INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®

University of Alberta

The Analysis and Prediction of Droughts in East Africa.

by

Henry Kayondo Ntale 

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

in

Water Resources Engineering

Department of Civil and Environmental Engineering.

Edmonton, Alberta

Fall 2001

I dedicate this thesis to
my son Jonathan, my daughters
and my wife Tuma.

Abstract

In this investigation on East African meteorological droughts, drought properties and patterns have been identified, three drought indices tested and modified, and two statistical teleconnection models developed for predicting its seasonal rainfall totals.

Using harmonic analysis, East Africa was delineated into 6 homogeneous rainfall zones, and the important rainfall seasons (in terms of % rainfall contribution to the annual rainfall) identified for each zone. Three drought indices (Palmer Drought Severity Index or PDSI, Bhalme Mooley Index or BMI, and Standardized Precipitation Index or SPI) were analyzed, modified where necessary, and compared in terms of their consistency in detecting the initiation, evolution, severity and termination of meteorological droughts in East Africa. It seems that SPI is more versatile and consistent than PDSI and BMI in tracking East African droughts. From 6-month and 12-month SPI data, East Africa was delineated into 7 drought homogeneous zones whose spatial boundaries bear a resemblance to the 6 homogeneous zones identified from harmonic analysis. From composites developed out of 22 El Niño and 13 La Niña events and 6-month SPI data, El Niño–Southern Oscillation (ENSO) has been found to exert an influence on the moisture regime of East Africa. The degree and temporal patterns of ENSO response vary between the above-identified droughts zones of East Africa. It seems northeastern Tanzania has the strongest response to ENSO. El Niño seems to exert a stronger influence on East Africa than La Niña. Two statistical models, combined Canonical Correlation Analysis-Simplex (CCA-Simplex) and projection pursuit regression (PPR)

models were developed to predict East African seasonal rainfall. CCA-Simplex is a linear model while PPR can model nonlinear associations. PPR performed better than the stand-alone CCA. By adjusting the prediction fields with 24 weights optimally determined by the Simplex algorithm, we found CCA-Simplex to consistently produce better forecasts than using un-weighted predictor fields. PPR-Simplex did not yield better results than the stand-alone PPR.

Acknowledgement

I am grateful to the Canadian Commonwealth Scholarship Program for having granted me a scholarship to undertake graduate studies in Canada. I am grateful to my supervisor, Dr. Thian Y. Gan for his guidance. I wish also to thank the other committee members especially Dr. Gehard Reuter and Dr. Andrew Bush, whose valuable suggestions and helpful advice improved the quality of this work.

I also thank my colleague, Dr. Joseph Mmbaga, for his assistance in getting me acquainted with the Supercomputer facility at the University of Alberta and also sharing his knowledge of UNIX and FORTRAN 90 with me. I acknowledge the support he and Oscar Kalinga extended to me during the final proof reading and editing of this thesis.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Overview

Since time untold, drought has affected mankind's livelihood worldwide, sometimes at catastrophic scale. Even though drought is a hydrologic extreme, it is closely related to our climate and its recurrence is beyond human control. There is probably no one single region in this world where drought has not affected man's livelihood at one time or the other.

An example of a disastrous drought in recent decades is the Sahelian drought, which had wide scale impact on Western Africa. This prolonged drought attracted the attention of many scientists. As a result, there is plenty of documentation on this drought and the general rainfall variability of the Sahel region (e.g., Nicholson, 1979, 1980, 1983; Sivakumar, 1991 and HAPEX, 1997).

Severe droughts have also occurred in some areas of East Africa. Table 2.1 lists some major drought events that have occurred in East Africa in the Twentieth century. Despite the recurrence of drought in East Africa, only a limited number of studies have been devoted to the understanding of the nature and origins of drought in this region (e.g. Trewartha, 1981 and Beltrando, 1990). The economy of East Africa largely depends on rain-fed agriculture and as such it is susceptible to variations in the amounts, distribution and timing of rainfall. Adverse effects brought about by prolonged periods of low or no rainfall could be mitigated if the rainfall deficits are monitored and advice to relevant authorities given on a timely basis.

Since time immemorial, human settlements have predominantly been in river valleys and as a result, attention has been more directed to flooding than to drought problems. Floods tend to capture more attention than droughts because the former is more dramatic, more sudden, and often bring about massive damage in a short period. Even then, in

quantitative terms, the cumulative effects of drought to some regions (e.g. West Africa and Northeast Africa) could be more catastrophic than flood impacts.

1.1 Statement of the Problem and Research Objectives

Given that only a limited number of studies have been conducted on droughts of East Africa, even though it is a recurrent problem, the theme of this research is about the statistical properties and physical mechanisms of meteorological droughts in East Africa and the development of two statistical-teleconnection models to predict seasonal droughts in the region.

Essentially, this investigation has five major objectives:

(1) Identify the major zones of homogeneous rainfall regimes for East Africa by harmonic analysis and investigate the seasonal rainfall of these rainfall zones.
(2) Determine the most suitable drought index to detect the initiation, evolution, termination and severity of meteorological drought in Eastern Africa. Then, based on the spatial and temporal characteristics of the selected drought index, classify East Africa into homogeneous drought zones.
(3) Investigate the possible influence of El Niño Southern Oscillation (ENSO) on East African moisture anomalies.
(4) Develop linear and nonlinear statistical predictive models for predicting East African seasonal rainfall.
(5) Investigate the underlying mechanisms behind the drought teleconnection patterns.

1.2 Significance of the study

Zoning East Africa into homogenous rainfall sub-regions is an important step in understanding the characteristics of East African meteorological drought. Each zone is assumed to have a unique ensemble of rainfall causative mechanisms and a unique

teleconnection structure with some large-scale climatic signals. It is therefore necessary to divide East Africa into homogeneous zones before investigating how the precipitation of these zones teleconnect with global climate signals.

Research on drought indices should lead to practical applications. Local, regional and national agencies dealing with the planning and management of drought control and the formulation of both short- and long-range drought mitigation policies may use the selected drought index to specify the severity of the drought and to estimate the expected social and economic damages over time. The drought index can also be used to indicate the onset, progression and end of droughts, and appropriate drought mitigation procedures.

There is yet to be a set of drought indices adequately tested in East Africa. Indexing of droughts in East Africa has been and is still sketchy, without a well thought out approach. There is yet a roster of drought indices systematically used to rank the severity of drought events in East Africa. Descriptions of drought events in the region are still ad-hoc. It would be beneficial to implement drought mitigation strategies with respect to drought indices of quantitative nature rather than subjective decisions of some individuals. Conversely, drought indices enable us to objectively implement various drought control strategies and measures.

If more accurate zoning of East Africa into regions of homogeneous drought characteristics can be achieved via an appropriate drought index, policy makers would be able to use the information for managing an ongoing droughts and for planning strategies against potential droughts. Further, accurate zoning is important because at a particular time, certain zones of the region may be experiencing droughts while other zones are not. With this knowledge, plans to transfer resources from the well-endowed zones to the drought prone areas could be charted out well in advance.

Results from a teleconnection investigation between East African drought and El Niño and Southern Oscillation (ENSO) could be used in the real time diagnostic monitoring of

East African droughts, especially during major ENSO events. This teleconnection could provide useful preliminary seasonal rainfall outlooks whenever warm or cold ENSO phases are forecasted or established.

The ultimate purpose in drought forecasting is to provide the society with sufficient warning of its occurrence for appropriate mitigative action to be taken. A timely preparation for drought could avert a significant proportion of the perilous impacts of drought when it occurs. A seasonal drought prediction model based on the concept of statistical teleconnection for East Africa has the potential to provide such timely warnings to those who live in this region.

1.3 Definition of Drought

Strictly speaking, there is no universally acceptable definition of drought. This is because (a) the concept of drought is not absolute but relative to uses and expectations, (b) drought (unlike floods) is not a distinct event - it neither has a well-defined start nor an end, and (c) drought is often the result of many complex factors acting and interacting within the environment.

Drought could occur in dry as well as in rainy areas. It is relative to the long-term or average water balance between rainfall and evapotranspiration in a particular area. The long-term condition is often perceived as "normal". Average rainfall alone does not provide an adequate measure of rainfall characteristic in a given region, especially in drier areas (Wilhite and Glantz, 1987). This is because the average value can be drastically affected by a few outliers. Drought is a "creeping phenomenon" (Glantz, 1987), making an accurate prediction of either its onset or end a difficult task. To most observers, it seems to start with a delay in the timing or a failure of the rains.

Although drought is difficult to define precisely, it can be regarded as the condition where there is a lack of sufficient water to meet demand requirements; which depend upon the distribution of plant, animal and human populations (Gibbs, 1975).

Drought is frequently defined according to a disciplinary perspective. Subrahamanyam (1967) identified six types of droughts; meteorological, hydrologic, climatological, atmospheric, agricultural and water management droughts; while Gibbs (1975) also included social-economic drought. Wilhite and Glantz (1987) suggested even though it may be useful to compartmentalize the various views of drought, the boundary separating these views is often vague. They grouped the above definitions into four types; namely meteorological, agricultural, hydrologic and social-economic droughts.

Meteorological drought is a period of substantially diminished precipitation duration and/or intensity. The commonly used definition of meteorological drought is an interval of time, generally of the order of months or years, during which the actual moisture supply at a given place consistently falls below the climatically appropriate moisture supply. Agricultural drought occurs when there is inadequate soil moisture to meet the needs of a particular crop at a particular time. Agricultural drought usually occurs after meteorological drought but before hydrological drought and can also affect livestock and other agricultural operations. Hydrological drought refers to deficiencies in surface and subsurface water supplies. It is measured in terms of streamflow, snowpack, and as lake, reservoir and groundwater levels. There is usually a time lag between a lack of rain or snow and measurable water in streams, lakes and reservoirs, making hydrological measurements later indicators of drought. Socio-economic drought occurs when physical water shortages start to affect the health, well being, and the quality of life of the people, or when the drought starts to affect the supply and demand of an economic product.

In terms of the above disciplinary drought definitions, meteorological drought usually precedes all other kinds of drought.

1.4 Scope of the Study

In this investigation, the term East Africa refers to three African countries; Uganda, Kenya and Tanzania even though the political boundaries of East Africa may not be consonant with the climatological zoning of the region. The lack of data in some parts of Africa prevents us from expanding the study area beyond these three countries.

This investigation is limited to meteorological drought, which precedes all other kinds of drought. Only dry spells more than a month in duration are considered in this investigation. It would have been desirable to track droughts using even shorter time units such as a daily unit because an affected drought region can return to normal condition with only a day's rainfall. However this is not practical given that most of the available raw climatological data is available in monthly time steps. Drought prediction (also known as drought warnings or outlooks) is generally limited to 3 to 6 months of lead-time, which is equivalent to one or two seasons in advance. Gridded precipitation data is used in developing the seasonal drought prediction models instead of station data. We realize that the resolution of the gridded rainfall data we used (2.5^0 latitude by 3.75^0 longitude) may at times be too coarse to capture mesoscale drought features. However most serious drought episodes in East Africa generally cover more extensive areas than this grid resolution.

This investigation is organized as follows: Chapter 2 discusses East African drought history, Climatology, and rainfall. Homogenous rainfall zones are delineated and their spatial and temporal characteristics studied. In Chapter 3, several drought indices are analyzed and experimented on East African data. The most appropriate drought index for East Africa is then selected and used to delineate this region into homogeneous drought zones. In Chapter 4, the influence of ENSO on East African rainfall based on harmonic analysis and other techniques is discussed. In Chapter 5, two statistical prediction models, Canonical correlation analysis (linear) and Projection pursuit regression (non-linear) are developed using sea surface temperature (SST) and sea level pressure (SLP) as predictors. The overall conclusion and summary are presented in Chapter 6.

Chapter 2 East African Precipitation Climatology and Drought History.

A general understanding of the rainfall climatology of East Africa is essential in analyzing its seasonal meteorological drought. In this chapter, we present a review of East African rainfall climatology and the history of droughts in the region. Further, zones of homogeneous rainfall regimes are identified. Each zone is assumed to have a unique ensemble of rainfall causative mechanisms and a corresponding unique teleconnection structure with some major climatic signals. Therefore, further drought analysis and predictions in Chapter 5 are presented in terms of these zones.

Past studies of East African rainfall using rotated principal components produced a highly segmented region with more than 26 rainfall climatological zones (Ogallo, 1989; Basalirwa, 1995; Basalirwa et al, 1999) which are not suitable for our investigation of East African droughts. The data and methodology used to arrive at our climatological zoning are briefly discussed in Section 2.4 and discussions of the results are presented in Section 2.5.

In East Africa, some seasons experience more rainfall than others and usually it is the failure of the critical rainfall seasons that cause the onset of drought. In section 2.5 we identify the important rain seasons for the various identified zones. These critical rainfall seasons are modeled in the statistical prediction models described in Chapter 5.

2.1 Rainfall Climatology.

East African rainfall exhibits great spatial and temporal variability (Ogallo, 1989; Nyenzi 1988), which cannot be shown in a simple map such as the mean annual rainfall of Figure 2.1a. This precipitation variability is partly due to the complex topography (Figure 2.1b), the existence of large inland lakes, the Indian Ocean in the east and the seasonal migration of the Inter-tropical Convergence Zone (ITCZ). These complexities produce diverse climates ranging from humid tropical to arid in East Africa.

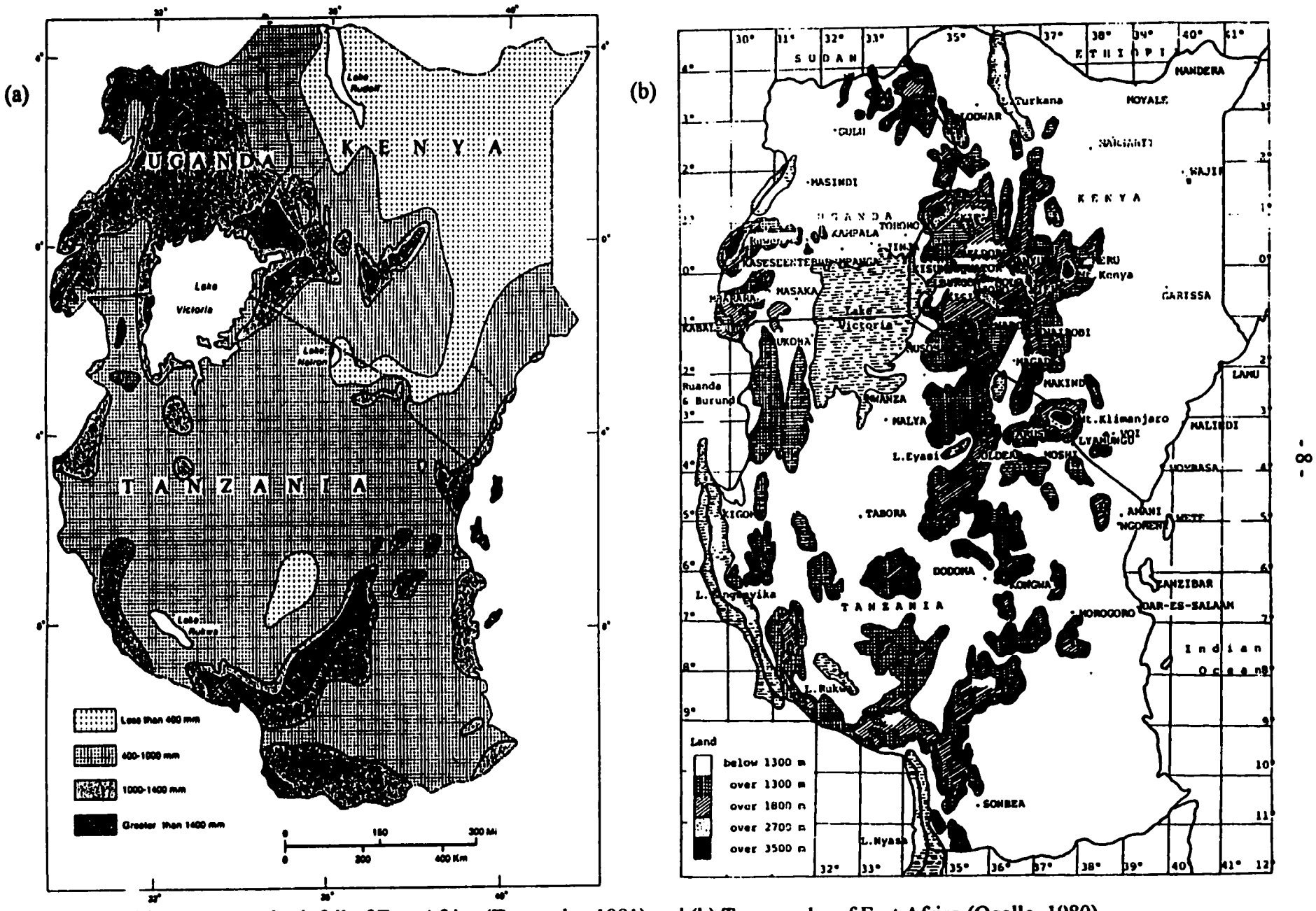


Figure 2.1 (a) Mean annual rainfall of East Africa (Trewartha, 1981) and (b) Topography of East Africa (Ogallo, 1980)

Humid conditions prevail in the highlands and near large water bodies, while dry areas include most of the northern and northeastern Kenya, central Tanzania and northern Uganda (Griffiths, 1972; Kenworthy, 1964)

In general, considering its equatorial position, East Africa does not receive much rainfall (Griffith, 1972). Trewartha (1981) attributed this rainfall deficiency to four factors, namely (1) the divergent character the Monsoon winds over extensive areas, (2) the modest thickness of the southwest monsoons over the highlands, (3) strong meridional winds occurring in all but the transitional seasons, a feature that limits the advection of sea moisture and reduces the orographic effect, and (4) a stable stratification of air aloft including a marked decline in moisture content.

It has been suggested that the East African seasonal rain-belts migrate northward and southward with the ITCZ, modified in some areas by local factors such as water bodies and highlands. Findlater (1971) and Ramage and Raman (1972) gave detailed accounts of the month to month variation of the low-level wind flow over the region. Using various rainfall anomaly scenarios, Okoola (1999) provided a detailed diagnostic investigation of the Eastern Africa Monsoon circulation during the Northern Hemisphere spring season.

During northern hemisphere (N.H.) spring and early summer, the ITCZ shifts northward and the southeast (SE) monsoon brings moisture into the region from the Indian Ocean, depositing part of their moisture over the highland areas and some of it in areas where thermal heating and wind convergence trigger rainfall (Findlater, 1971; Nyenzi, 1988). As the ITCZ moves further north, the south-easterlies turn to south-westerlies, maintaining a fairly strong diffluent low level jet over the coastal areas of East Africa. This period (July- September) is usually dry in most parts of Tanzania and Eastern Kenya, except in highland areas.

During the N.H. autumn, the northeast Trade winds flow into East Africa generally through two tracks (Findlater, 1971). One is a dry continental track from Arabia over Northern Africa and the other is a humid track over the Indian Ocean and Arabian sea.

During this time convective precipitation is common in most parts of the region. By December the northeastern monsoon is well developed, confining the low-level convergence to Southern Tanzania (Nyenzi, 1988).

2.1.1 Other Climatological Features

Temperature and evapotranspiration are also climatological features connected to droughts. The variation of the mean monthly air temperature throughout the year is very small, usually about 1° – 2°C in southwest Uganda and the southwestern parts of lake Victoria, with a maximum of 5°–6°C in southern Tanzania (Griffiths 1972). Nearly all stations report July or sometimes, August as the coldest month. The warmest month varies from February in Uganda to November and December in central and southern Tanzania. The absolute maxima do not usually exceed 40°C, and at an altitude of 2600 m is no higher than 25°C.

As expected, the mean temperatures in the region are very closely correlated to station elevation. Griffiths (1972) assumed approximately a fixed lapse rate and suggested the following linear relationships between temperature, T, in °C and the elevation h in meters;

$T = 34\text{-}0.0056h$ (mean annual maximum temperature °C)

$T = 24.5 - 0.0063h$ (mean annual minimum temperature °C).

On the average, the above relationships hold for most of the interior but they do not hold for most of the coastal areas as these regions are subject to distinct cooling by the land-sea breeze cycle.

The variation of the monthly mean dew point temperature is small in East Africa. A study of relative humidity (RH) values showed that the 06h00 readings (sunrise) in most stations of the region have an annual mean RH of 85-95%, except in northern Kenya with RH of 65-73%. At 12h00, RH readings are generally around 50%, except at the coast or

lakeside stations (65-70%) and in the very arid regions such as northeastern Kenya (35%).

The potential evapotranspiration (PET) exceeds annual rainfall in all of East Africa. Evaporation measured from 10 or 15-inch Evaporation pans ranges from an annual total of 1100 mm (117% of annual precipitation) in southwestern Uganda to 4500 mm (650% of annual precipitation) in northeastern Kenya. Altitude seems to be an important factor governing the rate of potential ET in this region (Griffiths, 1972).

2.2 Droughts in East Africa.

East Africa, like many other regions, has suffered from serious droughts in the past. Pre-colonial history of East Africa recounted in oral traditions demonstrates the impact historical droughts had on agricultural and pastoral societies in the region (Webster, 1979). Data relevant to determining the extent and severity of historical droughts in the region are scarce because of the lack of long-term instrumental climate records. There are also no high resolution proxy records such as tree rings and ice cores found in other climatic regions. Despite this, researchers have recently managed to reconstruct a 1,100 year rainfall drought climatology of the region based on lake-level and salinity fluctuations of Lake Naivasha (Kenya) inferred from palaeolimnological proxies (Verschuren et al., 2000). Their findings indicate that East Africa has alternated between contrasting climatic conditions, with significantly drier climate than today during the "Medieval Warm Period" (~ AD 1000 – 1270) and a relatively wet climate during the "Little Ice Age" (~ AD 1270 – 1850) interrupted by three prolonged dry episodes. Verschuren et al. (2000) also found strong chronological links between the reconstructed rainfall/drought climatology and the pre-colonial drought recounted in oral traditions.

Table 2.1 Major Drought Episodes in East Africa in the Twentieth century.

PERIOD	EXTENT	ZONE (see Fig. 2.5)	GENERAL COMMENTS	SOURCE
1899	Most parts of E. Africa, especially Northern Kenya	-	Lake Stephanie dried up.	Grove, 1974.
1900	Central Tanzania	6	More than 60,000 died	Kjekshus, H. 1996.
1949	Most of E. Africa, especially Sukuma Districts, Central Tanzania	2,3,4,5,6	1,500,000 cattle died or were hastily slaughtered out of a population of 2,500,000.	Baker, 1974
1965	Dry belt of Kenya	3	260,000 people affected.	USAID Disaster database.
1967	Karamoja, Uganda	3	25,000 people affected.	
1971	Wide spread in Kenya	3	1,500,000	
1977	Wide spread in Kenya	3	May: 100 killed, 20000 people affected.	
1979	Turkana District, Kenya.	3	40,000 people affected.	
1979	North and Northeastern Uganda especially Karamoja.	1	600,000 people affected.	
1984	Most of Kenya and Tanzania	3	Complete failure of Long rains. Worst drought in Kenya in 40 years.	Climate Monitor, 1984.
1984	Arid districts of Kenya	3	600,000 people affected.	USAID Disaster database.
1987	Karamoja, Uganda	3	331,000 people affected. Inadequate spring rains.	
1988	Northwestern Uganda	1	600,000 people affected. Poor seasonal rains.	
1990	North and north eastern Arid Districts of Kenya	3	1,200,000 people affected. Worst maize crop in 10 years.	
1991	North and north eastern Arid Districts of Kenya	3	2,700,000 people affected. Worst drought in 50 years. The NE region of Kenya was most severally affected.	
1992	North and north eastern Arid Districts of Kenya	3	Continuing drought.	
1996/97	Central Tanzania.	6	Late 96 - early 97: 50-year worst drought in Tanzania: Cities face major water shortages.	Climate Monitor, 1997.

Instrumental climate records show that East Africa has experienced several drought episodes in the past 100 years (Table 2.1), particularly in Kenya and Uganda identified as Zone 3 in Figure 2.5.

Perhaps the worst dry spell in the last 100 years occurred in 1898/99 in northern Kenya and southern Ethiopia. Lake Stephanie, which is just above the northern Kenya boarder, was several meters deep in 1895 – but later dried up completely in 1899, following two very dry years. By 1974 it still had not regained its former levels (Grove, 1974). During this dry spell, Lake Turkana in Northeastern Kenya also experienced a very dramatic decrease in areal extent, from which it has yet to fully recover. 1949 was also a particularly dry year in much of East Africa especially Tanzania. It is reported that only 40% of the 2.5 million cattle population in central Tanzania survived the 1949 drought (Baker, 1977). Droughts have occurred in East Africa in more recent years. The 1984 droughts in Kenya were among the worst in over 40 years (Climate Monitor, 1984). Tanzania too was adversely affected by the complete failure of the 1984 "long" rains, and again affected by a very severe drought in late 1996 to early 1997 (Climate Monitor, 1997). During this period, there was record low February rainfall in many parts of East Africa.

The most common livelihood practice in East Africa is subsistence agriculture and pastoralism, which are predominantly rain-dependent. Hence this region can easily be affected by modest meteorological droughts of short durations. The recent (Nov. 96 - May 97) drought wave in East Africa had a serious impact on its economy. The *East African* newspaper of Jan. 27th 1997 reported that it was one of the worst droughts to hit Tanzania in 40 years and it threatened to destroy its economy. Failure of the rains had resulted in severe limitations in crop cultivation and water shortages. Consequently, Dar es Salaam, the largest city in Tanzania with a population of about three million, faced an immense water shortage that forced many industries to shut down. The Uganda government newspaper, *The New Vision*, of 26th May 1997 reported that the 96/97 drought led to a doubling of the inflation rate with significant increase in the prices of staple foods in Uganda.

There are only a limited number of studies devoted to the understanding of the nature, distribution and origin of drought in Eastern Africa, despite its persistent recurrence in the region (e.g. Trewartha, 1981 and Beltrando 1990). The overall motivation of this investigation is the need for a better understanding of East African drought and how it teleconnects to large scale climatic signals.

2.3 Temporal Definition of East African Rainfall Seasons

This investigation focuses on meteorological drought analysis and prediction at seasonal time scales. Not all seasons are important in terms of annual rainfall contribution. Some areas regularly experience dry seasons. It usually takes the failure of one or more important rain seasons to bring about the onset of drought. Therefore for drought predictions to make sense, it is necessary to identify the season for each zone (Figure 2.5) critical in terms of annual rainfall contribution.

Previous studies such as that of Griffiths (1972) have pointed out the ambiguity as well as the complexity of rain seasons in this region. There is probably no single month during the year when all parts of East Africa are dry. The East African Meteorological Department (1963) generalized the rainfall distribution patterns in East Africa into four broad seasons.

Season 1 (December – February, DJF)

This period is usually dry for Uganda and Kenya, but generally wet for most of Tanzania. During this time, the ITCZ is far to the south outside the East African region. Any rainfall during this season in Uganda and Kenya are associated with regional features, such as the land-lake breeze in the shore areas of Lake Victoria.

Season 2 (March to May, MAM)

This main rainy season throughout Uganda and most of Kenya is referred to locally as the "long rains". It coincides with the presence of the moist southeast monsoons from the Indian Ocean in the region which convergence into the ITCZ.

Season 3 (June to August, JJA)

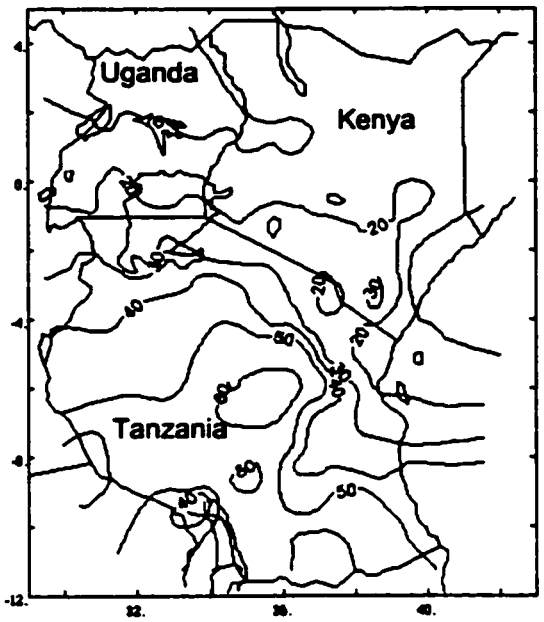
This season is relatively dry except in some parts of Northern Uganda. The rains in northern Uganda during this season are associated with the influx of the moist westerly Congo air mass controlled by the St Helena highs off south-west Africa (Basalirwa, 1995).

Season 4, (September to November, SON)

This rainfall of this season, locally known as the "short rains", are associated with the convergence into the ITCZ of the northeast monsoons controlled by the subtropical anticyclones over the Azores and the Arabian Peninsula.

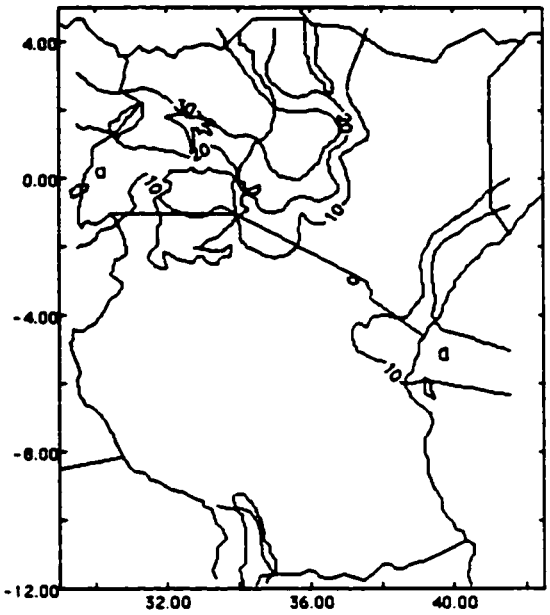
It should be noted that Seasons 1, 2, 3, and 4 roughly correspond to the northern Hemisphere winter, spring, summer and autumn respectively. These seasons were also used by Basalirwa (1995) in analyzing his climatological zones. However, other investigators such as Nyenzi (1988) and Ogallo (1983 and 1989) shifted the above seasons by one month in their analyses.

Figures 2.2 a-d show the seasonal % contributions towards the mean annual precipitation based on the seasons defined by the East African Meteorology Department (1963). Figure 2.2 (c) shows that northern Uganda and a few parts of western Kenya get a substantial portion of their rainfall from June to August (northern Hemisphere summer), which is probably due to the moist Congo mass. Most of Tanzania is dry during this season. Figure 2.2 (d) shows that the fourth season is quite important for Kenya and Uganda which receive approximately 30% or more of their annual rainfall during this period (September to November).



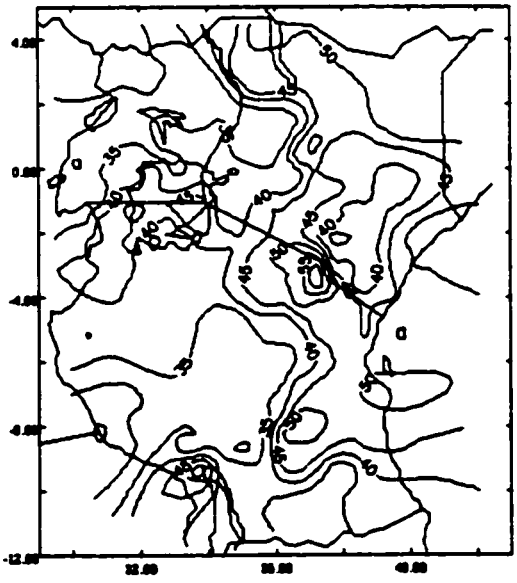
(a) Season 1: DJF ("Winter")

Tanzania's wet season



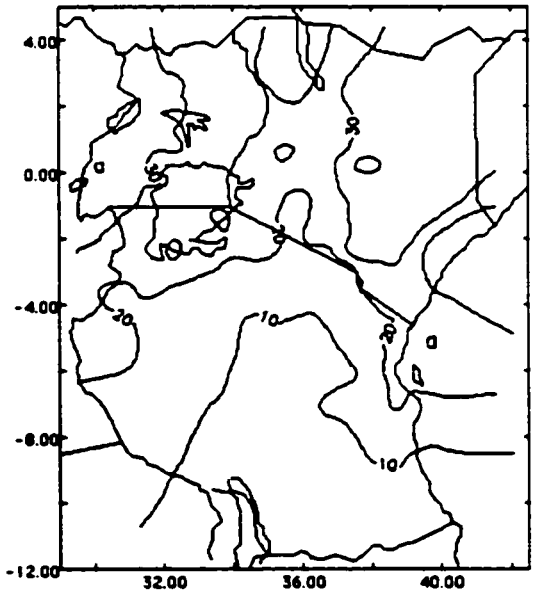
(c) Season 3: JJA ("Summer")

Normal Dry Season



(b) Season 2: MAM ("Spring")

(Long rains)



(d) Season 4: SON ("Autumn")

(Short rains)

Figure 2.2 (a) – (d) Percentage Seasonal Contributions to the Annual Rainfall for East Africa

The failure of two consecutive critical seasons in areas with bimodal rainfall, or the single critical rain season in zones with unimodal rainfall will result in water stress and eventually drought. Accurate prediction of the important rain seasons would therefore be a useful tool in mitigating the potential impacts of droughts in East Africa

2.4 Regionalization of East African Rainfall

When modeling a complex spatial phenomenon such as drought, it is advisable to first identify zones with homogenous characteristics. The identification of such homogeneous rainfall zones with unique patterns would be essential in developing a parsimonious seasonal prediction model for the region. Further, by studying the characteristics of the identified homogenous zones, we can also gain some understanding of the underlying mechanisms behind droughts that occurred in the region.

Several investigators have studied the spatial characteristics of seasonal rainfall in East Africa. Ogallo (1989) used rotated principal component analysis (RPCA) to analyze the spatial and temporal characteristics of seasonal rainfall in East Africa. He delineated 26 different homogeneous zones. Basalirwa (1995) and Basalirwa et al (1999) also used RPCA to do a more detailed analysis of Uganda and Tanzania respectively and came up with even more subdivisions.

While it is possible that the spatial distribution of East African rainfall may be so complex that it warrants many subdivisions, we chose a more modest approach, the harmonic analysis, that divided East Africa into only six homogenous precipitation zones. This was done partly to avoid excessive complexity. Ogallo (1983) studied the temporal fluctuations of seasonal rainfall patterns in East Africa using harmonic analysis but without delineating East Africa into homogenous rainfall zones. Potts (1971) also used harmonic analysis and the 1931-1960 data to group East Africa into climatological subdivisions. Many rainfall climatological extremes have occurred since 1960. Further, some of the climatological divisions he identified are not realistic, probably because of

the limited data he used. Now with almost a century of precipitation data available, we are probably in a better position to subdivide East Africa into more realistic precipitation zones than of Potts (1971).

2.4.1 Rainfall Data.

Mean monthly rainfall statistics were obtained for 135 stations scattered throughout East Africa (Figure 2.3) and many of these stations had about 97 years of data (1900 -1996). These 12 monthly mean values for the 135 stations were used in the harmonic analysis (section 2.3.2). In addition, a historical monthly precipitation dataset covering East Africa, for the period 1900 – 1996, gridded at 2.5° latitude and 3.75° longitude was used in the analysis. This gridded data was extracted from the global monthly precipitation dataset for land areas "gu23wld0098.dat" (Version 1.0) constructed and supplied by Dr Mike Hulme at the Climatic Research Unit, University of East Anglia, Norwich, UK. The grid points are shown in Figure 2.3

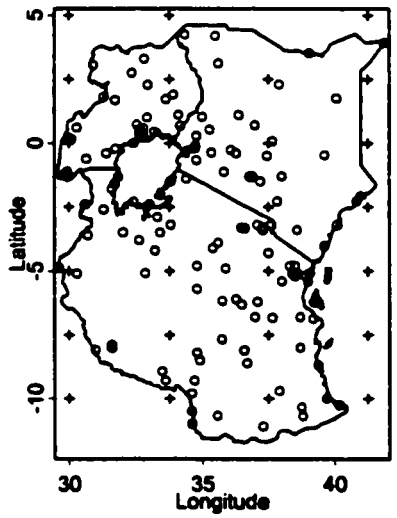


Figure 2.3 A map of East Africa showing the location of rainfall stations and the gridded dataset used in the analysis.

Hulme (1994) used Thiessen polygon weights to average Global Historical Climatology Network (GHCN) station data within 2.5° latitude by 3.75° longitude grid boxes. Where a

monthly station value was missing, he obtained an estimate by calculating the mean anomaly for that location derived from surrounding stations. This anomaly interpolation method required the station values to be converted into percentage anomalies from some reference period. The standard anomalies were then interpolated onto the missing station location using an inverse distance (with spherical adjustment), angular weighted method. For this interpolation, Hulme (1994) imposed a maximum percent anomaly value of 500. The interpolated percent anomaly was then converted back into a station mm estimate using that station's mean monthly precipitation total for the reference period. More details about the quality control of the global monthly precipitation dataset "gu23wld0098.dat" and the algorithm used to obtain the dataset can be obtained in Hulme (1994); Hulme and New (1997) and Hulme et al. (1998).

2.4.2 Harmonic Analysis

Harmonic analysis was used to delineate precipitation climatological divisions because this method has been proven to be effective for such analysis (e.g., Sabbagh and Bryson, 1962; Potts, 1971; Scott and Shulman, 1979; Kadioglu et al. 1999). For example, Scott and Shulman (1979) used harmonic analysis to investigate the areal and temporal characteristics of precipitation in the Northern United States. Such studies have demonstrated the efficiency of harmonic analysis in tracking spatial and temporal variations in rainfall occurrences for small and large areas (e.g., Kadioglu et al., 1999).

Theoretically, any function can be represented by an infinite series of sine and cosine waves, commonly referred to as the Fourier series. For discrete observed data, a finite number of series will generally be sufficient. For example, the sample mean, five sine and six cosine terms will be sufficient to completely describe the variation of a dataset containing 12 average monthly values. The determination of the finite sum of sine and cosine terms is called the "Harmonic analysis" (Panofsky and Brier, 1965).

The first or the "fundamental" harmonic has a period equal to the total length of the time series being analyzed. The second harmonic has a period equal to half the fundamental period; the third harmonic has a period of one third of the fundamental and so forth.

It can be shown that for *n* data points, x_j; j = 1, 2, . . . n can be fully represented by a summation of n/2 harmonics,

$$x_j = \bar{x} + \sum_{k=1}^{n/2} c_k \cos\left(\frac{2\pi k}{n}(t - t_k)\right) \quad \text{.....(1)}$$

where $\bar{x}$ is the sample mean, $c_k = \left(a_k^2 + b_k^2\right)^{1/2}$ where a_k and b_k are Fourier coefficients given by:

$$a_k = \frac{2}{n}\sum_{j=1}^{n-1} x_j \cos\left(\frac{2\pi jk}{n}\right) \quad \text{.....(2a)}$$

the last a_k term is

$$a_n = \frac{1}{n}\sum_{j=1}^{n-1} x_j \cos(2\pi j) \quad \text{.....(2b)}$$

$$b_k = \frac{2}{n}\sum_{j=1}^{n-1} x_j \sin\left(\frac{2\pi jk}{n}\right) \quad \text{.....(3)}$$

The last b_k term is always zero. The a_k terms are symmetric while the b_k terms are asymmetric about k=0 and $k = k_{max}$ respectively. The phase shift of the k^{th} harmonic in terms of months is denoted by t_k, as

$$t_k = \frac{n}{2\pi k}\arctan\left(\frac{a_k}{b_k}\right) \quad \text{.....(4)}$$

and

$$t_k = \frac{n}{2\pi k}\arcsin\left(\frac{a_k}{c_k}\right) \quad \text{.....(5)}$$

Solving Equation (5) will lead to two solutions but one of which will be equal to one of the two solutions of Equation (4). The common solution between Equations (4) and (5) is the correct solution.

Coherence is defined as

$$coh_k = \frac{\frac{1}{m}\sum_{i=1}^{m} \bar{v}_{i,k}}{\frac{1}{m}\sum_{i=1}^{m} |\bar{v}_{i,k}|} \quad \text{.....(6)}$$

Where v is a vector representing the magnitude and phase of the k^{th} harmonic vector. Thus coherence is the vectorial mean of the k^{th} harmonic vectors divided by the scalar, arithmetic mean of the harmonics for *m* events for a given grid location. Coherence can also be expressed as

$$coh_k = \frac{\left[\left(\sum_{i=1}^{m} v_{i,k} \cos\left(\frac{t_k n}{2\pi k}\right)\right)^2 + \left(\sum_{i=1}^{m} v_{i,k} \sin\left(\frac{t_k n}{2\pi k}\right)\right)^2\right]^{\frac{1}{2}}}{\sum_{i=1}^{m} |\bar{v}_{i,k}|} \quad \text{.....(7)}$$

The coherence ranges from 1 (when all the vectors have the same direction but not necessarily the same magnitude) to 0 when they have the same average magnitude in all directions. Coherence of the vectors can indicate the consistency of the composite data patterns.

A further measure of the reliability of the results is the amount of variance (*ve)* extracted by various harmonics,

$$ve_k = \frac{100c_k^2}{\sum_{i=1}^{n/2} c_i^2} \quad \text{.....(8)}$$

Further details of harmonic analysis can be obtained from Rayner (1971).

The magnitude, phase and variance for each of the harmonics obtained from the average monthly data for the 135 stations are presented in terms of contour maps (Figure 2.4).

Other than analyzing monthly averages, harmonic analysis is applied to multi-year data to get the temporal characteristics and year-to-year variation of the dominant harmonics as shown in Figure 2.7.

2.5 Discussion of Results

In many parts of our study site, the first two harmonics extracted most of the variance in the monthly mean data (Figure 2.4a&b). The 4^{th}, 5^{th} and 6^{th} harmonics generally account for very little of the variance in the data (in many cases less than 3%) and hence are discarded for further analysis.

2.5.1 Spatial Characteristics of the Data Harmonics

Figure 2.4a shows the regional pattern of the variance explained by the first harmonic as well as its phase in months. A phase of 2 means that the harmonic peaks during mid-February.

The 1^{st} harmonic represents the annual cycle of the data. There are two zones that are well represented by this harmonic (Figure 2.4a). The first zone includes most of the south, south west and west of Tanzania. In southwest Tanzania, the first harmonic extracts more than 90% of the variance and attains its peak in mid-February. This coincides with the passage of ITCZ. The second zone dominated by the first harmonic is in a form of a wedge like incursion starting from northern Uganda, dipping into western Kenya (Figure 2.4a). In this zone, the first harmonic reaches its peak between July and August which corresponds to the northern Hemisphere summer. The unimodal nature of the rainfall in zone 1 and 6 is demonstrated in Figure 2.6.

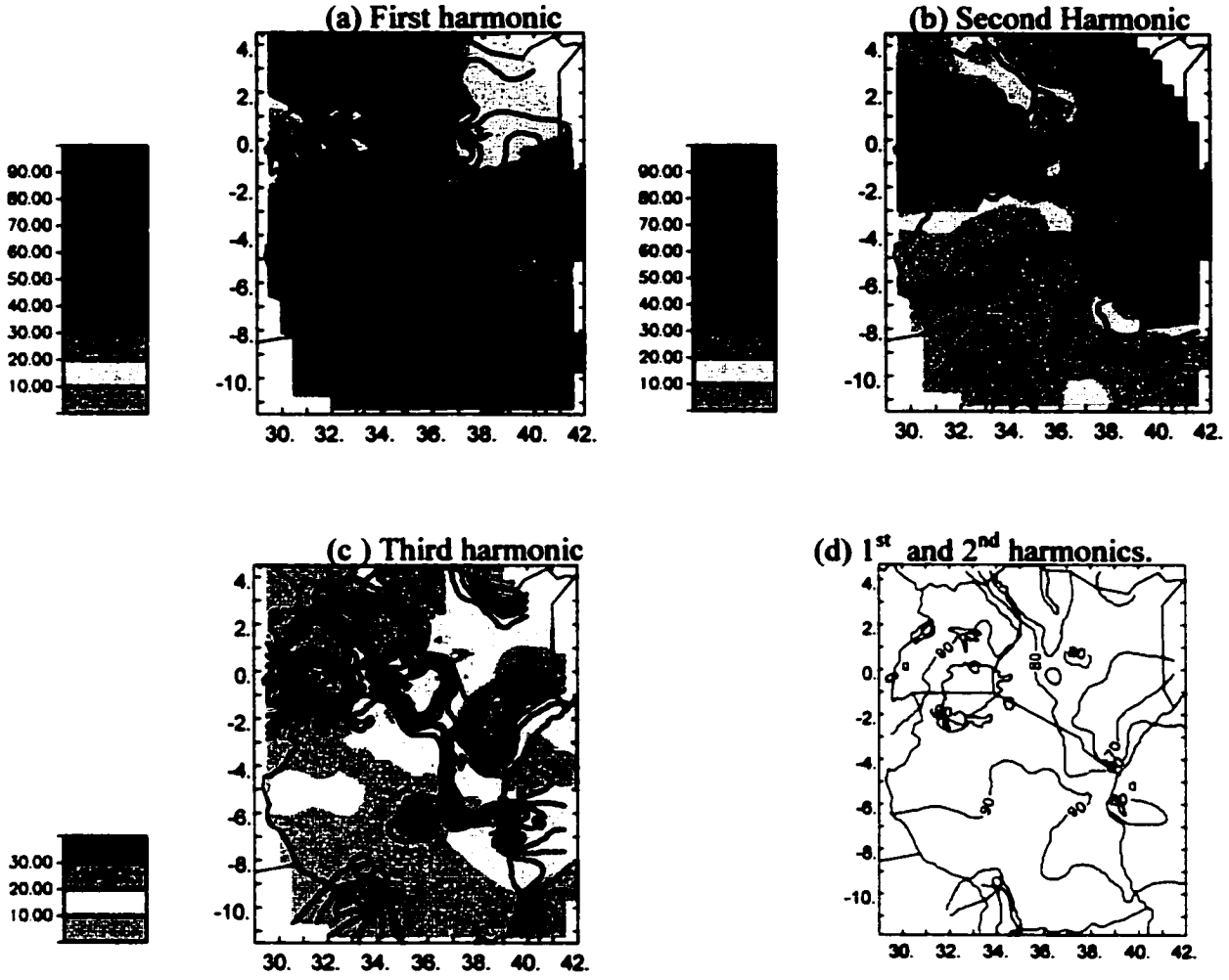


Figure 2.4 A Map of East Africa showing the % of variance explained by the (a) first, (b) second and (c) third harmonic and contours of their phase in month and (d) shows the total variance explained by the first two harmonics.

