FORECASTING SEASONAL STREAM FLOW IN ATHI RIVER BASIN USING GLOBAL AND REGIONAL CLIMATE PREDICTORS

CLIFFORD CLEMENT NYABERI OBIERO

A thesis submitted in partial fulfillment of the requirement for the degree of Doctor of Philosophy in the Department of Water and Environmental Engineering of Kenyatta University

November 2011

Oberi, Clifford
Forecasting seasonal stream flow in Athi

KENYATTA UNIVERSITY LIBRARY
DECLARATION

This thesis is my original work and has not been presented for a degree in any other university or any other award.

Clifford C. N. Obiero

Signature [signature] Date 24/11/2011

We confirm that the work reported in this thesis was carried out by the candidate under our supervision.

Dr. Simon M. Maingi
Department of Water and Environmental Engineering
Kenyatta University

Signature [signature] Date 24/11/2011

Prof. Laban J. Ogallo
IGAD Climate prediction and applications Centre (ICPAC), Nairobi

Signature [signature] Date 24/11/2011

Dr. Christopher Ondieki
Department of Geography
Kenyatta University

Signature [signature] Date 24/11/2011
DEDICATION

To my mother, late Sarah Kemunto
ACKNOWLEDGEMENT

I wish to extend my heartfelt thanks to my supervisors; Prof. L. J. Ogallo, Dr. S. Maingi and Dr. C. Ondieki for their immense guidance, advice, criticism, patience and devotion during this study. I also thank the government of the Federal Republic of Germany, through the Nairobi Daad office for sponsoring this study.

I wish as well to register my sincere thanks to all those who assisted me with softwares, data analysis and research papers. I wish to single out the following for their immense help; Messers Owiti and Omondi of ICPAC for their assistance in CCA and Mr. Gitau for assisting me with wavelet analysis script. Many thanks also go to Prof. J. Dracup (University of California, Los Angeles), Dr. A. Giannini (Columbia University) and Dr. F. Chiew (University of Melbourne) for their assistance.

Lastly, my sincere thanks go to my wife Lucy and my children, Moraa, Betty, Ray and Kemunto for their encouragement, support and understanding during this study.
TABLE OF CONTENT

Declaration ......................................................................................................................... ii
Dedication ........................................................................................................................... iii
Acknowledgement ............................................................................................................ iv
Abstract............................................................................................................................ xiv
List of Tables .................................................................................................................... ix
List of Figures .................................................................................................................. x

CHAPTER ONE .................................................................................................................. 1
1.0 INTRODUCTION ......................................................................................................... 1
1.1 Background of the study ............................................................................................... 1
1.2 Study area .................................................................................................................... 5
1.3 Problem statement ....................................................................................................... 7
1.4 General Objective ....................................................................................................... 8
  1.4.1 The specific objectives......................................................................................... 8
1.5 Justification for the study ........................................................................................... 9

CHAPTER TWO .................................................................................................................. 11
2.0 Literature Review ....................................................................................................... 11
2.1 Seasonal rainfall and stream flow variability in East Africa ........................................... 11
2.2 Design of rainfall and stream flow networks ............................................................... 15
2.3 Rainfall and stream flow variability during ENSO ...................................................... 17
2.4 Rainfall and stream flow variability during the Indian Ocean Dipole ......................... 22

CHAPTER THREE .............................................................................................................. 24
3.0 DATA AND METHODS ................................................................................................................. 24
3.1 Introduction ............................................................................................................................... 24
3.2 Estimation of missing data and data quality control ................................................................. 24
  3.2.1 Estimation of missing data methods ..................................................................................... 25
  3.2.2 The Linear Correlation method ......................................................................................... 25
3.3 Data quality of rainfall and stream flow data ........................................................................... 26
  3.3.1 Mass and double mass curve methods ................................................................................. 27
  3.3.2 Runs Test Method ............................................................................................................... 27
3.4 Design of a minimum rainfall and stream flow network ........................................................... 28
  3.4.1 Principal Component Analysis ............................................................................................ 29
    3.4.1.1 Number of Significant Principal Components ............................................................... 31
    3.4.1.2 Rotation of Principal Components ............................................................................... 31
    3.4.1.3 Significant factor loadings .......................................................................................... 32
    3.4.1.4 Communality analysis .................................................................................................. 33
3.5 Temporal characteristics of rainfall and stream flow ............................................................... 34
  3.5.1 Trend analysis ....................................................................................................................... 34
  3.5.2 Periodicity analysis .............................................................................................................. 35
    3.5.2.1 Wavelet analysis ......................................................................................................... 36
  3.5.3 PCA (T-mode) analyses ....................................................................................................... 39
3.6 Variability of rainfall and stream flow during ENSO and IOD .................................................. 39
  3.6.1 Composite analysis .............................................................................................................. 41
  3.6.2 Simple correlation method .................................................................................................. 43
  3.6.3 Canonical Correlation Analysis (CCA) .............................................................................. 43
    3.6.3.1 The Canonical Correlation Analysis Method ................................................................. 45
3.7 Rainfall and stream flow forecasting ......................................................................................... 47
3.7.1 Regression analysis method ................................................................. 47
3.7.2 Non-Parametric Seasonal Forecasting Model (NSFM) .......................... 48
  3.7.2.1 Parameter estimation of the model ............................................... 49
  3.7.2.2 Calibration and verification of the model ...................................... 50
  3.7.2.3 Nash–Sutcliffe coefficient of efficiency (E) .................................. 50
  3.7.2.4 The modified linear error in probability space (LEPS) score ........... 51

CHAPTER FOUR ........................................................................................................ 52

4.0 RESULTS AND DISCUSSION ....................................................................... 52

4.1 Introduction .................................................................................................. 52

4.2 Estimation of missing data .......................................................................... 52

4.3 Quality Control test Results ........................................................................ 54

4.4 Design of optimum rainfall and stream flow networks results .................. 57
  4.4.1 Results for spatial variability of MAM season rainfall using PCA ....... 57
  4.4.2 Spatial variability of SON season rainfall using PCA ......................... 62
  4.4.3 Stability analysis of spatial PC rainfall patterns ................................... 66
    4.4.3.1 Interstation rainfall correlations ............................................... 66
    4.4.3.2 Rainfall data split analysis ....................................................... 72
  4.4.4 Results for Spatial variability of MAM season stream flow ............... 78
  4.4.5 Spatial variability of SON season stream flow ................................... 81
  4.4.6 Stability analysis of PC stream flow patterns ...................................... 84
    4.4.6.1 Interstation stream flow correlation ........................................... 84
    4.4.6.2 Stream flow data split analysis ................................................ 84

4.5 Temporal characteristics of seasonal rainfall and stream flow ................. 90
  4.5.1 Cyclical analysis results derived from wavelet analyses .................... 93
  4.5.2 Results for temporal variability analysis of rainfall and stream flow .... 96

4.6 Results for rainfall and stream flow variability and teleconnections ......... 101
4.6.1 Results for spatial variability of rainfall and stream flow during El Niño ................................................................. 102
4.6.2 Results for variability of rainfall and stream flow during La Niña.....108
4.6.3 Evolution of ENSO during SOND season................................................................. 113
4.6.4 ENSO and rainfall /stream flow teleconnection ............................................... 121
4.7 Results for rainfall and stream flow spatial variability during IOD ..........122
  4.7.1 Results for rainfall and stream flow spatial analysis during positive IOD
  ............................................................................................................................... 122
  4.7.2 Results for rainfall and stream flow spatial analysis during negative IOD
  ............................................................................................................................... 126
  4.7.3 Results for the study of evolution of IOD during MAM and OND......129
4.8 Results for Canonical Correlation Analysis............................................................ 135
  4.8.1 Results for CCA models for MAM rainfall and stream flow ............135
  4.8.2 Results for CCA models for OND rainfall and stream flow ...............139
  4.9.1 Results for seasonal rainfall and stream flow regression analysis ....142
  4.9.2 Results for Non-parametric Seasonal Forecasting Model (NSFM) ......146
CHAPTER FIVE ............................................................................................................. 155
5.0 SUMMARY, CONCLUSIONS AND RECOMMENDATIONS ......................155
5.1 Introduction.............................................................................................. 155
5.2 Summary .............................................................................................. 155
5.3 Conclusion ............................................................................................ 159
5.4 Recommendations of the study................................................................. 160
6.0 REFERENCES .......................................................................................... 163
List of Tables

Table 1. Classification of the wet/dry rainfall and stream flow .........................42
Table 2. Rainfall stations used in the study ......................................................53
Table 3. River flow gauging stations used in the study ......................................54
Table 4a. Rotated principal components modes for MAM rainfall .......................58
Table 4b. Rotated principal components modes for SON rainfall .........................62
Table 5. The minimum rainfall network station for the Athi river basin .................78
Table 6. Minimum Stream flow network for Athi river basin ..............................89
Table 7. Spearman rank correlation coefficients for rainfall .............................93
Table 8. Results for T-mode PCA clusters of wet and dry years .......................96
Table 9. Results for T-mode PCA clusters of High and Low flow Years ...............97
Table 10. Regression models for Rainfall and Stream flow forecasting ...............142
Table 11. Results for forecasting SON rainfall using NSFM ..........................147
Table 12. Results for SON stream flow forecasting using NSFM ......................149
Table 13. Examples of forecast categorization (stream flow=2.5 m$^3$/s and 
          IOD=1.2) ..................................................................................................150
List of Figures

Figure 1. The Map of the athi river basin ................................................................. 6
Figure 2a. Distribution of the rainfall stations used in the study .............................. 55
Figure 2b. Distribution of the river gauging stations used in the study ....................... 55
Figure 3a. Mass curve for river gauging station at Kiu (3bb13)-heterogenous .......... 56
Figure 3b. Mass curve for rainfall station at Mombasa-homogenous ......................... 56
Figure 4a. Spatial patterns of the rotated pc mode 1 during mam rainfall ................. 59
Figure 4b. Spatial patterns of the rotated pc mode 2 during mam rainfall ................. 60
Figure 4c. Spatial patterns of the rotated third pc mode during mam rainfall .......... 61
Figure 5a. Spatial patterns of the rotated first pc mode for son rainfall ..................... 63
Figure 5b. Spatial patterns of the rotated PC mode 2 for SON rainfall ..................... 64
Figure 5c. Spatial patterns of the rotated PC mode 3 for SON rainfall ..................... 65
Figure 6a. Spatial patterns of the inter station correlations with reference to
Dagoretti (PC mode 1) station in the MAM rainfall .................................................. 67
Figure 6b. Spatial patterns of the inter station correlations with reference to
Rukanga (PC mode 3) station in the MAM rainfall .................................................. 68
Figure 7a. Spatial patterns of the inter station correlations with reference to Nong station (PC mode 1) in SON rainfall ................................................................. 69
Figure 7b. Spatial patterns of the inter station correlations with reference to Kwale station (PC mode 2) in SON rainfall ................................................................. 70
Figure 7c. Spatial patterns of the inter station correlations with reference to
Kitondo (PC mode 3) stations in the SON rainfall .................................................... 71
Figure 8a. Spatial patterns of the first rotated PC mode for even years in the MAM rainfall (1961-2004) ................................................................. 72
Figure 8b. Spatial patterns of the first rotated PC mode for odd years in the MAM rainfall (1961-2004) ................................................................. 73
Figure 9a. Spatial patterns of the first rotated PC mode for even years in the SON rainfall (1961-2004) ................................................................. 74
Figure 9b. Spatial patterns of the first rotated PC mode for odd years in the SON rainfall (1961-2004) ................................................................. 75
Figure 10. Homogenous divisions of stations derived from the seasonal rainfall RPCA patterns ........................................................................... 77
Figure 11a. The spatial patterns of first rotated PC mode derived MAM stream flow ..................................................................................... 79
Figure 11b. The spatial patterns of second rotated PC mode derived MAM stream flow .............................................................................. 80
Figure 12a. The spatial patterns of rotated first PC mode derived from SON stream flow ............................................................................... 82
Figure 12b. The spatial patterns of the rotated second PC mode derived from SON stream flow ................................................................. 83
Figure 13a. Spatial patterns of the inter station correlations with reference to station 3BC12 (MAM-PC mode 1) ......................................................... 85
Figure 13b. Spatial patterns of the inter station correlations with reference to station 3G03 (MAM-PC mode 2) ......................................................... 86
Figure 14. Stream flow homogenous zones derived from PC patterns .............................................................................................................. 88
Figure 15. Example of a rainfall time series: Wilson airport station April rainfall ......... 91
Figure 16. Streamflow hydrograph for Theta river at 3BD02 .................................................. 92
Figure 17. The wavelet spectrum for Mombasa November rainfall and Mean OND rainfall................................................................................................................................. 94
Figure 18. The wavelet spectrum for 3MH01 RGS Mean December Discharge and OND season discharge ................................................................. 95
Figure 19. Rainfall anomalies during the dry cluster of MAM 1970 ......................... 99
Figure 20. Rainfall anomalies during the wet cluster of MAM 1997 ......................... 99
Figure 21. Stream flow anomalies during the low flow cluster of SON 1984 .......... 100
Figure 22. Stream flow anomalies during the high flow cluster of SON 1981 .......... 100
Figure 23a. Spatial patterns for PC mode 1 (El Niño rainfall composite) .............. 103
Figure 23b. Spatial patterns for PC mode 2 (El Niño rainfall composite) .............. 104
Figure 24a. Spatial patterns for PC mode 1 (El Niño stream flow) .......................... 106
Figure 24b. Spatial patterns for PC mode 2 (El Niño stream flow composites) ....... 107
Figure 25a. Spatial patterns of the first PC mode during La Niña rainfall .......... 110
Figure 25b. Spatial patterns of PC mode 2 during La Niña rainfall .................. 111
Figure 25c. Spatial patterns of PC mode 3 during La Niña rainfall .................. 111
Figure 26a. Spatial patterns PC mode 1 during La Niña rainfall ...................... 112
Figure 26b. Spatial patterns PC mode 2 during La Niña stream flow ............... 112
Figure 26c. Spatial patterns PC mode 3 during La Niña stream flow ............... 113
Figure 27. The composite spatial patterns of JJA rainfall during year (0) of (a) El Niño (b) La Niña, X-axis is Longitude and Y-axis Latitude ......................... 115
Figure 28. The composite spatial patterns for SON rainfall during year (0) of (a) El Niño (b) La Niña ................................................................. 116
Figure 29. The composite spatial patterns for OND rainfall during year (0) of (a) El Niño (b) La Niña ................................................................. 117
Figure 30. The composite spatial patterns for DJF rainfall during year (+) of (a) El Niño (b) La Niña ................................................................. 118
Figure 31. The composite spatial patterns for OND stream flow during year (+) of (a) El Niño (b) La Niña ................................................................. 120
Figure 32. Spatial patterns of PC mode 1 during rainfall Positive IOD ............. 124
Figure 33. Spatial patterns of PC mode 2 during rainfall Positive IOD ............. 124
Figure 34. Spatial patterns of PC mode 1 during Stream flow Positive IOD ....... 125
Figure 35. Spatial patterns of PC mode 2 during Stream flow Positive IOD ....... 125
Figure 36. Spatial patterns of PC mode 1 during rainfall negative IOD .......... 127
Figure 37. Spatial patterns of PC mode 2 during rainfall negative IOD .......... 128
Figure 38. Spatial patterns of MAM (-) rainfall composite during (a) Positive IOD (b) Negative IOD, X axis shows latitudes and Y-axis Longitudes ............ 131
Figure 39. Spatial patterns of SON (0) rainfall composite during (a) Positive IOD (b) Negative IOD ................................................................................ 132
Figure 40. Spatial patterns of OND (0) rainfall composite during (a) Positive IOD (b) Negative IOD .......................................................... 133

Figure 41. Spatial patterns of OND (0) Stream flow composite during (a) Positive IOD (b) Negative IOD .......................................................... 134

Figure 42. The spatial patterns of CCA mode 1 for the correlation between (a) seasonal DJF Indian SST and MAM rainfall (b) rainfall and canonical vector (v) and (c) normalised temporal functions for CCA mode 1 patterns for rainfall and SST .......................................................... 137

Figure 43. The spatial patterns of CCA mode 1 for the correlation between (a) seasonal DJF Indian SST and MAM stream flow (b) stream flow and canonical vector (v) and (c) normalised temporal functions for CCA mode 1 patterns for stream flow and SST .......................................................... 138

Figure 44. The spatial patterns of CCA mode 1 for the correlation between (a) seasonal JJA Indian SST and OND rainfall (b) rainfall and canonical vector (v) and (c) normalised temporal functions for CCA mode 1 patterns for rainfall and SST .......................................................... 140

Figure 45. The spatial patterns of CCA mode 1 for the correlation between (a) seasonal JJA Indian SST and OND stream flow (b) stream flow and canonical vector (v) and (c) normalised temporal functions for CCA mode 1 patterns for stream flow and SST .......................................................... 141

Figure 46. Region VI rainfall forecast at Mombasa .......................................................... 144

Figure 47. Region IV rainfall forecast at Kibwezi .......................................................... 145

Figure 48. Region I stream flow forecast at 3BC15 ......................................................... 146

Figure 49. Stream flow Region I SON stream flow forecast (Above Normal) ............... 151

Figure 50. Stream flow Region I SON stream flow forecast (Normal flow) ............... 152

Figure 51. Stream flow Region I SON stream flow forecast (Below Normal flow) ....... 153
Abstract

Extreme weather events are associated with floods and drought which affect infrastructure, food security, water availability, sanitation and hydro power generation. The Athi river basin is prone to such extreme weather events. This study examines the spatial and temporal hydrological characteristics of the Athi river basin, their teleconnections with El Niño/Southern Oscillation (ENSO) and the Indian Ocean Dipole (IOD) and assesses the potential use of derived teleconnections for stream flow forecasting. The Principal Component analysis (PCA) method was used to design the minimum rainfall and stream flow networks for the basin. Spatial and temporal PCA modes were used to study the spatial characteristics of seasonal rainfall and stream flow. Trends were studied using graphical and Spearman rank correlation methods while periodicity was investigated using wavelet analysis. PCA and composite analysis were used to map linkages among extreme seasonal rainfall/stream flow patterns and ENSO and IOD evolution phases for the ‘long’ and ‘short’ rainy seasons. Simple linear correlation and Canonical Correlation Analysis (CCA) were used to delineate inter-linkages among rainfall, stream flow, ENSO and IOD. The potential of deriving seasonal rainfall and stream flow forecasts was investigated using the step-wise regression and the non-parametric seasonal forecasting (NSFM) models. PCA delineated the Athi river basin into six rainfall and three streamflow homogenous zones. The spatial characteristics of seasonal rainfall and streamflow showed a strong influence of the Inter Tropical Convergence Zone (ITCZ), land sea mesoscale circulation system, orography and land use systems. The PCA based areal rainfall and streamflow indices and communality analysis were used to determine the best representative stations for the homogenous zones. PCA (T-mode) delineated most of the wet/high flow and dry/low flow years observed from historical records including 1997/1984 which are some of the wettest/driest years on record in the basin. Although, the results showed significant rainfall and stream flow trends at some locations, it was difficult to associate the observed trends to climate change, due to limited data length. Wavelet analyses showed significant peaks centered at 2-3, 5-7 and 10-12 years which may be associated with Quasi-Biennial Oscillation, ENSO, solar cycles and decadal variability modes. Linkages between various modes of ENSO were also delineated including “El Niño Modoki” which is a manifestation of El Niño associated with dry instead of the normal wet conditions in the basin. Strong linkages of seasonal rainfall and stream flow with ENSO and IOD phases with time lags of 7-9 months and lag correlations of 0.8 - 0.9 were obtained. The composite results indicated that the wet/high flow and dry/low flow conditions in the basin could be associated with the evolution of ENSO and IOD phases. CCA method delineated the major Indian Ocean Sea Surface Temperature (SST) modes, which are associated with extreme March-May and September-November seasonal rainfall and streamflow. Regression and NSFM models showed good predictions skills for low and high flows using ENSO and IOD predictors. The results of this study provide tools for prediction, early warning of the extremes episodes, management, planning and operation of water resources systems.
CHAPTER ONE

1.0 INTRODUCTION

In this Chapter the background of the study, study area, problem statement, objectives and the justification of study are given.

