INTERNATIONAL JOURNAL OF CLIMATOLOGY
Int. J. Climatol. (2014)
Published online in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/joc.4131

RMetS
Royal Meteorological Society

Oceanic and atmospheric linkages with short rainfall season intraseasonal statistics over Equatorial Eastern Africa and their predictive potential

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

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

ABSTRACT: Despite earlier studies over various parts of the world including equatorial Eastern Africa (EEA) showing that intraseasonal statistics of wet and dry spells have spatially coherent signals and thus greater predictability potential, no attempts have been made to identify the predictors for these intraseasonal statistics. This study therefore attempts to identify the predictors (with a 1-month lead time) for some of the subregional intraseasonal statistics of wet and dry spells (SRISS) which showed the greatest predictability potential during the short rainfall season over EEA. Correlation analysis between the SRISS and seasonal rainfall totals on one hand and the predefined predictors on the other hand were initially computed and those that were significant at 95% confidence levels retained. To identify additional potential predictors, partial correlation analyses were undertaken between SRISS and large-scale oceanic and atmospheric fields while controlling the effects of the predefined predictors retained earlier. Cross-validated multivariate linear regression (MLR) models were finally developed and their residuals assessed for independence and for normal distribution. Four large-scale oceanic and atmospheric predictors with robust physical/dynamical linkages with SRISS were identified for the first time. The cross-validated MLR models for the SRISS of wet spells and seasonal rainfall totals mainly picked two of these predictors around the Bay of Bengal. The two predictors combined accounted for 39.5% of the magnitude of the SST changes between the July–August and October–November–December periods over the Western Pole of the Indian Ocean Dipole, subsequently impacting EEA rainfall. MLR models were defined yielding cross-validated correlations between observed and predicted values of seasonal totals and number of wet days ranging from 0.60 to 0.75, depending on the subregion. MLR models could not be developed over a few of the subregions suggesting that the local factors could have masked the global and regional signals encompassed in the additional potential predictors.

KEY WORDS predictability; equatorial Eastern Africa; intraseasonal statistics; wet and dry spells; Bay of Bengal; Indian Ocean Dipole; sea surface temperature

Received 31 October 2013; Revised 7 June 2014; Accepted 17 July 2014

1. Introduction

Precipitation remains the most difficult climate component for models to forecast, because it is closely related not only to the large-scale atmospheric dynamics but also to the small-scale weather systems, and is influenced by multiple atmospheric variables and local topography.

The annual cycle of equatorial Eastern African (EEA) rainfall is mostly bimodal in nature. The two wet seasons are locally referred to as the long and short rainfall seasons. The long rainfall period occurs within the March–April–May period and contributes much of the annual rainfall while the short rainfall season occurs from October to December. The two rainfall seasons that coincide with the transitions between the winter and summer monsoons are a result of the migration of the Intertropical Convergence Zone (ITCZ). The northward migration of the zonal arm of the ITCZ results in long rainfall season. The southward migration of the ITCZ zonal arm that results in the short rainfall season is more rapid than the northward migration, and thus the period of heavy rainfall is generally shorter and the average daily rainfall is lower. However, much of the interannual rainfall variability comes from the short rainfall season with the coefficient of variability of about 74% compared to 35% for the long rainfall season (Downing *et al.*, 2008). The recent droughts (1999–2001 and 2005–2006) and excessive rainfall leading to extensive flooding (1997) occurred in the short rainfall season. This calls for a thorough understanding of the climatic drivers of the short rainfall season, as well as an accurate and timely prediction of the anomalous rainfall conditions during this season.

National Research Council (2010) has highlighted three interrelated categories of predictability sources that exist within the climate system. The first one relates to particular variables that exhibit inertia or memory, such

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

as ocean heat content, in which anomalous conditions can take relatively long periods of time to decay. The second type of predictability source relates to patterns of variability or feedbacks. Coupling among processes in the climate system usually gives rise to characteristic patterns that explain some portion of the spatial and temporal variance exhibited by key climate variables such as temperature or precipitation. One such example is the El Niño-Southern Oscillation (ENSO) in which anomalous conditions in the tropical Pacific Ocean influence seasonal climate in the tropics around the globe. The final source of predictability is due to external forcing such as volcanic eruptions, changes in solar activity and the accumulation of greenhouse gases in the atmosphere among others.

Sea surface temperatures (SSTs) play an important role in modulating rainfall variability. Idealized SST anomalies have been used to force global and regional circulation models to simulate rainfall variability and study the physical mechanisms behind the variability over various regions (Reason, 2002; Li *et al.*, 2003; Misra, 2003; Moron *et al.*, 2003; Paeth and Friederichs, 2004). These studies and others based on observations (Janowiak, 1988; Walker, 1990; Jury and Pathack, 1993; Jury *et al.*, 1993; Mason, 1995; Shinoda and Kawamura, 1996; Reason and Lutjeharms, 1998; Biasutti *et al.*, 2004; Nicholson and Dezfuli, 2013) have suggested that the role played by SST in modulating rainfall variability is either quite direct, through enhanced convection over warm waters for instance, or indirect, through an alteration in the position of the ITCZ. Previous studies over East Africa have documented strong relationships between the interannual rainfall variability during the short rainfall season and SST over the global oceans. Ogallo (1988), Ogallo *et al.* (1988), Goddard and Graham (1999) and Indeje *et al.* (2000) among others have shown that the tropical part of the Pacific Ocean influence the EEA through ENSO teleconnections. ENSO remains the largest coupled ocean–atmosphere phenomenon resulting in climatic variability on interannual time scales (Godínez-Domínguez *et al.*, 2000).

The influence of the Indian Ocean on the interannual variability of East Africa rainfall is now better understood, with a strong relationship being found between the short rains and eastwest pressure and SST gradients across the equatorial Indian Ocean (Goddard and Graham, 1999; Saji *et al.*, 1999; Reason, 2001; Black *et al.*, 2003; Clark *et al.*, 2003; Hastenrath, 2007; Owiti *et al.*, 2008). These gradients reflect a coupled mode of variability which develops in boreal autumn, known as the Indian Ocean Zonal Mode or simply Indian Ocean Dipole (IOD). The difference between mean SST anomalies observed in tropical western Indian Ocean (50°E–70°E, 10°S–10°N) and tropical southeastern Indian Ocean (90°E–110°E, 10°S–Equator) has been used to quantify the zonal temperature gradient representative of the IOD (Saji *et al.*, 1999).

Nyakwada (2009) undertook principal component analysis (PCA) of the SST for each ocean basin separately and for Atlantic–Indian Oceans combined. The modes of variability that were highly correlated with the seasonal rainfall totals over Eastern Africa were identified and used as centre of action of the SST gradients. Although the relationship between the Atlantic Ocean and Eastern Africa rainfall remained not well understood, Nyakwada *et al.* (2009) have recently documented Atlantic–Indian Ocean Dipole index that suggests useful linkage with seasonal rainfall totals over the Eastern Africa region.

Building on earlier results by Mutai *et al.* (1998) which identified SST predictors for the East Africa short rainfall season, Philippon *et al.* (2002) developed a prediction model for the seasonal rainfall totals during this season. However, the prediction model did not have a sufficient lead time that could enable the predictors to be updated before the start of the rainfall season. Jury *et al.* (2009) found that the East African rainfall and zonal winds over the equatorial east Atlantic and West Indian Ocean had an in-phase relationship. A model for October–November–December (OND) seasonal rainfall developed using the central Indian Ocean zonal winds averaged over 3 months (JAS) was found to adequately hit 60% of the target categories but underpredicts the intensity of big events.

However, conspicuously missing are attempts to predict general features of rainfall distribution during the rainy season (e.g., number of wet/dry days, average length of wet/dry spells, frequency of prolonged wet/dry spells, etc.), which will hereafter be referred to intraseasonal statistics of wet and dry spells. Usman and Reason (2004) have already indicated that a season with above average rainfall may not be better than a below average season over an agricultural region if the rainfall are not well distributed in space and time. Crops perform better with evenly distributed 'light' rains than a few isolated 'heavy' rainfall interrupted by prolonged dry periods since consistency of minimum required rainfall as supplied by alternating wet and dry spells is more important than the total rainfall.

Recent studies have shown that some of these intraseasonal statistics of wet and dry spells, especially the number of rain days, are spatially more coherent and possibly more predictable than seasonal rainfall totals/anomalies in various parts of the world (Moron *et al.*, 2006, 2007; Robertson *et al.*, 2009; Gitau *et al.*, 2013). For the EEA short rains however, cross-validated simple regression models developed preliminary using Niño and IOD indices showed quite a modest performance (Gitau *et al.*, 2013). This necessitates the searching of additional potential predictors for the various intraseasonal statistics of wet and dry spells.

This article therefore seeks to improve on the existing knowledge of climate prediction over EEA by considering two main aspects. The first aspect is the increment of the lead time between the time the predictions are made and the start of the short rainfall season. This will enable the users of the climate predictions to consider the options available before the start of the rainfall season depending on the predictions made. The second aspect is the prediction of subregional intraseasonal statistics of the wet and dry spells in addition to the routine prediction of seasonal rainfall totals or anomalies. Although the aim is not to predict individual wet and dry spells, which refers to

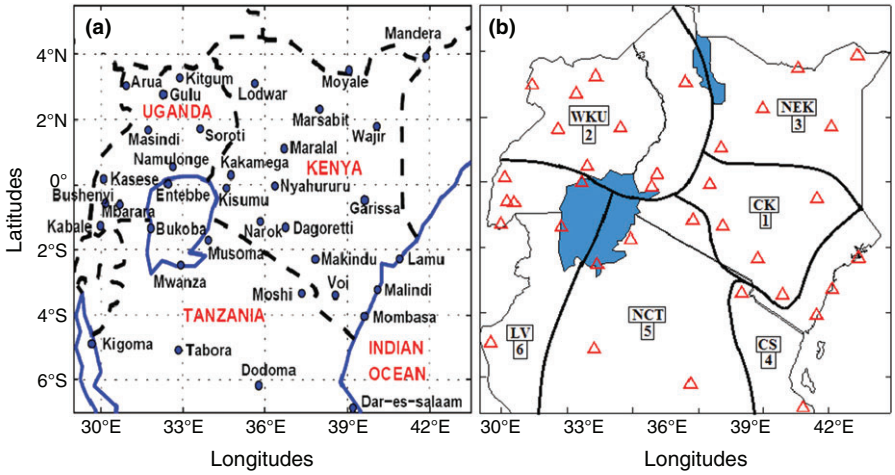

Figure 1. (a) Network of the East African rainfall stations used and (b) the near-homogeneous subregions based on the daily rainfall for the short rainfall season (Source: Gitau *et al.*, 2013.)

time-scales not addressed in this study, information on the likely performance of the rainfall within the season may help the users of climate predictions to make more informed decisions.

Section 2 describes the data used for the study (Section 2.1), the methods of determination of linkages between large-scale climate fields and intraseasonal statistics of wet and dry spells at subregional level (Section 2.2), and finally the development of linear regression models with sufficient lead time for prediction purposes (Section 2.3). Such models can be incorporated into the early warning systems. The results obtained from the various analyses are presented together with the discussions in Section 3. The final section highlights the major conclusions from the study and makes recommendations.

2. Data and methods

2.1. Datasets used

Quality-controlled daily rainfall data from 36 stations spanning for a period of 39 years (1962–2000) over EEA was used. This dataset was obtained from the archives of the Kenya Meteorological Department, IGAD Climate Prediction and Application Centre and Centre de Recherches de Climatologie. Figure 1(a) shows the spatial distribution of the EEA stations used in the study. The southern part of Tanzania (south of 7°S) was excluded from this study since it exhibits rainfall variations that are quite dissimilar to those of the other parts of East Africa (Camberlin and Philippon, 2002). The two rainfall seasons tend to merge together into a single season (unimodal regime) that spans from November to April. A study by Indeje *et al.* (2000) has further showed that central and southern parts of Tanzania have an opposite signal to the rest of East Africa with respect to the rainfall response to the ENSO phenomenon. In this study, only the period of the short rains (October–December) is considered.

