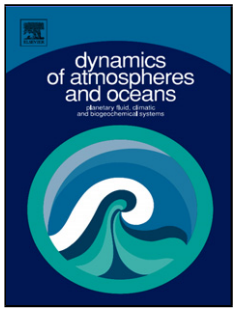


Title: Influence of Spatial and Temporal Scales on Statistical Analyses of Rainfall Variability in the River Nile Basin

Author: Charles Onyutha Patrick Willems

PII: S0377-0265(16)30116-6
DOI: http://dx.doi.org/doi:10.1016/j.dynatmoce.2016.10.008
Reference: DYNAT 1004

To appear in: *Dynamics of Atmospheres and Oceans*

Received date: 9-7-2016
Revised date: 25-10-2016
Accepted date: 26-10-2016

Please cite this article as: Onyutha, Charles, Willems, Patrick, Influence of Spatial and Temporal Scales on Statistical Analyses of Rainfall Variability in the River Nile Basin.Dynamics of Atmospheres and Oceans http://dx.doi.org/10.1016/j.dynatmoce.2016.10.008

Influence of Spatial and Temporal Scales on Statistical Analyses of Rainfall Variability in the River Nile Basin

Charles Onyutha[*1, 2] and Patrick Willems[1]

[1]Hydraulics Laboratory, KU Leuven, Kasteelpark Arenberg 40, 3001 Leuven, Belgium

[2]Faculty of Technoscience, Muni University, P.O. Box 725, Arua, Uganda

*****Corresponding Author (Charles Onyutha); E-Mail: charles.onyutha@bwk.kuleuven.be

Highlights:

- Variability in gridded rainfall was analyzed based on various spatio-temporal scales
- Empirical Orthogornal Function (EOF) was applied to analyze the rainfall variation
- As the grid size increased, the amount of variance explained by the EOF reduced
- As the data temporal resolution reduced, the amount of explained variance increased
- Rainfall variability was more explainable at regional than location-specific scale

Abstract:

In this study, empirical orthogonal function was applied to analyze rainfall variability in the Nile Basin based on various spatio-temporal scales. The co-occurrence of rainfall variability and the variation in selected climate indices was analyzed based on various spatio-temporal scales. From the highest to the lowest, the cumulative amount of variance explained by the first two principal components (PCs) for any selected size of the spatial domain was obtained for the annual, seasonal, and monthly rainfall series respectively. The variability in the annual rainfall of $1^{o}\times1^{o}$ spatial coverage explained by only the first PC was about 55% on average. However, this percentage reduced to about 40% on average across the study area when the size of the spatial domain was increased from $1^{o}\times1^{o}$ to $10^{o}\times10^{o}$. The variation in climate indices was shown to explain rainfall variability more suitably at a regional than location-specific spatial scale. The magnitudes and sometimes signs of the correlation between rainfall variability and the variation in climate indices tended to vary from one time scale to another. These findings are vital in the selection of spatial and temporal scales for more considered attribution of rainfall variability across the study area.

Keywords: Empirical Orthogonal Function (EOF); Variability Analyses; River Nile Basin; Climate Variability; Rainfall Variability; Meteorology

1 Introduction

Because the stationarity of weather for any extended period is highly unlikely, there exists the concern to investigate the effects of climate variability on hydro-meteorology. In line with the aforementioned concern, several recent studies can be found in literature; for instance As-syakur et al. (2016) investigated rainfall variability in the Maritime Continent, Li et al. (2015) studied the modulation of inter-annual variability of tropical cyclone activity, Onyutha (2016a) analyzed the changes in long-term evapotranspiration in the River Nile riparian countries, Donohue et al. (2015) assessed the loop current path variability, Zuo et al. (2016) examined the sub-seasonal variability and predictability of the Arctic Oscillation, etc.

