RMetS

Royal Meteorological Society

Spatial coherence and potential predictability assessment of intraseasonal statistics of wet and dry spells over Equatorial Eastern Africa

Wilson Gitau,[a,b,c]* Laban Ogallo,[b] Pierre Camberlin[c] and Raphael Okoola[a]

[a] *Department of Meteorology, University of Nairobi, Nairobi, Kenya*
[b] *IGAD Climate Prediction and Applications Centre (ICPAC), Nairobi, Kenya*
[c] *Centre de Recherches de Climatologie (CRC), Université de Bourgogne/CNRS, Dijon, France*

ABSTRACT: The aim of this study was to derive components of the intraseasonal rainfall variations from the daily rainfall in the Equatorial Eastern Africa region and assess their spatial coherence, a pointer to their potential predictability. Daily rainfall observations from 36 stations distributed over Equatorial Eastern Africa and extending from 1962 to 2000 were used. The March to May and October to December periods commonly referred to as the long and short rainfall seasons respectively were considered.

Seasonal and intraseasonal statistics at the local (station) level were first defined. The stations were also grouped into near-homogeneous (sub-regional) zones based on daily rainfall. Similarly, seasonal and intraseasonal statistics were then derived at sub-regional level using three different approaches. Inter-station correlation coefficients of the intraseasonal statistics at local levels were finally computed and plotted as box-plots.

For the two rainfall seasons, the two statistics showing the highest spatial coherence were the seasonal rainfall totals and the number of the wet days at sub-regional level. The local variance explained for these two variables, as an average over all the sub-regions, was more than 40%. At the bottom of the hierarchy were the mean rainfall intensity and frequency of dry spells of 5 days or more which showed the least coherence, with the local variance explained being less than 10% in each season. For each of the intraseasonal components of daily rainfall considered, the short rainfall season statistics were more coherent compared to the long rainfall season. Lag-correlations with key indices depicting sea-surface temperatures in the Pacific and Indian Oceans showed that the hierarchy between the rainfall statistics in the strength of the teleconnections reflected that of spatial coherence.

KEY WORDS   rainfall; spatial coherence; intraseasonal statistics; predictability; wet and dry spells; East Africa

Received 20 August 2011; Revised 1 October 2012; Accepted 6 October 2012

1. Introduction

The East Africa region is characterized by limited natural resources especially water, minerals and arable agricultural land. High population growth rate, poor agricultural practices, deforestation, abject poverty and high levels of unemployment are but some of the socio-economic challenges that face the region. The high population growth rate has led to people migrating into the semi-arid and arid land areas thereby affecting the ecosystems of the region and rendering them more vulnerable to disasters such as drought (Bryan and Southerland, 1989). The enormous socio-economic challenges have overstretched the limited natural resources leading to decline in environmental standards, land degradation, loss of biodiversity and increased vulnerability to man-made and natural hazards most of which are weather/climate related.

The economies of East Africa countries largely depend on rainfall-dependent agriculture. Over Kenya, e.g. the agricultural sector forms the main socio-economic activity accounting for up to 30% of the country's gross domestic product, 60% of the export earnings and is the largest source of employment (ICPAC, 2006). Most farmers practice rain-fed agriculture. Rainfall has therefore remained as the most important weather factor yet it displays the largest variability in both spatio-temporal distribution and magnitudes. This has been witnessed by the recent (2005–2006, 2010–2011) droughts recorded throughout the Horn of Africa, and the subsequent localized floods following the onset of rains. The spatial and temporal variability of rainfall over Eastern Africa at different time scales are due to complex topographical features and existence of large water bodies (Mukabana and Pielke, 1996; Indeje *et al.*, 2000; Oettli and Camberlin, 2005).

Many climate studies over the region have concentrated on the understanding of atmospheric processes and prediction of rainfall at seasonal timescale based on Sea Surface Temperature (SST) indices such as El

* Correspondence to: W. Gitau, Department of Meteorology, University of Nairobi, P.O. Box 30197, 00100 Nairobi, Kenya.
E-mail: wi.gitau@uonbi.ac.ke

Niño/Southern Oscillation (ENSO) and Indian Ocean Dipole (IOD) (Ininda, 1995; Mutai and Ward, 2000; Owiti *et al.*, 2008). The upper tropospheric temperature and geopotential variables have also been used (Njau, 2006). These studies have shown that over the Eastern Africa region, the short rainfall season [October to December (OND)] has higher predictability compared to the long rainfall season [March to May (MAM)]. The long rainfall season has been associated with complex interactions between many regional and large-scale mechanisms that generally induce large heterogeneities in the spatial rainfall distribution (Ogallo, 1982; Semazzi *et al.*, 1996; Okoola, 1998; Indeje *et al.*, 2000; Camberlin and Philippon, 2002) and virtually negligible correlation with ENSO (Ogallo, 1988; Ogallo *et al.*, 1988). The higher predictability of rainfall during the OND season is attributed to the strong teleconnections it displays with the regional and global climate anomalies (Ogallo, 1988; Ogallo *et al.*, 1988; Saji *et al.*, 1999; Black *et al.*, 2003; Owiti *et al.*, 2008).

However, studies to improve the understanding of the nature and characteristics of rainfall on intraseasonal timescales (from days to weeks) are still inadequate. Although, a number of studies have investigated intraseasonal convective variability and pentad mean rainfall characteristics (Okoola, 1998; Mutai and Ward, 2000; Camberlin and Okoola, 2003), a lot is still unknown on how intraseasonal distribution patterns constrain inter-annual variability. Reciprocally, applications of seasonal rainfall prediction would require further information on how the forthcoming season will behave in terms of rainfall distribution (such as spatio-temporal organisation of wet and dry spells, probability of long dry spells, intensity of the rains among others). The first step towards addressing this issue is to assess the potential predictability of the intraseasonal rainfall variability in terms of derivatives of wet and dry spells.

The scope of this study is therefore threefold; First, various seasonal and intraseasonal statistics of the wet and dry spells at both the local (station) and sub-regional levels were derived from the daily rain-gauge observations of the two main rainfall seasons separately. Second, the spatial coherence assessment (an indication of potential predictability) of the various seasonal and intraseasonal statistics of wet and dry spells over Equatorial East Africa was undertaken. Finally, the lag-relationships between these statistics and major climate predictors known to impact on East Africa rainfall were examined.

In Section 2, we present the data used for the study (Section 2.1.), methods used to delineate the study area into near-homogenous sub-regions (Section 2.2.), derivation of the seasonal and intraseasonal statistics of wet and dry spells at local (Section 2.3.) and sub-regional (Section 2.4.) levels and determination of spatial coherence of the intraseasonal statistics of wet and dry spells (Section 2.5.). Section 3 is dedicated to the presentation of the results obtained and discussions. In the final section, the major conclusions from the study are highlighted.
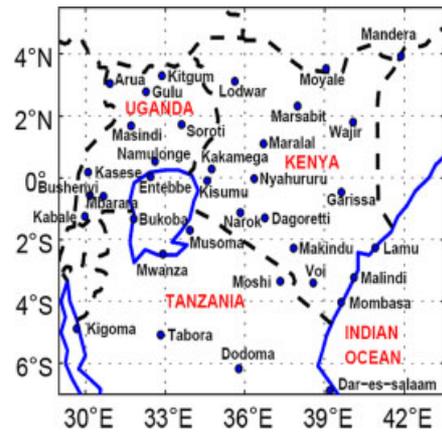
Figure 1. Network of East African rainfall stations used.

2. Data and methodology

2.1. Data sets used

Quality controlled daily rain-gauge observations covering 36 stations and evenly distributed over the Equatorial Eastern Africa region (Kenya, Tanzania, Uganda) with a long un-interrupted time series of 39 years (1962–2000) was used. This dataset was obtained from the archives of the Kenya Meteorological Department, IGAD Climate Prediction and Application Centre (Nairobi, Kenya) and Centre de Recherches de Climatologie (Dijon, France). The spatial distribution of the stations used in the study is shown in Figure 1. Southern Tanzania (south of $7°$ S) was excluded in the current study since it has one rainfall season that extends from November to May, contrary to most of the stations retained, which display bimodal (sometimes trimodal) rainfall regimes with major rainy seasons in MAM and OND.

SST indices depicting major modes of climatic variability were also used. These include the Niño indices reflecting ENSO variability, were downloaded from the Climate Prediction Centre (CPC) website, and the IOD index, representative of zonal SST gradients over the Equatorial Indian Ocean as documented by Saji *et al.* (1999); Black *et al.* (2003); Clark *et al.* (2003); Black (2005) and Owiti *et al.* (2008). The IOD index was obtained from Japan Agency for Marine-Earth Science and Technology (JAMSTEC) website, and is defined as the SST difference between the western and the eastern Equatorial Indian Ocean. The IGAD Climate Prediction and Application Centre have used these indices over the Greater Horn of Africa countries (of which the study region forms part of) for seasonal rainfall prediction. These indices are at monthly timescale and cover the period 1961–2000.

2.2. Regionalization of the study area into near-homogeneous sub-regions

The study region was first divided into several near-homogeneous sub-regions using the Varimax rotated principal component analysis (PCA) and simple correlation.

The classification of locations into spatial rainfall regimes with similar temporal rainfall characteristics aimed at reducing the local noise associated with observations from an individual location while extracting spatially coherent signals. This in return makes the interpretation of the results easier.

Since the daily rainfall distribution at each station was highly skewed, the daily rainfall totals were first transformed before the PCA was undertaken. Two approaches of the transformation that can be used are the square-root and logarithm transformations. Square-root transformation was used since it is easily applied unlike the logarithmic one, which gives some difficulties when zero rainfall amounts are considered. The square-root transformed daily rainfall series has already been used over the East Africa region (Bärring, 1988; Camberlin and Okoola, 2003). Juras (1994), Stephenson *et al.* (1999) and Fu *et al.* (2010) have indicated that square root transformation is beneficial in stabilizing the variance of sporadic rainfall time series, resulting in a good fit with a normal distribution. The square-root transformation basically reduces the weight attached to the most heavy rainfall events in the grouping of stations.

