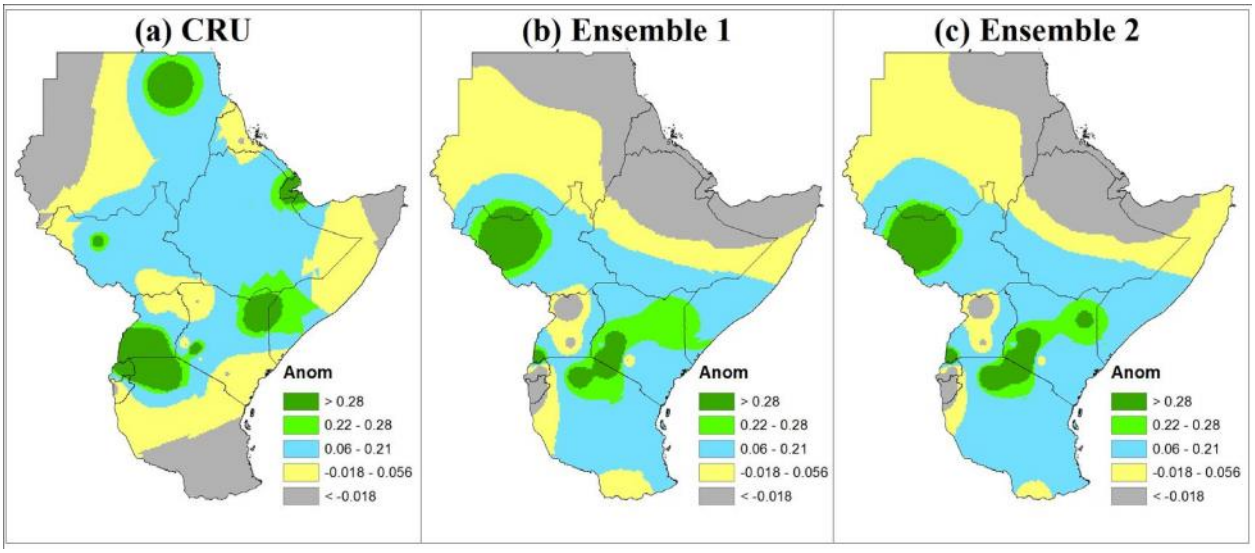

Figure 13: Spatial distribution of (a) CRU (b) ENSE 1, (c) ENSE 2 models output for 2000.

4.7 Enhanced and Depressed Rainfall Years

Analysis was done to select years with enhanced and depressed historical rainfall records. The seasonal long term mean was calculated for each of the 19 years in some few selected stations. Any year which recorded above 125% of seasonal long term mean was regarded as enhanced, while any year which recorded less than 75% of seasonal long term mean was regarded as a depressed year. The years 1986, 1987, 1991 and 1997 had enhanced rainfall records while the years 1984, 1988, 1995, 1998 and 1999 had depressed rainfall records. These definitions are according to World Meteorological Organization standards.

4.7.1 Temporal Analysis of the Ensemble Model Output for Years with Enhanced Rainfall

Figure 14 shows the observed trend of the observed rainfall, ENSE 1 and ENSE 2 models output for the years with enhanced historical rainfall period. The plotted seasonal rainfall anomalies were for few selected stations that had high correlations with the models output. The years analyzed were 1986, 1987, 1991 and 1997. The rainfall stations in the north i.e. Abuhamad, Asmara and Kigali fairly simulated the observed trend of the rainfall pattern as shown in Figures 14 (a), 14 (b) and 14 (d). Amongst the four years of enhanced rainfall, the year 1997 was well replicated by the ensemble model mostly by the stations in the Equatorial sector that had high correlations as shown in Figures 14 (c), 14 (e), 14 (f), 14 (g), 14 (i) and 14 (j). Previous studies by Indeje *et al.*, (2000) have shown that the year 1997 had a strong link to the ENSO event which is the main driver of the rainfall

pattern during the OND season. The ENSE 1 mimicked well the pattern of the observed rainfall compared to the individual models.

The years 1986, 1991 and 1997 were well replicated by both the ENSE 1 and ENSE 2 models output. Both the ensemble models output showed improvement in the forecast in terms of their inter-annual Variability. The ENSE 2 model output was better than the ENSE 1 model in some stations where both the individual and ENSE 1 model performed poorly as shown in Figures 11 (a), 11 (c) , 13 (b),13 (c) , 14 (g), 14 (h) and 14 (k).

Both the models output failed to reproduce the correct amplitude of the El Niño event. These results are similar to those reported by Bosire (2012) using the CFS model from Washington. The study found that both the individual models output and ensemble mean were not able to represent the correct amplitude of the El Niño event. Mutemi (2003) got similar results using ECHAM 4.5 showing that the correct amplitude linked to the inter-annual variability of El Niño episode 1997 was not well reproduced.