In the study area i.e. the Nile Basin which comprises the River Nile shared by 11 riparian countries (see **Figure 1**), majority of the people depend on rain-fed agriculture to support livelihoods. Due to the latitudinal movement of the inter-tropical convergence zone, the distribution of rainfall is spatially contrasted over the Nile Basin. The study area is characterized by humid, hyper-arid, and sub-humid conditions. Rainfall extremes occur in some parts of the Nile Basin thereby leading to flood and drought conditions the frequency of which tends to be regulated by the effect of climate variability. Therefore, the knowledge of historical changes in rainfall and their associated drivers is relevant for e.g. the management of agricultural practices, planning and operation of risk-based water resources applications, etc.

For the different regions of the Nile Basin, several studies have been conducted on the teleconnection of rainfall variability to large-scale ocean-atmosphere interactions (Nicholson and Entekhabi, 1986; Nyeko-Ogiramoi et al., 2013; Onyutha, 2016b; Onyutha and Willems, 2015a; Phillips and McIntyre, 2000; Taye and Willems, 2012; Tierney et al., 2013). In a study relevant for agricultural practices, Phillips and McIntyre (2000) investigated the El Niño-Southern Oscillation (ENSO) and inter-annual rainfall variability in Uganda. While considering all the River Nile riparian countries, Onyutha (2016b) applied nonparametric anomaly indicator method to derive variability in country-wide seasonal and annual rainfall. The author also examined the co-variation of rainfall with some climate indices including the Indian Ocean Dipole (IOD), Atlantic Multidecadal Oscillation (AMO) index, Pacific Decadal Oscillation (PDO) and Southern Oscillation Index (SOI). Using station-based observed rainfall at various locations across the entire study area, analyses of the variability in seasonal and annual rainfall were also conducted by Onyutha and Willems (2015a) based on the quantile perturbation approach. The authors used series of 40 years and above with missing data points not more than 10% to investigate the causes of rainfall variability with respect to the global Sea Surface Temperature (SST) and Sea Level Pressure (SLP) as well as 10 climate indices. Taye and Willems (2012) and Nyeko-Ogiramoi et al. (2013) applied quantile perturbation method to analyze the variation in rainfall extremes of the Blue Nile basin and Lake Victoria basin respectively. These authors also assessed the co-occurrence of the anomalies in the rainfall extremes and those of some climate indices such as the IOD, AMO, PDO and North Atlantic Oscillation (NAO) index. Tierney et al. (2013) applied the Empirical Orthogonal Function (EOF) to investigate the influence of the Indian Ocean on the multidecadal variability in East African hydro-climate. From the above studies, the selection of the following factors for analyses tended to differ: the variability computation method, spatial coverage, temporal scale, rainfall variability driver, etc. Generally, the above factors as well as the questionable

data quality (e.g. missing records), data limitation (e.g. short-term records), etc influence the analyses of rainfall variability and its attribution. Nonetheless, in all the above cited studies for the study area, little or no attention was given to investigate how rainfall variability analyses are influenced by any of the several aforementioned factors. It may be possible that the different factors may influence rainfall variability synergistically. However, for an in-depth investigation of their influence on rainfall variability, one factor can be assessed at a time (as done in this study) considering the variation in the spatio-temporal scales.

Some of the matrix methods for statistical analyses of structures in large datasets (as considered in this study) include the EOF, Maximum Covariance Analysis (MCA), Canonical Correlation Analysis (CCA), etc. As opposed to the use of space and time dimensions by the EOF to explain the maximum amount of variability in a given data matrix, the MCA and CCA consider two matrices with the same sampling dimension but of different structures. The application of the MCA especially for meteorological and oceanographic studies follows its popularization by Wallace et al. (1992) and Bretherton et al. (1992). However, worth noting is the extreme caution by Cherry (1996) that both the MCA and CCA tend to produce spurious spatial patterns which may not be physically meaningful or explainable. Eventually, the EOF was adopted for the analyses in this study. Moreover, the robustness of the EOF method for analyses of rainfall variability in the study area was already demonstrated in some studies such as Indeje et al. (2000), Ogallo (1988), Tierney et al. (2013), and Yeshanew and Jury (2007).