Owing to the high skewness of the daily rainfall totals, the data were initially square-root transformed (Bärring, 1988; Stephenson *et al.,* 1999; Camberlin and Okoola, 2003; Gitau *et al.,* 2013) to stabilizing the variance of the sporadic rainfall series resulting in a better fit with a normal distribution. Monte Carlo simulation method was used to determine the number of the principal components (PCs) to be retained and rotated, with 500 simulations done. All the eigenvalues were ranked and the 95th percentile considered as the 95% confidence threshold, to which the actual eigenvalues of the observed dataset were compared. Only six eigenvalues were higher than this threshold (Gitau *et al.*, 2013). Delineation of a near-homogeneous subregion 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). The boundaries between the subregions are just indicative but take into account the main topographical features. The six near-homogeneous subregions delineated during the short rainfall season as shown in Figure 1(b) were adopted (Gitau *et al.,* 2013).

The intraseasonal statistics of wet and dry spells indicated in Table 1 were initially computed at local/station level. The values so obtained were then averaged for all those stations constituting a given near-homogeneous subregion to obtain the subregional intraseasonal statistics of wet and dry spells (SRISS). The threshold of defining a wet day, the definition of a wet/dry spell and other details of this procedure together with its merits over the use of PCA scores can be obtained from Gitau *et al.* (2013). The dry periods before the first and after the last rainfall/wet spells were excluded in the computation of the intraseasonal statistics of dry spells. 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.

On the basis of the literature, a list of predefined predictors of EEA rainfall were defined, and the corresponding

Table 1. List of the seasonal and intraseasonal rainfall variables used in the study.

Statistic	Descriptive name	Definition	Units
SR	Seasonal rainfall	Total amount of rainfall received in a season	mm
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 spell	average duration of consecutive wet days	days
MD	Mean length of dry spell	average duration of consecutive dry days	days
3W	Three or more wet days	Frequency of wet spells of 3 days or more	

indices were extracted which depict El-Niño, IOD and different SST gradients in the tropics. These datasets were all at a monthly temporal resolution, and covered a period of 39 years (1962–2000).

The Niño indices (Niño 1+2, Niño 3, Niño 4 and Niño 3.4) are measures of the oceanic component of ENSO, which indicates the anomalous and sustained SST warming (or cooling) across the central and eastern tropical Pacific Ocean. The Niño indices were downloaded from the Climate Prediction Centre (CPC) of National Oceanic and Atmospheric Administration (NOAA) in the United States.

The Indian Ocean Dipole (IOD) refers to the occasional occurrences of see-saw SST anomalies over the southeastern and western parts of equatorial Indian Ocean. Analysis on the evolutional phases of IOD index by Owiti *et al.* (2008) indicated that the significant SST anomalies begin to appear around April, attain maximum peak around October/November and start decaying in January. Most cycles do not extend beyond 1 year. As such, the significant association between the IOD and Eastern Africa rainfall is strongest during the short (OND) rainfall season while the correlation values are generally not significant during the long (MAM) rainfall season. Comprehensive details of the IOD and its impact on rainfall in the Indian Ocean rim countries including Eastern Africa can be found in Saji *et al.* (1999), Webster *et al.* (1999), Ashok *et al.* (2003), Black *et al.* (2003), Clark *et al.* (2003), Behera *et al.* (2005), Black (2005), Behera *et al.* (2006), Owiti and Ogallo (2007), D'Arrigo and Smerdon (2008), Owiti *et al.* (2008) and Ummenhofer *et al.* (2009) among others. The data for the IOD were downloaded from the Japan Agency for Marine-Earth Science and Technology (JAMSTEC) Website.

Two of the SST gradients defined in Nyakwada (2009) were also used in this study, ZIND and ZPAC. The choice of these two SST gradients only among the many others is discussed later in Section 3.1 ZIND refers to the zonal SST gradient over the Indian Ocean related to the IOD but centred along the equator and which has been shown to have stronger relationships with September to December (SOND) rainfall over East Africa than the classical IOD (Nyakwada, 2009). The difference in the SST anomalies between the Western Pole located at 40°–60°E, 5°N–5°S and the Eastern Pole located at 80°–100°E, 5°N–5°S defines the ZIND SST gradient. ZPAC refers to zonal SST gradient across the equatorial Pacific Ocean, and is defined as the difference in the SST anomalies between the Eastern Pole located at 150°–180°E, 5°N–5°S and the Western Pole located at 120°–90°W, 5°N–5°S. These indices were obtained from the archives of the IGAD Climate Prediction and Application Centre in Kenya.

The Hadley centre Sea Surface Temperatures (HadSSTs) and European Centre for Medium-range Weather Forecast (ECMWF) re-analysis for 40 years (ERA 40) provided the oceanic and atmospheric variables respectively, from which the additional potential predictors were sourced. The HadSSTs used in this study are gridded to a horizontal spatial resolution of 1° latitude by 1° longitude, and covered the region between 45°N and 45°S, but spanned all longitudes (Rayner *et al.*, 2003). Over the Eastern Africa region, SSTs and SST-derived indices have regularly been used for various studies including the seasonal rainfall prediction (Ogallo *et al.*, 1988; Nicholson and Kim, 1997; Mutai *et al.*, 1998; Latif *et al.*, 1999; Indeje *et al.*, 2000; Black *et al.*, 2003; Nyakwada *et al.*, 2009; Diro *et al.*, 2011). This dataset was downloaded from UK Met Office Website.

ERA 40 dataset is gridded at a horizontal spatial resolution of 2.5° latitude by 2.5° longitude (Uppala *et al.*, 2005). Preliminary comparison of NCEP/NCAR and ERA40 re-analyses have shown that the ERA40 accounts for slightly higher variance of the radiosonde data observations for individual stations in equatorial Africa at most standard pressure levels compared to NCEP/NCAR re-analysis (Gitau, 2011). These results are consistent with those obtained in Stendel and Arpe (1997), Engelen *et al.* (1998); Annamalai *et al.* (1999) and Newman *et al.* (2000). From ERA 40, the zonal (u) component of wind vector and the specific humidity (q) were extracted. Zonal wind component was extracted at 925 mb, 700 mb and 200 mb levels representing the lower, middle and upper atmospheric levels while the specific humidity was confined in the former two levels. The ERA 40 dataset had been used over Eastern Africa with satisfactory results in Mukabana and Pielke (1996), and Okoola (1999a, 1999b) among others. The long lead predictions using atmospheric indices pose the question of the physical basis of the relationships. However, it should be recalled that atmospheric variability may reflect land and/or ocean surfaces, both having a relatively longer 'memory.' In such case, the atmospheric predictor can be viewed as a proxy of climate memory associated with these surface conditions (especially land) which cannot always be captured directly by available datasets. Previous studies have demonstrated the utility of these predictors, which also have the potential

to be simulated by General Circulation Models (GCMs). The ERA 40 dataset was downloaded from the ECMWF Website.

2.2. Linkages with large-scale climate fields

The identification and selection of suitable predictors (from July to August period) for each predictand (i.e. each of the intraseasonal statistics) was carried out in three successive steps:

(i) Identification of independent potential predictors within previously defined indices known to affect East African climate that is the SST gradients, Niño and IOD indices;
(ii) Identification of additional potential predictor indices based on total and partial correlations between large-scale oceanic and atmospheric fields and the predictands, and plausible physical/dynamic interpretation between the predictor and predictand in question; and
(iii) Actual predictor indices selection within the pool of potential predictor indices and final prediction model set up based on stepwise linear regression. This is discussed further in the next section.

The indices of Niño, IOD and SST gradients constitute the lists of the predefined predictors. Previous studies have shown significant association of these indices with seasonal rainfall totals during the short rainfall season (Ogallo, 1988; Black *et al.*, 2003; Black, 2005; Owiti and Ogallo, 2007; Owiti *et al.*, 2008, Nyakwada *et al.*, 2009). It has been proposed here that these predefined predictor indices may also have some predictive potential for the subregional intraseasonal statistics of wet and dry spells (Gitau *et al.*, 2013). To verify this, concurrent and lagged simple correlation analysis between the subregional intraseasonal statistics of wet/dry spells and these predefined predictor indices were first determined and the predictor index retained if the coefficient was significant at 95% confidence level.

Additional potential predictors were searched from the oceanic and atmospheric fields. HadSST constitutes the oceanic field while the atmospheric variables considered were the zonal wind component and specific humidity from the ERA 40. A preliminary study showed that the meridional component of wind vector was not associated with any of the subregional intraseasonal statistics of wet/dry spells and seasonal rainfall totals hence not used. Two approaches can be used to search and identify predictor (both oceanic and atmospheric) indices. The first approach involves plotting correlation maps with the predictand and extracting an index over a region showing high correlations. This uses the full resolution of the predictor field. The second approach uses pre-defined possible predictors either as regional indices computed from gridded data, or derived from a PCA. A stepwise procedure is then used to select indices which relate to the predictand. In this study, a modified version of the first approach was used. The oceanic field was initially nested as follows. Grids at 3° by 3°, covering the region (50°W–120°E, 30°S–30°N) were used for the oceans adjacent to Africa while coarser grids at 9° by 9° covered the region (180°W–180°E, 45°S–45°N), excluding the inner region. The rationale behind the nesting was that SST anomalies with large spatial extent at far distance may be expected to influence the East Africa climate just like SST anomalies with small spatial extent at close distance.

The two atmospheric variables were not nested as such. However, the predictor search was confined to region (50°W–120°E, 45°S–45°N). The choice of this region was based on the fact that it includes the subtropical anticyclones which control moisture fluxes towards East Africa. It also enabled the depiction of the wind features which directly affect East African climate such as the Indian Ocean monsoon, the Indian and Atlantic Ocean Walker-type circulation cells, the Tropical Easterly Jet, the Subtropical Westerly Jets among others. It is worthy to mention that there was an assumption that higher latitude (beyond 45°N/S) oceanic and atmospheric systems, at seasonal scale do not significantly influence the rainfall characteristics over the EEA.

The partial correlation between the predictands (OND seasonal rainfall totals as well as subregional intraseasonal statistics of wet/dry spells) and the July–August HadSST, atmospheric variables of zonal wind and specific humidity were then computed while controlling the influence of the predefined predictor indices that were significant at 95% confidence level. The rationale behind the use of partial correlation analysis was that, many large-scale climate fields are influenced by major modes of variability such as ENSO hence full correlation with East Africa rainfall may at times only reflect co-variations induced by the common forcing rather than a physical relationship. The partial correlation approach has been successfully used by Behera *et al.* (2005) in determining the effect of IOD (ENSO) on seasonal rainfall during short rainfall season over Eastern Africa while the effect of ENSO (IOD) is removed.

Partial correlation maps were then produced. It was from these maps that the highly correlated regions were identified and used to compute the new indices. The correlation boxes identified were at least 5° by 5°. This was to ensure that the predictor indices were less noisy, remained stable and under the assumption of memory effects do not vary too fast from the time the forecast is made until the target forecasting period. Mutai *et al.* (1998) have combined the UK Met. Office SST version 4 (MOHSST4) which are initially at 1° by 1° to form a 10° by 10° grid boxes to improve data coverage and reduce noise. Gong *et al.* (2003) have further demonstrated that spatial aggregation increases the skill of seasonal total precipitation forecasts.

The foregoing procedure yielded quite a large number of oceanic and atmospheric predictors. In this study, apart from the use of standard statistical methods, the selection of the additional potential predictors was also based on the physical/dynamical interpretation of the relationship with East Africa rainfall. Only those potential predictors with a plausible physical/dynamical relationship with the

predictands were retained and later used to generate the regression model.