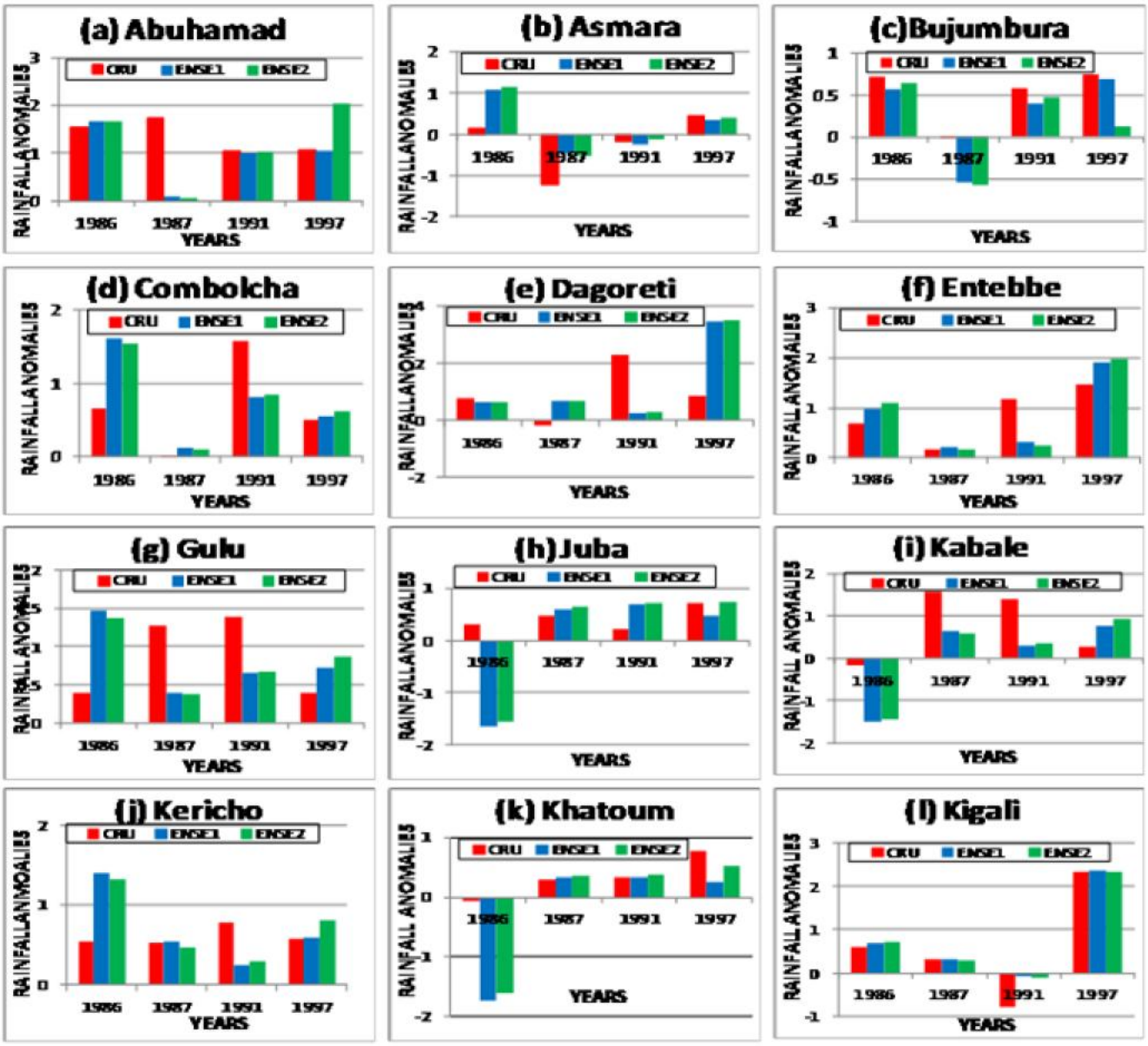

Figure 14: Ensemble model output and observed rainfall. Only years with enhanced rainfall records are plotted across the stations with high correlations. Red shading represents CRU, Blue shading represents ENSE 1 and green shading represents ENSE 2.

4.7.2 Ensemble Models Output for Years with Depressed Rainfall

Figure 15 shows the analysis between Ensemble models output and observed rainfall anomalies for six years with depressed rainfall records i.e. 1984, 1988,1995,1998,1999 and 2000. Only stations with significant correlations with the model output are plotted.

From the analysis of the results, the year 2000 had the strongest link to the ENSO period as previously reported by Indeje *et al.*, (2000). The ENSE 1 model output mimicked fairly well the observed rainfall pattern compared to the individual models output (Figure 15). Out of the six years used for investigation, the years 1984, 1988, 1999 and 2000 were well replicated by both the ENSE 1 and ENSE 2 model output. Both the ensemble models output showed some improvement in the forecast in terms of their inter-annual variability.

The ENSE 2 model was better than ENSE 1 model output in some stations where both the individuals and ENSE 1 models performed poorly as shown in Figures 14 and 15. The 2000 La Niña year was fairly well replicated by both the ensemble models. Both the individual and ensemble models did not reproduce correctly the amplitude of the La Niña event. These results are similar to those reported by Bosire (2012) using global model CFS from Washington in that both the models and ensemble mean were not able to represent the correct amplitude of the El Niño event.

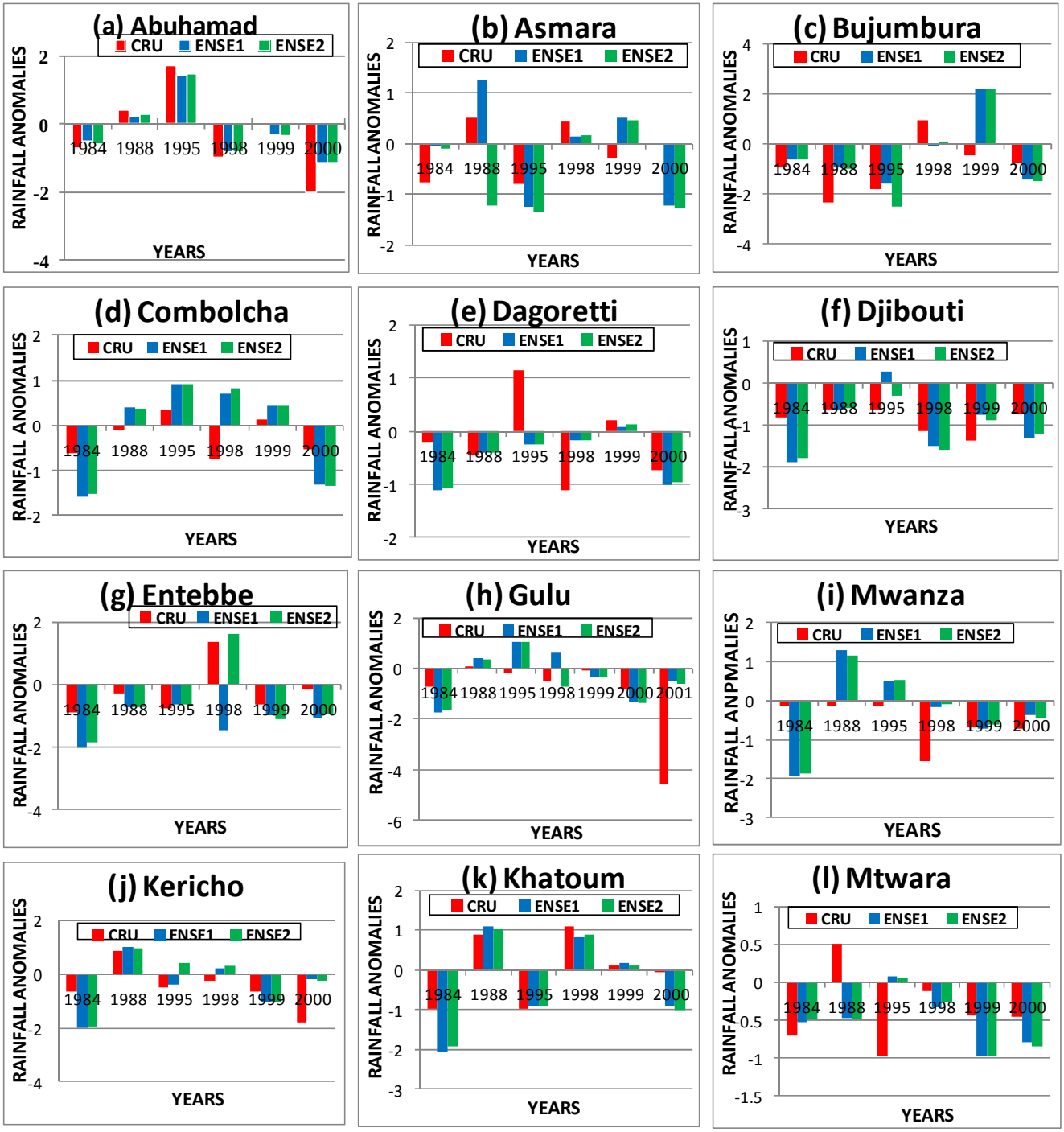

Figure 15: Ensemble model output and observed rainfall. Years with Depressed rainfall records are plotted across the stations with high correlations only. Red shading represents CRU, Blue shading represents ENSE 1 and green shading represents ENSE 2.