1.1 Background of the study

Water is essential to life on this planet. Inadequate or excessive water is a major natural threat to life. Water availability from rivers, lakes, underground aquifers and roof catchments are directly or indirectly linked to variability and change in climate elements especially precipitation. Abundant precipitation can lead to too much water and floods that often result into disasters such as flooding, water pollution, soil erosion, siltation, dam breaks, water related disease outbreaks and loss of crops and lives. Failure of precipitation leads to too little water and droughts which often lead to water scarcity, loss of vegetation, loss of livestock and wildlife, famine and general suffering of people. In the affected areas, land misuse, destruction of ecosystems, poor sanitation, reduced forest cover, desertification and other environmental degradation activities occur.

Kenya’s economy largely depends on agriculture, fisheries, tourism and agro-based industries. Agriculture contributes about 26% of the country’s Gross Domestic Product (GDP) as per the year 2003 and employs over 70% of the population (Library of congress, 2005). The dependence of agriculture on
rainfall makes agricultural production output highly variable and unpredictable. This is because rainfall is unreliable and is marked by complex space-time variability in the tropics (Ogallo, 1983).

Water resources in Kenya are unevenly distributed in space and time, a common phenomenon in most tropical countries. Her annual internal renewable water resources is less than 700 m$^3$ per capita per annum which is a chronic state of water shortage (GOK, 1999). While catchment precipitation determines water availability, water demand in any basin depends on specific water use activities like domestic use, agriculture, livestock, industrial and other socio-economic needs.

Unavailability of water undermines government and society efforts to improve the quality of life in a country. The Government of Kenya (GOK) has acknowledged the strong correlation between availability of water and the level of socio-economic development of her people (GOK, 1999). Thus, the government is focusing on provision of water for domestic use, agriculture, livestock development and industrial use in order to improve the social well being of the population and enhance economic development.

Kenya like many tropical countries is prone to extreme weather conditions, which include drought and floods. Major recent large scale drought events in Kenya occurred in 1984, 1991-93, 1999/2000, 2008/2009(GOK, 1999 and Prof., Ogallo, ‘Personal communication’). The major impacts of drought in
Kenya include lack of water, food, famine, hydropower rationing, health problems, mass migration of people and animals, loss of life and property among many other socio-economic problems. The occurrences of frequent and often prolonged droughts have greatly threatened the sustainable use of the natural resources and the very livelihood of Kenyans. In Kenya, the GDP growth rate reduced from 1.4% in 1999 to 0.7% in 2000 due to the 1999/2000 drought (GOK, 1999). The number of people threatened by starvation reached about 4.7 million that is, 16% of the country’s population.

While Kenya experiences perennial drought, the country also experience extreme floods in certain years. An example of this is the 1997-1998 year El Niño event that produced above normal rainfall in East Africa that lead to widespread flooding and diseases outbreak (Linthicum, et al., 1999). During 1997/98 year Masinga hydropower dam rose to 1057 m above sea level, the highest since it was built, the pasture in the arid and semi-arid lands, improved tremendously. However, the rains caused floods, soil erosion, landslides, damaged infrastructure and crops. Other recent major flood events in Kenya occurred in 1961/62, 1983 and 2006(GOK, 1999).

The cause of drought and floods events in Kenya are not only due to local and regional factors but also due to teleconnections with large scale systems associated with the fluctuations in the global general circulation system. One of these teleconnections phenomena is the El Niño Southern Oscillations (ENSO). ENSO events are often classified as warm (El Niño) and cold (La Niña).
Globally ENSO has also been associated with both floods and droughts in many regions (Philander, 1990; Glantz, et al., 1991 and Zubair, 2001).

In East Africa, ENSO response vary from one location to another and in many cases the implications for sub-regional climate patterns are lacking (Ogallo, 1983, 1988, 1989; Phillips and McIntyre, 2000; Indeje, et al., 2000). Other major teleconnection systems that are closely linked to ENSO include the Indian Ocean Dipole (IOD) and the Tropical Atlantic Ocean dipole (Saji, et al., 1999; Tourre, et al., 1999; Clark, et al., 2003; Rao, et al., 2002; Chang, et al., 1997; Mehta, 1997; Owiti, 2005 and Vaidya, 2005).

The complex spatial and temporal variability of rainfall in East Africa is due to topography, large inland water lakes, land use patterns, proximity to Indian Ocean among other factors. These factors make rainfall a difficult hydrological component of the river drainage basin in the region to study (Ogallo, 1983). While stream flow is due to rainfall input in a basin, its variability largely depends on the basin and channel characteristics. Minimal flows are observed during the dry seasons and are more pronounced in arid and semi-arid areas (Ondieki, 1993). Within a river basin, hydrologic processes are integrated into stream flow, which provides a natural filter for meteorological noise in precipitation (Piechota, et al., 1997).

The negative impacts associated with extreme weather events on water resources development and management can be reduced through good
understanding of the climate patterns of historical events like those associated with ENSO/IOD and their linkages with regional hydrological cycle. This type of study requires rainfall and stream flow data over a long period. In many parts of Kenya these data is missing. This is due to the high cost of maintaining rainfall and stream flow networks, use of unskilled personnel, vandalism, negligence and poor infrastructure.

Forecasting of rainfall and stream flow in advance is extremely important for the efficient management, planning and operation of water resources systems. It is now possible to provide some outlook of stream flow and rainfall events using ENSO and Indian Ocean Dipole indices for better management of water resources (Dettinger, et al., 2000; Chiew and McMahon, 2003; Owiti, 2005, Vaidya, 2005). This forms the foundation of this study.

1.2 Study area

The Athi river basin lies approximately within the latitude 0.8°S-5°S and longitude 36°E - 40°E and it covers approximately 68,837 km² (Figure 1). The basin is divided into 13 sub-basins, which comprise of the southern parts of the country east of the Rift Valley.

The land in the basin rises steadily from the coastal plain at sea level to Ngong hills on the southern slopes of Aberdare range. The major physical features in the basin include; Shimba hills, Nyika plateau, Kilungu hills, Mbooni hills,
Taita hills, Kilimambogo and Mt Kilimanjaro on the south-west basin boundary. At the upper part of the basin the land rises westwards from the foot of Oldonyo Sabuk towards Ngong hills.

Figure 1. The map of the Athi river basin

The basin is drained by the Athi river that rises from Ngong hills and drains eastwards before changing its course near Oldonyo Sabuk to drain southwards. Some of the major tributaries of the Athi river in the upper part of the basin include; Embakasi, Stony Athi, Nairobi, Mbagathi, Kiserian, Kamiti and Ruiru rivers. At the lower reaches it is joined by the Tsavo, which drains the eastern
slopes of Mt Kilimanjaro and Taita hills. The Tsavo and Athi rivers combine to form the Sabaki (Galana) which discharge its waters into the Indian Ocean.

The basin lies in the arid and semi-arid part of Kenya. Its climate is influenced by topography, ITCZ and proximity to the Indian Ocean among other factors (Asnani, 1993; Ininda, 1994; Ogallo, 1993). A relatively wet narrow tropical belt lies along the Indian Ocean coast. The area behind the coastal strip comprises of semi-arid and arid lands. The basin generally receives two rainy seasons, the long rains in March to May and the short rains between October and December (Asnani, 1993; Ininda, 1994). The mean annual rainfall in the area is about 550mm.

1.3 Problem statement
To design hydraulic structures, delineated flood plains and evaluate economic analysis of floods and drought, representative, continuous and evenly spread data is required. Such data is lacking in the Athi river basin.

Kenya is prone to extreme weather events which are unpredictable. These extreme events include floods and drought which damage infrastructure, hydro power and water rationing, loss of lives and affect food security. The extent of manifestation of these extreme events in the Athi river basin is lacking.

The impact of climate variability and change is a threat to human livelihood. Climate change affects global and regional weather systems shifting the onset
and cessation of rainfall seasons (Indeje, et al., 2000). The manifestation of ENSO and IOD phases and their variability due to climate change in the Athi river basin need to be investigated.

Water resources in Kenya are unevenly distributed in space and time, a common phenomenon in many tropical countries. The spatial and temporal variability of water resources in the Athi river basin need to be explored given its economic importance.

1.4 General Objective

The overall objective of this study is to examine the hydrological characteristics of the Athi river basin and their teleconnections with El Niño Southern Oscillation (ENSO) and Indian Ocean Dipole (IOD) events. The potential use of ENSO and IOD based predictors for obtaining rainfall and stream flow forecasts was also assessed. The specific objectives that were used to achieve the overall objectives of the study are highlighted in the next section.

1.4.1 The specific objectives

The specific objectives that were used to address the overall objective of the study include, to:

(i) Design a minimum rainfall and stream flow networks that can be used to monitor, predict and give early warning of extreme rainfall and stream flow regimes.
(ii) Establish the general linkages between the spatial and temporal characteristics of seasonal rainfall and stream flow.

(iii) Examine the linkages between the spatial characteristics of seasonal rainfall and stream flow in the basin during warm ENSO phase (El Niño) event, cold ENSO phase (La Niña) event and Indian Ocean SSTs represented by India Ocean Dipole (IOD).

(iv) Determine the potential use of ENSO and IOD indicators in obtaining long range rainfall and stream flow forecasts in of the Athi river basin.

1.5 Justification for the study

To design, manage, operate and plan sustainable methods of utilization of water resources in Kenya a good understanding of the characteristics of rainfall and stream flow is required. A study of this nature requires data which is reliable, homogenous, representative and continuous without gaps. A network to collect such data does not exist hence this study wishes to design rainfall and stream flow networks for the Athi river basin which can be used to collect data for study and management of water resources in the basin.

While several studies have been conducted on the link between rainfall and ENSO in East Africa limited work has been undertaken on stream flow. Rainfall in the tropics exhibits complex spatial and temporal variability hence
rainfall contains meteorological noise which makes it a very complex component to study (Ogallo, 1983). The hydrological process in the catchment filters the meteorological noise in rainfall making stream flow a stable component (Piechota, et al., 1997).

Climate information is critical in water resources systems management, thus this information can provide useful tools for future planning of harnessing local water resources and reduction of vulnerability risks associated with hydro-climate hazards. It is therefore important to understand the space time linkages between the basin hydrological characteristics, extreme rainfall and the associated systems such as ENSO and IOD.

In Kenya, the evolutions of various ENSO phases including onset, peak and withdrawal phases of both El Niño and La Niña events have been associated with extreme rainfall events (Indenje, et al., 2000. The patterns of the rainfall extremes vary significantly not only from season to season but also from region to region (Philips and McIntyre, 2000; Owiti, 2005; Mutemi, 2005; Indenje, et al., 2000, etc). In recent years the inclusion of Indian Ocean Dipole (IOD) has been observed to increase the predictability of the Eastern African rainfall especially during March-May season (Rao, et al., 2002; Mehta, 1997; Owiti, 2005 and Vaidya, 2005). Thus the use of ENSO and IOD derived forecasts can form useful tools in preparedness, operation, planning and management of water supply, hydropower and early warning systems. Some of these formed the main focus of this study.
CHAPTER TWO

2.0 Literature Review

In this Chapter related studies are reviewed. While several attempts have been made to understand rainfall variability in East Africa, only limited work has been done on stream flow variability.

2.1 Seasonal rainfall and stream flow variability in East Africa

The climate of the East African region is influenced by large, meso and micro scale systems (Ogallo, 1989). The regional systems, which include meso and micro scale systems, significantly modify the impacts of the large-scale climate systems. Teleconnections of the regional weather and climate systems with El Niño-Southern Oscillation (ENSO), Indian Ocean Dipole (Rao, et al., 2002; Black, et al., 2003; Owiti, 2005 and Vaidya, 2005), Tropical Atlantic Ocean Dipole (Tourre, et al., 1999; Chang, et al., 1997; Mehta, 1997), the 30-day oscillation and the Quasi-Biennial Oscillation (QBO) have also been observed in East Africa. Details of some of these studies can be obtained from Indeje et al. 2000, Philips and McIntyre, 2000; Owiti, 2005; Omondi, 2005; Gitau, 2005 and Mutemi, 2005.

The dominant rainfall patterns over East Africa are related to the meridional and zonal oscillation of the inter-tropical convergence zone (ITCZ), which brings rainfall approximately one month after the sun’s path and the plane of the equator coincide (Asnani, 1983). The passing of the overhead sun twice
yearly, results into a bimodal rainfall pattern, with the primary rainy season occurring in March-May (MAM) locally called the ‘long rainy season’ and a shorter secondary rainy season occurring in September-November (SON) also locally known as the ‘short rainy season’. However, during an ENSO event, the spatial and temporal distribution of rainfall takes another picture all together (Ogallo, 1988).

In order to develop empirical models for long-range prediction of rainfall in East Africa there is need to reduce the network of point rainfall stations by delineating the East Africa region into homogenous zones (Ogallo, 1989; Basalirwa, et al., 1993). Regionalization and averaging of rainfall over large but homogeneous regions have the advantage of reducing meteorological noise in the data as well as minimizing the number of variables which describe regional climate variability (Indeje, et al., 2000). The homogeneous rainfall regions could also be used in the verification of the numerical climate model runs over the East Africa region (Sun, et al., 1999).

Several attempts have been made to study the spatial variability of rainfall in East Africa. Some of the studies include those of Atwoki, 1975; Ogallo, 1980, 1983, 1988, 1989; Barring, 1988; Okoola, 1998; Nyenzi, 1992; Semazzi, et al., 1996; Indeje, et al., 2000; Ngala, 2000; Opere, 1999 and Mutemi, 2005. Ogallo, 1989 used rotated principal component analysis (RPCA) method to study the characteristics of seasonal rainfall patterns over East Africa for the
period 1922-1983. The results showed that seasonal shifts in the patterns of dominant RPCA modes were similar to the seasonal migration of rainfall patterns associated with the ITCZ. The influence of large water bodies especially Lake Victoria and the Indian Ocean wave outstanding throughout the year. Twenty-six homogenous regional zones were delineated using the spatial characteristics of the dominant eigenvectors. The solutions based on temporal correlation matrices clustered together some of the wet and dry episodes. Some of the T-mode map patterns were associated with ENSO events.

Indeje, et al., 2000 used empirical orthogonal function (EOF) and simple correlation analysis to study the spatial and temporal variability of East Africa seasonal rainfall. The spatial patterns of the first and second rotated eigenvectors for the seasonal rainfall anomalies for the MAM and OND seasons accounted for 24.6% and 52.7% of the rainfall variance for the two rainfall seasons respectively. The first EOF mode for the OND rainfall season displays an east/west dipole pattern over the region. The spatial patterns of the second EOF mode for MAM and OND rainfall accounted for 10.8% and 6.1% of the rainfall variance respectively. The methods delineated East African seasonal rainfall into eight homogenous zones. The results showed unique seasonal evolution patterns in rainfall during different phases of ENSO cycles.

While several studies have been done on spatial and temporal variability of rainfall in East Africa, relatively few attempts have been made to study stream flow variability. Among the attempts are those of Mutua, 1985; Ondieki, 1993; Opere, 1999; Ngala, 2000; Agwata, 2005; Murimi, 2005 and Mutie, et al., 2006. Ngala, 2000 regionalized annual maxima flow series for the period 1960-1994 and identified two homogenous zones for the Athi river basin. Opere, 1999 used daily stream flow data to study the spatial and temporal characteristics of stream flow in the Kenya using RPCA and delineated Kenya into 3 homogenous stream flow zones. Murimi, 2005 studied the environmental impact assessment of soil and water resources in the watershed of Lake Nakuru. The study indicated that there is a decreasing stream flow trend in the watershed from 1967 despite fairly constant rainfall during the wet season in
the area. The study concluded that land use patterns, surface and subsurface abstractions are responsible for the decreasing stream flow. Agwata, 2005 delimited the upper Tana river basin into drought severity and duration regions and obtained four homogenous zones. The study revealed drought recurrence durations of 2.5-3.7 and 6.3-12.5 years. The study also fitted various probability distribution functions to obtain the best probability distribution function to describe drought duration and severity and found out that the generalized normal distribution was the most suitable. Mutie, et al., 2006 evaluated the land use effects on the Mara river based on the geospatial stream flow model found that diminishing of the river flow is because of changes in land use patterns due to farming activities, logging and charcoal burning.

2.2 Design of rainfall and stream flow networks

Lack of data in developing countries has hampered planning and sustainable utilization of the available water resources. WMO, 1981 gives guidelines for minimum rain gauge densities for various areas and types of terrain. In the tropics it required that for every 900-3000km$^2$ there should be one rain gauge station. However, rainfall stations are established as *ad hoc* responses to particular problems and needs. This is evidenced in the precipitation network map for East Africa (Tomsett, 1969). The rain gauge network depicted closely resembles the population density. Such network is unlikely to capture all the synoptic characteristics of rainfall in a region.
Many methods of rain gauge design have been suggested, these include that of Wilm, *et al.*, 1939; Ganguli, *et al.*, 1951; Kohler, 1958; Alaka, 1970; Hutchison, 1974; Dymond, 1982; Basalirwa, 1991 and Basalirwa, *et al.*, 1993 among others. Basalirwa, 1991 recommended a method purposed by Ahuja, 1959 for a rain gauge network for Uganda with an error of 10%. The method gave 374 gauge stations and 1517 gauge stations by the method of Wilm, *et al.*, 1939. A general weakness in the methods cited was that the largest number of gauges was derived for either dry areas or for dry seasons. Basalirwa, *et al.*, 1993 designed a minimum rain gauge network for Uganda using Principal Component Analysis (PCA) approach and obtained a minimum of 14 rainfall stations for the country.