The 2nd harmonic is dominant in central and western Uganda, northeastern Kenya and to a lesser extent, north Eastern Tanzania (Figure 2.4b). This harmonic represents the bimodal component of rainfall in a given zone, and so all the aforementioned areas tend to have bimodal rainfall (Figure 2.6). Those areas that have rainfall peaks that are approximately equal and are six months apart are best represented by this harmonic. Often the peaks are separated by less than six months. In addition one of the seasonal peaks may at times be smaller than the other (e.g., long rain seasons of March-May and shorter rain season of September-October). Both the 1st and 2nd harmonics extract

significant variance. The 1st harmonic attenuates one of the two peaks of the 2nd harmonic. Stations in northeast Tanzania tend to be well represented by totals of the 1st and 2nd harmonics (see Zone 6, Figure 2.6)

The 3rd harmonic extracts significant variance in some drier areas of Kenya (Figure 2.4c). The variance of rainfall in drier climates is generally higher than that in the more moist climates. That may explain why 3 harmonics are needed to describe the annual rainfall curves in the drier parts of Kenya. Theoretically, the 3rd harmonic represents the tri-modal component in the rainfall. There are indeed some East African rift valley stations that have loosely defined tri-modal rainfall. However, southeastern Kenya, which is well represented by the third harmonic, does not have tri-modal rainfall. In this zone, the 3rd harmonic is probably mathematically necessary to better define the maxima location and the relative peak magnitudes.

The regions which were observed to have the same amount of variance extracted by the respective 1st, 2nd or 3rd harmonics and in addition share the same harmonic phase (see Figure 2.4) were grouped into respective homogenous divisions. Thus using the results from harmonic analysis, East Africa was delineated into six homogenous rainfall divisions as shown in Figure 2.5. It should be stressed that these regimes are based on the periodic components of the average monthly rainfall. The absolute magnitude of the rainfall is not a factor used in delineating the zones. The unshaded central region in Figure 2.5 does not seem to have unique rainfall characteritics discernable by the harmonic analyis probably because it shares the characteristics adjacent zones in varying degrees. Therefore, this central region is not considered a distinct precipitation zone because of the lack of clarity of its defining features. Examples of monthly rainfall histograms with fitted harmonics of representative stations located in the six delineated zones are given in Figure 2.6.

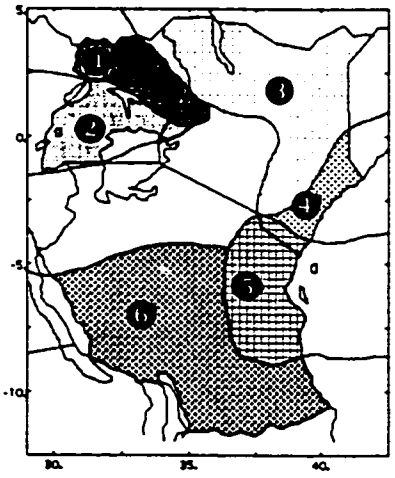


Figure 2.5 Six homogeneous precipitation zones for East Africa delineated by Harmonic Analysis.

2.5.2 Temporal Characteristics of the Zoned Rainfall

The previous harmonic analysis has been based on the mean monthly rainfall data. It could be argued that using mean monthly values can at times mask out important features in the signal. To address this concern, harmonic analysis was also applied to the gridded historical monthly precipitation data set of 1990-1996 (Hulme, 1994). We end up with 97 sets of harmonics for grid points from which the summary statistics are computed (Table 2.2).

For Zone 1, the ratio of the magnitude of the first harmonic to the second harmonic (h1/h2) was greater than 2 most of the time. The peak rainfall (phase) fluctuated from mid-June to mid-August and the variance explained by the second harmonic has always been less than that explained by the first harmonic except for 1949 (see Figure 2.7(a)) when the monthly rainfall distribution was bimodal instead of the usual unimodal feature found in Zone 1.

The monthly rainfall distribution in Zone 2 is mostly bimodal. Also, the 2nd harmonic explains most of the variance except for the 1942-48 period, during which the monthly rainfall distribution switched to the unimodal mode (see Figure 2.7b).

Table 2.2 Summary of the Harmonic Analysis of the aggregated six homogenous precipitation zones using gridded rainfall (Hulme, 1994) dataset.

Zone	Average Variance Explained by Harmonics (%)		Ave. Ratio of the 1st harmonic magnitude to the 2nd harmonic.	Mean of Harmonic Phase (months).		Rainfall mode
	var1	var2	|h1|/|h2|	ph1	ph2	
1	56	21	2.1	7.2	7.5	Almost unimodal April to August rainfall but with slightly reduced June rainfall.
2	14	56	0.51	5.4	8.1	Mainly Bimodal with exception of the 1948-1949 period.
3	15	48	0.59	4.4	9.1	Bimodal with very clear dry June to August season
4	36	32	1.2	5.6	9.8	Almost unimodal (with a sharp peak in May).
5	54	18	2.6	2.7	8.4	Bimodal but with often weak September to November rain season
6	84.1	6.4	4.6	1.8	4.1	Unimodal rainfall, mainly peaking early February.

It has been previously noted that the 3rd harmonic extracts significant (~30%) variance from the average mean rainfall of Zone 3 (Figure 2.4a). Temporal analysis actually shows that on many occasions, it is the second harmonic that extracts most of the variance; thus the rainfall of this region is actually bimodal. Overall there is considerable fluctuation in the h1/h2 ratio and the phase of the rainfall maxima in Zone 3. This is consistent with the general observation that rainfall fluctuations are more substantial in dry than in wet zones and Zone 3 is a relatively dry region.

According to Table 2.2, the 1st and 2nd harmonics in Zone 4 are generally of comparable magnitude each extracting about 35% of the total variance. Figure 2.7 shows that there is a substantial year-to-year fluctuation of variance extracted by the first two harmonics in this zone. The rainfall histogram of zone 4 (see Figure 2.6) shows a unimodal type of rainfall which normally peaks in May. It is probable that this zone experiences significant rainfall variations because of the land-sea breeze influence of the Indian Ocean. Figure 2.7 shows that there have been times when the variance explained by the 2nd harmonic is greater than the 1st harmonic such as 1981 to 1984, signifying a possible change of rainfall regime.

According to Table 2.2, Zone5 should have unimodal rainfall since over 50% of the variance is extracted by the first harmonic. However, the rainfall distribution in Zone 5 is more or less bimodal with the main rainfall peak occurring around late April (see Figure 2.6 and 2.7) and a second subdued peak occurring in November. Unlike Zone1 where the phase of both the first and second harmonic are almost the same, the phase of the second harmonic in Zone5 peaks when its first harmonic is getting close to its minimum. As a result, the combined effect of the two harmonics is a subdued, second seasonal peak in November. Like Zone 4, Zone 5 also exhibits considerable year-to-year variation in terms of variance explained by the first two harmonics (Figure 2.7e), which is probably due to local sea breeze effect from the Indian Ocean.

Among the six zones, Zone 6 exhibits the most stable rainfall pattern, with the % of variance extracted by the 1st harmonic rarely dropping below 80%. The rainfall peak occurs around late January or early February. The significant variation of the h1/h2 ratio for Zone 6 (Figure 2.7f) should be ignored since the 2nd harmonic, (h2) explains negligible variance compared to the 1st harmonic (h2).

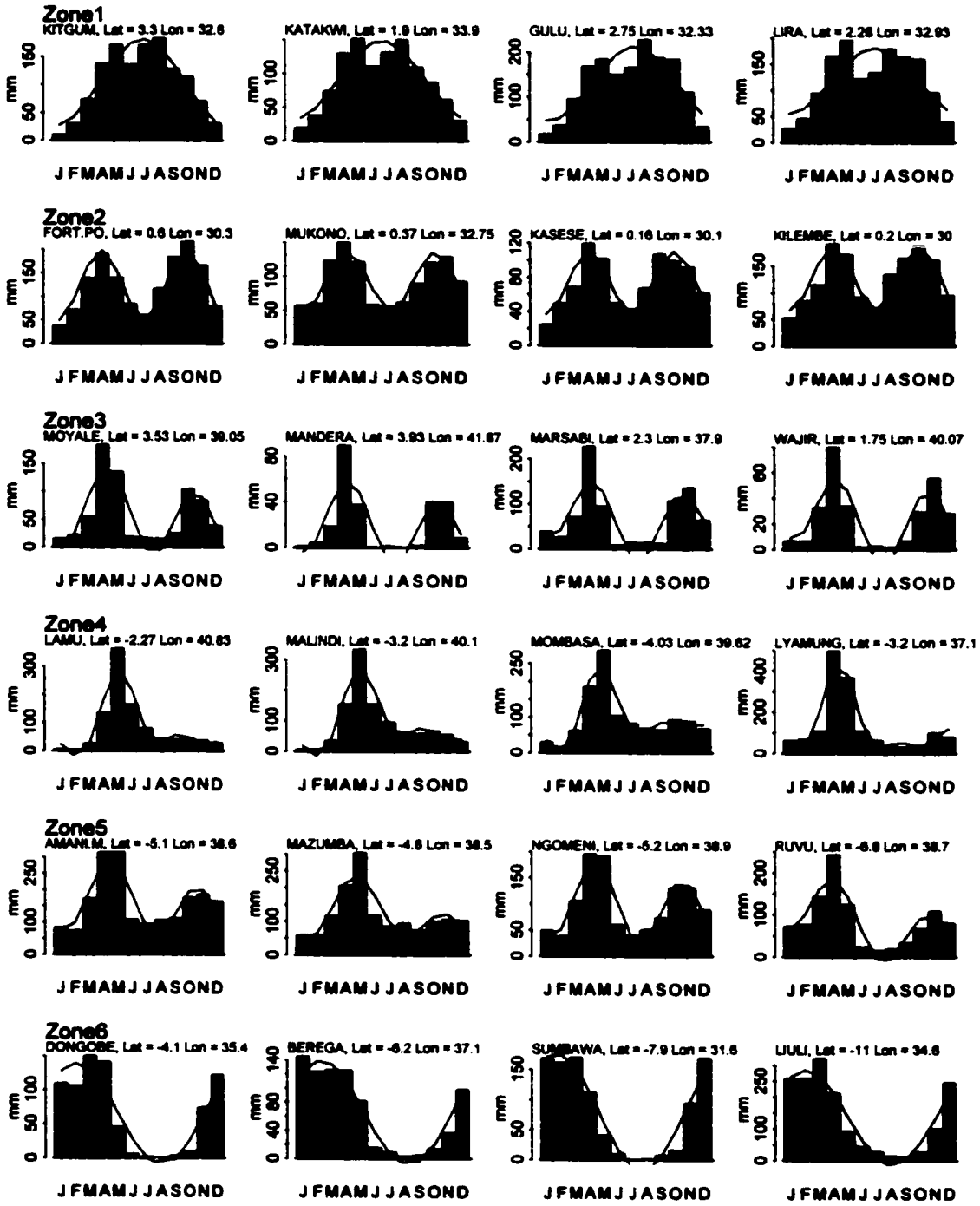


Figure 2.6 Histogram plots of mean monthly precipitation with fitted harmonics of selected stations of the 6 zones identified by harmonic analysis. Stations in Zone 1 and 6 are each fitted with only the first harmonic while those in zone 2 are fitted with only the 2nd harmonic. Zone 3 and 5 stations are fitted with the total of the 1st and 2nd harmonics while those in Zone 4 are fitted with the total of the 1st, 2nd and 3rd harmonics.

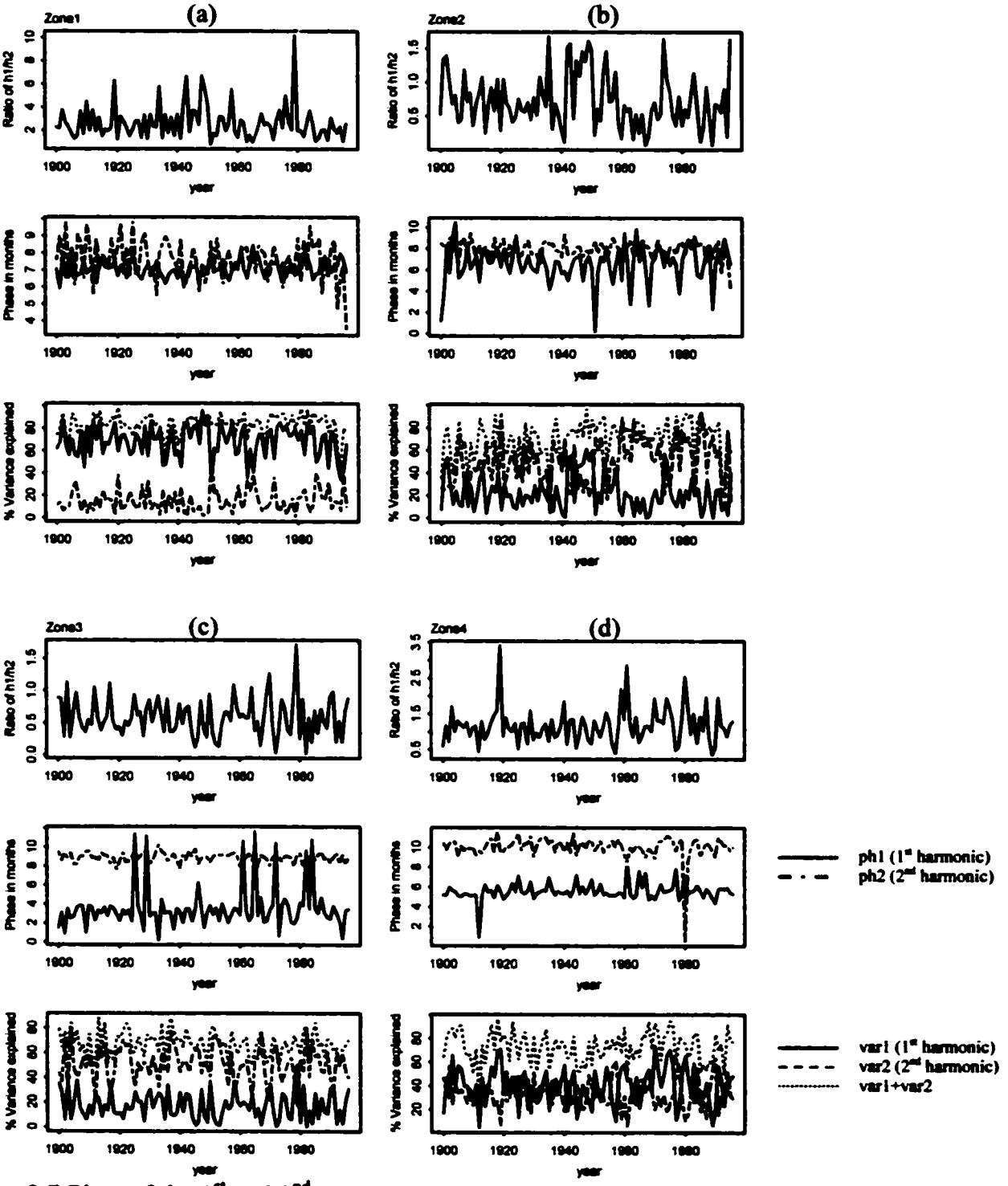


Figure 2.7 Plots of the 1st and 2nd Harmonics for the zonal averaged precipitation data of East Africa. The first row shows the ratio of 1st harmonic to the 2nd harmonic (h1/h2); the second row shows the phase of the 1st and 2nd harmonic and the third row shows the % of variance extracted individually by the 1st and 2nd harmonics as well as the sum of the first two harmonics. Each column corresponds to a specific zone.

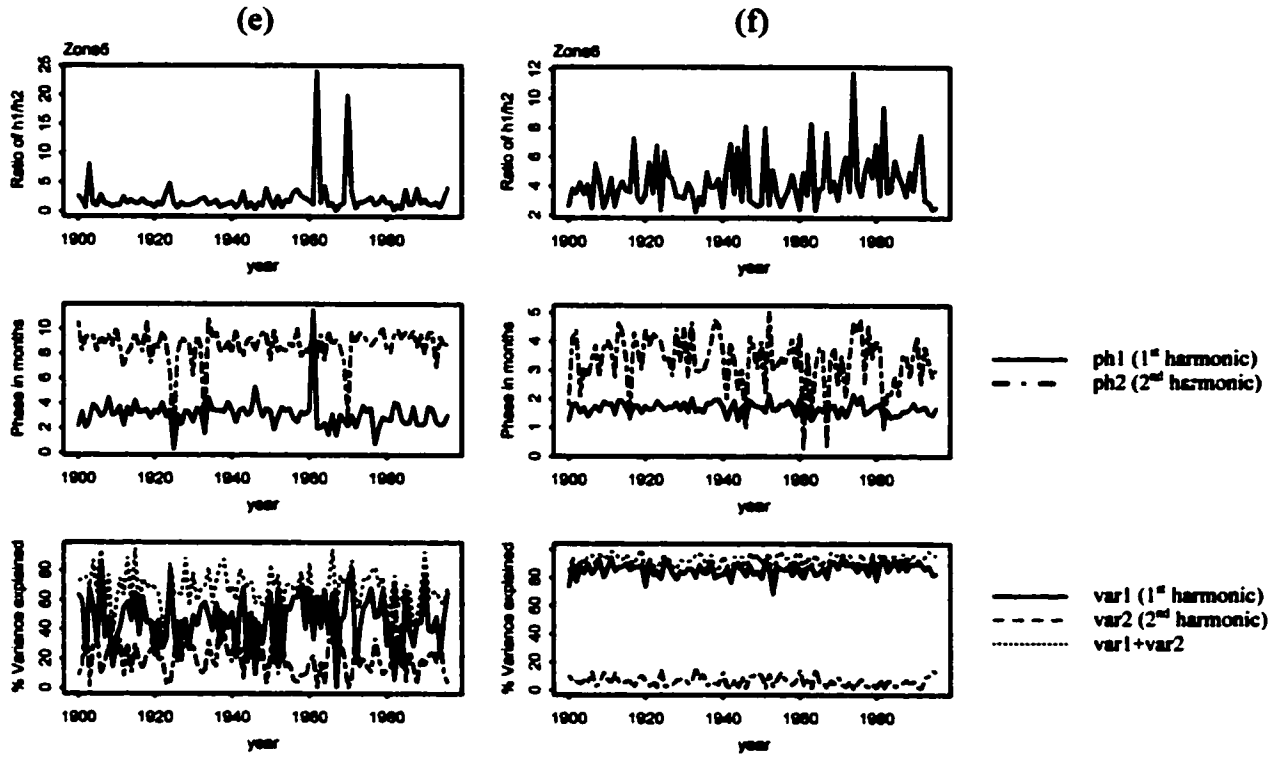


Figure 2.7 (Continued)

2.5.3 Important Rain seasons in the Various Homogeneous Rain Zones.

The identification of the important rain seasons in each of the delineated rainfall homogeneous zones in East Africa can lead to a better understanding of the drought incidents in the region. The rainy seasons provide water recharge critical to assisting the biosphere's survival through the normal dry period. In a region that heavily depends on rain–fed agriculture, any failure of the major rainy seasons will often be accompanied by massive crop failure which in turn can quickly impact the fragile economy of the region.

Table 2.3 shows the importance of certain rain seasons to some parts of East Africa. The importance of the various seasons in the delineated zones is deduced by analyzing the % contributions to annual rainfall (Figure 2.2) alongside the delineated rainfall

homogeneous zones (Figure 2.5) and the histogram plots of mean monthly precipitation (Figure 2.6).

Table 2.3 % of Rainfall Contributions of the major Rainfall seasons of East Africa subdivided into 6 zones (Figure 2.5) by harmonic analysis

Zone	Season				Total % Contribution
	DJF	MAM	JJA	SON	
1		✓	✓	✓	90%
2		✓		✓	70%
3		✓		✓	85%
4	✓	✓			80%
5	✓	✓			80%
6	✓	✓			90%

For example Zone 6 receives 90% of its annual rainfall in winter and early spring while the rest of the year is normally dry. The region has adjusted itself naturally to experience one long dry period every year starting from June up to November. The failure of the main seasons (DJF and MAM) will certainly lead to increased water stress during the normal dry period (JJA and SON), and even droughts under severe water deficit.

2.6 Summary and Conclusions

The spatial and temporal characteristics of East African rainfall have been identified in terms of 6 homogeneous zones via harmonic analysis applied to rainfall data grouped under four seasons by the East African Meteorological Department.

Based on the temporal structure of the first few harmonics of the 6 sets of zonally averaged rainfall data, we showed that the six zones possess rainfall properties that are generally unique and different from other zones. Among these zones, Zone 6 (most of central and southern Tanzania) probably has the most stable rainfall structure because the

phase of its first harmonic hasn't changed by more than a month in the last 90 years. This means that the peak rainfall always occurs between January and February for central and southern Tanzania.

The major rainy seasons of each zone, which contribute most to the total annual rainfall, were identified. It is clear that the failure of two consecutive critical seasons in areas with bimodal rainfall, or the single critical rain season in zones with unimodal rainfall will result in water stress and eventually drought. Accurate prediction of the important rain seasons would therefore be a useful tool in mitigating the potential impacts of droughts in East Africa

The statistical analysis and prediction of the seasonal rainfall for some of the critical rain seasons is the main focus of Chapter 5.

Chapter 3 Drought Indices and their Application to East African Conditions

3.1 Overview.

A variety of descriptors and definitions for drought have been used in the past. Gibbs (1975) and Wilhite and Glantz (1987) discussed drought definitions and their role in understanding the phenomenon. Drought indices, derived from massive amount of hydroclimatic data like rainfall, snowpack and stream flow are meant to provide a concise overall (and possibly comprehensive) picture of drought and they are often used for making decisions on water resources management and water allocations for mitigating the impact of drought.

The primary aim of this investigation was to analyze the properties of several popular drought indices and modify them where necessary and/or feasible in order to increase their effectiveness and dependability in detecting the initiation, evolution and termination of meteorological droughts in East Africa. Ideally, drought control and management strategies should be tied to a set of quantitative drought indices in order to reduce the subjective preferences of decision makers to a minimum. Therefore, an investigation on drought indexing should contribute to the decision making of drought control under an objective and impartial framework.

Table 3.1 shows some drought indices that are currently in use, or have been used in the past in various parts of the world. For an index to be effective, it should be derived locally, adapted to the climate of the territory and reflect a broad conceptual framework which comprehensively describes droughts in the region (Yevjevich et al., 1977). Along this line, we analyzed Palmer's Drought Severity Index (PDSI), Bhalme-Mooley Index and the Standardized Precipitation Index (SPI) and modified them where feasible in order to make them better indices for monitoring droughts in East Africa.

Table 3.1 Characteristics of current and past drought indices

Name of Index and its Developer	Input Data	Time Scale	Brief Description
1. Palmer Drought Severity Index (PDSI). Palmer (1965).	P, T, ET, L, RO	monthly	The PDSI is a soil moisture algorithm calibrated for relatively homogeneous regions. It is based on moisture inflow, outflow and storage. Many U.S. government agencies and states still rely on the PDSI to trigger drought relief programs.
2. Crop Moisture Index Palmer (1968).	P, T, ET, L, RO	weekly	A PDSI derivative, which reflects moisture supply in the short term across major crop producing regions.
3. Standardized Precipitation Index (SPI) McKee et al. (1993).	P	3 month – 48 month	An index based on the probability of precipitation for any time scale.
4. Surface Water Supply Index (SWSI) Shafer and Dezman (1982).	P, sn RO	monthly	The SWSI shares many characteristics with the PDSI but also takes into consideration the snow pack and the resulting runoff.
5. Reclamation Drought Index (RDI) Bureau of Reclamation (USA).	P	monthly	RDI is calculated on the river basin scale. Since index is unique to each river basin, interbasin comparison is not possible.
6. Bhalme Mooley Index (BMI) Bhalme and Mooley (1979).	P	monthly	The BMI models the percent departure of P from the long term averages using an algorithm similar to that of the PDSI.
7. Deciles First promoted by the Australian drought authorities who currently use it.	P	monthly	The Decile method groups monthly precipitation occurrences into deciles. By definition "much lower than" normal precipitation can't occur more often than 20 percent of the time.
8. Precipitation Anomaly Classification (PAC) Janowiak et al (1986).	P	monthly or yearly	The PAC is an improvement of the "Decile" Australian method of drought classification.
9. National Rainfall Index (NRI) Gommes and Petrassi (1994)	P	monthly	The NRI patterns abnormalities of precipitation on continental scale.
10. Percent of Normal (PN)	P	monthly	PN is obtained by dividing P with the normal value. It is a simple calculation well suited to the needs of TV weather people and general audiences.

Abbreviations: P – Precipitation, T – Temperature, ET – Evapotranspiration, L – Soil moisture, RO – Runoff, sn – snowpack.

Traditionally, the SPI algorithm involved the transformation of smoothed precipitation time series into standardized normal distribution variables. We demonstrated why this might not be appropriate especially since precipitation time series at 3-months scales or less are known to be typically skewed. We replaced the normal distribution in SPI algorithm with Pearson type III distribution (P3) in order to take care of the skewness characteristics of precipitation. Furthermore we introduced a regional "flood-index" method of obtaining the regional P3 curves in order to facilitate drought comparison across the regions.

Based on the information derived from the SPI drought index we regionalized East Africa into homogeneous drought zones.

3.2 Theoretical Basis of Selected Drought Indices

Among the drought indices listed in Table 3.1, we selected the PDSI, BMI and SPI for review in this investigation partly because these three are non-basin-specific indices that can be used for comparison of droughts in regions of varying climates. The theoretical background of these indices is first examined.

3.2.1 Palmer's Drought Severity Index (PDSI)

The empirical PDSI designed by Palmer (1965) is probably one of the most theoretically advanced, operational drought indices in use today (Titlow, 1987). It is also one of the few standardized indices that allows a direct comparison of index values between different climatological regions. It is also probably the most widely used drought index in the United States where it is computed weekly for 344 climatic divisions of the country and published in *The Weekly Weather and Crop Bulletin* (ISSN 0043-1974), a publication jointly prepared by the U.S. Department of Commerce, National Oceanic and Atmospheric Administration (NOAA) and the U.S. Department of Agriculture (USDA).

3.2.1.1 PDSI Algorithm

PDSI begins with the analysis of weekly or monthly soil moisture water budget. It assumes that evapotranspiration occurs close to the potential rate until a certain proportion of the available water is depleted, after which the actual evapotranspiration rate is less than the potential rate. Palmer (1965) used the following Equations to compute the moisture transfer from the first and subsequent underlying soil layers;-

$$L_s = \min\{S_s, (PE - P)\}, \quad \text{.....(1)}$$

$$L_u = \frac{\{(PE - P) - L_s\}S_u}{AWC}, L_u \leq S_u \quad \text{.....(2)}$$

where P = Precipitation

S_s = Available soil moisture stored in surface layer at the start of month

S_u = Available soil moisture stored in underlying layer(s) at the start of month

PE = Potential Evaporation for the month.

AWC = Combined available field capacity of both surface and underlying layers.

L_s = Moisture loss from the upper (surface) layer

L_u = Moisture loss from the underlying layer(s)

Palmer assumed that no runoff occurs until both layers reach field capacity. He estimated the monthly potential evaporation (PE) by the Thornthwaite (1948) method. He noted that the PE estimated by Thornthwaite method could be in error by as much as 100% or over on some individual days when compared to measured data. He found the average daily absolute error to be approximately 35%. Howeve, as the time period considered increases to about two weeks or longer, the average absolute error decreases to approximately 10 to 15%. Because of this, he suggested that the expected error of PE estimated from Thornthwaite's method was acceptable for the climatological analysis of monthly moisture requirements.

In addition, as part of water balance, Palmer also computed several other potential values such as the potential recharge (PR), potential loss (PL) and potential runoff (PRO). defined as follows:-

PR is the amount of moisture required to bring the soil to field capacity

$$PR = AWC - [S_s + S_u] \quad \text{.....(3)}$$

PL is the amount of moisture that could be lost from the soil to evapotranspiration or exfiltration during the period of zero precipitation,

$$PL = PL_s + PL_u \quad \text{.....(4)}$$

where $PL_s = \min(PE, S_s)$ .....(5)

$$PL_u = (PE - PL_s)\frac{S_u}{AWC}, \; PL_u \leq S_u \quad \text{.....(6)}$$

The potential runoff, PRO, is the potential precipitation minus potential recharge. Palmer assigned potential precipitation as being equal to AWC.

$$PRO = AWC - PR = S_s + S_u \quad \text{.....(7)}$$

Using the four potential quantities, PE, PR, PL and PRO, Palmer obtained four coefficients that were dependent on the climate of the area being analyzed.

$$\lambda_{1,j} = \frac{\overline{ET_j}}{\overline{PE_j}}, \quad \text{.....(8)}$$

$$\lambda_{2,j} = \frac{\overline{L_j}}{\overline{PL_j}} \quad \text{.....(9)}$$

$$\lambda_{3,j} = \frac{\overline{R_j}}{\overline{PR_j}}, \quad \text{.....(10)}$$

$$\lambda_{4,j} = \frac{\overline{RO_j}}{\overline{PRO_j}} \quad \text{.....(11)}$$

$j = 1, 2....12$

where $\overline{ET_j}, \overline{L_j}, \overline{R_j}$, and $\overline{RO_j}$ are monthly mean evapotranspiration, moisture loss, water recharge and runoff respectively.

The overhead bars signify that the coefficients are averages values of month j. From the above coefficients, Palmer computed the 'Climatologically Appropriate for Existing Conditions' (CAFEC) precipitation by the Equation

$$\hat{P}_{w,j} = (\lambda_{1,j}PE_{w,j} - \lambda_{2,j}PL_{w,j} + \lambda_{3,j}PR_{w,j} + \lambda_{4,j}PRO_{w,j}) \quad \text{.....(12)}$$

He then established the departure, d, between the precipitation of each month and the CAFEC value.

$$d_{w,j} = P_{w,j} - \hat{P}_{w,j} \quad \text{.....(13)}$$

where the subscripts w and j refer to year and month respectively.

In order to make temporal and spatial comparison of the depatures $d_{w,j}$ possible, Palmer (1965) introduced the concept of a moisture anomaly index Z:-

$$Z_{w,j} = K_j d_{w,j} \quad \text{.....(14)}$$

where K_j is

$$K_j = 1.5 \, \text{Log}_{10}\left(\frac{T_j + 2.8}{\overline{D}_j}\right) + 0.50 \quad \text{.....(15)}$$

where

$\overline{D}_j$ is the monthly average of the departures $d_{w,j}$ given in Equation 13.

$$T_j = \frac{(PE_j + R_j + RO_j)}{P_j + L_j} \quad \text{.....(16)}$$

T_j is a measure of the ratio of "moisture demand" to "moisture supply" for the month and region. Weighing the departures d_{wj} with the parameter K_j in Equation (14) facilitated the moisture deficit comparison among different areas and for different months. Equations (15) and (16) were derived using data from nine areas in the United States. The complexity and unusual form of the Equations 15 and 16 resulted from the difficulty Palmer (1965) had in deriving them (Alley, 1984). The monthly moisture index, from a moisture standpoint signifies the departure of the monthly weather from the "climatically normal" conditions for the month.

Palmer (1965) plotted cumulative Z versus the duration of the worst episodes of drought in his area of study. He obtained a linear relationship from which he deduced a drought severity Equation given as

$$X_j = 0.897X_{j-1} + 0.333Z_j \quad \text{.....(17)}$$

To use the above Equation, one needs to identify the starting month of a wet or a dry spell. He solved this by keeping track of three pseudo-indices X1, X2 and X3 given by

$$\left\{\begin{aligned} X1_j &= 0.897X1_{j-1} + 0.333Z_i \\ X2_j &= 0.897X2_{j-1} + 0.333Z_i \\ X3_j &= 0.897X3_{j-1} + 0.333Z_i \end{aligned}\right\} \quad \text{.....(18)}$$

which respectively represent conditions of 'wet spell becoming established', 'dry spell becoming established' and 'wet or dry spell that has become established'. X1 is restricted to nonnegative values, while X2 is the reverse. The values of X1 and X2 are set to zero when computations of Equation (18) violate these restrictions. He considered a drought to be established when $X2 \leq -1.0$, and a wet spell is considered established when $X1 \geq 1.0$. An established drought is considered to have certainly ended when the index reaches the "near normal" category which lies between –0.5 and +0.5. At this point X3 returns to zero.

The decision as to which of the three pseudo-indices becomes X depends on whether the dry/wet spells are incipient, established or ended. Often only one of the three pseudo-indices is non-zero, in which case X is set to the non-zero index. However, conflicting cases can arise and the appropriate index to use for X is not always obvious. In order to select a final value of X from X1, X2 and X3, Palmer(1965) devised a set of complicated operating rules which relied on computing X1, X2 and X3 over several months and then back tracking till a month of a known X was reached. Alley (1984) summarized these rules as follows:

(i) From a point of known established drought, assign X = X1 until X1=0.

(ii) Then assign X = X2 until X2 = 0

(iii) Repeat steps (i) and (ii) until a month is reached which already has an X value assigned to it.

(iv) If the pseudo-indices are such that the above rules cannot be conclusively used to select the X, then select the PDSI as X1 or X2 whichever has the largest absolute value, whenever X3 equals zero.

The practical calculation of X, which is the final PDSI, illustrates the complexity of the index. Because of this complexity, it is not uncommon for a time series of PDSI to exhibit large sudden changes.

3.2.1.2 Summary of the limitations of PDSI.

Several investigators such as Wilhite (1982), Alley (1984), Titlow (1987), Heddinghause and Sabol (1991) and Guttman (1998) have noted the shortcomings of PDSI:-

(a) Representative values of soil storage capacities are difficult to estimate accurately. The values used in the PDSI may be open to criticism. Doing a water balance in terms of two soil layers may not be representative enough for a location.

(b) The lumping of certain variables at monthly time periods may introduce significant errors. Precipitation and potential evaporation are often distributed throughout the month in such a way that both deficits and surplus periods could occur within a month.

(c) The assumption that runoff cannot occur unless soil moisture is at field capacity is not necessarily true. Horton overland flow could occur if rainfall intensity exceeds the rate of soil infiltration or return flow could occur in a concave slope. Also, no lag is incorporated in the Palmer model to account for the delay between the generation of excess water and the appearance of runoff.

(d) There is no justification to equate the potential precipitation with AWC.

(e) Potential evapotranspiration is estimated using the Thornthwaite method which is an approximate approach based on the temperature index and latitude only. According to this method, two locations of similar latitude and monthly temperatures would have the same PET even though they could be located on different continents, with different altitudes, land use, terrain features, etc.

(f) The PDSI is regional index developed out of the United States (Janowiak et al. 1986). It may not be justified to apply the weighing function and drought severity equations derived in the United States to other parts of the world.

(g) The abrupt switching among the pseudo indices X1, X2 and X3 as the value of PDSI has so far prevented incorporating some stochastic elements into the index.

3.2.1.3 Application of the PDSI outside the United States

Several investigators have applied the PDSI to different climates around the world. Bhalme and Mooley (1979) showed that PDSI in its original format failed to describe realistically the drought conditions in the Tropical India. However, Canceliere et al. (1996) found PDSI to be an appropriate indicator of droughts in the Mediterranean region. Briffa et al. (1994) analyzed the surface moisture variability across Europe based on PDSI. Using PDSI, Jones et al. (1996) reviewed the moisture availability of Europe simulated by the Hadley Center General Circulation Model. More recently, Scian and Donnari (1997) examined the PDSI in the semi-arid pampas region of Argentina.

Over the last three decades, little has been changed in the PDSI algorithm. Scian and Donnari (1997) used pan evaporation data instead of Thornthwaite data. Bhalme and

Mooley (1979) attempted to modify the PDSI coefficients to suit the Indian tropical conditions. Heddinghause and Sabol (1991) adjusted the operating rules for appropriating a value to X, when neither a wet nor a dry spell was known to be established. This made it possible to use the PDSI as an operational index.

Some modifications to PDSI are hereby suggested as an attempt to make it a more versatile drought-tracking index applicable to regions outside the US.

3.2.1.4 Suggested changes in the PDSI algorithm

3.2.1.4.1 Potential precipitation and Runoff

As mentioned, Palmer (1965) suggested that the potential precipitation could be taken as being equivalent to AWC. However, he felt uncomfortable with that assumption as stated that "*...it is not particularly an elegant way of handling the problem.*" He suggested that if he were to redo the analysis the computed potential runoff would be generally closer to reality if a constant value were assigned to the potential precipitation.

The value of potential rainfall is required only for the purpose of obtaining a potential runoff (PRO). The complication of determining this variable (potential rainfall) could be avoided by using an alternative method of obtaining PRO that does not require potential rainfall as an input, such as outlined below.

From Equations 4 to 6, potential loss (PL) is the amount of moisture that could be lost from the soil to evapotranspiration during periods of zero precipitation. Following the same logic, we can safely assume that potential runoff (PRO) in a given month cannot exceed the precipitation of that month. The limiting case would be when the soil is saturated and the evapotranspiration losses are negligible in comparison to the precipitation. In this case the PRO in a given month would be equivalent to the month's rainfall, and Equation 7 revised as

$$PRO_{w,j} = P_{w,j} \quad \text{.....(19)}$$

The long-term coefficient $\lambda_{4,j}$, of the actual runoff in relation to the PRO is still given in Equation 11.

Hence the CAFEC precipitation (Equation 12) could be modified as:

$$\hat{P}_{w,j} = (\lambda_{1,j}PE_{w,j} - \lambda_{2,j}PL_{w,j} + \lambda_{3,j}PR_{w,j} + \lambda_{4,j}P_{w,j}) \quad \text{.....(20)}$$

3.2.1.4.2 The Z index

Palmer (1965) introduced Equations 14 - 16 in order to make the departures from different locations and months comparable. We suggest that instead of using purely empirical transformations to force a functional relationship, a normalization of departures similar to that of Bhalme and Mooley (1980) would be sufficient to facilitate temporal and spatial comparison. We propose that the Z index in the PDSI be given as:-

$$Z_{w,j} = 100\frac{(d_{w,j} - \mu_j)}{\sigma_j} \quad \text{.....(21)}$$

$j = 1, 2, ..., 12$

where μ_j and σ_j are monthly departure means and standard deviations respectively.

3.2.1.4.3 The Drought Severity Equation

The drought severity recursive formula (Equation 17) is a critical part of the PDSI algorithm. However, the constants in Equation 17 do vary with location because the progression of droughts varies from place to place. Though several investigators (such as Scian and Donnari, 1997; Cancelliere et al, 1996), have used the original drought severity equation in areas other than the US with some success, we feel that they could have done better if they had changed the coefficients of the severity equation to better reflect the local conditions. To demonstrate the need to change the coefficients, we hereby formally

derive Equation 17. The plot of the cumulative Z for the worst drought periods of various durations often gives a straight line as Palmer (1965) indicated (see Figure 3.1).

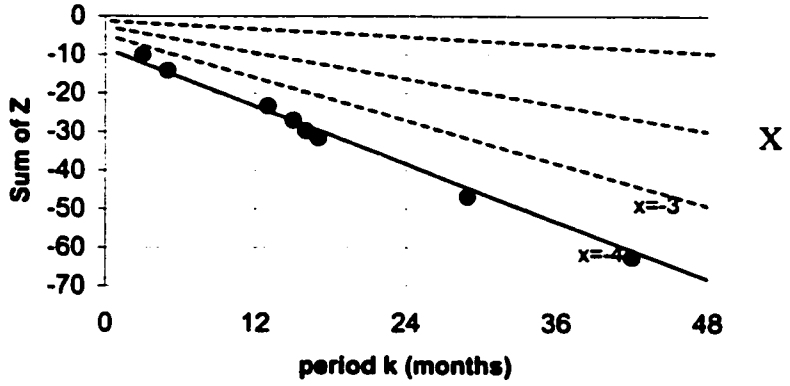


Figure 3.1 Plot of cumulative Z vs. time t for worst drought episodes using a set of Iowa and Western Kansas data obtained by Palmer(1965). X is the drought severity.

In the example shown in Figure 3.1, the driest 42-month period had a cumulative Z of -61. The solid line represents extreme drought. Along this line the drought severity X is arbitrarily fixed at –4. The vertical distance (ordinate) between this extreme drought line and the t-axis at Z=0 is divided into four equal lengths by drawing three dashed lines representing severe, moderate and mild droughts. It can be shown that the Equation of the extreme drought line is given by

$$X_k = \frac{\sum_{t=1}^{k} Z_t}{(a + kb)} \qquad \text{.....(22)}$$

where X_k is the drought intensity of the k^{th} month

$\sum_{t=1}^{k} Z_t$ is the accumulated moisture index over duration k

a = -c/4 and b = -m/4 where c and m are the intercept and slope of the extreme drought line.

Equation (22) provides only a partial drought severity expression because it is based on the cumulative sum of Z. As Palmer (1965) and Bhalme and Mooley (1980) pointed out, the cumulative procedure of accounting for duration of the dry period can be misleading.

Apparently according to Equation (22) the drought intensity X_k in part depends on the order of the month, k, whose index is being computed. For example, the same moisture index Z_t could have a different contribution to the drought index depending on whether it is the 2nd or 6th month in the computation sequence. To solve this problem, it is necessary to revise Equation (22) such that the incremental drought intensity contribution for each successive month is evaluated independent of the order k, in which it occurred.

For the initial month, let $X_o = 0$. The contribution of the next successive month can be obtained by setting k=1 in Equation 22 to obtain

$$X_1 = \frac{Z_1}{(a+b)} \quad \text{.....(23)}$$

Then the change in X is given by

$$X_1 - X_0 = \Delta X = \frac{Z_1}{(a+b)} \quad \text{.....(24)}$$

In successive months a negative value of Z will be required to maintain the existing dry spell. The magnitude of this Z that is required depends on the magnitude of the previous drought index X.

Hence

$$\Delta X_k = \frac{Z_k}{(a+b)} + \theta X_{k-1} \quad \text{.....(25)}$$

where

$$\Delta X_k = X_k - X_{k-1} \quad \text{.....(26)}$$

It should be noted that in Equation 25, the Z index of each successive month contributes to the previous drought index on an incremental basis and is independent of k. To determine the constant θ we use Equation 22 for t = k-1 and k, with the drought index X_k and X_{k-1} kept at some constant value ϕ. On solving for Z_k we obtain the Equation

$$Z_k = \sum_{t=1}^{k} Z_t - \sum_{t=1}^{k-1} Z_t = b\phi \quad \text{.....(27)}$$

Using Equation 25 and 27 we can then solve for the constant θ.

$$\theta = \frac{-b}{(a+b)} \quad \text{.....(28)}$$

When we substitute for θ and ΔX_k in Equation 25, we obtain (after re-arranging),

$$X_k = \left(\frac{a}{a+b}\right) X_{k-1} + \frac{Z_k}{(a+b)} \quad \text{.....(29)}$$

and since a = -c/4 and b = -m/4 with c and m as the intercept and slope of Figure 3.1 then

$$X_k = \left(\frac{c}{c+m}\right) X_{k-1} + \frac{-4Z_k}{(c+m)} \quad \text{.....(30)}$$

Equations 29 or 30 are the drought severity relationships that are used recursively to track drought conditions in the PDSI. In Palmer's (1965) analysis a and b were 2.691 and 0.309 respectively. There is little basis to use the above values for regions different from those Palmer (1965) investigated. Analysis of East African data has shown that different zones could have diverse drought progression characteristics and hence different regression coefficients in a plot of cumulative Z vs. extreme drought length period t. Figure 3.2 illustrates the difference between the extreme drought line as proposed by Palmer (1965) and the extreme drought lines for selected East African stations.

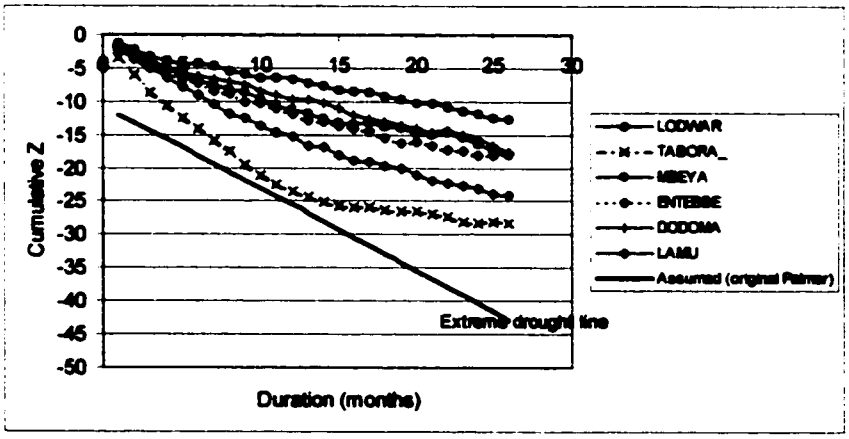


Figure 3.2 Plot of most negative cumulative moisture index Z vs. duration for periods ranging from two to thirty months in selected East African stations.