4.7.3 Inter-Annual Variability of Observed and the Ensemble Model Output Rainfall for 1983-2001

Figure 16 shows the year to year inter-annual variability between the observed rainfall and model ensembles for the years 1983 to 2001. The training period for the ensemble model was 10 years (1983-1992) while the testing period was 9 years (1993-2001). Linear correlations and inter-annual variability revealed strong relationships between the observed rainfall and ensemble models output. Regions that exhibited strong relationships with the observed rainfall were confined to the Equatorial region of the GHA as shown in Figures 16 (c), 16 (e), 16 (h), 16 (i), and 16 (j).

The best relationships between the ensemble models and observed rainfall were found during the transition between wet and dry regimes, or entirely within the dry or wet seasons. For example the years 1997 and 2000 had high inter-annual variability between the ensemble and observed rainfall (Figure 16). The variability in seasonal rainfall was quite well represented by the model output in the years 1997 and 2000 which were El Niño and La Niña years respectively. The Ensemble model output failed to replicate the observed pattern for the year 1996 in several regions. The ensemble models output replicated the observed pattern fairly well compared to the performance of the single models as shown earlier in Figure 10.

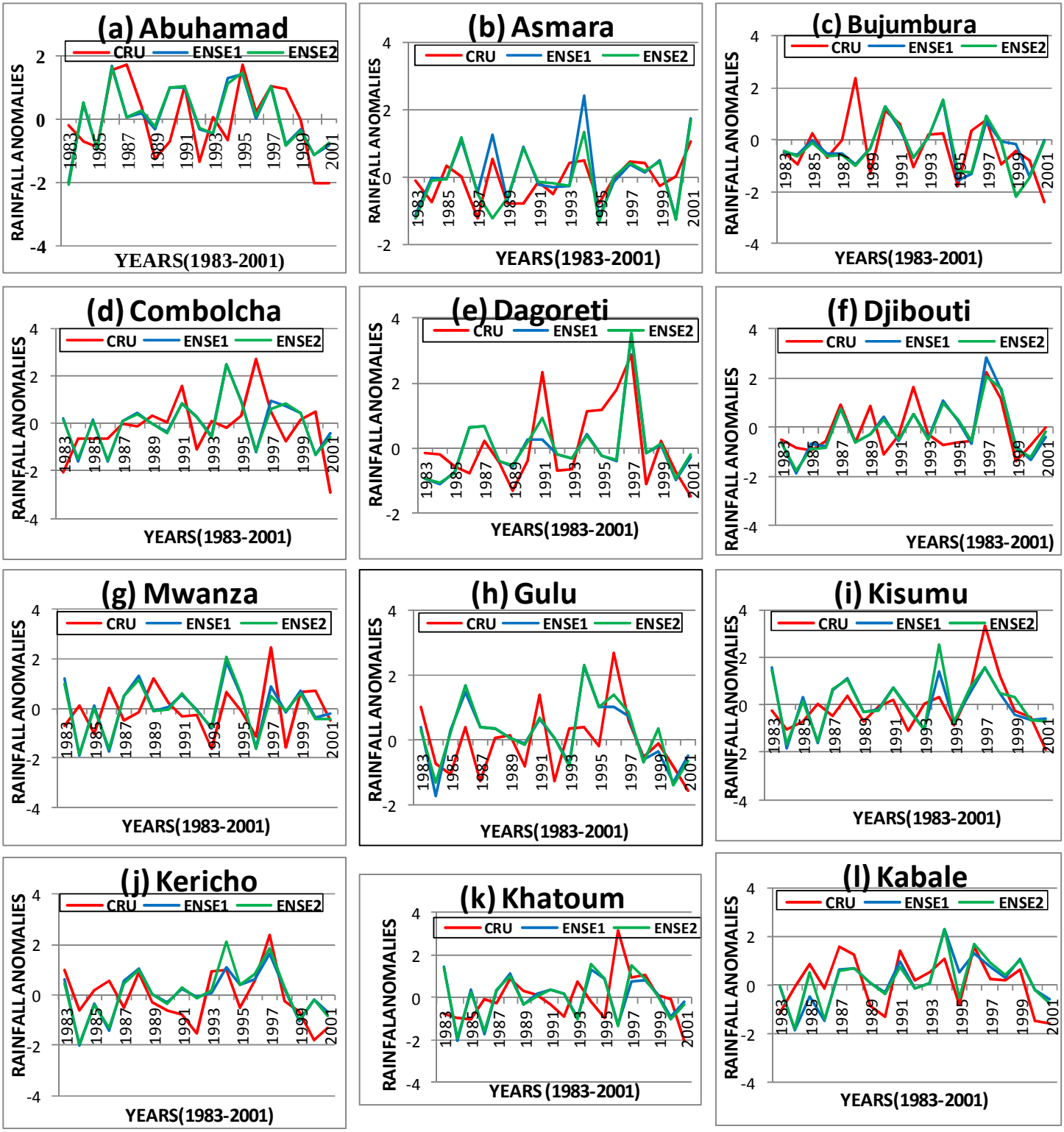

Figure 16: Inter-annual variability of rainfall anomalies between the observed and the Ensemble model output for the years (1983-2001). Only stations with high correlations are shown. Red shading represents CRU, blue shading represents ENSE 1 and green shading represents ENSE 2.

4.8 Correlation and Categorical Statistics between Observed and Ensemble Model Output Based on Simple composite and Weighted Average

This subsection presents the results obtained from correlation and various categorical score for the Ensemble models output.

4.8.1 Correlation between Observed Rainfall and Ensemble Models Output

The first Ensemble (ENSE 1) model output based on simple composite was correlated with observed rainfall to establish the linear relationship between the new model generated and the observed rainfall. To further improve the skill of the Ensemble model, the ENSE 2 model was developed by weighted average method according to Equation 10.

Table 8 shows the correlation coefficients calculated for each ensemble models output. The green shading in the table indicates those correlations which are positive and significant while yellow shading shows correlation coefficients which are negative and significant at 95% interval level. There was an improvement on the number of significant and positive correlations. Five stations in the northeastern part of the region i.e., Asmara, Djibouti, Khartoum, Wajir and Wau had correlations which were significant. The correlation coefficients from the Ensemble model output were compared with those from the mean of individual models output.

From Table 8, there was an improvement in the correlation coefficient values especially in the northern sector of the GHA region. For example in Table 4 the coefficients for the stations Abuhamad, Khartoum, Asmara are in the range of -0.5 to 0.4 but in Table 8 the coefficients are in the range of 0.2 to 0.7.

The coefficients increased especially for stations around the Equatorial sector. For example in Table 4 stations like Dagoreti, Entebbe, Gulu, Kericho and Kabale the correlation coefficients are in the range of -0.5 to 0.6. In Table 8 the coefficients are in the range of 0.3 to 0.7. The improvements in the coefficients were also realized from the ENSE 1 to ENSE 2 model output. For example the coefficients for the stations Abuhamad, Khartoum, Asmara and Juba for the ENSE 1 are 0.21, -0.7, 0.56 and 0.68. For the ENSE 2 models output the correlation coefficients are 0.23, 0.75, 0.56 and 0.73. The stations in the Equatorial sector i.e. Dagoreti, Entebbe, Gulu, Kericho and Kabale the coefficients for the ENSE 1 are 0.63, 0.61, 0.76, 0.71, 0.62 and 0.33. For the ENSE 2 the coefficients from the same stations are 0.65, 0.55, 0.78, 0.64, 0.62 and 0.37.