Hydrological networks for stream flow monitoring are designed to monitor physiographic, climatic, hydrologic, biological and chemical features in a river basin. The density of the network depends on the nature of the terrain and water resources on the population creating a water demand (Shaw, 1994). In England and Wales, it was proposed that there should be 400 primary river gauging stations, equivalent to 1 station in every 375km² (Shaw, 1994). The ultimate design and establishment of a river gauging network depends on data requirement, the hydrological characteristics of the area and the cost-benefit relationship of the network.
2.3 Rainfall and stream flow variability during ENSO

Large scale circulation patterns are affected by ocean-atmospheric phenomena and a particularly well documented phenomenon is the El Niño Southern oscillation (ENSO). When warm water pools usually present in the western tropical Pacific Ocean shifts eastward towards the coast of Peru such case is often referred to as El Niño. La Niña is used to refer to cases when this shift is dominated by a pool of cold sea water. The term southern oscillation (SO) is used to describe the planetary atmospheric motion in response to El Niño phenomena. It involves seesaw in sea surface pressure between Indonesia and the southeast Pacific. Due to close association between El Niño /La Niña and SO, the two events are often collectively referred to El Niño Southern oscillation ENSO (Rasmusson and Wallace, 1983).

The ENSO phenomenon comprises a warm phase, termed El Niño and a cold phase called La Niña (Philander, 1990). The ENSO phenomenon has been linked to climatic anomalies throughout the world (Philander, 1990; Diaz and Markgraf, 1992; Allan, et al., 1996). Indicators of ENSO activity that are commonly used include; the Southern Oscillation Index (SOI) and equatorial Pacific Sea Surface Temperature (SST) over various locations commonly referred to as Niño 1, Niño 2, Niño 3 and Niño 4 respectively. The ENSO phenomenon is known to be a fundamental and quasi periodic feature of the ocean-atmosphere system, with periods ranging from 2 to about 7 years (Rasmusson and Carpenter, 1983; Halpert and Ropelewski, 1992). No two ENSO events are completely a like; they evolve according to a consistent
pattern but differ in timing, intensity, extent and duration. Quin, *et al.*, 1978 categorized El Niño events as: Strong, moderate, weak or very weak.

Several studies have been done on the link between rainfall and ENSO. Walker, 1923 pioneered the study of linking ENSO events and large-scale precipitation patterns so as to predict variations in the Indian Monsoon rainfall (Walker and Bliss, 1937). Recently, precipitation anomalies during the evolution of ENSO events have been also extensively analyzed over the low and middle latitudes (Rasmusson and Carpenter, 1982, 1983; Bradley, *et al.*, 1987; Ropelewski and Halpert, 1986, 1987; Jin, *et al.*, 2005). Jin, *et al.*, 2005 studied the relationship between SOI and observed precipitation in South Korea and Japan, the study showed significant correlation at 1% at lag time of 4 months during strong La Niña event category of SOI > 2.

Cayan and Peterson, 1989 investigated the influence of north Pacific atmospheric circulation on stream flow in the western United States. They pointed out that stream flow anomalies might be predicted two seasons in advance using the southern oscillation index and one season in advance by using the Pacific North America index. Koch, *et al.*, 1991 indicated that knowledge of the SOI in the late summer and fall seasons preceding the water supply forecast period (April-September) may provide guidance as to the nature of stream flow in the coming water year. Kuhnel, *et al.*, 1990 compared rainfall and runoff in South Eastern Australia and in the Southern United
States. They found that the Southern Oscillation (SO) signal that translated to rainfall and runoff is stronger in Australia than in the United States.

Kahya and Dracup, 1993 studied the stream flow patterns in relation to ENSO in the United States. They found out that there is a strong mid-latitude hydrologic response to the tropical ENSO phenomena. Piechota, et al., 1997 used PCA, Cluster analysis and Jackknife analysis to investigate the spatial and temporal modes that dominate stream flow variability in the western USA in response to ENSO events. They found out that stream flow anomalies occur after the onset of the ENSO event. The first three principal components explained more than 50% of temporal stream flow variance. Dracup and Kahya, 1994 studied the relationship between USA stream flow and La Niña events. They found that La Niña events were of opposite effect as compared to El Niño and there existed a significant relationship between stream flow and La Niña based on hyper geometric distribution.

Piechota, et al., 1997 used ENSO to forecast seasonal stream flow in eastern Australia. They found out that SOI is a better predictor of July-September and October-December stream flow while SST was a better predictor of January-March and April-June stream flow. Chiew, et al., 1998 studied the link and potential of forecasting of Australian rainfall, stream flow and drought using ENSO. They found that the link between rainfall, stream flow and ENSO was statistically significant in most parts of Australia. However, the link was not sufficiently strong to consistently predict rainfall and stream flow accurately.
Some of the studies that have been done in East Africa to study the relationship between rainfall and ENSO include those of Ogallo, 1988; Mutai, et al., 1998; Indeje, et al., 2000; Farmer, 1980; Hutchinson, 1992 and Philips and McIntyre, 2000. Ogallo, 1988 found peak lag zero correlation of $-0.6$ between SOI and October-November-December (OND) rainfall along the coast of southern Kenya and northern Tanzania. Similar correlation values relate SOI and September-December rainfall along the Kenyan Coast (Farmer, 1988) whereas higher values of $-0.8$ describe the relationship of SOI to the OND rainfall in Somalia (Hutchison, 1992). Indeje, et al., 2000, showed that a unique seasonal evolution patterns in rainfall exist during the different phases of the ENSO cycles. This affects the monthly and seasonal rainfall patterns in East Africa. In some regions ENSO shifts the onset and cessation of rainfall.

Mutai, et al., 1998 using SSTs from the Pacific Indian and Atlantic Oceans in a multiple linear regression prediction scheme for the East African short rains, found that variability in equatorial Pacific SSTs is the lead predictor, with equatorial Indian Ocean and Southern Atlantic ocean surface temperature playing a smaller, although significant role. Multiple regression values ranged from 0.56 to 0.78 depending on the period tested in regressions with July-August-September (JAS) SST predicting OND rainfall.

The analysis of SST-rainfall relationships as a function modality has resulted in significant relationships for sites with an August peak in rainfall. Ogallo, 1988 found that the relationship between SOI and JAS rainfall were strongest over
Uganda and western Tanzania and were opposite in sign from that relating SOI to regional OND rainfall. Philips and McIntyre, 2000 studied ENSO and interannual rainfall variability in Uganda and found out that Pacific Ocean Niño 3 region sea-surface temperature (SST) from July to September are significantly correlated with August-September rainfall (correlation coefficient $r = -0.75$) and the following November-December rainfall $(r = 0.57)$ but with opposite signs. The study separated rainfall stations in Uganda into those in the unimodal and bimodal zones. They found out the ENSO events were different in the two zones. In the unimodal zone, El Niño events are associated with a depression of the August rainfall peak, but a lengthening of the season. At bimodal zones, there is very little impact on August rainfall but November rainfall is enhanced in El Niño years and depressed in La Niña years.

While a lot of research has been done in East Africa to explore rainfall-ENSO relationship at regional level, relatively less attention has been directed to the investigation of the effects of ENSO on stream flow in different sub-regions of East Africa. This is because the rainfall runoff phenomenon integrates the effects of the hydrologic processes in the watershed (Kahya and Dracup, 1993; Piechota, et al., 1997)). The dynamic non linear relationships between precipitation and stream flow consists of four important intermediate processes; evaporation, transpiration, infiltration and storage (surface and subsurface). Effective precipitation plays an important role in stream flow variability (Kuhnel, et al., 1990).
2.4 Rainfall and stream flow variability during the Indian Ocean Dipole

The Indian Ocean Dipole (IOD) mode is a climate mode that occurs interannually in the tropical parts of the Indian Ocean. The dipole was first discovered by Saji, *et al.*, 1999 and Webster, *et al.*, 1999. As the name dipole suggests it has simultaneous variation in the East and West parts of Indian Ocean which is rooted in subsurface equatorial dynamics (Rao, *et al.*, 2002). During positive IODM events, sea surface temperature drops in the South Eastern part of Indian Ocean off north West Australia, East coast of Japan and throughout Indonesia while sea surface temperature rises in the West Equatorial Indian Ocean off the Eastern coast of Africa from northern Madagascar to Somalia coast and the reverse occurs during negative IODM events. Positive IODM events are associated with increased rainfall activities in East Africa and the reverse during negative IODM events (Black, *et al.*, 2003).

Vaidya, 2005 observed that there was a strong ENSO influence on interannual variability of precipitation and aridity in all rainfall seasons in East Africa while, IOD influence was evident in the June-September season which is the driest season over East Africa. Beltrando and Camberlin, 1993 described the teleconnections between ENSO, Indian Ocean surface field and Ethiopian summer and Somalia autumnal rainfall. Black, et al., 2003 have shown that extreme East Africa short rains (September-November) are associated with Indian Ocean SST anomalies and IOD and that enhanced rainfall in East Africa (e.g. 1961) may be an integral part of IOD. Recent studies have shown significant improvement in the prediction of seasonal rainfall with the inclusion of IOD indices (Owiti, 2005). IOD indices will therefore be used as predictors in this study.

The next section presents the methods that were used to address the various objectives of the study.
CHAPTER THREE

3.0 DATA AND METHODS

In this chapter the data and methods used in this study are discussed.

3.1 Introduction

Rainfall data for 34 rainfall stations and streamflow data from 18 river gauging stations in the Athi river basin were obtained from the Kenya Meteorological Department, Dagoretti corner and Ministry of Water and irrigation at Maji house respectively. The SOI Troup index was obtained from the National Climate centre of the Australia Bureau, Niño 3, Niño 3.4 and Niño 4 were obtained from NOAA/CPC, USA. The data was subjected to preliminary data analysis namely estimation of missing data and quality control analysis. These methods are presented first.

3.2 Estimation of missing data and data quality control

The quality and quantity of data used in any research is very important. In many parts of Kenya like in many developing countries, there exist many gaps in the data collected. The factors, which yield to these gaps, include breakdown of equipment, change of observations equipment, negligence on the observer, poor communication network, vandalism of the observation station, and changes of river courses among other reasons.
3.2.1 Estimation of missing data methods

To estimate missing data many techniques are used (WMO, 1983). However, stations with more than 10% of missing data were discarded (WMO, 1983). The methods used include; the arithmetic means, Thiessen Polygon, Isohyetal, Linear and correlation. The details of these methods are available in Shaw, 2004, Ross, 1987 and WMO, 1983 among others.

In this study the long term seasonal arithmetic mean was used to estimate missing monthly stream flow data, since stream flow is persistent with low temporal variability (Kachroo, 1992). A year was divided into four north-hemisphere seasons thus December-January-February (DJF), March-April-May (MAM), June-July-August (JJA) and September-October-November (SON). The seasonal stream flow mean of the months without missing data was used to fill in the missing data of the month with missing data in the season. To estimate missing monthly rainfall records, the linear correlation method was used and it is discussed next.

3.2.2 The Linear Correlation method

The method assumes that the neighbouring gauges receive rainfall from the same rainstorm. If station $A$ has the highest correlation with the gauging station with missing data $B$, the correlation method assumes that a linear correlation exists. The linear correlation coefficient, $r_{AB}$ at the pair of rain gauges $A$ and $B$ is determined by (WMO, 1983);
Where \( i \) is the \( i^{th} \) observation. If the long term monthly mean rainfall at stations \( A \) and \( B \) are \( \bar{A} \) and \( \bar{B} \) respectively, then the missing rainfall record at the \( j^{th} \) observation, given that \( A_j \) and \( B_j \) are the known and unknown records respectively. \( B_j \) was constructed using:

\[
B_j = \frac{A_j \bar{B}}{\bar{A}}
\] (2)

The accuracy of the estimated records were tested using the Chi-square test at 95% confidence level.

3.3 Data quality of rainfall and stream flow data

The quality of data is a priority for any research. Change of equipment, stations and observation routines affect data quality. Because of these changes it is hardly expected that hydro meteorological records will remain comparable over a long period of time at all stations. It is important therefore that climatological records are homogenous before they are used in any study. Some of the methods used for data quality studies include mass curves, correlation methods, trend analysis, non-parametric methods among others. In this study the mass curve, double mass curve and runs method were used to investigate the quality of data. These methods are discussed next.
3.3.1 Mass and double mass curve methods

In these methods, cumulative records (mass curve) or residual mass curve are plotted against time. Homogenous records will give a straight-line graph while heterogeneous ones will deviate significantly from the straight line.

In the double mass curve method one or more stations situated within the same climatic zone and known to have homogenous records are chosen. Temporal variations in these records and those of the test station are then compared. A plot of corresponding cumulative values is done to obtain a double mass curve. Like in the mass curve, a straight line indicates homogeneity otherwise it indicates heterogeneity. The heterogeneous records can be adjusted using the slopes of these curves. Details on mass curves methods can be obtained from Ogallo, 1981, 1988 and WMO, 1970, 1986 among others.

3.3.2 Runs Test Method

Graphical techniques for determining homogeneity are generally incomplete because they lack a criterion for determining whether the data is homogenous or not. A more objective test is preferred which can provide a criterion for accepting or rejecting a hypothesis of homogeneity on the basis of probability of occurrence. Such tests include: Chi-square test, Kolmogorov-Smirnov test, signs test, Mann-Kendall test, runs test and many other parametric and non-parametric tests. In this study the runs test method was used. The runs test method is a non-parametric test which is used to determine whether a particular
series is homogenous or whether there is a trend of periodic oscillations of the mean (Thom, 1969; Siegel, 1957). A run is defined as a number of sequences above or below some central value (for example mean or median). A count is made of the number of runs above the mean and below the mean in a naturally ordered series. If the series is homogenous the difference between the number of runs above and below the mean should be zero within the limit of a probable error. In this method, the mean of the data is computed first, all the other data points are then categorized as being above or below the mean. Then the number of runs \( r \) is determined. The significance of the observed number of runs is obtained using the modified normal variate \( Z \) given as

\[
Z = \frac{|r - \mu_r| - 0.5}{\sigma_r} \quad (3)
\]

Where \( \mu_r \) is the mean and \( \sigma_r \) is the standard deviation given by

\[
\mu_r = \frac{2n_1n_2}{n} + 1 \quad (4)
\]

\[
\sigma_r = \frac{2n_1n_2(2n_1n_2 - n)}{n_2(n_2 - 1)} \quad (5)
\]

Where \( n \) is the data length usually more than 40 years, while \( n_1 \) and \( n_2 \) are the number of observations above and below the mean respectively. In this study annual rainfall and annual stream flow data were subjected to the runs test and their homogeneity tested at 95 % confidence limit.

3.4 Design of a minimum rainfall and stream flow network

The quality of areal rainfall and stream flow data from a monitoring network depends on the instrument's site, the skill of the observer and the ability of the
collected data sample to represent all the spatial and temporal characteristics of
the given region. Establishment of dense networks to achieve high levels of
accuracy in areal rainfall and stream flow estimates is very expensive and is
not only unaffordable but also unsustainable in developing countries. A
cheaper rationale is the establishment of rainfall and stream flow monitoring
networks which provide data representative of all the unique spatial and
temporal characteristics of a region. A minimum monitoring network would be
a good approach to achieve this.

A minimum monitoring network includes in its design considerations of the
need to obtain rainfall and stream flow data from each distinct climatological
subdivision in the region of study. This can be achieved only if the spatial
extent and number of homogenous rainfall and stream flow regions are known
_a priori_, to ensure their representation in the designed minimum network. In
this study the Principal Component Analysis (PCA) approach was used to
develop minimum rainfall and stream flow networks for the Athi river basin.

### 3.4.1 Principal Component Analysis

Pearson (1902) first proposed the Principal Component Analysis (PCA)
method. The method identifies linear transformations of the data set that
concentrate as much of the variance as possible into a small number of
variables. A related and more complex technique, called factor analysis was
introduced by Spearman, 1904a and b and extended by Hotelling, 1936 who
formulated mathematical equations for defining unique solutions. The first applications to Meteorology were by Fakuoka, 1951; Lorenz, 1956; Holmstrom, 1963 and Obukhov, 1960. The Lucid exposition of PCA by Kutzbach, 1967 was instrumental in promoting the use of this technique in climate research.

The PCA method can be specified in at least six basic operational modes depending on which parameters are chosen as variables, individuals and fixed entities (Richman, 1986). The modes, which have been widely applied in climate research, are the S and T modes. A parameter for study is fixed and a correlation data matrix between stations over a set of period (S-mode) is obtained. A correlation data matrix can also be generated between periods (time) over a set of stations (T-mode). The S-mode yields a group of locations in terms of change over time while the T-mode yields a group of periods with similar spatial patterns. The PCA model is given by (Richman, 1986).

\[ Z_j = \sum_{k=1}^{m} a_{jk} F_k \quad (j = 1, 2 \ldots \ldots N) \quad (6) \]

where \( Z_j \) - The standardized variable \( j \)

\( F_k \) - Hypothetical factor \( k \) (Principal component)

\( N \) - number of variables

\( m \) - number of common factors.

\( a_{jk} \) - Loading of the variable \( j \) on factor \( k \).
3.4.1.1 Number of Significant Principal Components.

Many researchers have proposed methods of determining the number of significant principal components (PCs). These include Kaiser Criterion, the Scree method, the logarithm of the eigenvalue methods and sampling errors of the eigenvalues method. Details of these methods are given in Kaiser, 1959; Harman, 1976; North et al., 1982 and Richman, 1986. Some of the recent application of these methods to determine significant principal components in rainfall and stream flow includes those of Ogallo, 1988, 1989; Indeje, et al., 2000; Opere, 1999 and Ngala, 2000. In this study the Kaiser criterion was used to determine the number of significant principal components (PCs).

3.4.1.2 Rotation of Principal Components

In order to reduce ambiguities, which accompany direct solutions and obtain stable PCA patterns, the PCA solutions are rotated. Several criteria are used to adjust the frames of reference of the eigenvectors through rotation (Ogallo, 1989). The most common methods of rotation are orthogonal and oblique rotations. In orthogonal rotations the references axes are maintained at right angles, while in the oblique case the components are partially correlated. The types of the orthogonal rotations include the Varimax, Quantimax and Equimax methods. In this study orthogonal rotation was employed and the Varimax method was used. The Varimax method defines simple factors as one with ones and zeros in the column. In expanded form equation (6) may be written as
\[ Z_1 = a_{11} F_1 + a_{12} F_2 + \ldots + a_{1m} F_m \]
\[ Z_2 = a_{21} F_1 + a_{22} F_2 + \ldots + a_{2m} F_m \]
\[ Z_n = a_{m1} F_1 + a_{m2} F_2 + \ldots + a_{mn} F_m \]  \hspace{1cm} (7)

If factor 1, \( F_1 \) is a significant factor on variables \( Z_1, Z_3, Z_4, Z_6 \) and \( Z_n \) only and is not significant on variables \( Z_j, \ j = 2, 5, 7, \ldots, n-1 \), then coefficients of loading \( a_{11}, a_{31}, a_{41} \) and \( a_{mj} \) are approximately equal to one and all other coefficients in column one (i.e. \( a_{21}, a_{51}, a_{71}, a_{81}, \ldots, a_{(n-1)1} \)) are equivalent to zero. This is equivalent to maximizing the variance of the squared loading in each column.