This study was therefore specifically aimed at investigating the influence of spatial and temporal scales on the analyses rainfall variability. The influence of the variation in spatio-temporal scales on the linkage of rainfall variability to the large-scale ocean-atmosphere interactions was also investigated.

2 Data and Considerations

2.1 Rainfall Data

Global gridded ($0.5^{o}\times0.5^{o}$) daily rainfall data of the Princeton Global Forcings (PGFs) (Sheffield et al., 2006) were obtained online at http://hydrology.princeton.edu/data/pgf/0.5deg/ [accessed: 12th February, 2016]. Initially the PGF-based daily rainfall series were from 1948 to 2006 and of $1^{o}\times1^{o}$ spatial resolution. However, the series were recently extended to cover the period 1948–2008, and their spatial resolution refined to $0.5^{o}\times0.5^{o}$ grid. The PGFs are observational-reanalysis hybrid i.e. derived by combining the NCEP–NCAR reanalysis dataset (Kalnay et al., 1996) with a number of observational-based products including the CRU TS2.0, the GPCP, the TRMM, and the NASA Langley Research Center SRB (see Sheffield et al. (2006) for the details of the data description). It may be possible (though not investigated in this study) that the bias and errors in the rainfall data from the various sources even out through their hybridization to form the PGF-based series.

The location of point rainfall was marked in each of the selected countries or regions labeled (1) to (6) in **Figure 1**. For each region, grids were selected in an increasing order from $1^{o}\times1^{o}$ to $10^{o}\times10^{o}$ (see **Figure 2**) such that the location of the point rainfall fell at the center of the area covered by each spatial domain. Because the PGF data obtained were of $0.5^{o}\times0.5^{o}$ grid resolution, the total number of grid points from $h^{o}\times h^{o}$ spatial scale was $2h^2$. For instance, there were $(2\times10)^2 = 400$ grid points for the

$10^{o}\times10^{o}$ spatial domain of each region. Point rainfall at the center of the $1^{o}\times1^{o}$ spatial domain was also considered. The extents of the grid or spatial domains for each region are summarized in **Table 1**.

The daily rainfall at each grid point was converted to other temporal resolutions including monthly, seasonal and annual time scales. For the seasonal rainfall, only the months covering the rainy seasons i.e. from March to May (MAM) for regions 1-2, and June to September (JJAS) for regions 3-5 were considered. Whereas MAM is the main rainy season for region 1-2 in the equatorial region (Nicholson, 1996) as well as 6 (Egypt), the "long" rains for regions 3-5 occur in JJAS (Onyutha and Willems, 2015a).

Due to their robustness and the long period they cover, data from the PGFs have been employed in several variability-related studies e.g. Hoell et al. (2015), Zeng and Cai (2016), etc. Zeng and Cai (2016) used the PGF data to investigate the evapotranspiration temporal variability under the influence of climatic and terrestrial storage. Based on the PGF data, Hoell et al. (2015) investigated the forcing of monthly precipitation variability in the Southwest Asia. In this study, the validity of the PGF series for variability analyses was first assessed based on daily rainfall obtained at 10 stations (**Table 2**) selected over different regions. Stations labeled 1-5 and 6-10 in **Table 2** were selected from the equatorial region, and the Ethiopian Highlands respectively. These regions were considered because they normally have high rainfall variability and totals in excess of 1000 mm annually. Daily rainfall data from the meteorological stations 1-4 (**Table 2**) were obtained from the database of the FRIEND/Nile (River Nile basin Flow Regimes from International, Experimental and Network Data). The details on the FRIEND/Nile project can be found via http://www.unesco.org/new/en/cairo/natural-sciences/hydrology-programme/friendnile/ [accessed: 22nd August, 2016]. Furthermore, daily rainfall data at stations 6-10 (**Table 2**) were obtained online from the Global Historical Climatology Network (GHCN) (Menne et al., 2012a, 2012b) through the link http://www.ncdc.noaa.gov/oa/climate/ghcn-daily/ [accessed: 11th June, 2014]. The data from the meteorological stations 1-4 and 6-10 used in this study were adopted following their quality-control by Onyutha and Willems (2015b) and Onyutha (2016c) respectively. Finally, long-term daily rainfall data from Station 5 in **Table 2** were obtained from the Ministry of Water and Environment, Uganda. At each of the selected locations, the daily PGF rainfall series were also extracted.