These observations showed improvements from the individual model output analysis. It is evident from the findings that the GPCs models output have high skill over the Equatorial sector and low skill in the Northern and southern sectors of the GHA region.

Table 8: Correlation between Observed, ENSE 1 and ENSE 2 model output Rainfall anomalies respectively. Green shading represents positive significant correlation, yellow shading represents negative significant and unshaded represents insignificant correlation coefficients at 95% interval level.

Stations	ENSE 1	ENSE 2
Abuhamad	0.21	0.23
Asmara	0.56	0.26
Bujumbura	0.59	0.64
Combolcha	0.29	0.43
Dagoretti	0.63	0.65
Djibouti	-0.35	0.55
Entebbe	0.61	0.66
Gulu	0.76	0.78
Juba	0.68	0.73
Kabale	-0.35	-0.37
Kericho	0.62	0.62
Khartoum	-0.70	0.75
Kigali	0.64	0.64
Kisumu	0.71	0.76
Lamu	0.71	0.76
Lodwar	0.71	0.76
Makindu	0.69	0.73
Mtwara	0.71	0.70
Mwanza	0.29	-0.54
Narok	0.55	0.61
Wajir	-0.39	-0.59
Wau	0.32	0.30

4.8.2 Distribution of Correlation Coefficients for the Ensemble Models Output

Figure 17 shows the spatial distribution of correlation coefficients for the ENSE 1 and ENSE 2 models output over the GHA. This was done to show the regions within the study domain where the correlation indices were high and those regions over the study domain where the correlation were low during the OND season for the ENSE 1 and ENSE 2 models output.

High correlation values were concentrated around the Equatorial sector. Other regions that showed high correlation with the ensemble models were central Ethiopia and the Kenyan Coastal strip (Figure 17 b). Correlation values ranging between 0.4 and 0.5 were evenly distributed over the most sectors of the study region as shown in Figure 17 (b). Low coefficients were observed on the northern part of the Kenya, Sudan, north eastern part of

Somalia and southern part of Tanzania of the study domain. This could be attributed to large scale systems like ITCZ which are the main drivers of seasonal rainfall over the Equatorial sector. The results revealed an improvement in the ability of the ensemble models to replicate the climate features around these sectors better with high skill and accuracy than the individual models.

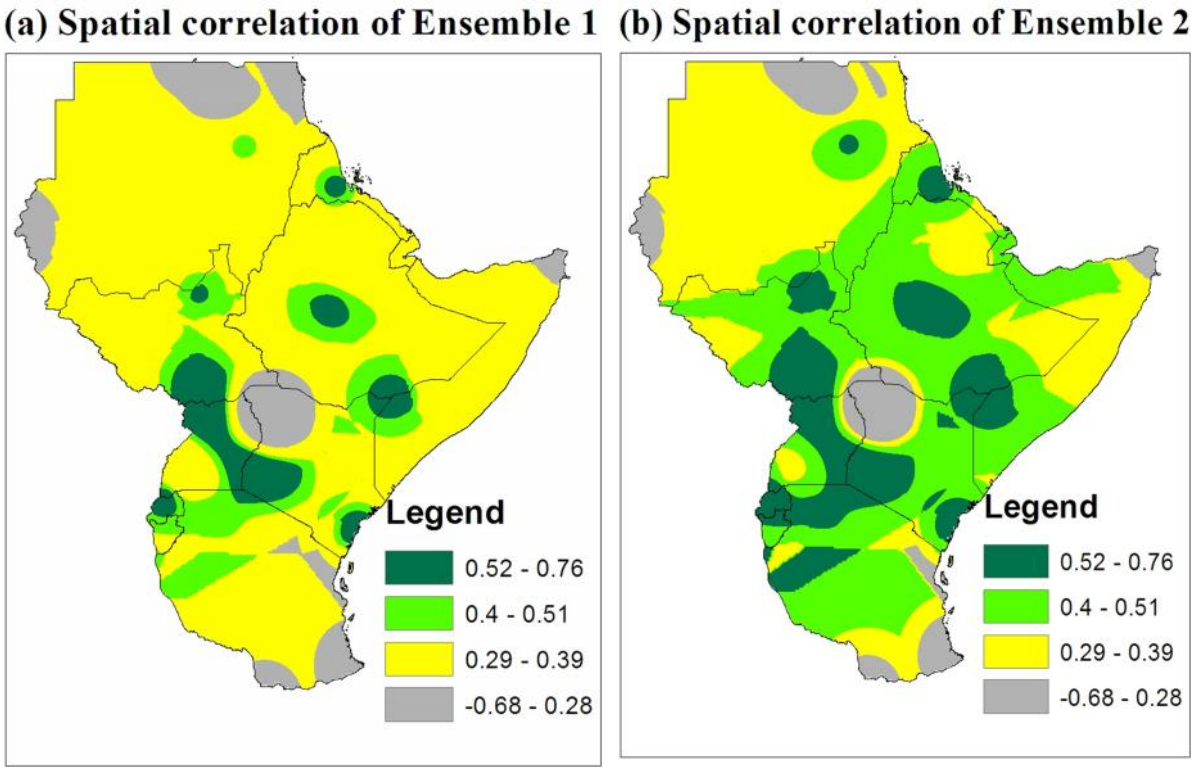

Figure 17: Distribution of correlation coefficients for (a) ENSE 1 and (b) ENSE 2 models output over the study Domain. The indices range between -0.68 and 0.76

4.8.3 Categorical Statistics for the Ensemble Models

Table 9 shows the results for Percent correct, probability of detection (PoD), Bias, Heidke skill score (HSS) and the false alarm ratio (FAR) calculated from the 3 by 3 contingency Table for the ENSE 1 and ENSE 2 models output.

From the analysis of Percent correct in Table 9, the number of correct forecasts significantly increased compared with those from the individual model output (Table 7). For example the number of cases where the models predicted rainfall events correctly was above 50% in 8 stations out of the 19 stations for the individual models. The number of percent correct above 50% was observed in 6 stations for the Ensemble models as shown in Table 9. Improvement in terms of correct forecasts was noted in ENSE 2 than in ENSE 1. The models performance

improved for stations in the northern sector from 32-53% to 37-58%, as shown in Tables 7 and 9 respectively. This shows the ability of the Ensemble models to improve the accuracy and skill of the seasonal rainfall forecasts.

From the analysis of the Heidke Skill Score (HSS), none of the models presented had values close to the perfect score of 100%. The score was high especially for stations in the Equatorial sector for the individual models (e.g., Dagoretti, Kericho and Lamu) as shown in Table 7. For the ensemble models the HSS values improved across some stations with at least seven stations obtaining values above 50% in the Equatorial sector as shown in Table 9. This is compared to only five stations getting above 50% for the individual models (Table 7).