### 3.4.1.3 Significant factor loadings

The factors loadings obtained by PCA are similar to correlation coefficients and therefore they can be treated the same way. Many users of PCA have widely, used as a rule of thumb, that loading having values \( \pm 0.3 \) or greater are taken as significant for sample size greater than 100. Because of the uncertainty surrounding the assessment of errors in small samples, Burt and Banks, 1952 formula is used to obtain significant factor loading thus

\[ SEL = SEC \sqrt{\frac{n}{n + 1 - m}} \]  \hspace{1cm} (8)

Where \( SEL = \) Standard error of loading

\[ SEC = \text{Standard error of correlation} \]

\( m = \) number of variables (years) in the analysis period
\( n \) = the factor number during factor extraction

The standard errors of correlation are obtained from Child, 1990.

The stability of spatial patterns was checked against inter stations correlations of both rainfall and stream flow respectively. The results were then used to delineate the basin into homogenous rainfall and stream flow zones. Using communality analysis, physical characteristics and location of stations, representative stations for each zone were obtained.

3.4.1.4 Communality analysis

In order to ensure that the station, which represents the individual homogenous zone, is realistic and could represent most of the regional information, the principal of communality was used. The communality of each variable \( Z_j \) given by \( h_j^2 \), is given by Nie, et al., (1970) as;

\[
h_j^2 = \sum_{k=1}^{m} a_{jk}^2 \quad (j = 1,2,\ldots, N)
\]

(9)

With \( a_{jk}, m \) and \( N \) defined in equation (6).

The communality of a station represents the degree of association it has with other stations in the data set. The best representative stations would therefore be the highest communality station of a homogenous zone. When two or more stations have the same communality other factors such as centrality of the station in the region and topographic characteristics of the zone are considered.
In this study the PCA methods was used to design the minimum rain gauge and stream flow network and to study the spatial and temporal variability of seasonal rainfall and stream flow in the Athi river basin. The next section presents the methods that were used to study the temporal characteristics of seasonal rainfall and stream flow.

3.5 Temporal characteristics of rainfall and stream flow

The rainfall and stream flow data were subjected to various temporal analyses that included analyzing the patterns of trends, cyclical and PCA (T-mode) analysis. Representative stations for rainfall and stream flow data in each PCA homogenous zones were used for the study.

3.5.1 Trend analysis

Trend analysis is aimed at detecting if there exist random or persistent change on the hydrological time series, which influence hydrological behaviour and consequently the water resources of the river basin.

The existence of increasing or decreasing trend in hydrological components can be induced by either natural and human induced factors such as in rainfall together with other climatic factors; changes of land use, etc (Mahera and Kolvy-Mahera, 1990; Giakoumakis and Baloutsos, 1997). Other factors could include change of observation site and change of equipment that are addressed through data quality control analyses.
The use of rank correlations tests is robust and the departures from the Gaussian normal distribution will not be of serious concern (WMO, 1966). In this study spearman rank correlation method was applied to study trend in monthly and seasonal rainfall and stream flow in the Athi river basin. The spearman rank correlations is calculated from

\[ \Gamma_s = 1 - \frac{6 \sum_{j=1}^{N} di^2}{N(N^2 - 1)} \]  

Where, \( di = k_i - 1 \), \( k_i \) is the rank of the series \( x_i \), and \( N \) the total number of observations. For \( N > 8 \), the value of \( \Gamma_s \) is tested for significance by computing the statistic \( t \) defined by

\[ t = \Gamma_s \left( \frac{N - 2}{1 - \Gamma_s^2} \right)^{\frac{1}{2}} \]  

The \( t \)-value is compared with probability points of students' \( t \)-distribution with \( N - 2 \) degree of freedom.

### 3.5.2 Periodicity analysis

The most common method of examining cycles in any time series has been spectral analyses (Jenkins and Watts, 1968; Ogallo, et al., 1984; Ondieki, 1996; Opere, 1999). Spectral analysis is simply a Fourier transformation of autocorrelations analysis. Spectral estimates in spectral analysis can be determined using three basic methods, the auto covariance transform, the fast Fourier transform and the maximum entropy method (Jenkins and Watts, 1968).
Recent periodic analysis approaches have however been based on wavelet analysis (Daubechies, 1992). In this study wavelet analysis was used to study periodicity in rainfall and stream flow in the Athi river basin. The method and some of its advantages are highlighted in the next section.

3.5.2.1 Wavelet analysis

Wavelet analysis is a tool for analyzing non-stationary variance at many different frequencies (Daubechies, 1992) within a geophysical time series (Torrence and Compo, 1998; Smith, et al., 1998; Labat, et al., 2000). Wavelets are a set of limited duration waves, also called daughter wavelets, because they are formed by dilations and translations of a single prototype wavelet function \( \psi(t) \), where \( t \) is real valued, called the basic or mother wavelet (Castleman, 1996). The mother wavelet designed to oscillate like a wave, is required to span an area that sums to zero and die out rapidly to zero as \( t \) tends to infinity to satisfy the so-called admissibility condition i.e.

\[
\int \psi(t) dt = 0 \quad (t \to \infty) \tag{12}
\]

In this study to compute the wavelet power, the Morlet wavelet (the non dimensional frequency \( w_0 = 6 \)), was used because its structure resembles that of a rainfall time series, given by (Torrence and Compo 1998);

\[
\psi(t) = \pi^{-1/4} e^{i w_0 t} e^{-t^2 / 2} \tag{13}
\]
Examples of other wavelet functions include the Paul, Mexican hat and derivative of Gaussian (DOG), details are given in Torrence and Compo (1998).

The continuous wavelet transform $W_n$ of a discrete sequence of observations $x_n$ is defined as the convolution of $x_n$ with a scaled and translated wavelet $\psi(\eta)$ that depends on a non-dimensional time parameter $\eta$ (Torrence and Compo, 1998);

$$W_n(s) = \sum_{n=0}^{N-1} x_n \psi^* \left[ \frac{(n' - n)\delta t}{s} \right]$$  \hspace{1cm} (14)

Where $n$ is the localized time index, $s$ is the wavelet scale, $\delta t$ is the sampling period, $N$ is the number of points in the time series and the asterisk indicates the complex conjugate. Since complex wavelets lead to complex continuous wavelet transform, the wavelet power spectrum, defined as, $|W_n(s)|^2$ is a convenient description of the fluctuation of the variance at different frequencies. Further, when normalized by $\sigma^{-2}$ (where $\sigma^2$ is the variance) it gives a measure of the power relative to white noise, since the expectation value for a white noise process is $\sigma^{-2}$ at all $n$ and $s$.

To determine significance levels for wavelet spectrum an appropriate background spectrum was chosen. The many geophysical phenomena, an appropriate background spectrum is either white noise (with a flat Fourier
spectrum) or red noise (increasing power with decreasing frequency) (Torrence and Compo, 1998). It has been shown on average that the local wavelet power spectrum is identical to the Fourier power spectrum given by

\[
P_k = \frac{1 - \alpha^2}{1 + \alpha^2 - 2\alpha \cos \left( \frac{2\pi k}{N} \right)}
\]

(15)

Where \( k = 0, \ldots, N/2 \) is the frequency index. By choosing an appropriate lag-1 autocorrelation equation (15) can be used to model a red-noise spectrum. If \( \alpha = 0 \) in equation (15) then it models a white noise spectrum.

If \( x_n \) is a normally distributed random variable, then both the real and imaginary parts of \( \hat{x}_k \) are normally distributed (Chatfield, 1989). Hence \( |\hat{x}_k|^2 \) is chi-square distributed with two degrees of freedom, denoted by \( X^2_2 \) (Jenkins and Watts, 1968). In order to determine the 95% confidence level, the background spectrum (15) is multiplied by the 95\textsuperscript{th} percentile value of \( X^2_2 \) (Gilman, \textit{et al}, 1963). The confidence interval at each scale can be used to construct confidence contours. In this study the 95% confidence limit was used to study the periodicity of seasonal rainfall and stream flow in the Athi river basin.
3.5.3 PCA (T-mode) analyses

S-mode PCA analysis method and an introduction to PCA was presented in section 2.3.1. In this study T-mode PCA analysis was also performed using the same concepts as for S-mode. In PCA T-mode analysis a correlation data matrix is generated between periods (time) over a set of stations. The PCA T-mode analysis yields a group of years with similar spatial patterns. The years with similar spatial patterns will tend to cluster on to similar PC modes by having large factor loadings on them. The years which cluster on to similar PC modes experience almost similar spatial rainfall anomalies (Ogallo, 1989). In this study the time coefficients of the dominant PC modes were used as a regional rainfall and stream flow anomaly index for the Athi river basin region. Areal rainfall and stream flow averages during the individual years clustered together were computed using rainfall and stream flow records from the regional stations. An attempt was also made to relate some of rainfall and stream flow spatial patterns so climatic systems. Similar studies which have used this method include those of; Ogallo, 1989; Okoola, 1998; Opere, 1999 among others.

3.6 Variability of rainfall and stream flow during ENSO and IOD

In this section, the methods used to study spatial variability of rainfall and streamflow in the basin during ENSO and IOD phases are provided. Both the cold and warm ENSO phases as well as the positive and negative IOD phases were used. The ENSO indices used were derived from the standard ENSO
indices namely SOI, Niño 3 SST and Niño 4 SST. The Troup SOI (Troup, 1965) was obtained from the national climate centre of the Australian Bureau while Niño 3 SST (5° N, 60°E - 5°S, 90°W) and Niño 4 SST (5° N, 160°E - 5°S, 150°W) were obtained from NOAA/CPC. It should be noted that several new ENSO indices are now available that include multivariate indices (Wolter, 1987; Wolter and Timlin, 1993) but this study used the indices from the national climate centre of the Australian Bureau in order to enable easy comparison with some of the previous work that have been undertaken in the region.

The Indian Ocean dipole is quantified by the difference between the grid boxes (50° E- 70° E, 10°S-10°N and 90°E-110°E, 10°S-0°) Saji, et al., 1999. The Indian Ocean Dipole is calculated as the difference between SSTs averaged over western and eastern Indian Ocean as given by Saji, et al., 1999 as

\[
\text{IOD}_W = \text{Average SST} (50° E- 70° E, 10°S-10°N), \quad \text{IOD}_E = \text{Average SST} (90°E-110°E, 10°S-0°) \quad \text{and} \quad \text{IOD Index} = \text{IOD}_W - \text{IOD}_E
\]

The Indian Ocean dipole has positive and negative events. During the positive event the SST drops in the south eastern part of Indian Ocean off the north coast of Australia, eastern coast of Japan and Indonesia, while SST rises in the western equatorial of Africa, from northern half of Madagascar to the northern edge of Somalia (Merchant, et al., 2006). The conditions occur during the negative event. The IOD data used in this was from 1961-2004. Recent
applications of the IOD to study rainfall variability in East Africa include; Gitau, 2005; Owiti, 2005; Vaidya, 2005 and Bamanya, 2007.

The composite and S-mode PCA methods were used to study spatial similarities in the patterns of rainfall and stream flow during ENSO and IOD phases. Simple correlation analysis was used to examine teleconnections between seasonal rainfall/stream flow, ENSO and IOD indices. Finally, canonical correlations analysis (CCA) method was used to study teleconnections between seasonal rainfall and stream flow in the Athi river basin with Indian Ocean SSTs.

3.6.1 Composite analysis

The method involves identifying and averaging one or more categories of fields of a variable selected according to their association with ‘Key’ condition (Folland, 1983). The results of these composites are then used to generate hypotheses for patterns, which may be associated with the individual scenarios. The advantage of composite fields is that they are easy to interpret.

A number of studies have shown that the results of composite analysis are similar to those of correlation methods (Ward, 1992). This method has been used to study climatic parameters in East Africa, recent examples include those of Mataria and July, 1992; Ininda, 1995 and Okoola, 1998.
In this study the composites of rainfall and stream flow derived from the warm and cold ENSO and positive and negative IOD events were used to study the spatial variability of rainfall and stream flow during ENSO/IOD while those derived from the MAM and SON rainfall and stream flow seasons were used to identify wet, dry, normal, high flow and low flow years in the Athi river basin.

The method of categorization rainfall and stream flow into wet/high and dry/low flows involves the use of a normalization index $q$ expressed as:

$$ q = \frac{100}{\bar{y}} y $$

(16)

Where $y$ represents the observation while $\bar{y}$ the long term mean. This categorization is based on the current drought indices used at IGAD climate prediction and application centre (ICPAC).

**Table 1. Classification of the wet/dry rainfall and stream flow**

<table>
<thead>
<tr>
<th>SCENARIO TYPE</th>
<th>ANOMALY RANGE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wet/High flow</td>
<td>$q &gt; 125$</td>
</tr>
<tr>
<td>Normal</td>
<td>$75 \leq q \leq 125$</td>
</tr>
<tr>
<td>Drought/Low flow</td>
<td>$q \leq 75$</td>
</tr>
</tbody>
</table>

$q$ values for both warm and cold ENSO and positive and negative IOD phases were subjected to principal component analysis separately. The significant principal components were used to study the rainfall and stream flow variability during the positive and negative ENSO and IOD phases.
3.6.2 Simple correlation method

The simple correlations coefficient ($r_k$) between the predictand $y_i$ and any predictor ($x_{i-k}$) at time lag $k$ is defined as

$$r_k = \frac{\sum_{t=1}^{m} X_{t-k} Y_t}{\left[ \sum_{t=1}^{m} Y_t^2 \sum_{t=1}^{m} X_{t-k}^2 \right]^{1/2}}$$ (17)

Where $X_t = x_i - \bar{x}$ and $Y_t = y_i - \bar{y}$, $r_k$ ranges between -1 to 1. It is positive or negative if the variables are positively or negatively correlated, zero if there is no correlation and $r=1$ for perfect correlation. The student $t$-statistic was used to determine whether $r$ is significant at 95% confidence limit. The $t$-statistic is given by Ross, 1987 as

$$t = \left[ \frac{(m - 2) r^2}{1 - r^2} \right]^{1/2}$$ (18)

$m$ - is the number of years covered and $r$ is the correlation coefficient

In this study seasonal rainfall and stream flow data representative stations in each homogenous region were independently correlated with ENSO indices at various time lags.

3.6.3 Canonical Correlation Analysis (CCA)

Canonical correlation analysis (CCA) provides much more powerful teleconnection analyses tools between two data sets. The technique isolates the
linear combination of data from the left field (e.g. rainfall or stream flow) and the linear combination of data from the right field (e.g. SST) that have the maximum correlation coefficients. There are two approaches to CCA (Anderson, 1958 among others) a classical treatment and an approach due to Barnett and Preisendorfer, 1987 in which the time series of each field is filtered by projection onto a leading subject of its EOFs and then the maximum correlation between linear combinations of the filtered time series of the two fields is sought. In this study the CCA approach by Barnett and Preisendorfer, 1987 and Graham, 1990 was used. Recent application of this approach in East Africa includes those of Indeje, et al., 2000; Mutemi, 2005 and Omondi, 2005.

Due to close linkages between East African regional rainfall and Indian Ocean SST (Ogallo, 1988, 1989; Bazira and Ogallo, 1999; Mutai, et al., 1989; Phillips and McIntyre, 2000 and Indeje, et al., 2000), in this study Indian Ocean SST modes were used in the CCA model. CCA was used to isolate the most dominant modes of the co-variability between two variables (rainfall or stream flow on the left hand and Indian Ocean region SST on the right hand) and secondly to determine the degree of connection between pairs of patterns in the co-variability of the two variables. The rainfall or stream flow field \( Z \) over the Athi river basin is presumed to be influenced by the variability of Indian Ocean SST \( Y \). Both \( Y \) and \( Z \) have equal temporal dimension \( (nt) \) and spatial dimension \( (ny, nz, ny \neq nz) \) respectively. The CCA assumes that for a given SST pattern there exist corresponding rainfall pattern.
A reduced number of linear combinations of \( Z \) and \( Y \) are determined such that corresponding pairs of those linear combinations have a maximum possible correlation and these sequences of the new pairs (or patterns) are orthogonal to the projections of the variable on any other identifiable patterns.

### 3.6.3.1 The Canonical Correlation Analysis Method

Consider two data sets \( Y_{t,y} \) (e.g. SST) and \( Z_{t,z} \) (rainfall, stream flow, etc). If \( U \) and \( V \) denote the canonical vectors of \( Y \) and \( Z \) respectively which are maximally correlated i.e.

\[
U = YR, \quad V = ZQ \tag{19}
\]

\[
\text{Corr} \ (U, V) = \frac{\sum UV}{\sqrt{\sum U^2} \sqrt{\sum V^2}} = \text{Max} \tag{20}
\]

\( U \) and \( V \) are projected values of \( Y \) and \( Z \) respectively

Solving the system of equations for \( R \) and \( U \) we obtain

\[
\text{Var}(U) = \frac{\sum U}{nt-1} = \frac{U^T U}{nt-1} = \frac{(YR)^T YR}{nt-1} = \frac{R^T Y^T YR}{nt-1} \tag{21}
\]

\[
\frac{Y^T Y}{nt-1} = S_{yy} \quad \text{(The covariance matrix of \( Y \))}
\]

Therefore

\[
\sum \bar{y} = \bar{R} \bar{S}_{yy} R, \quad \sum \bar{v} = \bar{Q} \bar{S}_{zz} R \quad \text{and} \quad \sum UV = \bar{R} \bar{S}_{yz} Q \tag{22}
\]
Where

\[
S_{yy} = \frac{Y^T Y}{nt - 1}, \quad S_{zz} = \frac{Z^T Z}{nt - 1}, \quad S_{yz} = S_{zy} = \frac{Y^T Z}{nt - 1}
\]  

\[
\text{Corr} (U, V) = \frac{R^T S_{yz} Q}{\left( R^T S_{yy} R \times Q^T S_{zz} Q \right)^{1/2}} = \text{Max} \tag{23}
\]

To avoid \( R \) and \( Q \) being large matrices, we set the constraint that total variance must be 1 i.e.

\[
R^T S_{yy} R = 1, \quad Q^T S_{zz} Q = 1 \tag{24}
\]

Let \( F(R, Q) = R^T S_{yz} Q \) be maximized subjected to \( R^T S_{yy} R = 1 \) and \( Q^T S_{zz} Q = 1 \). To solve the problem two Lagrange multipliers \( \mu \) and \( \lambda \) are used hence

\[
F(R, Q) = R^T S_{yz} Q - \mu (Q^T S_{zz} Q - 1) - \lambda (R^T S_{yy} R - 1) \tag{25}
\]

\[
\frac{\partial F}{\partial R} = 0 \quad \text{and} \quad \frac{\partial F}{\partial Q} = 0 \tag{26}
\]

\[
\left( S_{yy}^{-1} S_{yz} S_{zz}^{-1} S_{zy} - \mu^2 I \right) R = 0 \,. \tag{27}
\]

Equation (30) is an eigenvalue problem where \( \mu^2 \) are the eigenvalues of \( S_{yy}^{-1} S_{yz} S_{zz}^{-1} S_{zy} \), \( R \) are the eigenvectors and \( I \) is the identity matrix. Note that \( Q = S_{zz}^{-1} S_{zy} R \). In CCA, \( R \) and \( Q \) are called the canonical spatial patterns and \( \mu \) is called the canonical correlation coefficient between \( U \) and \( V \).
3.7 Rainfall and stream flow forecasting

One of the main objectives of the study was to identify unique linkages between seasonal rainfall and stream flow with ENSO indices. This link could be used for rainfall and stream flow forecasting for management, planning and operation of water resources systems in the Athi river basin. In this study an attempt was made using regression and Non-parametric Seasonal Forecasting Model (NSFM) methods to obtain a long-range rainfall and stream flow forecasting for SON rainfall and stream flow.