Different authors have suggested different methods that can be used to determine the number of principal components (PCs) that should be retained for rotation (Kaiser, 1959; Castell, 1966; North *et al.*, 1982; Overland and Preisendorfer, 1982). In this study, the number of the PCs to be retained and rotated was determined from the Monte Carlo simulation method. The Monte Carlo method simulates a statistical model under the assumption that a given null hypothesis H_o is true (von Storch and Zwiers, 1999). A matrix of random values of the same size as the observed data is generated, in which the temporal auto-correlation found in the observed time series is preserved. PCA was computed on this matrix, and the eigenvalues stored. This procedure was repeated 500 times. All the eigenvalues are ranked and the 95th percentile considered as the 95% confidence threshold, to which the actual eigenvalues of the observed data set are compared. All eigenvalues higher than the threshold were considered significant.

The time series obtained from Varimax rotated PCA was correlated with the stations' rainfall data and stations with significant correlation coefficient were identified. Delineation of a near-homogeneous sub-region was accomplished by identifying the stations with the largest correlation with the rotated PC time series associated with the given eigenvector of the daily rainfall in a season (Ogallo, 1980; Indeje *et al.*, 2000). To consider further the uncertainty from the regionalization of precipitation zones that might affect the predictability of the intraseasonal statistics, some stations having relatively low loadings coefficients on any rotated PC were excluded from their respective sub-regions. The spatial coherence of the intraseasonal statistics of wet and dry spells, determined using inter-station correlation coefficients as described in Section 2.5. below, was then compared with and without these stations.

2.3. Local intraseasonal statistics of wet and dry spells

A threshold for delineating wet and dry days is required when analysing spells of rainfall since the frequency distribution of the length of the wet/dry spells is highly skewed and depends on the selected threshold (Bärring *et al.*, 2006).

Some studies have used the standard observational threshold of 0.1 mm since it is the usual precision of rain-gauges (Moon *et al.*, 1994; Martin-Vide and Gomez, 1999; Dobi-Wantuch *et al.*, 2000). Perzyna (1994) used a threshold of 2.0 mm in order to remove any events featuring less rainfall and with very little significance in the river flow due to losses by interception and evaporation. Ceballos *et al.* (2004) have used a threshold of 10 mm since rainfall below this amount have only small effect on the soil water-content at a depth greater than 5 cm from the surface (Ceballos *et al.*, 2002). Recent studies have mainly used more than one threshold for delineating the dry/wet days in observational records (Ceballos *et al.*, 2004; Gitau *et al.*, 2008).

In this study, a 1.0 mm threshold is used to delineate the wet days from the dry ones. Lower thresholds (less than 1.0 mm) are more vulnerable to measurement errors associated with light rainfall and readily evaporate given the higher evapo-transpiration rate at the study region (Ogallo and Chillambo, 1982; Douguedroit, 1987; Lazaro *et al.*, 2001; Frei *et al.*, 2003).

Different authors have considered different definitions of the wet/dry spells. Two main examples of such definitions are given here. Peña and Douglas (2002) defined wet (dry) spells as days when 75% or more (35% or less) of the stations along the Pacific side of Nicaragua, Costa Rica, and Panama records rainfall. However, most authors define wet and dry spells locally. Ogallo and Chillambo (1982) defined a wet (dry) spell of length *i* as a sequence of *i* wet (dry) days preceded and followed by a dry (wet) day. It is this later definition, which is used in this study.

The occurrence of a wet or dry day is a mutually exclusive event (Chapman, 1998; Dobi-Wantuch *et al.*, 2000). The wet and dry spells of varying lengths at local level were tallied and organised into a frequency distribution table as described in Gitau *et al.* (2008). From the frequency distribution tables, various seasonal and intraseasonal statistics of the wet and dry spells (as shown in Table 1) were computed for each season and at each station. These include statistics on the frequency of wet and dry days, of a 3-days wet and 5-days dry spells, on the average and maximum length of wet and dry spells, and on the seasonal rainfall amount and average intensity per rain-day.

It is worth clarifying at this point that in order to determine the above intraseasonal statistics of the wet and dry spells, the dry periods before the first and after the last wet day were excluded. This was in order to avoid the long dry spells that occur at the beginning and at the end of the rainfall period, and which belong to the preceding and following dry seasons respectively. To accomplish this and since the date of onset and cessation

Table 1. The various seasonal and intraseasonal statistics of wet and dry spells.

Statistic	Descriptive name	Definition	Units
SR	Seasonal rainfall	Total amount of rainfall received in a season	mm
MI	Mean intensity	Average rainfall on a wet day	mm/day
NW	Wet days	Number of wet days in a season	days
ND	Dry days	Number of dry days in a season	days
MW	Mean length of wet	Average duration of consecutive wet days	days
MD	Mean length of dry	Average duration of consecutive dry days	days
LW	Longest wet	Maximum number of consecutive wet days	
LD	Longest dry	Maximum number of consecutive dry days	
3W	3 or more wet days	Frequency of wet spells of 3 days or more	
5D	5 or more dry days	Frequency of dry spells of 5 days or more	

of the rainfall period were not predetermined, the dry spells before (after) the first (last) wet spells for each rainfall season were excluded from the dry spell analysis. The onset and cessation date of the rainfall season were not predetermined because these statistics were meant to match the targets of the Regional Climate Outlook Fora, which consider the MAM and OND seasons as a whole. Additionally, the use of actual onset/cessation dates operationally would require that these variables be separately predicted.

2.4. Sub-regional intraseasonal statistics of wet and dry spells

Different methods can be used to calculate the sub-regional intraseasonal statistics of wet and dry spells (SRISS) based on daily rainfall from several locations. The near-homogeneous sub-regions over the study area were delineated earlier in Section 2.2. At sub-regional level, the local noise associated with observations from individual locations is minimized.

Figure 2 shows a schematic diagram of three different approaches that can be used to compute the SRISS. The first method involves computing the local intraseasonal statistics of wet and dry spells (LISS) at individual locations, which are then averaged for a specific near-homogeneous zone to obtain the SRISS. In the second method, the observed daily rainfall amounts from all the stations constituting a given sub-region (near-homogeneous zone) are first averaged and the SRISS derived from the sub-regional averaged rainfall. The final method involves using the PCA scores as obtained from regionalization for each sub-region (near-homogeneous zone). In the case of the PC score, the threshold that could correspond to the 1.0 mm threshold which was used at the station level was chosen. To that end, the daily rainfall data for all the n stations constituting a near-homogeneous sub-region are grouped together into a single column and sorted in ascending order (starting with the smallest). The percentile p, corresponding to the value of 1.0 mm was then obtained. The PC scores were also sorted in ascending order and the position (p/n) obtained. The PC score threshold used is the value that corresponds to the p/n position rounded off upwards. This PC score value corresponds to 1.0 mm threshold used for station data.

The three approaches available for computing the SRISS (Figure 2) were then compared, by considering the mean value of each statistic, and the representativeness of local temporal variations through correlation analyses.

2.5. Spatial coherence and potential predictability

The extraction of a spatial coherent signal (if any) is an important step towards the assessment of the potential predictability of a given climate variable. Potential predictability may be inferred from the spatial coherence analysis of sub-regional scale anomalies based on the hypothesis that any large scale climate forcing such as the SST would tend to give a rather spatially uniform signal (Haylock and McBride, 2001). It is hypothesized that low spatial coherence of any of the seasonal and intraseasonal statistics indicates that the signal is localized and thus the sub-regional potential predictability is reduced, since any large scale forcing may be masked by stronger local effects.

From the intraseasonal statistics derived at local level earlier in Section 2.3., the inter-station correlation coefficients of a given statistic were computed. These are temporal correlations between the time series of all possible couples of stations within a given sub-region. The results obtained are represented as box-plot which gives the spatial coherence for a given sub-region. To derive the spatial coherence of each intraseasonal statistic for the whole study region, the inter-station correlation coefficients of a given intraseasonal statistic for all the near-homogeneous sub-regions were also plotted as a single box-plot (Figure 8). A higher median correlation would imply higher potential predictability of the intraseasonal statistic in question. A smaller box-length would indicate that all couples of stations show similar levels of correlation.

Alternatively, the total local variance of each intraseasonal statistics that is accounted for by the sub-regional statistics was also evaluated. This was achieved by computing the simple inter-station correlation coefficient between the sub-regional intraseasonal statistics and the individual local intraseasonal indices. The correlation coefficients obtained over all the near-homogeneous sub-regions were averaged and this relates the sub-regional intraseasonal statistics and the individual local ones. The total local variance explained was the square of this single simple correlation coefficient.
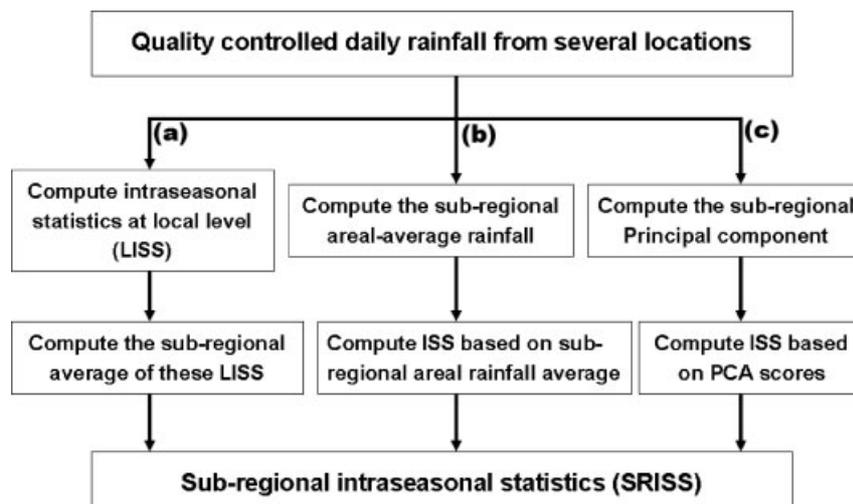
Figure 2. Schematic diagram on different approaches of calculating sub-regional intraseasonal statistics of wet and dry spells.