2.3. Development of linear regression models

The MLR approach is a common method in seasonal climate prediction over East Africa region and has provided seasonal forecasts with useful skills (Mutai *et al.*, 1998; Camberlin and Philippon, 2002; Korecha and Barnston, 2007; Nyakwada, 2009). In the forward stepwise MLR approach, each predictor variable was entered into the regression model in an order determined by the strength of their association with the predictand. The effect of adding each predictor was assessed and the predictor retained if it contributed significantly to the variance explained by the model.

Significant intercorrelation between the predictors leads to multi-collinearity which means that the predictors are non-orthogonal. This results into lack of the model's accuracy and may lead to unclear interpretation of the regression coefficients as measures of original effects (Mc Cuen, 1985). It further imposes the problem of redundancy and unnecessary loss of degrees of freedom especially when large numbers of correlated predictors are used (Krishna Kumar *et al.*, 1995). In this study, cross-correlations between potential predictors were carefully checked and only independent variables were used to generate the regression model.

A popular measure of the strength of association in linear regression between the observation and the model output is the coefficient of determination R^2, defined as the proportion of variability in the outcome variable explained by the model. However, a serious problem with this measure is that it can substantially overestimate the strength of association when the number of predictors p, is not small relative to the number of observations n. The adjusted coefficient of determination overcomes this problem (Liao and McGee, 2003). The adjusted coefficient of determination, in the forward stepwise MLR analysis, discourages incorporating additional predictors that will make little marginal changes in the unexplained variance. The adjusted R^2 accounts for the number of the predictors in the model and only increased if the new predictor improves the model more than would be expected by chance. The number of predictors to be retained in the final MLR model was thus determined from the adjusted R^2 of the model. Delsole and Shukla (2002) and Nyakwada (2009) have observed that fewer predictors tend to produce better models than those developed using large numbers of predictors. Details of regression principles can be obtained from Kendall and Stuart (1961), Kendall (1976) and Wilks (2006) among other authors.

The cross-validation method was used to test the performance of the developed MLR models. This method involved temporarily discarding observations from the dataset and then estimating the discarded observations. The estimated values are then compared with the discarded value (Isaaks and Srivastaka, 1989; Barnston *et al.*, 1996; Wilks, 2006; Camberlin *et al.*, 2014). To undertake cross validation, the time series of the training and validation samples are initially standardized separately, for each predictor and for the predictand. The regression model is then built using these anomaly time series by estimating the regression coefficients. The predicted values are finally multiplied by the standard deviation and added to the mean of the predictand time series. In each time, three observations were discarded and then estimated from the developed models. The discarding of one observation at a time was found to be sensitive to the outliers.

The linear error in probability space (LEPS) score developed by Ward and Folland (1991) and later refined by Potts *et al.* (1996) was used to assess the skill of the regression models developed (Jolliffe and Stephenson, 2003; Zhang and Casey, 2000).

A good MLR model requires that the residuals (the difference between the actual observations and the forecasted values) are independent and have a normal distribution (Nayagam *et al.*, 2008). The Durbin–Watson (DW) statistic which checks the significance of the assumption that the residuals for successive observations are uncorrelated/independent was used to determine whether the residuals were independent (Makridakis *et al.*, 1998). One sample Kolmogorov–Smirnov test was used to ascertain that the residuals were normally distributed.

3. Results and discussions

3.1. Linkages with predefined SST predictors

The relationship between the predefined SST predictors averaged for July–August period and the OND seasonal rainfall totals and subregional intraseasonal statistics of wet and dry spells (SRISS) are illustrated by Figure 2(a)–(f). The seasonal rainfall totals (Figure 2(a)) and SRISS of wet spells (Figure 2(b)–(d)) have a positive lagged association with the Niño indices. In terms of the Niño indices, the highest correlations are found with Niño 3.4 and the lowest ones for Niño 1+2. Other significant relationships are generally obtained with SST gradients across the equatorial Indian Ocean (IOD and ZIND), though the correlations are often higher for IOD (Figure 2(a)–(d)). The zonal SST gradient across the equatorial Pacific Ocean (ZPAC) has significant correlation with seasonal rainfall totals and other SRISS, though slightly lower than that of Niño indices (not shown). However, multi-collinearity assessment shows that ZPAC has a highly significant (at 99% confidence levels) negative association ($|r| > 0.7$) with the Niño 1+2, Niño 3 and Niño 3.4 indices hence is not discussed here. Other SST gradients derived by Nyakwada (2009) do not have significant associations with the subregional intraseasonal statistics of wet and dry spells during the short rainfall season and are therefore not included in this study.

The anomalous warm conditions during the boreal autumn over the Niño regions induce variations in the Walker circulation, with anomalous ascending motion over EEA and anomalous descending motion over the Maritime Continent and Southern Africa. The anomalous

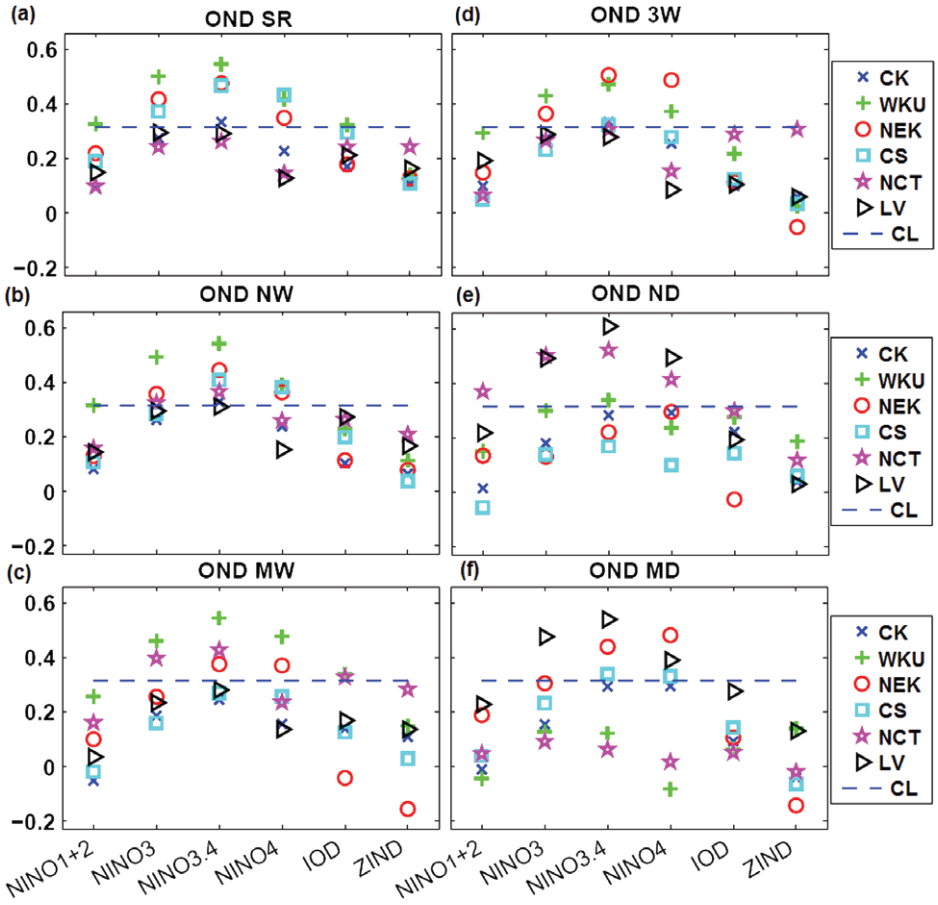

Figure 2. Correlation coefficient between predefined predictors averaged for July–August (*x*-axis) and areal-averaged October–November–December (a) seasonal rainfall totals, (b) number of wet days, (c) mean length of wet spells, (d) frequency of three wet days or more, (e) number of dry days, and (f) mean length of dry spells, over the six rainfall subregions. The correlation coefficient in (e) and (f) has been multiplied by −1 for consistence and ease in comparison. CL shows the 95% confidence level threshold.

ascending (descending) motions tend to bring wet (dry) conditions over EEA (Maritime continent and southern Africa). Variations of the zonal SST gradients across the Indian Ocean (IOD and ZIND) also affect Walker circulation and the associated vertical motion over East Africa. However, their relationship with the SRISS is generally weaker compared to the Niño indices.

The number of wet (NW) days in a season and the seasonal rainfall (SR) totals showed the strongest association with the predefined predictors while the number of dry days (ND) in a season and the mean duration (MD) of the dry spells had the least association (Figure 2). These results are consistent with the spatial coherence and potential predictability assessment which showed the former two to be spatially more coherent and thereby more potentially predictable than the later two (Gitau *et al.*, 2013). It is shown that the greater spatial coherence is a result of a more or less uniform response to the global ENSO signals.

Subregion 2 (Figure 1(b)) which covers western Kenya and most parts of Uganda (as represented by Kakamega, Kisumu and Lodwar; Arua, Kitgum, Masindi, Namulonge and Soroti stations, respectively, and abbreviated by WKU), displays the strongest and statistically significant (at 95% confidence levels) lagged correlation coefficients between the subregional seasonal rainfall totals and the intraseasonal statistics of wet spells on one hand and the Niño indices on the other hand (Figure 2(a)–(d)). This is consistent with Ntale and Gan (2004) who have shown that despite southern Uganda and the Lake Victoria Basin receiving rainfall almost throughout the year due to the influence of Lake Victoria (LV), the area has reasonable El Niño influence.

Over the highland areas such as subregion 1 in central Kenya (CK); parts of the subregion 5 in northern and central Tanzania (CT); and parts of subregion 6 (LV area) over southwestern Uganda, there seems to be a weaker response to the ENSO phenomenon since these areas are subjected to stronger local orographic influence which tends to curtail the ENSO signal (Ntale and Gan, 2004).

The association of the subregional intraseasonal statistics of dry spells with the predefined predictors is rather diverse (Figure 2(e) and (f)). In many cases, the correlations are low and insignificant, but there are exceptions. Subregion 6 in southern Uganda and western Tanzania (as represented by Bushenyi, Entebbe, Kabale, Kasese and Mbarara; and Bukoba and Kigoma stations, respectively) has strong lagged correlation coefficients between the SRISS of dry spells and the predefined SST predictors.

Table 2. Brief description of the additional potential predictors for the short rainfall (OND) season and their location details.

Index name	Description	Location details (°)	
		Longitude	Latitude
BoBEN	SST index over Bay of Bengal extending to west coast of Malaysia and Indonesia	83–90 E	12–17 N
SEHAW	SST index on the southeast of Hawaii in the Pacific Ocean	140–120 W	10–25 N
SWAFRC	Specific humidity index at 700 mb level located at Angola coast on south-western Africa and extending to Atlantic Ocean on the west and Zambia to the east	5–15 E	25–15 S
SINDS	Zonal wind index at 925 mb level to the south of the Bay of Bengal near the southern tip of India subcontinent	70–90 E	5–10 N

For instance, a correlation of −0.6 is found between the Niño 3.4 index on one hand and the number of dry days and the mean length of the dry spells in a season on the other hand, suggesting the increase in the number and length of the dry spells during La Niña years (Figure 2(e) and (f)).

From the strong significant correlations with the predefined predictors, two independent indices with strong significant lagged correlations with the seasonal rainfall totals and SRISS are chosen. These are the Niño 3.4 and ZIND indices whose average values for July–August are not related ($r = 0.16$), yet they are associated with seasonal rainfall totals and most of the SRISS (Figure 2(a)–(f)). They can be thought of as representing the pre-season (July–August) SST conditions associated with the ENSO and IOD climate signals.