Equations 17, 29 and 30 can all be represented in the form

$$X_k = \varphi X_{k-1} + \varepsilon Z_k \quad \text{.....(31)}$$

where ϕ is c/(c+m), ϵ is –4/(c+m) and Z is the moisture index. The coefficients "c" and "m" are the intercept and slope of the extreme drought line as illustrated in Figures 3.1 and 3.2. Equation (31) can be readily identified as a first order autoregressive model (AR1). However, PDSI is not a pure AR(1) process because the final value X at times abruptly shifts between pseudo indices X1, X2 and X3 according to some predefined rules. The coefficients of regression determine the magnitude of the coefficients of the drought severity index (Equation 30). If the intercept c is much less than the sum of the intercept and the slope (c+m), then the coefficient of the first term in Equation (31) would be small. This would mean that for the particular region, the previous conditions (or the previous month drought index) carries less weight when computing the present index. It is reasonable to assume that the magnitude of these coefficients and hence progression of a dry or wet spell will vary from place to place. The variation of these coefficients for several East African stations is shown in Table 3.2 alongside the values established by Palmer (1965).

In most cases, the plot of Z vs. duration is almost linear as shown by the consistently high correlation values in Table 3.2. It is noted that even though ϕ does not change much, all the computed values are less than the value established by Palmer (1965). The computed ϵ parameter varies significantly from what Palmer (1965) established for North America. In most of the cases ϵ is four or five times the original Palmer parameter of 0.333. Application of Equation (31) with varying ϕ and ϵ parameters to the same moisture index time series (Z_k, k = 1, 2, ..., n) will produce different PDSI time series. However, even though ϕ and ϵ are obtained locally from the individual station extreme-drought-line (see Figures 3.1 and 3.2), the resulting PDSI time series should theoretically be comparable across stations because all are calibrated relative to the driest conditions existing at each individual location.

Table 3.2 Examples of coefficients of regression (c and m) and resulting drought severity terms for several stations compared with the original PDSI terms.

Station Name	Latitude	Longitude	Intercept c	Slope m	Correlation ρ	Eqn. (31) parameters	
						ϕ	ϵ
Lodwar	3.12	36	-1.899	-0.412	-0.992	0.822	1.731
Mandera	3.93	42	-1.867	-0.565	-0.998	0.768	1.645
Wajir	1.75	40	-1.491	-0.443	-0.996	0.771	2.068
Kisumu	-0.10	35	-3.508	-0.576	-0.989	0.859	0.979
Garissa	-0.47	40	-1.797	-0.435	-0.993	0.805	1.792
Lamu	-2.27	41	-3.782	-0.861	-0.983	0.815	0.861
Mombasa	-4.03	40	-2.728	-0.417	-0.989	0.868	1.272
Kigoma	-4.88	30	-3.95	-0.812	-0.991	0.830	0.84
Dodoma	-6.17	36	-2.384	-0.551	-0.995	0.812	1.363
Mtwara	-10.27	40	-4.739	-0.642	-0.971	0.881	0.743
Original PDSI			**-10.764**	**-1.236**	-	**0.897**	**0.333**

Thus, in the modified PDSI algorithm, we propose that the ϕ and ϵ parameters in Equation 31 be obtained locally for each station. There is little basis for using the same ϕ and ϵ values for all stations unless all the stations exhibit the same extreme drought characteristics, which is clearly not the case in East Africa.

3.2.2 Bhalme Mooley Index

Bhalme and Mooley (1980) developed a numerical drought index for assessing the drought intensity using precipitation only. The computational details of this index and PDSI are generally similar with a few differences. Bhalme Mooley substituted the moisture index Z (which is obtained from surface moisture balance computations) in Palmer's algorithm with a simpler monthly rain index M which is computed from the rainfall data only. Furthermore, the Bhalme Mooley index does not involve the

simultaneous tracking of pseudo indices of dry, wet or un established spells. This makes it a much easier index to compute and than Palmer's PDSI.

To obtain the Bhalme and Mooley index, first compute the long term monthly rainfall mean (μ), standard deviation (σ) and coefficient of variation (CV) and then the rainfall anomaly which is the normalized departure from the long term mean. The moisture index (M) is normalized rainfall anomaly divided by the CV.
Hence,

$$M = 100\frac{(x-\mu)}{\mu}.\frac{\mu}{\sigma}$$

$$= 100\frac{(x-\mu)}{\sigma} \quad \text{.....(32)}$$

Bhalme and Mooley (1980) noted that comparison of the moisture index (M) for different locations and months was acceptable 'within reasonable limits'.

The drought severity equation is obtained using the same arguments as Palmer (1965) did from the plot of accumulated monthly index M vs. period (months); i.e. Equations 22-30 hold with M being used instead of Z.

3.2.3 Standardized Precipitation Index (SPI).

The understanding that a deficit of precipitation has different impacts on the groundwater, reservoir storage, soil moisture, snowpack and stream flow led McKee et al. (1993) to develop the SPI. The purpose of the SPI is to assign a single numeric value to precipitation which can be compared across regions with markedly different climates. McKee et al. (1993) defined SPI as the number of standard deviations that the observed cumulative rainfall at a given scale would deviate from the long-term mean. Since the cumulative precipitation may not be normally distributed, a transformation is first applied to the data so that the transformed precipitation approximately follows a normal distribution.

The SPI index was designed to explicitly express the fact that it is possible to simultaneously experience wet conditions on one or more time scales, and dry conditions at other time scales. For example, soil moisture conditions which typically respond to precipitation on a relatively short scale may get quickly depleted in a brief drought spell while stream flow, ground water and reservoir storage which is affected by longer term precipitation anomalies may still be relatively normal.

The Colorado Climate Center, the Western Regional Climate Center and the National Drought Mitigation Center use SPI to monitor the current state of drought in the United States (Edwards and McKee, 1997).

The SPI calculation is based on the long-term precipitation record for a desired period. To compute SPI, a monthly precipitation time series is "smoothened" with a moving window of width equal to the time scale desired. For example, a 3-month SPI would use a moving window of a 3-month width. In their study, Edwards and McKee (1997) selected a 3-month SPI for a short-term or seasonal drought index, a 12-month SPI for an intermediate- drought index and a 48-month SPI for a long-term drought index. The window is *non-centered* in the sense that its first coefficient is the lag 0 coefficient so that the filtered series depends only on past and present values of the original series. For example, if a 3-month window is used, the new smoothed series $x'_t, t = 1,2,\cdots,n$ are given by.

$$x'_t = \frac{1}{3}\sum_{i=0}^{i=2} x_{t-i} \qquad \text{.....(33)}$$

The smoothed data is broken into 12 time series each containing data for a given month for all the available years. A probability distribution is fitted to each of these 12 sets of time series. For this purpose McKee et al (1993) used the gamma distribution which is useful for describing skewed hydrologic variables (such as precipitation) without the need for log-transformation (Chow et al. 1988). Its probability density function is defined as,

$$g(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} \quad \text{for } x > 0. \qquad \text{.....(34)}$$

where $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1}e^{-y}dy$ (the gamma function), and α and β are the shape and scale parameters respectively.

However, it should be noted that distributions other than the gamma can also be used, as long as they fit the data adequately.

The maximum likelihood estimators for the parameters α and β are (McKee and Edwards, 1997)

$$\hat{\alpha} = \frac{1}{4A}\left(1 + \sqrt{1 + \frac{4A}{3}}\right) \quad \text{.....(35)}$$

$$\hat{\beta} = \frac{\overline{x}}{\hat{\alpha}} \quad \text{.....(36)}$$

where: $A = \ln(\overline{x}) - \frac{\sum \ln(x)}{n}$

n = number of observations.

From the parameters α and β, the cumulative probability G(x) corresponding to an observed precipitation is estimated as

$$G(x) = \int_0^x g(x)dx \equiv \frac{1}{\hat{\beta}^{\alpha}\Gamma(\hat{\alpha})}\int_0^x x^{\hat{\alpha}-1}e^{-x/\hat{\beta}}dx \quad \text{.....(37)}$$

substituting $t = x/\hat{\beta}$,

$$G(x) = \frac{1}{\Gamma(\hat{\alpha})}\int_0^x t^{\hat{\alpha}-1}e^{-t}dt \quad \text{.....(38)}$$

The incomplete gamma function of Equation 37 was solved using a statistical package called SPLUS.

The gamma function is undefined for x = 0 yet it is possible to get months with no precipitation. To take care of this, Edwards and McKee (1997) suggested that the actual probability of non-exceedance H(x) should be given by

$$H(x) = q + (1 - q)G(x) \quad \text{.....(39)}$$

where q is the probability of x = 0. If *m* is the number of zeros in a particular precipitation time series, q can be estimated by *m/n* where *n* is the sample size (Edwards and McKee, 1997). The inverse normal (Gaussian) function is applied to the non-exceedance probability H(x) to obtain the SPI.

From our experience, the smoothing of the precipitation time series will often eliminate or greatly decrease the number of months with zero precipitation. For example, when analyzing six-month SPI, having a zero in the smoothed time series is only possible if there are six consecutive months with zero precipitation. This is generally a rare possibility for our study area. Hence the problem of zero values seldom arises in our analysis, and the value of q in Equation 36 is set to zero.

The transformation from the non-exceedance probability H(x) to the SPI is an equiprobability transformation whereby a variate is transformed from one distribution (i.e. Gamma) to another variate with a distribution of a prescribed form (Gaussian normal) such that both the new and old variable have the same probability of non-exceedance in their respective distributions.

In the example illustrated in Figure 3.3, a 3-month precipitation amount (March, April and May) is converted to an SPI value with mean of zero and variance of one. The dotted line on the left side of Figure 3.3 denotes the empirical cumulative probability of the actual three-month precipitation totals, while the continuous smooth curve denotes the fitted gamma distribution. The curve on the right hand side of Figure 3.3 is a cumulative curve distribution of the standard normal distribution. In this example MAM precipitation of 300 mm corresponds to an SPI value of 0.39. It should be noted that both values, 300 mm and 0.39, have the same non-exceedance probability of 0.65.

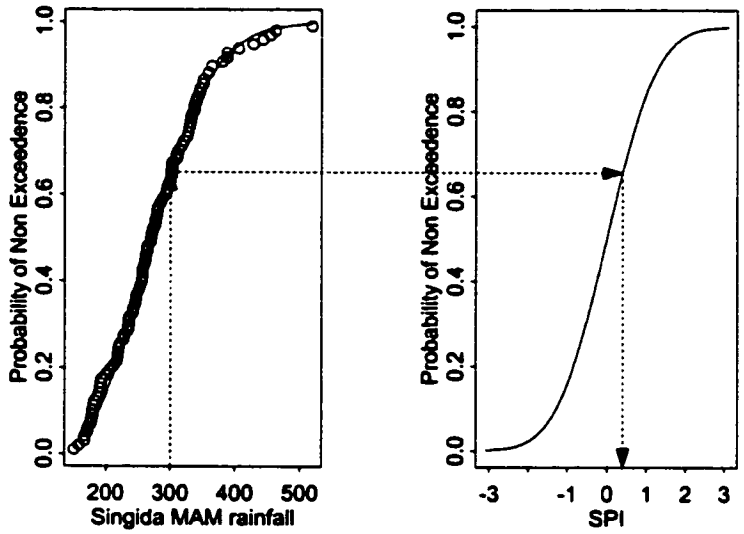


Figure 3.3 The SPI of Singida March - May precipitation based on 1900-1996 data.

Figure 3.4 shows 6, 12 and 48-month SPI of Singida, a location in Central Tanzania. It is evident from this figure that the frequency, duration and intensity of drought at a given location is dependent on the time scale. The 48-month SPI curve is relatively smoother than the 12-month SPI curve, while the 6-month SPI is relatively rugged.

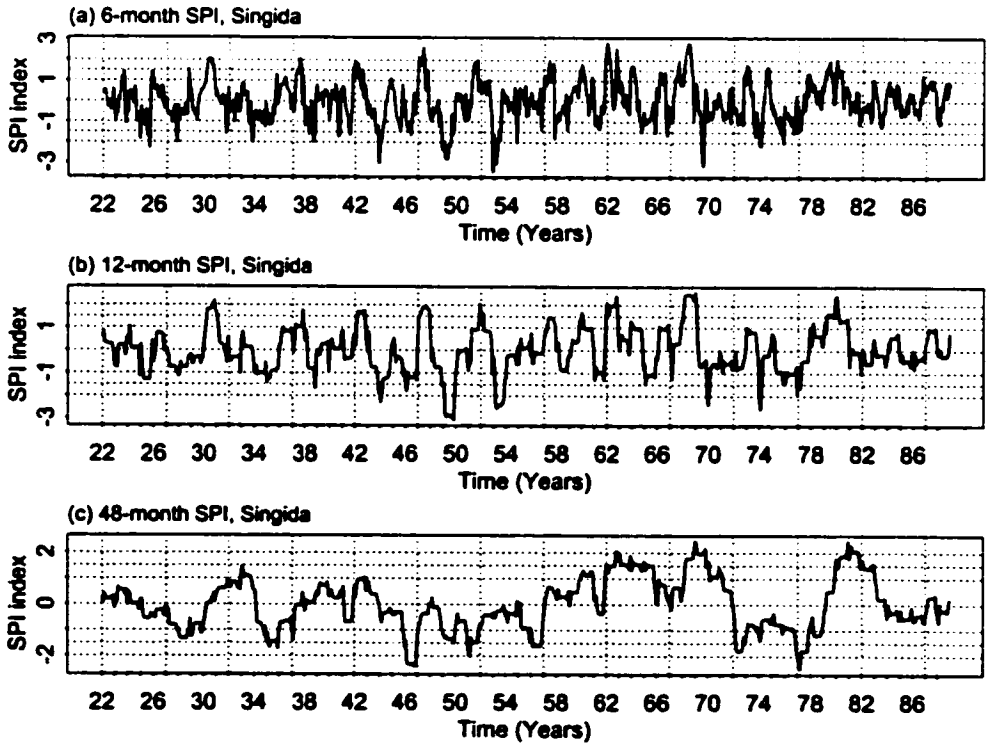


Figure 3.4 Time Series of 6, 12 and 48-month SPI of Singida, Central Tanzania.

The SPI values for a given location can vary significantly as much as 3 SPI units when computed at different time scales. For example, in July and August 1969, the 6-month SPI values for Singida were –2.7 and –3.0 respectively indicating a severe drought while the 12-month SPI values for the same location were 0.7 and 0.7 in both months, indicating near normal conditions. The 48-month SPI for the two months were 1.95, indicating a very wet period.

McKee et al. (1993) used the classification system shown in Table 3.3 to assign drought categories with respect to SPI. McKee et al. (1993) also defined the criteria for a drought event for various time scales. A drought event is considered to have occurred any time the SPI is continuously negative and reaches an intensity where the SPI is –1.0 or less. The event ends when the SPI becomes positive. Therefore, each drought event has a well defined duration.

Table 3.3 Classification of SPI values (McKee et al, 1993)

SPI Values	Drought Category
2.0 and above	Extremely wet
1.5 to 1.99	Very wet
1.0 to 1.49	Moderately wet
-0.99 to 0.99	Near normal
-1.0 to –1.49	Moderately dry
-1.5 to –1.99	Severely dry
-2.0 and less	Extremely dry

3.2.3.1 Suggested Improvements in the SPI algorithm.

The SPI algorithm has not been revised ever since it was introduced by McKee (1993). There are several properties of the index that merit some revision in order to make it a better index. These briefly are discussed below.

There is still disagreement regarding the type of probability distribution that should be fitted to the precipitation data. A probability distribution is needed to fit to the precipitation data because the SPI algorithm requires the exceedance probability of each

data point in the smoothed precipitation time-series. The most commonly used distribution has been the gamma distribution. It should be noted that, index values corresponding to a given precipitation amount may vary depending on the chosen distribution chosen to fit to the data.

We propose using a suitable plotting position formula to estimate the empirical exceedance probabilities. This is more preferable if the data set is reasonably long, say more than 90 years of data so that we can estimate exceedance probabilities ranging from 0.01 to 0.99. A plotting position is said to be suitable for a given distribution if the estimates of both the quantiles and non-exceedance probabilities obtained using the plotting formula are unbiased (Hydrology Handbook, 1993). There are several popular plotting position formulae in hydrology, which attempt to achieve an almost quantile-unbiased fit for different distributions (Hydrology Handbook, 1993). In general, most plotting position formulae can be written as

$$q_i = \frac{i - a}{n + 1 - 2a} \qquad \text{.....(40)}$$

where i is the rank order and n is the total length of the series. Cunnane (1978) discussed the application and characteristics of these formulae and recommended a=0.40 which proved to be suitable for estimating unbiased quantiles for a wide range of distributions. The most popular plotting position is the Weibull plotting formula (a =0 in Equation 40) which has been shown to be suitable for estimating unbiased exceedance probabilities for all distributions (Hydrology Handbook, 1993). Cunnane's formula is suitable for skewed distributions while Weibull's formula is suitable for Uniform distributions.

The estimation of non-exceedance probability using plotting positions formula is a non-parametric procedure since only data ranks are used. Therefore, by using a plotting position formula to estimate the non-exceedance probabilities of the 12 smoothed sub-precipitation series, we transform the SPI algorithm into a non-parametric SPI index. (Each of the 12 sub-series corresponds to a unique combination of lunar months e.g. for

6-month SPI you could have January to June, or February to July. There are always twelve such sub-series regardless of the length of the smoothing window.) Figure 3.5 shows a comparison of SPI obtained by parametric and non-parametric methods for Singida station. The latter is obtained by using Cunnane's formula to estimate the probabilities of non-exceedance. Non-parametric methods have attractive properties – the most obvious being their ability to handle disproportionate outliers.

Figure 3.5 shows that there is hardly any difference between the indices obtained by the two respective methods except for very extreme events. There are noticeable differences at extreme events because the Weibull formula cannot estimate exceedance probabilities less than 1/(n+1) or greater than n/(n+1) whereas one can estimate exceedance probabilities outside the above limits using a parametric distribution. In any case the actual exceedance probability associated with the least and largest observations are random variable with mean 1/(n+1) and n/(n+1) respectively and a standard deviation 1/(n+1) (Hydrology Handbook, 1993). However, once the SPI has exceeded the extreme threshold of –2.0 (corresponding to extreme severe drought), it is no longer relevant to know how far it dips further since all are extremely severe droughts anyway.

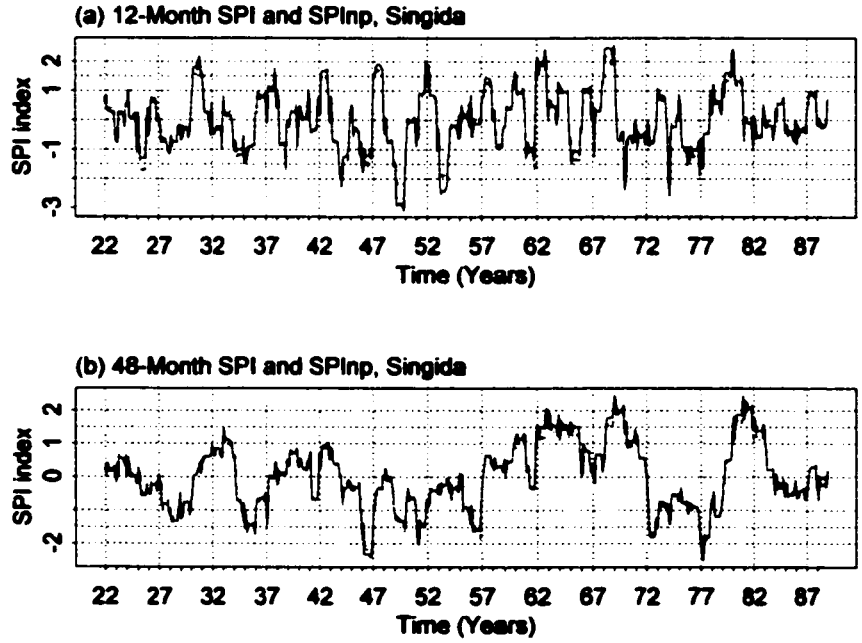


Figure 3.5 Time series of the (a) 12-month and (b) 48-month SPI and nonparametric SPI (SPInp) of Singida. The red dotted line is the non-parametric SPI.

Another point of disagreement in the SPI methodology is the length of the base period used in calibrating the distribution functions. Edwards and McKee (1997) used a fixed base of 30 years to fit the necessary gamma parameters. This ensured that the historical SPI values did not change with time.

We propose using a flexible time base to fix the probability distributions. Although the SPI values may change with time, the change is not so dramatic so as to jump from drought to wetness.

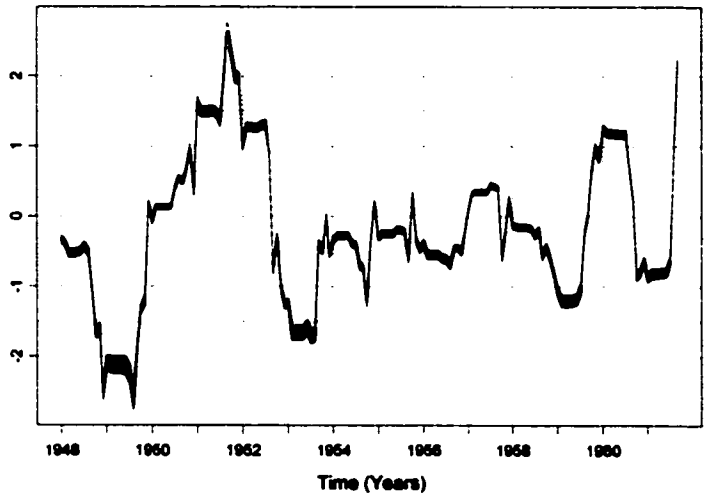


Figure 3.6 Thirty six (36) virtually indistinguishable time series of 12-month SPI for Singida (1945 – 1960) obtained using a calibration period varying from 1900-1961 to 1900 – 1996.

Figure 3.6 shows the minor variations in SPI, which occur as the length of the calibrating period grows from 62 to 97 years. Apparently, at least for East African data, changing the calibration period does not lead to significant change in the level of drought severity.

Drought is a relative phenomenon. A drought index ought to be normalized with respect to time if it is to provide a meaningful estimate of drought. The dynamic nature of an SPI index based on a flexible calibration period may be better suited to capture the subtle changes of the drought properties such as the relative drought intensity. The intensity of a

past drought event may increase (decrease) relative to the average conditions if the general climatic trend is shifting towards wetness (dryness).

At the core of the SPI algorithm is the transformation of the non-exceedance probability of occurrence of smoothed precipitation totals to the normal distribution domain. The original SPI represents a Z-score, or the number of standard deviations above or below that an event is from the mean. This assumption of normality is not true for SPI of short time scales of analysis. Rainfall aggregate series composed of 6-month totals or less are usually positively skewed. The degree of skewness generally increases as the climate becomes drier (Glantz and Katz 1977). Transforming such positively skewed variables to the normal distribution domain will cause undesirable distortion, especially in the right tail of the distributions. We propose to work around this problem by transforming the variables using an appropriate regional growth curve instead of the simple normal distribution.

The regional growth curves are developed using the regional index-flood procedure (Hosking and Wallis, 1996). The term "index flood" arose because early applications of the procedure were to flood data in hydrology, but the method is also applicable to drought analysis. A brief description of the procedure is provided below. More detailed descriptions of the method can be found in the National Environment Research Council (1975) or Handbook of Hydrology (1993).

Suppose that the data are available at N sites, where the i^{th} site has a sample size n_i, and observed data P_{ij}, $j = 1,\ldots, n_i$. Let $P_i(F)$, $0<F<1$, be the quantile function of the frequency distribution at site i. We may then re-write

$$P_i(F) = \mu_i q(F), \quad i=1,\ldots,N \qquad (41)$$

where μ_i is the index flood which is the mean of the at-site frequency distribution, and q(F) is the regional growth curve, a dimensionless quantile function which is common at

every site. Conversely, q(F) is the quantile function of the regional frequency distribution of the dimensionless, rescaled data, $P_{ij}/\mu_i, j = 1,\cdots,n_i, i = 1,\cdots,N$.

The index-flood procedure is premised on a number of assumptions. The method assumes that observations at any given site are identically distributed, and are serially independent. Furthermore, observations at different sites should be independent. The frequency distributions at different sites should be identical apart from a scale factor. All these assumptions, especially the last one, are only approximately attained by hydrologic data such as rainfall.

Since it is desirable that the modified SPI be comparable across the entire East African region, we applied the regional analysis to the whole of East Africa as one region. After smoothing the monthly data series with a window representing the scale of interest, the individual grid data was divided by the respective grid means to obtain the dimensionless grid-based, variables. A distribution was then fitted to the rescaled data using L-moments. We used L-moments instead of the traditional method of moments because the latter have been shown to have some undesirable properties (Wallis et al., 1974; Dalén, 1987; Hosking and Wallis, 1997). It has been shown that the ordinary moment estimators of the skewness and kurtosis have algebraic bounds that depend on the sample size. For a sample of size *n* the bounds of skewness are (Hosking and Wallis, 1997)

$$|\gamma| \le n^{\frac{1}{2}} \quad \text{.....(42)}$$

For example, the skewness of a two parameter lognormal distribution is given by

$$\gamma = 3[e^{\sigma_y^2} - 1]^{\frac{1}{2}} + [e^{\sigma_y^2} - 1]^{\frac{3}{2}} \quad \text{.....(43)}$$

where σ_y is the variance. Thus, when σ_y=1, the skewness of the two parameter lognormal distribution is 6.18. However using ordinary moment estimators to estimate the skewness of a sample of size n=20 drawn from this distribution will have an upper bound of 4.47 or 65% of the population value. L-moments do not suffer from such limitations (Hosking and Wallis, 1997).

It was necessary to establish whether it was reasonable to apply regional analysis to the whole of East Africa as one homogenous zone. This was done by assessing whether the at-site (grid) variation of the sample L-moments was consistent with what would be expected of a homogenous region. This homogeneity test is described in Hosking and Wallis, (1997) and Hosking (2000). Using the software provided by Hosking (2000), we found that considering East Africa as one region did approximately satisfy the homogeneity condition that the grids' frequency distribution for the smoothed, aggregated monthly rainfalls were identical apart from a grid-specific scale factor.

Five three-parameter distributions (generalized logistic, generalized extreme value, generalized Pareto, Pearson type III and Wakeby) were fitted to the smoothed aggregated East African monthly rainfall data using the FORTRAN software provided by Hosking (2000). We found the Pearson type III (P3) to be acceptable to 30 out of the 31 grid points. There is considerable acceptance and application of the P3 distribution or its variant, the Log Pearson type III (LP3) distribution in water resources investigations (Vogel and McMartin, 1991). P3 can assume a wide range of distribution shapes including the Gamma and Normal distribution that can be viewed as special cases of P3, given as:-

$$f(x) = \frac{|\beta|}{\Gamma(x)}[\beta(x-m)]^{\lambda-1}e^{-\beta(x-m)} \quad \text{.....(44)}$$

where β, λ and m are the scale, shape and location parameters respectively. When $\beta>0$, x is positively skewed with m as the lower bound, i.e. $m \leq x \leq +\infty$. Similarly, m is the upper bound of a negatively skewed P3 random variable. Most precipitation time series are positively skewed.

The parameters β, λ and m are related to the first three moments of the random variable x as follows:

$$\mu = m + \frac{\lambda}{\beta} \quad \text{.....(45)}$$

$$\sigma^2 = \frac{\lambda}{\beta^2} \quad \text{.....(46)}$$

$$\gamma = \frac{2\beta}{|\beta|\lambda^{1/2}} \quad \text{.....(47)}$$

When $\beta > 0$ and m = 0, the P3 distribution reduces to the gamma distribution. According to Equation (47), as the shape parameter λ goes to infinity, the skewness coefficient γ goes to zero, in the process converging to a normal distribution. For $\lambda = 1$, and $\gamma = 2$, P3 becomes the two parameter exponential distribution. The parameters of the P3 regional distribution curves, estimated using L –moments, are shown in Table 3.4. A sample of the regional curves are shown in Figure 3.7

Table 3.4 The Regional P3 distribution parameters obtained from the precipitation grid totals of different durations.

Scale of Analysis (months)	Standard deviation σ	Skewness γ	Shape parameter, λ
2	0.56	1.22	2.686
3	0.46	1.02	3.820
6	0.30	0.71	8.007
9	0.24	0.57	12.495
12	0.20	0.51	15.561
16	0.18	0.49	16.817
24	0.15	0.47	17.773
36	0.12	0.44	20.624

As expected, the standard deviation of the regional curves decrease as the scale of analysis is increased. The scaling factor μ used in Equation 45 is equivalent to the at-site mean. For a long duration, say 36 months, the mean value of 36 consecutive months of rainfall totals is relatively big. When the smoothed series are divided by this big number, the resulting series are bound to be small, and the spread about the mean is also small.

Table 3.4 shows the gradual increase of the shape parameter λ as the scale of analysis is increased. A normal distribution has a $\lambda = \infty$. It would seem that monthly rainfall series which are smoothed by a large enough window (> 24) approximate well to the normal distribution. This is in consonance with the Central Limit Theorem, which states that the sample mean of identically distributed variables is approximately normal (Larsen and

Marx, 1981). What a large smoothing window does is to create a new series composed of consecutive sample means, the sample size being equivalent to the size of the smoothing window. The larger the window size, the more approximately normal the sample means will be.

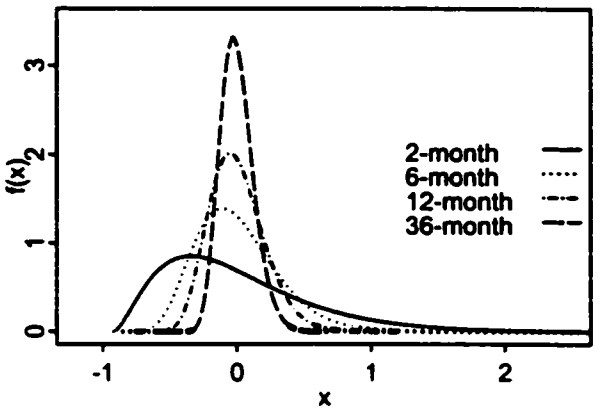


Figure 3.7 Regional P3 probability density functions used to develop the modified SPI.

From Figure 3.7, we observed that the regional probability distribution function for short time scales of analysis, say 6-months or less are clearly positively skewed. The use of normal distribution ought to be discouraged for such skewed variables because it would introduce unnecessary distortion at the tail ends.

For the P3 distributions, a quantile-unbiased plotting position to be used to estimate quantiles depends on the shape parameter β and hence skewness in addition to the rank i and sample size n. Therefore, Equation 40 cannot be used to obtain P3 plotting positions. Nguyen et al. (1989) developed an approximate plotting position for the P3 distribution given as follows:

$$q_i = \frac{i - 0.42}{n + 0.3\gamma + 0.05} \quad \text{.....(48)}$$

where γ is the skewness. This formula is suitable for skewness in the range $-3 \le \gamma \le 3$ and samples size in the range of $5 \le n \le 100$. These ranges are appropriate for our analysis and hence Equation 48 was used to obtain the P3 plotting positions that were then transformed into the standardized P3 variates.

The use of Equation 48 requires the estimation of the skewness coefficient γ. The method of moment estimator of the skew coefficient is usually defined as

$$\hat{\gamma} = \frac{1}{s^3}\left[\sum_{i=1}^{n} x_i^3 / n - 3\bar{x}s^2 - \bar{x}^3\right] \quad \text{.....(49-a)}$$

where $s = \left[\sum_{i=1}^{n} x_i / n - \bar{x}^2\right]^{\frac{1}{2}}$ ..... (49-b)

Several investigators (e.g. Bobee and Robitaille, 1975; Tasker and Stendinger, 1986) have shown that Equation 49 is often an unsuitable skewness estimator because of problems associated with small-sample bias. Bobee and Robitaille (1975) proposed an unbiased skewness estimator given as

$$\gamma_u = \bar{\gamma}\left[\left(1 + \frac{6.51}{n} + \frac{20.2}{n^2}\right) + \left(\frac{1.48}{n} + \frac{6.77}{n^2}\right)\bar{\gamma}^2\right] \quad \text{.....(50)}$$

where $\bar{\gamma}$ is the mean of the sample distribution of the sample skewness for a sample of size n from a P3 distribution (usually replaced by $\hat{\gamma}$ since only one sample is typically available).

Figure 3.8 shows an example of the original 6-month SPI compared with the modified 6-month SPI for the stations of Tabora and Singida in Central Tanzania. As expected, differences between the original and modified SPI mainly shows in the extreme values since the distribution resulted from approximating a P3 with a normal distribution increases towards the tails. Typically, the P3 based SPI will have smaller negative extremes than the normal-distribution-based SPI. This is especially so for short scale SPI series say 6-month SPI as shown above. For example, the original (normal distribution based) 6-month SPI for Tabora was –2.87 in May 1949. During that same dry month, the 6-month modified (P3 based) SPI is computed to be –1.86 standard deviations from the mean. It is likely that the modified SPI gives a more realistic picture since the 6-month precipitation totals are positively skewed.

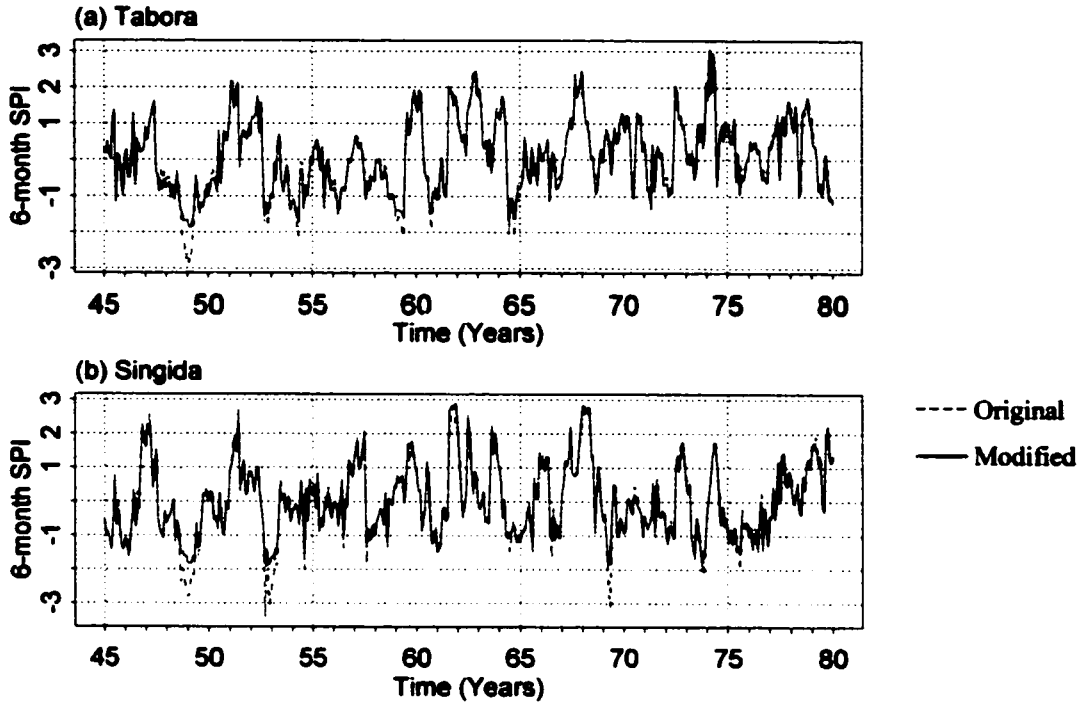


Figure 3.8 Illustration of the modified and original SPI for (a) Tabora and (b) Singida, both located in Tanzania, East Africa.

McKee et al. (1993) arbitrarily chose "clean" SPI threshold values ranging from 2.0 to –2 in steps of 0.5 to assign the severity of drought categories drought categories (Table 3.3). For example, he associated an SPI value between –1 and 1 with normal or near normal conditions. This would mean that precipitation totals with the probability of exceedance ranging form 0.16 to 0.84 are associated with normal conditions. Such a drought classification scheme cannot work with the modified P3 based SPI index because standardized P3 variates corresponding to a given probability vary in magnitude according to the shape factor of the regional distribution curve. In the modified index, each scale of analysis should have its own classification thresholds depending on the shape of regional distribution governing the precipitation totals.

We propose a new drought classification table based on probabilities of non-exceedance shown in Table 3.5. This table assumes that the extreme wet and dry events are those with the non-exceedance probabilities of 0.98 and 0.02 respectively.

There is a noticeable difference in the threshold of the extreme moisture conditions. The number of standard deviations less than the mean required to signal an extreme drought condition (with a non exceedance probability of 0.02) progressively increase from 1.37 for the 2-month SPI to 1.81 for 36-month SPI. On the other hand the number of standard deviations (from the mean) required to reach extremely wet conditions decrease from 2.63 to 2.28 for the same respective SPI indices. It would seem that if we increase the scale of analysis n in n-month SPI index, say n = 60, we would eventually approach the normal-distribution threshold values shown in Table 3.5. However, it probably doesn't make sense to use a 60-month SPI in a drought analysis.

Table 3.5 SPI-Drought classification based on probabilities of exceedance.

Drought Classification	Probability of non-exceedance	Pearson Type 3 based SPI				Normal dist. based SPI
		2-month	6-month	12-month	36-month	
Extremely wet	> 0.98	> 2.63	2.41	2.31	2.28	2.05
Very wet	0.95 – 0.9799	1.91 – 2.63	1.82 – 2.41	1.78 – 2.31	1.76 – 2.28	1.64 – 2.05
Moderately wet	0.80 – 0.9499	0.73 - 1.91	0.79 – 1.82	0.81 – 1.78	0.81 – 1.76	0.84 – 1.64
Near normal	0.20 – 0.7999	-0.84 - 0.73	-0.86 – 0.79	-0.86 – 0.81	-0.86 – 0.81	-0.84 – 0.84
Moderately dry	0.05 – 0.1999	-1.24 - -0.84	-1.42 - -0.86	-1.49 - -0.86	-1.51 - -0.86	-1.64 - -0.84
Very dry	0.02 – 0.0499	-1.37 - -1.24	-1.66 - -1.42	-1.77 - -1.49	-1.81 - -1.51	-2.05 - -1.64
Extremely dry	< 0.02	< -1.37	< -1.66	< -1.77	< -1.81	< -2.05

The differences highlighted in Table 3.5 arise from the skewed nature of the precipitation totals, which gradually decreases as the size of the smoothing window is increased. However, these subtle differences between the original and the new P3-based SPI index might be quite important especially where the SPI index variables are to be processed further.

3.3 Inter-Comparison of the Drought Indices

Although there is no one drought index that is inherently superior than others in all circumstances, some indices are better suited than others for certain regional applications.

It is the purpose of this section to explore which index might have an edge over the rest for monitoring East African drought.

3.3.1 Original and Modified PDSI

A sample output from the original and modified PDSI algorithms for three locations is shown in Figure 3.7. In general it was observed that the original Palmer algorithm did not produce a realistic index for the drier northern East African locations such as Lodwar (Figure 3.7). For Lodwar, the unmodified method produced an index, which portrayed an erroneous scenario that this place suffered from continuous drought starting from 1960 right up to 1985. The modified method corrected this problem.

Analysis of the detailed output of the original PDSI algorithm showed that R (recharge), and RO (runoff) were typically zero in dry areas such as Lodwar. Unlike in the wetter areas, the potential loss (PL) in semiarid areas was usually zero or a very small amount that did not significantly decrease the climatologically appropriate precipitation (CAFEC) value, $\hat{P}_{w,j}$ (see Equation 12). Consequently, the CAFEC precipitation was always much bigger than the available small amounts of precipitation. This seemed to be the primary cause of the unrealistic perpetual drought episode in dry areas such as Lodwar. Revisions in the computation of the CAFEC precipitation outlined in Equation 19 and 20 seemed to solve this problem in the modified PDSI algorithm.

It is also noted that the departure weighing coefficient (K_j) obtained by the empirical relation given in Equation 15 always returned values that are less than 0.6 for dry places such as Lodwar while the same equation returned values that are in the range of 1.0 to 2.0 for moderately wet areas.

The differences between the original and modified PDSI for the wetter regions are less pronounced. This is probably because shifting between the pseudo-indices X1, X2, X3 re-initializes the system such that differences between the original and modified PDSI do

not propagate along the entire time series. The extreme drought values resulting from the modified PDSI algorithm seem to be less than those obtained from the original PDSI algorithm (See Figure 3.9-c), probably due to using the revised at-site drought recursive formula (Equations 29-31).

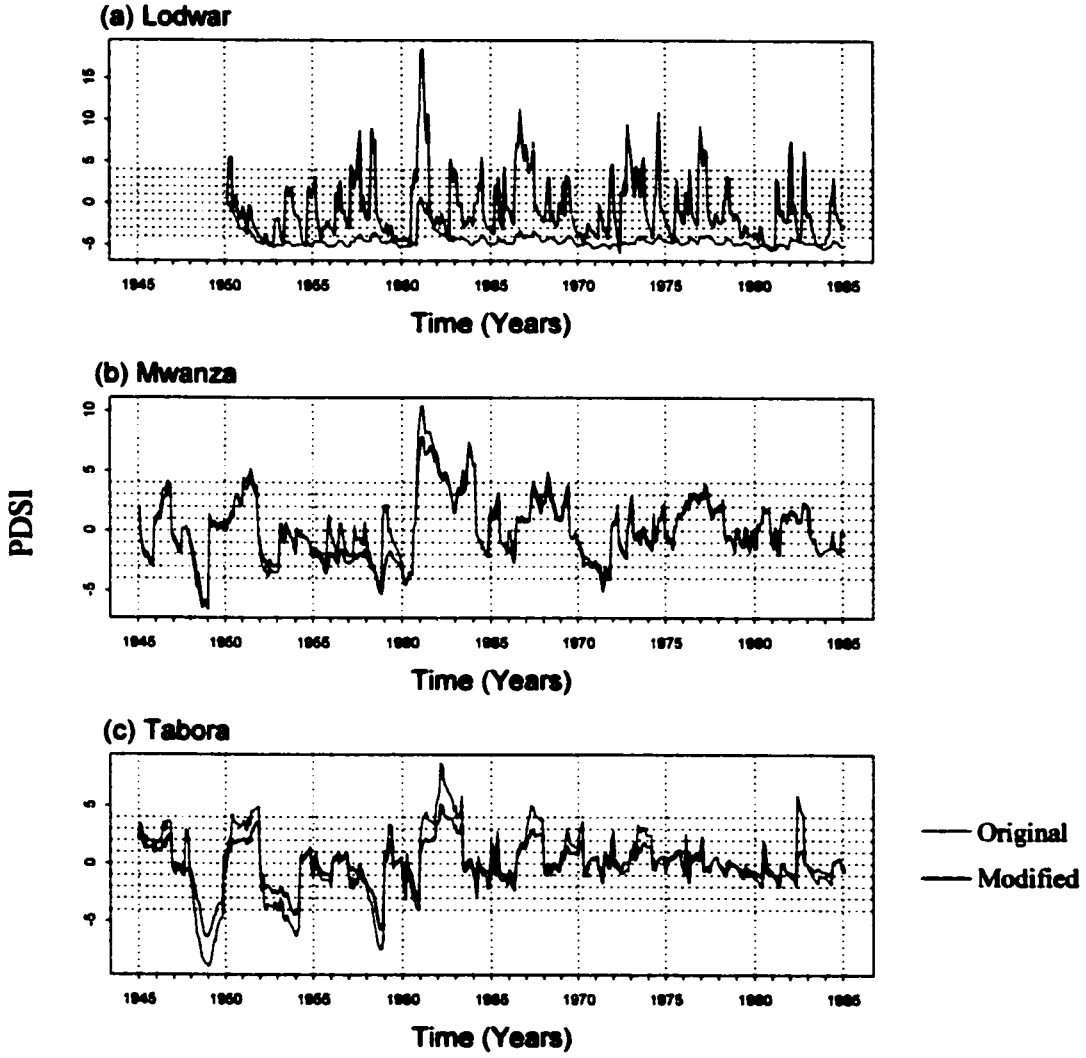


Figure 3.9 Typical time series of the original and modified PDSI for three locations.

In both the original and modified cases the magnitude of the extreme positive PDSI values was observed to be way beyond +4, which is the threshold for extreme wet events. The original PDSI recursive formula was developed with respect to extremely dry events that were given a threshold of –4 and assuming symmetrical distribution, extremely wet

events would have a threshold of +4. However, this assumption of symmetric distribution is not necessarily true.

It is observed that in most cases the modified algorithm tends to smooth out some fluctuations seen in the original algorithm. The modified PDSI, just like the original, is not very sensitive to the actual evaporation values because rainfall dominates the moisture balance accounting used to compute the index. In this regard, we tested the modified PDSI first using the Thornthwaite procedure to compute the monthly evaporation, and second, the average monthly pan-measurements. We did not find much difference between the PDSI obtained from the two types of data. It would seem therefore that the evapotranspiration input in the modified PDSI algorithm serves as a kind of datum on which the rainfall variability can be transposed. So far there has been no report of large variations in monthly evapotranspiration time series comparable to that found in monthly rainfall time series. It is worth noting that Karl (1986) also did not find any significant change in the United States PDSI series when he forced the potential evapotranspiration to a constant equal to the long term monthly potential evapotranspiration.

Even though the main PDSI recursive formula appears as first order autoregressive (AR1) model (see Equation 31) it is not a purely AR1 because the index at times abruptly shifts between pseudo indices X1, X2 and X3. Nevertheless, due to Equation 31, PDSI generally exhibits some AR1 characteristics such as an exponential autocorrelation function. The recursive relationship of Equation 31 suggests that the PDSI for a month is linearly related to the previous month. Spectral analysis however reveals the existence of five to nine-year long-term memory in the PDSI data. This is similar to the spectral results obtained by Guttman (1998) when he analyzed United States PDSI data. The long-term memory could be a result of the nature of the water balance accounting approach upon which the index is based. It is difficult to justify the existence of long-term memory of 5 or more years in drought mechanisms. This leads us to conclude that extra caution has to be taken when interpreting both the modified and the original PDSI index.