From the analysis of Bias score for below normal, normal and above normal categories, the perfect score of 100% was achieved in 17 instances for the model Ensembles, as shown in Table 9. The cases of ensemble model giving forecast nearing almost perfect score was achieved in ten instances. The values shown in Table 9 indicate that the cases of over forecasting and under forecasting greatly reduced for stations in the Equatorial sector compared to the values obtained for the Individual models (Table 7).

The analysis from the Probability of Detection (PoD) score for the ensemble models output indicates that, 12 instances predicted above 50% for the normal category, 15 instances predicted above 50% for the above normal category and 8 instances predicted above 50% for below normal category. PoD gives the proportion of rainfall events successfully forecast by a model. For a good forecast the PoD score is 100%. From Table 9 most of the stations around the Equatorial sector recorded score above 50%, while stations in the northern and southern regions obtained values above 50% in most instances indicating an improvement in the skill of the forecast using ensemble approach. The ENSE 2 model at these regions successfully forecasted more than half of the rainfall events. The improvement in the forecast skill by ensemble model shows the model ability to resolves correctly the systems that influence rainfall over the Equatorial, Northern and Southern sectors of the study domain well.

The results for False Alarm Ratio (FAR) score for the below normal and above normal categories indicate that most stations around the Equatorial region recorded score less than 50% while those in the northern and southern sector of the region slightly scored below 50% in few cases. Stations that recorded score more than 50% reduced in the ensemble models than for the individual ones as shown in Tables 7 and 9. For a good forecast FAR score is

0%. From Table 9 the score was above 50% in few instances compared to when the individual models were used. This indicated that when the model forecasted above normal or below normal rainfall the observation was not in the same category. The Ensemble models across the stations recorded less than 50% in most instances except for a few stations to the north of the GHA.

Table 9: Percent correct (%), POD (%), FAR (%), BIAS (%) &HSS (%) for the ENSE 1 and ENSE 2 models for Below Normal (BN), Normal (N) and Above Normal (AN) categories.

Stations	Ensemble	Percent Correct	POD			FAR		BIAS			HSS
			BN	N	AN	BN	AN	BN	N	AN	
Abuhamad	ENSE1	58	50	57	67	40	42	83	100	117	30
	ENSE2	58	50	57	67	40	42	83	100	117	42
Bujumbura	ENSE1	47	33	71	86	43	40	67	50	83	28
	ENSE2	51	17	71	80	41	37	50	50	83	43
Asmara	ENSE1	47	67	43	67	32	42	83	100	117	31
	ENSE2	59	50	43	67	29	41	100	100	100	39
Combolcha	ENSE1	36	50	43	50	50	62	100	71	130	36
	ENSE2	37	47	43	51	50	61	67	86	120	42
Dagoretti	ENSE1	26	50	29	50	25	67	70	86	1	36
	ENSE2	42	47	42	53	21	63	71	93	130	47
Djibouti	ENSE1	47	33	86	83	12	37	83	71	150	52
	ENSE2	47	33	67	82	13	33	100	78	150	55
Entebbe	ENSE1	52	33	57	67	15	33	33	133	133	35
	ENSE2	51	32	56	63	15	30	50	123	110	41
Gulu	ENSE1	31	42	71	50	33	20	83	100	117	31
	ENSE2	32	42	50	61	29	21	83	129	83	40
Juba	ENSE1	26	67	71	67	20	43	83	57	100	51
	ENSE2	31	59	50	68	32	40	100	43	123	53
Kabale	ENSE1	37	0	57	50	33	23	50	130	100	32
	ENSE2	42	0	57	60	32	31	50	121	100	37
Kericho	ENSE1	58	57	71	50	20	40	67	103	83	36
	ENSE2	58	59	33	86	50	25	67	100	67	45
Khatoum	ENSE1	37	50	14	50	47	54	83	90	100	45
	ENSE2	39	50	23	51	63	50	83	43	120	49
Kigali	ENSE1	36	50	14	50	29	33	117	57	50	36
	ENSE2	42	67	50	67	21	32	103	57	52	47
Lamu	ENSE1	37	83	43	67	29	43	100	43	120	41
	ENSE2	45	50	50	67	30	41	100	57	130	51
Lodwar	ENSE1	47	33	50	67	30	47	83	57	130	28
	ENSE2	49	33	57	69	31	40	100	53	1.2	37
Mtwara	ENSE1	37	33	57	50	0	0	133	71	100	49
	ENSE2	63	47	59	50	12	10	133	71	100	52
Mwanza	ENSE1	37	33	43	33	60	33	83	100	160	24
	ENSE2	42	33	43	33	25	29	100	71	133	37
Wajir	ENSE1	32	17	33	43	50	83	83	114	100	55
	ENSE2	33	17	33	43	33	25	100	100	100	58
Wau	ENSE1	53	33	71	50	50	25	67	129	100	32
	ENSE2	53	33	71	50	40	30	67	129	100	37

The Ensemble models had better score than the individual model output (Table 7). ENSE 2 models had better skill than the ENSE 1 model. From Table 7, cases of over forecasting and under-forecasting were many in individual models output especially for most stations in the Northern and Southern sectors. The ENSE 1 and ENSE 2 models had close to 9 instances of perfect score. The ENSE 2 had many instances where the score was above 50% (Table 9).

There are at least ten instances when the PoD score is above 70%. These stations like Bujumbura, Gulu, Kericho and Djibouti are those in the Equatorial and Northern sectors. The score are low for some stations in the Southern sector of the GHA as shown in Tables 7 and 9 implying the models inability to detect the signal of rainfall over these sectors. Washington, Melbourne, Montreal and Moscow models had better skill and accuracy. The skill of the Ensemble models was better than the skill of each of the models output (Table 9). In general several models were able to indicate the direction of the anomalies during the extremely wet and dry years but such extremes were underestimated or overestimated in some cases.

CHAPTER FIVE

5.0 SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

This chapter provides a summary of the results obtained from the various methods used to achieve the objectives of the study. The chapter also provides the conclusions drawn and the recommendations made.

5.1 Summary

The overall objective of the study was to improve the seasonal rainfall forecasting using global models as multi-model ensembles during the OND season over the GHA region. The data used in the study included the gridded rainfall data from the Climate Research Unit, University of East Anglia and model hindcast data from eight global producing centers from 1983 to 2001. A total of 25 rainfall stations were used; these stations were selected based on the homogeneous rainfall zones adopted for the study. The methodology employed for the study involved graphical plots for the spatial analysis, correlation analysis, time series analysis, regression analysis, model output statistics (MoS), simple and weighted averages and categorical statistics.

The spatial patterns of the individual model output from Washington, Montreal, Melbourne and CPTEC centers were closest to the observed rainfall pattern. The largest departure from observations in this season was observed in the northern and southern sectors of the GHA. These models were able to indicate the direction of the anomalies during the extremely wet and dry years but such extremes were under estimated or over estimated in most cases.

The analysis for the correlation coefficients showed higher coefficients for the ensemble model output than for the individual models. The coefficients were higher for most of the stations in the Equatorial sector (5°N to 5°S) than for the stations in the northern and southern sectors of the GHA. This is an indication that the ensemble model was able to resolve better the different features that enhance convection over the Equatorial sector during the OND season.