3.2. Linkages with additional potential predictors

Concurrent and lagged partial correlation analysis between the seasonal rainfall totals and SRISS on one hand and the oceanic and atmospheric variables on the other hand while controlling the effects of significantly correlated predefined indices (Niño 3.4 and ZIND) identified several common potential predictors. Only four potential predictors have a plausible physical/dynamical relationship with the predictands and are briefly described in Table 2.

The association of the four additional (oceanic and atmospheric) predictors with the seasonal rainfall totals and SRISS during the short rainfall season are summarized by Figure 3(a)–(f). Consistent with the predefined predictors, the seasonal rainfall totals and the SRISS of wet spells are more coherent in their responses to these predictors (Figure 3(a)–(d)). All the SRISS of the wet spells respond more or less uniformly to any predictor identified for all the subregions. The insignificant relationship found for some of the coefficients is attributed to the fact that for the sake of simplicity, total correlations are shown in Figure 3(a)–(f), while the identification and selection of the additional potential predictors is based on partial correlation analysis, after the effect of significantly correlated predefined predictors have been removed.

The SRISS of dry spells are somehow diverged in their responses to the additional potential predictors identified and often have insignificant association (at 95% confidence level) as shown in Figure 3(e)–(f). However, there are several exceptions. One such example is the atmospheric predictor SINDS that has a generally consistent response with the SRISS of dry spells (Figure 3(e)–(f)). These show that the response of the intraseasonal statistics of dry spells may not be uniform for any given oceanic or atmospheric signals.

In the subsequent paragraphs, these four additional potential predictors are described in details and a physical interpretation on how the predictor influenced the SRISS for which it is significantly correlated provided.

BoBEN, the first additional potential predictor, is an SST index averaged over July–August period over the Bay of Bengal extending to west coast of Malaysia and Indonesia (Table 2). The area coverage with significant correlation coefficient reduces as the OND season approaches and is confined to the eastern Indian Ocean during the OND period. Correlation analysis with the global SST shows that this index has no signal over the tropical Pacific Ocean, which is further confirmed by the insignificant correlation coefficients between the Niño indices and this index (Table 3).

The computation of the correlation coefficients between the changes in the magnitude of the Western and the Eastern Poles of the IOD from July to August and OND periods on one side and the BoBEN index on the other side shows that, while there is no significant correlation with the Eastern Pole, the BoBEN index accounts for 28% of the changes in the magnitude between the July–August and OND periods over the Western Pole of the IOD. Thus warm conditions in the northeastern Indian Ocean, as portrayed by BoBEN index are expected to result in the strengthening of the Indian Ocean Walker circulation cell in boreal autumn, with anomalous ascending motion in the east and descending motion in the west. A SST warming over the index location is likely to reinforce the circulation anomalies associated with the negative phase of the IOD/ZIND that is characterized by warm (cold) conditions in the eastern (western) Indian Ocean. BoBEN index is located slightly to the north of the Eastern Pole of the IOD/ZIND, and it actually displays a significant correlation (though weakly) with the ZIND index (Table 3) during the OND period. The justification for retaining this and other additional potential predictors that have significant relationship with the predefined indices shall be discussed towards the end of this section.

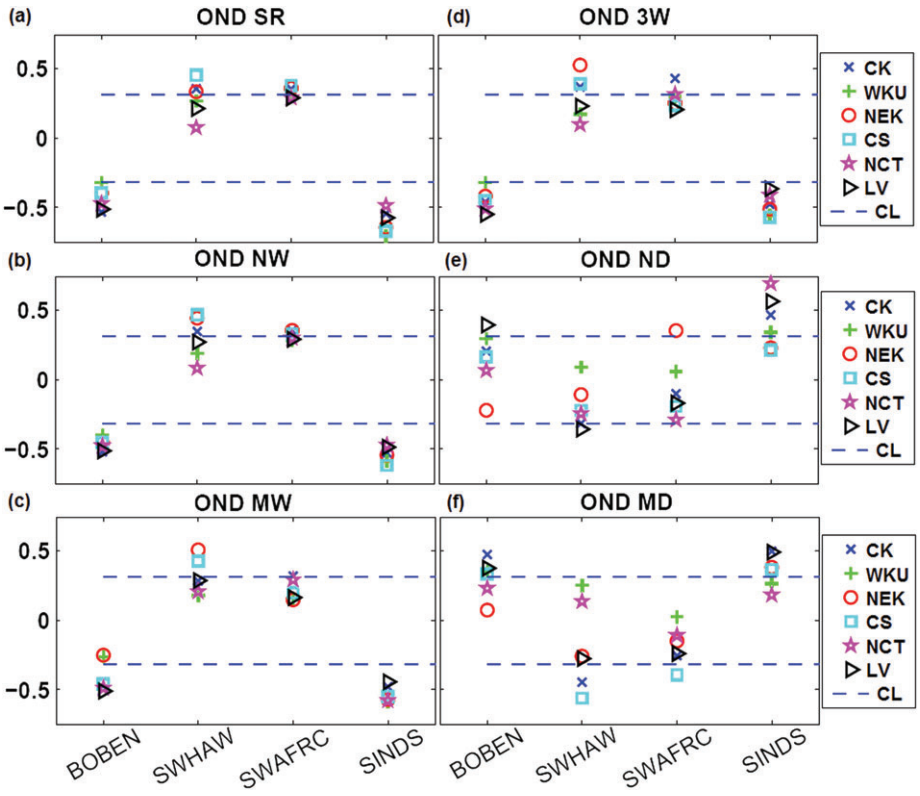

Figure 3. Correlation coefficient between additional potential predictors averaged for July–August (*x*-axis) and areal-averaged October–November–December (a) seasonal rainfall totals, (b) number of wet days, (c) mean length of wet spells, (d) frequency of three wet days or more, (e) number of dry days, and (f) mean length of dry spells, over the six rainfall subregions. CL shows the 95% confidence level threshold.

Table 3. Correlation coefficients between the four additional potential predictors averaged over July–August (Jul–Aug) and October–December (OND) period and some predefined predictors averaged over the OND period.

		Niño 1+2	Niño 3	Niño 4	Niño 3.4	IOD	ZIND
BoBEN	Jul–Aug	0.28	−0.02	0.02	−0.16	−0.12	−0.27
	OND	0.03	−0.15	−0.18	−0.23	−0.21	***−0.39***
SEHAW	Jul–Aug	0.14	0.09	***0.62***	0.28	0.05	−0.10
	OND	***0.36***	0.24	***0.49***	0.28	0.19	−0.03
SWAFRC	Jul–Aug	0.06	0.07	0.07	0.08	0.24	0.23
	OND	0.20	0.22	0.01	0.14	***0.35***	***0.36***
SINDS	Jul–Aug	***−0.57***	***−0.66***	***−0.61***	***−0.70***	***−0.45***	−0.07
	OND	***−0.72***	***−0.69***	***−0.60***	***−0.69***	***−0.61***	***−0.32***

The numbers in bold italics indicate that the correlation coefficients are significant at 95% confidence levels.

BoBEN has significant negative association over most of the subregions with the seasonal rainfall totals and all SRISS of the wet spells (Figure 3(a)–(d)). Over southern Uganda, northwestern and western Tanzania (LV, subregion 6), it has significant positive association with all the SRISS of the dry spells. It also has significant positive relationship with the mean duration of dry spells over the western parts of the study domain (WKU and LV), CK and southeastern lowlands as well as the coastal strip (CS) of Eastern Africa, as shown in Figure 3. The strengthening of the Indian Ocean Walker circulation cell results in the reduction of seasonal rainfall totals, NW days and the mean frequency of the wet spells of 3 days or more over Eastern Africa. The mean duration of the dry (wet) spells is also increased (decreased).

The independence of BoBEN from ENSO, coupled with the fact that it shows significant relationship with different atmospheric variables (zonal wind component and specific humidity) over East Africa during the OND season (not shown), justified its retention as a potential predictor.

SEHAW, the next additional potential predictor is an SST index averaged over July–August period and located southeast of Hawaii in the Pacific Ocean (Table 2). It is clearly distinct from the core ENSO region and has strong persistence from July–August through to October–December. This index has a significant positive signal with SST over the central equatorial Indian Ocean during the OND period and with zonal winds at 925 mb level over the Indian Ocean closer to the East Africa coast that appeared in July–August, and grew in September

through to December (not shown). This index may be related to the fifth rotated empirical orthogonal factor (REOF5) earlier identified by Mutai *et al.* (1998). This REOF had strong weights in Pacific and Indian Oceans, but was independent of the SOI. A large-scale signal in Africa rainfall and near-surface marine divergence over the Indian Ocean was related to this REOF. It therefore suggests that some of the East Africa rainfall variability that is independent of the SOI could not also be attributed to local-scale chaotic features (Mutai *et al.,* 1998).

Over the CS of Kenya and Tanzania and northeastern Kenya (NEK), SEHAW has significant positive association with seasonal rainfall totals and all the SRISS of wet spells. The seasonal rainfall totals, NW days and mean frequency of wet spells of 3 days or more over CK and southeastern lowlands of Kenya also have significant positive association with SEHAW. Significant negative associations were observed for number of dry days over southern Uganda, northwest and western Tanzania (LV); and the mean duration of the dry spells over CK and southeastern lowlands of Kenya and CS of Kenya and Tanzania as shown in Figure 3.

Results of correlation analysis with the predefined predictor indices showed that SEHAW is significantly correlated with Niño 4 from July–August to October–December and with Niño 1.2 during the OND season only (Table 3). It is thus expected that this index depicts SST conditions which are associated with some ENSO events, and which results into a subsequent warming of the Indian Ocean in the northern autumn (Cadet, 1985). Its independence from the Indian Ocean zonal SST gradient indices (Table 3) and significant relationship with several atmospheric variables around East Africa during the OND season justifies the retention of SEHAW index as an additional potential predictor.

SWAFRC, the third additional potential predictor is a specific humidity index at 700 mb level over the Angola coast on southwestern Africa and extending to Atlantic Ocean on the west (at 5°E) and Zambia to the east (Table 2) averaged over July–August period. SWAFRC has a significant positive relationship with the specific humidity at 700 mb level over Arabian Sea, Red sea, most parts of northern Africa, equatorial Atlantic Ocean and southern Indian Ocean around latitude 30°S during July–August period. In September, the signal weakens and seems a bit noisier. During the OND period, the index has well-defined signal over central and Eastern Africa and equatorial Indian Ocean extending to southern Indian Ocean. Mutai *et al.* (1998) identified a weak positive (though not statistically significant) relationship between southeastern Atlantic SST anomalies and East African short rains. The modest effect of the southeastern Atlantic on East Africa rainfall was through influencing the strength and moisture content of the airmass that flows into Equatorial Africa from the Atlantic Ocean (Mutai *et al.*, 1998).

Enhanced low- to mid-tropospheric moisture over these areas, when advected to East Africa may result in wet conditions. Concurrent and lagged correlation analysis with the global SSTs did not show any persistent signal over the three global oceans. This is further confirmed by the weak correlation coefficients between SWAFRC on one hand and Nino and Indian Ocean indices (Table 3), which emphasizes the strength of SWAFRC as an independent potential predictor of East African rainfall.

Over the eastern sector of the study region (CK, NEK and CS as shown in Figure 1(b)), SWAFRC has a significant positive association with seasonal rainfall totals and NW days (Figure 3(a) and (b)). The mean duration of the wet spells and mean frequency of wet spells of 3 days or more over CK and southeastern lowlands of Kenya (CK) have significant positive association with SWAFRC. A significant positive correlation with number of dry days over NEK, an arid and semi-arid area is also noted as shown in Figure 3(e).