The PDSI indices obtained by using the modified algorithm compared well with the historic climatic records. The worst recorded drought year in East Africa in the 20th century was 1949. Most places in East Africa received annual rainfall in the 20th or less percentile in that drought year. During that year, the monthly PDSI was less than –3.0 for most of Tanzania and Eastern Kenya indicating that 1949 was a severe drought year. Comparison with plots of the original PDSI show that the modified algorithm was generally better at capturing the well known documented droughts like that of 1949 (Figure 3.10).

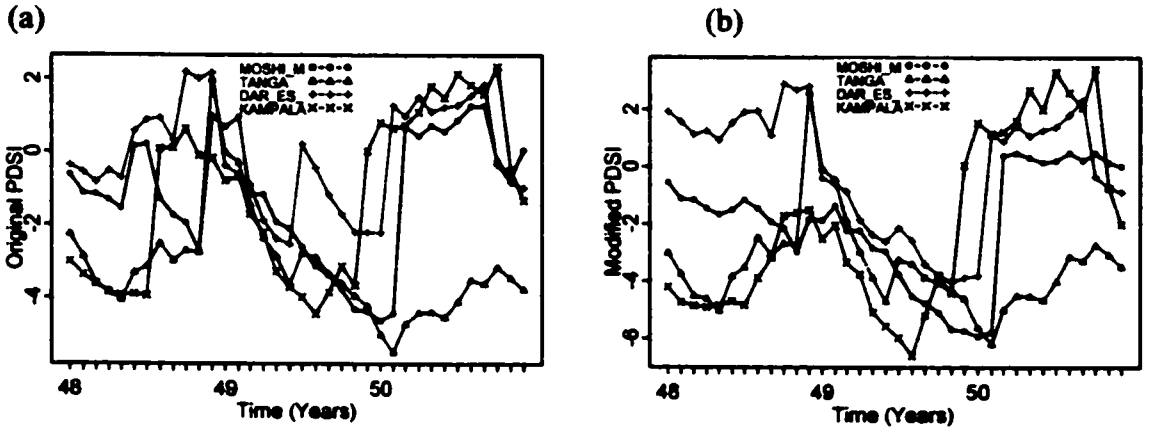


Figure 3.10 shows (a) original and (b) modified PDSI of selected stations during the period of Jan 1948 to Dec 1950.

Although both the original and modified PDSI indices have the same trends at some of the stations, at other stations there are some important differences. For example, the modified PDSI series show that Dar es Salaam experienced severe dry conditions just like the rest of the other areas during May-Oct of 1949. On the other hand, according to the original PDSI series, Dar es Salaam had near-normal moisture conditions during the May-July 1949 period which is not correct since we know that Dar es Salaam (average rainfall 1100mm) was very dry in 1949 because that year it received the lowest annual rainfall (438mm) in the 1900-1997 record.

3.3.2 Comparison of the Modified PDSI with the Bhalme – Mooley Index.

There seems to be a high correlation between the modified PDSI and the Bhalme – Mooley Index (Figure 3.11). The high correlation (greater than 0.8 in many places) tends to suggest that most of the variability in the PDSI is due to precipitation alone.

Some parts of East Africa, namely southern Tanzania and a part of western Kenya highlands show a lower correlation between the two indices probably because in these locations, the evaporation regime significantly modifies the drought characteristics. As the distance from the equator increases, seasonal variations in temperature and evaporation tend to increase. From this perspective, the Bhalme-Mooley index which does not consider the evapotranspiration variable, may fail to capture drought evolution in locations further than ±6° in latitude away from the Equator. On the other hand, it is worth to note that the areas where there is least correlation are mainly highlands areas (western Kenya highlands and southwestern Tanzania highlands) where evaporation computed by the Thornthwaite temperature index method is likely to be inaccurate.

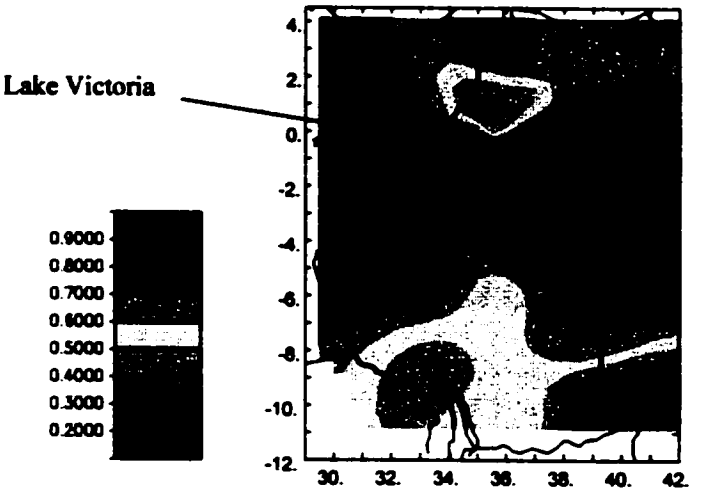


Figure 3.11 . The Spatial correlation between PDSI and the Bhalme-Mooley Index for East Africa.

Although the Bhalme-Mooley index avoids many of the shortcomings of the PDSI algorithm, it still shares with it one disadvantage in that it simplifies the occurrence of droughts to a recursive, AR1 process. Unlike the PDSI which may at times abruptly shift

among three pseudo indices, the Bhalme–Mooley index is a purely AR1 process. A sudden shock in the series may take a very long to die down depending on the coefficient in the first order process (Figure 3.12). In this regard, PDSI is more realistic than the Bhalme-Mooley index, because it does not have as much memory. It should be recognized that a drought lasting many months or years may end abruptly with one or two month's intense rainfall.

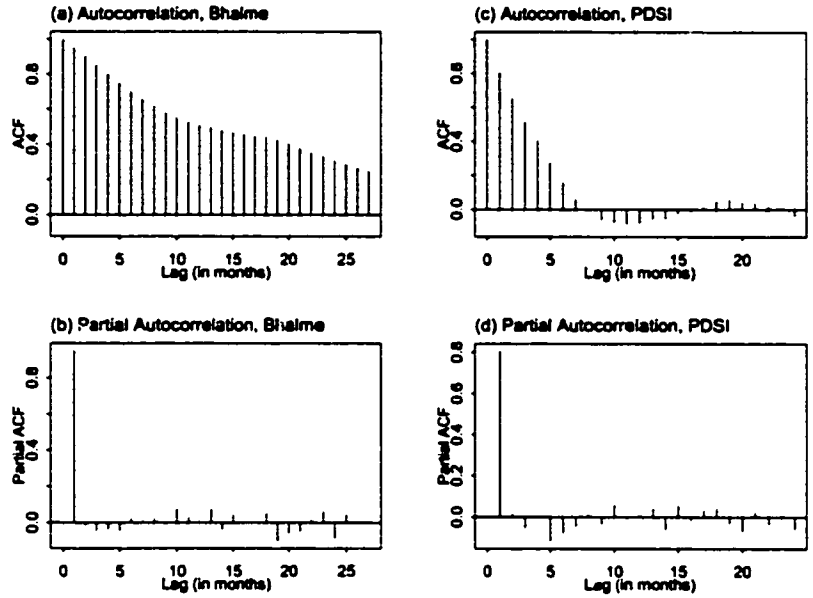


Figure 3.12 Autocorrelation and Partial autocorrelation of the Bhalme-Mooley index (a and b) and that of the modified PDSI (c and d) for Kampala, Uganda. The dotted horizontal lines represent the interval within which the correlations are significant at 95% confidence level.

The lack of any significant partial autocorrelation in both the Bhalme-Mooley and the PDSI is a further demonstration of typical AR1 properties. However, since the PDSI is not a pure AR1 process, its autocorrelation drops much faster than it does for the Bhalme Mooley index.

3.3.3 Advantages of the SPI Index over the other indices.

It is desirable that we obtain the most suitable index to detect the initiation, evolution and termination of droughts in East Africa. However, we should be realistic and recognize

that there is no one index that can perfectly track everything about the four dimensional (three spatial, one temporal) variation of the main climate variables. Furthermore, indices are at best site and application-dependent.

Several investigators such as Yevjevich et al. (1977) and Redmond (1991) discussed criteria that can be used to gauge the suitability of a drought index. Based on these quantitative and qualitative criteria among others, we are persuaded to adopt the SPI as the most suitable index for monitoring droughts in East Africa. In the following sections, we explain the advantages of the SPI in greater detail.

3.3.3.1 Adaptation to the Local Climate

Yevjevich et al. (1977) suggested that in order for a drought index to be effective, it should be derived locally and adapted to the local climate. The modified SPI, which incorporates a form of regional analysis, satisfies this qualitative criterion. Although the modified PDSI and Bhalme-Mooley indices may also be calibrated locally, the algorithm of these latter two indices still retains some relationships that were obtained empirically for a different region. For example, although Palmer (1965) assumed that a plot of the worst negative cumulative moisture index Z vs. duration is approximately a straight line because of the empirical observations he made for the North American study area, there is little basis to believe that it will always be a straight line elsewhere (see Figure 3.1 and 3.2)

3.3.3.2 Flexible Temporal Normalization.

A good drought index, in addition to being standardized spatially, should have the ability to attain unprecedented values if extraordinary climate behavior occurs in the future. The index's temporal normalization with respect to background climatology should be a continuous process. Of the available drought indices, the SPI best approaches this behavior of being "open-ended". Future values of the SPI are not bounded. On the other

hand, by virtue of their algorithms, the PDSI and Bhalme-Mooley indices are theoretically bounded by a lower value of –4. For example the original PDSI parameters were obtained basing on the extreme droughts in American Great Plains that occurred in the 1930s. The fixed calibration period of the PDSI makes it unsuitable for monitoring the occurrence of exceptional climatic variations.

3.3.3.3 Spatial Invariability

A particularly desirable attribute of a drought index is the property of spatial invariability. The index should be able to present the same information regardless of the spatial location of the site being investigated. This is fundamental if inter-zonal comparison of drought conditions is to be meaningful. One way we can investigate this consistency is by examining the spectra properties of the signals (Guttman 1998).

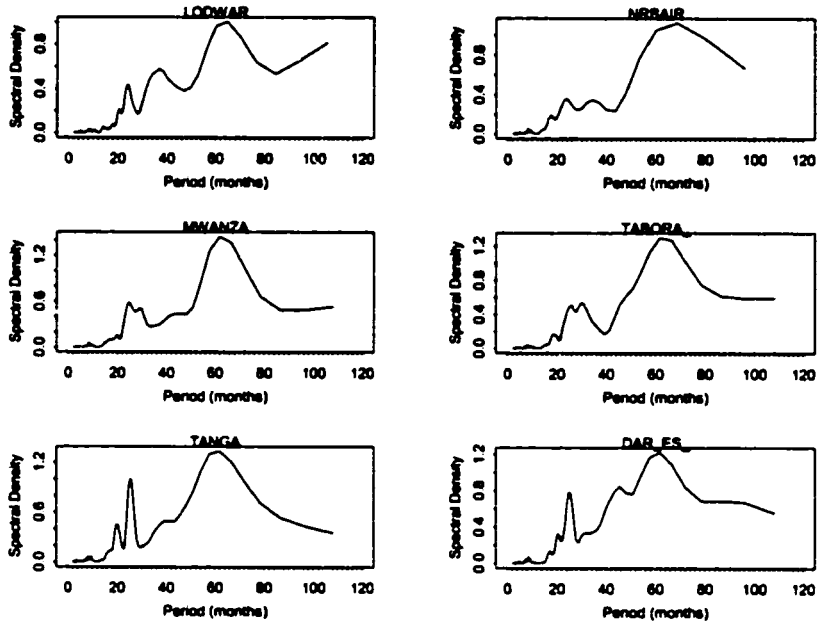
An inter-site comparison of the power spectra of both PDSI and SPI show important differences in the uniformity of the spectra patterns. The power spectra of SPI show more consistent characteristics than that of PDSI (Figure 3.12). This observation is in agreement with the results obtained by Guttman (1998) who compared PDSI and SPI for continental United States. PDSI's relatively tedious procedure involving multiple algorithm variables could lead to inconsistent conclusions as reflected in the diverse nature of power spectra plots shown in Figure 3.13 (b). In contrast, the relatively straightforward nature of SPI involving only precipitation produce relatively consistent results across East Africa. This consistency of SPI enables us to compare results between stations located in the same region easily. However, comparing drought conditions between different stations using PDSI has to be done in a more cautious manner.

3.3.3.4 Flexibility of multiple time scale analysis.

Unlike the rest of drought indices, SPI can be tailored to varying time periods according to a user's interests. Hydrological variables such as soil moisture, stream flow and

reservoir storage respond to precipitation shortage at different time scales. The 12-month SPI is suitable for tracking intermediate term droughts while 24 or 36-month SPI is suitable for tracking long-term droughts.

(a)



(b)

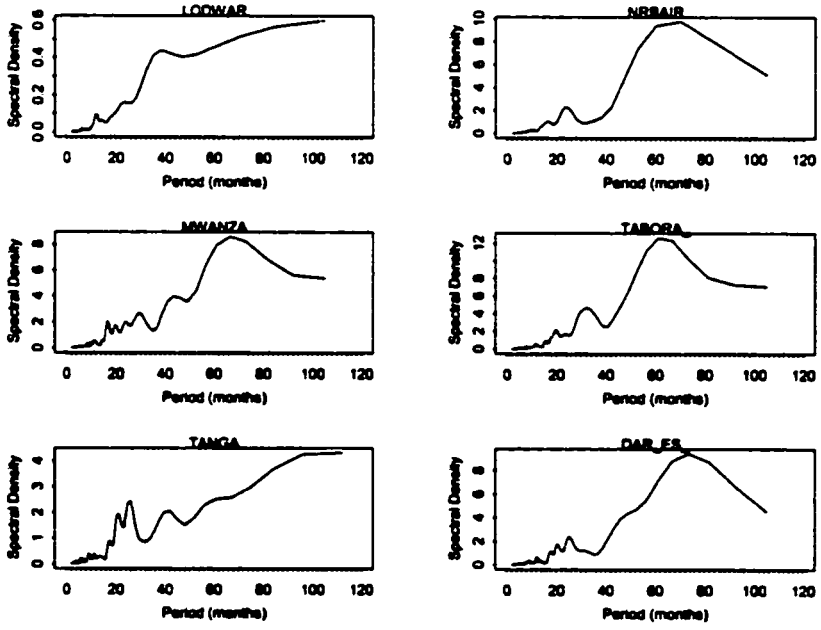


Figure 3.13 Spectral density estimates for (a) 12-month SPI (b) Modified PDSI of selected East African stations.

Our results show that the modified PDSI is strongly correlated to 12-month SPI for most stations and 11-month SPI for a few stations. Figure 3.14 shows an example of the correlation between the modified PDSI and 2-month to 36-month SPI (both original and modified). This shows that the PDSI is probably not suitable for tracking droughts at times scales shorter than a calendar year since it is highly correlated with the 12-month SPI.

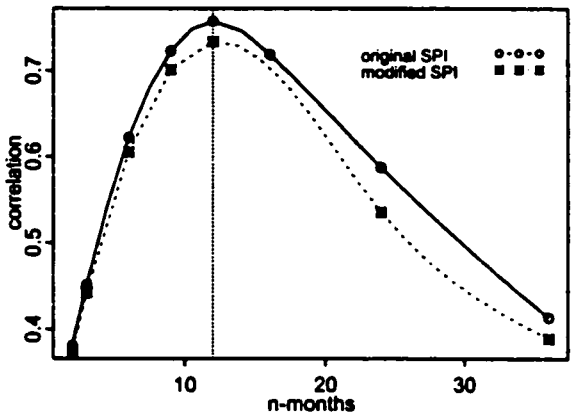


Figure 3.14 Correlation between PDSI for Tabora station and the n-month SPI where n ranges from 2 months to 36 months. The dashed curve refers to the modified P3-based SPI while the dashed vertical line marks the 12-month time scale (n=12).

Most current drought indices use a monthly or longer time scale as a unit. Byun and Wilhite (1999) suggest that a daily unit should be used because the water amount of an affected drought region can return to normal with only a day's rainfall. For example, if there were heavy rains only on 1st December and 31st January, sixty days of no precipitation from 2nd December to 30st January may not be detected by a monthly index in spite of serious damage arising from 60 continuous days of no rainfall.

We submit that minimal alterations in the SPI algorithm may permit the computation of the SPI using daily units. There has been no known previous attempt to calculate the SPI in that manner. Our investigations reveal that it might even be preferable to use daily data where it is available as the computed index tends to be more responsive and always slightly leads the one computed from monthly data (See Figure 3.13). The only difference between the two variants of the indices is that one uses a window of a specified number

of months to smoothen the data while the other uses a window having an equivalent number of days to smoothen the daily data.

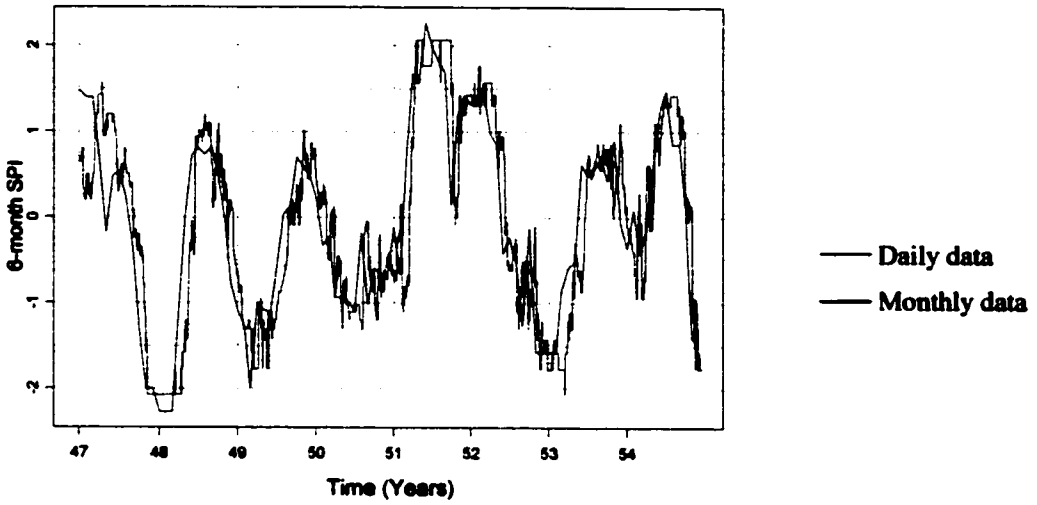


Figure 3.15 Comparison of 6-month SPI for Mbarara, Uganda, obtained by (i) using monthly data and a moving six-month average filter and (ii) daily data and a moving filter of 180 days.

3.3.3.5 Data Requirements and Availability

For any drought index to be generated and used in an operational mode, the input data needed must be readily available. The PDSI requires long time series of precipitation, and evapotranspiration as well as the spatial distribution of the soil moisture properties which are hard to come by. Long record time series of evaporation data are not readily available in East Africa, especially in more remote areas, hence limiting the application of PDSI.

On the other hand the SPI is based on precipitation data only. Although it is a precipitation-only index, at 12-month analysis scales it correlates well with the PDSI, which involves extensive moisture balance operations. Since East African precipitation data is reasonably available, it is advisable to adopt the SPI rather than the PDSI for it

would not be possible to compute the PDSI in many parts of East Africa due to a lack of evapotranspiration and soil moisture data of reasonable length.

3.3.3.6 Interpretability

A good drought index should be easy to interpret if it is to become an effective tool in monitoring droughts. Of the three indices analyzed in this investigation, SPI is the easiest to interpret because an n-month SPI value for a given location is the number of P3 standard deviations from the mean for that location at that particular period, looked at n-month time scale perspective. For example, if the 12-month SPI of a location in November is 1, it means that the amount of rainfall totals for the preceding 12 months until November equal to the mean plus one standard deviation. No such straightforward interpretation can be derived for PDSI nor Bhalme Mooley index.

As the intended audience gets wider, and probably less knowledgeable about the index details, proper interpretation of the index values become less critically dependent on the detailed understanding of the index algorithm, caveats, limitations and assumptions involved. It is our view that although the SPI may not be straight forward to understand and compute, it is much easier than the rest of the drought indices considered in this investigation to interpret.

3.3.3.7 Probabilistic nature.

Another advantage of SPI over the other two indices is its probabilistic properties, which can be used quantitatively, to determine the probability that a drought event will come to an end in a given period. This can be useful in risk and decision analysis. Based on the previous SPI drought index of a given month, we can compute the quantity of the precipitation that would be required in the ensuing month to bring the SPI value to the "near-normal" range. The probability of having this quantity of precipitation could then be calculated based on the historical datasets thereby providing information which would

be useful to decision makers. Alternatively, the computed amount of precipitation required to end a drought in the ensuing month could be compared to forecast products from an empirical or dynamic seasonal forecast model to give a reasonable outlook assessment whether an ongoing drought is likely to end in the ensuing month.

3.3.4 Application of the Modified SPI index.

The application of the modified SPI index in East Africa is illustrated for the 1991/92 period when north and northeastern Kenya experienced a severe drought that affected more than 2 million people. Figure 3.16 shows how the modified SPI index could have been used to track such a drought. A perusal of Figure 3.16 reveals a start of a drought cell in northeastern Kenya by May 1991. In 1991 there was general failure of the long (March – May) rains in both northeastern Uganda and northwestern Kenya (Climate Monitor, 1994). In May 1991, the lowest 6-month SPI index was approximately –1.0 in this region indicating moderately dry conditions. By March 1992, conditions in north and northwestern Kenya had worsened into the extremely–dry threshold of –2.0. By this time, maize-crop and livestock had suffered serious damage from the prolonged dryness (Climate Monitor, 1995).

It was not until later in September 1992 that the moisture conditions returned to normal. The spatial maps of the 6-month SPI index are generally in agreement with what was recorded in the weather archives (Climate Monitor, 1995 & 1996). The use of SPI would have facilitated decision planners in the region to better demarcate the spatial distribution and severity of the drought episode and in the process assist in the formulation of a better drought mitigation strategy for the region.

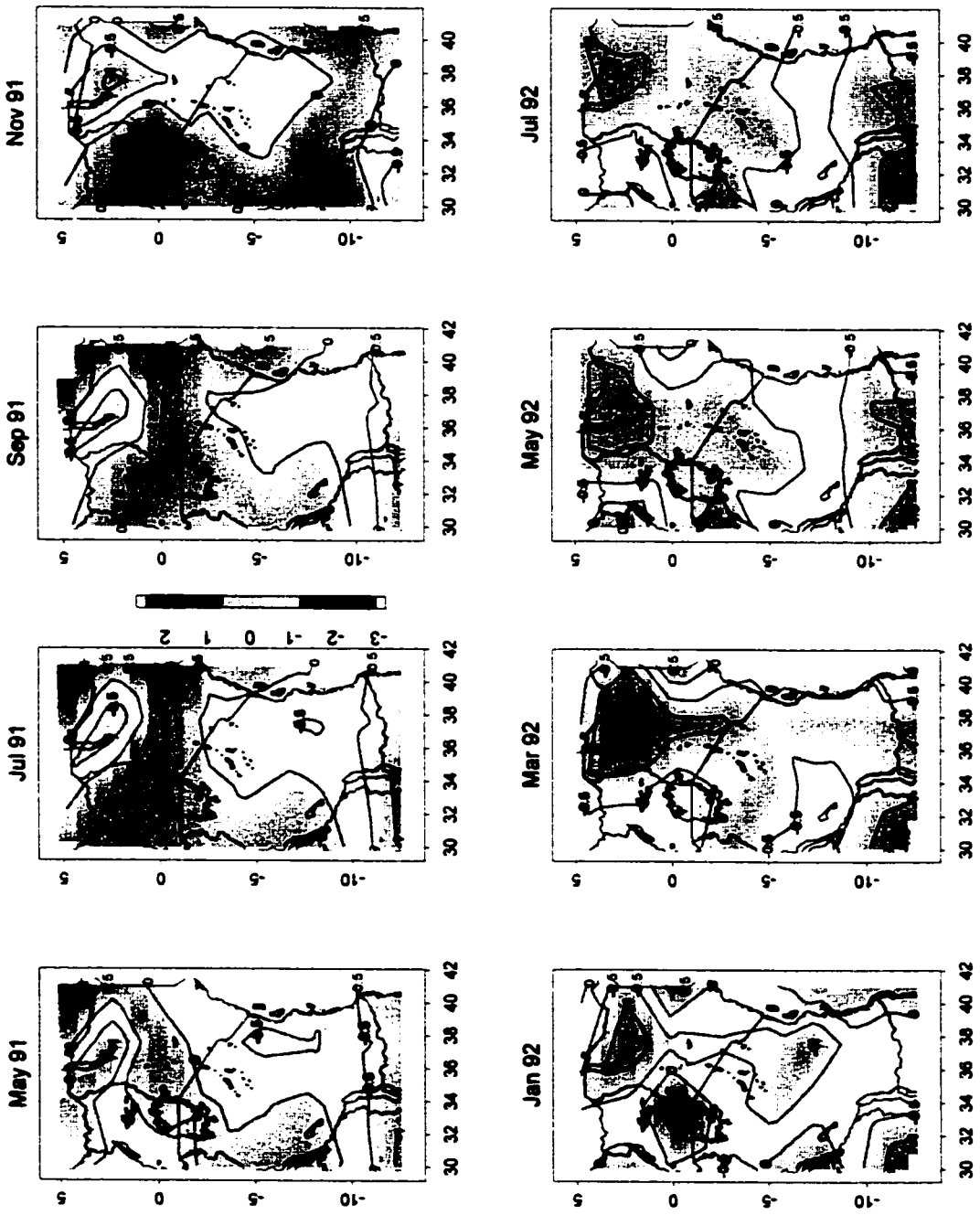


Figure 3.16 The drought of 1991/92 in Kenya as tracked by the (modified P3 based) 6-month SPI.

- 79 -

3.4 Regionalization of East Africa into drought zones using the Drought Indices

Drought management and mitigation plans for future droughts rely significantly on the knowledge about the boundaries of drought zones and their characteristics, because for a region such as East Africa, some areas may be experiencing severe water stress while others may be undergoing wet spells. The delineation of such spatial drought patterns also facilitate a more timely transfer of resources from the endowed zones to the drought prone zones. This knowledge can also enable policy planners formulate better settlement and land use policies for various zones with due consideration to the drought risk component.

Methods such as eigen-techniques and harmonic analysis have been used to delineate homogenous zones according to some variables of interest. Each method has its advantages and shortcomings. Harmonic analysis is suitable for signals that are predominantly periodic while eigen-techniques such as the Empirical Orthogonal Functions (EOF) have been used in cases where periodicity is not a feature of the data or when it is desirable to reduce the data dimensionality.

Here we use the rotated Principle Component Analysis (PCA) on the SPI covariance matrix to delineate East Africa into homogenous drought zones. PCA consists of transforming a large number of variables into (usually) a smaller number of orthogonal variables that represent common causes of manifested changes. The method sorts correlated data into a hierarchy of statistically independent modes of variation that progressively explain less and less of the total variance. Richman (1986) demonstrated the inherent disadvantages of un-rotated principal components (PCs) when PCA is used to depict individual modes of variation of data matrices in exploratory analyses. The major disadvantage is that the topography of the un-rotated PCs is primarily determined by the shape of the domain and not the co-variation of the data (Richman, 1986). In this investigation we use the popular orthogonal varimax rotation of the PCs to circumvent this problem. A brief mathematical derivation of the PCA is given in Chapter 5.

The data used in this analysis is the 6-month and 12-month SPI drought index computed from East African rainfall gridded data for the periods 1900-1996. We conjecture that the 6-month SPI index is appropriate to track seasonal drought of spells lasting 3 months to 8 month. McKee et al. (1993) suggest that 12-month SPI is suitable for tracking intermediate range drought. We used a composite data set which includes only months when severe drought conditions occurred in at least one grid location in the East African region (i.e. 6-month SPI ≤ -1.42).

We only retain the first few principal components explaining a substantial amount of the total variance. The cut off is based on the Kaiser Criterion and the Scree test. The Kaiser criterion (Kaiser 1960), is probably the most widely used guideline in deciding the number of PCs to be retained. This criterion suggests that the PCs to be selected should be those with eigenvalues greater than one. The second procedure, the Scree test, is a graphical method by Cattell (1966) where the eigenvalues are plotted against their rank. The point on the curve where the smooth decrease of eigen values appears to level off signify the number of significant PCs to be retained (see Figure 3.17).

3.4.1 Application of Rotated PCA on SPI

The results of the unrotated PCA analysis are shown in Table 3.6. Only the first 10 principal components, which account for more than 75% of the total variance, are shown.

Table 3.6 Results of PCA analysis on composite drought SPI data.

Eigen rank	6-month SPI			12-month SPI		
	Eigen Value	% Variance Explained		Eigen Value	% Variance Explained	
		Of total	Cumulative		Of Total	Cumulative
1	6.71	21.43	21.43	6.21	20.76	20.76
2	5.14	16.43	37.86	5.16	17.28	38.04
3	2.92	9.34	47.21	2.85	9.54	47.58
4	1.79	5.72	52.92	1.85	6.18	53.76
5	1.61	5.15	58.08	1.75	5.84	59.61
6	1.35	4.30	62.37	1.34	4.49	64.09
7	1.15	3.66	66.04	1.29	4.31	68.40
8	1.10	3.51	69.55	1.06	3.56	71.96
9	0.95	3.03	72.58	0.99	3.30	75.26
10	0.89	2.83	75.41	0.78	2.63	77.88

It is noted that using the Kaiser guideline would mean retaining too many PCs for some of which contribute less than 5% of the total variance. On the other hand using the scree test would only allow four PCs to be retained which account for only 53% of the total variance. We settle for a compromise by retaining the first seven PCs that explain 66% and 68% of the 6-Month SPI and 12-month SPI respectively.

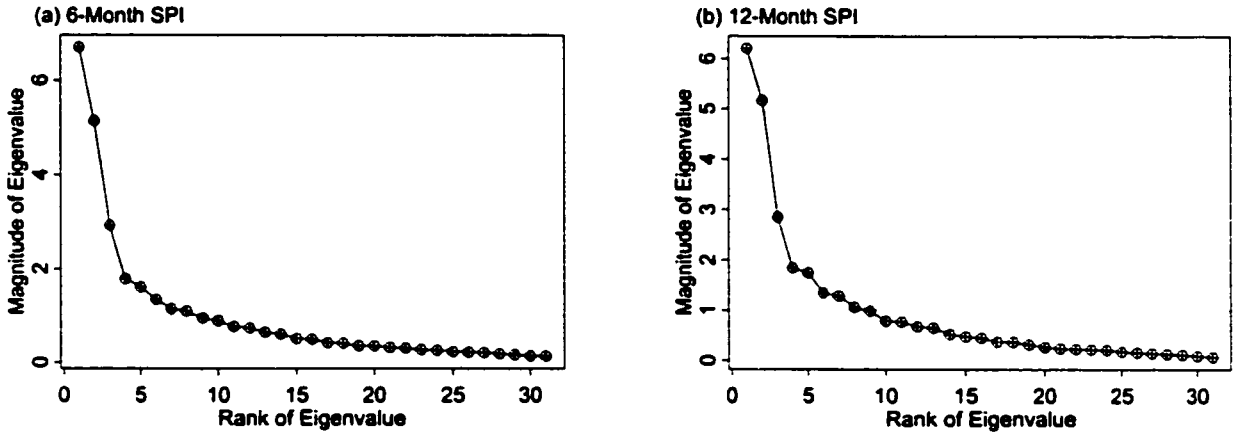


Figure 3.17 Scree test plots of the eigenvalues obtained from applying PCA to (a) 6 month SPI composite drought data, (b) 12-month SPI composite drought data. The horizontal line shows the cutoff according to Kaiser criterion.

The SPI is derived from rainfall, which in turn is influenced by synoptic and regional factors. Thus, some degree of similarities between the SPI patterns of neighboring locations should be expected. Such similarities are not accounted for in the PCA solutions making some of the derived PCA solutions unrealistic. Rotating the retained PCs minimizes the ambiguities associated with the initial PCA solutions. The retained PCs were rotated according to the varimax method. The rotated solutions are shown in Figure 3.18 (a) and (b)

A perusal of the PCA loadings in Figures 3.18 (a) and (b) show that generally similar loading patterns are obtained for both the 6-month and the 12-month SPI analysis. For example, most of Uganda seems to be loaded under the first component as shown in Figure 3.18(a). The same area is observed to be loaded with the third component in Figure 3.18(b). It would seem that in this case, PCA extracts the same patterns in both cases but under different components.

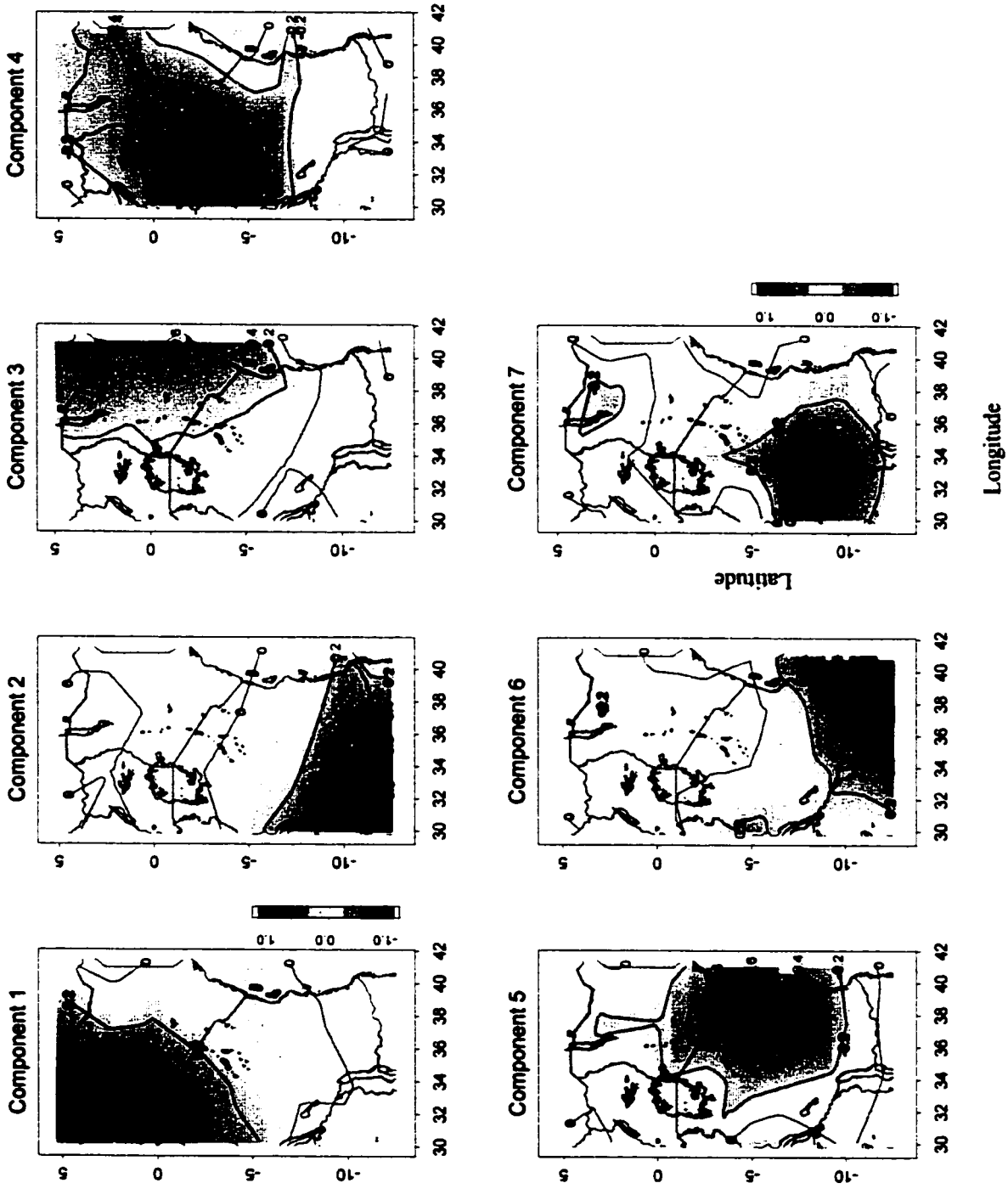


Figure 3.18 (a) Loading of the seven rotated components obtained from PCA analysis of 6-month SPI drought data.

- 83 -

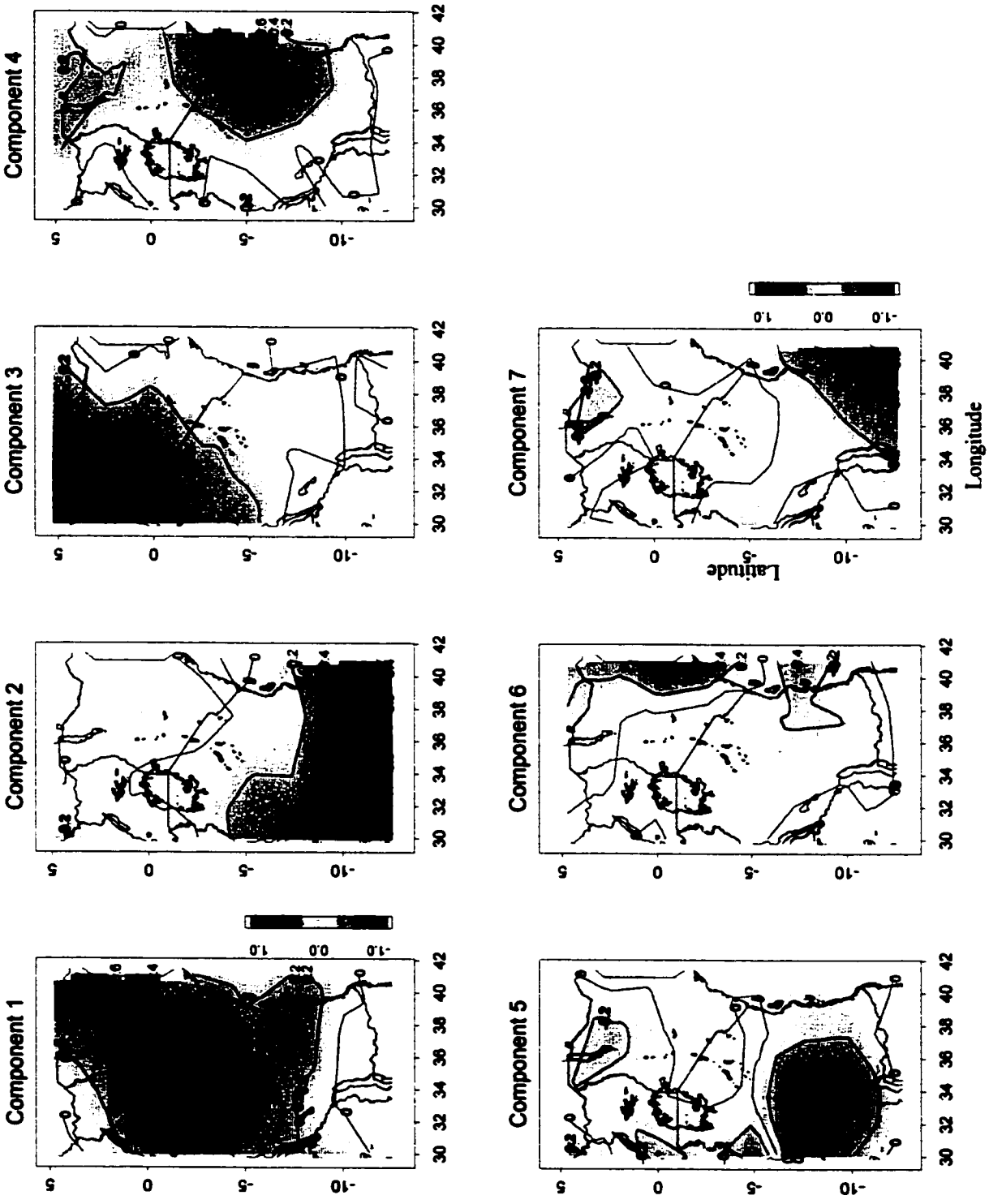


Figure 3.18 (b) Loading of the seven rotated components obtained from PCA analysis of 12-month SPI drought data.

- 84 -

3.4.2 The PCA Delineated Drought zones.

Figure 3.19 shows the PCA delineated homogeneous drought zones (labeled A to G). The spatial boundaries of these zones are inferred from the PCA loading patterns obtained in Figure 3.18 (a) and (b). There is high correlation of the drought indices within the grids of a given homogeneous zone, indicating possible common drought causing factors. These factors are possibly closely related to the nature of the rainfall generating mechanisms in the specific zones which range from synoptic to regionally induced meso-scale systems. The factors that control rainfall over East Africa include the position, orientation, and intensity of the ITCZ, subtropical anticyclones, Indian Ocean cyclones, monsoon winds, Sea Surface temperatures, the jet streams, and several local factors like topography and the influence of large water bodies.

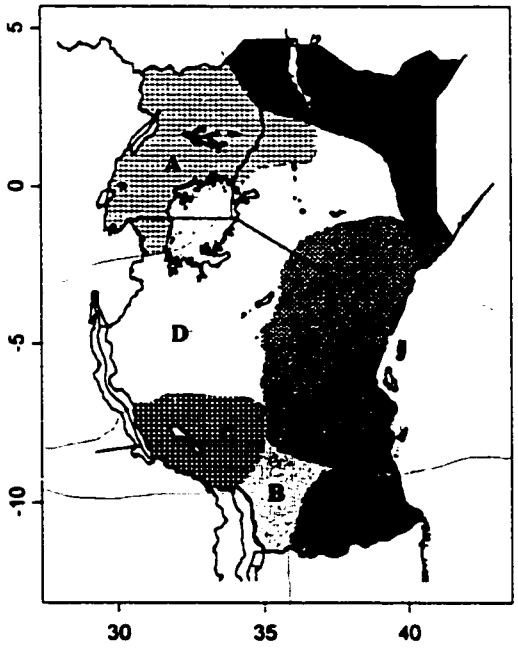


Figure 3.19 PCA Delineated Drought zones for East Africa based on 6 and 12-month SPI.

The drought zones shown in Figure 3.19 bear some resemblance to the homogenous precipitation zones delineated by harmonic analysis (Figure 2.5). This is to be expected since it is likely that during a drought episode, rainfall is likely to be uniformly low in a given homogeneous precipitation zone giving rise to similar drought conditions throughout the zone. For example zone 1 and 2 (Uganda) in Figure 2.5 seem to make up region A in Figure 3.19, while the semi arid north and northeastern Kenya are zone 3 and region C in Figure 2.5 and 3.19 respectively. The similar precipitation based zones obtained by using widely different techniques demonstrate the overall validity of the results obtained.

Most of Zone C (eastern and northeastern Kenya) is a basically semi-arid region receiving less than 400 mm of annual rainfall. The effect of droughts on such dry regions can be disastrous. The average moisture conditions in such dry lands are such that the biosphere is already living on or quite close to the minimum amount of water required to survival, and so any further reduction in water supply often creates an extremely stressful environment for survival. It is therefore not surprising to find that most of the severe historical droughts documented in East Africa have occurred in zones C. The difference between zone B and H is probably brought about by the proximity and influence of the Indian Ocean.

Zone A covers most of Uganda and parts of western Kenya. Apart from the northeastern tip, most of Uganda receives substantial and reliable amount of rainfall throughout the year in excess of 1000 mm. The frequency of documented droughts with severe impacts in this zone is less when compared to other zones. The Congo airmass seems to play an important role in enhancing the long rains in this region (Okoola, 1999 and Trewartha, 1981). Years that had weak westerly incursions and subsequent strong flows of the Southeasterlies reaching deep into East Africa (as far as zone A) have been observed to be particularly dry years in zone A (Trewartha, 1981).

Zone D covers most of central and northern Tanzania as well as southwestern Kenya while zone G covers the southern Highlands of Tanzania. Most parts of zone F receive

relief rainfall unlike the surrounding zones which rely on convectional rainfall. Zone D has one dry and wet season which follows the passage of the ITCZ. The main moisture source for this zone are the moist southeasterly monsoons. Any atmospheric disturbance that weakens the southeasterly flow into this zone during the wet season will bring about diminished rainfall throughout the zone.

3.4.2.1 Interrelation of the drought zones

We carried out a simple correlation analysis between the index values of a selected drought zone with that of other zones, conditioned on the drought state of the selected zone (Table 3.7)

Table 3.7 Inter – zone correlation matrix of 12-month SPI between the delineated homogenous drought zones.

Zone experiencing drought conditions	A	B	C	D	E	F	G
A		**-0.35**	0.04	0.29	0.08	**-0.18**	**-0.26**
B	**-0.29**		0.04	0.04	0.07	0.25	0.34
C	0.25	-0.05		0.41	0.53	-0.01	0.21
D	0.56	-0.07	0.13		0.31	**-0.21**	-0.06
E	0.18	0.04	0.22	0.43		0.04	0.19
F	0.19	0.12	0.07	0.18	-0.03		0.02
G	0.06	0.29	0.17	0.37	0.29	0.23	

Correlations (ρ) within the range $-0.18 \geq \rho \geq 0.18$ are statistically significantly at 95% significant level ($n \approx 120$ months). These have been shaded gray in the table. In addition, the significant negative correlations are shown in bold figures.

The first column of Table 3.7 contains the zones whose 12-month SPI drought states are the reference for computing the correlations. For example, all the months for zone C (which is Eastern Kenya) whose drought index values indicated a moderate drought

conditions or worse (i.e. 12-month SPI equal or less than -1) were selected. The aggregate values for the other zones corresponding to these selected months were obtained and correlated with the zone C values. The correlation results indicated that whenever zone C is experiencing drought, zone A (most of Uganda) is likely to be in drought too (ρ=0.25), even though C and A are not spatially adjoining zones. Interestingly the analysis results showed that such relationship did not necessarily hold the other way round (Table 3.7).

From Table 3.7, it would seem that zone B (southwest East Africa) is least correlated with the drought state of the other regions. The observation that zone A (Uganda) is negatively correlated to the southern parts of East Africa (zone B, F and G) could be put to use in drawing up drought mitigation plans. Since Uganda and the later zones are seldom in drought at the same time, one could act as a source of surplus resources whenever the other is experiencing drought stress. Other relationships of interest would include that between zones D and A or D and F. The drought conditions in Zone D (central and northern Tanzania as well as southwestern Kenya) are positively correlated to those in Zone A (Uganda) and E (eastern Tanzania) but are negatively correlated with those in Zone F (southeastern Tanzania).