Finally, simultaneous and lagged simple correlations between the SRISS and climate indices (Niño and IOD) were determined at 95% confidence level. Previous studies have shown significant association of these indices with rainfall totals especially during the short rainfall season over the Eastern Africa region (Ogallo, 1988; Black *et al.*, 2003; Black, 2005; Owiti *et al.*, 2008). It is proposed here that these climate indices may also have some predictive potential for the SRISS. Since the five climate indices were significantly correlated to each other, simple linear regression models were developed using the one of climate indices that showed the highest correlation. These models were cross-validated by leaving out one observation each time, followed by the development of the model and estimation of the observation left out earlier. The root mean square error for the cross-validated simple linear regression models was finally computed for each of these models.

3. Results and discussions

3.1. Regionalisation of the study area into near-homogeneous sub-regions

The essence of the regionalisation was to group the daily rain-gauge stations which aided in the derivation of the sub-regional intraseasonal statistics. With the exception of Bärring (1988) whose study was restricted to Kenya, none of earlier studies have zoned the Equatorial Eastern Africa region into near-homogeneous rainfall sub-regions based on the observed gauged daily rainfall (Ogallo, 1980; Indeje *et al.*, 2000). The results obtained from this study were nevertheless compared with those of other studies that used observed gauged rainfall data at other timescales.

Application of the Varimax rotated PCA and simple correlation to the square-root transformed quality controlled daily rainfall data yielded six near-homogenous rainfall sub-regions for both the long and the short

rainfall seasons as shown by Figures 3(a) and (b), respectively. This simply means that only six PCs were found to be significant at 95% confidence level according to Monte Carlo testing. However, the stations at the near-homogeneous zones are different for each rainfall season. The boundaries between the sub-regions (Figure 3) are just indicative but take into account the main topographical features. It should be stressed that PCA produces patterns of rainfall variability rather than patterns of actual rainfall since the data after square-root transformation were normalized before the procedure was carried out, which includes removing mean rainfall (Williams *et al.*, 2007). The regionalisation was therefore based on the occurrence and intensity of daily rainfall events, so that stations/locations that receive rainfall under related synoptic conditions fall within the same sub-region (Tennant and Hewitson, 2002).

The six near-homogeneous zones derived from the daily rainfall series over Equatorial East Africa during the long rainfall season (Figure 3(a)) are:

1 CK – Central and Western Kenya (represented by Dagoretti, Kakamega, Kisumu, Lodwar, Maralal, Narok and Nyahururu stations);
2 CS – Coastal strip of Kenya and Tanzania (represented by Dar-es-salaam, Lamu, Malindi and Mombasa stations);
3 NK – Northeastern Kenya (represented by Mandera, Marsabit, Moyale and Wajir stations);
4 LV – Western Tanzania and Southern Uganda (represented by Bukoba, Bushenyi, Kabale, Kigoma, Mbarara, Musoma and Mwanza stations);
5 EH – Southeastern lowlands of Kenya and Northeastern Tanzania (represented by Dodoma, Garissa, Makindu, Moshi, Tabora and Voi stations); and
6 WU – Western Uganda (represented by Arua, Entebbe, Gulu, Kasese, Kitgum, Masindi, Namulonge and Soroti stations).
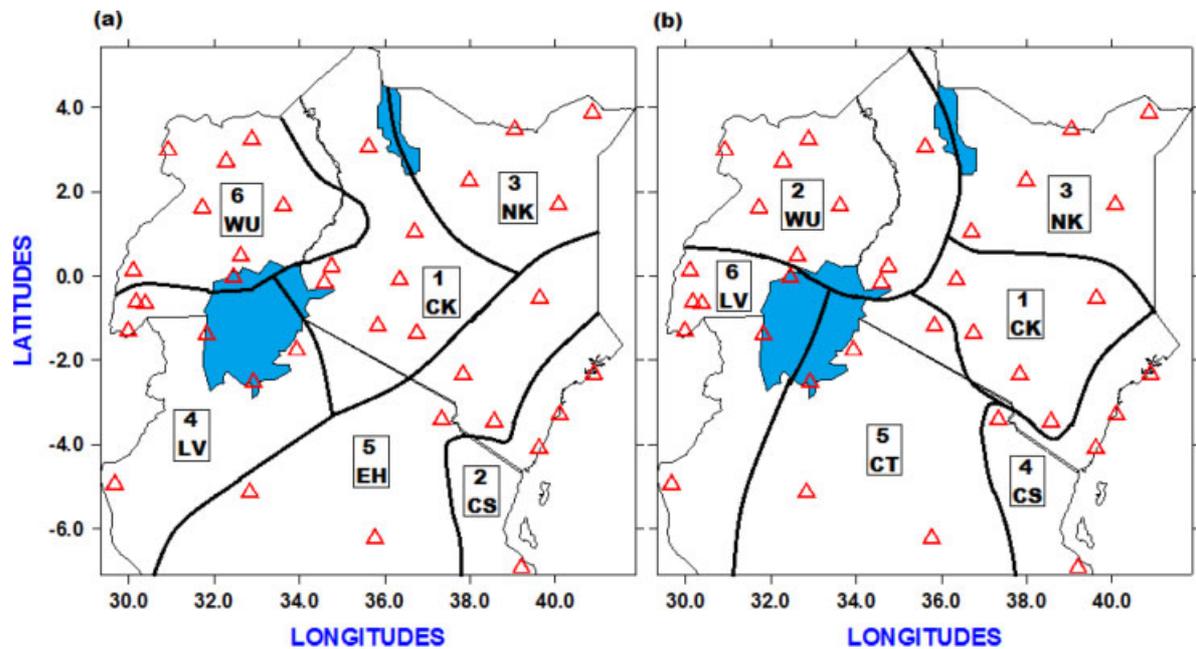
Figure 3. The near-homogeneous daily rainfall sub-regions for the (a) long and (b) short rainfall seasons over Equatorial Eastern Africa.

Two stations, Lodwar and Arua (located on the northern fringe of study region) had low loadings for the long rainfall season on any PCs. They were therefore individually grouped with the closest sub-region delineated. The consequence of this on spatial coherence will be discussed later in the relevant section.

The six near-homogeneous zones delineated for the daily rainfall series during the short rainfall season (Figure 3(b)) are:

1 CK – Central Kenya and southeastern lowlands (represented by Dagoretti, Garissa, Makindu, Nyahururu and Voi stations);
2 WU – Western Kenya and most parts of Uganda (represented by Arua, Gulu, Kakamega, Kisumu, Kitgum, Lodwar, Masindi, Namulonge and Soroti stations);
3 NK – Northeastern Kenya (represented by Mandera, Maralal, Marsabit, Moyale and Wajir stations);
4 CS – Coastal strip of Kenya and Tanzania (represented by Dar-es-salaam, Lamu, Malindi, Mombasa and Moshi stations);
5 CT – Central and Northern Tanzania (represented by Dodoma, Musoma, Mwanza, Narok and Tabora stations); and
6 LV – Western of Lake Victoria and western Tanzania (represented by Bukoba, Bushenyi, Entebbe, Kabale, Kasese, Kigoma and Mbarara stations).

The spatial patterns of the near-homogeneous zones for the two seasons are quite different with the exception of northeastern Kenya and the coastal strip of Kenya and Tanzania (Figure 3(a) and (b)). The variations in the spatial patterns for the near-homogeneous rainfall sub-regions for the two rainfall seasons may point to the different oceanic and atmospheric dynamics responsible for the behaviour of climate during the various seasons of the year (Camberlin and Philippon, 2002). Ininda (1995) advocated for a monthly analysis of the long rainfall season since it is not as homogeneous as the short rainfall season.

Although, the spatial pattern of near-homogeneous sub-regions obtained in this study was quite similar to those constructed by Ogallo (1988) and Indeje et al. (2000), there were slight variations. These variations were attributed to the fact that the spatial patterns assessed from daily rainfall data arisen from both interannual and intraseasonal variability, whereas the use of seasonal and annual totals provided information on interannual variability only.

Collectively, the six significant PCs accounted for 36.67% and 36.26% of the total variance in the daily rainfall during the long and short rainfall seasons, respectively (Table 2). The low percentage of the total variance of daily rainfall explained may be attributed to the noise in the daily rainfall that is smoothed out when summed to obtain seasonal and annual totals. It is also due to the fact that the synoptic and mesoscale systems that mainly influence rainfall at a daily time scale only affect a few stations at a time, but cannot be captured at seasonal or higher timescales. Station rainfall data are only representative of some area of variable size and shape surrounding the rain gauge (Huffman et al., 1997). Such point measurements of a spatially variable parameter can be highly erratic relative to the rest of the area, and be a poor representation of the effect of the large-scale processes. This is particularly true of daily data.

3.2. LISS: mean patterns

The spatial patterns of the mean length of wet and dry spells during the long rainfall season are shown by

Table 2. The eigenvalues, variance and cumulative variance explained in the daily rainfall for the long (MAM) and short (OND) rainfall seasons.

Principal component	Eigenvalue		Variance explained (%)		Cumulative variance explained (%)	
	MAM	OND	MAM	OND	MAM	OND
1	2.70	2.54	7.49	7.04	7.49	7.04
2	2.51	2.50	6.97	6.96	14.46	14.00
3	2.22	2.21	6.16	6.13	20.62	20.13
4	2.05	2.09	5.70	5.80	26.32	25.93
5	1.89	1.98	5.26	5.49	31.59	31.42
6	1.83	1.74	5.08	4.83	36.67	36.26

Figure 4(a) and (b), respectively. In these figures, the kriging method of interpolation was used to extrapolate over areas where the data was not adequate especially over northern Kenya and Tanzania. Details on this interpolation method can be found in Cressie (1990, 1991). Longer wet spells are reported over the coastal strip of East Africa, northern Tanzania closer to Mt. Kilimanjaro and next to Lake Victoria, central and western Kenya (Figure 4(a)). On the other hand, longer dry spells are confined to northern and eastern Kenya, Central Tanzania and northeastern part of Uganda (Figure 4(b)). These patterns relatively conform to those of the mean seasonal rainfall amounts with drier (wetter) areas tending to show longer (shorter) dry spells and shorter (longer) wet spells. Other LISS had similar spatial patterns hence they are not shown.