The last additional potential predictor, SINDS wind index, is identified when the partial correlation analysis is undertaken between the July and August zonal component of the low level wind field (925 mb level) and NW days during the OND season while controlling the effects of July–August predefined predictors (Niño 3.4 and ZIND) and one of the additional SST potential predictor (SEHAW). Significant negative correlations are found over the southern tip of Indian subcontinent; on the border of Europe, northern Africa and northern Atlantic Ocean; and on southern Atlantic around 20°S on the South America coast (Figure 4(a)). It is from the significant negative correlation area around the southern tip of the Indian subcontinent that the SINDS index is extracted (Figure 4(b)).

The July–August values of the 925 mb SINDS zonal wind index has a significant negative association with zonal wind at 925 mb over East Africa and extending into western Africa coast and Gulf of Guinea from July–August through to October–December (Figure 5(a)–(c)). Though the local significant correlation with the zonal wind over northern Indian Ocean seems to die out after September (Figure 5(b)), Figure 6(a)–(c) suggests that the enhanced low-level monsoon winds as portrayed by SINDS index modulates the SST by cooling around the index location initially in July–August. The modulation spread to northern, central and western parts of Indian Ocean closer to the Western Pole of the IOD/ZIND from September through to December (Figure 6(b) and (c)). The cooling of the SST around the Western Pole of the IOD/ZIND is associated with anomalous subsidence and drier conditions over Eastern Africa. Correlation analysis with the global SST further shows that this signal is significantly but negatively correlated with SST over the Niño regions from July–August through to October–December (Figure 6(a)–(c)) and over much of the northern and western Indian Ocean during the October–December period (Figure 6(c)). This is further confirmed by the strong negative significant correlation coefficients obtained with the Niño and Indian Ocean indices (Table 3).

Similar to BoBEN index, further computation of the correlation coefficients between the changes in the magnitude of the Western and the Eastern Poles of the IOD from July to August and OND periods on one side and the SINDS index on the other side shows that, while there is no

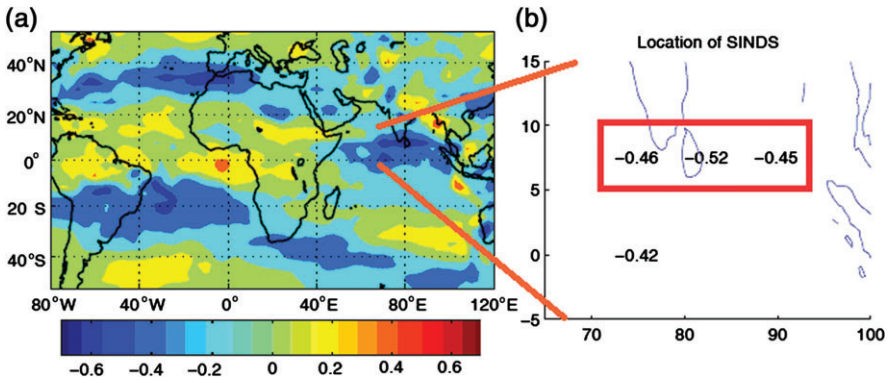

Figure 4. (a) A map of partial correlation coefficients between July and August u-wind at 925 mb and number of wet days during OND over subregion 4 while controlling Niño 3.4, ZIND and SEHAW indices. (b) Location of the significant partial correlation coefficients over the southern tip of the Indian subcontinent. The rectangle shows the location from which SINDS index was extracted.

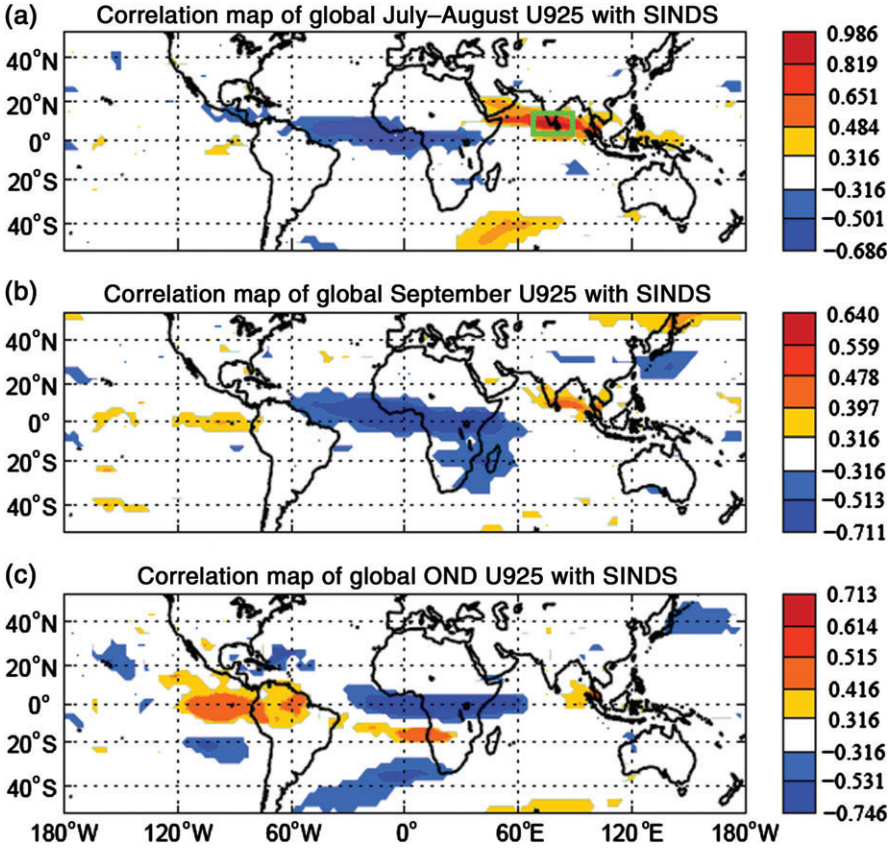

Figure 5. Map of significant correlation between southern tip of Indian subcontinent (SINDS) zonal wind index and global U925 for (a) July–August, (b) September and (c) October–December. The rectangle in (a) shows the approximate location of SINDS zonal wind index computed for July–August period from 1962 to 2000.

significant correlation with the Eastern Pole, the SINDS index accounts for 13% of the changes in the magnitude between the July–August and OND periods over the Western Pole. This is again confirmed by the increase in both the spatial coverage and the magnitude of the significant correlation areas over western Indian Ocean (Figure 6(a)–(c)). This simply shows that the SINDS index is a one of the precursor for the rapid warming/cooling of the Western Pole of the IOD. Previous studies (Saji *et al.,* 1999; Webster *et al.,* 1999; Black *et al.,* 2003; Clark *et al.,* 2003) have already indicated that while the Eastern Pole of the IOD develops quite early in the year, the Western Pole evolves quite rapidly.

The drier conditions result in the reduction in the magnitude of seasonal rainfall totals and SRISS of the wet spells as well as an increase in the magnitude of the SRISS of dry spells. SINDS has significant association with the rainfall totals and all the SRISS of the wet and dry spells. With the seasonal rainfall totals and SRISS of wet spells, the whole study area has significant negative association with SINDS (Figure 3(a)–(d)). Over the whole study area, SINDS has significant positive association with number of

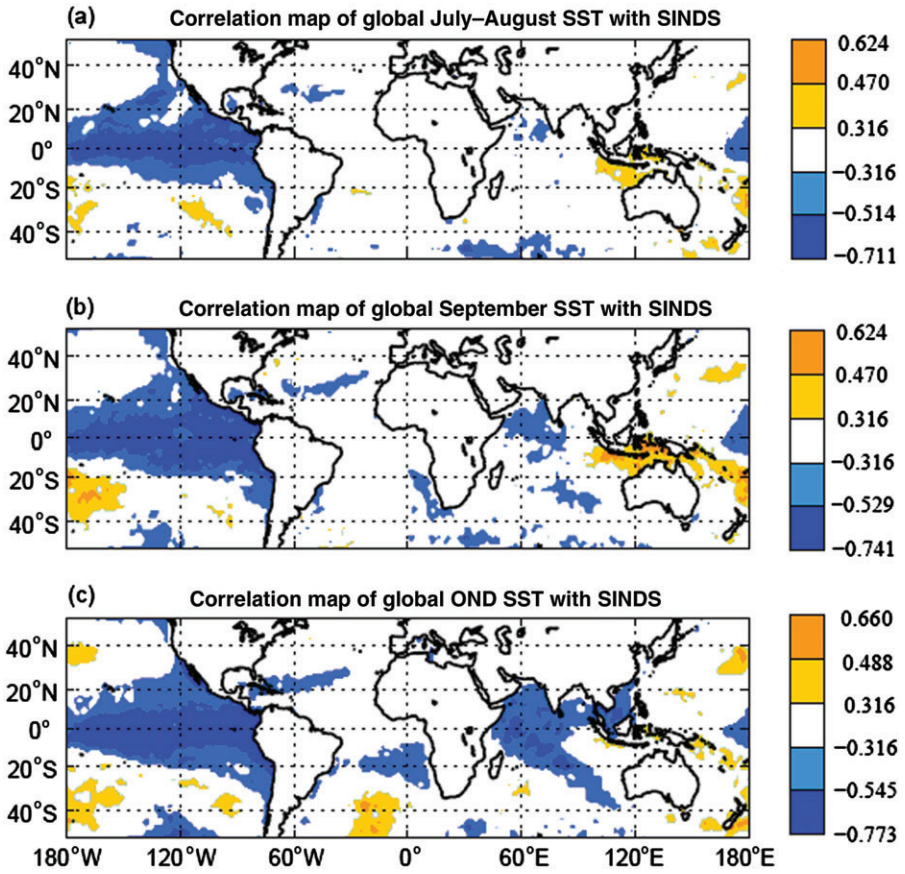

Figure 6. Map of significant correlation between southern tip of Indian subcontinent (SINDS) zonal wind index and global SST for (a) July–August, (b) September, and (c) October–December.

dry days in a season (Figure 3(e)) except over NEK and CS of Kenya and Tanzania ; and mean duration of dry spells in a season (Figure 3(f)) except over western Kenya and most parts of Uganda (WKU) and Northern and Central Tanzania (NCT).

In their study on the prediction of the East African OND rains, Philippon *et al.* (2002) also found that an atmospheric index, taken in September and describing the Indian monsoon intensity had some predictive skill, in addition to more traditional SST predictors. The present study further demonstrates a partly independent predictive skill from Asiatic monsoon dynamics as early as in July–August period.

The justification for the retention of SINDS despite the strong association with SST over Niño regions and the other predefined predictors (Table 3) is discussed here. The total and partial correlation coefficients between the SINDS wind index and NW days over the six subregions while controlling the effects of other predictors averaged for July–August period are shown in Table 4.

First, the NW days have significant negative (positive) total correlation coefficient with SINDS (Niño 3.4) over the six subregions and significant positive total correlation coefficient with ZIND (IOD) over two (four) subregions only. Moreover, the magnitudes of the correlation coefficients with SINDS are higher than those of Niño 3.4, ZIND and IOD in all the six subregions (Table 4). Secondly, significant negative partial correlation coefficients between NW days and SINDS while controlling the effect of ZIND and IOD individually are also obtained over the entire region (Table 4). Significant negative partial correlation coefficients are also obtained, however for two subregions only, when Niño 3.4 predictor is controlled. Controlling the combined effects of Niño 3.4, ZIND and SEHAW predictors, significant negative partial correlation coefficients between the NW days and SINDS wind index are obtained over the same two subregions [namely CK and the East African coast (CS)]. This therefore means that despite the strongly significant total correlation coefficient between SINDS wind index and the predefined predictors (Table 3), the SINDS wind index provides additional predictive information on the NW days that could not be captured by the Niño 3.4, ZIND and SEHAW predictors in two out of the six subregions (Table 4). The case where the Niño 3.4, IOD and SEHAW are controlled should be treated with caution since Niño 3.4 and IOD are not independent predictor indices. Similar remarks applied to the other additional potential predictor indices which have significant correlation with the predefined predictors.