Such relationships between the various drought zones could be useful in designing improved drought monitoring and mitigation programs. Observations from historical rainfall records indicate that the whole of East Africa has never been in drought at the same time. Even during major historical droughts when large percentages of East Africa suffered from severe dry spells, some regions were unaffected with good or even above average rainfall.

For example 1949 and 1984 were among the driest years in East Africa in the last century. The 1949 drought was particularly severe with wide spread loss of animal life. Figure 3.20 (a) and (b) show the annual rainfall percentiles during those respective years. It is readily observed that some East African regions still experienced well above average rainfall during those drought years. Comparing the spatial extent of the 1984 drought and

the 1949 drought (Figure 3.17), it is obvious that the areal extent of the two drought events are quite exclusive of each other. The 1984 drought was particularly harsh in zone A (Uganda) and some areas of zone C (Kenya). The impacts of the 1984 drought in zone A were widespread but were not as severe as those felt by the dry lands in zone C. Ironically, the rest of the other zones received normal to above normal rainfall that year.

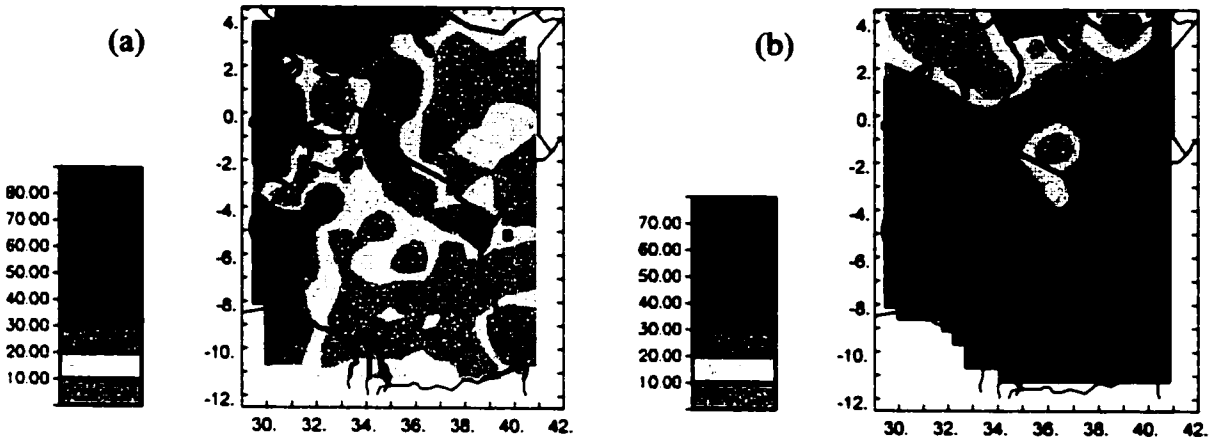


Figure 3.20 Annual Rainfall in terms of percentiles during the extreme drought years of (a) 1949 and (b) 1984.

3.5 Summary and Conclusions

In this investigation we analyzed the properties of several drought indices with the objective of identifying the most suitable index to detect and monitor meteorological droughts in East Africa. The indices that we analyzed included the Palmers Drought Severity Index (PDSI), Bhalme-Mooley Index, and the Standardized Precipitation Index (SPI). These indices were selected mainly because theoretically they can be applied to regions of varying climates.

Our analysis shows that the original PDSI algorithm, designed for the US, did not give reasonable results for some parts of East Africa – especially the drier parts. The PDSI algorithm was partly modified to make it more applicable to the climate of East Africa. In particular, a new method was suggested to compute the potential runoff, the Z index and the recursive formula that is perhaps the core of PDSI. The modified PDSI algorithm

produced more realistic results than the unmodified PDSI for all the varying climates found in East Africa.

The Bhalme-Mooley index whose input is rainfall only, was found to be highly correlated to that of the modified PDSI. This suggested that for East African drought, most of the variability in both indices is due to precipitation alone. Thus the Bhalme-Mooley index, which does not involve complex water balance computations, could be as informative as the complex PDSI.

Our analysis showed that the SPI is a relatively more suitable index for monitoring droughts in East Africa because of its flexibility to track droughts at different time scales according to a user's interests, its spatial invariability, its suitable statistical properties, the ease of computation and its ability to extract more or less the same information contained by the temporally fixed PDSI. If required, the SPI can even be computed at daily intervals provided that suitable daily records are available.

We introduced two modifications to improve SPI. First, instead of fitting a gamma distribution to the "smoothed" precipitation data, we used an unbiased, P3 plotting position formula (Nguyen, et al. 1989) to reduce the effects of outliers on the results. Second, we proposed getting the final SPI by transforming the non-exeedence probabilities into standard P3 variates using regional parameters derived from the flood index method, instead of using a Gaussian normal distribution that typically introduces distortion in the tails for skewed data (such as most precipitation data). The modified SPI produced results that are more representative of East Africa's drought conditions than the original SPI of Mckee, et al. (1993).

Using the rotated principal component analysis (RPCA) on 6-month and 12-month SPI, we delineated East Africa into 7 drought homogenous zones. The spatial distribution of these zones bore some resemblance to the precipitation zones delineated using harmonic analysis (Chapter 2).From the inter-zone correlation between the homogenous drought zones, it seems that some zones in East Africa are likely to experience drought

simultaneously, while others are likely to undergo opposite experience, e.g. droughts in some zones but wet spells in others.

In conclusion,

1. The original Palmer Drought Severity Index, though popular and widely used in North America, is not suitable for use in East Africa.
2. The SPI is likely more suitable drought index for tracking and monitoring droughts in East Africa. The SPI has been modified to take into consideration the skewed nature of precipitation data.
3. East Africa can be divided into several drought zones that are closely related to the homogeneous precipitation zones. There is no one single drought episode that affects the whole of East Africa and spreads to all the zones concurrently.

Chapter 4 East African Rainfall Anomalies and Drought Index Patterns in Association with El Niño/Southern Oscillation.

4.1 Background

Part of this investigation investigates the extent of El Niño – Southern Oscillation's (ENSO) influence on the moisture regime of Eastern Africa. ENSO is probably the most notable climate variation existing at interannual scales (Rasmusson and Wallace, 1983). Precipitation anomalies in relation to ENSO events have been extensively analyzed and documented for different parts of the world, e.g. Rasmusson and Carpenter (1983) and Ropelewski and Halpert (1986, 1987).

El Niño arises from a persistent low-pressure zone in the western Pacific and an equally persistent high-pressure zone over the Eastern Pacific. These coupled pressure systems at times seesaw back and forth producing phenomenal climatic perturbations that affect many regions of the world (Rasmusson and Carpenter, 1983). Normally, trade winds are driven from the eastern to the western Pacific by differences in atmospheric pressure, which move water westward along the Equator, maintaining a buildup of warm water and creating a higher sea level elevation in the western Pacific Ocean. During the onset of El Niño (SOI drops to a negative value), the atmospheric pressure in the eastern Pacific decreases, trade winds weaken, the warm pool extends eastwards and piles off the coast of Peru and Southern Ecuador. This is the warm phase of ENSO. When conditions opposite to the above occur, it is called the La Niña event or the cold phase, which is an amplification of "normal" conditions (Rasmusson and Carpenter, 1983).

Previous studies have indicated some relationships between East African rainfall and the Southern Oscillation (Ropelewski and Halpert, 1996; Ogallo, 1988; Nicholson and Kim, 1997) such that warm events of ENSO tend to be associated with above average rainfall events and vice-versa. Ogallo (1988) used lag correlation and factor analysis to

show that there is significant negative correlation between Southern Oscillation and seasonal rainfall over parts of East Africa during October – December. Nicholson and Kim (1997) studied the rainfall response over the entire Africa to ENSO episodes in the Pacific. They identified southeastern Africa as one of the 15 sectors where ENSO appears to modulate rainfall.

However, much remains to be studied about the effects of global, regional and local circulations on rainfall generating mechanisms in East Africa and how they are connected to ENSO.

This part of the investigation has 3 primary objectives:

(1) By applying harmonic analysis and bootstrap re-sampling technique to ENSO and non-ENSO composites of rainfall anomalies and 6-month SPI drought index, identify homogeneous response regions of East Africa to ENSO;

(2) Using the 6-month index time series formulated for the homogeneous regions identified in (1), perform a detailed analysis of the ENSO effects on the East African Rainfall at seasonal time scales;

(3) Further analyze the regional and seasonal dependency of ENSO effects on East African Rainfall from boxplots of the 6-month SPI between ENSO and non-ENSO affected seasons.

4.2 Data and Methodology

The rainfall data for this analysis was derived from the Hulme (1994) archive ('gu23wld0096.dat', Version 1.0, constructed and supplied by Dr Mike Hulme at the Climatic Research Unit, University of East Anglia, Norwich, UK). This is a historical monthly precipitation dataset for global land areas from 1900 to 1996, gridded using 2.5° latitude by 3.75° longitude resolution. Using this gridded dataset of 31 grids (Figure. 4.1), the standardized precipitation Index (SPI) is computed from a modified procedure of Edwards and McKee (1997). Details of this modified non-parametric SPI index are given

in Chapter 3. In this investigation, East Africa refers to Uganda, Kenya and Tanzania (Figure 4.1).

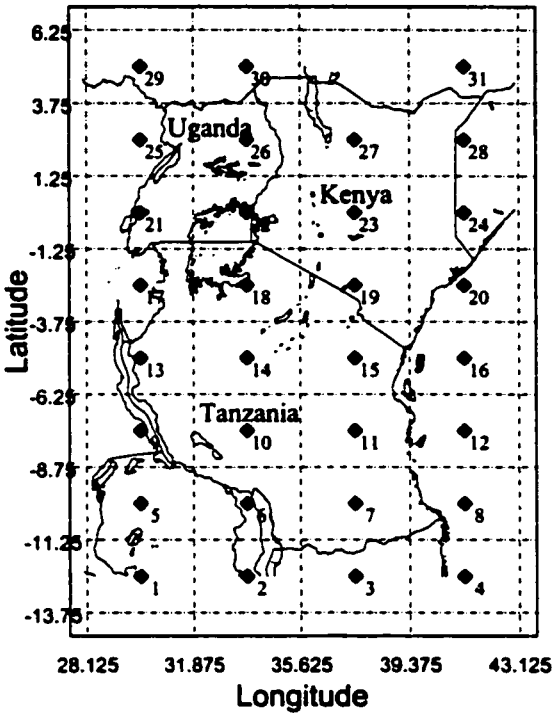


Figure 4.1 Study area showing the rainfall grid used in the study

A composite dataset comprising of data recorded during ENSO events which are identified by climatic indices such as NINO3 or SOI (Southern Oscillation Index) exceeding certain threshold values (Rasmusson and Carpenter, 1983; Ropelewski and Jones, 1987; Piechota and Dracup, 1996) is formed. Data falling in-between ENSO episodes are not included in the composite dataset. Table 4.1 shows the years when 22 El Niño and 13 La Niña events occurred within the 1900-1996 period.

Table 4.1 A list of the major ENSO events that occurred within the 1900-1996 period.

El Niño	La Niña
1902, 1905, 1911, 1914, 1918, 1923, 1925, 1930, 1932, 1939, 1941, 1951, 1953, 1957, 1963, 1965, 1969, 1972, 1976, 1982, 1986, 1991.	1904, 1909, 1916, 1924, 1928, 1938, 1950, 1955, 1964, 1970, 1973, 1975, 1988.

Harmonic analysis, which assumes that any function can be represented as the sum of a series of different sine curves, is applied to the ENSO composites to identify coherent regions. Harmonic analysis had been used by Dracup and Kahya (1994) to study streamflow variability in the United States, and by Halpert and Ropelewski (1992) to investigate the global and regional scale precipitation and temperature patterns associated with ENSO.

4.2.1 Application of Harmonic Analysis

Theoretically, any function can be represented by an infinite series of sine and cosine waves, commonly referred to as the Fourier series. For discrete observed data, a finite number of series will generally be sufficient. For example, the sample mean, five sine and six cosine terms will be sufficient to completely describe the variation of a dataset containing 12 average monthly values. The determination of the finite sum of sine and cosine terms is called the "Harmonic analysis" (Panofsky and Brier, 1965).

The first or the "fundamental" harmonic has a period equal to the total length of the time series being analyzed. The second harmonic has a period equal to half the fundamental period; the third harmonic has a period of one third of the fundamental and so forth. The theoretical details of this analysis are discussed at length by many authors, e.g. Panofsky and Brier (1958) and Rayner (1971). Some of these details are presented in Chapter 2.

By applying harmonic analysis to the ENSO composites, we assume that (a) the composite events can be represented by a summation of cosine and sine curves, (b) the first few harmonics extract most of the variance, and (c) the characteristics of the ENSO response does not drastically change with time for a given location.

Based on the events identified by Rasmusson and Carpenter (1983), ENSO composites were formed for the two-year period starting on July preceding the episode, designated as

July (-) in this investigation. By this convention the evolution of a typical El Niño episode would depend on getting positive SST anomalies off the coast of Peru in December (-) with the maximum anomalies occurring between January (0) and December (0). The ENSO composites were prepared from standardized precipitation anomalies and six-month Standardized Precipitation Index (SPI_6) for both El Niño and La Niña events, giving rise to four major composite datasets, which were subjected to harmonic analysis. The phase and amplitude of the first harmonic were presented on harmonic dials (Figure 4.2).

The harmonic analysis was applied to each ENSO event individually and an average of the harmonic series for all the ENSO events at grid locations was computed. Although the average harmonic could be obtained by applying the analysis on the average of the ENSO events, the former was preferred because by comparing the individual harmonics obtained from the ENSO events, it could show the coherency of the analysis results.

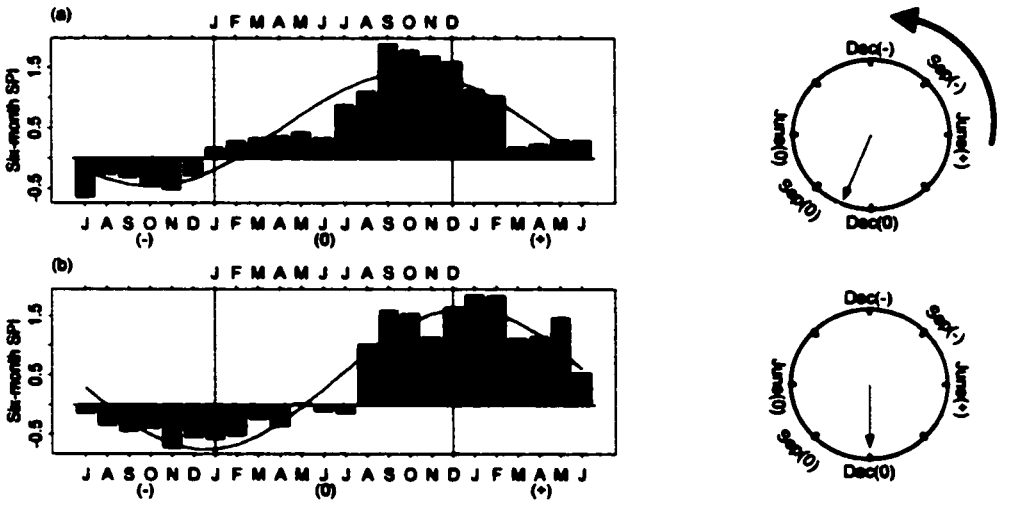


Figure 4.2 The 1941 ENSO composite data for (a) Northeastern Kenya and (b) Northeastern Tanzania, fitted with the first harmonics. The harmonic dials on the right show the magnitude and phase of the first harmonics. –1,0 and +1 refer to months in the years prior to, during and following an ENSO event.

4.2.2 Identification of the ENSO Response Periods

For climate predictions, it is beneficial to delineate the months that show an increased modulation of the drought index (or precipitation) by the ENSO phenomena. An average 36-month ENSO aggregate composite is obtained for each region identified in Figure 4.3. The response periods of maximum ENSO modulation are intervals in which the composite variable is consistently greater than the threshold level set at 90% significant level of a randomly re-sampled composite for at least three months. Typically, there will be a positive and a negative ENSO response corresponding to the minimum and maximum harmonics arising out of the harmonic analysis technique. However, by carefully inspecting the composite, it is often possible to select either the negative or the positive responses as the dominant ENSO response.

If the variable is assumed to be normally distributed, the significant levels can be obtained from standard normal tables. Alternatively, threshold values can be established using bootstrap re-sampling which does not require the normality assumption, which was what we chose in our analysis.

4.2.2.1 The Bootstrap re-sampling procedure

Traditionally, researchers have relied on the central limit theorem and a Gaussian normal approximation to obtain confidence intervals. This approximation is only valid if the variable being analyzed, or some known transformation of it, is asymptotically normal. Since 6-month SPI rainfall anomalies are mostly not normally distributed, it is better to use a bootstrap re-sampling procedure that can accommodate both normal and non-normal distributions. In the bootstrap technique, n new samples, each of the same size as the observed data are drawn with replacement from the observed data. The statistics are calculated for each new set of data, yielding a bootstrap distribution of the data. The fundamental assumption of bootstrapping is that the observed data are representative of the underlying population. By re-sampling data from the observed sample, the process of

sampling observations from the population is mimicked. Further details of bootstrapping can be obtained from Efron (1982), Efron and Tibishrani (1993) and SPLUS 4 (1997).

The database of El Niño (La Niña) composite is established from 22(13) 36□month events, and each of which consist of 12 months prior to the El Niño (La Niña) event, 12 months of the actual warm (cold) ENSO phase and 12 months after which the ENSO signal completely dissipated. The bootstrap composites are obtained from randomly selecting 22 events (each event represents 3 consecutive years) of data with replacement from the available 96 years of data. Thus the yearly structure of the data is maintained. This is repeated 500 times for each of the regions. The empirical statistics of the bootstrap composites are computed to obtain the 90% confidence interval within which we expect the mean of a randomly re-sampled aggregate composite to lie (Piechota and Dracup, 1996). This interval is then used to identify the ENSO period of maximum response. The results from this analysis are shown in Figures 4.4 to 4.6.

4.2.3 Boxplots of Seasonal 6-month SPI Conditioned on ENSO Phase

The shift in the probability distribution of the 6-month SPI index in relation to ENSO is examined by separately computing the lower quartile, median and upper quartile of the seasonal index during El Niño, La Niña and non ENSO years and compare them using boxplots (Figure 4.8). The 6-month SPI data were first prepared on a regional basis, and then the data grouped under ENSO influenced and non-ENSO seasons.

4.3 Discussion of Results

4.3.1 Harmonic Vector Maps

The results of harmonic analysis on El Niño composites composed of standardized rainfall anomalies and six-month Standardized Precipitation Index (SPI6) data are shown in Figure 4.3, (a-c) while (d-f) show the corresponding results for La Niña composites.

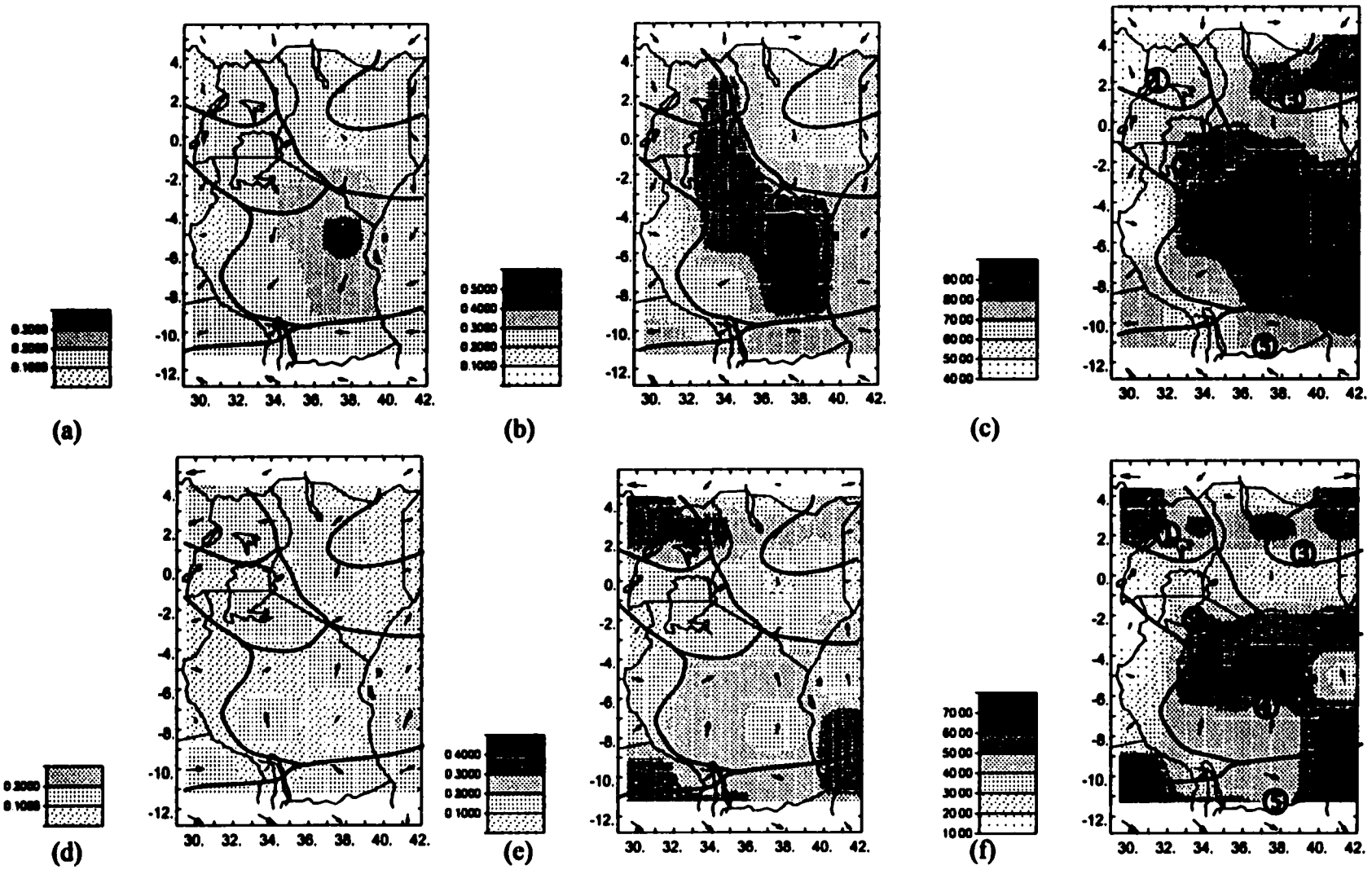


Figure 4.3 Plots showing (a) vector coherence of rainfall anomaly, (b) vector coherence of 6-month SPI, (c) variance (*ve*) extracted by the first Harmonic of El Niño composites; and the corresponding plots, (d) to (f) of La Niña composites. The graded shading represents either the vector coherence ((a), (b), (d) & (e)) or variance *ve* ((c) & (f)) while the size and direction of arrows represents the magnitude and phase of the first harmonic respectively. The five zones identified in each plot signify areas with homogenous ENSO response identified from harmonic analysis.

Based on Figure 4.3, we delineated five candidate regions that have distinct response patterns to ENSO signals (Table 4.2) assessed in terms of the magnitude, direction and general consistency of the harmonic vectors. The zones of El Niño influence delineated from rainfall anomaly composites are generally similar to those obtained from six-month SPI composite data (see Figure 4.3 (a) - (b), (d) - (e)). From these figures, it seems the influence of La Niña in East Africa is relatively modest compared to El Niño. Nicholson and Selato (2000) also did not find strong and coherent La Niña influence in East Africa.

Table 4.2 East African regions that have unique El Niño response patterns.

Region	Description
1	Northern Uganda including the Uganda – Sudan border area
2	Southern Uganda and much of Lake Victoria basin
3	Northeastern Kenya
4	Northeastern quadrant of Tanzania bordering the Indian Ocean
5	Southern Tanzania

Region 4 has the highest vector coherence (>60%) as well as the highest % of variance extracted by the 1st harmonic (>90%) which implies this zone has the most homogeneous relationship with the El Niño phenomena partly due to its proximity to the Indian Ocean. A somewhat similar zoning or pattern of influence seem to apply to La Niña composites albeit at a lower degree of clarity partly because the influence of La Niña is much more modest compared to El Niño.

Some areas of East Africa, e.g. central Kenya and southwestern Tanzania, do not seem to respond to the ENSO phenomena, probably because these areas are subjected to stronger local orographic influence than the global ENSO signals, e.g. central Kenya and southwestern Tanzania that are highland areas. In terms of ENSO response, vector maps in Figure 4.3 (b) & (c) show northwestern Tanzania to be almost totally independent of the rest of East Africa.

Region 1 has been different from other East African regions (e.g., Ogallo, 1988) partly because it is believed to receive its moisture from the tropical South Atlantic and moist Congo air mass while most of East Africa gets its moisture from the Indian Ocean via the northeasterly and northwesterly monsoons. Region 2, which includes the Lake Victoria Basin, seems to be reasonably influenced by El Niño even though it is generally wet all year round because of the influence of Lake Victoria. The statistically significant ENSO response periods are determined by the procedure outlined in Section 2.2.1 above. The response periods for the five regions are shown in Figure 4.4.

With respect to 90% bootstrap confidence limits, it seems that all five regions have identifiable periods of significant ENSO response with Region 5 having the most distinct (in terms of SPI magnitude) ENSO season (Figure 4.4). There is a general lag between the maximum ENSO phase and a region's climatic response indicated by the 6-month SPI. The degree of this phase lag depends on the region and the type of ENSO phase. Region 1's response in the El Niño year is almost opposite to that of the La Niña signal. The maximum response for Region 2 seems to occur in the winter at the end of a warm event year, while the response for Regions 3 and 4 reaches the peak in the last four months of an El Niño year. The maximum response for region 5 occurs between the first six months after an El Niño year.

There is a strong positive response for region 5 that occurs in the first six months after a La Niña event. On the other hand there is hardly any La Niña response in region 2. For regions 3 and 4 there is virtually no lag in responding to La Niña events. Region 3 and 4 are generally similar with respect to timing and phase of the ENSO response (Figure 4.4). The subtle difference between these two regions lies in the strength of the response. Both these regions receive their moisture from the Monsoon winds, which deposit moisture inland as they change direction due to a combination of Coriolis force and orographic channeling effects.

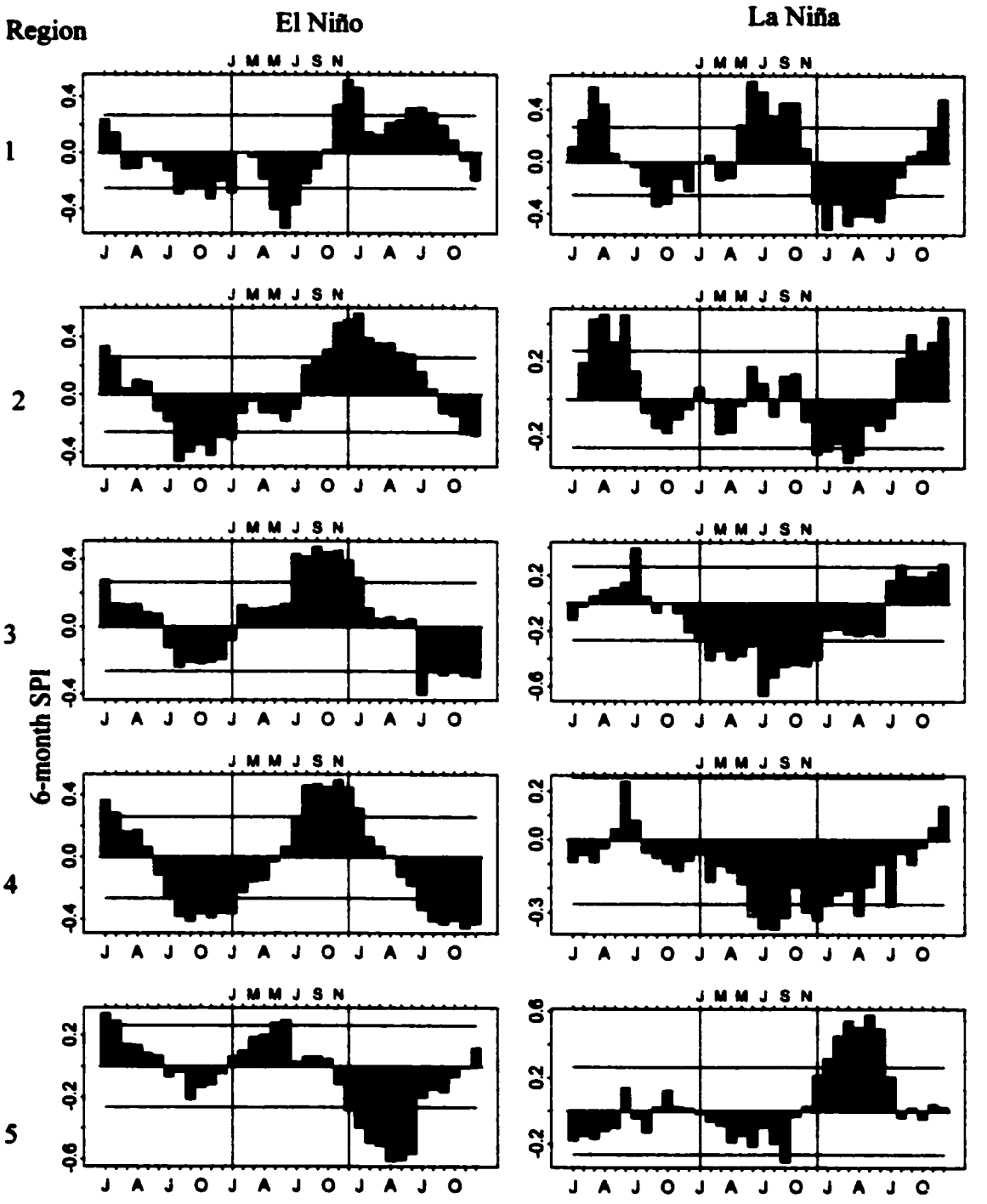


Figure 4.4 ENSO aggregate composites (3-year period) for 5 East African regions based on the six-month SPI. The two horizontal lines enclose 90% of the randomly re-sampled composites obtained by bootstrap methods. The 2 vertical lines represent the 12-month period of maximum ENSO activities. The histogram represents the 6-month SPI index averaged either from 22 episodes of El Niño, or 13 episodes of La Niña

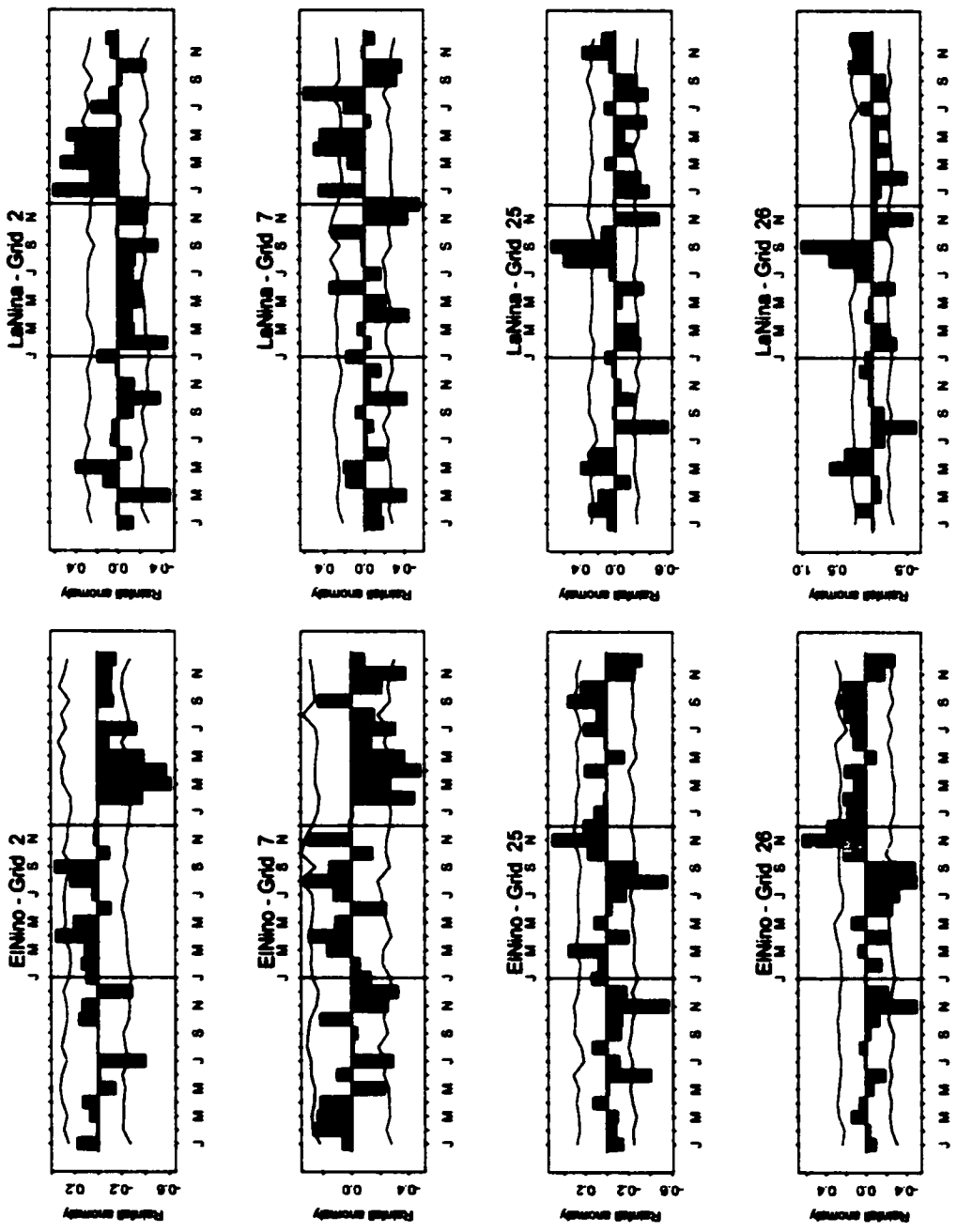


Figure 4.5. Samples of ENSO aggregate composites (3-year period) based on standardized rainfall anomalies for selected grids. The two horizontal lines enclose the 90% of the randomly re-sampled composite obtained by the bootstrap methods, while the 2 vertical lines represent the 12-month period of maximum ENSO activities.

- 103 -

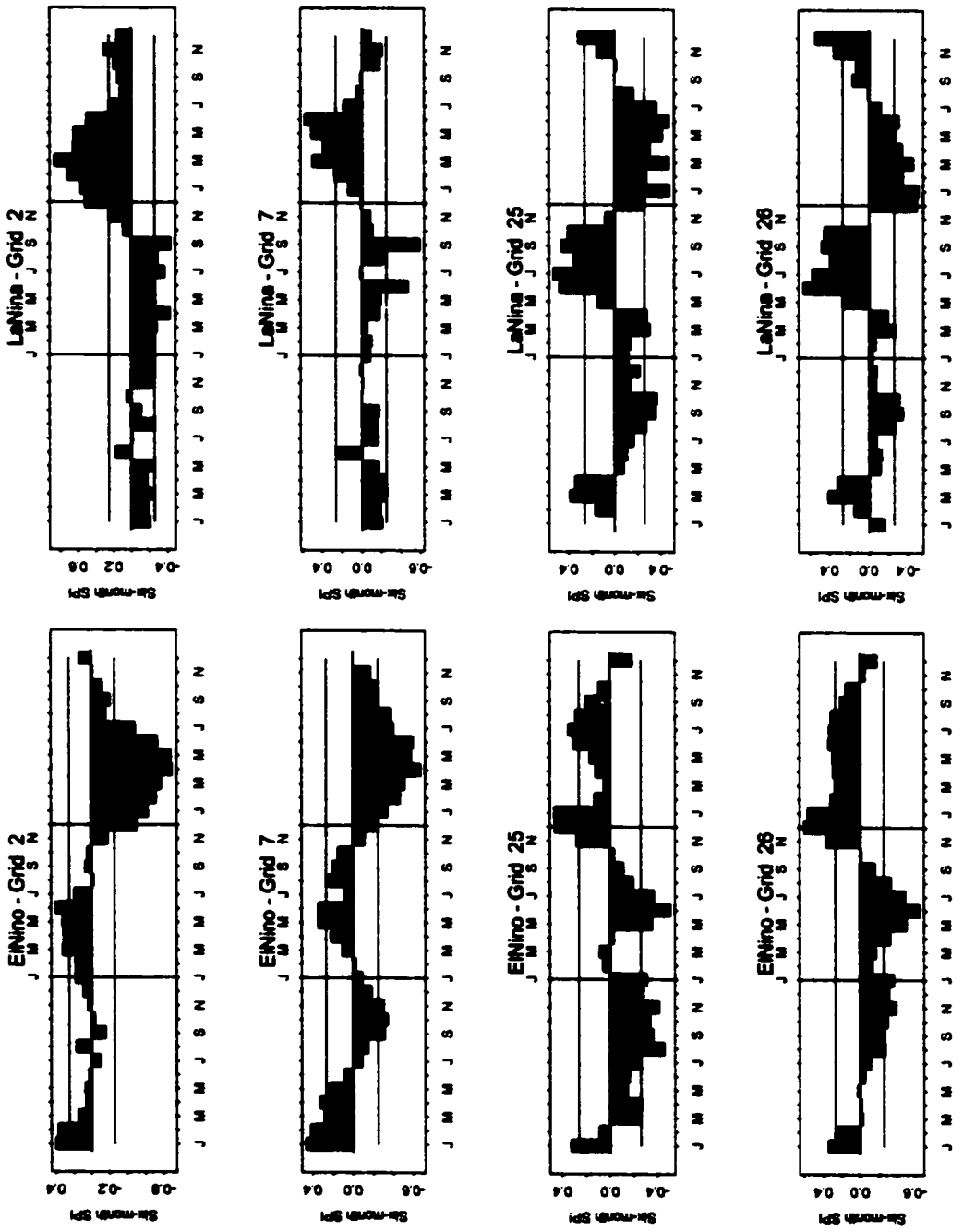


Figure 4.6. Samples of ENSO aggregate composites (3-year period) based on six-month SPI for selected grids. The two horizontal lines enclose the 90% confidence interval for the mean of a randomly re-sampled composite obtained by bootstrap methods, while the 2 vertical lines represent the 12-month period of maximum ENSO activities.

- 104 -

Regions 2 and 5 generally respond late to both cold and warm ENSO events (Figure 4.4). It is interesting to note that Lake Victoria basin (in Region 2) does show some positive feedback for November, December and January. The normal rainy seasons of this region are the northern hemisphere spring and autumn. It would seem that El Niño links the autumn rains, enhance the winter precipitation and the next spring "long rains" into one prolonged wet spell for Region 2. Region 5 seems to have more in common with the general South African circulation than the East Africa climatology, as observed by Nicholson and Kim (1997), and others.

The La Niña episodes are fewer than El Niño events. Even then, it is clear from Figure 4.4 that generally wet conditions prevail at the end of La Niña events for Region 5. It is also interesting to note that the El Niño 36-month aggregate composite for region 5 is almost a mirror reflection of its La Niña counterpart (see Figure 4.4) where the maximum response always occurs in the first half of the year after the ENSO year.

The exact mechanisms that control the co-variability of East African rainfall and ENSO are not well understood but our conventional knowledge of the general circulation suggests two primary factors. The first factor is associated with the eastward shift of the Walker circulation away from Central Africa towards East Africa during warm ENSO episodes. This tends to enhance rainfall for the short rains. ENSO is associated with the buildup of warm SST anomalies over the Indian and Atlantic Ocean basins, which would bring rainfall to some parts of East Africa (CLIVAR, 1999). The anomalously warm SST in the western Indian Ocean generates east-west SLP gradients that drive the observed near surface winds. This scenario of SST anomalies, which usually builds up in ENSO years, may account for much of the ENSO-East Africa teleconnection. On the other hand, a second mechanism that tends to counter the effect of the eastward shift in the walker circulation is the reduced surface pressure due to the presence of higher SST over the Indian Ocean adjacent to East Africa during warm ENSO events (CLIVAR, 1999). The reverse departures in surface pressure gradient probably result in offshore wind anomalies that defer the

transport of moisture towards some parts of East Africa, which probably contributes to reduced rainfall in Region 5 during warm ENSO episodes.

Figures 4.5 and 4.6 show a sample of ENSO composite aggregates for individual grids based on standardized rainfall anomalies and 6-month SPI respectively. The confidence intervals for the rainfall anomalies are not straight lines because they were established by bootstrap re-sampling methods on a month-by-month basis. Using precipitation anomaly data is less desirable than the six-month SPI because the former contain more data noise (Figures 4.5 and 4.6). However, it is still possible to observe some ENSO influence using the precipitation anomalies, although the response patterns are somewhat less defined when compared to those found in Figure 4.6 for the 6-month SPI, which with less data noise is able to show a more definite response to ENSO effects.

4.3.2 Index Time Series

Index time series (ITS) for the five regions (Figure 4.7) were formed to gauge the efficiency of the ENSO based seasonal zoning with respect to the seasons identified. In Region 3, the months of July to January within an El Niño year tend to have higher than normal 6-month SPI values that are statistically significant at 90% level. Therefore the ITS of say, Region 3, consists of 97 values of 6-month SPI averaged from July to the following January for the 1900 to 1996 period and these values are meant to show whether the delineated ENSO season actually exhibits behavior expected in ENSO years. The number of months and the months selected to form the ITS differ from region to region, and may differ between El Niño and La Niña ITS, depending on the months that respond positively (or negatively) to ENSO at 90% significance level (see Table 4.3).

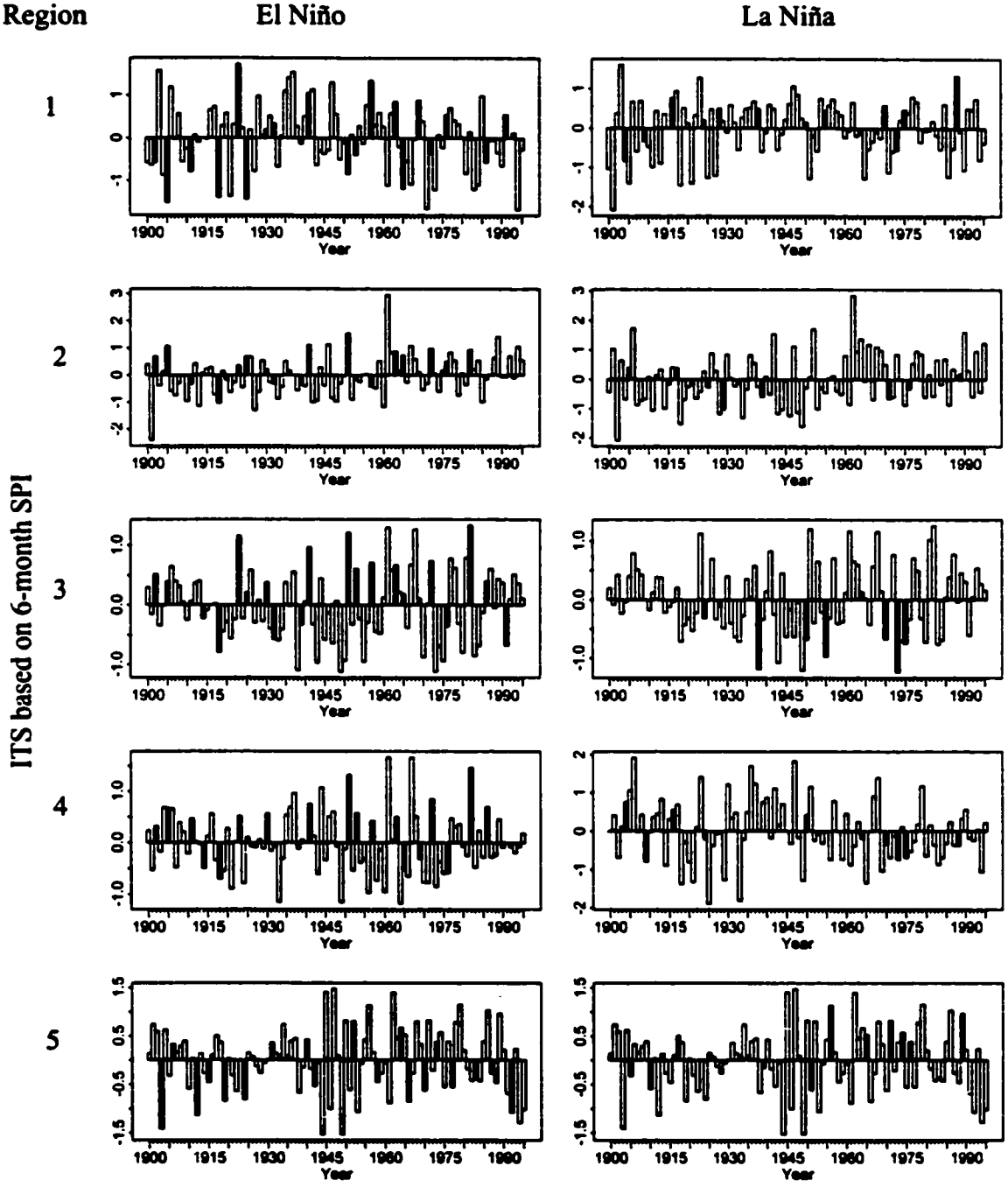


Figure 4.7. Index Time Series (ITS) based on 6-month SPI indices from 1900 to 1996 for Regions 1 to 5. The months that make up the identified ENSO response period change from region to region and between El Niño and La Niña. For example, the ITS of say, Region 3, consists of 97 values of 6-month SPI averaged from July to the following January for the 1900 to 1996 period (Figure 4.4). The dark columns represent either El Niño or La Niña years.