3.7.1 Regression analysis method

A regression model expresses the mathematical relationship between the independent variables (predictors) and the dependent variable (Predictand). A linear regression model comprising a set of $k$ parameters $a_0, a_1, \ldots, a_k$ that relate one or more predictors $x_1, x_2, \ldots, x_k$ to a single predictand $y$ is given by

$$y = a_0 + a_1 x_1 + a_2 x_2 + \ldots + a_k x_k$$  \hspace{1cm} (29)

The regression can take several forms. When a simple linear relationship assumption is made between the dependent (predictor) and independent (predictand) a simple linear regression model is obtained, while when the predictors are more than one a multiple regression model is obtained. If the predictors are introduced step by step, a step-wise regression model is obtained. The significance of the predictors in the regression equation was tested using various tests that include the variance ratio test ($F$-test) among
many other tests. The details of regression analysis and recent applications in East Africa include (Ross, 1987; Bazira and Ogallo, 1999; Indeje; 2000, Owiti, 2005 among others).

3.7.2 Non-Parametric Seasonal Forecasting Model (NSFM)

The model uses a non-parametric approach to forecast continuous exceedance probabilities of available thresholds for example stream flow. The method uses linear discriminant analysis to empirically fit the data without making any prior assumption of the model distribution structure. The variable (stream flow, rainfall, etc) is separated into three categories; below normal, normal and above normal. After separating the data into three categories, the posterior probability of each stream flow category conditioned on the predictor variable are calculated from the Bayes' probability theorem (Davis, 1986)

$$\Pr\left( \frac{Q_i}{x} \right) = \frac{p_i f_i(x)}{\sum_{i=1}^{k} p_i f_i(x)}$$  \hspace{1cm} (30)

Where $Q_i$ = category $i$ stream flow, $p_i$= prior probability of category $i$ stream flow (30% for below normal and above normal and 40% for normal): $f_i(x)$ = probability density functions (PDF) of the predictor variable in category $i$ and $x$ = observed value of the predictor variable.
To calculate \( f_i(x) \) at some given predictor value \( x \), it is necessary to fit the probability density function to three subsets of the predictor data. The PDF is estimated non-parametrically.

### 3.7.2.1 Parameter estimation of the model

A commonly used non-parametric method for parameter estimation of a PDF is the kernel density estimation method. The kernel density estimate is defined as

\[
f^\hat{\cdot} = \frac{1}{hn} k \left( \frac{x - x_i}{h} \right)
\]

Where \( x_i \) \( \ldots \ldots \) \( x_i \) is a set of \( n \) observations; \( k() \) is the kernel function and \( h \) is band width. Several types of kernel functions, such as rectangular, Epanechnikov and triangular are available, however, studies have shown that the choice of \( k() \) is secondary to the choice of the bandwidth \( (h) \). Silverman (1986) found the optimal width \( (h) \) of the Gaussian kernel density estimator using adaptive measure is given by

\[
h_i = 0.9 A_i n_i^{-\frac{1}{2}}
\]

\[
A_i = \min(\sigma_i, \text{inter-quartile range}) / 1.34
\]

Where \( \sigma_i \) = standard deviation of the predictor data in each subset \( i \) and \( n_i \) number of observations in each subset. Details of this method are given in Chiew and Siriwardena, 2005.
3.7.2.2 Calibration and verification of the model

For the model to be used it must be calibrated and verified using data and its forecasting skill, established. The skill of the non-parametric seasonal forecasting model is established using two criteria (i) Nash-Sutcliffe coefficient of efficiency ($E$) and (ii) The modified linear error in probability space (LEPS) score (Piechota, et al., 2001).

3.7.2.3 Nash–Sutcliffe coefficient of efficiency ($E$)

The efficiency criterion is by Nash and Sutcliffe, 1970 as:

$$E = \frac{\sum_{i=1}^{m} (Q_{oi} - Q_{av})^2 - \sum_{i=1}^{m} (Q_{oi} - Q_{si})^2}{\sum_{i=1}^{m} (Q_{oi} - Q_{av})^2}$$

where $Q_{oi}$ is measured runoff, $Q_{si}$ is simulated runoff and $Q_{av}$ is average measured runoff, $m$ is number of observations. This measure provides a measure of the agreement between deterministic average forecast runoff and the recorded runoff. A higher $E$ is indicates a better agreement between the forecast and the recorded (observed) values. A value 1.0 indicates that the average forecast is exactly the same as the actual values. As general rule of the thumb an $E$ value greater than 0.1 indicates that there is some skill in the forecast.
3.7.2.4 The modified linear error in probability space (LEPS) score

The LEPS score measures, in a probabilistic sense, the distance between forecast and observed values and is defined as

$$S^* = 3(1 - |p_f - p_o| + p_f^2 - p_f + p_o^2 - p_o) - 1$$

(35)

Where \( p_f \) and \( p_o \) are the cumulative probabilities of the forecast and observed values found from the climatology non-exceedance curve. For an exceedance probability forecast the average LEPS skill score is given as (Piechota, et al., 2001)

$$LEPS = \frac{\sum_{i=1}^{m} 100 S^*}{\sum_{i=1}^{m} S^*_n}$$

(36)

Where \( S^*_n \) depends on whether \( s^* \) is positive or negative. Details of this method are given in Piechota, et al., 2001.

A higher LEPS score indicates a higher model forecasting skill. A LEPS score of greater than 10 is generally considered a good skill (Chiew and Siriwardena, 2005). In this study the model was used to investigate its ability to forecast rainfall and stream flow at various lead times in the Athi river basin. The next chapter gives results and discussions.
CHAPTER FOUR

4.0 RESULTS AND DISCUSSION

This chapter presents the results and discussions of the study obtained using the methods discussed in Chapter three.

4.1 Introduction

The preliminary data analysis results are presented first followed by, PCA, trend analysis, periodicity, teleconnections and finally forecasting results.

4.2 Estimation of missing data

Tables 2 and 3 give the respective listing of rainfall and river flow stations and their corresponding percentages of missing data in the Athi river basin which were used in this study. The stations with more than 10 percent of missing data were excluded from any further analysis because they did not meet the WMO criteria (WMO, 1983). The rainfall and streamflow missing data were estimated using the linear correlation method and seasonal means respectively. Finally, the spatial distribution of rainfall and stream flow stations used in this study is given in figures 2a and 2b.
Table 2. Rainfall stations used in the study

<table>
<thead>
<tr>
<th>STATION NAME</th>
<th>CODE</th>
<th>DATA LENGTH</th>
<th>% OF MISSING DATA</th>
</tr>
</thead>
<tbody>
<tr>
<td>NGONG FOREST STATION</td>
<td>9136009</td>
<td>1961-2004</td>
<td>5</td>
</tr>
<tr>
<td>DOONDU ESTATE</td>
<td>9136018</td>
<td>1961-2004</td>
<td>2</td>
</tr>
<tr>
<td>NATIONAL AGRI. LABS. NAIROBI</td>
<td>9136025</td>
<td>1961-2004</td>
<td>5</td>
</tr>
<tr>
<td>KARURA FOREST STATION</td>
<td>9136027</td>
<td>1961-2004</td>
<td>3</td>
</tr>
<tr>
<td>THIKA KARAMAINI</td>
<td>9136029</td>
<td>1961-2004</td>
<td>3</td>
</tr>
<tr>
<td>KIGWA ESTATE</td>
<td>9136037</td>
<td>1961-2004</td>
<td>3</td>
</tr>
<tr>
<td>RUIRU JACARANDA COFFEE RES. STN.</td>
<td>9136084</td>
<td>1961-2004</td>
<td>2</td>
</tr>
<tr>
<td>NAIROBI EASTLEIGH AERODROME</td>
<td>9136087</td>
<td>1967-2004</td>
<td>4</td>
</tr>
<tr>
<td>NAIROBI WILSON AIRPORT</td>
<td>9136130</td>
<td>1961-2004</td>
<td>3</td>
</tr>
<tr>
<td>DAGORETTI CORNER MET STATION</td>
<td>9136164</td>
<td>1961-2004</td>
<td>2</td>
</tr>
<tr>
<td>KIKUYU AGRICULTURE OFFICE</td>
<td>9136165</td>
<td>1961-2004</td>
<td>2</td>
</tr>
<tr>
<td>J.K.I.A.MET.SATION.</td>
<td>9136168</td>
<td>1961-2004</td>
<td>3</td>
</tr>
<tr>
<td>MACHAKOS, MATILIKU</td>
<td>9137028</td>
<td>1961-2004</td>
<td>6</td>
</tr>
<tr>
<td>DONYO SABUK KIANZAVI F.C.S.LTD.</td>
<td>9137054</td>
<td>1961-2004</td>
<td>12</td>
</tr>
<tr>
<td>KYUSYANI CHIEF'S OFFICE</td>
<td>9137092</td>
<td>1961-2004</td>
<td>10</td>
</tr>
<tr>
<td>MBOONI FOREST STATION</td>
<td>9137099</td>
<td>1961-2004</td>
<td>2</td>
</tr>
<tr>
<td>KITONDO FOREST STATION</td>
<td>9137101</td>
<td>1961-2004</td>
<td>3</td>
</tr>
<tr>
<td>MAKindu MET. STATION</td>
<td>9237000</td>
<td>1961-2004</td>
<td>2</td>
</tr>
<tr>
<td>KIBWEZI, DWA PLANTATION LTD.</td>
<td>9237002</td>
<td>1961-2004</td>
<td>3</td>
</tr>
<tr>
<td>TAVETA ZIWANI SISAL ESTATE</td>
<td>9337081</td>
<td>1961-2004</td>
<td>16</td>
</tr>
<tr>
<td>VOI METEOROLOGICAL STATION</td>
<td>9338001</td>
<td>1961-2004</td>
<td>3</td>
</tr>
<tr>
<td>D.C.'S OFFICE, WUNDANYI</td>
<td>9338003</td>
<td>1961-2004</td>
<td>2</td>
</tr>
<tr>
<td>VOLRUKANGA-KASIGAU</td>
<td>9338018</td>
<td>1961-2004</td>
<td>2</td>
</tr>
<tr>
<td>MAZERAS RAILWAY STATION</td>
<td>9339000</td>
<td>1961-2004</td>
<td>19</td>
</tr>
<tr>
<td>GANZE DISPENSARY</td>
<td>9339012</td>
<td>1961-2004</td>
<td>15</td>
</tr>
<tr>
<td>JILORE FOREST STATION</td>
<td>9339045</td>
<td>1961-2004</td>
<td>5</td>
</tr>
<tr>
<td>MSABAHA AGROMET STATION</td>
<td>9340007</td>
<td>1961-2004</td>
<td>2</td>
</tr>
<tr>
<td>MALINDI METEOROLOGICAL STATION</td>
<td>9340009</td>
<td>1961-2004</td>
<td>9</td>
</tr>
<tr>
<td>KWALE AGRICULTURAL DEPARTMENT</td>
<td>9439001</td>
<td>1961-2004</td>
<td>2</td>
</tr>
<tr>
<td>MSAMBWENI DISTRICT OFFICE</td>
<td>9439014</td>
<td>1961-2004</td>
<td>8</td>
</tr>
<tr>
<td>KINANGO AGRIC. OFFICE KWALE</td>
<td>9439015</td>
<td>1961-2004</td>
<td>8</td>
</tr>
<tr>
<td>MOMBASA PORT REITZ AIRPORT</td>
<td>9439021</td>
<td>1961-2004</td>
<td>4</td>
</tr>
<tr>
<td>VANGA LUNGA LUNGA CAMP</td>
<td>9439046</td>
<td>1961-2004</td>
<td>9</td>
</tr>
</tbody>
</table>
Table 3. River flow gauging stations used in the study

<table>
<thead>
<tr>
<th>Station Name</th>
<th>Station code</th>
<th>Area (sq. km)</th>
<th>Data Length</th>
<th>% of missing Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ruaraka</td>
<td>3BA10</td>
<td>53</td>
<td>1960-1984</td>
<td>10</td>
</tr>
<tr>
<td>Tigoni</td>
<td>3BA17</td>
<td>17.7</td>
<td>1960-1991</td>
<td>5</td>
</tr>
<tr>
<td>Ruaraka</td>
<td>3BA18</td>
<td>51.4</td>
<td>1960-1995</td>
<td>3</td>
</tr>
<tr>
<td>Tusoga</td>
<td>3BA39</td>
<td>5.18</td>
<td>1960-1982</td>
<td>8</td>
</tr>
<tr>
<td>Ruiru</td>
<td>3BB10</td>
<td>41.4</td>
<td>1960-1995</td>
<td>10</td>
</tr>
<tr>
<td>Kiu</td>
<td>3BB11</td>
<td>41.4</td>
<td>1960-1995</td>
<td>10</td>
</tr>
<tr>
<td>Kiu</td>
<td>3BB13</td>
<td>103</td>
<td>1960-1994</td>
<td>7</td>
</tr>
<tr>
<td>Ruiru</td>
<td>3BC07</td>
<td>73</td>
<td>1965-1992</td>
<td>3</td>
</tr>
<tr>
<td>Bathi</td>
<td>3BC09</td>
<td>52.5</td>
<td>1960-1994</td>
<td>10</td>
</tr>
<tr>
<td>Ruiru</td>
<td>3BC12</td>
<td>357</td>
<td>1960-2000</td>
<td>3</td>
</tr>
<tr>
<td>Komothai</td>
<td>3BC13</td>
<td>103</td>
<td>1960-1995</td>
<td>7</td>
</tr>
<tr>
<td>Gatamayu</td>
<td>3BC15</td>
<td>65.5</td>
<td>1960-1995</td>
<td>10</td>
</tr>
<tr>
<td>Theta</td>
<td>3BD02</td>
<td>53</td>
<td>1960-1995</td>
<td>3</td>
</tr>
<tr>
<td>Ndarugu</td>
<td>3CB02</td>
<td>65.1</td>
<td>1960-1999</td>
<td>10</td>
</tr>
<tr>
<td>Athi munyu</td>
<td>3DA02</td>
<td>5724</td>
<td>1960-1996</td>
<td>5</td>
</tr>
<tr>
<td>Athi Mavindini</td>
<td>3F02</td>
<td>4521</td>
<td>1952-1996</td>
<td>3</td>
</tr>
<tr>
<td>Tsavo</td>
<td>3G02</td>
<td>7252</td>
<td>1949-1991</td>
<td>5</td>
</tr>
<tr>
<td>Mzima</td>
<td>3G03</td>
<td>306</td>
<td>1951-1990</td>
<td>8</td>
</tr>
</tbody>
</table>

4.3 Quality Control test Results

The rainfall data mass curve showed homogeneity at many locations (figure 3b) while river flow data mass curves revealed heterogeneity at several gauging stations (Figure 3a). The results for the runs test showed appreciable heterogeneity significant at 95% confidence limit in rainfall data for Msambweni District office and Kibwezi, Dwa Plantation stations. The test also showed that there was appreciable heterogeneity significant at 95% in confidence limit in river flow data at Athi-Munyu, Ruiru (3BC07), Kiu (3BB13) and Theta river gauging stations. The data from rainfall and river flow stations with significant heterogeneity were subjected to double mass curve analysis and adjustments were made accordingly. The quality controlled data formed the foundation for this particular study.
Figure 2a. Distribution of the rainfall stations used in the study

Figure 2b. Distribution of the river gauging stations used in the study
Figure 3a. Mass curve for river gauging station at Kiu (3BB13)-Heterogenous

Figure 3b. Mass curve for rainfall station at Mombasa-Homogenous
4.4 Design of optimum rainfall and stream flow networks results

The design of optimum rainfall and stream flow in Athi river basin were obtained by subjecting seasonal rainfall and stream flow data to Principal Component Analysis (PCA). Details of the methods were presented in section 2.3. The rainfall and stream flow minimum network patterns were obtained for the major rainfall season namely March-April-May (MAM) and September-October-November (SON). The 95% confidence level indicated that all factor loadings equal or greater than 0.4 were significant. The principal component modes of significant factors were used to describe the spatial and temporal characteristics of seasonal rainfall and stream flow in the Athi river catchment. In this chapter only a few PC mode maps for MAM and SON seasons are given.

4.4.1 Results for spatial variability of MAM season rainfall using PCA

Table 4a gives the PCA results for MAM rainfall season. Only three Principal Components (PCs) had eigen values greater than 1. Thus only three PC modes were significant which accounted for 76% of the season’s rainfall variance. Figures 4a-c, shows the spatial patterns of the factor loadings of the first, second and third PC modes for the MAM rainfall season. The first PC mode exhibited dominancy over the central and north western part of the basin while the second PC mode was dominant at the coast. The third PC mode on the other hand showed some dominancy over the central part of the basin and parts east of Mt. Kilimanjaro.
The MAM season is the major rainy season over East Africa and is normally referred to as the 'long rains' season. Due to heating by the overhead sun during the season larger part of the Athi river basin is dominated by the ITCZ-related rainfall events and the major rainfall activities are close to the equator (Figure 4a). The first PC mode normally represents the mean rainfall characteristics of the basin because it accounts for the largest variance. The second PC mode has maximum loadings in the coastal region (Figure 4b), while the third PC mode is dominant in areas that the first two PC modes had least loadings, especially eastern part of Kilimanjaro (Figure 4c).