Categorical statistics score showed higher skill for the ensemble models than for the individual models output. The skill and accuracy of the forecasts was enhanced especially in the sectors where individual models had low skill by the use of MoS technique for

downscaling. There was improvement in the forecasting skill noted in the northern sector of the region when ENSE 2 model was used for forecasting. The skill and accuracy of the models were relatively higher during the start of ENSO event and becomes low towards the end of ENSO period.

5.2 Conclusions

The model output from Washington, Montreal, Melbourne and Moscow showed better skill than the rest of the models for seasonal prediction. The models have shown high skill over the Equatorial than over the northern and southern sectors of the GHA.

The models can be used for forecasting with high skill and accuracy at the onset of the OND season since the skill was higher at the start and becomes low at the end of the ENSO events. The models were not able to resolve the convective pattern over the regions they underestimated or overestimated forecasts during the El Niño and La Niña events.

The weighted average method was able to minimize the errors in the individual model and allowed for an optimal linear combination of the individual model forecasts by taking account of the relative skill of each model hence the improved forecasts skill observed using the ensemble models.

The reliable and accurate forecasts would improve early warning systems in the various sectors like Agriculture, Transport, Hydrology, Geothermal, within the GHA region.

5.3 Recommendations

Further research using all the 12 global producing centre models needs to be done to assess the improvement in the skill of the forecast. There is need for a similar research with 12 global models for MAM season to contrast and compare the performance of the model output for the OND and MAM seasons.

To get better resolutions of the local features in the region the observed stations data should be included for analyses. This reduces the error that could be inherent in the CRU datasets during interpolation of the rainfall data.

A detailed research on the systems that influence climate over the Northern and Southern sectors of the GHA is necessary. This knowledge would improve the understanding of the systems that are dominant over these regions and hence better forecasting methodologies and

tools for the GHA region. The physics and the configuration of the models should be studied in detail to enhance the understanding of the global models and their suitability in the seasonal predictions over the GHA region.

Since the ensemble models significantly improved the skill of the forecasts over the Equatorial sector, this study therefore recommends the use of the ensemble models developed for seasonal prediction over the Equatorial sector of the GHA for the OND season.

There are many ways to combine model output to generate an Ensemble model. In this study simple average and weighted mean was used. Other techniques like Bayesian method which are relatively accurate can be employed for the study.

6.0 REFERENCES

AchutaRao, K. M., and Sperber, K. R., 2006: ENSO Simulation in Coupled Ocean-Atmosphere Models: Are the Current Models Better? *Clim Dyn*, **27**, 1-15.

ACMAD, 2010: Learning forum on improving the accessibility and usability of climate information, Nairobi. *www.africa-adapt.net/media/resources/334/ACMAD.pdf*.

Anyah, R., and Otieno, V., 2012: CMIP5 simulated climate conditions of the Greater Horn of Africa. *Clim. Dyn*, doi 10.1007/s00382-012-1549 **35**, 455-463.

Anyah, R., and Semazzi, F. H. M., 2006: Climate variability over the Greater Horn of Africa based on NCAR AGCM ensemble. *Theor. Appl. Climatol.* **86**, 39–62.

Arnell, N. W., Hudson, D. A., and Jones, R. G., 2003: Climate change 256 scenarios from a regional climate model: Estimating change in runoff 257 in southern Africa, *J. Geophys. Res.*, **108**(D16), 4519, doi: 10.1029/ 2582002JD002782.

Arribas, A., and coauthors, 2011: The GloSea4 Ensemble Prediction System for Seasonal Forecast. *Mon Wea Review*, **139**, 1891-1910.

Boer, G. J., 2009: Changes in Inter-annual Variability and Decadal Potential Predictability under Global Warming, *J. Climate*, **22**, 3098–3109.

Boer, G. J., and Lambert, S. J., 2001: Second order space–time climate difference statistics; *Clim Dyn*, **17,** 2/3, 213-218.

Bosire, E.N., 2012: Assessment of the predictability of rainfall on seasonal timescales over east Africa using the Climate Forecast System Model, *MSc* Thesis, *Dept. of Meteorology, University of Nairobi.*

Bowden, J. H., 2004: Recent and Projected Climate Variability during the Seasonal Rains of the Greater Horn of Africa. *MSc. Thesis, Marine, Earth,and Atmospheric Science, North Carolina State University, 213 pp.*

Chavas, D., 2008: Seasonal climate prediction dissemination to rural farmers in sub-Saharan Africa: A "bottom-up" perspective and the emergence of the mobile Massachusetts Institute of Technology. *Clim Dyn*, **17**, 213–218.

Cottrill, A., and Coauthors, 2013: Seasonal Forecasting in the Pacific Using the Coupled Model POAMA-2. *Wea forecasting*, ***28***, 668–680. doi: http://dx.doi.org/10.1175/WAF-D-12-00072.1 .

Déqué, M., 2001: Seasonal predictability of tropical rainfall: probabilistic formulation and validation. Tellus, **53A**, 500-512.

Doblas-Reyes, F. J., 2012: Proceedings of ECMWF Seminar on Predictability in the European Seasonal prediction over Europe and Atlantic Regions from Days to

Years. *Institut Català de Recerca i Estudis Avançats, Barcelona, Spain Institut Català de Ciències del Clima (IC3), Doctor Trueta 203, 08005 Barcelona, Spain.*

Doblas-Reyes, F.J., Hagedorn, R., Palmer, T.N., and Morcrette, J. J., 2006: Impact of increasing greenhouse gas concentrations in seasonal ensemble forecasts. ECMWF Technical Memorandum No. 476, 9, *ECMWF, Reading,UK*.

Funk, C., Senay, G., Asfaw, A., Verdin, J., Rowland, J., Michaelson, J., Eilerts, G., Korecha, D., and Choularton, R., 2005: Recent drought tendencies in Ethiopia and Equatorial-subtropical eastern Africa. *Washington DC, FEWS-NET.*

Giorgi, F. and Francisco, R., 2000: Uncertainties in regional climate change prediction: A regional analysis of ensemble simulations with the HADCM2 coupled AOGCM. *Clim Dyn.,* **16**, 169-182.

Gitau, W., 2011: Diagnosis and predictability of Intra-seasonal characteristics of wet and dry spells over Equatorial East Africa. *PhD Thesis, University of Nairobi.*

Hagedorn, R., Palmer, T. N., Shutts, G. J., Doblas-Reyes, F.J., Jung T., and Leutbecher, M., 2005: Representing model uncertainty in weather and climate prediction. *Annual Review of Earth and Planetary Sciences,* **33**,163-193.

Heikkila, U., Sandvik. A., Sorterberg, A., 2010: Dynamical downscaling of ERA-40 in complex terrain using WRF regional Climate model. *Clim Dyn*. Doi: 10.1007/s00382-010-0928-6.