The ITS in Figure 4.7 confirm most of the previous findings. Most El Niño years in Region 2, 3 and 4 show a positive ITS value while Regions 1 and 5 had negative ITS values. The three worst droughts in Region 5 occurred during El Niño events as shown previously. As expected, the ITS values based on La Niña are less informative compared to El Niño whose influence is relatively strong except in Region 3, where the ITS shows that three worst droughts were all associated with La Niña events.

Table 4.3 Months used in the formation of the ITS series. (0) denotes ENSO year while (+) denotes post-ENSO year.

Zone	El Nino	La Nina
1	May (0) – July (0)	June (0) – October (0)
2	October (0) – April (+)	January (+) – April (+)
3	July (0) – January (+)	July (0) – December (0)
4	August (0) – January (+)	June (0) – September (0)
5	January (+) – June (+)	January (+) – June (+)

4.3.3 Shift in the Distribution of 6-month SPI under ENSO Influence

Boxplots in Figure 4.8 clearly show the shifts caused by ENSO in the percentile distributions of 6□month SPI in East Africa. These shifts are quite significant for some regions in certain seasons. The boxplots for Region 1 show that for the summer, dry conditions are more likely to occur during El Niño than non-ENSO years. In contrast, summers of La Niña years are likely to be wet in this region. The median drought indices of summer La Niña is as high as the 75th percentile drought indices of the summer El Niño. The opposite occurs during the winter seasons when Region 1 is likely to be wetter during El Niño years. Essentially, La Niña winters in Region 1 seem to be much drier than El Niño winters. These results agree with those obtained from ENSO aggregate composites, which show that for Region 1 La Niña (El Niño) events exert a negative (positive) influence on the winter precipitation after the ENSO year.

Conversely Region 2 (Lake Victoria Basin and surrounding areas) is likely to have wetter conditions during the spring long rains after an El Niño year compared to non-ENSO years. The 50th percentile of spring long rains in an El Niño year are as high as the 75th percentile of a non-ENSO year. Relatively wet winters after an El Niño year shown here agree in conformity with aggregate composites in Figure 4.4.

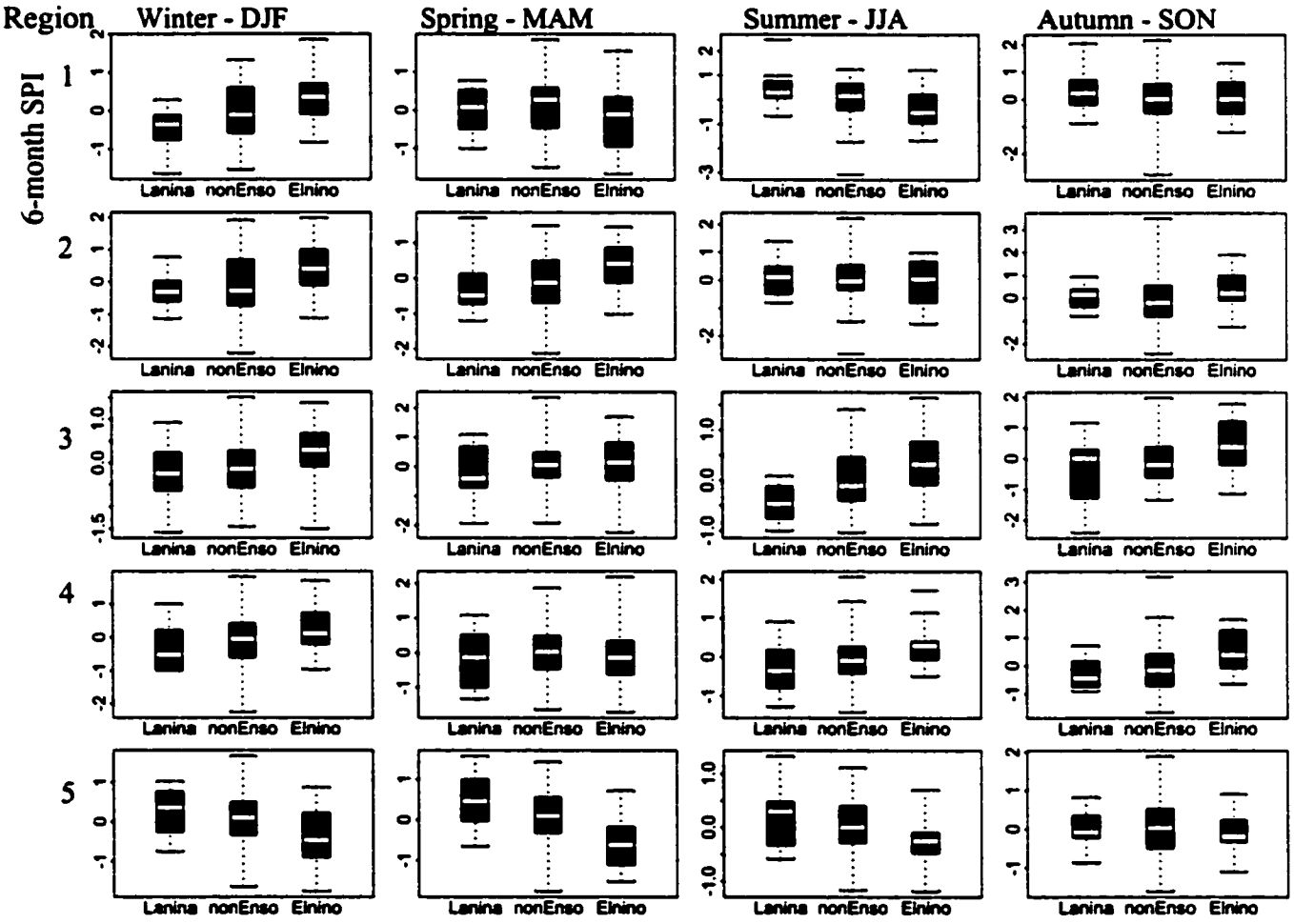


Figure 4.8. Boxplots of 6-month SPI for El Niño, La Niña and non-ENSO periods for Regions 1 to 5. The middle line in the box represents the median; the top and bottom of the box represents 75th and 25th percentile, while the top and bottom brackets represent the maximum and minimum values, respectively.

There is little or no influence of ENSO on the autumn short rains in region 1, 2 and 5. This is in contrast to the findings by Nicholson and Kim (1997) who suggested that

short rains are enhanced in equatorial East Africa during ENSO years, but the spring long rains are reduced. The difference could be attributed to several factors. First, we used the bootstrap methods to identify statistically significant anomalies during ENSO response periods, whereas Nicholson and Kim only relied on the location of peaks and troughs in the first harmonic to identify the periods of maximum and minimum ENSO responses. Second, their definition of equatorial Africa covers almost the entire East Africa, while we subdivided East Africa into 5 regions. Third, their definition of rain seasons differ from that of East African Meteorological Department (1963) by a month, e.g., their autumn or "short rains" season is OND instead of SON; in other words, their results show that rainfall in OND and JFM are enhanced but slightly reduced in AMJ. A better representation of the rain seasons in East Africa should be that we adopted in this analysis; e.g., in many parts of East Africa the long rains are in March to May (MAM). Based on the above factors, we believe our results are more representative than that of Nicholson and Kim (1997).

The El Niño summer, autumn and winter seasons of regions 3 and 4 are clearly wetter than those of non-ENSO years. Conversely, La Niña summers and autumns are more likely to be dry compared to non-ENSO years. This response is more critical to region 3, which receives about 45% of its annual precipitation during the second half of the year than region 4, which gets only 15% of precipitation in the same period. The ENSO response of region 5 seems to be mostly opposite of other regions. Winter and spring after an El Niño event seem to be particularly dry while those after La Niña events are likely to be wet. ENSO seems to exert little influence on autumn short rains in Region 5.

It is clear from these boxplots that the ENSO response is both season and region dependent and some seasons are more strongly affected than others during an ENSO event. These observed influence of ENSO might be used in diagnostic monitoring of East African droughts, in anticipating rainfall anomalies whenever warm or cold ENSO phases are established.

4.4 Summary and Conclusions

This investigation examined the influence of ENSO on East African moisture regime represented by the 6-month Standardized Precipitation Index (SPI) using harmonic analysis, composites of 22 El Niño and 13 La Niña events respectively, and 90% confidence limits established from bootstrap re-sampling. It was found that ENSO responses in East African rainfall are region dependent, and the influence of El Niño is stronger and opposite that of La Niña. Five regions of unique ENSO responses have been identified, such that Region 4 has the highest ENSO vector coherence (>60%) and also the highest percentage of variance extracted by the first harmonic (>90%). High vector coherence probably means that the ENSO response has remained fairly constant and un-modulated by any low frequency variability. However, areas such as the central highlands of Kenya and southeast highlands of Tanzania have no discernable ENSO response probably because their rainfall generating mechanisms are more localized than synoptic.

ENSO responses are also season dependent. The length and temporal locations of these identified "ENSO seasons" vary from region to region, but they could be useful for preparing East Africa against potential droughts, particularly for Regions 4 and 5, which seem to have the most consistent and distinct ENSO responses. Region 5 (southern Tanzania) experiences positive (negative) response under La Niña (El Niño) influence during January and June of the post ENSO year. Southern Uganda and much of Lake Victoria basin show some significant positive ENSO response for November, December and January. La Niña usually precedes drought in most of East Africa except for Northern Uganda and Southern Tanzania, where El Niño is a precursor of droughts.

From the index time series (ITS) formed out of 6-month SPI, we found that Regions 2, 3, and 4 have positive ITS values during El Niño years while regions 1 and 5 have negative ITS values during warm ENSO years. Boxplots show a shift in the distribution of 6-month SPI between ENSO and non-ENSO affected seasons. For

Region 1 (northern Uganda), summers in El Niño years are dry but the ensuing winters are often wet compared to non-El Niño years. All winter responses are positive to El Niño except for Region 5, which has a negative (positive) winter El Niño (La Niña) response. The boreal spring ENSO response is mixed except for regions 2 and 5, which have positive and negative El Niño response respectively.

Our results did not concur with that of Nicholson and Kim (1997), who suggested that rainfall in equatorial Africa is enhanced in the short rains (autumn) season of the ENSO year, but reduced during the ensuing spring rains. Our findings show more details about the East Africa □ ENSO teleconnection dipole structure established by Ropeleweski and Halpert (1987) in their pioneer work. Further, the difference could be partly attributed to Nicholson and Kim's (1997) definition of the equatorial Africa that covered almost the entire East Africa in one region, whereas we divided East Africa into 5 regions; their definition of rain season that differed from ours by a month (e.g., OND instead of SON), and the different approach used in obtaining the maximum (minimum) ENSO response periods. For reasons given in Section 3.2, our results are probably more representative than that of Nicholson and Kim (1997).

Chapter 5 Analysis and Prediction of East African Seasonal Droughts Using Canonical Correlation and Projection Pursuit Regression

Meteorological droughts are extended periods of anomalously low rainfall. The emphasis of this investigation is on the analysis and prediction of seasonal droughts in East Africa using (i) a Canonical Correlation Analysis model (CCA) with climatic inputs systematically and optimally weighted by Simplex (Nelder and Mead, 1965), and (ii) a nonlinear projection pursuit regression model.

There have been several studies that teleconnected climatic signals such as Southern Oscillation and Sea Surface Temperatures with East African rainfall (Ogallo, 1988; Ropeleweski and Halpert, 1996; Nicholson and Kim, 1997). However, as far as we know, there is no study that translates this knowledge into an operational rainfall prediction model in East Africa. Further, past application of CCA are predominantly based on one type of climatic signals as input, or occasionally based on two types of signals combined in an ad hoc manner.

Even though the dynamics of climatic phenomena are predominantly nonlinear, linear statistical models such as CCA are usually preferred over non-linear models because their simplistic nature makes them relatively easy for users to understand and apply. Rather than driving the CCA with only one climatic signal as has been done in the past, we optimize the weights applied to two input climatic signals (SST and SLP) to CCA by an automatic, direct search algorithm called Simplex (Nelder and Mead, 1965). The aim is to improve the forecast skill of CCA based on optimally weighted SST and SLP since both input variables play a part in the rainfall climatology of East Africa.

Beyond a linear statistical model such as CCA, this study also investigates the Projection Pursuit Regression (PPR) model, which is capable of modeling both linear and nonlinear relationships as a predictive tool for East African drought.

5.1 Overview of Canonical Correlation Analysis (CCA).

Canonical correlation analysis (CCA) has been used to forecast the surface climate in various places of the world such as the continental USA (Barnston, 1994), winter rainfall in Hawaii (Chu and He, 1994), May-June (Mei-Yu) rainfall for Taiwan (Chu, 1998) and seasonal precipitation in tropical Pacific Islands (He and Barnston, 1996; Yu et al., 1997). Each of the above investigators implemented a slightly different flavor of CCA to fit the location being studied e.g. the manner in which the input variables were preprocessed and put together.

Although CCA was developed in the 1930s by Hotelling (1936), its potential for geophysical applications was noted only in the 1960s by Glahn (1968). However, CCA did not gain popularity until the late 1980's. For example, Nicholls (1987) used CCA to analyze Darwin pressure, Tahiti pressure, southeastern Australian rainfall and Willis Island air temperature.

Many studies have shown that SST is one of the most important boundary forcings on the atmospheric circulation. By using the dominant eigenmodes of the Pacific SST as a predictor, Chu (1998) was able to forecast by 4 to 6 months lead-time Mei-Yu rainfall for Taiwan with moderate skill. He used dominant eigenmodes of the SST in order to reduce the large dimensionality of the Pacific SST data set.

Canonical correlation can be used to analyze complete as well as composite data sets. In the latter, certain years of data are selected according to some criterion, e.g., when a base index such as a rainfall index or Southern Oscillation Index (SOI) exceeds a certain threshold. The mean values of various fields corresponding to those selected

years are calculated (e.g. Piechota et. al. 1997; Pan and Oort, 1983). This approach may be appropriate if the events being teleconnected can be considered as a sequence of discrete "episodes" separated by periods within which the variations are of less interest. In this investigation, we use CCA to analyze a composite data set consisting of the driest (bottom 10%) SON rain seasons of the available 1900-1997 data.

5.1.1 The Canonical Correlation Analysis (CCA) Model

CCA is the most general form of correlation analysis among a family of correlation techniques. A multiple regression is a specifically one dependent variable correlated with multiple independent variables, while a product moment correlation is even more restrictive, where one dependent variable is correlated only with one independent variable. However, CCA can be used to investigate the inter-correlation between two multivariate datasets. The derivation of CCA (mainly after Glahn, 1968) is briefly summarized below.

Suppose there are n observations of p variables X_i (i= 1, 2, ... p) and q variables Y_i (i= 1, 2, ... q) in a $(p + q)$ dimensional space that can be arranged in a matrix form as ${}_nX_p$ and ${}_nY_q$. The variables X_i and Y_i have means $\overline{X}_i$ and $\overline{Y}_i$ respectively and deviations from the mean are given by $x_i = X_i - \overline{X}_i$ and $y_i = Y_i - \overline{Y}_i$.

New variables ${}_nx_pA_i$ and ${}_ny_qB_i$ (i =1, 2, ..., r) where r is less than or equal to the smaller of p and q, can be formed such that their means are zero. The directional matrices ${}_pA_r$ and ${}_qB_r$ are the canonical vectors of x and y respectively.

The variance - covariance matrices are given as

$$ {}_pS_{11p} = \frac{1}{n}{}_p x'_n x_p \qquad (1) $$

$$ {}_pS_{12q} = {}_pS'_{21q} = \frac{1}{n} {}_px'_n y_q \quad \text{.....(2)} $$

$$ {}_qS_{22q} = \frac{1}{n} {}_qy'_n y_q \quad \text{.....(3)} $$

The relationship between the new variables ${}_nx_pA_i$ and ${}_nyB_i$ can be shown to be:-

$$ {}_rA'_pS_{11p}A_r = I_r \quad \text{.....(4)} $$

$$ {}_rB'_qS_{22q}B_r = I_r \quad \text{.....(5)} $$

$$ {}_rA'_pS_{12q}B_r = \Lambda_r \quad \text{.....(6)} $$

where I_r is a unit matrix, A′ and B′ is a transpose of A and B respectively.

$$ \Lambda_r = \begin{bmatrix} \xi_1 & & & \\ & \xi_2 & 0 & 0 \\ & 0 & . & 0 \\ & 0 & 0 & . \\ & & & \xi_r \end{bmatrix} \quad \text{.....(7)} $$

where $\xi_1 \geq \xi_2 \geq \ldots \geq \xi_r$.

Equations (4) and (5) state that the variance of each of the new variables is unity and each is uncorrelated with all others in its respective set. Equation (6) and (7) state that each element of ${}_nx_pA_i$ is uncorrelated with that of ${}_ny_qB_i$ except for i = j when the correlation is ξ_i. The ξ_i^2 are known as the latent roots which indicate the size of patterns that are common to both data sets. The occurrence of a single non-zero latent root indicates that by appropriate weighting of the predictand ${}_nX_p$ and predictors ${}_nY_q$, one set of the variables can be projected into the space of the other along a straight line. The existence of two non-zero roots indicates that each set projects linearly into a two-dimensional plane of the other set and so on. The square roots of the latent values are known as the canonical correlations.

To obtain the canonical correlations and the associated weights, Equation 6 is maximized within the limiting conditions imposed by Equations 4 and 5. The steps involved are as follows:-

A Lagrangian multiple of Equation 6 is formed, thus

$$L = {}_rA'_pS_{12q}B_r - \theta({}_rA'_pS_{11p}A_r - I_r) - \eta({}_rB'_qS_{22q}B_r - I_r) \quad \text{.....(8)}$$

Taking partial derivatives with respect to A and B and equating to zero we obtain Equation (9) and (10).

$${}_pS_{12q}B_r = 2\theta\,{}_pS_{11p}A_r \quad \text{.....(9)}$$

$$({}_qS_{12p})'A_r = 2\eta_q\,S_{22q}B_r \quad \text{.....(10)}$$

where θ and η are Lagrange multipliers.

Equation (9) and (10) are pre-multiplied with ${}_rA'_p$ and ${}_rB'_q$ respectively, yielding:

$${}_rA_pS_{12q}B_r = 2\theta I_r \equiv 2\eta I_r \quad \text{.....(11)}$$

Solving Equations 9 and 10 yields the following relationships:

$${}_pS^{-1}_{11p}S_{12q}S^{-1}_{22q}S_{21p}A_r = 4\theta^2\,{}_pA_r \quad \text{.....(12)}$$

$${}_qS^{-1}_{22q}S_{21p}S^{-1}_{11p}S_{12q}B_r = 4\theta^2\,{}_qB_r \quad \text{.....(13)}$$

It can be seen from the format of Equations 12 and 13 that ${}_pA_r$ and ${}_qB_r$ are eigenvectors. Hence, ${}_pA_r$ and ${}_qB_r$ can be found from

$$({}_pS^{-1}_{11p}S_{12q}S^{-1}_{22q}S_{21p} - \xi_i^2\,{}_pI_p)A_r = 0 \quad \text{.....(14)}$$

$$({}_qS^{-1}_{22q}S_{21p}S^{-1}_{11p}S_{12q} - \xi_i^2\,{}_qI_q)B_r = 0 \quad \text{.....(15)}$$

The ξ_i satisfy the determinantal equations

$$\left|{}_pS^{-1}_{11p}S_{12q}S^{-1}_{22q}S_{21p} - \xi^2\,{}_pI_p\right| = 0 \quad \text{.....(16)}$$

$$\left|{}_qS^{-1}_{22q}S_{21p}S^{-1}_{11p}S_{12q} - \xi^2\,{}_qI_q\right| = 0 \quad \text{.....(17)}$$

Equations 16 and 17 are the fundamental canonical equations.

The eigenvectors ${}_pA_r$ and ${}_qB_r$ can be scaled up or down by any arbitrary constants and still satisfy Equations 14 and 15. On the other hand it should be noted that only a unique set of ${}_pA_r$ and ${}_qB_r$ can satisfy Equations (4), (5) and (6). Most computational packages return normalized eigenvector solutions, which may not be necessarily the correct solutions for Equations (4) to (6). Appropriate scaling factors for the eigen vectors can be obtained by substituting the preliminary eigen solutions into Equation (4) and (5). For example if the preliminary eigenvector is ${}_rA_p^*$, then

$$({}_rA_p^*)'S_{11}{}_rA_p^* = \psi \quad \text{.....(18)}$$

The correct eigenvector solution is obtained by multiplying ${}_rA_p^*$ with the diagonal matrix $\psi^{-1/2}$.

The predictand variables ${}_ny_q$ can be predicted in a least-square sense by the predictor variables ${}_nx_q$ using the following equation (Glahn, 1968),

$${}_n\hat{y}_q = {}_nx_pA_r\Lambda_rB_q'S_{22q} \quad \text{.....(19)}$$

It is possible that not all the canonical correlations ξ_i (i = 1, 2, . . r) are significantly different from zero. Typically, ξ_i of small magnitude (≤ 0.3) are of little value, as they usually account for very little variability in the data. The significance of these correlations can be examined using the Bartlett's Chi-Square test. If only φ numbers of the canonical correlations (or latent roots) are significant, we can replace ${}_r\Lambda_r$ by ${}_\varphi\hat{\Lambda}_\varphi$ where the latter contains only the first φ latent roots. Equation (19) is then modified to:

$${}_n\hat{\hat{y}}_q = {}_nx_pA_\varphi\hat{\Lambda}_\varphi B_q'S_{22q} \quad \text{.....(20)}$$

The prediction Equation (19) may be expressed in terms of ${}_nX_p$ and ${}_nY_q$ if desired, such that

$$ {}_n\hat{Y}_q = {}_nX_pA_r\Lambda_rB_q'S_{22q} - {}_n\overline{X}_pA_r\Lambda_rB_q'S_{22q} + {}_n\overline{Y}_q \quad \text{.....(21)} $$

If the original variables ${}_nX_p$ and ${}_nY_q$ were standardized, Equation (21) will reduce to (19) since ${}_n\overline{X}_p$ and ${}_n\overline{Y}_q$ are null matrices.

5.1.2 Principal Component Analysis (PCA) and CCA.

It is highly advisable to "reduce" the input data by the EOF or PCA before inputting them to the CCA (Yu et al., 1997; Shabbar and Barnston, 1996) and it is advisable to retain only a limited number of dominant principal components as input to CCA. This pre-processing reduces the large *(p+q)* dimensional space of the model input data into that of retained EOF modes. Moreover, by retaining the dominant modes of variability within each data set, relatively minor data noise is being filtered out. However it is possible that by pre-processing the data using PCA may destroy some correlation between the input variables.

An important advantage of using PCA prior to CCA (which has not been explicitly mentioned by others) is that pre-orthogonalizing the data greatly simplifies the process of solving the canonical determinant Equations (16) and (17). PCA transforms the data into orthogonal variables that are not correlated to each other. The covariances (S_{11} and S_{22}) of the orthogonal variables are all identity matrices whose inverse are also identity matrices. This means that when PCA is applied to the data fields, it is no longer necessary to find the inverse of S_{11} and S_{22} in Equations (16) and (17), thereby avoiding that extra computational work which may at times be error prone if the data matrices are ill-conditioned.

A brief description of PCA is as follows:- Let x be the (n x p) demeaned observed data matrix where n is the number of cases and p the number of variables. Then the j^{th} principal component T_j of the sample of p variate observations is the linear compound

$$T_j = e_{1j}x_1 + \ldots + e_{pj}x_p \quad \ldots\ldots(22)$$

whose coefficients e_{ij} are the elements of the characteristic vectors of the sample covariance matrix S (or correlation matrix R if one opts for the correlation input) corresponding to the j^{th} eigen value λ_j. In other words, the matrix x is transformed into the (n x p) matrix T scores and (p x p) matrix E according to

$${}_nT_p = {}_nX_pE_p \quad \ldots\ldots(23)$$

subjected to two conditions: (1) the transformed variables are uncorrelated, and (2) each transformed variable accounts for the maximum total variance in the original data set in the given direction of the eigen vector. The transformational matrix E is orthogonal (i.e. ${}_pE_pE_p^t = I_p$) and is the normalized eigenvector of the covariance matrix, S, of the original variables.

In some literature, T is referred to as empirical orthogonal functions (EOF) which are mutually uncorrelated, and the matrix E is referred to as empirical orthogonal weights -EOW (Burke, 1996). The covariance of EOF is a diagonal matrix containing the eigen values of S in descending order i.e.

$$\frac{1}{(n-1)}{}_pT_n^tT_p = {}_p\lambda_p \equiv {}_pE_p^tS_pE_p \quad \ldots\ldots(24)$$

where ${}_p\lambda_p$ is the diagonal eigen-value matrix. The matrices T and E obtained from the above EOF analysis may be sometimes re-scaled as follows

$${}_n\vec{t}_p = {}_nT_p\lambda_p^{-\frac{1}{2}} \quad \ldots\ldots(25)$$

$${}_pL_p = {}_pE_p\lambda_p^{\frac{1}{2}} \quad \ldots\ldots(26)$$

The new variable ${}_n\bar{t}_p$ is referred to as principal components or PCs and has zero mean and unit variance. The variable L is referred to as component loadings. The loadings represent the association between the PC components and the original variables. Each element in matrix L shows the amount of variance explained by a given PC in a given variable.

Un-rotated PCA solutions have been shown to suffer from four deficiencies, namely, domain shape dependence, sub-domain instability, sampling problems and inaccurate portrayal of the physical relationships embedded within the input matrix (Richman, 1986). On the other hand, un-rotated solutions have several advantages – namely; the ability to extract maximal variance from the dataset, their spatial and temporal orthogonality and their pattern insensitivity to the number of PC's retained. These characteristics make un-rotated principal component solutions more appropriate for situations where pure data reduction is being sought (Richman, 1986). Hence in the CCA – PCA model, only un-rotated PCA components are used.

The setup of our model necessitated the calculation of principle components of a large predictor data set several thousand times. Obtaining PCs of a large dataset is a computer intensive process because of the iterative procedure involved in obtaining the eigen solution of the large covariance matrix associated with the data. However if the data matrix is "fat and short" (i.e. n<<p), matrix manipulations may be done which can cut down the computer time to less than a hundredth of the time it would otherwise take. In such a situation, the first *n* PCs can be obtained by finding the eigen solution of the pseudo-covariance matrix S′:-

$$S' = {}_n x_p x'_n \frac{1}{n} \qquad \text{.....(27)}$$

The mean of X must be subtracted from the matrix *before* transposing it, giving rise to S′ which is a different covariance matrix from S_{11} in Equation 1. The eigenvalues have to be scaled by a factor equal to the number of variables divided by the number

of cases (*p/n*). In addition, the correct eigenvector is obtained by pre-multiplying the eigen solution of S′ in Equation 27 with x, and then normalizing the product. The final solution will contain the correct first n eigenvectors.

In terms of the principal components, the input data sets can be written as;

$$ {}_{n}x_{p} = {}_{n}(\vec{x}_t)_p [(\lambda_x)_p^{1/2} (E_x)'_p] \quad \text{.....(28)} $$

$$ {}_{n}y_{q} = {}_{n}(\vec{y}_t)_q [(\lambda_y)_q^{1/2} (E_y)'_q] \quad \text{.....(29)} $$

where $(E_x)'$ and $(E_y)'$ are the transpose of the PCA eigen vectors of the predictors and predictand respectively, λ_x and λ_y their respective eigen values, and ${}_{n}(\vec{x}_t)_p$ and ${}_{n}(\vec{y}_t)_q$ their respective time coefficients or PCA component. The square bracketed quantities in Equations 28 and 29 represent the spatial component modes of the predictors and predictand respectively.

When using PCA pre-filtered data in the CCA model, ${}_{n}(\vec{x}_t)_i$, $(i \le p)$ and ${}_{n}(\vec{y}_t)_j$, $(j \le q)$ are used in Equations 1 to 3 instead of ${}_{n}x_p$ and ${}_{n}y_q$. The fundamental canonical Equations 16 and 17 are then used to obtain the canonical vectors. An extra step is now required to convert the predicted time components ${}_{n}(\hat{y}_t)_j$ to the original units of the input predictand data ${}_{n}y_q$. This is achieved by using the orthogonal characteristics of EOF modes. Hence,

$$ {}_{n}\hat{y}_{q} \approx [{}_{n}(\vec{x}_t)_i A^{*}_{r} \Lambda^{*}_{r} B'^{*}_{j} S^{*}_{22j}][(\lambda_y)_j^{1/2} (E_y)'_q] \quad \text{.....(30)} $$

Several terms in Equation 30 are star-superscripted to differentiate them from the quantities of a similar Equation 19. The covariance matrix S^{*}_{22} need not be included in Equation 30 because it is an identity matrix (due to the orthogonal properties of the EOF modes). The input matrix ${}_{n}(\vec{x}_t)_i$ has to be estimated using the predictor EOF loading pattern. If the new predictor information is ${}_{m}X_p$ then the least squares estimate for the new predictand variables ${}_{m}Y_q$ is given as

$$ {}_m\hat{Y}_q = [{}_mX_p - {}_m\overline{X}_p][{}_p(E_x)_p(\lambda_x)^{-\frac{1}{2}}_{xeof}]_{xeof} A_r \Lambda_r B'_{yeof} [(\lambda_y)^{\frac{1}{2}}_{yeof} (E_y)'_q] + {}_m\overline{Y}_q \quad \text{.....(31)} $$

where *xeof* and *yeof* are the number of PCA modes retained for the input into the CCA analysis.

It is desirable to show the relationship between the canonical variates and the original variables. The correlation between the original predictor data and the predictor canonical variates is given by:-

$$ {}_pg_r = \left[\left({}_p(E_x)_p(\lambda_x)^{-\frac{1}{2}}_{xeof} \right)' (\sigma_x)^{-1}_p \right]' {}_{xeof}A_r \quad \text{.....(32)} $$

while that between the original predictand data is given by

$$ {}_ph_r = \left[\left({}_q(E_y)_q(\lambda_y)^{-\frac{1}{2}}_{yeof} \right)' (\sigma_y)^{-1}_p \right]' {}_{yeof}B_r \quad \text{.....(33)} $$

where σ_x and σ_y are diagonal vectors containing the standard deviations of the respective predictor and predictand variables.

The canonical maps **g** and **h** are not unit vectors nor are they mutually orthogonal. For ease of interpretation they may be normalized such that the sum of **g'** or **h'** components for a given canonical mode is equal to unity. Using the normalized canonical maps we may gauge the relative importance of the predictor data. However, if the predictor data set has too many variables, normalizing the g map might obscure the relationships since the normalized vectors are likely to have very small values.

5.1.3 Application of the CCA model to the East African Seasonal rainfall data.

5.1.3.1 The Data Fields.

The principle seasonal rainfall dataset for this analysis was extracted from the 'gu23wld0098.dat' (Version 1.0) archive constructed and supplied by Dr Mike Hulme

at the Climatic Research Unit, University of East Anglia, Norwich, UK (Hulme 1994). This is a historical monthly precipitation dataset for global land areas from 1900 to 1998, gridded using 2.5° latitude by 3.75° longitude resolution. The location of this rainfall grid is shown in Figure 5.1.

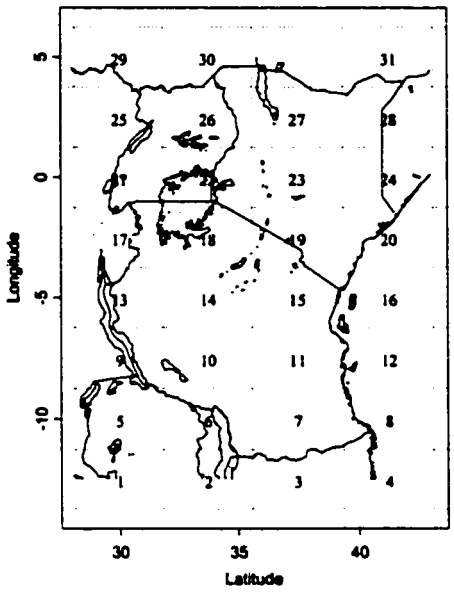


Figure 5.1 Map of East Africa showing the rainfall grids.

The focus of this investigation is drought analysis and prediction at one or more seasonal lead times. Hence, it is necessary to clearly delineate the rainfall seasons in the region. Previous studies such as that of Griffiths (1972) have pointed out the ambiguity as well as the complexity of rain seasons in the region, e.g., there is no single month during the year when all parts of East Africa are dry. A description of the various rain seasons is given in Chapter 2.

Various homogenous rainfall zones are probably modulated by different sets of climatic factors and so it is important to delineate the extent of those areas that seem to share common seasonal rainfall patterns.

In this investigation we chose harmonic analysis (Chapter 2) as the method to regionalize East African rainfall. Basalirwa (1995) and Ogallo (1989) used PCA to

carry out the above analysis. Griffiths (1972) had earlier on arrived at some descriptive homogeneous zones by compositing 12 East African maps each depicting the location of the 50mm isohyet for a given month (Jan to Dec). The results of the studies by the above three authors produced too many fragmented regions (in excess of 20 distinct zones) that were unsuitable for our analysis. On the other hand, the application of harmonic analysis to East African monthly rainfall yielded only six major zones (see Figure 2.5). Details of this analysis are given in Chapter 2. Assessing the skill of the statistical prediction models are in terms of these six zones. It is probable that droughts may be more predictable in some zones than others.

Some seasons in some parts of East Africa do not contribute much rainfall. The normally dry periods may not constitute droughts since they are not anomalous. Rather, it is the failure of the critical wet rain seasons that trigger off drought conditions (see Chapter 2). Based on Table 2.1's the important rain seasons in East Africa, it was decided that our focus should be on modeling the March – May (MAM) and September – November (SON) rain periods. In East Africa, these seasons are known as the "long" and the "short" rains respectively. These two seasons contribute more than 70% of the annual rainfall in many parts of East Africa.

The predictor fields in this investigation were the Sea Surface Temperature (SST) and Sea level Pressure (SLP). The global SST 5° x 5° grid data for the period 1856-2000 was obtained from the Climatic research Unit of the University of Anglia, UK. This data is expressed as anomalies with respect to 1961-1990 as the base period. The SLP data was obtained from the Global Mean Sea Level Pressure dataset (GMSLP2.1) provided by the UK Meteorology office. GMSLP2.1 is a global, observed monthly historical mean sea level dataset. The dataset begins in 1871 and is composed of 5° x 5° gridded values. The quality control of this dataset is discussed by Basnett and Parker (1997). We transformed the GMSLP2.1 data into sea level pressure anomalies using the 1961-1990 climatology as the base period. Thus both predictor fields were anomalies with respect to the 1961-1990 period. In this investigation, we focused on

the teleconnection between the predictor fields of the Indian and Atlantic oceans and East African rainfall. The selected predictor zones are shown in Figure 5.2.

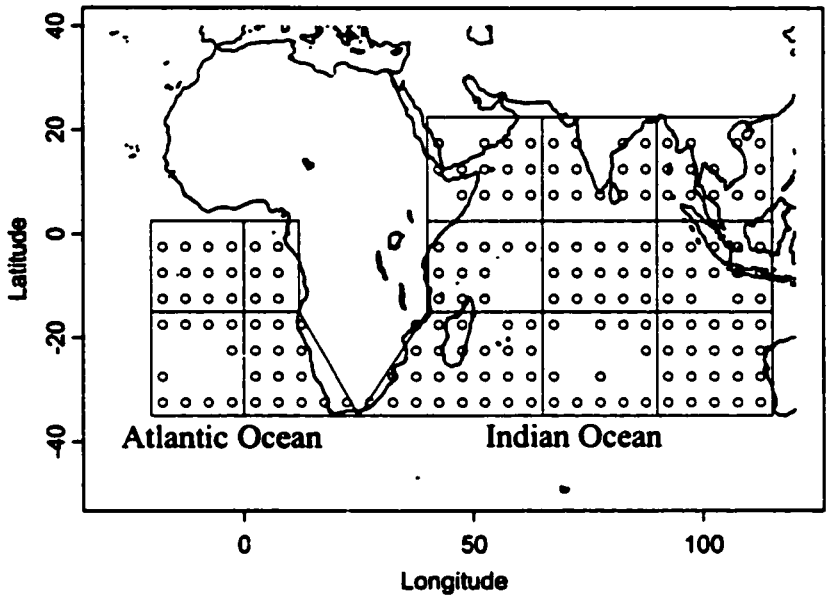


Figure 5.2 shows the Ocean zones used in the teleconnection model. The dots show the actual location of the continuous (1900-1997) anomaly SST grid data used in the model. The large boxes outline the subzones whose data is further weighed using coefficients obtained by the Simplex algorithm of Nelder and Mead (1965).

Considerable effort was spent ensuring that only predictor grids with complete data were entered in the model (Figure 5.2). Where necessary, an interpolation scheme was used to estimate the missing SST or SLP grid data based on surrounding observed values. Grid points that had more than 20% of their data missing were excluded from the model. In particular, the 1942-45 years had many missing SLP and SST values especially in the central-southern Indian Ocean and southern Atlantic Ocean.

5.1.3.2 The Specific CCA Model Setup

An inherent lead-time is built within the model. This lead time or the amount of time between the end of the latest predictor season and the end of the three month

predictand season can be varied from three to twelve month periods at 3-months increments. However, our investigation was confined to one season lead-time prediction of March-May (MAM) and September-November (SON) rainfall only.

The predictor and predictand fields were arranged in such a way so as to incorporate both space and time evolution of the climate system that gives rise to the predictive skill. For example, using two previous seasons, March to May and June-August, to predict September-November rainfall probably accounts for some temporal evolution of the teleconnection mechanisms that may take six months or less. Thus the total dimension of the predictor matrix was 554 or (144 SST grid points + 133 SLP grid points) x 2 seasons. Other investigators such as Barnett and Preisendorfer (1987) used four previous seasons to predict any given seasons. We used only two previous seasons to predict any given season primarily to decrease the matrix dimensions of the problem.

Moving windows of 60 and 44-year width were used to predict the long and short East African rains beginning in 1960 to 1989. For example, at the calibration stage the 60-year moving window, 1921-1980 March – May and June to August predictor data was used to predict 1980 SON rainfall. The prediction skill between the predicted and the observed data was computed. Weights were applied to the predictor field and the whole process repeated until the prediction skill was maximized. The Simplex algorithm by Nelder and Mead (1965) (here after referred to as NMS) was used to optimize the predictor's pre-weighting scheme. Finally, at the validation stage, the rainfall seasons for 1990 – 1997 period were predicted using the optimally weighted predictor fields. Several widths of the moving window were tested. In particular, a 44-year moving window generally returned better results than the 60-year moving window.

The NMS algorithm searches for optimal values (weights) in the parameter space of size m (without using any partial derivatives) so as to minimize a scalar objective function of least-square form,

$$\overbrace{\sum_{i=1}^{n}(x_{pred\_i} - x_{obs\_i})^2}^{min} \quad \text{.....(33)}$$

For NMS, the search in the m-dimensional parameter space is executed iteratively through 3 operations; expansions, contractions and reflection, until the objective function is minimized. More details of the NMS algorithm can be found in Nelder and Mead (1965) or more recently in Lagarius et al. (1998).

The predictor field used in our model was a combination of SST and SLP anomalies. These variables have a different units and a different range of magnitudes. This could impact the nature of the extracted principal components that are used in the CCA model along side the predictand PCA components. Furthermore, several ocean zones are poorly sampled relative to others. This is particularly true for the southern portion of both the Atlantic and Indian oceans (Figure 5.2). The zones that are better sampled probably contribute more to the major extracted PCA components than the poorly sampled ocean zones. Shabbar and Barnston (1996) sought to overcome this by applying relative weights to the various predictor fields. They apportioned arbitrarily heavier weights to those predictor fields that they conjectured to be most important to the model in an adhoc manner.

The application of NMS should be more superior and objective than that of Shabbar and Barnston (1996).. As far as we know, this is the first attempt to systematically search for optimal predictor weights in CCA applications.

The zones assigned with larger weights by the Simplex probably contribute more to the variability of the first several principal components used in the CCA model. The entire predictor field was divided into 13 zones (Figure 5.2). The SLP and SST anomalies in each of these zones were weighed with separate coefficients. In our case, the parameter space is of size 24 (m = 24) i.e., 13 for SST and 11 SCP. Two of the zones in the southern portion of the oceans had insufficient SLP data; so they were excluded. The 24 coefficients were iteratively optimized by minimizing the least

square difference between the observed and predicted rainfall fields of East Africa. By minimizing the least square difference, we were indirectly maximizing the forecast skill of the model at calibration stage. The validity of these weights was then tested at the validation stage using data independent of the calibration experience.

5.2 Projection Pursuit Regression (PPR) Model

The statistical approach to modeling a response function can be broadly classified into parametric (e.g. linear and canonical regression) and nonparameteric (e.g. k-nearest neighbor method) methods, as broadly represented by Equation 34,

$$E(Y \mid X = x) = m(x_1, \ldots, x_p) + \varepsilon \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (34)$$

where Y and X are predictand and predictor, and ε is a random term error. PPR is a nonparametric regression method that models the regression surface as a sum of general smoothing functions of linear combinations of the predictor variables in an iterative manner.

The first projection pursuit algorithm was proposed by Friedman and Tukey in 1974 for exploratory data analysis. Later, Friedman and Stuetzle (1981) applied projection pursuit to regression problems, thus giving birth to the PPR technique. Friedman (1985) improved several algorithmic features and extended the approach to multiple response regression is addition to single response regression. It is the multiple response version of PPR that shall be of interest to us.

Although PPR has been around for some time, its application to real world data is still in its early stages. Klinke and Grassman (1998) suggest that the lack of established standard guidelines about the choice of parameters could somehow be discouraging potential users from applying the method. Also interpreting the fit of the model is very difficult.

Klinke and Grassman (1998) discussed the application of PPR in scientific literature ranging from such diverse fields as chemistry to military science. Recently, Chan and Shi (1999) used PPR to develop a prediction scheme for forecasting the summer monsoon rainfall over South China. PPR can be expected to perform better than linear regression in situations where there are substantial nonlinearities in the responses on the predictor variables, especially if the nonlinearities can be approximated reasonable well by a few ridge functions (Friedman, 1985).

5.2.1 Outline of the PPR Model

A brief description of PPR is as follows: suppose there are n observations of p variables X_j (j= 1, 2, ... p), and q variables Y_j (j= 1, 2, ... q) in a *(p + q)* dimensional space that can be arranged in a matrix form as ${}_nX_p$ and ${}_nY_q$. The variables have means $\overline{X}_j$ and $\overline{Y}_j$, respectively and deviations from the mean are given by $x_{ij} = X_{ij} - \overline{X}_j$ and $y_{ij} = Y_{ij} - \overline{Y}_j$, (i= 1, 2, ... n).

The multiple response linear regression model may be written as:-

$$ {}_n y_q = {}_n x_p \omega_q^T + {}_n \varepsilon_q \quad \text{.....(35)} $$

where ω is a matrix consisting of regression coefficients and ε is the random term error with zero mean. Equation (35) is similar to the canonical prediction Equation (19). In general, this linear function is estimated by the conditional expectation of y given particular values of x. The expected value of Y is estimated by the sample mean. Equation (35) may be approximated by

$$ {}_n \hat{y}_q = ({}_n x_p \alpha_q^T)\beta_q \quad \text{.....(36)} $$

where α consists of normalized ω, β is a scaling factor arising from the norming of ω and $\hat{y}$ are "fitted values".

The relation between $\hat{y}$ and the projection $({}_n x_p \alpha_q^T)$ is linear. If q = 1, the relationship would be a straight line. PPR allows this relationship to vary by allowing the predictand variables to be modeled as a linear combination of a sum of univariate ridge functions ϕ_m, (m =1,2, . .M_o) of predictor projections $({}_n x_p \alpha_q^T)$. The ϕ_m functions are smooth, unrestricted parametrically and are constrained to have zero mean and unit variance. Thus, the PPR model with M_o terms is given by

$$ {}_q\hat{y}_n = {}_q\overline{Y}_n + {}_q\left[\sum_{k=1}^{M_o}\hat{\beta}_{kj}\hat{\phi}_k(\hat{\alpha}_k^T x)\right]_n \quad \text{.....(37)} $$

where $\hat{\phi}_k$ are the smooth nonparametric ridge functions.

It can be shown that the PPR model (Equation 37) can model interactions between the predictor variables. For example suppose that

$$ E[y \mid x_1, x_2] = x_1 x_2 \quad \text{.....(38)} $$

This is described for Equation 37 with $\overline{Y}$ =0, M_o=2, $\beta_1=\beta_2={}^1/_4$, α_1^T=(1,1), α_2^T=(1,-1), $\phi_1(t) = t^2$ and $\phi_2(t) = -t^2$.

Then $\phi_1(\alpha_1^T x) = (x_1 + x_2)^2 \equiv x_1^2 + 2x_1x_2 + x_2^2$ .....(39a)

$\phi_2(\alpha_2^T x) = -(x_1 - x_2)^2 \equiv -x_1^2 + 2x_1x_2 - x_2^2$ .....(39b)

so that $\left[\sum_{k=1}^{2}\beta_k\phi_k(\alpha_k^T x)\right] = x_1 x_2$ .....(39c)

The ability to model such interactions makes PPR a versatile modeling tool. It should be noted however that the parametrically unrestricted functions ϕ_m(m =1,2, . .M_o), are univariate. Therefore, PPR can be expected to perform well only in those

situations where the nonlinear relationships can be approximated reasonably well by a few ridge functions that vary in only one direction (Friedman, 1985).

The parameters $\hat{\alpha}$, $\hat{\beta}$ and functions $\hat{\phi}$ can be estimated by minimizing the distance $L_2 = [y - \hat{y}]^2$. Minimizing this distance is equivalent to minimizing the sum of squared residuals (SSR).