<table>
<thead>
<tr>
<th>Mode</th>
<th>Eigenvalue</th>
<th>Variance (%)</th>
<th>Cumulative Variance (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10.57</td>
<td>48.03</td>
<td>48.03</td>
</tr>
<tr>
<td>2</td>
<td>4.00</td>
<td>18.17</td>
<td>66.21</td>
</tr>
<tr>
<td>3</td>
<td>2.14</td>
<td>9.72</td>
<td>75.92</td>
</tr>
</tbody>
</table>

Table 4a. Rotated principal components modes for MAM rainfall
Figure 4a. Spatial patterns of the rotated PC mode 1 during MAM rainfall
Figure 4b. Spatial patterns of the rotated PC mode 2 during MAM rainfall
Figure 4c. Spatial patterns of the rotated third PC mode during MAM rainfall

The first PC mode can be associated with the influence of the ITCZ on the basin's rainfall while the second mode can be associated with the effect of the mesoscale system related to land/sea breeze of the Indian Ocean. The dominancy of the third PC on the eastern part of Kilimanjaro may reflect the influence of orography on rainfall in the basin.
4.4.2 Spatial variability of SON season rainfall using PCA

In this season three PC modes had eigen values greater than one and accounted for 81% of the total rainfall variance during this season. The first PC mode was dominant over northwestern part of the basin, while the second mode was dominant over the coast. The third PC mode was dominant over the northern part of the basin near Ol donyo Sabuk and eastern part of Mt. Kilimanjaro (Figures 5a, 5b and 5c).

<table>
<thead>
<tr>
<th>Mode</th>
<th>Eigenvalue</th>
<th>Variance (%)</th>
<th>Cumulative Variance (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9.20</td>
<td>41.81</td>
<td>41.81</td>
</tr>
<tr>
<td>2</td>
<td>5.24</td>
<td>23.82</td>
<td>65.63</td>
</tr>
<tr>
<td>3</td>
<td>3.34</td>
<td>15.16</td>
<td>80.78</td>
</tr>
</tbody>
</table>

Similar rainfall systems seem to be driving the rainfall patterns in this season as in the MAM season, this is reflected from the spatial patterns of the three PC modes and the mean rainfall for SON rainfall season.
Figure 5a. Spatial patterns of the rotated first PC mode for SON rainfall
Figure 5b. Spatial patterns of the rotated PC mode 2 for SON rainfall
Figure 5c. Spatial patterns of the rotated PC mode 3 for SON rainfall
4.4.3 Stability analysis of spatial PC rainfall patterns

The stability of map patterns derived from RPCA has been a question of many investigations. Methods of determining the robustness and consistency of the PC patterns have been discussed by North, et al., 1982; Richman, 1986 among others. In order to determine the stability of the stream flow patterns, inter-station correlation and split data analysis methods ((Ogallo, 1989) were used in this study.

4.4.3.1 Interstation rainfall correlations

For each season and each significant mode, the station with the highest PC loading was identified. These stations were then used as reference points in plotting inter station correlations. The inter-station correlations were obtained using simple correlation method. Spatial correlation patterns centred at the reference stations were generated and their patterns compared to those obtained by spatial patterns of PC loadings of the corresponding mode. Figures 6a, 6b and 7a, 7b, 7c show the spatial correlation maps for significant PC modes dominant during the two main rainy seasons of MAM and SON. The map patterns were quite similar to those obtained by factor loadings of significant modes. It therefore means that the PC spatial patterns are stable and can effectively be used to explain the spatial variability of seasonal rainfall in the Athi river basin.
Figure 6a. Spatial patterns of the inter station correlations with reference to Dagoretti (PC mode 1) station in the MAM rainfall
Figure 6b. Spatial patterns of the inter station correlations with reference to Rukanga (PC mode 3) station in the MAM rainfall
Figure 7a. Spatial patterns of the inter station correlations with reference to Ngong station (PC mode 1) in SON rainfall
Figure 7b. Spatial patterns of the inter station correlations with reference to Kwale station (PC mode 2) in SON rainfall
Figure 7c. Spatial patterns of the inter station correlations with reference to Kitondo (PC mode 3) stations in the SON rainfall
4.4.3.2 Rainfall data split analysis

To investigate the stability of the spatial patterns further, the seasonal rainfall data were divided into (a) even and odd observation years, (b) 1961-1981, (c) 1982-2003 and (d) Monthly rainfall. The results for the split data analysis are given in Figures 8a-8b and 9a-9b. The spatial patterns showed remarkable coherency indicating stability of the rainfall patterns when compared with the results from the previous sections.

Figure 8a. Spatial patterns of the first rotated PC mode for even years in the MAM rainfall (1961-2004)
Figure 8b. Spatial patterns of the first rotated PC mode for odd years in the MAM rainfall (1961-2004)
Figure 9a. Spatial patterns of the first rotated PC mode for even years in the SON rainfall (1961-2004)
Figure 9b. Spatial patterns of the first rotated PC mode for odd years in the SON rainfall (1961-2004).
These results show that the first three significant PC modes derived from MAM and SON rainfall seasons exhibited stability over the basin and can be used to describe the spatial variability of rainfall in the Athi river basin. Thus, the basin experiences large spatial variability of seasonal rainfall with high rainfall in the northern part, the coastal part near the Indian Ocean and the eastern parts of Mt. Kilimanjaro and Mt. Ol donyo Sabuk. Similar results were obtained by Ogallo, 1980, Beltrando, 1990, Nyenzi, 1990, Nicholson, 1996 and Indenje, et al., 2000. The first PC mode which is the dominant mode accounts for the highest rainfall in both the MAM and SON seasons. The dominant mode which is the major driving mechanism of rainfall in East Africa is associated with the zonal and meridional oscillation of the ITCZ (Ogallo, 1989; Indeje, et al., 2000). The high rainfall variance explained by PC modes during SON season shows high dependence of rainfall in this season on the major rainfall driving systems like the ITCZ, ENSO, IOD among others, hence fluctuations in the driving mechanisms affect SON rainfall more than MAM rainfall. The low variance explained by the first PC mode of MAM rainfall suggest that other factors apart from global teleconnections play a major role in modulating the ITCZ (Indeje, et al., 2000). Figure 10 gives the homogenous rainfall zones that were derived from the integrated PC patterns while the regions are discussed next.

Region I- is centred around the urban environment; Region II- Peri-urban region in the neighbourhoods of the Ngong hills; Region III- is largely semi-arid area lands with very low annual rainfall with large year to year variability;
Region IV- is close to the Indian Ocean and in the neighbourhoods of Mt. Kilimanjaro on the western part of the region; Region V- is in the North coast region; while Region VI- includes Central and south coast areas.

Figure 10. Homogenous divisions of stations derived from the seasonal rainfall RPCA patterns
The communality analysis method (averaging the MAM and SON rainfall communalities) and location of the stations were used to identify representative stations in each homogenous rainfall zones hence a minimum rainfall network was established for the Athi river basin (Table 5).

Table 5. The minimum rainfall network station for the Athi river basin

<table>
<thead>
<tr>
<th>Regions</th>
<th>Stations</th>
<th>Representative Station</th>
</tr>
</thead>
<tbody>
<tr>
<td>I and II</td>
<td>Wilson Airport, JKIA, Ngong Forest, National Agr. Lab, Karura Forest, Thika, Kigwa, Dagoretti Comer, Kikuyu</td>
<td>Wilson Airport</td>
</tr>
<tr>
<td>III</td>
<td>Mbooni, Kitondo, Machakos</td>
<td>Mbooni</td>
</tr>
<tr>
<td>IV</td>
<td>Makindu, Kibwezi, Rukanga, Voi</td>
<td>Kibwezi</td>
</tr>
<tr>
<td>V</td>
<td>Msabaha, Malindi</td>
<td>Msabaha</td>
</tr>
<tr>
<td>VI</td>
<td>Kwale, Kinango, Mombasa</td>
<td>Kwale</td>
</tr>
</tbody>
</table>

4.4.4 Results for Spatial variability of MAM season stream flow

The results of spatial characteristics for MAM and SON season stream flow are presented. Like in the case of rainfall, the three PC modes for MAM stream flow were significant explaining about 72% stream flow variance in the basin, with the first, second and third modes explaining 50%, 17.7% and 10.5% of stream flow variance in the season respectively. The spatial patterns of the
loadings of the significant MAM stream flow PC modes are quite similar to those of MAM season rainfall (Figures 11a-b).
Figure 11b. The spatial patterns of second rotated PC mode derived MAM stream flow
4.4.5 Spatial variability of SON season stream flow

During the SON season stream flow, three PC modes were significant and accounted for 79% of the total stream flow variance in the Athi river basin, with the first, second and third modes explaining 53%, 15% and 11% of stream flow variance respectively. Figure 12(a) shows SON stream flow spatial patterns for PC mode 1. The mode has positive significant loadings in the central, north, northwest and northeast part of the basins. The patterns in this mode are quite similar to those of MAM and SON rainfall patterns. The similarity could be associated to the strong correlation between SON rainfall and stream flow in the basin. Figure 12(b) shows spatial patterns for SON stream flow mode 2, which had significant negative loadings at the coastal part of the basin.
Figure 12a. The spatial patterns of rotated first PC mode derived from SON stream flow
Figure 12b. The spatial patterns of the rotated second PC mode derived from SON stream flow
4.4.6 Stability analysis of PC stream flow patterns

Like in the case of rainfall data, the stability of spatial patterns derived from PC modes were investigated using inter-station correlation and split data analysis.

4.4.6.1 Interstation stream flow correlation

Like in the case of rainfall, interstation correlations were mapped using stream flow data with reference locations centred on the stations with the highest PC mode loading (Figures 13a-b). The patterns show a remarkable coherency with the rotated PC modes patterns which is reflective of the stability of the rotated PC modes.

4.4.6.2 Stream flow data split analysis

Like in the case of rainfall split analyses were undertaken with stream flow records for (i) odd and even year stream flow data, (ii) 1960-1978 and 1979-1995 data and (iii) monthly data. The spatial patterns of the three dominant PC modes were quite close to those derived from the complete set of stream flow records which is reflective of the stability of the PC mode patterns that were derived from stream flow records. Figure 14 shows stream flow homogenous zones derived from seasonal PC mode patterns. Unlike rainfall that had six homogenous regions only three homogenous zones could be delineated from steam flow records. This could be due to interactions within some of the heterogeneous rainfall zones through stream flow linkages and the stability of
stream flow due to filtering of meteorological noise in rainfall by hydrological processes in the catchment.

Figure 13a. Spatial patterns of the inter station correlations with reference to station 3BC12 (MAM-PC mode 1)
Figure 13b. Spatial patterns of the inter station correlations with reference to station 3G03 (MAM-PC mode 2)
4.4.6.3 Homogenous Stream flow zones

Figure 14 shows the homogenous stream flow derived using rotated principal components patterns and interstation correlation patterns. Stream flow Region I- includes the upper Athi river basin and regions drained by the Nairobi, Kamiti, Stony Athi and Mbagathi rivers. The hydrology of the region is mostly influenced by topography, drainage pattern and soils. Region II-This includes the middle part of the Athi river basin with the Yatta plateau and the coastal lowlands. Hydrology of the area is determined by topography which influences the drainage pattern, the soils and the land use pattern. While Region III- consist of the south western part and coastal part of the basin. Its hydrology is influenced by topography due to Mt Kilimanjaro, Taita and Shimba hills which dictate the drainage pattern, land use patterns and soils.

4.4.6.4 Stream flow communality analysis

The communality analysis method was used to identify representative stations in each homogenous stream flow zones. Table 6 shows the minimum stream flow network for the Athi river basin.
Figure 14. Stream flow homogenous zones derived from PC patterns
Table 6. Minimum Stream flow network for Athi river basin

<table>
<thead>
<tr>
<th>Region</th>
<th>Station</th>
<th>Representative Station</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>3BA17</td>
<td>Gatamayu (3BC15)</td>
</tr>
<tr>
<td></td>
<td>3BA18</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3BB10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3BB11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3BB13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3BC09</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3BC12</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3BC13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3BC15</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3BD02</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3CB02</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>3DA02</td>
<td>Athi Mavindini (3F02)</td>
</tr>
<tr>
<td></td>
<td>3F02</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>3G03</td>
<td>Mzima (3G03)</td>
</tr>
</tbody>
</table>

It may be concluded from these results that the PC modes were able to delineate distinct homogenous zones for Athi river basin. The patterns delineated five and three homogenous regions for rainfall and stream flow respectively. Similar approaches have been used by Indenje, *et al.*, 2000; Ogallo, 1989; Basalirwa, *et al.*, 1993; Opere, 1999 and Ngala, 2000 in East Africa and Dyer, 1977 for South Africa among others. The results of this study compare well with those of Ogallo, 1989; Ngala, 2000 and Indeje, *et al* 2000. The results showed that rainfall has a major influence on the stream flow patterns over various homogenous regions. Topography, changes in land use, and other local factors were other key factors that influenced some differences in the homogenous rainfall/stream regions. These results provide useful information that can be used to strengthen rainfall and stream flow observations systems and in formulation of empirical rainfall and stream flow forecasting models for the homogenous zones in the Athi river basin. It also
helps to reduce the number of models that could be developed for climate risk reduction in the catchment including monitoring, planning, management, prediction, early warning and coping with climate variability as well as adaptation of the catchment water resources to future climate variability and changes.

4.5 Temporal characteristics of seasonal rainfall and stream flow

In this study, the graphical and Spearman rank correlation methods were used to examine trends in monthly and seasonal rainfall and stream flow in the Athi river basin. To study periodicity in rainfall and stream flow in the Athi river basin the wavelet analysis method was used.

Results for graphical methods indicated that although trends were discernible in the interannual rainfall patterns at some locations, the most significant temporal fluctuations were recurrences of high and low rainfall values in all locations in the Athi river basin (Figures 15). Some of the recurrent large/low events of rainfall seem to occur within warm and cold ENSO years (Ogallo, 1989; Indenje, 2000; Owiti, 2005; Mutemi, 2005).

The results for Spearman rank correlation method are given in Table 7. The results revealed significant rainfall trend at 95% confidence limit at Kitondo (October and MAM season rainfall), Kinango (MAM season rainfall) and Wilson (April rainfall) stations. The significant negative rainfall trend at
Wilson airport station in April rainfall is shown in figure 15. This trend could be due declining rainfall trend in the basin due to recurrence of drought events.

![Wilson Airport April rainfall](image)

**Figure 15.** Example of a rainfall time series: Wilson airport station April rainfall

The results for stream flow trend using the Spearman rank correlation method showed that there was significant stream flow trend at 95% confidence limit in SON season at 3BD02 and 3BC13 river gauging stations (**Figure 16**). Like in the case of rainfall, the most significant temporal fluctuations in stream flow were recurrences of high and low stream flow values at all gauging locations.
Figure 16. Streamflow hydrograph for Theta river at 3BD02

The results from trend analysis identified some significant trends in rainfall and stream flow at some locations. There were no homogeneity in the spatial patterns of the observed trend, and there was no clear evidence of climate change for the catchment rainfall/stream flow trends delineated from the study.
Table 7. Spearman rank correlation coefficients for rainfall

<table>
<thead>
<tr>
<th>Station</th>
<th>April rainfall</th>
<th>October rainfall</th>
<th>MAM rainfall</th>
<th>SON rainfall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wilson</td>
<td>-0.36*</td>
<td>-0.07</td>
<td>-0.12</td>
<td>0.04</td>
</tr>
<tr>
<td>JKIA</td>
<td>-0.23</td>
<td>0.02</td>
<td>-0.07</td>
<td>-0.01</td>
</tr>
<tr>
<td>Ngong</td>
<td>-0.18</td>
<td>-0.21</td>
<td>-0.15</td>
<td>-0.07</td>
</tr>
<tr>
<td>Kiambu</td>
<td>-0.17</td>
<td>-0.21</td>
<td>-0.12</td>
<td>0.03</td>
</tr>
<tr>
<td>National Ag.</td>
<td>-0.23</td>
<td>-0.10</td>
<td>-0.19</td>
<td>0.16</td>
</tr>
<tr>
<td>Karura</td>
<td>-0.18</td>
<td>-0.12</td>
<td>-0.24</td>
<td>-0.00</td>
</tr>
<tr>
<td>Thika</td>
<td>-0.18</td>
<td>-0.23</td>
<td>-0.12</td>
<td>0.00</td>
</tr>
<tr>
<td>Kigwa</td>
<td>-0.067</td>
<td>-0.19</td>
<td>-0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>Dagoret</td>
<td>-0.25</td>
<td>-0.17</td>
<td>-0.17</td>
<td>-0.01</td>
</tr>
<tr>
<td>Kikuyu</td>
<td>-0.18</td>
<td>-0.20</td>
<td>-0.16</td>
<td>0.15</td>
</tr>
<tr>
<td>Mbooni</td>
<td>-0.29</td>
<td>-0.20</td>
<td>-0.28</td>
<td>-0.14</td>
</tr>
<tr>
<td>Kitondo</td>
<td>-0.32</td>
<td>-0.34*</td>
<td>-0.39*</td>
<td>-0.20</td>
</tr>
<tr>
<td>Machakos</td>
<td>-0.04</td>
<td>0.07</td>
<td>-0.07</td>
<td>0.08</td>
</tr>
<tr>
<td>Kibwezi</td>
<td>-0.2</td>
<td>-0.17</td>
<td>-0.18</td>
<td>0.00</td>
</tr>
<tr>
<td>Makindu</td>
<td>-0.07</td>
<td>-0.22</td>
<td>-0.09</td>
<td>0.01</td>
</tr>
<tr>
<td>Msabaha</td>
<td>-0.06</td>
<td>0.01</td>
<td>0.00</td>
<td>-0.14</td>
</tr>
<tr>
<td>Malindi</td>
<td>-0.06</td>
<td>-0.19</td>
<td>-0.07</td>
<td>-0.00</td>
</tr>
<tr>
<td>Rukanga</td>
<td>0.09</td>
<td>0.01</td>
<td>-0.02</td>
<td>0.12</td>
</tr>
<tr>
<td>Voi</td>
<td>0.12</td>
<td>-0.15</td>
<td>-0.07</td>
<td>0.02</td>
</tr>
<tr>
<td>Kwale</td>
<td>0.07</td>
<td>0.05</td>
<td>0.18</td>
<td>-0.06</td>
</tr>
<tr>
<td>Kinango</td>
<td>0.2</td>
<td>-0.16</td>
<td>0.45*</td>
<td>-0.06</td>
</tr>
<tr>
<td>Mombasa</td>
<td>0.14</td>
<td>-0.19</td>
<td>0.29</td>
<td>-0.23</td>
</tr>
</tbody>
</table>

*Significant trend at 95% confidence

4.5.1 Cyclical analysis results derived from wavelet analyses

Results for rainfall and stream flow wavelet analyses showed significant peaks centered at 2-3, 5-7 and 11 years (Figures 17 and 18). The 2-3 year peak is associated with Quasi-Biennial Oscillation (QBO), the 5-7 year peak associated with ENSO while the 10-12-year peak is associated with the solar cycle and decadal variability (Ogallo, 1980; Ogallo, et al, 1994; Nicholson and Nyenzi, 1990 and Nicholson, 1996). The 2-3 year peak was common in MAM and OND season rainfall and stream flow while the 5-7 year peak was only discernible in November and OND season rainfall and stream flow. The major ENSO events of 1970, 1980 and 1997 were clearly discernible. The Solar cycle was also observed at many locations in OND season rainfall and stream flow.
(Figure 16). Similar observations have been made in East Africa rainfall by Ogallo (1980) and Indeje, et al., (2000) among others.