Therefore, the plausible physical explanation on how SINDS relates to East Africa rainfall and the fact that it provides additive predictive information despite its strong association with the Niño and Indian Ocean indices provides a strong case for the retention of this index as

Table 4. Total and partial correlation coefficients between areal-averaged number of wet days (NW) in subregional zones 1 to 6 and southern tip of India subcontinent (SINDS) 925 mb zonal wind index, while controlling predefined predictors and one additional potential predictor (SEHAW) for July–August period.

	Total correlation with NW				Partial correlation between NW & SINDS while controlling				
	SINDS	Niño 3.4	ZIND	IOD	Niño 3.4	ZIND	IOD	Niño 3.4, SEHAW &	
								ZIND	IOD
CK	**−0.53**	**0.41**	0.20	0.29	**−0.37**	**−0.52**	**−0.46**	**−0.32**	−0.29
WKU	**−0.59**	**0.59**	0.24	**0.39**	−0.30	**−0.59**	**−0.50**	−0.31	−0.29
NEK	**−0.54**	**0.52**	0.20	0.30	−0.29	**−0.54**	**−0.48**	−0.21	−0.19
CS	**−0.62**	**0.48**	0.22	**0.38**	**−0.45**	**−0.62**	**−0.54**	**−0.38**	**−0.34**
NCT	**−0.47**	**0.42**	**0.32**	**0.40**	−0.27	**−0.47**	**−0.35**	−0.31	−0.26
LV	**−0.49**	**0.45**	**0.34**	**0.44**	−0.27	**−0.49**	**−0.36**	−0.25	−0.19

Bold numbers indicate that the coefficient is significant at 95% confidence level.

an additional potential predictor during the short rainfall season.

Though BoBEN and SINDS indices appear to be located closer to each other, the two indices are independent since correlation coefficient between them was found to be 0.05 which is insignificant at 95% confidence level. This means that the two indices provide independent predictive information. These two additional predictor indices could be associated with the Asian monsoon. The Asian monsoon is connected to East Africa via the Somali Jet. Weaker winds may result in warmer SST in the Western Indian Ocean at the end of the monsoon season, which subsequently increases convection and affect the Indian Ocean Walker circulation. Interestingly, the two predictor indices combined accounts for 39.5% of the SST changes in the magnitudes between the July–August and OND periods over the Western Pole of the Indian Ocean Dipole (IOD).

3.3. Regression models for subregional intraseasonal statistics of wet and dry spells

The methodology used to develop the MLR models at subregional level and assess the performance of these models was discussed in Section 2.3.

3.3.1. Seasonal rainfall totals

Table 5 illustrates on how the final number of predictors to be retained is determined for the seasonal rainfall totals over subregion 1 (CK and southeastern lowlands of Kenya) based on the R-adjusted consideration with the forward stepwise regression technique. The predictor that is most strongly associated with seasonal rainfall totals over subregion 1 is SINDS with a correlation coefficient of 0.54 and adjusted correlation coefficient of 0.53. In the cross-validated mode, this predictor has a correlation coefficient of 0.47. In the second step, predictor BoBEN is picked. The two predictors have a multiple correlation coefficient of 0.74 with seasonal rainfall totals while the adjusted correlation coefficient was 0.73. In the cross-validated mode, the two predictors have a multiple correlation coefficient of 0.69 with seasonal rainfall totals. In the third step SWAFRC is picked, the fourth step gives SEHAW.

Table 5. Forward stepwise fitting of the multivariate regression model for OND areal-averaged seasonal rainfall totals over subregion 1 (CK).

Step	Predictor included	Multiple correlation coefficient		
		R	Adjusted R	R_{cv}
1	SINDS	0.54	0.53	0.47
2	BoBEN	0.74	0.73	0.69
3	SWAFRC	0.79	0.77	0.73
4	SEHAW	0.81	0.78	0.75

A close look at this table shows that the multiple correlation coefficient for the developed MLR model and its adjusted correlation coefficient as well as multiple correlation coefficient for the cross-validated model has been increasing at each step. The four additional potential predictors can thus be used to develop the MLR model for seasonal rainfall totals over this particular subregion. However, multi-collinearity assessment further shows that SINDS and SEHAW are significantly inverse correlated ($r = -0.41$) at 95% confidence level. Thus, only one out of these two predictors should be used to avoid the inflation of the variance and loss of degrees of freedom (Krishna Kumar *et al.*, 1995). Hence the regression model developed for the seasonal rainfall totals over subregion 1 is based on the first three predictors shown in Table 5. The predictors to be retained for other subregions and the SRISS are similarly obtained.

Table 6 summarizes the predictors used in the developed and cross-validated MLR models and the skill score of the models for the seasonal rainfall totals. The developed models capture the direction of the observation quite well though at times the magnitudes are not attained. From a list of four additional potential predictors and two predefined predictors, two sets of predictors' combinations are adequate to describe the interannual variability of the seasonal rainfall totals over the six subregions (Table 6). The atmospheric predictor SINDS and oceanic predictor BoBEN are common to all the MLR models. It should be observed from Table 6 that none of the models picks the Niño 3.4 index as a predictor while ZIND index is only picked once.

Table 6. The list of predictors' combination and skill of regression models for areal-averaged seasonal rainfall totals during the short rainfall season.

Subregion	Predictors	Multiple correlation coefficient between observed and predicted		LEPS score (%)	Durbin–Watson statistic
		R	R_{cv}		
CK	BoBEN, SWAFRC, SINDS	0.79	0.73	43	1.40
WKU	BoBEN, SWAFRC, SINDS	0.80	0.75	42	2.11
NEK	BoBEN, SWAFRC, SINDS	0.79	0.63	40	1.89
CS	BoBEN, SWAFRC, SINDS	0.82	0.70	45	1.65
NCT	BoBEN, SINDS, ZIND	0.69	0.62	32	2.40
LV	BoBEN, SWAFRC, SINDS	0.78	0.68	44	1.26
Mean value		0.78	0.69	41	

Table 7. The list of predictors' combination and skill of regression models for areal-averaged number of wet days during the short rainfall season

Subregion	Predictors	Multiple correlation coefficient between observed and predicted		LEPS score (%)	Durbin–Watson statistic
		R	R_{cv}		
CK	BoBEN, SWAFRC, SINDS	0.77	0.70	42	1.43
WKU	BoBEN, SINDS	0.69	0.63	33	1.86
NEK	BoBEN, SWAFRC, SINDS	0.75	0.65	34	1.50
CS	BoBEN, SWAFRC, SINDS	0.79	0.70	43	1.55
NCT	BoBEN, SWAFRC, SINDS	0.69	0.63	35	1.84
LV	BoBEN, SINDS	0.69	0.60	35	2.04
Mean value		0.73	0.65	37	

This does not mean that Niño 3.4 index (a representative of the ENSO indices) is not related to Equatorial Eastern Africa seasonal rainfall totals, but rather the predictive signal in ENSO is contained in the other predictors from the Indian Ocean region.

On average, the Linear Error in Probability Space (LEPS) skill score of 41% is attained for the six models (Table 6). Since the LEPS skill score value for all the regression models are positive, it means that the models output (forecast) are much better than climatology. Results further show that the residuals from all the models are normally distributed. Over western Kenya and most parts of Uganda (WKU) and NCT, the model residuals have negative autocorrelation (the value was greater than 2) while the rest of the study area has positive autocorrelation (the value was less than 2). Over the northern sector of the study, the model residuals are independent of each other (Farebrother, 1980). Over the rest of Kenya, northern, eastern and central Tanzania (CK, CS and NCT), the test is inconclusive while the residuals over the southern Uganda and western Tanzania (LV) has positive first-order autocorrelation. This means that the regression models can be improved further by adding an autoregressive term.

In the subsequent sections, only those subregions for which the multiple correlation coefficient between the time series of observed and cross-validated MLR model output is equal or greater than 0.5 are discussed since only such models can be incorporated in operational forecasting (Philippon *et al.,* 2009).

3.3.2. Number of wet and dry days in the season

From a list of three predictors, two combinations of predictors are adequate to describe the interannual variability of the NW days over the six subregions (Table 7). Figure 7(a)–(f) show the time series plots for the developed and cross-validated MLR models and the actual observations for the NW days in a season. The figures show that the models developed capture the peaks quite well but not so well for the lows. Several peaks associated with the El Niño events are well captured by the models over all the subregions. These include the El Niño events of 1968, 1972, 1982, 1994 and that of 1997. Equally, the low seasons associated with the La Niña events of 1970 and 1998 are also well captured by most of the developed models.

The multiple correlation coefficient between the cross-validated MLR model outputs and the actual observations of the NW days in a season for the six subregions range from 0.60 to 0.70, with an average of 0.65 (Table 7). According to the LEPS skill score, an average value of 37% is obtained for the six cross-validated MLR models. It should be observed that the skill of the multiple correlation coefficient and the LEPS skill score for the cross-validated MLR models for the NW days are comparable to those obtained for the seasonal rainfall totals (Table 6) though slightly lower. This is consistent with the spatial coherence results which showed that the two were almost equally potentially predictable (Gitau *et al.*, 2013).

The residuals from the six cross-validated MLR models are normally distributed. Compared to the tabulated critical values, the model residuals over western block of the

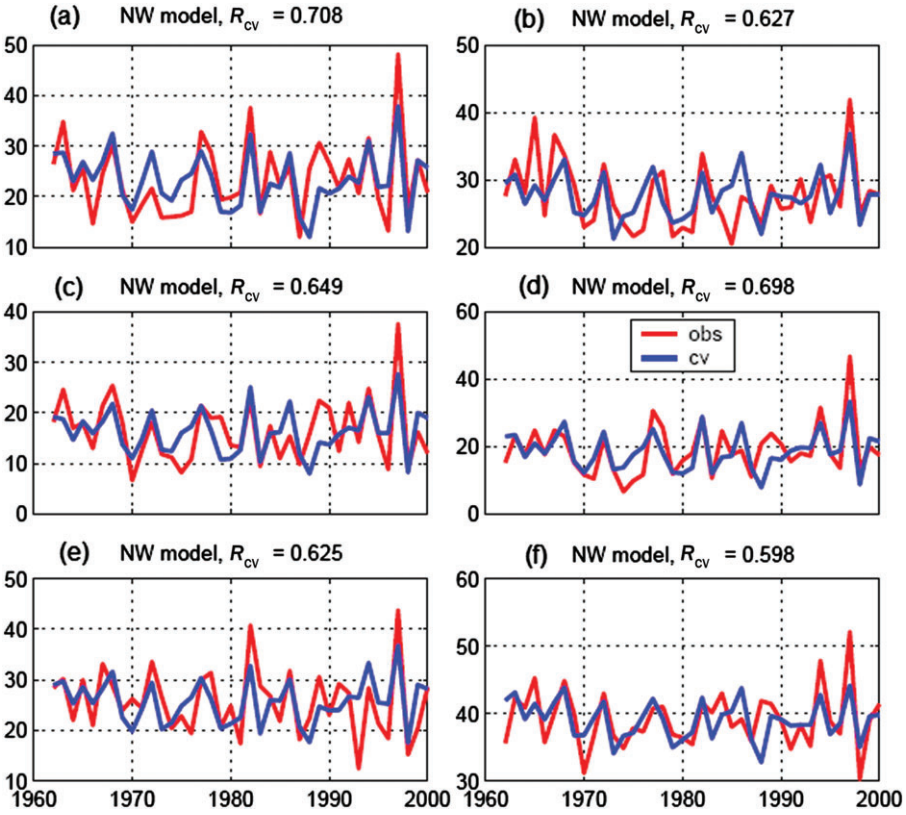

Figure 7. Time series plot of the observed (obs) and cross-validated model (cv) estimates for October–November–December areal-averaged number of wet days (NW) over (a) Central highlands and southeastern lowlands of Kenya – CK, (b) Western Kenya and most parts of Uganda – WKU, (c) northeastern Kenya – NEK, (d) Coastal strip of Kenya and Tanzania – CS, (e) Northern and Central Tanzania – NCT, and (f) Western of Lake Victoria and western Tanzania – LV. R_{cv} shows the multiple correlation coefficient for cross-validated model.