Hence a more formal PPR formulation is

$$\min_{\beta_j,\alpha_j,\phi_j;j=1,\cdots,q} L_2(\beta,\alpha,\phi,x,y)$$
$$\text{given} \quad \alpha_j^T\alpha_j = 1 \qquad \text{.....(40)}$$
$$E[\phi_j] = 0$$
$$Var[\phi_j] = 1, \; j = 1,\cdots,q$$

Equation (40) above cannot be minimized simultaneously for all parameters. However if certain parameters are fixed, the optimal values of others are easily solved for, an optimization strategy that Friedman(1985) suggested.

S-plus 4(1997) determines the model parameters by minimizing the sum of squares residues (SSR) as shown below. This method is similar to Freidman's (1985) technique.

$$SSR_q(m) = \sum_{j=1}^{q}\sum_{i=1}^{n}\left[y_{ij} - \overline{Y}_j - \sum_{k=1}^{M_o}\hat{\beta}_{kj}\hat{\phi}_k(\hat{\alpha}_k^T x)\right]^2 \qquad \text{.....(41)}$$

A difficulty sometimes arises in trying to choose the number of terms (M_o) to retain. Standard model selection criteria like the generalized cross validation (GCV) have been used (Klinke and Grassman, 1998). Friedman (1985) remarked that adding more terms than necessary and using a backward stepwise model selection helps to avoid getting stuck in a local maxima. PPR approximations are dense in the sense that any function of p variables can be closely approximated by ridge function expansion of Equation (37) for large enough M_o (Friedman, 1985).

Increasing the number of terms M_o decreases the bias (model specification error) at the expense of increasing the variance of the model and parameter estimates. Since the expected squared error, ESE, is a sum of these two effects, there is an optimal value of M_o. The relative importance of a term is given by

$$I_\ell = \sum_{j=1}^{q} \left|\hat{\beta}_{j\ell}\right|, \quad \ell = 1, \cdots, k+1 \qquad \text{.....(42)}$$

The most important terms are the ones with the largest I_ℓ and the corresponding values of $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\phi}$ are used as initial conditions in minimizing $SSR_q(k)$. A good model selection can be made through the examining the multivariate response fraction of the unexplained variation,

$$e_q^2(k) = \frac{SSR_q(k)}{\sum_{j=1}^{q}\sum_{i=1}^{n}\left[y_{ij} - \overline{Y}_j\right]^2} \qquad \text{.....(43)}$$

for various numbers of minimum terms (k) retained.

Further computation details of the algorithm can be obtained from Friedman and Suetzle (1981), Friedman (1985), Chan and Shi, (1999), Morton (1989) and S-Plus 4(1997).

5.2.2 Application of the PPR model.

The input data to PPR was prepared much the same way as that for CCA. Again in order to decrease the computation work and to filter out unnecessary data noise, we reduced the dimensions of the predictor fields using PCA. This preprocessing step, though necessary, could introduce some undesirable linear constraints in the PPR model. However, all the PPR case studies listed by Klinke and Grassman, (1998) which involved more than 40 variables employed some form of dimension reduction.

5.3 Results and Discussions

5.3.1 Results from the CCA model.

The standard PCA truncation rules were used to decide on the number of PCs retained and subsequently entered into the CCA model. For the case where a moving window of 60 years was used, a total of 7 predictand and 11 predictor PCs were retained (Figure 5.3). The combined SLP – SST predictor field had more than 30 eigenvalues greater than unity. However, the scree plot (Figure 5.3c) shows that the inflexion point occurs at around the sixth eigenvalue. As a compromise, only the first 11 PCs were chosen to represent the combined predictor field.

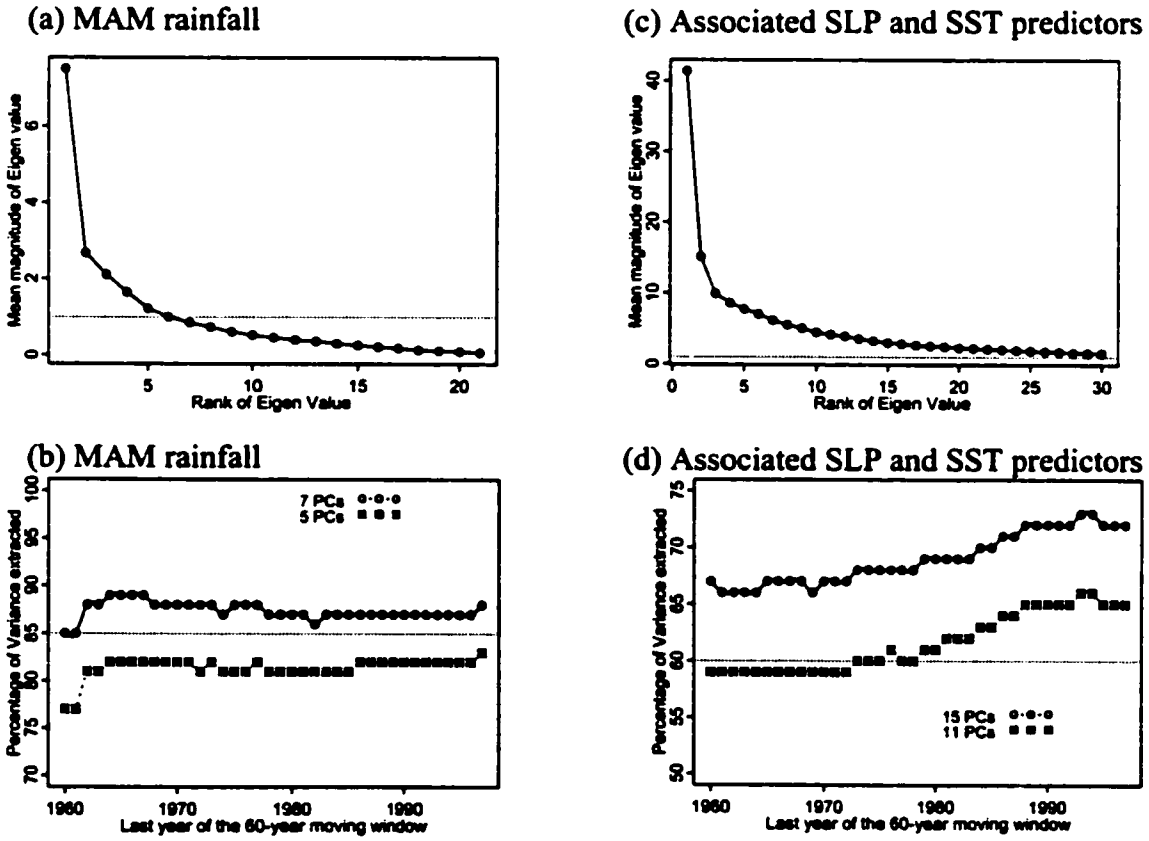


Figure 5.3 (a) shows the eigenvalues arranged in descending order (scree plot) for the MAM rainfall, (b) the temporal change of % variance extracted by 5 and 7 PCs, (c) shows the eigenvalues for the combined predictor field comprised of the previous Sept-Nov (SON), Feb-Dec(DJF), SST and SLP fields, and (d) temporal change of % variance extracted by 11 and 15 PCs for predictor field.

While the variance extracted by the PCs from the predictand field remains relatively steady as the 60-year moving window progressed from 1900-1959 to 1938-1997, that extracted by the predictor PCs monotonically increased during the 80' and apparently leveled off in the 90's. In general, the predictand PCs used in the CCA model explained about 87% of the variance in the original variables while the predictor PCs explained about 63% of the variance in the original, weighted and combined SLP and SST anomaly fields.

For the MAM rainfall, the % of variance explained by the selected PCs remain about constant as the last year of the 60-year window moves from 1960 to 1990, but the % of variance explained by the selected PCs for SLP and SST increases with the moving window. The weights of the SLP - SST anomaly field, obtained by the Nelder – Mead Simplex algorithm were generally close to unity with some exceptions (Figure 5.4).

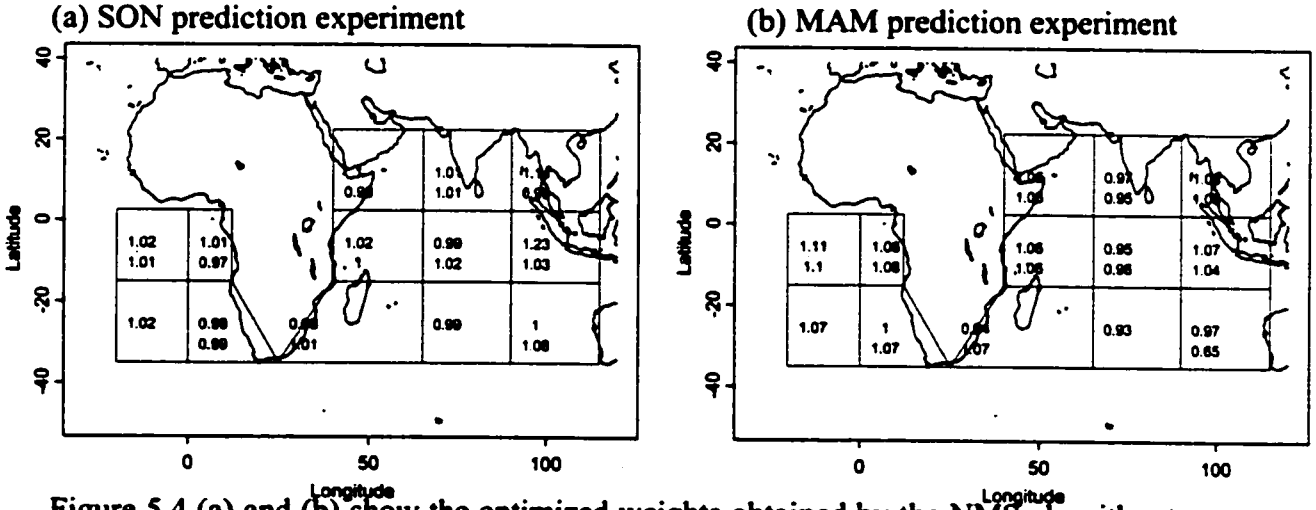


Figure 5.4 (a) and (b) show the optimized weights obtained by the NMS algorithm to the 1900 – 1986 calibration data that comprise of previous 6 months of SST and SLP fields data for SON and MAM prediction experiments respectively. The first number in each sector is the optimized SST weight while the second number is the SLP weight.

The optimized SST weights for the September – November (SON) prediction experiments in the east and northeastern sectors of Indian Ocean (close to Indonesia) were bigger (approx 1.2) when compared to the rest (0.98 – 1.01). It should be noted that ideally, if the predictor dataset is composed of variables of the same units, and is

of a relatively uniform sampling density, then no weights applied to the predictor field can improve the skill of the model. Attempts to find optimized weights for such a case would return weights equal to unity.

For the MAM prediction experiments, the weights that were noticeably different from others were for the SLP around the southeastern Indian Ocean sector, which borders Australia. To this sector, NMS found a relatively small weight (0.65) to reduce its contribution to the overall extracted principal components. On the other hand, a slightly bigger weight (1.11) is observed in the Atlantic ocean sector suggesting that MAM East African rainfall is teleconnected to eastern Atlantic Ocean (Gulf of Guinea) as previously suggested by several researchers such as Trewartha (1981) and Okoola (1999).

The canonical roots (correlation between the predictor and predictand time series) obtained from the model were generally modest and seemed to vary with the width of the moving window used in the prediction experiments. Using the 60-year and 44-year moving windows, the first canonical root averaged 0.7 and 0.9 for the SON prediction experiments respectively (Figure 5.5).

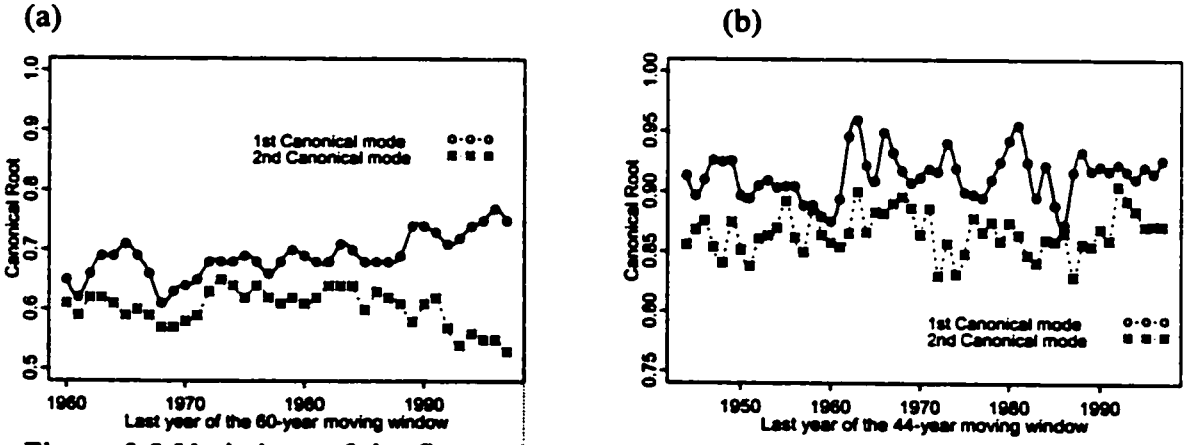


Figure 5.5 Variations of the first and second canonical roots for the SON prediction experiments using a (a) 60-year moving window, and (b) 44-year moving window.

The change in magnitudes of the canonical roots as the width of the moving window is varied shows that the relationship between the predictor and predictand changes

with time. Interpreting the canonical roots should be done cautiously because it is possible to have raw datasets that have nothing in common yet their canonical projections have some form of relationships.

5.3.1.1 CCA Model Prediction Skill.

The prediction skill of the CCA model was tested by assessing the closeness of the predicted values to the observed records at the validation stage where the test data used were independent of the calibration experience. The statistics used were correlation (ρ), root mean square error (RMSE) and the Hanssen-Kuipers skill score (Hanssen and Kuipers, 1965), which is also called the true skill static that is based on categorical forecasts. The correlation (ρ) is given as

$$\rho = \frac{\sum_{k=1}^{n}\left(obs_k - \overline{obs}\right)\left(pred_k - \overline{pred}\right)}{\left[\left(\sum_{k=1}^{n}\left(obs_k - \overline{obs}\right)^2\right)\left(\sum_{k=1}^{n}\left(pred_k - \overline{pred}\right)^2\right)\right]^{\frac{1}{2}}} \quad \text{.....(44)}$$

where obs_k and $pred_k$ are the observed and predicted values, $\overline{obs}$ and $\overline{pred}$ their respective means and n the sample size.

The root mean square error (RMSE) is estimated by

$$RMSE = \left(\frac{1}{n}\sum_{k=1}^{n}\left(obs_k^2 - pred_k^2\right)\right)^{\frac{1}{2}} \quad \text{.....(45)}$$

The Hanssen-Kuipers (HK) skill score, which does require that the data is normally distributed, is a popular skill measure for categorical forecasts. To compute this score, the predicted and observed rainfall values are grouped into categories say "Dry", "Near Normal" and "Wet". Tercile percentages of below 33%, 33% – 66% and above 66% may be used to define the categories which are used to relate the observed and the predicted seasonal rainfall in a square contingency table (Table 5.1).

Then the HK score is defined as

$$HK = \frac{H - E_c}{T - E_m} \quad \text{.....(46)}$$

where H is the total number of correct forecasts, T is the total number of forecasts obtainable with a perfect forecast model, E_c is the number of correct hits expected by chance and E_m is marginal number of correct [observation] hits expected by chance.

Table 5.1 An example of a square contingency table prepared for grid 15 SON prediction experiment at the calibration stage.

		Predicted Categories			
		Dry	Near-normal	Wet	Total
Observed categories	Dry	8	4	2	14
	Near-normal	2	7	5	14
	Wet	4	3	8	15
	Total	14	14	15	43

For an I x I contingency table, the HK score may be expressed in terms of probabilities as follows:-

$$HK = \frac{\sum_{i=1}^{I} p(obs_i, pre_i) - \sum_{i=1}^{I} p(obs_i) * p(pre_i)}{1 - \sum_{j=1}^{J} [p(obs_j)]^2} \quad \text{.....(47)}$$

For example, in Table 5.1 the probability of the correct forecasts $p(obs_i,pre_i)$ is the sum of the diagonal elements of the three way contingency table over the total number of forecasts (23/43). The probability of getting the correct forecasts by chance is the product of the marginal probabilities $p(obs_i)p(pre_i)$ since random forecasting implies statistical independence between forecasts and observed events.

The HK score values range from -1 to +1 with the latter corresponding to perfect scores. The skill expression as given in Equation 47 possesses a number of appealing characteristics. Both random forecasts and constant forecasts receive the same zero score. Furthermore, the contribution made to the HK score by a correct category forecast increases as the event becomes more or less likely to occur. Table 5.2 shows

a summary of the skill measures obtained from the model runs while Figure 5.6 shows the spatial correlation of the predicted vs. the observed seasonal SON and MAM rainfall at the validation stage. Results of 21 of the 31 grids shown in Figure 5.2 are retained for discussion, while the rest that fall of outside the three East African countries (Uganda, Kenya and Tanzania) are discarded.

Table 5.2 A summary of the skill measures obtained from the model runs using a 44-year moving window. The "un-weighted" results refer to runs where no weight obtained from the NMS algorithm was applied to the predictor fields. In all cases a combined SST-SLP predictor field was used.

Zone	Grid No.	Weights Optimized by Simplex						Un-Optimized Weights					
		Calibration (1944-1986)			Validation (1987-1997)			(1944-1986)			(1987-1997)		
		RMSE	ρ.	HK	RMSE	ρ.	HK	RMSE	ρ.	HK	RMSE	ρ.	HK
1	25	1.12	0.07	-0.08	1.27	0.13	-0.24	1.2	-0.10	-0.08	1.37	-0.31	-0.10
	26	1.08	0.03	-0.05	1.20	0.16	0.31	1.2	-0.26	-0.19	1.17	0.17	0.18
2	18	1.30	0.22	0.13	0.82	0.25	0.04	1.4	0.09	0.13	0.84	-0.11	0.31
	21	1.07	0.12	0.09	1.22	0.21	-0.10	1.1	0.00	0.02	1.31	-0.16	-0.24
	22	1.26	0.20	-0.05	0.95	0.20	-0.38	1.3	0.14	0.06	0.99	-0.09	0.31
3	23	1.21	0.20	0.06	1.23	0.34	0.31	1.2	0.12	0.02	1.61	-0.23	-0.10
	24	1.35	0.20	0.16	0.84	0.44	0.31	1.4	0.06	0.09	1.32	-0.23	0.04
	27	1.13	0.21	0.16	1.16	0.24	0.18	1.1	0.23	0.09	1.44	-0.25	-0.38
	28	1.16	0.18	-0.08	0.93	0.42	0.18	1.1	0.23	0.13	1.02	0.21	0.18
4	16	1.17	0.36	0.13	1.29	0.13	0.04	1.3	0.17	0.13	1.31	0.09	0.04
	20	1.28	0.26	0.20	1.14	0.09	0.04	1.4	0.01	0.13	1.24	-0.09	-0.10
5	11	1.26	0.31	0.30	0.89	0.33	0.31	1.3	0.14	0.37	0.97	0.00	-0.10
	12	1.17	0.17	0.30	0.87	0.03	0.45	1.2	0.23	0.16	0.72	0.28	-0.10
	15	1.17	0.33	0.09	1.08	0.28	0.31	1.2	0.17	0.16	1.22	0.03	-0.10
	19	1.22	0.37	0.13	0.91	0.47	0.31	1.3	0.27	0.13	1.09	-0.07	0.04
6	6	1.02	0.47	0.30	0.63	0.45	-0.10	1.2	0.13	0.06	0.8	-0.10	-0.10
	7	1.09	0.39	0.13	0.88	0.07	-0.04	1.2	0.12	0.20	0.71	0.40	0.07
	8	1.06	0.42	0.20	1.28	-0.04	0.18	1.1	0.19	0.20	0.97	0.49	0.04
	9	1.25	0.12	0.06	0.74	0.22	0.18	1.2	0.15	0.16	0.65	-0.11	0.04
	10	1.22	0.21	-0.10	0.73	-0.01	0.04	1.3	0.07	-0.04	0.65	-0.08	-0.24
	14	1.27	0.05	0.02	0.83	0.44	0.04	1.2	0.14	0.16	0.59	0.25	-0.10

In this case, the calibration stage represents the 1944-1986 period where the 24 weights applied to the predictor (SST and SLP) fields were optimized by the NMS algorithm. To test the validity of these weights derived by NMS, the same weights were applied to CCA using the 1987-97 period of data not used in the calibration stage and the results are tabulated under validation in Table 5.2.

A moving window was used so that if necessary, the calibration period could include data that exhibit significant inter-annual and inter-decadal variabilities or data that possess other desired characteristics. Because of this feature, it was not possible to use the common cross validation methods to test the skill of the model. To ensure credibility of the calibrated CCA, its prediction skill is tested with a dataset independent of the calibration experience.

Table 5.2 shows that for the selected number of PCs applied in the CCA model, the use of optimal weights determined by NMS at the calibration stage improve the prediction results even at the validation stage (compare results under validation and un-weighted in Table 5.2). In particular, Zone 5 which comprises of grid 11, 12, 15 and 19 (figure 5.1) shows higher Hanssen-Kuipers (HK) skill scores, correlation, and lower RMSE. This area corresponds to northeastern Tanzania and south-central Kenya. For a model to have useful predictive value, it should have positive HK scores. The root means square error (RMSE) is generally of the magnitude of 1 standard deviation. Zone 3, and 4 also perform generally well with positive HK scores. Zone 6 generally does not perform well at the validation stage, as the correlation values for grids 7,8 and 10 are less than or equal to zero. The statistics in Table 5.2 indicated that a CCA model which does not use predictor fields weighted with optimized coefficients would have no or little prognostic value.

Figure 5.6 shows the correlation maps for the validation stage of both SON and MAM experiments while Figure 5.7 shows predicted vs. observed standardized SON season rainfall anomalies for selected grids.

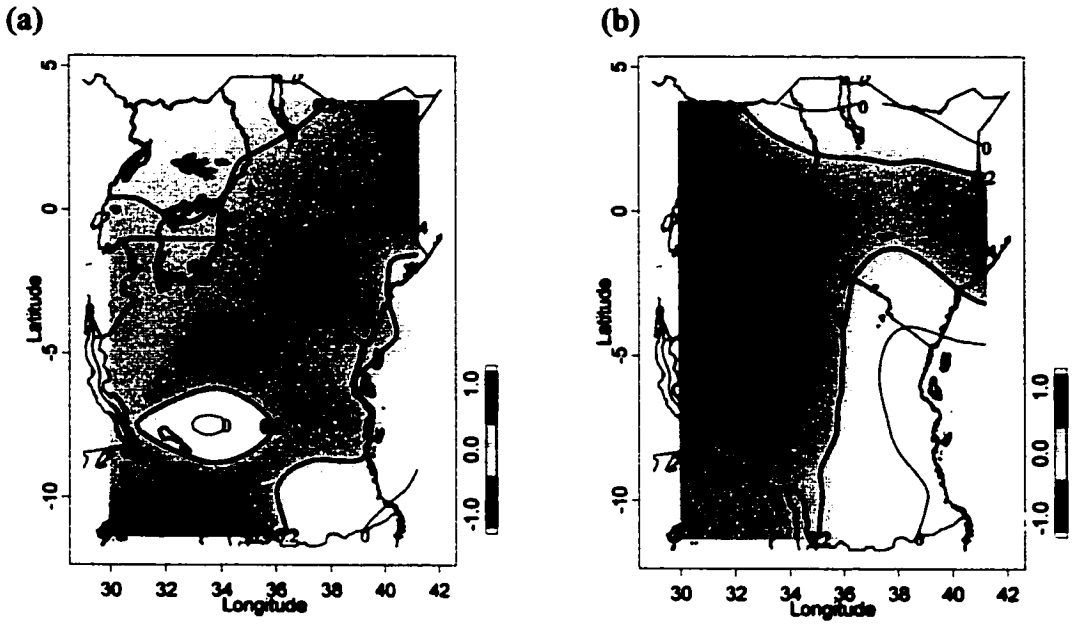


Figure 5.6 Correlation between the 1987-97 (validation) predicted and observed (a) SON and (b) MAM standardized precipitation values using the CCA model.

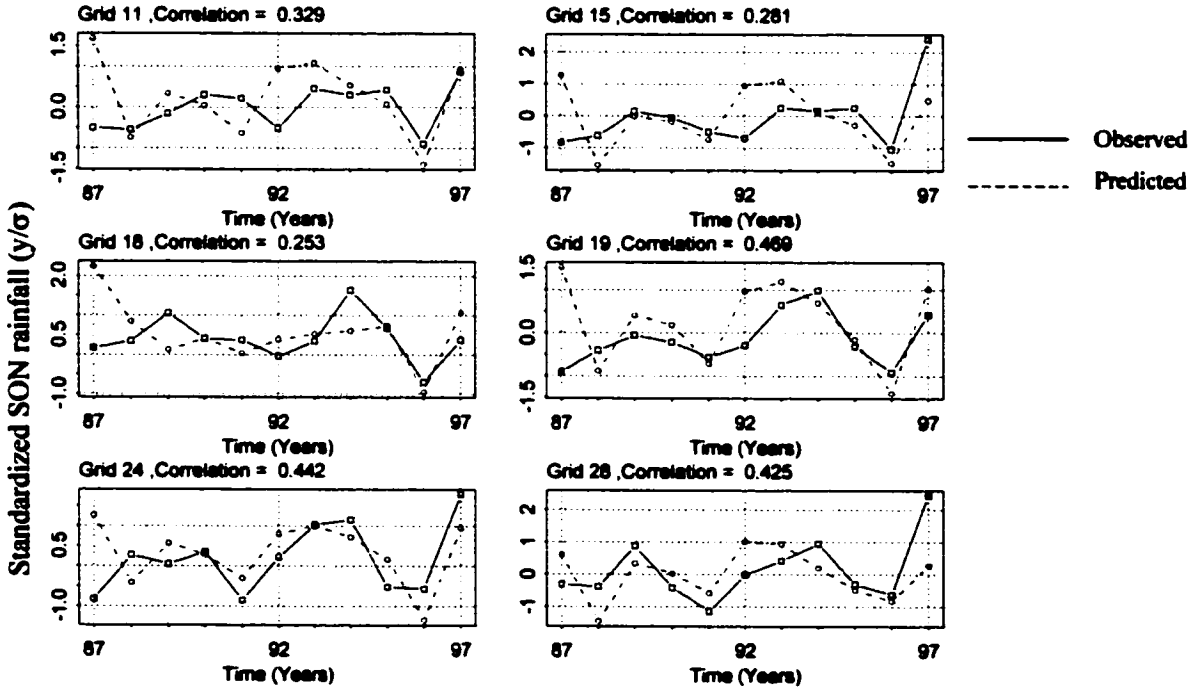


Figure 5.7 An example of the one season-lead SON prediction for selected grids (see Figure 5.1 for the grid location).

The MAM season seems to be better predicted than the SON seasons in the western parts of East Africa and to a lesser extent in eastern Kenya. Elsewhere it is poorly predicted. On the other hand correlation for September-November prediction experiments is more than 0.25 for most parts of East Africa with the exception of Uganda and southern Tanzania. It would seem that the model could neither predict MAM rainfall nor SON for southeastern Tanzania. From Figure 5.7, the 1996 SON seasonal rainfall seems to have been well predicted by the CCA model. However the 1987 SON seasonal rainfall (not shown) was poorly predicted in all the above selected grids.

The practical application of the combined CCA – NMS (Nelder & Mead Simplex) system is demonstrated in Figure 5.8 which compares the observed (Figure 5.8a) versus the corresponding predicted (Figure 5.8b) SON rainfall maps for the year 1988. In addition, this is a validation result that further demonstrates the predictive skill attainable from this CCA-NMS system calibrated by using 1900 – 1960 data.

(a) (b)

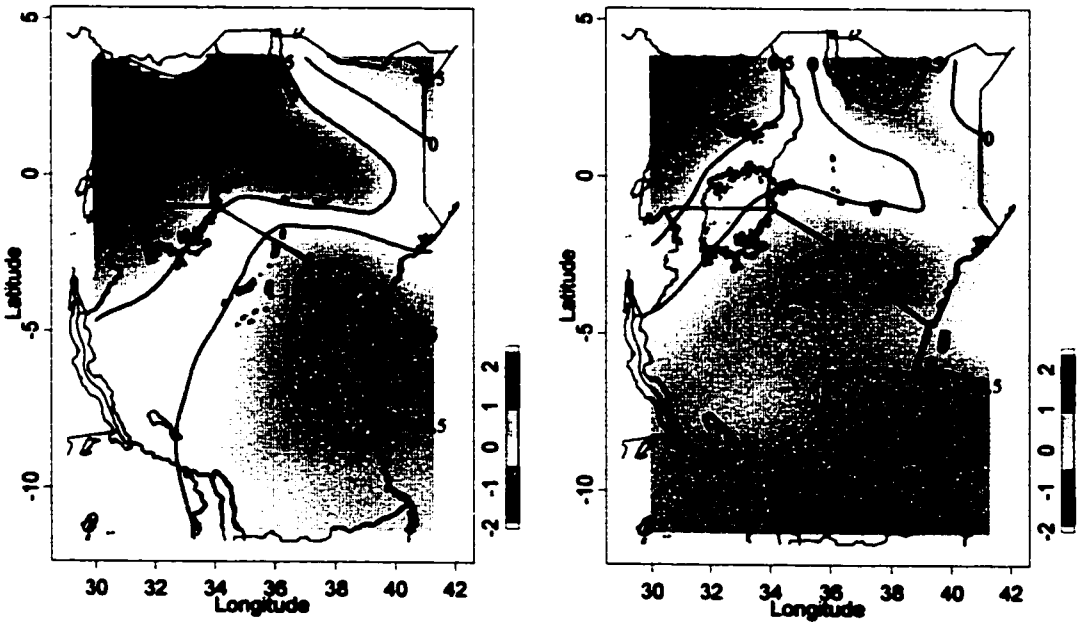


Figure 5.8 Maps showing the (a) observed and (b) predicted SON standardized seasonal rainfall for the year 1988.

Although there are some differences in the respective standardized seasonal values, it is obvious that the general spatial distribution for the two maps is similar. Based on the observed field of the 1988 SON season, Uganda and a portion of western Kenya extending into the central Kenya highlands experienced wetter than normal conditions while most of Tanzania and northern and northwestern Kenya experienced dry conditions. These patterns are also found in the predicted fields.

5.3.1.2 CCA Model Diagnostics

Canonical Correlation maps (Equations 32 and 33) between the canonical time series (precipitation) and the original variables (SST & SLP) are useful CCA diagnostic model outputs that may be used to explain the basic relationship between the predictor and predictand datasets. By studying the magnitude and sign of the loadings in these maps, one may learn how the linear components of the predictor and predictand fields co-vary with each other.

We subjected the ten driest September-November (SON) rain seasons (which is the bottom 10% of the available 1900-1997 dataset), together with the corresponding SST and SLP fields of the previous season (JJA) to a CCA analysis to illustrate their (linear) relationships during dry conditions. Using the above composite dataset was premised on the assumption that low SON rainfall and the associated SLP/SST fields could be considered as a sequence of discrete "episodes" separated by periods within which the variations were of less interest.

It should be pointed out that a dry SON season may not necessarily be part of a drought for several reasons. First, it could be following a MAM or JJA wet spell such that there is still plenty of moisture to make up for the shortfall. Secondly, the rains may sometimes be merely delayed by other synoptic factors to, say, late November or

early December. Nevertheless, modeling dry conditions is a good starting point in understanding drought phenomena.

The canonical correlations for the first three modes were all greater than 0.8, which means all three modes were probably important. However, Figure 5.9 only shows the third-mode canonical correlation maps (commonly referred to as *gmap* for the predictor and *hmap* for the predictand) that were obtained from the CCA analysis of the ten driest SON (Figure 5.8) and MAM (Figure. 5.9) seasons respectively. The predictor *gmaps* for each field and time lag had to be carefully deconstructed from the combined predictor fields.

The third canonical mode *hmap* shows negative correlation coefficients for most of the western parts of East Africa. This is of interest since we are mostly interested in the anomalously low rainfall patterns. During this season, the northeasterly trade winds flow into East Africa generally through two tracks (Findlatter, 1971). One is a dry continental track from Arabia and the other is a humid track over the Indian Ocean and Arabian Sea. Inspection of the third mode, June-August (JJA) SST gmap reveals that there is a narrow coastal band starting from Madagascar reaching the northeastern coast of Africa and extending all the way to west coast of India that is also negatively correlated to the canonical (precipitation) time series. A buildup of cold SSTs during the June – August period in this ocean sector probably reduces the amount of moisture advected by the north easterlies giving rise to anomalously low SON rain seasons.

From the JJA SLP *gmap* (Figure 5.9 c) we can observe that a build up of the SLPs in the south-southwest Indian Ocean sector probably reduces the possibility of low SON totals of parts of eastern Tanzania and southeastern Kenya.

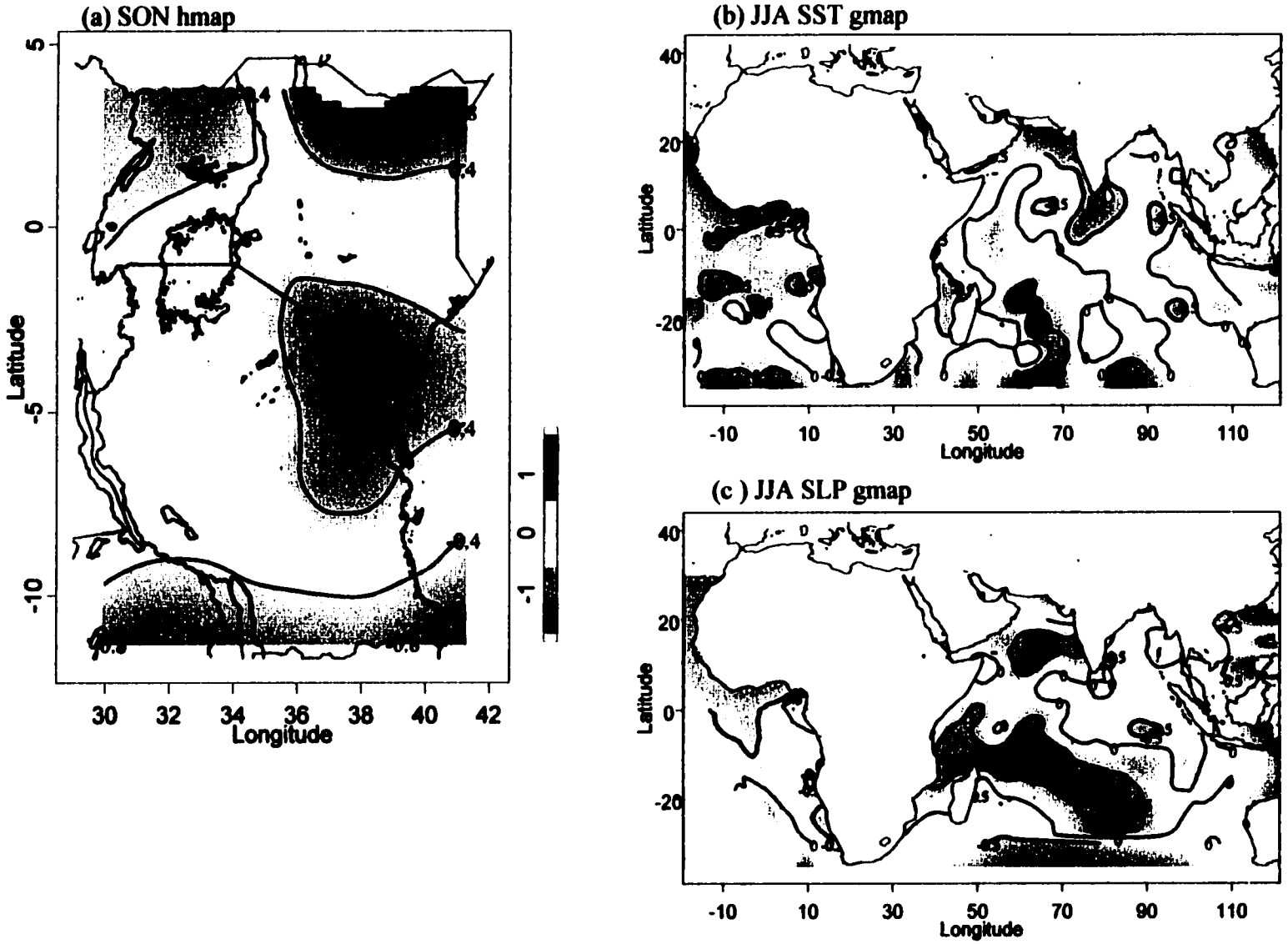


Figure 5.9 Canonical maps of the third mode obtained from analyzing the ten driest September-November (SON) seasons in the 1900-1997 period as predictands with the corresponding JJA SST and SLP fields as the predictors. The SON hmap represents the correlation between the CCA and observed rainfall time series component while the SST (SLP) gmap represents the correlation between the CCA's SST (SLP) component and observed SST (SLP) data.

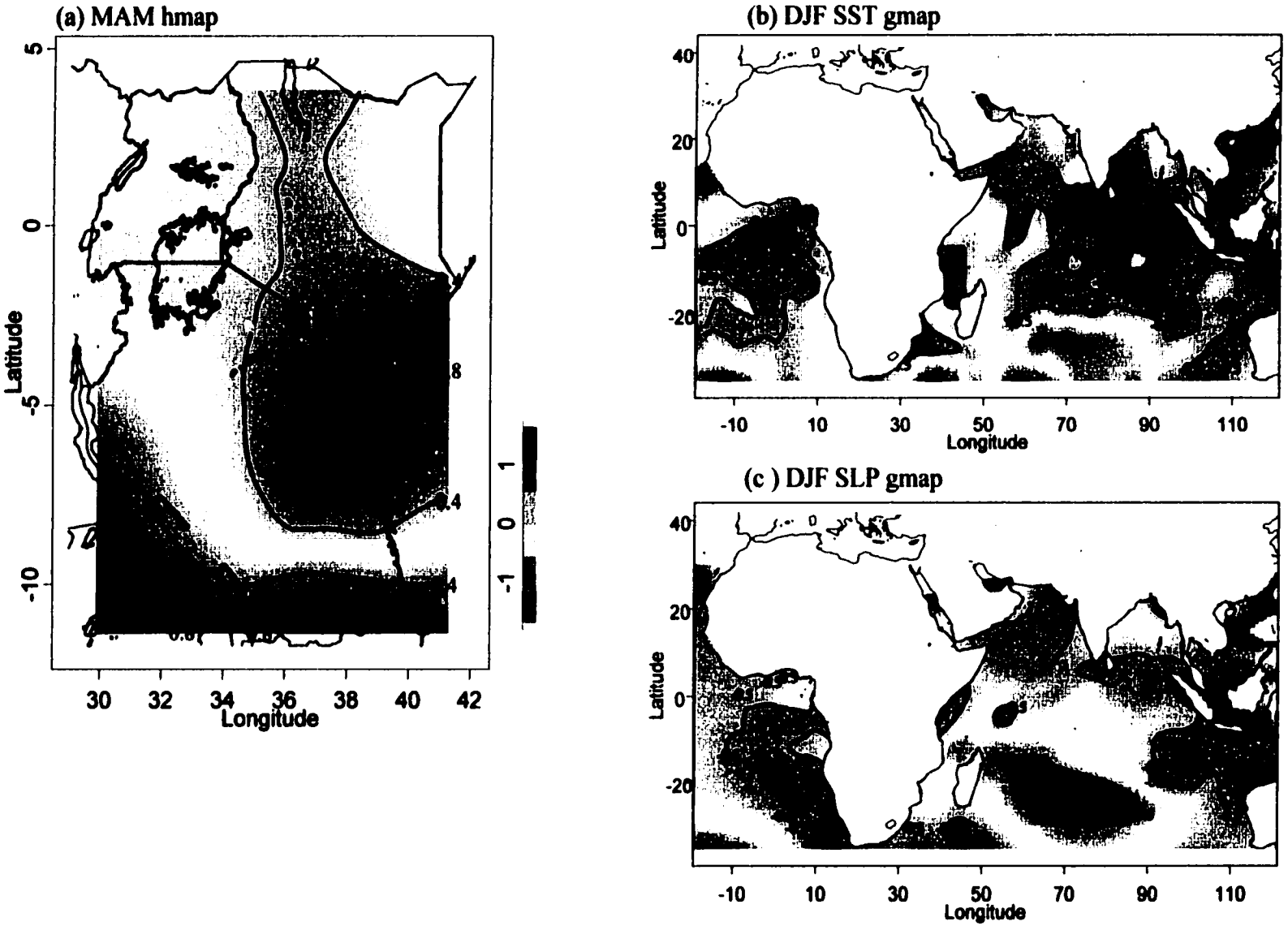


Figure 5.10 Canonical maps of the third mode obtained from analyzing the 10 driest March-May (MAM) Seasons in the 1900-1997 period as predictands with the corresponding DJF SST and SLP fields as the predictors.

- 146 -

Similarly, a comparison of the low-MAM hmap and gmaps (Figure 5.10) shows an association between low MAM seasonal rainfall over many parts of East Africa and a general build up of low SSTs in the adjacent Indian Ocean and the Gulf of Guinea (Atlantic Ocean) during the previous December-February period.

During this season, the ITCZ passes over equatorial East Africa, on its northward swing following the solar cycle. The southeasterly monsoons deposit part of the moisture carried over from the Indian Ocean into the interior lands. It has been suggested that a strong Congo airmass flow into the region increases instabilities of the convergence zone thus increasing the likelihood of precipitation (Okoola, 1999). Low SSTs in the Gulf of Guinea could probably affect the strength and moisture content of the Congo airmass that converges into East Africa. The Congo airmass seems to play an important role in enhancing the long rains in Kenya and Uganda (Okoola, 1999 and Trewartha, 1981). For those years that had weak westerly incursions and subsequent strong flows of the southeasterly flows into the interior of East Africa have been known to be particularly dry (Trewartha, 1981).

Inspection of the *gmap* (Figure 5.10c) indicates an opposite association between a positive SLP pressure zone in southern Indian Ocean and low East African MAM rainfall. A high pressure zone in this ocean sector is an indication of strong southeasterly monsoons which have been shown to be associated with weaker MAM rainfall. This observation is in agreement with findings by Trewartha (1981) and Okoola (1999) who demonstrated the relationship between strong easterly monsoons and dry conditions over East Africa.

We are aware that prediction of precipitation at seasonal time scales is subject to errors because our climatic system is very complex and subjected to the influence and interaction of variables other than SST and SLP considered in our investigation. High frequency instabilities, local conditions and a combination of interacting factors not accounted for in the CCA – NMS system will all impact the accuracy of the statistical model. However, based on our results, the statistical model could on the average predict

30 to 40% of the variability of the seasonal precipitation at 3-months lead time in East Africa.

5.3.2 Results from the PPR model.

A summary of the PPR prediction experiments results is given in Table 5.3. We used a 5 term model (M_o = 5, Equation 37) for we found that increasing M_o further did not improve the results.

In general, the results of the PPR runs without the NMS weights were relatively modest (highest ρ=0.33) but better than those of CCA runs without the application of NMS weights (Table 5.3). PPR performed poorly in zone 6 (southern Tanzania) and modest in zone 1 (Uganda) The calibration results shown in Table 5.3 appear to be close to the results obtained by running PPR with un-weighted predictor fields. On the other hand, validation results are poorer when compared to the calibration results and the case of un-weighted predictors. From this observation alone, it would seem that there is no advantage gained in iteratively optimizing the predictor weights inputs for the PPR model. This could be due to a number of factors. First, by applying PCA to the predictor field, we inevitably introduced some linear constraints to PPR in the model which in turn could have affected the overall accuracy of the PPR algorithm.

Secondly, the smooth nonparametric functions ϕ_m(m =1,2, . .M_o) vary according to the input variables into the PPR model. Thus, whenever different weights are applied to the predictor fields in the PPR model, different ridge univariate functions ϕ_m(m =1,2, . .M_o) are computed by the model. Thus although the NMS scheme optimizes the root mean square error (RMSE), combining it with PPR could produce unpredictable results because each time the predicted values are being computed with changing ridge functions ϕ_m (Equation 37). On the other hand, for the CCA model there is a fixed relationship between the predictand and the predictor variables (Equation 31) which could allow the NMS algorithm to better systematically search for optimized predictor weights.

Table 5.3 A summary of the skill measures obtained from the PPR model runs using a 44-year moving window. The "un-weighted" results refer to runs where no weight obtained from the NMS algorithm was applied to the predictor fields. In all cases a combined SST-SLP predictor field was used.