(*The thick black contour encloses regions greater than 95% confidence using a red noise background. The parabola shows the cone of influence.)

Figure 17. The wavelet spectrum for Mombasa November rainfall and Mean OND rainfall
Results from wavelet analyses provided evidences of some cyclical modes of rainfall and stream flow over the Athi river catchment that can provide tools for medium and long term catchment planning and management of water use activities.
4.5.2 Results for temporal variability analysis of rainfall and stream flow

The results for temporal characteristics of seasonal rainfall and stream flow derived from PCA (T-mode) in the Athi river basin are discussed in this section. Details of this method are given in section 2.4.3. The PCA T-mode solutions of the first and second principal component in each MAM and SON seasons were used to cluster together years with similar rainfall and stream flow spatial patterns in the Athi river basin (Ogallo, 1989). The wet/high flow years had large positive factor loadings while dry/low flow years had large negative factor loadings. Tables 8 and 9 give clusters of wet/high and dry/low years obtained for MAM and SON rainfall and stream flow records for the two major rainy seasons in the Athi river basin.

<table>
<thead>
<tr>
<th>Table 8. Results for T-mode PCA clusters of wet and dry years</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAM Rainfall</td>
</tr>
<tr>
<td>---------------------</td>
</tr>
<tr>
<td>SON Rainfall</td>
</tr>
</tbody>
</table>
Table 9. Results for T-mode PCA clusters of High and Low flow Years

<table>
<thead>
<tr>
<th>Stream flow</th>
<th>Clustered years</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAM Stream flow</td>
<td></td>
</tr>
<tr>
<td>SON Stream flow</td>
<td></td>
</tr>
</tbody>
</table>

It was observed from the clusters and historical records that not all wet/high and dry/low flow years were included in Tables 8 and 9. This indicates that the first and second principal component modes could not effectively cluster all the wet/high flow and dry/low flow years in the period 1960-2005. This could be due to large spatial variability observed in East African rainfall (Ogallo, 1989). However, many of the anomalous rainfall and stream flow years clustered together are representation of some of the major floods/droughts years in the region some of which associated with different ENSO phases. A few of the wet years which could not be reflected in the stream flow analysis correspond to years when localised rainfall extremes were observed over selected parts of the basin. This is very common in areas, which are dominated highly convective rainfall system like the Athi river basin and East Africa at large.

Figures 19–22 show examples of rainfall and stream flow anomalies during the wet/high and dry/low flow years in the Athi river basin. The spatial map
patterns show distinct differences during the wet/high flow and dry/low flow seasons. During the MAM season negative peak anomalies are found near the Indian Ocean part of the basin for the dry years while wet year clusters had peak positive anomalies in the upper part of the Athi river basin and with peak negative anomalies in the lower part of the basin near the Indian Ocean. This pattern could be due to the influence and fluctuations of the zonal arm of the ITCZ and monsoonal wind systems over the western part of Indian Ocean (Ogallo, 1989).

During the SON season for both low and high flows peak negative rainfall anomalies are well distributed all over the basin while peak positive anomalies are also well spread over the basin (Figures 21-22). This pattern could be due to the influence of SON rainfall and stream flow by the weakening and strengthening of the monsoonal wind. Similar spatial patterns have been observed by Ogallo(1989).
Figure 19. Rainfall anomalies during the dry cluster of MAM 1970

Figure 20. Rainfall anomalies during the wet cluster of MAM 1997
Figure 21. Stream flow anomalies during the low flow cluster of SON 1984

Figure 22. Stream flow anomalies during the high flow cluster of SON 1981
The results of this study obtained from PCA T-mode and composite analyses were able to delineate most of the past extreme wet/high flow, and dry/low flow events experienced in the Athi river basin. These results can be used in mapping various climate and stream flow risks zones in the basin.

Observed also in this study includes unique El Niño modes which are associated with reversal impacts of El Niño over some locations, thus droughts are observed instead of the traditional floods (Ashok, et al., 2007). This types of El Niño is referred to by some scientist El Niño ‘Modoki’ during which a unique concentration of warming over central Pacific equatorial basin during some El Niño events (Ashok, et al., 2007).

4.6 Results for rainfall and stream flow variability and teleconnections

This section presents results for spatial variability of rainfall and stream flow during ENSO years, evolution of ENSO during the short rainfall season of September-October-November-December (SOND) and the rainfall/stream flow global teleconnections. The spatial variability was investigated using 24 month rainfall and stream flow composites and Principal Component Analysis (PCA), the evolution of ENSO during SOND season was investigated using spatial patterns of rainfall and stream flow composites during the pre-ENSO, ENSO and post ENSO years while the strength of the teleconnections was studied using linear correlation analysis. Studies, which have used similar methods, included those by Ropelewski and Halpert (1986), Kahya and Dracup (1993)
and Chiew, et al., (1998). Details of the methodology are discussed in sections 2.3.1, 2.5 and 2.5.1.

4.6.1 Results for spatial variability of rainfall and stream flow during El Niño

The results for rainfall variability using PCA extracted two significant Principal Components (PC) modes. The modes (PC1 and PC2) accounted for 46% and 23% of the total rainfall variance respectively (Figures 23a and 23b). The few PC modes extracted could be due to large spatial coherence and homogeneity of rainfall during ENSO years as observed in previous studies (Ogallo, 1988; Phillips and McIntyre, 2000; Indeje, et al., 2000; Mutemi, 2005). The first mode is significant in the northern part of the basin while the second mode is significant in the central and coastal parts of the basin. The first mode patterns could be associated with the amplification of ITCZ while the second mode spatial patterns could be associated with the amplification of the mesoscale (land/sea breeze) systems, during El Niño.
Figure 23a. Spatial patterns for PC mode 1 (El Niño rainfall composite)
Figure 23b. Spatial patterns for PC mode 2 (El Niño rainfall composite)
The results for stream flow variability study using PCA during El Niño extracted three significant PC modes which accounted for approximately 68% of the total stream flow variance in the Athi river basin (Figures 24(a) and 24(b)). The spatial patterns for stream flow mode 1 show that the mode is significant in the northern part of the basin, the spatial patterns for mode 2 show that it is significant in the middle part of the basin, while mode 3 is significant in the lower part of the basin near the Indian Ocean. The dominancy of mode 1 in the northern part of the basin could be associated with amplification of the ITCZ during El Niño while the dominancy of modes 2 and 3 could be associated with the amplification of micro and mesoscale circulation systems, during the same period.

In order to study the stability of the PC modes for rainfall and stream flow composites spatial patterns during El Niño, inter-station rainfall and stream flow data correlations were used. The results showed remarkable coherency with PC spatial patterns. This shows that the PC modes are stable hence they can be used to explain the spatial variability of rainfall and stream flow during El Niño in the Athi river basin.
Figure 24a. Spatial patterns for PC mode 1 (El Niño stream flow)
Figure 24b. Spatial patterns for PC mode 2 (El Niño stream flow composites)
4.6.2 Results for variability of rainfall and stream flow during La Niña

The results for rainfall PCA extracted three significant modes. The first mode (PC 1) accounted for 60.1% the second mode (PC 2) accounted for 14.6% and the third mode (PC 3) accounted for 8.1%. The three principal component modes accounted for a total of 83.6% of the total rainfall variance during La Niña. The high total rainfall variance accounted by PC modes during La Niña rainfall shows the influence of ENSO on rainfall variability in the Athi river basin and could be associated with high spatial coherence and homogeneity of rainfall during ENSO.

Figures 25(a)-25(c) show the spatial patterns for PC modes 1-3 during La Niña rainfall. PC mode 1 is significant in the northern part of the basin. The proximity to the equator indicates that this mode could be associated with the influence of the meridional oscillation of the ITCZ. PC mode 2 is only significant in the central part of the basin while mode 3 is significant in the central part of the basin and along the coastal strip near the Indian Ocean. Modes 2 and 3 could be associated with the modulation of orographic and land/sea breeze circulation systems during ENSO.

The PCA results for stream flow composites during La Niña extracted three PC modes. The first mode (PC1) accounted for 53%, the second mode (PC2) accounted for 12% while the third mode (PC3) factor accounted for 9% of the stream flow variance. The three modes accounted for approximately 74% of
the total stream flow variance during La Niña period. **Figures 26(a), 26(b) and 26(c)** show the spatial patterns of PC modes 1-3. Like for rainfall during La Niña, the first mode had significant positive factor loading in the middle and northern part of the Athi river basin. This mode could be associated with the modulation of the ITCZ during La Niña. PC mode 2 had significant positive loadings in the northern part of the basin near Ol Donyo Sabuk. This mode could be associated with modulation of orographic effect during La Niña. PC mode 3 has significant negative factor loadings on the southern part of the basin and along the coastal strip. This mode could be associated with modulation of the land sea meso scale circulation system during La Niña.

The stability of rainfall and stream flow spatial patterns during La Niña was investigated using the inter stations correlations coefficient by choosing the station with the largest PC loading as the reference station and split data analysis. The inter station correlation coefficients were mapped to produce spatial patterns. The results showed a remarked coherency with PC spatial patterns, which shows that the PC patterns are stable hence suitable to explain the rainfall and stream flow patterns during La Niña.
Figure 25a. Spatial patterns of the first PC mode during La Niña rainfall
Figure 25b. Spatial patterns of PC mode 2 during La Niña rainfall

Figure 25c. Spatial patterns of PC mode 3 during La Niña rainfall
Figure 26a. Spatial patterns PC mode 1 during La Niña stream flow

Figure 26b. Spatial patterns PC mode 2 during La Niña stream flow
In this section, the results of the study of the seasonal evolutions of El Niño and La Niña events before and after SOND months are presented. Thus the composite of rainfall and stream flow patterns in the previous and preceding months to the peak rainfall periods were investigated. For example rainfall and stream flow conditions in JJA and DJF seasons were integrated when interpreting OND patterns thus the rainfall and stream flow patterns before and after OND period were considered. This integrated approach was used in order to incorporate soil moisture, flow conditions, general vulnerability to flood and drought risks together with other associated water needs in the basin.

Figure 26c. Spatial patterns PC mode 3 during La Niña stream flow

4.6.3 Evolution of ENSO during SOND season
Figures 27-31 show examples of the spatial patterns of SOND rainfall composites patterns during El Niño and La Niña. The composite patterns show that during El Niño year, JJA season rainfall is enhanced and wide spread throughout the basin and depressed during La Niña (figure 27). A similar pattern is repeated during SON and OND seasons where rainfall activities are greatly enhanced during El Niño and depressed during the La Niña (figures 28-29). The same pattern continues till the month of February of the post ENSO event (figure 30).

These results show that there is increased rainfall amount and spread during most El Niño years while there is depression of the short rains during most La Niña years. Similar results have been obtained by Ogallo (1988); Indenje, et al., (2000); Mutai, et al., (1999) and Mutemi (2005). However, reversed signals have been observed in some ENSO events where regions which normally receive surplus rainfall during El Niño receive depressed rainfall. This has been given a special name El Niño ‘Modoki’ (Ashok, et al., 2007; 2009).

The Athi basin receives 550 mm of rainfall per year hence enhanced rainfall and stream flow during some of these El Niño events can be harvested for both domestic and farming purposes hence mitigating the impact of La Niña in the basin.
Figure 27. The composite spatial patterns of JJA rainfall during year (0) of (a) El Niño (b) La Niña, X-axis is Longitude and Y-axis Latitude.
Figure 28. The composite spatial patterns for SON rainfall during year (0) of (a) El Niño (b) La Niña
Figure 29. The composite spatial patterns for OND rainfall during year (0) of (a) El Niño (b) La Niña
**Figure 30.** The composite spatial patterns for DJF rainfall during year (+) of (a) El Niño (b) La Niña.
The stream flow patterns show that like in the rainfall case stream flow is enhanced during most El Niño events and depressed during most La Niña events (Figure 31). The stream flow patterns during the JJA, SON and DJF seasons of the ENSO event years were closely influenced by observed seasonal rainfall patterns, with some time lags. The delayed response to ENSO in stream flow could be associated to effect of basin characteristics which include land use, shape, size of sub catchments, evaporation, hydraulic conductivity, topography and geology.
Figure 31. The composite spatial patterns for OND stream flow during year (+) of (a) El Niño (b) La Niña.
4.6.4 ENSO and rainfall/stream flow teleconnection

In this section the results for seasonal rainfall/ENSO and seasonal stream flow/ENSO teleconnections during ENSO phases are presented. To quantify the strength of the teleconnections during ENSO, lag correlations of the linear correlation between rainfall, stream flow, SOI, Niño 4 and Niño 3 SSTs were calculated. The correlations were calculated using 2-month and 3-month running means of ENSO indicators.

The results for SOI and rainfall correlation during El Niño show that the wet period in October-January is significantly correlated at 95% confidence limit with December-January–February average SOI at $r = 0.834$ (lag 7) and June-July average SOI at $r = -0.849$ (lag 2). The results also indicated significant correlations between July-February stream flow during El Niño and SOI in the Athi river basin stream flows. The highest significant lag correlation was between September-October average SOI and July-October stream flow which was about -0.9 with time lags extending as far back as 9 time lags. Similarly, results for stream flow and SOI correlation during La Niña indicated that there was significant correlation between February-March SOI and May-June stream flow. The correlation value was as high as 0.97 which was significant at 99% confidence level. Similar studies have shown there is a strong correlation between ENSO and rainfall/stream flow (Kahya and Dracup, 1994; Gutierrez and Dracup, 2001; Zubair, 2003) with lag correlation of up to 9 months. The
ENSO-rainfall/stream flow relationship can be used in rainfall and stream flow forecasting (Chiew, et al., 1999; Piechota, et al, 2000; Chiew, et al., 2002).

4.7 Results for rainfall and stream flow spatial variability during IOD


4.7.1 Results for rainfall and stream flow spatial analysis during positive IOD

Using 24-month rainfall composites and S-mode PCA, three Principal component (PC) modes were extracted using the Kaiser criterion. The three PCA modes accounted for a total of 63% of the total rainfall variance. For stream flow, two PCA modes were extracted which accounted for 68% of the total stream flow variance. The high percentage of rainfall and stream flow variance accounted by PC modes can be probably be associated to the impacts of IOD on rainfall systems and water resources variability in the Athi river basin. Linkages between IOD and East African rainfall have been discussed in many past studies including that of Gitau (2005); Owiti (2005); Vaidya (2005) and Bamanya (2007).
Figure 32 shows the spatial patterns of rainfall PC mode 1 during positive IOD events. The patterns show high PC loading of $\geq 0.5$ in the northern part of the basin for mode 1 while the spatial patterns for rainfall PC mode 2 shows high loading of $\geq 0.55$ in the central and lower part of the basin with peak loadings along the coastal strip (figure 33). Similar spatial patterns are depicted by stream flow PC modes (figures 34 and 35). The stability of the rainfall and stream flow PC mode patterns during IOD were investigated using interstation data correlation analysis. The results showed that the PC spatial patterns are stable and can be used to explain rainfall and stream flow spatial variability in the Athi river basin during IOD. The spatial patterns for PC mode 1 for rainfall and stream flow could be associated with amplification of the regional rainfall processes during a positive IOD event. PC mode 2 patterns could be associated with meso scale processes including topography, and sea-land breeze, together with other processes that have not been captured by the first mode.
Figure 32. Spatial patterns of PC mode 1 during rainfall Positive IOD

Figure 33. Spatial patterns of PC mode 2 during rainfall Positive IOD
Figure 34. Spatial patterns of PC mode 1 during Stream flow Positive IOD

Figure 35. Spatial patterns of PC mode 2 during Stream flow Positive IOD
4.7.2 Results for rainfall and stream flow spatial analysis during negative IOD

Figures 36 and 37 shows spatial patterns of PCA modes for rainfall in the Athi river basin derived from 24-month rainfall composites during the negative phase of IOD. The rainfall PCA modes show high factor loadings of $\geq 0.5$ in the northern part of the basin for mode 1 while mode 2 shows high factor loading $\geq 0.7$ in the central part of the basin. Similar patterns are depicted by stream flow however, the patterns are not well defined like in rainfall. This kind of pattern could be associated with the influence of topography, soils and land use patterns on stream flow. These patterns show the influence of the large scale systems such the ITCZ and their interactions during the negative phase of IOD. In general the spatial patterns obtained during negative IOD were opposite to those observed with positive IOD.
Figure 36. Spatial patterns of PC mode 1 during rainfall negative IOD
Figure 37. Spatial patterns of PC mode 2 during rainfall negative IOD
4.7.3 Results for the study of evolution of IOD during MAM and OND

The evolution of positive IOD events in the Athi river basin were studied using 24-month composites of MAM and OND seasonal rainfall and stream flow. The composites were abbreviated as MAM (-), OND (-), MAM (0), JJA (0), SON (0), OND (0) and DJF (+) where (-) is the year preceding, (0) the peak year and (+) the year following the peak IOD event.

Figure 38a shows rainfall spatial patterns corresponding to MAM (-) composite during the positive IOD event while figure 38b shows rainfall spatial patterns corresponding to MAM (-) composite during the negative IOD event. The composite spatial patterns during the positive event show wet anomalies on the eastern part of the basin with negative anomalies to the western parts of the region. The rainfall spatial patterns during negative IOD show reduced wet conditions and enhanced dry conditions in the basin. Similarly, the basin experiences enhanced wet conditions during positive IOD and dry conditions during negative IOD events during the SOND season (Figures 39-40). The stream flow spatial patterns closely resemble those of rainfall with enhanced flow conditions during positive IOD and depressed flows (enhanced low flows) during negative IOD in the basin (Figure 41).