Hulme, M., Jenkins, G. J., Lu, X., Turnpenny, J. R., Mitchell, T. D., Jones, R. G., Lowe, J.,Murphy, J. M., Hassell, D., Boorman, P., McDonald, R. and Hill, S., 2002: Climate Change Scenarios for the United Kingdom: *The UKCIP02 Scientific Report, Tyndall Centre for Climate Change Research, School of Environmental Sciences, University of East Anglia, Norwich, UK, 120.*

Indeje, M., and Semazzi F., 2000: Relationships between QBO in the lower Equatorial Stratospheric Zonal Winds and East African Seasonal Rainfall. *J. Meteorol. Atmos. Phys.* **23**,227-244.

Indeje, M., F.H.M and Ogallo, L.A., 2000: ENSO signals in East African Rainfall and their predictions potentials.*Int.J.Climatol.**20** 19-46.*

Ininda, J., 2008: Towards Improvement of seasonal rainfall forecasting through model output statistics (MOS) and downscaling of ECHAM Forecasts over Tanzania. *J. Kenya Meteorol. Soc.,* **2** (2), 96-104.

IPCC 2007: Africa Climate Change 2007: Impacts, Adaptation and Vulnerability and Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. *Cambridge University Press,Cambridge UK*, 433-467**.**

IRI 2005: Regional Climate Prediction and Risk Reduction in the Greater Horn of Africa. Final Report to the USAID Office of Foreign Disaster Assistance from the International Research Institute for Climate Prediction (IRI), the Earth Institute at Columbia University June 15, 2005.

ISDR 2008: Linking Disaster Risk Reduction and Poverty Reduction Good Practices and Lessons Learned. *A Publication of the Global Network of NGOs for Disaster Risk Reduction.*

Jin, E. K., Kinter J. L., Wang, B., Park, C. K., Kang, I. S., Kirtman, B. P., Kug, J. S., Kumar, T., S., V., Luo, J. J., Schemm J., Shukla, J., and Yamagata, T., 2008: Current status of ENSO prediction skill in coupled ocean-atmosphere models, *Climate Dyn*, **31**, 647-664.

Jolliffe, I.T. and Stephenson, D.B., 2003: Forecast Verification: A Practitioners guide in Atmospheric Science, *Wiley*, 240.

Jones, R., 2005: Senate Hearing Demonstrates Wide Disagreement About Climate Change. FYI Number 142, *American Institute of Physics*.

Kalnay, E., 2003: Atmospheric Modelling, data assimilation and Predictability. *Cambridge University Press, Cambridge. ISBN 0 521791790.*

Kenya Meteorological Department 2013: Review of Rainfall during the Long Rains and the Outlook for the June-July-August Season 2013. *Ministry of Environment and Mineral Resources, Republic of Kenya.*

Kirtman, B. and Pirani, A., 2009: The State of the Art of Seasonal Prediction Outcomes and Recommendations from the First World Climate Research Program (WCRP) Workshop on Seasonal Prediction, *BAMS*, DOI: 10.1175/2008BAMS2707.1

Krishnamurti, T. N., Kishtawal, C. M., Zhang, Z., LaRow, T. E., Bachiochi, D. R., Williford, C.E., Gadgil, S. and Surendran, S., 2000: Improving tropical precipitation forecasts from a multi analysis super ensemble; *J. Climate*, **13**, 4217- 4227.

Krishnamurti, T.N., Basu, S., Sanjay, J. and Gnanaseelan, C., 2008: Evaluation of several different planetary boundary layer schemes within a single model, a unified model and a multimodel superensemble. *Tellus*, **60A**, 42-61.

Kristler, R., Kalnay, E., Collins, W., Saha, S., White, G., Wollen, J., Chelliah, M., Ebisuzaki, W., and Koussky, V., 2001: The NCEP-NCAR 50year reanalysis: Monthly means CD ROM and documentation.*Bull.Amer.Meteor.Soc*., **82** 247-268.

Landman, W. A. and Goddard, L., 2002: Statistical recalibration of GCM forecasts over southern Africa using model output statistics. *J. Climate*, **15**, 2038-2055.

Latif, M., Sperber K., Arblaster J., Braconnot P., Chen D., Colman A., Cubasch U., Cooper C., Delecluse P., De Witt D., Fairhead L., Flato G., Hogan T., Ji M., Kimoto M., Kitoh A., Knutson T., Le Treut H., Li T., Manabe, S., Marti, O., Mechoso C., Meehl G., Roeckner E., Sirven J., Terray L., Vintzileos, A., Wang, B., Washington, W., Yoshikawa, I., Yu, J. and Zebiak S.,2001; The El Nino simulation inter-comparison project (ENSIP). *Clim Dyn.*, **18,** 255–276.

Lo, J. C.-F., Yang, Z.-L., and Pielke, R. A., 2008: Assessment of three dynamical climate downscaling methods using the weather research and forecasting (WRF) model. *J. Geophys. Res*., **113**, D09 112, doi: 10.1029/2007JD009216.

Marengo, J.M., Cavalcanti, I. F., Satyamurty, P., Trosniko,v I., Nobre, C. A., Bonatti, J. P. , Camarg, H., Sampaio, G., Sanches, M. B., Manzi, A. O., Castro, C.A.C., D"Almeida, C., Mechoso C.R**.**, Robertson, A.W., Barth N., Davey, M.K., Delecluse, P., Gent, P.R., Ineson, S., Kirtman, B., Latif M., Le Treut, H., Nagai, T., Neelin, J.D., Philander, S.G.H., Polcher, J., Schopf, P.S., Stockdale, T., Suarez, M.J., Terray, L., 2003:Assessment of regional seasonal rainfall predictability using the CPTEC/COLA atmospheric GCM. *Climate Dyn*.,**21**, 459–475.

Marengo, J. A., Soares W. R., Saulo, C., and Nicolini, M., 2004: Climatology of the low level jet east of the Andes as derived from the NCEP-NCAR reanalyses: Characteristics and temporal variability. *J. Climate*, **17**, 2261-2280.

Min, S.-K., and Hense, A., 2006a: A Bayesian approach to climate model evaluation and multi-model averaging with an application to global mean surface temperatures from IPCC AR4 coupled climate models. *Geophys. Res. Lett*., **33**, L08708, doi: 101029/2006GL025779.

Muhati, F., Opijah, F. J. and Ininda. J., 2007: Relationship between ENSO Parameters and the Trends and Periodic Fluctuations in East African Rainfall. *J. Kenya Meteorol. Soc*.**1** (1), 20-43.

Murphy, J. M., Sexton D. M. H., Barnett D. N., Jones G. S., Webb M. J., Collins M. and

Stainforth, D. A., 2004: Quantification of modeling uncertainties in a large ensemble of climate change simulations; *Nature*, **430**, 768–772.

Mutai, C. C. and Neil, M., W., 2000: East African rainfall and the tropical circulation convection on intra-seasonal to inter-annual time scales. *J. Climate,* **3**, 3915–3939.

Mutemi, J. N., 2003: Climate anomalies over East Africa associated with various ENSO evolutions Phases. PhD Thesis,*University of Nairobi*.

Nicholson, E.S., 2013: A Review of Recent Studies on the Rainfall Regime and Its Inter-annual Variability. *Earth, Ocean, and Atmospheric Sciences Department, Florida State University, Tallahassee, FL 32306.* doi.org/10.1155/2013/453521.