Zone	Grid No.	Weights Optimized by Simplex						Un-Optimized Weights					
		Calibration (1944-1987)			Validation (1988-1997)			(1944-1987)			(1988-1997)		
		RMSE	ρ.	HK	RMSE	ρ.	HK	RMSE	ρ.	HK	RMSE	ρ.	HK
1	25	1.2	0.11	-0.03	1.6	0.01	-0.21	1.1	0.08	-0.02	1.6	-0.48	-0.52
	26	1.1	0.24	-0.01	1.4	0.15	-0.36	1.1	0.22	-0.06	1.4	-0.28	-0.21
2	18	1.2	0.32	0.15	1.8	-0.09	0.00	1.3	0.24	-0.02	0.7	0.36	0.00
	21	1.2	-0.01	-0.03	1.5	0.06	-0.21	1.2	-0.01	0.05	1.4	-0.12	-0.06
	22	1.1	0.43	0.05	1.7	0.14	0.09	1.3	0.22	-0.06	1.1	-0.10	-0.21
3	23	1.1	0.30	0.11	1.9	0.06	-0.06	1.1	0.27	-0.02	1.6	-0.07	-0.06
	24	1.4	0.20	0.08	1.7	0.06	0.09	1.4	0.03	-0.19	1.1	-0.30	-0.21
	27	1.1	0.35	0.05	1.6	0.14	0.24	1.0	0.49	0.18	1.5	-0.49	-0.06
	28	1.2	0.12	-0.06	1.4	0.14	0.09	1.3	0.03	0.05	1.5	-0.30	-0.21
4	16	1.3	0.27	0.15	2.1	-0.20	-0.21	1.3	0.20	0.05	1.5	-0.31	0.24
	20	1.3	0.28	0.01	2.0	-0.19	-0.06	1.3	0.20	0.08	1.2	-0.32	0.24
5	11	1.4	0.18	-0.02	2.1	-0.26	0.07	1.4	0.13	-0.02	0.9	-0.13	-0.05
	12	1.4	0.04	0.01	1.7	-0.07	0.24	1.3	0.04	-0.04	1.0	-0.24	-0.36
	15	1.2	0.28	0.09	2.0	-0.15	0.09	1.3	0.12	-0.05	1.4	-0.30	-0.06
	19	1.2	0.43	0.15	2.0	0.04	0.09	1.3	0.26	0.01	1.0	0.09	-0.21
6	6	1.5	-0.13	0.08	1.4	-0.20	-0.06	1.3	-0.05	-0.06	1.1	-0.45	-0.06
	7	1.6	-0.02	0.11	1.5	0.02	0.09	1.4	-0.06	-0.04	1.1	-0.29	-0.36
	8	1.5	-0.13	0.08	1.7	0.01	0.09	1.3	-0.11	-0.09	1.3	-0.30	-0.21
	9	1.3	0.16	0.11	1.3	0.43	0.24	1.3	0.08	-0.13	0.9	0.21	0.09
	10	1.6	0.09	0.01	1.1	0.09	-0.21	1.4	-0.01	0.03	1.1	0.01	-0.21
	14	1.1	0.40	-0.02	1.6	0.12	0.09	1.1	0.44	-0.01	0.7	0.50	0.39

Another possible explanation why PPR-NMS did not perform so well is that PPR is suitable for modeling nonlinearities that can be approximated reasonably well by a few ridge functions or functions that vary in only one direction (Friedman, 1985). It could be

that climate nonlinearities are too complex to be approximated by a couple of such ridge functions.

It is worth noting that Chan and Shi (1999)'s attempt to use PPR did not yield satisfactory real – time prediction results of the summer Monsoon rainfall in south China even though they did not reduce their data by principal component analysis (they used only 10 predictor variables).

5.3.3 Origin of Skill in the Statistical Prediction models

Climate variability can generally be attributed to internal and external forcings. External forcings such as sunspots, solar flares, varying solar output, volcanic eruptions and industrial sources have not been shown to play any significant role in short-term climate predictions. This is partly because such external forcings are unpredictable and so the variability generated by them are unpredictable. As a result, the mechanisms that facilitate short-term climate prediction mostly originate from internal forcings (NRC, 1998).

Internally forced atmospheric variability could be classified into three distinct categories: first are those arising from high frequency forcings of the slow components, second are the interactions of the internal variations within the same components that individually would not exert such an effect, and third is the coupling and feedback between fast and slow frequency components of the climate system. The third mode is perhaps the most common mechanism that can be used to explain and predict climatic anomalies such as droughts. When a fast component, e.g. wind transfer of moisture, is coupled to a slow component, e.g. ocean surface temperatures, and the latter exerts a significant control over the regional climate pattern, the longer time scales of the slow component can be capitalized for seasonal or more distant forecast of the fast component which would otherwise be impossible.

Hasselmann (1976) introduced a theory explaining the relevance of different climatic time scales to climate prediction. His theory asserts that through various instabilities, the atmosphere produces high frequency variability that presents itself as weather. When a slower reacting reservoir, such as the ocean, is forced by such a high frequency variability, it basically damps out the high frequency variability over time scales such as seasons. From this perspective, the "Hasselmann mechanism" probably accounts for a great deal of the observed climatic variability as well as modeled variability of long-term climate simulations (NRC, 1998).

Since our combined CCA-NMS system predicts reasonable droughts at 1 season lead-time, it seems that Hasselmann's theory can at least partly account for the cause of East African droughts much as it accounts for the observed global climatic variability elsewhere.

5.3.3.1 Predictability of the Seasonal rainfall.

The predictability of a nonlinear, dissipative system can be estimated from the Lyapunov exponent of the system. For climate variables, they generally range from four days to at most two weeks (Zeng et al., 1992). Moreover, our experience also tells that realistic weather predictions are in the order of several days to at best 2 weeks. It seems natural to question why is it possible to predict droughts several months ahead?

We are aware that such kind of predictions at seasonal time scales are subjected to errors because our climatic system is very complex and subjected to the influence and interactions of many more variables than SST and SLP. On the other hand, if the boundary conditions of the atmosphere, such as SST, can be reasonably predicted at a certain time scale, we will have some information about the statistics of the atmosphere at that time scale (NRC, 1998). The precise state of the atmosphere at any instant cannot be predicted because it is subject to high frequency variability and could vary in equilibrium with the predicted boundary conditions. However, under a given set of boundary

conditions, the "average" condition of the atmosphere can generally be predicted. Therefore, it is possible to predict the seasonal precipitation over a region, even if we cannot say on what specific day the precipitation will fall. Thus, while it is not possible to predict the storms tracks which will develop 90 days in the future, due to predictability limits, it is possible to predict the "average" quantity of the precipitation that will fall in the next three months if a certain set of boundary conditions such SSTs and SLPs are given or reasonably predicted.

The CCA-NMS and PPR models functions on the basis of the above paradigm. The difference is that instead of predicting a boundary condition (e.g. SST or SLP) at a given season and deducing the state of atmosphere (seasonal precipitation) at that season, we develop empirical relationships between current the boundary condition (SST & SLP) and the precipitation field at the next season. If the statistical model can capture the essential co-variability between the boundary condition and an atmospheric state variable (precipitation) at seasonal time scale, reasonable predictions of the "average" conditions of this state variable relative to the boundary conditions can be made.

Sufficient information should be available from past climate records to establish the necessary statistical relationships between the predictor and predictand fields. This statistical relationship could vary depending on the period of calibration data used. A moving window would provide the flexibility of using a data size of the period desired, especially if we wish to use the data that exhibit some low frequency (inter-annual to inter-decadal) variability. As shown in the wavelet analysis plot (Figure 5.11), a 7-year frequency that was statistically significant occurred in Zone 4 East African rainfall between 1965 and 1983.

It will be interesting to explore the possible effect of using calibration data that possess low frequency variability on the model prediction skill of say, CCA or PPR in future work.

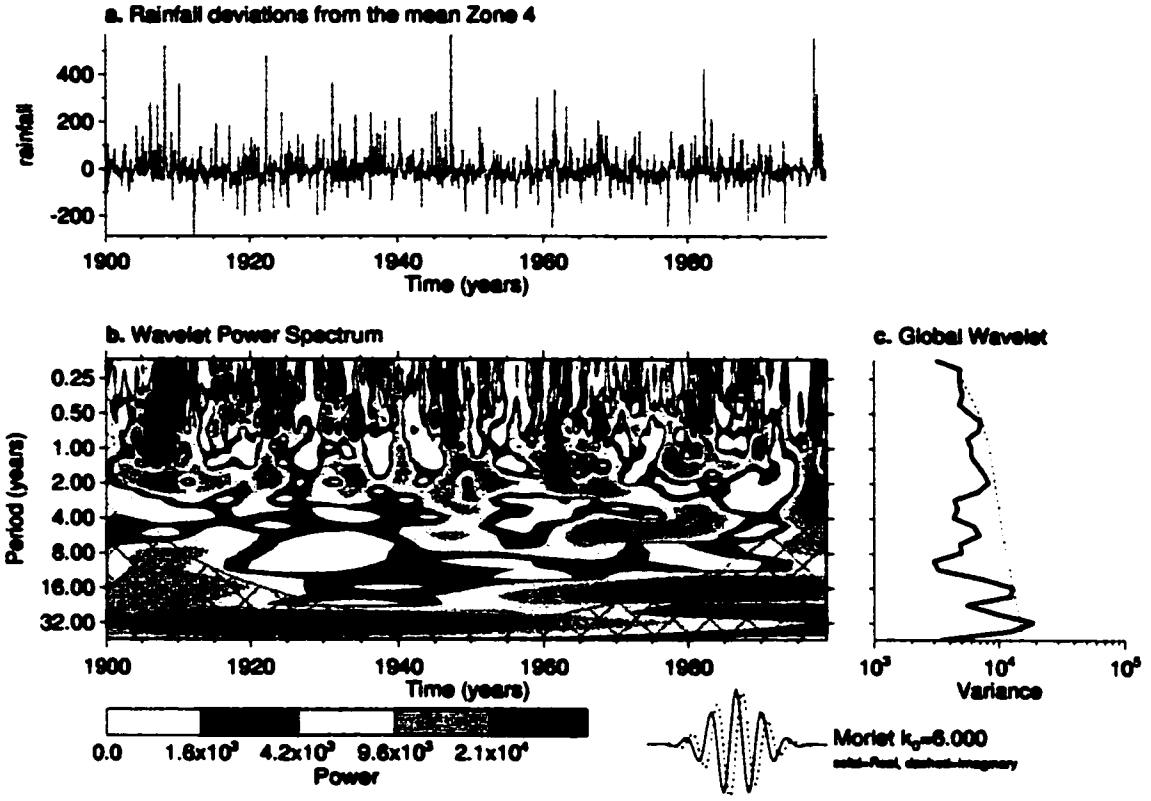


Figure 5.11 An example of a Wavelet power plot, (Compo and Terrance, 1998) for East African Rainfall (Zone 4) showing some inter-annual variability between 1963-1985.

5.4 Summary and Conclusions.

Two statistical models, CCA-Simplex and PPR, were used to predict the standardized seasonal rainfall totals of East Africa at 3-months lead-time using a combination of Sea Level Pressure (SLP) and Sea Surface Temperature (SST) anomaly fields of the Indian and Atlantic Oceans. For the former, the predictor fields were weighted with 24 coefficients and then "reduced" by principle components analysis. The weighting coefficients applied to the input data of CCA were automatically optimized using the Nelder and Mead (1965) Simplex algorithm with respect to minimizing the least - square difference between the observed and simulated rainfall totals (objective function). A moving window was used to flexibly choose the data size and period desired for the calibration experience.

Most of the weighting coefficients were close to unity except for a relatively large weight for SSTs in the east and northeastern sector of the Indian Ocean (SON prediction experiments) and relatively low weights for the southeastern Indian Ocean (MAM prediction experiments).

Our analysis showed that applying the optimized weights to the predictor fields produced better MAM and SON seasonal rain forecasts than a direct application of the same predictor fields to CCA. Similar experiments with CCA using only un-weighted predictor fields faired poorly, as indicated by high RMSE, negative correlations and Hanssen-Kuipers skill scores that were negative or close to zero. As far as we know, this is the first attempt where such weights are assigned to the predictor fields in a systematic manner. Previously, such weights were assigned in an adhoc manner (e.g. He and Barnston 1996; Shabbar and Barnston, 1996).

Among the 6 rainfall zones delineated in East Africa, predictor and predictand relationships obtained through canonical roots varied temporally and spatially, and also with the calibration data size and period. Among the 6 zones, northeastern Tanzania and south-central Kenya (Zone 5) had the best SON prediction experiment results with both validation correlation coefficients and Hanssen-Kuipers skill scores exceeding +0.3. The MAM season was better predicted in the western parts of East Africa and to a lesser extent in eastern Kenya. Elsewhere the MAM season was predicted poorly.

By applying CCA to a composite set of the ten driest SON and MAM seasons in the 1900-1997 period respectively, we demonstrated the possible association between dry conditions in East Africa and SST/SLP fields of the Indian and Atlantic Oceans. The CCA correlation maps showed that low SON rainfall in East Africa is associated with cold SSTs off the Somali Coast and the Benguela coast which are upwelling regions. A buildup of the low July – August (JJA) SSTs in this part of the Indian Ocean probably reduces the amount of moisture advected by the northeasterlies, resulting in anomalously low SON rain totals.

Similar analysis showed that low MAM rainfall is associated with a buildup of low SSTs in areas of the Indian Ocean adjacent to East Africa as well as the Gulf of Guinea off the coast of West Africa. Strong southeasterly monsoons have been previously associated with weak MAM rainfall in East Africa (Trewartha, 1981). Further, low MAM rainfall could also be attributed to a high SLP zone in southern Indian Ocean where a build up of high pressure anomaly means that the south easterlies are likely to be stronger than normal, thereby penetrating deep into East Africa and in the process preventing the humid Congo westerly from depositing its moisture in East Africa.

We conjecture that most of the skill of the CCA model might be explained by the theory advocated by Hasselman (1976). The relatively high frequency atmospheric instabilities (including seasonal rainfall anomalies) are damped by the relatively lower frequencies of the slower reacting reservoir (Oceans). The statistical models probably capitalize on the mechanistic control the slow components have over the fast components, to make distant forecasts of the latter.

Even though the PPR can model nonlinear associations, it performed poorer than the linear CCA-NMS model. First reducing the predictor dimensions by the linear principle component analysis could have partly attributed to the poor performance of the PPR. Alternatively, it could be that the climate non-linearities are too complex to be well approximated by a few ridge functions that vary only in one direction in PPR.

In the CCA-Simplex system, we assumed a linear relationship between the predictor and predictand fields. We confined the predictor fields to SST and SLP of Indian and southwest Atlantic Oceans. However fields from the Pacific have been shown to teleconnect with East African rainfall especially during ENSO years. Further, the SSTs of Tropical Pacific correlate strongly with the SSTs of Indian Ocean with a 3-month lead (Nyenzi, 1988). It may be worthwhile including data of the Pacific Ocean as part of predictor fields in future work.

Chapter 6 Summary and Conclusions

The theme of this investigation was about the statistical properties of East African meteorological droughts, drought indices, and two statistical teleconnection models developed to predict seasonal drought in the region.

In Chapter 2, we reviewed East African rainfall climatology, its historical droughts, and used harmonic analysis to delineate East Africa into 6 homogeneous rainfall zones, even though more zones were delineated in the past (e.g. Ogallo, 1989; Basalirwa, 1995). Among these zones, it seems that central and southern Tanzania have the most stable rainfall structure. We also identified the important rain seasons (in terms of % contribution to the annual rainfall) that vary from zone to zone. It seems water stress or drought would occur under the failure of two consecutive important rain seasons in areas with bimodal rainfall characteristics, or even one critical rain season in areas with unimodal rainfall characteristics, since East Africa generally depends on rain-fed agriculture heavily.

In Chapter 3, we analyzed, modified (when necessary), and compared the performance of 3 drought indices (Palmer Drought Severity Index or PDSI, Bhalme-Mooley Index or BMI, and Standardized Precipitation Index or SPI) in detecting the initiation, evolution, termination and severity of meteorological drought events in East Africa.

After finding the original PDSI (designed for the US) performed poorly in some parts of East Africa, we modified PDSI's approach to compute the potential runoff, the Z index and its recursive formula, and obtained more realistic results for most of East Africa.

The Bhalme-Mooley index, whose input is only rainfall data, was found to produce results that are highly correlated to those of the modified PDSI. This suggests that precipitation data alone accounts for a large part of the variability of East African

droughts. Therefore, for East Africa, a simpler drought index like BMI could be as informative as the relatively complex PDSI. However, the third drought index SPI is even more versatile and suitable for East Africa than the above two for several reasons.

SPI is easy to compute, yet flexible enough to track droughts at time scales according to a user's interest, and could extract more or less the same information as the relatively complex PDSI that doesn't possess such a flexibility in time scale. We also introduced two modifications to improve SPI. First, instead of fitting a gamma distribution to the "smoothed" precipitation data, we used an unbiased, semi-parametric, plotting position formula to reduce the effect of outliers on the results. Second, we proposed transforming the final SPI values using a regional Pearson Type III (P3) distribution instead of using a gaussian normal distribution that for skewed data (such as most precipitation data) may introduce considerable distortion in the tails. The regional flood-index method was used to obtain the regional parameters of P3. The results we obtained are different from the original SPI.

Using a rotated principal component analysis (RPCA) on 6-month and 12-month SPI, we delineated East Africa into 7 drought homogenous zones. The spatial distribution of these zones somewhat resemble the precipitation zones delineated by harmonic analysis in Chapter 2. From the inter-correlation between the homogenous drought zones, it seems that some zones are likely to experience a given drought spell simultaneously, while others could experience a wet spell. For example, most of Uganda seemed to experience a wet (dry) spell whenever southwestern Tanzania underwent dry (wet) conditions. Such information on inter-zone correlation could be used to transfer resources between the zones as part of the drought mitigation strategies.

In Chapter 4, using composites of 22 El Niño and 13 La Niña events based on rainfall anomalies and 6-month SPI, we used harmonic analysis to investigate the possible relationship between ENSO and East African moisture regime. Out of 5 identified regions that show unique responses to ENSO, northeastern Tanzania has the highest ENSO response coherence (>60%) and also the highest % of variance extracted by the

first harmonic (>90%). Conversely, the central highlands of Kenya and the southeastern highlands of Tanzania have no discernable response to ENSO probably because their rainfall mechanisms are highly localized.

Furthermore, ENSO's influence is also season dependent. For example, southern Tanzania has the maximum (minimum) response to La Niña (El Niño) between January and June of the year after the ENSO event. Southern Uganda and Lake Victoria basin shows significantly positive ENSO response for November to January. Analysis of 6-month SPI, based Index Time Series produced results that were similar to those based on ENSO composites thus confirming the finding that various parts of East Africa have different temporal patterns of ENSO response.

The regional and seasonal dependency of ENSO response are further demonstrated via boxplots comparing the probability distribution of 6-month SPI for non-ENSO and ENSO "seasons". The influence of ENSO resulted in obvious shifts to the probability distributions of ENSO seasons as against that of non-ENSO seasons. Among the ENSO seasons, El Niño exerts a stronger influence than La Niña on East African rainfall, winter response was generally positive for El Niño (except for southern Tanzania) and negative for La Niña. The spring response to ENSO is generally less clear cut, except for southern Tanzania and Lake Victoria basin with negative (positive) response to La Niña (El Niño).

Our results on the ENSO response in East Africa did not concur with that of Nicholson and Kim (1997) who suggested that rainfall in equatorial Africa is enhanced in the short rains season (northern hemisphere autumn rains) during ENSO years, but reduced in the following spring rains. We believe the difference in results are at least partly attributed to Nicholson and Kim defining East Africa as only one region, while we subdivided East Africa into 6 homogeneous rainfall regions. Given that our zoning of East Africa is of higher resolution and likely more representative of the spatial variability of its rainfall mechanism, our results should be more representative than that of Nicholson and Kim (1997). Also, at sub-regional resolution, we believe our results provide more details than

the simple dipole structure of Ropeleweski and Halpert (1987) on East Africa ENSO teleconnection.

In Chapter 5, we developed a combined canonical correlation analysis (CCA) – Simplex modeling system to predict the standardized, seasonal rainfall totals at 3-month lead time using a combination of sea level pressure (SLP) and sea surface temperature (SST) anomalies as predictor fields. The predictor fields were adjusted with weights optimally determined by the direct search algorithm, Simplex of Nelder and Mead (1965) via the calibration of CCA based on a least square objective function of the predictand field. Before applying CCA, the dimension space of weighted predictor and predictand fields was reduced by unrotated principal component analysis (PCA). The weights were generally close to unity with some exceptions. However, our results show that applying these optimized weights to the predictor fields produced consistently better MAM and SON seasonal rain forecasts than un-weighted predictor fields. For the later, the Hansen-Kuipers skill scores (HK) were either negative or close to zero.

To compare with CCA, a linear statistical model we also applied the same predictor fields (but not adjusted with weights from Simplex) to a nonlinear statistical model called projection pursuit regression (PPR). The prediction skill of PPR assessed by correlation coefficients and HK scores turned out to be poorer than the combined CCA-Simplex system. This partly demonstrates the importance of adjusting the predictor fields with such optimal weights.

As far as we know, this is the first attempt where such weights assigned to the predictor fields are optimized in a systematic manner. Previously, application of the field weights had been done in an adhoc manner (e.g. He and Barnston 1996; Shabbar and Barnston, 1996). We have demonstrated that using the Simplex optimized weights improves the traditional CCA model.

Based the CCA- Simplex model output, northeastern Tanzania and south central Kenya had the best SON prediction results of both correlation coefficients and HK > +0.3. The

MAM prediction was better in the western parts of East Africa and to a lesser extent in eastern Kenya.

By applying CCA to a composite set of the driest 10 percentile SON and MAM seasons for 1900-1997 respectively, we were able to diagnose possible mechanisms between dry conditions in East Africa and SST/SLP fields in the Indian/Atlantic Oceans. The CCA correlation maps showed that low SON rainfall in East Africa is associated with a build up of cold SSTs during July – August (JJA) in the Indian Ocean off the Somali Coast and the Benguela coast which probably reduces the amount of moisture advected by the northeasterlies to East Africa.

Similar analysis showed that low MAM rainfall is associated with a buildup of low SSTs in areas of the Indian Ocean adjacent to East Africa as well as the Gulf of Guinea off the coast of West Africa. Strong southeasterly monsoons have been previously associated with weak MAM rainfall in East Africa (Trewartha, 1981). Our analysis showed that low MAM rainfall is probably also associated with high SLP in southern Indian Ocean which means that the south easterlies are likely to be stronger than normal, thereby penetrating deep into East Africa and in the process preventing the humid Congo westerly from depositing its moisture in East Africa.

We believe that most of the predictive skill of CCA and PPR can be explained by the theory advocated by Hasselman (1976). According to this theory, the relatively high frequency atmospheric instabilities (including seasonal rainfall anomalies) are damped by the relatively low frequencies of the slow reacting reservoir (Oceans). By capitalizing on the mechanistic control of the slow frequency components over the fast frequency components, it is possible for statistical models such as CCA to make forecast of the latter up to several months of lead time.

Further, the combined CCA-Simplex model was premised on the assumption that there exists some linear relationships even though most of the atmospheric processes are non-

linear. Such simplification can at times achieve reasonable prediction skill of 0.3 to 0.5 correlation with the observed.

Wavelet analysis of a sample of East African monthly rainfall series showed that the structure of rainfall series in terms of the underlying frequencies generally varies with time thus suggesting the existence of inter-annual to inter-decadal variability. Therefore, we selected a moving window of arbitrary width in the prediction experiments to provide the flexibility to choose the data size and period as desired.

Lastly, our contributions to East African drought research include identifying the causative factors to triggering the droughts, spatial and temporal patterns of droughts, the development of drought indices, and statistical teleconnection models to predict droughts. These contributions should have practical and economical values to East Africa.

6.1 Suggested Future Work

There is still much more to be studied and understood on the complex, multi-disciplinary Drought phenomenon. Below are some recommendations for further research related to this work:

1. Among the three indices analyzed, we found SPI to be better suited for tracking the initiation, evolution and cessation of meteorological droughts in East Africa. There is need to explore further SPI time scales suitable for tracking East African droughts, which could vary from one rainfall zone to the next. SPI should also be modified to make it suitable for tracking hydrological, agricultural and social economical droughts. In this case, beyond precipitation, we anticipate requiring a few more hydrologic variables such as evaporation, streamflow, ground water table etc to track these few types of droughts. It might be possible to also include water use/demand variables to enhance the overall objectivity of drought control/mitigation decisions based on drought indices.

2. It is unclear whether anthropogenic causes such as greenhouse gas emissions could exacerbate the frequency and severity of droughts in say East Africa. Landuse changes and settlement patterns are some such anthropogenic factors that should be investigated in relation to droughts in East Africa.

3. Since ENSO response is both season and region dependent, there is a need to further investigate in detail other climatic and possibly topographic factors that together contribute to the spatial and temporal variability of ENSO response in East Africa. In particular, we need to find out why long rains (MAM) ENSO response is not as clear cut as the short rains (SON) in East Africa.

4. For the combined CCA-Simplex and PPR models, we need to explore the potential contributions of predictors such as SST/SLP data of the Pacific Ocean, besides those of Indian and Atlantic Ocean already used in this investigation.

References

Alley, M. W., 1984. The Palmer Drought Severity Index; Limitations and assumptions. *J. Clim. Appl. Meteo.*, Vol. 23, p. 1100 - 1109.

Baker, S.J.K., 1977. 'A background to the study of Drought in East Africa' in Drought in Africa II edited by Dolby, D., Church, R.J.H and Bezzaz, F., *African Environment Special Report*, No. 6, p. 74 - 81.

Barnett, T. P. and Preisendorfer, R., 1987. Origins and Levels of Monthly and seasonal Forecast Skill for United States Surface Air Temperatures Determined by Canonical Correlation Analysis. *Mon. Wea. Rev.*, Vol. 115, p. 1825 – 1851.

Barnston, A. G. 1994. Linear statistical short term climate predictive skill in the Northern hemisphere. *J. Climate*, Vol. 5, p. 1514 -- 1564.

Basalirwa, C. P. K., 1995. Delineation of Uganda into Climatological Rainfall zones using the method of Principal component analysis, *Int. J. Climatol.*, Vol. 15, p. 1161 - 1177

Basalirwa, C. P. K., Odiyo, J.O, Mngodo, R. J., and Mpeta E.J., 1999. The Climatological regions of Tanzania based on rainfall characteristics. *Int. J. Climatol.*, Vol. 19, p. 69 – 80.

Basnett, T.A. and Parker, D.E., 1997. Development of the Global Mean Sea Level Pressure Data Set GMSLP2, Hadley Centre Climate Research Technical Note CRTN 79

Beltrando, G., 1990. Space time variability of rainfall in April and October- November over East Africa during the period 1932 - 1983, *Int. J. Climatol.*, Vol. 10, p. 691 - 702.

Bhalme, H. N. and Mooley, D. A. 1980. Large Scale Droughts/Floods and Monsoon Circulation. *J. Clim. Appl. Meteo.*, p. 1197 - 1207

Bhalme, H. N. and Mooley, D. A., 1979. 'On the performance of Modified Palmer Index.', *Proc. International Symposium*; *Hydrological Aspects of Droughts*, Vol. 1, New Delhi, India, p. 373 - 383.

Briffa, K. R., Jones, P. D., and Hulme, M. 1994. Summer moisture variability across Europe, 1892 - 1991: An analysis based on the Palmer Drought Severity Index. *Int. J. of Climat.*, Vol. 14, No. 5, p. 475 -506

Burke Jr., T. T., 1996. *Characterization and predictability of Droughts in the Midwest.* PhD Thesis, Purdue University, USA. 210p.

Byun, R. H., and Wilhite, D. A, 1999. Objective Quantification of Drought Severity and duration. *J. of Climate,* Vol. 12, p. 2747-2755.

Cancelliere, A., Rossi, G. and Ancarani, A., 1996. Use of Palmer Index as a Drought indicator in Mediterranean regions. *Proc. 'From Flood to Drought' 5 - 7 August, 1996, IAHR, African Division*, Suncity, South Africa.

Chan, C. L. J., and Shi, J.-E., 1999. Prediction of the Summer Monsoon Rainfall Over South China, *Int. J. Climatol.*, Vol. 19, p. 1255-1265.

Chow, V. T., Maidment, R. D. and Mays, W. L., 1988. *Applied Hydrology*. McGraw-Hill Series in water resources engineering. 572 p.

Chu, P.-S. and He, Y., 1994. Long range prediction of Hawaiian winter rainfall using canonical correlation analysis. *Int. J. Climatol.*, Vol. 14, p. 659-669.

Chu, P.-S., 1998. Short term climate prediction of Mei-Yu rainfall for Taiwan Using Canonical Correlation Analysis. *Int. J. Climatol.*, Vol. 18, p. 215 – 224.

Climate Monitor, 1984. Country by country listing of exceptional climatic events. Ed. D. A. McKinlay, Climatic Research Unit. University of Anglia, UK. Vol. 13, No. 3.

Climate Monitor, 1994. Ed. D. A. McKinlay, *Climatic Research Unit.* University of Anglia, UK. Vol. 20, No. 1-5.

Climate Monitor, 1995. Ed. D. A. McKinlay, *Climatic Research Unit.* University of Anglia, UK. Vol. 21, No. 1-5.

Climate Monitor, 1997. Country by country listing of exceptional climatic events. Ed. D. A. McKinlay, Climatic Research Unit. University of Anglia, UK Vol. 26, No. 1.

CLIVAR, 1999. *Climate research for Africa*, International CLIVAR Project Office, WCRP Informal Report No. 16/1999, ICPO Publication Series No.29.

Cunnane, C., 1978. Unibiased plotting positions – A review., *J. Hydrol.*, Vol. 37, no.3/4, p. 205-222.

Dalén, J. ,1987. Algebraic bounds on standardized sample moments. *Statistics and Probability Letters*, Vol. 5, p. 329 – 331.

Dracup, J.A., and Kahya, E., 1994. The relationships between U.S. streamflow and La Niña events, *Water Resources Research*, 30(7), p. 2133-2141.

East African Meteorological Department, 1963. *Climatic seasons of East Africa*, East African Meteorological Department, Report No. 8, 4 p.

Edwards, C. D. and McKee, B. T., 1997. *Characteristics of 20^{h} century drought in the United States at multiple time scales.* Atmospheric Science Paper No. 634, Climatology Report No. 97-2, Dept. of Atmospheric Sciences, Colorado State University.

Efron, B. and Tibishirani, R. J., 1993. *An Introduction to the Bootstrap.* Chapman & Hal, San Francisco.

Efron, B., 1982. The Jackknife, the Bootstrap and other re-sampling plans, CBMS-NSF Regional Conference Series in Applied mathematics, *Soc. Ind. and Appl. Maths,* Vol. 38, Philadelphia, PA, 104p.

Findlater, J., 1971. Mean monthly airflow at low levels over the Western Indian Ocean. *Geophysical Memoirs,* No. 115, 16, 1-38.

Friedman, H. J., 1985. Classification and multiple regression through projection pursuit., Technical Report LCS012, Department of Statistics, Stanford University.

Friedman, H. J., and Stuetzle, W. 1981. Projection Pursuit Regression. *J. American Statistical Association.* Vol. 76, No. 376, p.817 – 823

Friedman, H. J., and Turkey, J. W., 1974. A projection pursuit algorithm for exploratory data analysis. *IEEE Transactions on Computers*, Vol. C – 23, p. 881 –889.

Gibbs, W. J., 1975. Drought - its Definitions, delineations and Effects, *Special Environment Report*: *Drought*, Vol. No. 5, WMO, Geneva, p.

Glahn, H. R., 1968. Canonical correlation and its relationship to discriminant analysis and multiple regression. *J. Atmos. Sci.*, Vol. 25, p. 23 - 31

Glantz, H. M, 1987. 'Drought and Economic Development in Sub-Saharan Africa.', in *Planning for Drought: Towards a Reduction of Society Vulnerability*, edited by Wilhite, D. A., Easterling, E. A. and Wood, D. A., Westview Press Inc., Colorado, p. 297 - 316

Glantz, M.H., and Katz. R.W., 1977. When is a drought a drought? Nature. Vol. 267, p. 192-193.

Gommes, R. and Petrassi, F., 1994. Rainfall variability and drought in Sub-Saharan Africa since 1960. Agrometeorology Series, Working paper No. 9, Food and Agricultural Organization, Rome, Italy.

Griffiths, J. F., 1972. Eastern Africa, Climates of Africa. *World Survey of Climatology*, Vol. 10, Elsevier, Amsterdam, 604 p.

Grove, A. T., 1974. Desertification in the African Environment, *Africa Review*, Vol. 73, No. 291, p. 137 – 151.

Guttman, B. J, 1998. Comparing the Palmer Drought Index and the Standardized Precipitation Index. *J. American Water Resources Association*. Vol. 34, No. 1, p. 113 –121.

Halpert, M. S., and Ropeleweski, C. F., 1992. Surface temperature patterns associated with the Southern Oscillation, *J. Climate*, 5, p. 577-593.

Hanssen, A.W., and Kuipers, W.J.A. 1965: On the relationship between the frequency of rain and various meteorological parameters. De Bilt, Koninklijk, Nederlands, *Meteorologist Institua*, Mededelingen en Verhandelingen, Vol. 81, p. 2-15.

HAPEX 1997. HAPEX-Sahel - De la naissance à l'âge adulte, from birth to adulthood. *J. of Hydr.*, Special Issue, Volume 188/189 No. 1 - 4.

Hasselmann, K., 1976. Stochastic Climate models: I. Theory. *Tellus*, Vol. 28, p. 459 - 479

Hasselmann, K., 1976. Stochastic Climate models: I. Theory. *Tellus*, Vol. 28, p. 473 - 485

He, Y. and Barnston, A. G., 1996. Long lead forecasts of seasonal precipitation in the tropical Pacific Islands using CCA. *J. Climate*, Vol. 9 p. 2020 – 2035.

Heddinghaus, R. T. and Sabol, P., 1991. A review of the Palmer's Drought Severity Index and Where do we go from here? Preprints, *6th Conference on Applied Climatology*, 17th -22nd January, Anaheim, CA, p. 242-246.

Hosking, J. R. M., 2000. *Fortran routines for use with the method of L-moments: Version 3.03*. Research Report, IBM Research Division T. J., Watson Research Center, NY, 30 p.

Hosking, J. R. M., and Wallis, J. R., 1997. *Regional Frequency Analysis: An approach based on L-moments.* Cambridge University Press, 224 p.

Hotelling, H., 1936. Relations between two sets of variates. *Biometrika*, Vol. 28, p. 139 – 142.

Hulme, M. and New, M., 1997. The dependence of large-scale precipitation climatologies on temporal and spatial gauge sampling, *J.Climate,* Vol. 10, p. 1099- 1113.

Hulme, M., 1994 Validation of large-scale precipitation fields in General Circulation Models p.387-406 in, *Global precipitations and climate change* (eds.) Desbois, M. and Desalmand,F., *NATO ASI Series*, Springer-Verlag, Berlin, 466p.

Hulme, M., Osborn, T. J. and Johns T.C., 1998. Precipitation sensitivity to global warming: Comparison of observations with HadCM2 simulations *Geophys. Res. Letts.*, Vol. 25, p. 3379-3382.

Hydrology Handbook, 1993. David R. Maidement, Editor in Chief, McGraw-Hill Inc, p. 18.20 – 18.21.

Janowiak, J. E., Ropeleweski, C. F. and Halpert, M. S, 1986. The precipitation anomaly Classification: A method for monitoring regional precipitation deficiency and excess on a Global scale. *J. Clim. Appl. Meteo.*, Vol. 25 p. 565 - 573.

Jones, P. D., Hulme, M., Brigga, K. R., Jones, C. G., Mitchell, J. B. and Murphy, J.M., 1996. Summer Moisture Availability over Europe in the Hadley Center *General Circulation Model based on the Palmer Drought Severity Index.* Vol. 16, No.2. p. 155 – 162

Kadioglu, M., Ozturk, N., Erdun, H. and Sen, Z., 1999. On the Precipitation climatology of Turkey by harmonic analysis. *Int. J. Climatol.*, Vol. 19, p. 1717-1728.

Karl, T. R., 1986. The sensitivity of the Palmer Drought Severity Index and Palmer's Z-Index to their calibration coefficients including potential evapotranspiration. J. Climate Appl. Meteor., Vol. 25, p. 77-86.

Kenworthy, M. J., 1964. Rainfall and water resources of East Africa. *in "Geographers and the Tropics: Liverpool essays."* Ed. By Steel et al. Longmans, London. UK. p. 111-137

Kidson, J. W., 1977. African rainfall and its relation to the upper air circulation. *J. Atmos. Sci.*, Vol. 38, p. 441-456.

Kjekshus, H., 1996. *Ecology Control and Economic Development in East African History, The case of Tanganyika 1800 – 1980.* Fountain publishers, Kampala, 222 p.

Klinke, S., and Grassman, J., 1998. Projection Pursuit Regression and Neural Networks. Discussion paper 98/17 SFB 373, Humboldt Universtät, Berlin, 46p.

Lagarius, C. J., Reeds, A. J., Wright, H. M. and Write, E. P., 1998. Convergence properties of the Nelder-Mead simplex method in low dimensions. *SIAM J. Optim.*, Vol. 9, No.1, p. 112-147.

Larsen, R. J., and Marx, M. L., 1981. *An Introduction to Mathematical Statistics and its applications*. Prentice-Hall, Inc., Englewood Cliffs, N.J., 524p.

McKee, T. B., Doesken, N. J. and Kleist, J., 1993. The relationship of drought frequency and duration to time scales. Preprints, *8th Conference on Applied Climatology*, 17th -22nd January, Anaheim, CA, p. 179-184.

Morrison, F. D., 1976. *Multivariate Statistical methods.* 2nd Ed. McGraw Hill, 415 p.

Morton, C. S, 1989. *Interpretable Projection Pursuit*, Ph.D. Thesis, Stanford University, Ca.

Nelder, J. A. and Mead, R., 1965. A simplex method for function minimization, *Comput. J.*, Vol. 7., p. 308-313.

Newel, R. E. and J. Kidson, 1984: African mean wind changes between Sahelian wet and dry periods, *J. of Climat.*, Vol. 4, p. 22 – 33.

Nguyen, V.T.V., In-na, N., and Bobee, B., 1989. New plotting position formula for Pearson Type III distribution. *J. Hydraulic Enginering*, Vol. 115 (6), p. 709 – 730.

Nicholls, N., 1987. The use of Canonical Correlation to Study Teleconnection. *Mon. Wea. Rev.*, Vol. 115, p. 393-399.

Nicholson, S. E. 1979. 'Revised rainfall series for the West African sub Tropics', *Mon. Wea. Rev.* 107, p. 620-623.

Nicholson, S. E. 1980. 'The nature of rainfall fluctuations in subtropical West Africa, *Mon. Wea. Rev.* 108, p. 473-487.

Nicholson, S. E. 1983. 'Sub-Saharan rainfall in the years 1976-1980: evidence of continued drought', *Mon. Wea. Rev.* 111, p. 1646-1654.

Nicholson, S. E. and Kim, J. 1997. The relationship of El Niño – Southern Oscillation to African rainfall. *Int. J. Climatol.*, Vol. 17, p. 117 – 135.

Nicholson, S. E. and Selato, J. C., 2000. The influence of La Niña on African rainfall. *Int. J. Climatol.*, Vol. 20, p. 1761 – 1776.

NRC (National Research Council), 1998. Decade-to-Century-Scale Climate Variability and Change: *A Science Strategy*. National Academy Press, Washington D.C., 142 p.

Nyenzi, S. B., 1988. *Mechanism of East African rainfall Variability*, Ph.D. Thesis, The Florida State University, USA 184 p.

Ogallo, L. J., 1980. Regional classification of East African rainfall stations into homogeneous groups using the method of principal component analysis. *Statistical Climatology, Development in Atmospheric Sceince*, Vol. 13, Elsevier Scientific Publishing company, p. 253 – 266.

Ogallo, L. J., 1983. Temporal fluctuations of seasonal rainfall patterns in East Africa. *Mausam*, Vol. 35, p. 175-180.

Ogallo, L. J., 1988. Relationship between seasonal rainfall in East Africa and Southern Oscillation. *Int. J. Climatol.*, Vol. 8, p. 34 – 43.

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

Okoola, E. R., 1999. A diagnostic study of the Eastern Africa Monsoon circulation during the Northern Hemisphere spring season. *Int. J. Climatol.*, Vol. 19, p. 143-168.

Palmer, W. C. , 1965. *Meteorological Drought.*, Research Paper, Vol. 45, U.S. Dept. of Commerce, Weather Bureau, 58 p.

Palmer, W. C., 1968. Keeping track of crop moisture conditions, nationwide; the new Crop moisture Index, *Weatherwise*, 21: p. 156-161.

Pan, Y. H., and A. H. Oort, 1983. Global climate variations associated with sea surface temperatures anomalies in the eastern equatorial Pacific Ocean for the 1958-73 period. *Mon. Wea. Rev.*, Vol. 111, p. 1244 – 1258.

Panofsky, H. A., and Brier, G. W., 1965. *Some applications of statistics in meteorology.* Penn State University Press, College Park, Pa.

Piechota, C. T., Dracup, A. J. and Fovell, G. R., 1997. Western US stream flow and atmospheric circulation patterns during El Niño – Southern Oscillation. *J. Hydr.*, Vol. 201 p. 249 – 271.

Piechota, T. C, and Dracup, J. A, 1996. Drought and regional hydrologic variation in the United States: Associations with the El Niño-Southern Oscillation. *Water Resour. Res.*, Vol. 32, p. 1359-1373

Potts, A., 1971. Application of harmonic analysis to the study of East African rainfall data. *J. Trop. Geog.*, Vol. 34, p. 31-42.

Ramage, C. S, and Raman, C.V., 1972. *Meteorological atlas of the international Indian Ocean expedition.* Vol. 2, Upper air. National Science foundation.

Rasmusson, E. M., and Carpenter, T. H., 1983. The relation between Eastern equatorial Pacific Sea surface temperature and rainfall over India and Sri Lanka, *Mon. Weather Rev.*, Vol. 111, p. 517 – 528.

Rasmusson, E. M., and Wallace, J. M., 1983. Meteorological aspects of El Niño/Southern Oscillation, *Science*, 222, p. 1195 – 1202.

Rayner, J. N., 1971. *An introduction to Spectral Analysis*, Pion, London, 174p.

Redmond, K. 1991. Climate Monitoring and Indices. *Proceeding of the Seminar and Workshop*: Drought Management and planning, May 30-June1, 1990, Denver, CO, p. 29-33.

Richman, B. M., 1986. Rotation of principal components. *J. of Climatol.*, Vol. 6, p.293-334.

Ropeleweski, C. F., and Halpert, M. S., 1986. North American precipitation and temperatures patterns associated with El Niño/Southern Oscillation, *Mon. Weather Rev., Vol. 114*, p. 2352-2362.

Ropeleweski, C. F., and Halpert, M. S., 1987. Global and regional scale precipitation patterns associated with El Niño/Southern Oscillation, Mon. *Weather Rev.*, Vol. 115, p. 1606 – 1626.

Ropeleweski, C. F., and Halpert, M. S., 1996. Quantifying Southern Oscillation – Precipitation relationships. *J. Climate,* Vol. 9, p. 1043 – 1059.

Sabbagh, M. E., and Bryson, R. A., 1962. Aspects of the precipitation climatology of Canada investigated by the method of Harmonic analysis. *Annals, Association of American Geographers*, Vol. 52, p. 426-440.

Scian, B. and Donnari, M., 1997. Analysis of the Palmer Drought Severity Index in the Semi-arid Pampas Region. *Int. J. of Climatol.*, Vol. 17, No. 3. p. 313 - 321

Scott, C. M., and Shulman, M. D., 1979. An Areal and temporal analysis of precipitation in the United States. *J. Appl. Meteor.*, Vol. 18, p. 627 – 633.

Shabbar, A. and Barnston, G. A. 1996. Skill of seasonal climate forecasts in Canada using canonical correlation analysis. *Mon. Wea. Rev.*, Vol. 124, p. 2370 - 2385.

Shafer, B. A. and Dezman, L. E., 1982. Development of a Surface Water Supply Index (SWSI) to assess the severity of drought conditions in snowpack runoff areas. *Proceedings of the Western Snow conference*, 164-175.

Sivakumar, M. V. K., 1991. Drought Spells and Drought Frequencies in West Africa. *International Crop Research Institute for the Semi Arid Tropics (ICRISAT)*, Research Bulletin No. 13, p. 181.

S-PLUS 4, 1997. *Guide to Statistics*, Data Analysis Products Division, Mathsoft, Seattle.

Subrahmanyam, V. P., 1967. Incidence and Spread of Continental Drought. *World Meteorological Organization*, Geneva, Switzerland.

Thornthwaite, C. W., 1948. An approach towards a rational classification of Climate. *Geog. Rev.*, Vol. 38, p. 55 - 94.

Titlow, J. K., 1987. A precipitation-Based Drought Index for the Deleware River basin. *Publications in Climatology*. Vol. 40. No. 2, 68 p.

Torrence, C, and Compo, G. P., 1998. A Practical Guide to Wavelet Analysis. *Bull. Amer. Met. Soc.*, Vol. 79, No. 1, p. 61–78.

Trewartha, G. T. 1981. *The Earth's Problem Climates*. 2nd Ed., The University of Wisconsin Press, 372 p.

USAID, 1992. Disaster database compiled by Labat – Anderson, Inc, under contract to the USAID Office of the US Foreign Disaster Assistance. Contact; suzanne@usaid.gov

Verschuren, D., Laird, R. K., and Cumming, F. B., 2000. Rainfall and Drought in equatorial East Africa during the past 1,100 years, *Nature*, Vol. 403, p. 410-410.

Vogel, W. R., and McMartin, E. D. (1991). Probability Plot Goodness-of-fit and skewness estimation procedures for the Pearson Type 3 Distribution. *Water. Resources Res.*, Vol. 27, No. 12, p. 3149-3158.

Wallis, J. R., Matalas, N. C., and Slack, J. R., 1974. Just a moment! *Water. Resources Res.*, Vol. 10, p. 211-219.

Webster, J. B., 1979. *Chronology, Migration and Drought in Interlacustrine Africa*, Longman and Dalhousie Univ. Press, Dalhouisie, p. 1-37.

Wilhite, D. A and Glantz, H. M, 1987. 'Understanding the drought phenomenon: The role of definitions.' in *Planning for Drought: Towards a Reduction of Society Vulnerability*, edited by Wilhite, D. A., Easterling, E. A. and Wood, D. A., Westview Press Inc., Colorado, p. 11- 27.

Yevjevich, V., Hall, W. A. and Salas, J. D., 1977. *Drought Research Needs.* Water Resources Publication, Colorado, 276 p.

Yu, Z.-P., Chu, P.-S. and Schroeder, T. A. 1997. Predictive skills of seasonal to annual rainfall variations in the US affiliated Pacific islands: canonical correlation analysis and multivariate principal component regression approaches. *J. Climate*, Vol. 10, p. 2586 – 2599.

Zeng, X., Pielke, R.A. and Eykholt, R., 1992. Extracting Lyapunov exponents from short time series of low precision. *Modern Phys. Lett. B*, Vol. 6, p.55-75.