Notable differences were found when the mean lengths of the wet spells for the long and short rainfall seasons were compared. The longer wet spell that was observed over the coastal strip during the long rainfall season was notably absent during the short rainfall season (not shown). The longer wet spell observed over the western parts of Lake Victoria during the long rainfall season extended into parts of western Uganda and most parts of western Tanzania during the short rainfall season. The mean length of a wet spell was slightly longer during the long rainfall season over most parts of study domain, which may help to explain why the long rainfall season receives slightly more rainfall than the short rainfall season over most parts.

In the case of the mean length of the dry spell, the spatial patterns for the long and short rainfall seasons were almost similar. However, the dry spells for the short rainfall season were slightly longer compared with the long rainfall season.

Two near-homogeneous sub-regions remain relatively unchanged between the long and the short rainfall seasons. They are the coastal strip of Kenya and Tanzania (CS) and the northeastern parts of Kenya (NE) in Figure 3(a) and (b). The coastal strip (CS) was used to compare the local seasonal and intraseasonal statistics during the two rainfall seasons. The results show that the long rainfall season records more (less) number of wet (dry) days and the longest wet spell, the mean lengths of the wet (dry) spells are longer (shorter), and higher (lower) frequency of wet (dry) spells of 3 (5) days or more are higher (lower) compared to the short rainfall. The mean seasonal rainfall amounts were more during the long rainfall season.

3.3. SRISS

Figure 5(a)–(d) illustrates two ways of deriving the SRISS by using the coastal strip of Kenya and Tanzania (sub-region 2) as an example. Figure 5(a) shows a line graph of the PC score for this sub-region during the MAM 1977 season. In this instance, the 1.0 mm threshold used elsewhere corresponds to a PC score of -0.214. The distribution of the wet days obtained from the PC score with a blue dot representing a wet day over the sub-region is shown by Figure 5(b). Figure 5(c) shows the distribution of the wet days at local (station) level where a red dot represents a wet day at a station. Figure 5(d) shows the distribution of wet days obtained by averaging the rainfall amounts and plotting the resultant series while maintaining the 1.0 mm threshold.

The local and sub-regional statistics obtained from Figure 5(a)–(d) are shown by Table 3. The table shows that there is an outright bias if the daily rainfall amounts from the individual stations are averaged and then used to determine the sub-regional statistics (second last column in Table 3). For instance, while the other two methods give 31.5 and 29 as the number of wet days in a season (NW), this approach gave 49 wet days. This approach tends to overestimate the components of the wet statistics while underestimating those of the dry statistics. The method of temporal averaging daily rainfall time-series (approach (b) as per Figure 2) before generating ISS is therefore found unsuitable and the SRISS derived from it were disregarded from further analysis. Averaging the intraseasonal statistics obtained at the local level to obtain areal-averaged intraseasonal statistics on the other hand gives results that are consistent with those of the PCA score analysis. Bärring *et al.* (2006) have shown that the threshold for delineating wet/dry days on area-average are quite different as compared to when using the point observational data. They found out that by using the threshold of 1.0 mm to delineate the wet and dry days on the point observations, the threshold had to be adjusted in order to obtain the same results as those of point observations.

The interannual time-series of the various intraseasonal statistics obtained at the local level (for each station) were
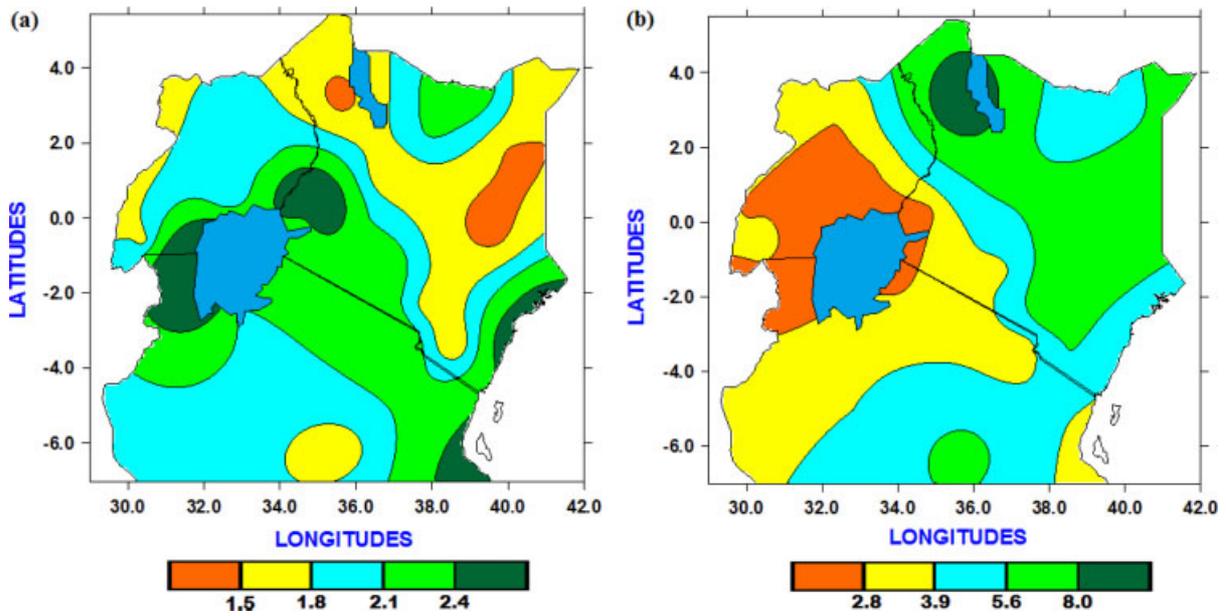
Figure 4. Spatial patterns of the mean length of (a) wet and (b) dry spells in days for the long rainfall season (1962–2000).
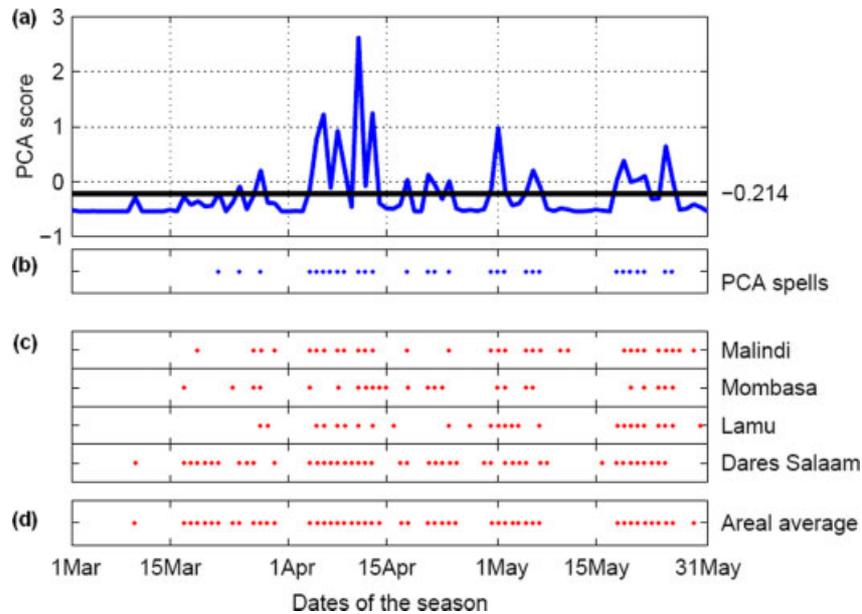
Figure 5. The temporal distribution of wet and dry spells during the MAM 1977 over the coastal strip of Kenya and Tanzania (sub-region 2) at local and sub-regional levels. (a) The PC score time series, (b) the distribution of wet and dry spells from the PCA score time series, (c) the distribution of wet and dry spells at four individual stations and (d) the distribution of wet and dry spells derived the averaged daily rainfall of the four stations. The x-axis is the dates of the season and is common to the four graphs. A blue dot represents a wet day over the sub-region. A red dot represents a wet day at a station.

correlated with those obtained from the PCA scores and those obtained by areal-averaging of the intraseasonal statistics from all stations within a given sub-region. The aim was to assess how representative of the local rainfall distribution were the two types of the sub-regional indices. In both rainfall seasons, seasonal rainfall totals (SR) and number of wet days have the highest correlation coefficients while the frequency of dry spells of 5 days or more have the least coefficient for the two types of sub-regional intraseasonal indices. In the case of areal-averaged SRISS, both seasons show significant correlations for all the components considered. However, more outliers are also observed in these correlations as compared to the PCA-based SRISS. On the whole, the correlation coefficient values obtained for the PCA-based SRISS are lower, which simply means that the PCA-based data was less representative of the local rainfall distribution of the wet and dry spells and the derived seasonal and intraseasonal statistics. However, both were retained in the rest of the study because the average

Table 3. The local and sub-regional seasonal and intraseasonal statistics of wet and dry spells for MAM 1977 over the coastal strip of Kenya and Tanzania.

	Local scale				Sub-regional scale		
ISS	Malindi	Mombasa	Lamu	Dar-es-salaam	Areal-average of LISS	Average rainfall	PCA score
SR	298.4	196.9	243.8	525.4	316.1	316.1	
MI	9.63	8.20	9.38	11.68	9.72	6.45	
NW	31	24	26	45	31.5	49	29
ND	41	47	38	32	39.5	32	37
MW	2.21	1.85	2.00	3.75	2.45	4.90	2.42
MD	3.15	3.92	3.17	2.91	3.29	3.56	3.36
LW	4	5	5	10	6	11	6
LD	7	13	10	7	9.25	10	10
3W	6	3	3	6	4.5	6	5
5D	4	4	3	2	3.25	2	3

values of the statistics remain close to each other, as shown above for the coastal strip (Table 3). Ogallo *et al.* (1988) have used both the PC-based and arithmetic areal-average SR totals indices at near-homogeneous zones over East Africa region to study their teleconnections with the global SST anomalies.