Nyakwada, W., Ogallo, L., and Okoola, R., 2009**:** The Atlantic-Indian Ocean Dipole and Its Influence on East African Seasonal Rainfall, *J.Meteorol. Rel. Sci*., **3,** 21–35.

Ogallo, L. A., 1989: The spatial and temporal patterns of East African- rainfall derived from principal component analysis; *Int. J. Climatol.,* **9,** 145–167.

Ogallo, L. A., Pierre, B., Jean. Simon, M., and Stephen J. C., 2008: Adapting to climate variability and change: *The Climate Outlook Forum process WMO bulletin* **57** (2).

Okoola, R. E., Ininda, J.M., and Camberlin, P., 2009: Wet periods along the East Africa Coast and the extreme wet spell event of October 1997. *J.Meteorol. Rel. Sci***.,** ***2(1)*** *65–81*.

Okoola, R. E.,1998. Characteristics of the Inter-Tropical Convergence Zone over Equatorial Eastern Africa based on station rainfall records. *J. Meteorol. Rel. Sci.*, **3**, 61-101.

Okoola, R., Camberlin, P., and Ininda, J., M. 2008: Wet periods along the East Africa Coast and the extreme wet spell event of October 1997**.** *J.Meteorol. Rel. Sci***.,** **2**(1) 65–81.

Omeny, P., Ogallo, L. and Okoola, R., 2008: East Africa rainfall variability associated with the MJO, *J.Meteorol. Rel. Sci*., **2**, 105–114.

Omondi, A., P., 2010: The teleconnections between decadal rainfall variability and global sea surface temperatures and simulation of future climate scenarios over East Africa. *PhD thesis University of Nairobi.*

Omondi, P., Ogallo L. and Okoola, R., 2009: Decadal Rainfall Variability Modes in observed Rainfall Records over East Africa and their Predictability using Sea Surface Temperature **.***J.Meteorol. Rel. Sci.,* **3**, 37–54.

Owiti, Z. and Weijun, Z., 2012: East African rainfall seasonality based on modulated annual cycle (MAC) *Int. Journal of Phy. Sc.*, **7**(24), 3050-3061.

Owiti, Z., Ogallo, L. and Mutemi, J., 2008: Linkages between the Indian Ocean Dipole and East African Seasonal Rainfall Anomalies, *J.Kenya Meteorol. Soc.*, **2**(1), 3 –17.

Palmer T.N., Brankovic C. and Richardson, D.S., 2000, A probability and decision-model analysis of PROVOST seasonal multi-model ensemble integrations, *Q. J. R. Meteorol Soc.*, **126** (567), 2013-2033.

Palmer, T. N., Alessandri, A., Andersen, U., Cantelaube, P., Davey, M., Alessandri, A.**,** Gualdi, S., Andersen, U., Feddersen, H., Cantelaube, P., Terres, J-M., Davey, M., Graham, R., Délécluse, P., Lazar, A., Déqué, M.,. Guérémy, J-F., Díez, E., Orfila, B., Hoshen, M., Morse, A. P., Keenlyside, N., Latif, M., Maisonnave, E., Rogel, P., Marletto, V. and Thomson, M. C., 2004: Development of a European multi-model ensemble system for seasonal to interannual prediction (DEMETER). *Bull. Amer. Meteor. Soc.*, **85**, 853-872.

Plisnier, P.D., Serneels, S. and Lambin E. F., 2000: Impact of ENSO on East African ecosystems: A multivariate analysis based on climate and remote sensing data. *Global Ecology and Biogeography* **9**, 481-497.

Robock, A., 2000: Volcanic Eruptions And Climate, *Reviews of Geophysics*, **20**, 191-219.

Saha, S., Wu, X., Wang, J., Nadiga, S., P., Behringer, D., Yu-Tai, H.,Hui-ya, C., Pena-Mendez, M., Qin, Z., Chen, M., and Becker, E., 2012: The NCEP Climate Forecasts System Version 2 . *J. Climate.*

Shongwe, M. E., Geert, J. O., Bart, H., Maarten, A., 2011: Projected Changes in Mean and Extreme Precipitation in Africa under Global Warming. Part II: East Africa. *J. Climate*, **24**, 3718–3733. doi: http://dx.doi.org/10.1175/2010JCLI2883.1

Stenchikov, G., Hamilton, K., Stouffer, R. J., Robock, A., Ramaswamy, V., Santer, B., and Graf, H.-F., 2006: Arctic Oscillation response to volcanic eruptions in the IPCC AR4 climate models. *J.Geophys Res.*, **111**, doi: 10.1029/2005JD006286.

Taylor, K. E., 2001: Summarizing multiple aspects of model performance in a single diagram, *J. Geophys. Res.*, **106***(*D7*)*, 7183–7192, *doi:*10.1029/2000JD900719.

Timothy, N. S., Oscar, A., Boer, G., Michel, D., Yaqui, D., Arun, K., Krishna,K., Willem, L., Simon, M., Paulo, N., Adam, S., Ose, T., and Yun, W.,T., 2009: Understanding and Predicting Seasonal to Interannual Climate Variability - the producer perspective. *www.**wmo**.int/wcc3/sessionsdb/documents/WS3_WP_capability.pdf.*

UNISDR, 2012: Disaster Risk Reduction in Africa. *Africa Informs Special Issue on Drought 2012.*

Van, L. H., Gerald, A. M., Arblaster M. J., and Branstator, G., 2008: A Coupled Air–Sea Response Mechanism to Solar Forcing in the Pacific Region. *J. Climate*, **21**, 2883–2897.

Wang, B., Kang I. S., and Lee J. Y., 2004: Ensemble simulations of Asian-Australian monsoon variability by 11 GCMs. *J. Climate*, **17,** 803–818.

Wang, B., Lee, J.-Y., Kang, I.-S., Shukla, J., Park, C.-K., Kumar, A., Schemm, J., Cocke, S., Kug, J.S., Luo, J.-J., Zhou T., Wang B. X. Fu., Yun W.-T., Alves O., Jin E. K., Kinter J., Kirtman, B., Krishnamurti, T., Lau, N.C., Lau, W., Liu P., Pegion, P., Rosati, T., Chubert, S., Stern, W., Suarez, M., Yamagata, T., 2008: Advanced and prospectus of seasonal prediction: Assessment of the APCC/CliPAS 14-model ensemble retrospective seasonal prediction (1980-2004), *Climate Dyn*., **33**, 93-117.

Wilks, D., 2006: Statistical Methods in the Atmospheric Sciences, *Academic Press*, 592.

Williams, C. A., and Hanan, N. P., 2011: ENSO and IOD teleconnection for African ecosystems; evidence of destructive interference between climate oscillations. *Biogeosciences. www.biogeosciences.net/8/27/2011/doi:10.5194/bg-8-27-2011*.

WMO, 2007: Social and economic Benefits of weather, Madrid conference statement and actionplan.*http://www.wmo.int/pages/themes/wmoprod/documents/madrid07_ActionPlan_wE.*

Yang, W., Seager, R., and Cane, M. A., 2013: The East African Long Rains in Observations and Models. *Columbia University,Palisades, New York. USA International Research Institute for Climate and Society*.