Table 8. The list of predictors' combination and skill of regression models for areal-averaged number of dry days during the short rainfall season

Subregion	Predictors	Multiple correlation coefficient between observed and predicted		LEPS score (%)	Durbin–Watson statistic
		R	R_{cv}		
NCT	SWAFRC, SINDS	0.74	0.69	30	1.79
LV	BoBEN, SEHAW, Niño 3.4	0.75	0.68	40	2.17

study area (WKU, NCT and LV) are not significantly autocorrelated. For the eastern block of the study area (CK, NEK and CS), the test for significant autocorrelation is inconclusive.

In the case of number of dry days in a season, only two subregions [inland Tanzania and southern Uganda (NCT and LV)] are shown since the rest had a multiple correlation coefficient of less than 0.5 between the observations and the model outputs from the cross-validated MLR model (Table 8). We observe that much lower prediction skills is found for the number of dry days compared to that obtained for the NW days. This is due to the fact both variables depend not only the intraseasonal distribution of the rainfall but also on the length of the rainy season. A longer rainy season generally experiences both a greater absolute and relative frequency of wet days. By contrast a longer rainy season tends to be associated with a lesser relative frequency of dry days; hence an increase in the length of the season has an inverse (mechanical) effect to potentially increase the absolute number of dry days.

The good skill of the models for the number of dry days over NCT and LV (Table 8) indicate that they can complement the models for the NW days (Table 7) that have the lowest values over the same subregions, especially if the onset and cessation days have been determined earlier. An assessment of the cross-validated MLR models and the residual analysis clearly indicate that the models over these two subregions are quite robust and can be incorporated for operational uses.

3.3.3. Mean duration of wet and dry spells for the season

The multiple correlation coefficient for the cross-validated models for the mean duration of wet spells are high (Table 9), though slightly lower than those of the seasonal

Table 9. The list of predictors' combination and skill of regression models for areal-averaged duration of wet spells during the short rainfall season

	Predictors	Multiple correlation coefficient between observed and predicted		LEPS score (%)	Durbin–Watson statistic
		R	*R*cv		
CK	BoBEN, SWAFRC, SINDS,	0.72	0.67	34	1.75
WKU	BoBEN, SINDS	0.63	0.57	30	1.93
CS	BoBEN, SINDS	0.70	0.59	35	1.42
NCT	BoBEN, SINDS, ZIND	0.78	0.71	51	2.25
LV	BoBEN, SINDS	0.66	0.58	32	1.87

rainfall totals (Table 6) and ranges between 0.57 and 0.71. Most of the additional potential predictors picked by these models are similar to the ones retained for seasonal rainfall totals (Table 6) and NW days in a season (Table 7). The LEPS skill score ranges from 30% to 51% for the five MLR models. The lowest multiple correlation coefficients and the lowest LEPS skill scores are obtained over the western sector of the study area (WKU and LV), which is closely followed by the CS of Kenya and Tanzania. The model residuals from the five subregions were normally distributed. In four of these models, the residuals were independent of each other while over the CS of Kenya and Tanzania, the test was inconclusive.

Only three subregions have their cross-validated MLR models achieve multiple correlation coefficients of equal or more than 0.5 between the time series of cross-validated model and the actual observations of mean duration of dry spells. The lowest correlation (0.54) is over southern Uganda and western Tanzania (LV), followed by CK and southeastern lowlands with 0.60, while the highest correlation (0.67) obtained over the CS of Kenya and Tanzania. Niño 3.4 index alone is adequate to describe the interannual variability of the average duration of dry spells during the short rainfall season over southern Uganda and western Tanzania (LV). A further assessment showed that the model residuals over the three subregions were normal distributed and independent of each other.

3.3.4. Frequency of wet spells of 3 days or more

The multiple correlation coefficient for the cross-validated models over the five subregions ranges between 0.53 and 0.69. The lowest multiple correlation coefficients are observed over western Kenya and most parts of Uganda (WKU) as well as the CS of Kenya and Tanzania. The LEPS skill score ranges between 25 % and 41% for the five MLR cross-validated models, with the lowest value again observed over western Kenya and most parts of Uganda (WKU). The residuals from these cross-validated MLR models are normally distributed.

4. Conclusions and recommendations

In summary, this study explores additional potential predictors from the large-scale (oceanic and atmospheric) climatic fields which have plausible physical/dynamical relationship with the seasonal rainfall totals and intraseasonal statistics of wet and dry spells during the short rainfall season at subregional levels over Equatorial East Africa region. Four such potential predictors are identified.

The fitting of the cross-validated multiple linear regression (MLR) models for the seasonal rainfall and intraseasonal statistics of wet spells mainly picked the two of the additional potential predictors around Bay of Bengal mostly. These predictors are BoBEN (SST index over the Bay of Bengal extending to west coast of Malaysia and Indonesia and averaged over July–August period) and SINDS (a zonal wind component index at 925 mb to the south of the Bay of Bengal and near the southern tip of Indian subcontinent averaged over the same period). Though BoBEN and SINDS indices appear to be located closer to each other, the two indices are statistically independent. The two predictor indices combined accounted for 40% of the magnitude of the SST changes between the July–August and OND periods over the Western Pole of the Indian Ocean Dipole (IOD). Thus they are among the precursor indicators for the formation of the IOD/ZIND index especially the Western Pole that usually evolves very fast in time and closer to the OND season over Equatorial Eastern Africa.

Consistent with earlier studies, most of the cross-validated MLR models are developed using two or three predictors, and occasionally one predictor. Fewer predictors tend to produce better models than those developed using large numbers of predictors. An assessment of the models developed showed that the LEPS skill score is positive for all the models, which indicates that they performed much better than the climatology and are very robust. The residuals from the models are confirmed to be independent and having a normal distribution. Occasionally, the test for independence is inconclusive. This occurs when the calculated value lies between the lower and upper boundary of the critical values. The skills of the models at some subregions for the some intraseasonal statistics are not high enough to justify their discussion and future incorporation in operational uses.

For the first time, this study has produced cross-validated MLR models for various intraseasonal statistics of wet and dry spells in additional to the routine seasonal rainfall models developed by the IGAD Climate Prediction and Applications Centre (ICPAC) and National Meteorological and Hydrological Services (NMHS). Consistent with

earlier studies which have shown the spatial coherence of the seasonal rainfall and NW days to be almost equal, the skills of the models developed for these two variables are similar thus confirming the spatial coherence tests.

This study has relied mostly on linear statistical methods that provide important tools for comparing model predictions and observations and subsequently identifying model deficiencies. Recent research has however demonstrated that non-linear methods can yield statistically significant increases in prediction skill when compared to traditional linear techniques. The study therefore recommended non-linear alternatives should be explored to augment the prediction skills of the models discussed here. This study also recommends the use of other climatic fields such as the sub-surface heat content in the search of additional potential predictors.

Acknowledgements

This article was part of research for the award of Doctor of Philosophy by the lead author (W.G.). Most part of this work was done when the lead author (W.G.) was attached to the Centre de Recherché de Climatologie, Université de Bourgogne for nine months under the sponsorship of Embassy of France in Kenya for which he is greatly indebted.

References

Annamalai H, Slingo JM, Sperber KR, Hodges K. 1999. The mean evolution and variability of the Asian summer monsoon: comparison of ECMWF and NCEP-NCAR re-analyses. *Mon. Weather Rev.* **127**: 1157–1186.

Ashok K, Guan Z, Yamagata T. 2003. Influence of the Indian Ocean Dipole on the Australian winter rainfall. *Geophys. Res. Lett.* **30**(15): 1821, doi: 10.1029/2003GL017926.

Barnston AG, Thiao W, Kumar V. 1996. Long-lead forecasting of seasonal precipitation in Africa using CCA. *Weather Forecast.* **11**: 506–520.

Bärring L. 1988. Regionalization of daily rainfall in Kenya by means of common factor analysis. *Int. J. Climatol.* **8**: 371–389.

Behera SK, Luo J, 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. *J. Clim.* **18**: 4514–4530.

Behera SK, Luo J-J, Masson S, Rao SA, Sakuma H, Yamagata T. 2006. A CGCM study on the interaction between IOD and ENSO. *J. Clim.* **19**: 1688–1705.

Biasutti M, Battisti DS, Sarachik ES. 2004. Mechanisms controlling the annual cycle of precipitation in the tropical Atlantic sector in an atmospheric GCM. *J. Clim.* **17**: 4708–4723.

Black E. 2005. The relationship between Indian Ocean sea-surface temperature and East African rainfall. *Philos. Trans. R. Soc. 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. *Mon. Weather Rev.* **103**: 74–94.

Cadet DL. 1985. The Southern Oscillation over the Indian Ocean. *J. Climatol.* **5**: 189–212.

Camberlin P, Okoola RE. 2003. The onset and cessation of the "long rains" in eastern Africa and their interannual variability. *Theor. Appl. Climatol.* **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. *J. Clim.* **15**: 1002–1019.

Camberlin P, Gitau W, Oettli P, Ogallo L, Bois B. 2014. Spatial interpolation of daily rainfall stochastic generation parameters over East Africa. *Clim. Res.* **59**: 39–60.

Clark CO, Webster PJ, Cole JE. 2003. Interdecadal variability of the relationship between the Indian Ocean zonal mode and East African coastal rainfall anomalies. *J. Clim.* **16**: 548–554.

D'Arrigo R, Smerdon JE. 2008. Tropical climate influences on drought variability over Java, Indonesia. *Geophys. Res. Lett.* **35**: L05707, doi: 10.1029/2007GL032589.

Delsole T, Shukla J. 2002. Linear prediction of Indian monsoon rainfall. *J. Clim.* **15**: 3645–3658.

Diro GT, Grimes DIF, Black E. 2011. Teleconnections between Ethiopian summer rainfall and sea surface temperature: part II. *Clim. Dyn.* **37**: 121–131.

Downing C, Preston F, Parusheva D, Horrocks L, Edberg O, Semazzi F, Washington R, Muteti M, Watkiss P, Nyangena W. 2008. Kenya: climate screening and information exchange. Report no. AEA/ED05603 2, DFID, London.

Engelen RJ, Wittmeyer IL, Stephens GL. 1998. Assessment of reanalysis hydrology and energy budgets: water vapour and radiative fluxes. In *Proceedings of First International WCRP Conference on Reanalysis*, Silver Spring, MD, World Climate Research Programme, WCRP-104 (WMO/TD-876), 175–178.

Farebrother RW. 1980. The Durbin-Watson test for serial correlation when there is no intercept in the regression. *Econometrica* **48**(6): 1553–1563.

Gitau W. 2011. *Diagnosis and Predictability of Intraseasonal Characteristics of Wet and Dry Spells over Equatorial East Africa*, PhD thesis, Department of Meteorology, University of Nairobi, Nairobi, 260 pp.

Gitau W, Ogallo LA, Camberlin P, Okoola RE. 2013. Spatial coherence and potential predictability assessment of intraseasonal statistics of wet and dry spells over equatorial Eastern Africa. *Int. J. Climatol.* **33**: 2690–2705.

Goddard L, Graham NE. 1999. The importance of the Indian Ocean for simulating precipitation anomalies over eastern and southern Africa. *J. Geophys. Res.* **104**: 19099–19116.

Godínez-Domínguez E, Rojo-Vázquez J, Galván-Piña V, Aguilar-Palomino B. 2000. Changes in the structure of a coastal fish assemblage exploited by a small scale gillnet fishery during an El Niño–La Niña event. *Estuar. Coast. Shelf Sci.* **51**(6): 773–787.