3.4. Spatial coherence and potential predictability

For relatively homogeneous sub-regions, the spatial coherence analysis provides a measure of potential predictability (Moron *et al.*, 2006). An illustration of the within-the-region (inter-station) differences in the inter-annual variability of two intraseasonal statistics is shown by Figure 6(a) and (b). Figure 6(a) shows that all the 5 stations making up the sub-region 1 (central highlands and southeastern lowlands of Kenya) during the short rainfall season display similar year-to-year variations in the standardized number of wet days. The standardised sub-regional number of wet days derived from both PCA and areal-averaging replicate these variations very well. This reveals that the number of wet days in a season is a spatially very coherent variable over this sub-region.

For the same season and over the same sub-region, the duration of the longest dry spell between individual locations and at the sub-regional level are quite contrasted (Figure 6(b)). This simply suggests that there is high (low) potential to predict the number of wet days (duration of the longest dry spell) over the central highlands and the southeastern lowlands of Kenya during the short rainfall season.

Further comparison of the intraseasonal statistics, sub-regions and seasons in terms of spatial coherence is carried out based on the computation of inter-station correlation coefficients over the period 1962–2000. To illustrate this, Figure 7(a) and (b) shows the inter-station correlation coefficients of intraseasonal statistics of wet and dry spells at two sub-regions during the short rainfall season. For sub-region 1 (central highlands and southeastern lowlands Kenya), only the seasonal rainfall totals (SR), the mean length of the dry spells (MD), the frequency of wet spells of 3 days or more (3W) and the number of wet days in a season (NW), had significant

correlation coefficients between almost all the stations, though quite low. For the other variables, significant correlations are restricted to a few station couples (top whiskers and crosses on Figure 7(a)). On the other hand, the coastal strip of Kenya and Tanzania (sub-region 4), had significant correlation coefficients on the seasonal rainfall totals (SR), the duration of the longest wet spells (LW) and the number of wet days in a season (NW) although they were marginally significant (Figure 7(b)). Similar observations were made during the long rainfall season though the significance of correlation coefficient was slightly lower.

During the long rainfall season, when all the intraseasonal statistics for the individual sub-regions were considered, only two variables were significant namely the seasonal rainfall totals (SR) and number of wet days in a season (NW). The marginally significant correlation coefficients were only noted over the sub-region 1 (central highlands and western Kenya) and sub-region 6 (most parts of Uganda).

When all the inter-station correlation coefficients from all the sub-regions were assembled together, the seasonal rainfall totals (SR) and number of wet days in a season (NW) had the greatest spatial coherence during the two rainfall seasons (Figure 8(a) and (b)). The frequency of dry spells of 5 days or more (5D) and the mean rainfall intensity (MI) were found to have the lowest values for both seasons. The median values for the inter-station correlation coefficients during the short rainfall season were slightly higher as compared to those of long rainfall season for any chosen intraseasonal statistic. During the short rainfall season, the number of wet days in a season is even slightly more coherent than the seasonal rainfall totals.

A box-plot of all the inter-station correlation coefficients for all the sub-regions shows that merging all the inter-station correlation coefficients has the net effect of reducing the median value of the inter-station correlation coefficient (Figure 8(a) and (b)). Despite this decrease, a few variables still have significant correlation coefficients. For both rainfall seasons, these variables are the seasonal rainfall totals (SR) and number of wet days in a season (NW) only. In addition, during the long rainfall
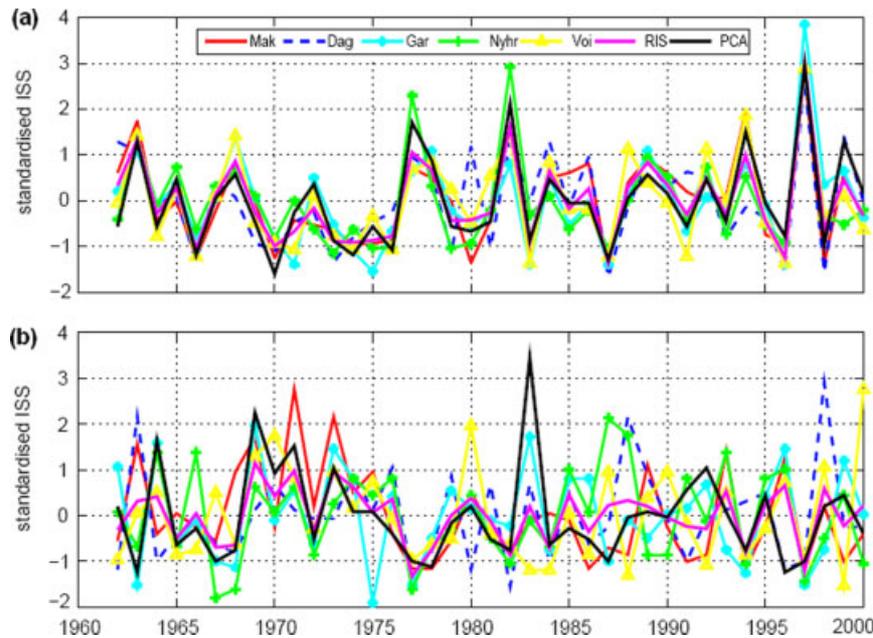
Figure 6. The standardized (a) number of wet days in a season and (b) duration of the longest dry spell over central highlands and southeastern lowlands of Kenya (sub-region 1) during the short rainfall (OND) season for the sub-region as a whole (RIS and PCA) and for the individual stations (Makindu, Dagoretti, Garissa, Nyahururu and Voi) which belongs to this sub-region.
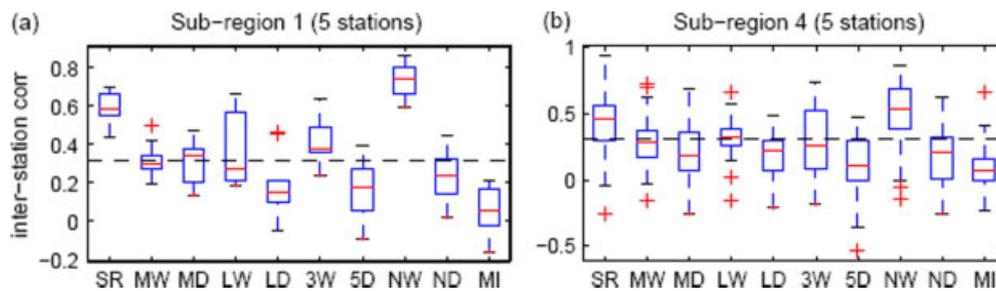
Figure 7. The inter-station correlation of the various seasonal and intraseasonal statistics of wet and dry spells over (a) central highlands and southeastern lowlands of Kenya with 5 stations (sub-region 1), and (b) coastal strip of Kenya and Tanzania with 5 stations (sub-region 4) during the short rainfall (OND) season. The dotted line across shows 95% confidence level threshold.

season, the marginally significant coherent variables were the mean length of dry spells (MD) and number of dry days (ND). During the short rainfall season, the additional variables were mean length of wet spells (MW), duration of the longest wet spell (LW) and frequency of wet spells of 3 days or more (3W). This meant that the spatial coherence (hence potential predictability) is reasonably high in a few sub-regions for these variables. Given the relatively higher spatial coherence of inter-annual anomalies of rainfall frequency compared to seasonal rainfall and mean daily rainfall intensity, recent work in the tropics have suggested that the rainfall frequency at the station scale is more seasonally predictable than the later two (Moron *et al.*, 2006, 2007; Robertson *et al.*, 2009). This has been attributed to the fact that tropical mesoscale convective clusters can produce large differences in rainfall intensity over short distances (Moron *et al.*, 2006, 2007).

The sensitivity analysis to the exact delineation of the sub-regions, by considering stations having relatively low

PCA loadings coefficients, shows that while the inclusion of such stations may lower the inter-station correlation for some of the intraseasonal statistics, the box-plots of the inter-station correlation coefficients remain relatively unchanged for most of the intraseasonal statistics. This may serve to illustrate that though there is uncertainty in the grouping of the stations, it was only marginally. For most of the intraseasonal statistics of the wet and dry spells computed, the spatial coherence results are very weakly affected by the inclusion or exclusion of any given stations within a sub-region.