This study has delicately delineated seasonal evolutions of the catchment rainfall and stream flow patterns during extreme wet and dry conditions during IOD phases, which can be used in planning, management, operation and
development of various water projects in the catchment including hydropower and irrigation applications.
Figure 38. Spatial patterns of MAM (-) rainfall composite during (a) Positive IOD (b) Negative IOD. X axis shows latitudes and Y-axis Longitudes
Figure 39. Spatial patterns of SON (0) rainfall composite during (a) Positive IOD (b) Negative IOD
Figure 40. Spatial patterns of OND (0) rainfall composite during (a) Positive IOD (b) Negative IOD
Figure 41. Spatial patterns of OND (0) Stream flow composite during (a) Positive IOD (b) Negative IOD
4.8 Results for Canonical Correlation Analysis

This section presents the results obtained by canonical correlation analysis method for the predictability of MAM and OND seasons rainfall and stream flow for the Athi river basin using Indian Ocean SST. The details of this method are outlined in section 2.5.3.

4.8.1 Results for CCA models for MAM rainfall and stream flow

Figure 42 shows the spatial patterns for the DJF Indian Ocean SST CCA mode 1 for prediction of MAM rainfall over the Athi river basin. Mode 1 CCA has positive loading over most of the Indian Ocean with some negative loading over southern western parts extending to near the western part of Australia. The time series of the rainfall and SST coefficients derived from CCA mode 1 is given in figure 42c. The correlation coefficient between the two time series is about 0.47. It is evident from the time series that the extremely large rainfall anomalies were not well simulated by the first CCA mode. It was noted that when only January Indian Ocean SST was used as a predictor for MAM rainfall the CCA coefficient for mode 1 increased to as high as 0.6. The strong time lagged correlation between Indian Ocean SST and MAM rainfall in the Athi river basin can be used to provide seasonal rainfall prediction in the basin. These results could be useful in planning, management and operation of water resources systems in the basin. Similar patterns have been reported by Indenje (2000) and Mutemi (2005).
**Figure 43** shows the results that were obtained when using DJF Indian Ocean SST CCA mode 1 as a predictor of MAM stream flow with a correlation coefficient of 0.58. The patterns show negative correlations in the northern part of Indian Ocean and north west part of the Athi river basin. Positive correlations are located in the part of the Indian Ocean and in the middle and lower parts of the Athi river basin. The temporal scores of stream flow and SST show a coherent trend. Similar patterns are depicted by CCA modes for MAM stream flow and using January SST as a predictor. However, the CCA coefficient for mode 1 was 0.37. The consistency of the patterns shows the stability of the relationship between DJF SST and MAM stream flow in the Athi river basin.
Figure 42. The spatial patterns of CCA mode 1 for the correlation between (a) seasonal DJF Indian SST and MAM rainfall (b) rainfall and canonical vector (v) and (c) normalised temporal functions for CCA mode 1 patterns for rainfall and SST.
Figure 43. The spatial patterns of CCA mode 1 for the correlation between (a) seasonal DJF Indian SST and MAM stream flow (b) stream flow and canonical vector (v) and (c) normalised temporal functions for CCA mode 1 patterns for stream flow and SST.
4.8.2 Results for CCA models for OND rainfall and stream flow

In the predictability of OND rainfall in the Athi river basin using JJA SST as a predictor only one significant CCA mode was obtained. The CCA coefficient for the mode 1 was 0.4(Figure 44). In this mode the Indian Ocean dipole pattern is clearly discernible with a positive centre near the East African coast and a negative centre in western and southwestern regions.

The stability of the predictability of OND rainfall using JJA Indian Ocean SST was further investigated using July SST and only one CCA mode was significant. The patterns were also coherent with those of JJA SST. The Indian dipole is more clearly discernible. Similar observations have been made by Landman and Mason (1999) and Omondi (2005).

OND stream flow CCA patterns were similar to those for rainfall (Figure 45). The strong correlation indicates that there is a strong relationship between JJA Indian Ocean SST and OND stream flow in some parts of the Athi river basin. The results from the study also revealed that the CCA correlation values were higher for stream flow records compared to the actual rainfall. These could due to some smoothing effects of catchment rainfall when stream flow data are used. The significant lagged correlations indicate a potential of using Indian Ocean SST in prediction of OND season rainfall and stream flow in the Athi river basin. These
results like in the case of MAM season rainfall and stream flow could be used for planning, management and operation of water resources systems in the Athi river basin.

(a) CCA-1 JJA SST

(b) CCA-1 OND Rainfall

Figure 44. The spatial patterns of CCA mode 1 for the correlation between (a) seasonal JJA Indian SST and OND rainfall (b) rainfall and canonical vector (v) and (c) normalised temporal functions for CCA mode 1 patterns for rainfall and SST.
Figure 45. The spatial patterns of CCA mode 1 for the correlation between (a) seasonal JJA Indian SST and OND stream flow (b) stream flow and canonical vector (v) and (c) normalised temporal functions for CCA mode 1 patterns for stream flow and SST.
4.9 Results for Multiple regression and NSFM models

In this section results obtained using multiple linear regression and the non-parametric seasonal forecast models are given and discussed.

4.9.1 Results for seasonal rainfall and stream flow regression analysis

The independent variables (predictors) for the multiple linear regression models used were Niño 3.4 index, IOD, and SOI at various lags. These variables were introduced step wise into the regression model due to some potential interrelationships among some of the predictors. Table 10 shows results for long range prediction potential for both rainfall and stream flow in the basin.

Table 10. Regression models for Rainfall and Stream flow forecasting

<table>
<thead>
<tr>
<th>Region</th>
<th>Predictand</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>OND Rain</td>
<td>-0.5IOD_May+0.3IOD_July-0.2SOI_Sept</td>
</tr>
<tr>
<td>3</td>
<td>OND Rain</td>
<td>-0.5IOD_Apri+0.3SOI_May-0.4SOI_Jul</td>
</tr>
<tr>
<td>1</td>
<td>OND Flow</td>
<td>0.5IOD_March+0.2IOD_Jul+0.1SOI_Feb-0.4SOI_March+0.5SOI_April+0.5Niño3.4_June</td>
</tr>
</tbody>
</table>

Figures 46-48 give some examples of the results that were obtained from multiple linear regression models. The Figures show that the models gave realistic fits of
both rainfall and stream flows. Some of the extreme rainfall and stream flow values were however under estimated or not well captured by the models. This may be a reflection that there others factors that could be driving the catchments rainfall and stream flow apart from the predictors used in the study. Such predictors whose effects might have not been reflected in the study include the complex regional systems, ITCZ, tropical cyclones activities, monsoonal winds, Maden Julian Oscillation (MJO), solar cycles among many others. The results from the study however show that multiple linear regression models can provide realistic OND rainfall and stream flow outlooks that can be used for long range forecasts for early warning, planning, management and operation water resources systems in the Athi river basin.
Figure 46. Region VI rainfall forecast at Mombasa
Figure 47. Region IV rainfall forecast at Kibwezi
Figure 48. Region I stream flow forecast at 3BC15

4.9.2 Results for Non-parametric Seasonal Forecasting Model (NSFM)

This section presents results for the Non-parametric Seasonal Forecasting Model (NSFM) that were based on the skill of providing forecasts for grouped categories of normal, above normal and below normal occurrences of the hydrological
variables in the Athi river basin. The two parameters that were used to assess the skill of the NSFM models were the Nash–Sutcliff coefficient of efficiency (E) modified linear error in probability space (LEPS) score. The independent variables (predictors) were still the same like in the case of the multiple linear regression models.

Examples of the results that were obtained from the study are given in Table 11. The table shows that best forecasting results were obtained when July–August IOD was used as a predictor in forecasting SON rainfall in region VI.

**Table 11. Results for forecasting SON rainfall using NSFM**

<table>
<thead>
<tr>
<th>Lead Time</th>
<th>Correlation With SON rainfall</th>
<th>Calibration</th>
<th>Verification</th>
<th>Region</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>E</td>
<td>LEPS</td>
<td>E</td>
</tr>
<tr>
<td>June-July Niño</td>
<td>0.4</td>
<td>0.11</td>
<td>11.1</td>
<td>-0.03</td>
</tr>
<tr>
<td>3.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>July-Aug IOD</td>
<td>0.4</td>
<td>0.37</td>
<td>23</td>
<td>0.04</td>
</tr>
<tr>
<td>June-July IOD</td>
<td>0.44</td>
<td>0.11</td>
<td>5.8</td>
<td>-0.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Table 12 provides other results that were obtained by using NSFM when the predictor values were extended to include August data for the prediction of September–November rainfall and stream flow. The results showed that there is an improvement in prediction skills as the predictor time lags reduces. This poses a*
challenge in that predictions are not available on time and the use of such options would require the predictors to be predicted first. This would create further errors and uncertainties in the forecasts derived.
Table 12. Results for SON stream flow forecasting using NSFM

<table>
<thead>
<tr>
<th>Lead time</th>
<th>Stream flow Region</th>
<th>Predictor</th>
<th>Correlation</th>
<th>Calibration</th>
<th>Verification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>E</td>
<td>LEPS</td>
<td>E</td>
</tr>
<tr>
<td>June-July</td>
<td>II</td>
<td>IOD</td>
<td>0.38</td>
<td>0.06</td>
<td>5.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>June-July Flow</td>
<td>0.02</td>
<td>7.6</td>
<td>-0.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IOD+ June-July Flow</td>
<td>0.38</td>
<td>18.06</td>
<td>-0.24</td>
</tr>
<tr>
<td>July-August</td>
<td>I</td>
<td>July-August Flow</td>
<td>0.4</td>
<td>0.01</td>
<td>16.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IOD</td>
<td>0.11</td>
<td>5.79</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IOD+ July-August Flow</td>
<td>0.46</td>
<td>25.7</td>
<td>-0.02</td>
</tr>
<tr>
<td>Jan-Feb-March Flow</td>
<td>III</td>
<td>Jan-Feb-March Flow</td>
<td>0.5</td>
<td>0.4</td>
<td>41</td>
</tr>
</tbody>
</table>
Table 13. Examples of forecast categorization (stream flow=2.5 m$^3$/s and IOD=1.2)

<table>
<thead>
<tr>
<th>Category</th>
<th>Flow</th>
<th>IOD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Above normal</td>
<td>1.86</td>
<td>0.91</td>
</tr>
<tr>
<td>Normal</td>
<td>1.49</td>
<td>0.38</td>
</tr>
<tr>
<td>Below normal</td>
<td>1.29</td>
<td>-1.15</td>
</tr>
</tbody>
</table>

Table 13 shows a tercile categorization of stream flow using NSFM. The results obtained for forecasting above normal, normal and below normal SON flows are given in figures 49-51. The results show that the model over estimates above normal flow, under estimates normal flow and forecasts fairly well below normal flow.
Figure 49. Stream flow Region I SON stream flow forecast (Above Normal)
Figure 50. Stream flow Region I SON stream flow forecast (Normal flow)
Figure 51. Stream flow Region I SON stream flow forecast (Below Normal flow)

From these results it may be concluded that although realistic SON rainfall and stream flow forecasts can be obtained from time lagged ENSO, IOD, and other
related predictors using linear multiple regression and NSFM approaches, not all extremes in catchment rainfall and stream flow could be derived from such models. The use of other complex models which incorporate the catchment rainfall and stream flow characteristics are therefore recommended for future studies.
CHAPTER FIVE

5.0 SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

This section provides the summary of the major results and conclusions from the study. It also highlights some recommendations from the study.

5.1 Introduction

The details of the results of the study were presented and discussed in chapter four, only a brief summary is presented in this section.

5.2 Summary

Data quality tests for this study using the runs test showed that most data was homogenous apart from the rainfall stations at Msambweni District Office and Kibwezi Dwa plantation rainfall stations and river gauging stations at Athi–Munyu (3DA02), Ruiru(3BC07) and Theta (3BD02) which showed significant heterogeneity at 95% confidence limit.

The significant S-mode PCA and communality analysis results were used to delimit the Athi basin into six-rainfall homogenous zones and three stream flow homogenous zones. The homogenous rainfall zones delimited closely resembled the climatologically zones of the basin. The stream flow homogenous zones were slightly different from the homogenous rainfall zones. The results from spatial
analysis of rainfall indicated significant seasonal shifts in the patterns of the dominant principal component modes, which closely resembled the seasonal migration patterns of the ITCZ which is a major rainfall inducing factor in the region. The influence of the proximity of the Indian Ocean which is a significant mesoscale circulation system in the basin was also evident in all the seasons. Ogallo (1988), Okoola (1996) and Indeje, et al., (2000), observed similar results. The effects of other local factors like the 'heat island effect' in Nairobi and orographic effect especially those due to Mt. Kilimanjaro on the southwestern part of the basin were evident in some seasons. Similar results were depicted by stream flow, however the effect of stream flow recharge from the snow capped Mt. Kilimanjaro was pronounced during the dry seasons of December-January-February and June-July-August.

The results from T-mode rainfall and stream flow were able to cluster wet/ high flow and dry/ low flow years observed from historical records in the basin. Some of those clusters included 1997 and 1984, which are known as some of the past wettest/driest years on record within the catchment. However, not all the wet, dry, high flow and low flow years were clustered. This could be due to complex topography and other local factors like land use patterns.

The temporal characteristics of rainfall and stream flow in the Athi river basin in terms of trends and periodicity were further studied using graphical, runs test and
The results showed significant rainfall trends at Wilson airport, Kitondo, Kinango and Machakos rainfall stations. Significant stream flow trends were observed at Theta-Kiambu (3BD02) and Komothai-Kiambu (3BC13) river gauging stations. The results for rainfall and stream flow wavelet analyses showed significant peaks centered at 2-3, 5-7 and 11 years which are associated with Quasi-Biennial Oscillation (QBO), ENSO and solar cycles respectively. Ogallo, (1980) and Indeje, et al., (2000), observed similar peaks.

The results from CCA showed that the technique was able to isolate unique co-variability modes between rainfall and stream flow for the major rainfall seasons in the Athi river basin using Indian Ocean SST fields. The existence of significant positive CCA SST weights over western Indian Ocean around the East African coast extending to the Arabian sea and negative weights over the eastern extreme equatorial Indian Ocean have been associated with above normal rainfall over East Africa during MAM and SON rainfall (Ogallo, 1989; Mutemi, 2005). The results showed that there exists a potential of forecasting MAM and SON rainfall and stream flow using Indian Ocean SST.

The results obtained using composite analysis delineated linkages between various modes of El Niño, IOD and their impacts on catchment rainfall and stream flow over the Athi river basin, including cases when the warming is concentrated over
the central equatorial Pacific Ocean region, which is sometimes referred to as “El Nino Modoki”. Parts of the region have been observed to receive rainfall anomalies that are opposite to the traditional El Nino years. The results for rainfall/stream flow teleconnections showed strong linkages with El Nino Southern Oscillation (ENSO) and Indian Ocean Dipole (IOD). The teleconnections showed strong time lagged linkages extending as back as 7-9 months and lag correlation values of 0.834 to 0.899. The strong time lagged linkages among ENSO/IOD and Athi river basin rainfall and stream flow were also observed from the results of composite analysis. The composite results indicated that wet/high flow and dry/low flow conditions could be associated with ENSO/IOD phases in the Athi river basin.

Finally, an attempt was made to develop long-range forecasting models for both rainfall and stream flow based on the ENSO indicators. The results showed that multiple linear regression models could provide realistic OND rainfall and stream flow outlooks that can be used for long-range forecasts for early warning, planning, management and operation of water resources systems in the Athi river basin. The results obtained using NSFM showed that the model over estimates above normal flow, under estimates normal flow and forecasts fairly well below normal flow.
5.3 Conclusion

In this study for the first time the linkages between rainfall, stream flow and phases of ENSO and IOD were delineated. The results showed a strong influence of ENSO and IOD indices on seasonal rainfall and stream flow over the Athi river basin including cases when the warming is concentrated over the central equatorial Pacific Ocean region which is sometimes referred to as “El Nino Modoki” when parts of the East African region have been observed to receive rainfall anomalies that are opposite to the traditional El Nino years. These results could be used in the planning, operation, management and development of various water projects in the catchment including hydropower and irrigation applications.

This study delineated the Athi river basin into six homogenous rainfall zones and three homogenous stream flow zones. The PCA based areal rainfall and stream flow indices and communality analysis was further used to determine the best representative rainfall/stream flow station for each of 5 rainfall and 3 stream flow homogenous zones. Finally, a rain gauge and stream flow network was designed of 5 rainfall stations and 3 river gauging stations respectively for data collection in the Athi river basin. This information can be used to equip and manage few stations which can provide adequate data for the basin.

The canonical correlation analysis was for the first time used to delineate the regions and Indian Ocean SST modes that are associated with extremes in the Athi
river basin seasonal rainfall/stream flow extremes during March-May (MAM) and September (SON) rainfall seasons. This information is vital in mapping disaster and climate risk zones in the Athi river basin for mitigation, rescue and adaptation mechanisms.

Finally, for the first time using multiple linear regression and non-parametric seasonal forecasting models were used to develop forecasting models based on ENSO and IOD indicators for SON rainfall and stream flow for the basin. The results gave good prediction skills both at the calibration and verification stages. These findings can be used for planning, management and forecasting water resources systems in the Athi river basin.

5.4 Recommendations of the study
Finally, recommendations were made for scientists for further studies, and policy markers.

This study has attempted to design rainfall and stream flow networks for the Athi river basin using rainfall data from 1961-2004 and stream flow data from 1960-2000. There is need to design a rainfall and a stream flow network for Kenya using a longer data length. This will not only minimize the cost of running and maintenance of data networks but also ensure efficient and effective data collection.
In order to understand the rainfall-stream flow response in a basin the land use patterns are very important. In this study, a linear relationship between rainfall/stream flow and ENSO/IOD indicators was assumed. It is important future studies to explore a non-linear rainfall-stream flow relationship that takes into account aspects like evaporation and land use.

The influence of global SST on rainfall in East Africa is well documented (Ogallo, 1988, Philips and McIntyre, 2000, etc) however; the influence of global SST on stream flow has not been well studied. This study attempted to study the influence of Indian Ocean SST on Athi river rainfall and stream flow using CCA. There is need to study the influence of Atlantic and Pacific Ocean SST on stream flow in East Africa.

The impact of climate variability and change is a global concern. This study recommends that a study should be done to explore the impact of climate change not only on rainfall but also on water security in Kenya.

There is need to formulate a policy for data collection in Kenya. The policy should look at design of data collection networks and use of modern technologies in data collection. The government should provide funds for purchase and installation of Radar and Satellite systems for data collection. Funds should also be set aside for capacity building and acquiring of research resources.
The public should be made aware on how to use this information for optimum utilization of water resources. The City and Municipal councils should use the long-range prediction forecasts for management of water supply systems and repair and unclogging of drainage systems to avoid destruction of property during floods. Kenya electricity generating company (Kengen) should use long-range forecast information for management of hydropower generation reservoirs and setting up of emergency power supply systems during La Niña. The Ministry of Agriculture should educate the public on the use of conservation techniques for storage of water during Normal or El Niño years for use during La Niña years and set up a disaster response unit to handle extreme weather event disasters.
6.0 REFERENCES