Gong X, Barnston AG, Ward MN. 2003. The effect of spatial aggregation on the skill of seasonal precipitation forecasts. *J. Clim.* **16**: 3059–3071.

Hastenrath S. 2007. Circulation mechanisms of climate anomalies in East Africa and equatorial Indian Ocean. *Dyn. Atmos. Oceans* **43**: 25–35.

Indeje M, Semazzi FHM, Ogallo LJ. 2000. ENSO signals in East African rainfall and their prediction potentials. *Int. J. Climatol.* **20**: 19–46.

Isaaks EH, Srivastaka RM. 1989. *Applied Geostatistics*. Oxford University Press: New York, NY.

Janowiak JE. 1988. An investigation of interannual rainfall variability in Africa. *J. Clim.* **1**: 165–179.

Jolliffe I, Stephenson DB. 2003. *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. John Wiley and Sons: New York, NY.

Jury MR, Pathack B. 1993. Composite climatic patterns associated with extreme modes of summer rainfall over southern Africa: 1975–1984. *Theor. Appl. Climatol.* **47**: 137–145.

Jury MR, Pathack B, Waliser D. 1993. Satellite OLR and microwave data as a proxy for summer rainfall over subequatorial Africa and adjacent oceans. *Int. J. Climatol.* **13**: 257–269.

Jury MR, Matari E, Matitu M. 2009. Equatorial African climate teleconnections. *Theor. Appl. Climatol.* **95**: 407–416.

Kendall MG. 1976. *Time Series*. Charles Griffin: London.

Kendall MG, Stuart A. 1961. *Advanced Theory of Statistics*. Charles Griffins: London.

Korecha D, Barnston AN. 2007. Predictability of June-September rainfall in Ethiopia. *Mon. Weather Rev.* **135**: 628–650.

Krishna Kumar K, Soman MK, Rupa Kumar K. 1995. Seasonal forecasting of Indian summer monsoon rainfall. *Weather* **50**: 449–467.

Latif M, Dommenget D, Dima M, Grotzner A. 1999. The role of Indian Ocean sea surface temperature in forcing East African rainfall anomalies during December–January 1997/98. *J. Clim.* **12**: 3497–3504.

Li S, Robinson WA, Peng S. 2003. Influence of the North Atlantic SST tripole on northwest African rainfall. *J. Geophys. Res.* **108**(D19): 1–16, doi: 10.1029/2002JD003130.

Liao JG, McGee D. 2003. Adjusted coefficients of determination for logistic regression. *Am. Stat.* **57**: 161–165.

Makridakis S, Wheelwright SC, Hyndman RB. 1998. *Forecasting: Methods and Applications*. John Wiley and Sons: New York, NY, 241–310.

Mason SJ. 1995. Sea-surface temperature-South African rainfall associations, 1910–1989. *Int. J. Climatol.* **15**: 119–135.

Mc Cuen RH. 1985. *Statistical Methods for Engineers*. Prentice Hall: Englewood Cliffs, NJ, 439 pp.
Misra V. 2003. The influence of Pacific SST on the precipitation over southern Africa diagnosed from an AGCM. *J. Clim.* **16**: 2408–2418.
Moron V, Philippon N, Fontaine B. 2003. Skill of Sahel rainfall variability in four atmospheric GCMs forced by prescribed SST. *Geophys. Res. Lett.* **30**(23): 2221, doi: 10.1029/2003GL018006.
Moron V, Robertson AW, Ward MN. 2006. Seasonal predictability and spatial coherence of rainfall characteristics in the tropical setting of Senegal. *Mon. Weather Rev.* **134**: 3248–3262.
Moron V, Robertson AW, Ward MN, Camberlin P. 2007. Spatial coherence of tropical rainfall at the regional scale. *J. Clim.* **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. *Mon. Weather Rev.* **124**: 224–243.
Mutai CC, Ward MN, Colman AW. 1998. Towards the prediction of the East Africa short rains based on sea-surface temperature coupling. *Int. J. Climatol.* **18**: 975–997.
National Research Council. 2010. *Assessment of Intraseasonal to Interannual Climate Prediction and Predictability*. The National Academies Press: Washington, DC.
Nayagam LR, Janardanan R, Ram Mohan HS. 2008. An empirical model for the seasonal prediction of southwest monsoon rainfall over Kerala, a meteorological subdivision of India. *Int. J. Climatol.* **28**: 823–831.
Newman M, Sardeshmukh PD, Bergman JW. 2000. An assessment of the NCEP, NASA and ECMWF reanalyses over tropical west Pacific Warm Pool. *BAMS* **81**: 41–48.
Nicholson SE, Dezfuli AK. 2013. The relationship of rainfall variability in western Equatorial Africa to the tropical oceans and atmospheric circulation. Part I: the boreal spring. *J. Clim.* **26**: 45–65.
Nicholson S, Kim J. 1997. The relationship of the El Niño–Southern Oscillation to African rainfall. *Int. J. Climatol.* **17**: 117–135.
Ntale HK, Gan TY. 2004. East African rainfall anomaly patterns in association with El Niño/Southern Oscillation. *J. Hydrol. Eng.* **9**: 257–268.
Nyakwada W. 2009. *Predictability of East African Seasonal Rainfall with Sea Surface Temperature Gradient Modes*, PhD thesis, Department of Meteorology, University of Nairobi, Nairobi.
Nyakwada W, Ogallo LA, Okoola RE. 2009. The Atlantic-Indian Ocean Dipole and its influence on East African seasonal rainfall. *J. Meteorol. Relat. Sci.* **3**: 21–35.
Ogallo LJ. 1980. Regional classification of the East African Rainfall stations into homogeneous groups using the method of Principal Component Analysis. *Stat. Climatol. Devel. Atmos. Sci.* **13**: 255–266.
Ogallo LJ. 1988. Relationship between seasonal rainfall in East Africa and Southern Oscillation. *Int. J. Climatol.* **8**: 31–43.
Ogallo LJ, Janowiak JE, Halpert MS. 1988. Teleconnection between seasonal rainfall over Eastern Africa and global sea surface temperature anomalies. *J. Meteorol. Soc. Jpn.* **66**: 807–822.
Okoola RE. 1999a. A diagnostic study of the eastern Africa monsoon circulation during the northern hemisphere spring season. *Int. J. Climatol.* **19**: 143–168.
Okoola RE. 1999b. Midtropospheric circulation patterns associated with extreme dry and wet episodes over equatorial Eastern Africa during the Northern Hemisphere spring. *J. Appl. Meteorol.* **38**: 1161–1169.
Owiti Z, Ogallo LA. 2007. Dynamical aspects of Indian Ocean Dipole and East African rainfall anomaly linkages. *J. Meteorol. Relat. Sci.* **1**(1): 14–19.
Owiti ZO, Ogallo LA, Mutemi J. 2008. Linkages between the Indian Ocean Dipole and East African seasonal rainfall anomalies. *J. Meteorol. Relat. Sci.* **2**(1): 3–17.
Paeth H, Friederichs P. 2004. Seasonality and time scales in the relationship between global SST and African rainfall. *Clim. Dyn.* **23**: 815–837.
Philippon N, Camberlin P, Fauchereau N. 2002. Empirical predictability study of October-December East African rainfall. *Q. J. R. Meteorol. Soc.* **128**: 2239–2256.
Philippon N, Martiny N, Camberlin P. 2009. Forecasting the vegetation photosynthetic activity over the Sahel: a model output statistics approach. *Int. J. Climatol.* **29**: 1463–1477.
Potts JM, Folland CK, Jolliffe IT, Sexton D. 1996. Revised "LEPS" scores for assessing climate model simulations and long-range forecasts. *J. Clim.* **9**: 34–53.
Rayner NA, Parker DE, Horton EB, Folland CK, Alexander LV, Rowell DP, Kent EC, Kaplan A. 2003. Global analyses of sea surface temperature, sea ice and night marine air temperature since the late nineteenth century. *J. Geophys. Res.* **108**(D14): 4407, doi: 10.1029/2002JD002670.
Reason CJC. 2001. Subtropical Indian Ocean SST dipole events and southern African rainfall. *Geophys. Res. Lett.* **28**: 2225–2227.
Reason CJC. 2002. Sensitivity of the southern African circulation to dipole SST patterns in the south Indian Ocean. *Int. J. Climatol.* **22**: 377–393.
Reason CJC, Lutjeharms JRE. 1998. Variability of the south Indian Ocean and implications for southern African rainfall. *S. Afr. J. Sci.* **94**: 115–123.
Robertson AW, Moron V, Swarinoto Y. 2009. Seasonal predictability of daily rainfall statistics over Indramayu District, Indonesia. *Int. J. Climatol.* **29**: 1449–1462.
Saji NH, Goswami BN, Vinayachandran PN, Yamagata T. 1999. A dipole mode in the tropical Indian Ocean. *Nature* **401**: 360–363.
Shinoda M, Kawamura R. 1996. Relationships between rainfall over semi-arid Southern Africa, geopotential heights and sea surface temperatures. *J. Meteorol. Soc. Jpn.* **74**(1): 21–36.
Stendel M, Arpe K. 1997. Evaluation of the hydrological cycle in reanalyses and observations. Report No. 228, Max-Planck-Institut fur Meteorologie, Hamburg, Germany, 52 pp.
Stephenson DB, Rupa Kumar K, Doblas-Reyes FJ, Royer JF, Chauvin F, Pezzulli S. 1999. Extreme daily rainfall events and their impact on ensemble forecasts of the Indian monsoon. *Mon. Weather Rev.* **127**: 1954–1966.
Ummenhofer CC, England MH, McIntosh PC, Meyers GA, Pook MJ, Risbey JS, Gupta AS, Taschetto AS. 2009. What causes southeast Australia's worst droughts? *Geophys. Res. Lett.* **36**: L04706, doi: 10.1029/2008GL036801.
Uppala SM, Kallberg PW, Simmons AJ, Andrae U, Da Costa Bechtold V, Fiorino M, Gibson JK, Haseler J, Hernandez A, Kelly GA, Li X, Onogi K, Saarinen S, Sokka N, Allan RP, Andersson E, Arpe K, Balmaseda MA, Beljaars ACM, Vande Berg L, Bidlot J, Bormann N, Caires S, Chevallier F, Dethof A, Dragosavac M, Fisher M, Fuentes M, Hagemann S, Holm E, Hoskins BJ, Isaksen L, Janssen PAEM, Jenne R, McNally AP, Mahfouf JF, Morcrette JJ, Rayner NA, Saunders RW, Simon P, Sterl A, Trenberth KE, Untch A, Vasiljevic D, Viterbo P, Woollen J. 2005. The ERA-40 re-analysis. *Q. J. R. Meteorol. Soc.* **131**: 2961–3012.
Usman MT, Reason CJC. 2004. Dry spell frequencies and their variability over southern Africa. *Clim. Res.* **26**: 199–211.
Walker ND. 1990. Links between South African summer rainfall and temperature variability of the Agulhas and Benguela system. *J. Geophys. Res.* **95**: 3297–3319.
Ward NM, Folland CK. 1991. Prediction of seasonal rainfall in north Nordeste of Brasil using eigenvectors of sea-surface temperature. *Int. J. Climatol.* **11**: 711–743.
Webster PJ, Moore AM, Loschnigg JP, Leben RR. 1999. Coupled ocean–atmosphere dynamics in the Indian Ocean during 1997–98. *Nature* **401**: 356–360, doi: 10.1038/43848.
Wilks DS. 2006. *Statistical Methods in the Atmospheric Sciences*, 2nd edn. Associated Press: London, 627 pp.
Zhang H, Casey T. 2000. Verification of categorical probability forecasts. *Weather Forecast.* **15**: 80–89.