The percentage of the total local variance explained by the sub-regional seasonal and intraseasonal statistics of wet and dry spells (SRISS) during the long (MAM) and short (OND) rainfall seasons for all the sub-regions combined is shown by Figure 9. The variance considered is again that of interannual variations over the period 1962–2000. The figure clearly shows that the seasonal rainfall totals (SR) and the number of wet days in a
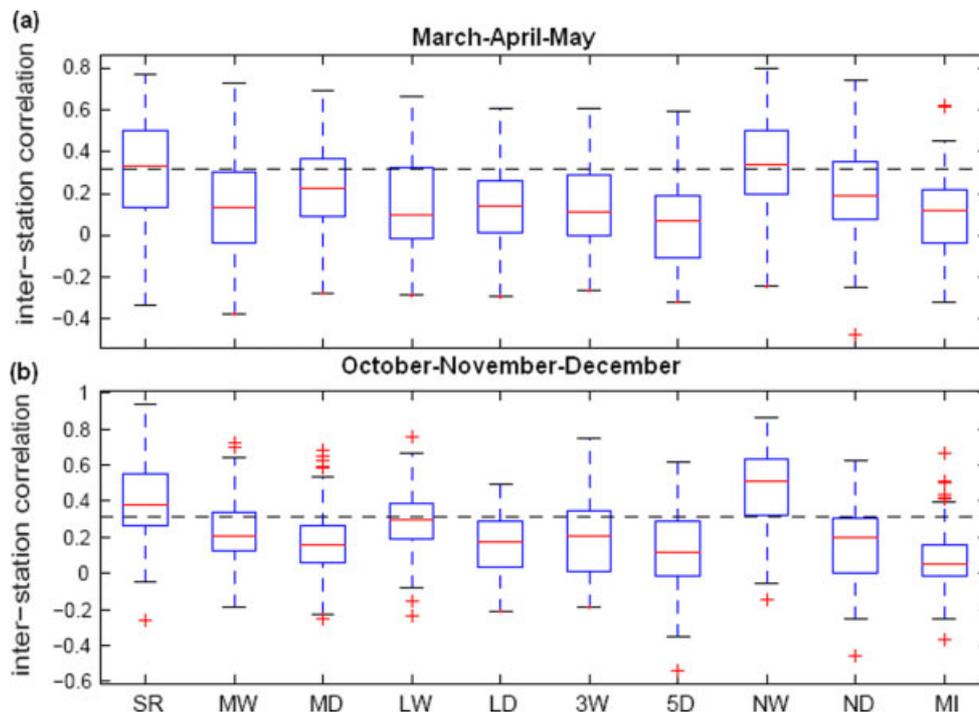
Figure 8. Box plot of inter-station correlation coefficients (per sub-region) for all stations within the study region for the (a) long (MAM) and (b) short (OND) rainfall seasons. The dotted line across shows 95% confidence level threshold.

season (NW) have higher potential predictability during the two rainfall seasons. The percentage of the local variance explained for the whole study area during the two rainfall seasons was between 30 and 60% for these two statistics from both the PCA-based and areal-average based SRISS. The frequency of dry spells of 5 days or more (5D), the duration of the longest dry spell (LD) and the mean rainfall intensity (MI) have the lowest coherence, and hence the least potential predictability. Some of the variables for the areal-average based SRISS, like the duration of longest wet spells (LW), mean length of the wet spells (MW), frequency of wet spells of 3 days or more (3W) and mean length of the dry spells (MD) explain a reasonably high percentage of variance (35–40%) for OND, which makes us expect some level of predictability. Consistent with previous studies which have shown the seasonal rainfall totals for the short rainfall season to be highly predictable and with significant teleconnection with well-known global and regional climate signals (Ogallo, 1988; Ogallo *et al.*, 1988; Indeje *et al.*, 2000; Black *et al.*, 2003; Black, 2005; Owiti *et al.*, 2008), the SRISS have higher potential predictability for the short rainfall season as compared to the long rainfall season.

In some years, anomalies of usually weakly coherent intraseasonal statistics may display similar values at all stations. This is exemplified on Figure 6(b) by the years 1972, 1977, 1978 and 1997, during which all stations in central Highlands and southeastern Kenya have anomalies of the same sign for the longest dry spell, despite the fact that on average this variable is the least spatially coherent (Figure 9). Other examples from other

intraseasonal statistics of wet and dry spells were also found (not shown), which suggest that spatial coherence (and presumably potential predictability) is also time-dependent.

It was found that the PCA-based SRISS explained very low percentages of total local variance (Figure 9). They remain below 20% (10%) for all the intraseasonal statistics apart from the seasonal rainfall totals and number of wet days in a season during the short (long) rainfall season. This could be due to the fact that the spatial signature of each PC has a much larger spatial extent than the sub-region to which it has been associated with. In other words, the PCA-based SRISS are not strictly sub-regional. The results further suggest that sub-regional indices of seasonal rainfall totals and intraseasonal statistics derived from areal-average are more representative than those derived from the PC scores.

The percentage of the variance of local random series explained by the area-average SRISS was also determined. This was accomplished by generating random Gaussian time series, and aggregating them by computing the average. The number of stations in each sub-region was however maintained. The percentage of the local variance was then computed. This was repeated 500 times and the 95th percentile extracted. It is the percentage of local variance which is exceeded only 5 times out of 100 based on random time-series. This 95% confidence level is 17% for MAM and 19% for OND. The slight difference in thresholds between MAM and OND is due to the fact that the number of stations in each sub-region is slightly
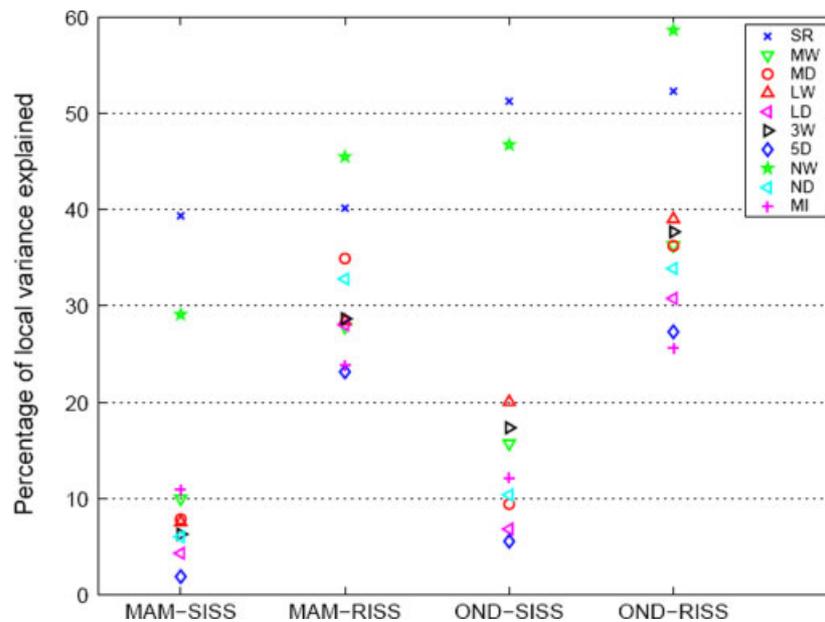
Figure 9. The local variance explained by sub-regional seasonal and intraseasonal statistics of wet and dry spells derived from PCA scores (SISS) and from areal-averaging (RISS) during the long (MAM) and short (OND) rainfall season.

different between the two seasons. All the SRISS values computed from the real data (Figure 9) surpass these thresholds which mean that the spatial coherence in all cases is significant. In other words, there is a climate signal in all the variables. However, for some variables like the frequency of dry spells of 5 days or more (5D) and the mean rainfall intensity (MI), the percentage of local variance explained is only marginally higher than the 95% confidence level threshold.

3.5. Linkages with the climate indices

Ropelewski and Halpert (1987), Janowiak (1988), Ogallo (1988), Ogallo et al. (1988) and Indeje et al. (2000) have shown that there exists a strong ENSO signal in the seasonal rainfall totals during the short rainfall season. Black et al. (2003), Black (2005), Behera et al. (2005) and Owiti et al. (2008) have shown the connection between seasonal rainfall totals and the IOD for the same season.

The lag-relationship between these two large-scale climate signals and the various SRISS in East Africa is examined for the short rain season, and for all the six sub-regions (Z1–Z6 on Figure 10). The connections between the SRISS for the long rainy season and the same climate signals are weak and not discussed here. With the consideration for a sufficient lead-time, while further noting that the lagged correlation coefficients beyond June are insignificant, a two-month average (July-August) for the climate indices was used. Figure 10(a) confirms the relationship between the seasonal rainfall totals on one hand and global and regional climate indices (Niño and IOD) on the other hand. Figure 10(b)–(j) shows the association between the different SRISS and the same SST indices, with coefficients multiplied by −1 for panels g–j for ease of comparison. The seasonal rainfall totals

(Figure 10(a)) and SRISS of wet spells (Figure 10(b)–(e)) have a positive lagged association with the Niño indices. The highest correlations were noted for Niño 3.4 and the lowest ones were those of Niño $1+2$ and IOD. The anomalous warm conditions during the boreal summer and autumn over the Niño regions induce changes in the Walker circulation, with anomalous ascending motion over Equatorial Eastern Africa and anomalous descending motion over the maritime continent and southern Africa. The anomalous ascending (descending) motions tend to bring wet (dry) conditions over Equatorial Eastern Africa (Maritime continent and southern Africa).

The association of the mean rainfall intensity and the intraseasonal statistics of the dry spells with the climate indices were rather diverse (Figures 10(f)–(j)). In many cases, the correlations are low and insignificant, but there are exceptions. The mean frequency of dry spells of 5 days or more shows the weakest control by Niño and IOD indices (Figure 10(j)). The mean rainfall intensity (Figure 10(f)), the mean duration of dry spells (Figure 10(h)) and number of dry days (Figure 10(g)) follow in that order.

Interestingly, the duration of the longest dry spell (LD) shows a relatively strong and coherent/uniform response to Niño and IOD indices (Figure 10(i)) that is similar to the seasonal rainfall totals and SRISS of wet spells (Figures 10(a)–(e)). This means that a very long dry spell is likely to occur throughout Equatorial East Africa during the short rainfall season with La Niña conditions, with potentially adverse effect on crops and other rain-fed activities. The weak control of the mean rainfall intensity by the ENSO and IOD indices and the low spatial coherence observed earlier maybe attributed to the fact that tropical mesoscale convective clusters
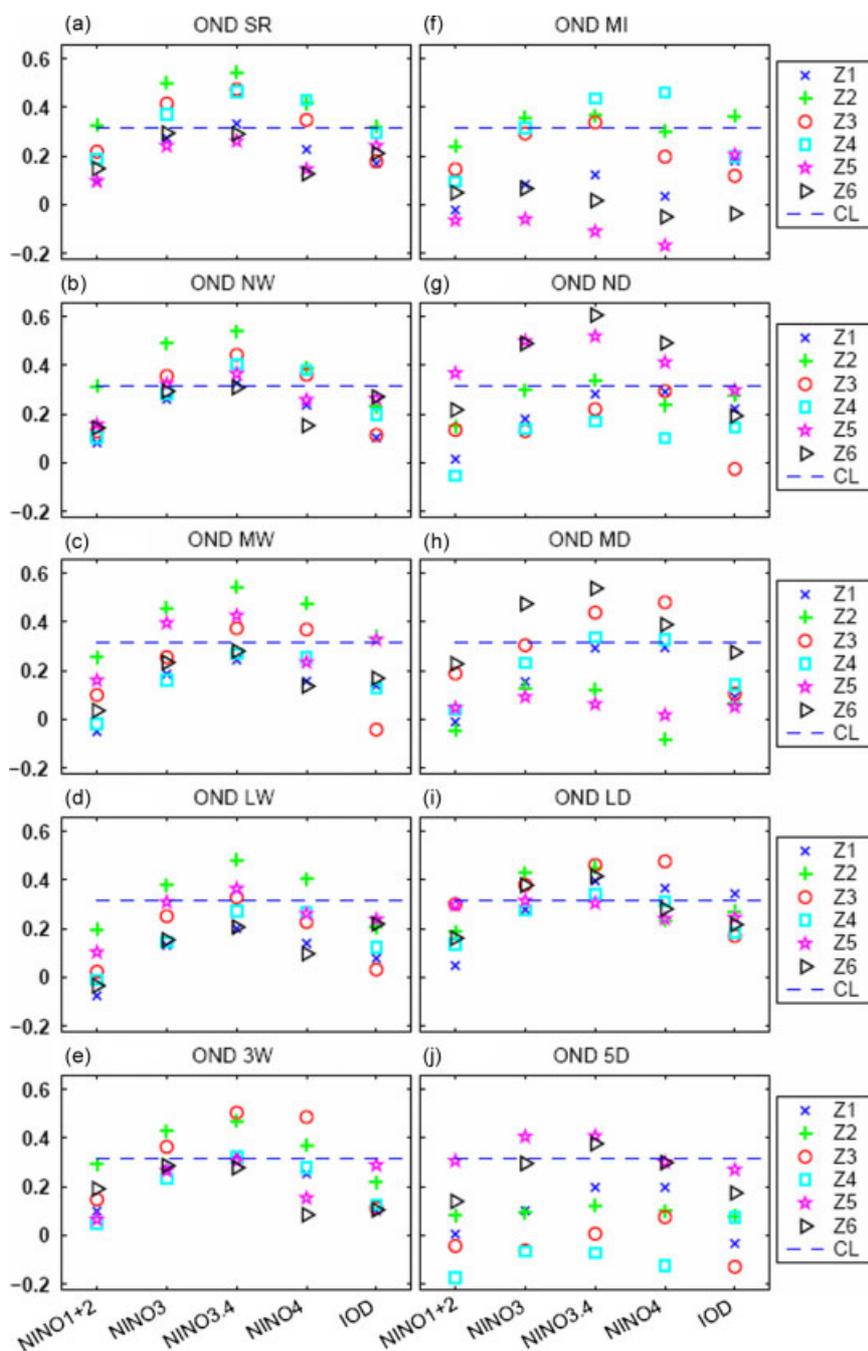
Figure 10. Correlation coefficients between climate indices averaged for July-August (x-axis) and areal-averaged October-November-December (a) seasonal rainfall totals, (b) number of wet days, (c) mean length of wet spells, (d) longest wet spell, (e) frequency of 3 wet days or more, (f) mean rainfall intensity, (g) number of dry days, (h) mean length of dry spells, (i) longest dry spell, and (j) frequency of 5 dry days or more, over the six rainfall sub-regions Z1–Z6. The correlation coefficient in figures (g), (h), (i), and (j) has been multiplied by minus one for ease in comparison. CL shows the 95% confidence level threshold.

produce large differences in rainfall intensity over short distances (Moron *et al.*, 2006, 2007).

The correlation coefficients between the various intraseasonal statistics of the wet and dry spells including seasonal rainfall totals and the cross-validated simple linear regression models output are shown in Table 4. Although significant prediction skill was displayed for some rainfall sub-regions (especially Z2 which covered Uganda and Western Kenya), the results clearly indicated that the correlation coefficients were lower compared with those in Figure 10. Most of these correlation coefficients were not significant at 95% confidence levels. Table 4 further showed the root mean square errors for the cross validated models were quite high. Even for number of wet days in a season (NW) and seasonal rainfall totals (SR) that had shown the highest spatial coherence, the regression models had significant correlation at sub-region 2 only.

Table 4. The correlation coefficients between the various intraseasonal statistics of wet and dry spells during the short rainfall season and the cross-validated simple linear regression model outputs. The numbers in bold indicate that the correlation coefficient was significant at 95% confidence level. The numbers in the brackets show the computed root mean square errors for the cross-validated simple linear regression model. The dashes indicate that the model could not be developed for that zone.

Rainfall statistics	Rainfall sub-regions					
	Z1	Z2	Z3	Z4	Z5	Z6
SR	0.100 (106.359)	**0.441** (63.375)	0.258 (103.667)	0.264 (143.970)	–	–
NW	0.066 (7.642)	**0.455** (4.357)	0.275 (5.965)	0.201 (7.483)	0.184 (6.384)	–
ND	–	0.194 (5.842)	–	–	**0.423** (5.273)	**0.545** (3.758)
MW	–	**0.461** (0.242)	0.190 (0.314)	–	0.246 (0.381)	–
MD	–	–	**0.383** (3.022)	0.166 (3.143)	–	**0.465** (0.297)
LW	–	0.391 (1.289)	0.095 (1.557)	–	0.225 (1.741)	–
LD	0.278 (3.893)	**0.336** (3.813)	**0.378** (4.593)	0.187 (5.777)	–	**0.319** (1.827)
3W	0.076 (1.240)	**0.348** (0.813)	**0.399** (0.770)	0.060 (1.407)	–	–
5D	–	–	–	–	0.277 (0.715)	0.243 (0.650)
MI	–	0.227 (1.208)	0.114 (2.454)	**0.352** (2.488)	–	–

These partial results show that despite the spatially most coherent variables displaying the greatest predictability potential as reflected from the lag-correlation with key SST indices, the regression models develop performed poorly as indicated by the computed root mean square error. A search for additional potential indices that have physical/dynamical linkages with the various SRISS is necessary. This will aid in the prediction of the occurrence of these statistics. However, this was beyond the scope of the current study.

4. Conclusions and recommendations

Application of the rotated PCA and simple correlation analysis on the square-root transformed quality-controlled daily rainfall observations showed that the occurrence and amounts of daily rainfall over Equatorial Eastern Africa can be broadly classified into six near-homogeneous rainfall regimes during both the MAM and OND rainfall seasons. The spatial patterns obtained are somehow similar to those obtained in other studies that have used seasonal or annual rainfall data. However, there are significant spatial differences in the patterns for the individual seasons, a pointer to the different atmospheric and oceanic dynamics responsible for the behaviour of climate during the various seasons of the year over the study area. The low percentage of total variance of daily rainfall explained (about 36%) was attributed to the fact that for daily rainfall observations, both the intraseasonal and interannual variability are in play while at higher timescales such as month and seasonal, only the interannual variability is considered.

The study has compared three possible methods of deriving seasonal and intraseasonal statistics for each season at sub-regional (near-homogeneous zone) level. It has concluded that SRISS obtained from averaging the daily rainfall amounts from the individual stations are the most unrealistic and thus should not be used. Although the PCA-based and arithmetic areal-averaged SRISS gave similar results, correlations between the local and the sub-regional seasonal and intraseasonal statistics for a given sub-region were lower for the PCA-based SRISS, an indication that statistics derived this way are less representative of the local rainfall distribution. The PCA-based SRISS explained less than 20% (10%) of the local variance for all the SRISS apart from the seasonal rainfall totals and number of wet days in a season during the short (long) rainfall season. By contrast, during the two rainfall seasons, the arithmetic areal-averaged SRISS explained a percentage of local variance of SRISS for any statistic which is significantly higher than that obtained by chance, which denotes spatial coherence and hence some degree of potential predictability.

At the sub-regional level, it has been shown that there is higher probability to predict certain intraseasonal rainfall components than others. Consistent with previous studies, the number of wet days in a season was the spatially most coherent SRISS and closely followed by the seasonal rainfall totals. The frequency of dry spells of 5 days or more, the mean rainfall intensity and the duration of the longest dry spell were the spatially least coherent SRISS and hence likely to be the least predictable. Consistent with earlier studies, which considered only a fraction of East Africa and a more limited number of rainfall variables, the intraseasonal statistics of wet and dry spells and seasonal rainfall totals during the short rainfall season are more coherent and potentially more predictable compared to those of the long rainfall season. The hypothesis that a higher spatial coherence reflects a higher predictability tends to be confirmed by preliminary results on lag-relationships between SRISS and two SST predictors depicting ENSO and the IOD. The most coherent variables generally display the largest correlations with these SST indices. However, the linear regression models developed performed poorly since only one SST index was used in each model.

The study recommends that for each sub-region, an assessment of how much of the spatial coherence comes from the intraseasonal variability and how much comes from the interannual variability be undertaken. It may happen that the interannual variations are strongly similar

(when wet years occur simultaneously at all stations) whereas the day-to-day variability does not agree much between the stations, or vice versa.

The study suggests that for those SRISS with marginally significant spatial coherence, an alternative approach would be to derive the intraseasonal statistics of the wet and dry spells at local level first, which are then used to regionalise the study area into near-homogeneous sub-regions. From this, some large-scale signal (if any) may thus emerge and possibly enable to detect additional predictors.

Acknowledgements

References

Bärring L. 1988. Regionalization of daily rainfall in Kenya by means of common factor analysis. *International Journal of Climatology* **8**: 371–389.

Bärring L, Holt T, Linderson ML, Radziejewski M, Moriondo M, Palutikof JP. 2006. Defining dry/wet spells for point observations, observed area averages and regional climate model gridboxes in Europe. *Climate Research* **31**: 35–49.

Behera SK, Luo JM, Masson S, Delecluse P, Gualdi S, Navarra A, Yamagata T. 2005. Paramount Impact of the Indian Ocean Dipole on the East African short rains: A CGCM study. *Journal of Climate* **18**: 4514–4530.

Black E. 2005. The relationship between Indian Ocean sea-surface temperature and East African rainfall. *Philosophical Transactions of the Royal Society A* **363**: 43–47.

Black E, Slingo JM, Sperber KR. 2003. An observational study of the relationship between excessively strong short rains in coastal East Africa and Indian Ocean SST. *Monthly Weather Review* **103**: 74–94.

Bryan, R. B. and R.A. Southerland, 1989. Erosion and soil conservation in semi-arid tropical region: Rift Valley Province, Kenya. *Sixth International Soil Conservation Conference*, Ethiopia, 6–18 November 1989.

Camberlin P, Okoola RE. 2003. The onset and cessation of the "long rains" in eastern Africa and their interannual variability. *Theoretical and Applied Climatology* **75**: 43–54.

Camberlin P, Philippon N. 2002. The East African March–May rainy season: Associated atmospheric dynamics and predictability over the 1968–1997 period. *Journal of Climate* **15**: 1002–1019.

Castell RB. 1966. The scree-test for the number of factors. *Multivariate Behavioral Research* **1**: 245–276.

Ceballos A, Martinez-Fernandez J, Santos F, Alonso F. 2002. Soil water behaviour of sandy soils under semiarid conditions in the Duero Basin (Spain). *Journal of Arid Environments* **51**: 501–519.

Ceballos A, Martínez-Fernández J, Luengo-Ugidos MÁ. 2004. Analysis of rainfall trends and dry periods on a pluviometric gradient representative of Mediterranean climate in the Duero Basin, Spain. *Journal of Arid Environments* **58**: 215–233.

Chapman TG. 1998. Stochastic modeling of daily rainfall: the impact of adjoining wet days on the distribution of rainfall amounts. *Environmental Modeling and Software* **13**: 317–324.

Clark CO, Webster PJ, Cole JE. 2003. Interdecadal variability of the relationship between the Indian Ocean Zonal Mode and East African coastal rainfall anomalies. *Journal of Climate* **16**: 548–554.

Cressie NAC. 1990. The origins of Kriging. *Mathematical Geology* **22**: 239–252.

Cressie NAC. 1991. *Statistics for Spatial Data*. John Wiley and Sons, Inc. 900 pp: New York.

Dobi-Wantuch I, Mika J, Szeidi L. 2000. Modeling wet and dry spells with mixture distributions. *Meteorology and Atmospheric Physics* **73**: 245–256.

Douguedroit A. 1987. The variation of dry spells in Marseilles from 1865 to 1984. *Journal of Climatology* **7**: 541–551.

Frei C, Christensen JH, Deque M, Jacob D, Jones RG, Vidale PL. 2003. Daily precipitation statistics in regional climate models: Evaluation and intercomparison for the European Alps. *Journal of Geophysical Research* **108**: 4124. DOI: 1029/2002DJ002287.

Fu G, Viney NR, Charles SP. 2010. Evaluation of various root transformations of daily precipitation amounts fitted with a normal distribution for Australia. *Theoretical and Applied Climatology* **99**: 229–238.

Gitau W, Ogallo L, Mutemi JN. 2008. Intraseasonal characteristics of wet and dry spells over Kenya. *Journal of Kenya Meteorological Society* **2**(1): 19–30.

Haylock M, McBride J. 2001. Spatial coherence and predictability of Indonesian wet season rainfall. *Journal of Climate* **14**: 3382–3887.

Huffman GJ, Adler RF, Arkin P, Chang A, Ferraro R, Gruber A, Janowiak J, McNab A, Rudolf B, Schneider U. 1997. The Global Precipitation Climatology Project (GPCP) combined precipitation dataset. *Bulletin of the American Meteorological Society* **78**(1): 5–20.

ICPAC, 2006. Factoring climate products for improved agriculture production and food security for East Africa region. In *Proceedings of Regional Capacity Building and Training Workshop*. 21–28 August 2006, Nairobi, Kenya.

Indeje M, Semazzi FHM, Ogallo LJ. 2000. ENSO signals in East African Rainfall and their prediction potentials. *International Journal of Climatology* **20**: 19–46.

Ininda JM. 1995. *Numerical simulation of the influence of sea surface temperature anomalies on the East African seasonal rainfall*. PhD thesis, Department of Meteorology, University of Nairobi, Kenya.

Janowiak JE. 1988. An investigation of interannual rainfall variability in Africa. *Journal of Climate* **1**: 240–255.

Juras J. 1994. Some common features of probability distributions for precipitation. *Theoretical and Applied Climatology* **49**: 69–76.

Kaiser HF. 1959. The varimax criterion for analytic rotation in factor analysis. *Psychometrika* **23**: 187–200.

Lazaro R, Rodrigo FS, Gutierrez L, Domingo F, Puigdefabregas J. 2001. Analysis of a 30-year rainfall record (1967–1997) in semi–arid SE Spain for implications on vegetation. *Journal of Arid Environments* **48**: 373–395.

Martin-Vide J, Gomez L. 1999. Regionalisation of Peninsular Spain based on the length of dry spells. *International Journal of Climatology* **19**: 537–555.

Moon SE, Ryoo SB, Kwon JG. 1994. A Markov chain model for daily precipitation occurrence in South Korea. *International Journal of Climatology* **14**: 1009–1016.

Moron V, Robertson AW, Ward MN. 2006. Seasonal predictability and spatial coherence of rainfall characteristics in the tropical setting of Senegal. *Monthly Weather Review* **134**: 3248–3262.

Moron V, Robertson AW, Ward MN, Camberlin P. 2007. Spatial coherence of tropical rainfall at the regional scale. *Journal of Climate* **20**: 5244–5263.

Mukabana JR, Pielke RA. 1996. Investigating the influence of synoptic-scale monsoonal winds and mesoscale circulations on diurnal weather patterns over Kenya using a mesoscale numerical model. *Monthly Weather Review* **124**: 224–243.

Mutai CC, Ward MN. 2000. East African rainfall and the tropical circulation/convection on intraseasonal to interannual timescales. *Journal of Climate* **13**: 3915–3939.

Njau LN. 2006. *Diagnostics and predictability of East African rainfall with tropospheric circulation parameters*. PhD thesis, Department of Meteorology, University of Nairobi, Kenya.

North GR, Bell TL, Cahalan RF, Moeng FJ. 1982. Sampling errors in the estimation of empirical orthogonal functions. *Monthly Weather Review* **110**: 699–706.

Oettli P, Camberlin P. 2005. Influence of topography on monthly rainfall distribution over East Africa. *Climate Research* **28**: 199–212.

Ogallo LJ. 1980. Regional classification of the East African Rainfall stations into homogeneous groups using the method of Principal Component Analysis. *Statistical Climatology: Developments in Atmospheric Science* **13**: 255–266.

Ogallo LJ. 1982. Quasi-periodic patterns in the East African rainfall records. *Kenya Journal of Science and Technology* **A3**: 43–54.

Ogallo LJ. 1988. Relationship between seasonal rainfall in East Africa and Southern Oscillation. *International Journal of Climatology* **8**: 31–43.

Ogallo LJ, Chillambo AW. 1982. The characteristics of wet spells in Tanzania. *East African Agricultural and Forestry Journal* **47**(4): 87–95.

Ogallo LJ, Janowiak JE, Halpert MS. 1988. Teleconnection between seasonal rainfall over Eastern Africa and Global Sea surface temperature anomalies. *Journal of the Meteorological Society of Japan* **66**: 807–822.

Okoola RE. 1998. Spatial evolutions of the active convective patterns across the equatorial eastern Africa during Northern Hemisphere spring season using outgoing longwave radiation records. *Meteorology and Atmospheric Physics* **66**: 51–63.

Overland JE, Preisendorfer RW. 1982. A significance test for principal components applied to a cyclone climatology. *Monthly Weather Review* **110**: 1–4.

Owiti Z, Ogallo LA, Mutemi J. 2008. Linkages between the Indian Ocean Dipole and East African Seasonal Rainfall Anomalies. *Journal of Kenya Meteorological Society* **2**(1): 3–17.

Peña M, Douglas MW. 2002. Characteristics of wet and dry spells over the Pacific side of Central America during the rainy season. *Monthly Weather Review* **130**: 3054–3073.

Perzyna G. 1994. Spatial and temporal characteristics of maximum dry spells in southern Norway. *International Journal of Climatology* **14**: 895–909.

Robertson AW, Moron V, Swarinoto Y. 2009. Seasonal predictability of daily rainfall statistics over Indramayu district, Indonesia. *International Journal of Climatology* **29**: 1449–1462.

Ropelewski CF, Halpert MS. 1987. Global and regional scale precipitation patterns associated with El Niño/Southern Oscillation. *Monthly Weather Review* **115**: 1606–1626.

Saji NH, Goswami BN, Vinayachandran PN, Yamagata T. 1999. A dipole mode in the tropical Indian Ocean. *Nature* **401**: 360–363.

Semazzi HFM, Burns B, Lin NH, Schemm JE. 1996. A GCM study of the teleconnections between the continental climate of Africa and global sea-surface temperature anomalies. *Journal of Climate* **9**: 2480–2497.

Stephenson DB, Rupa Kumar K, Doblas-Reyes FJ, Royer JF, Chauvin F, Pezzulli S. 1999. Extreme daily rainfall events and their impacts on ensemble forecasts of the Indian monsoon. *Monthly Weather Review* **127**: 1954–1966.

von Storch H, Zwiers FW. 1999. *Statistical Analysis in Climate Research*. Cambridge University Press pp: 494, Cambridge.

Tennant WJ, Hewitson BC. 2002. Intra-seasonal rainfall characteristics and their importance to the seasonal prediction problem. *International Journal of Climatology* **22**: 1033–1048.

Williams CJR, Kniveton DR, Layberry R. 2007. Climatic and oceanic associations with daily rainfall extremes over southern Africa. *International Journal of Climatology* **27**: 93–108.