Validity of the PGF rainfall was checked in two ways. Firstly, the correlation between the observed or station-based rainfall and the PGF series was analyzed at various time scales. Secondly, the coefficient of variation (CV) or measure of variability in the observed series was compared with that of the PGF series. The null hypothesis H_0 of no correlation between the observed and PGF rainfall was generally rejected at the significance level of 5% for the various time scales (**Table 2**). Furthermore, as the temporal resolution of the rainfall series increases i.e. from annual through seasonal to monthly time scale, the CV is also shown to increase for both the observed and PGF rainfall series (**Figure 3**). The scatter points are shown to fall along the bisector. In other words, the PGF series reproduce the variability of the observed rainfall fairly well for the various time scales (as demonstrated for Station 8 in **Figure 4 a-d**) especially for seasonal and annual time scales. Therefore, based on results shown in **Table 2** and **Figure 3**, the quality of the PGF data was deemed adequate for the purpose of this study.

2.2 Climate Indices

It is known that the variability of rainfall in the equatorial region of the study area is driven by the variation in SST of the Indian Ocean or SLP in the North Atlantic Ocean (Onyutha and Willems, 2015a; Tierney et al., 2013). For the northern part of the Nile Basin, the rainfall variability is linked to the variation of the SST over the Atlantic and Pacific Oceans (Jury, 2010; Taye and Willems, 2012). In the same line, four climate indices relevant for the analyses of rainfall variation in the study area were selected as briefly explained next.

a) The IOD is the anomalous SST difference between the western (50°–70°E and 10°S–10°N) and south eastern (90°–110°E and 10°S–0°N) equatorial Indian Ocean. Monthly IOD data from 1900 to 2011 were obtained from the website of the Japan Agency for Marine-Earth Science and Technology via http://www.jamstec.go.jp/frcgc/research/d1/iod [accessed: 20th January, 2014].

b) The AMO index is defined as the SST averaged over 25°–60° N, 7°–70°W minus the regression on global mean temperature (van Oldenborgh et al., 2009). The AMO index for the period 1900-2012 was obtained through the climate explorer of the Koninklijk Nederlands Meteorologisch Instituut via http://climexp.knmi.nl/data/iamo_hadsst2.dat [accessed: 29th January, 2013].

c) The PDO is the leading principal component of North Pacific Ocean (pole ward of 20° N in the Pacific Basin) (Mantua et al., 1997). The PDO index from 1900 to 2011 was obtained from website of the Joint Institute for the Study of the Atmosphere and Ocean via the link http://jisao.washington.edu/pdo/PDO.latest [accessed: 30th January, 2013].

d) The NAO index is the normalized SLP difference between Reykjavik, Gibraltar and Azores (Jones et al., 1997). The NAO index from 1900 to 2012 was obtained through the website of the National Oceanic and Atmospheric Administration at http://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/NAO/ [accessed: 16th April, 2016].

3 Methodology

3.1 EOF Analyses of Rainfall Variability

To investigate the effect of spatial resolution on variability analyses, EOF was applied to the rainfall of various spatial domains shown in **Figure 2** for each region. Furthermore, for the rainfall of each selected size of the spatial domain, the EOF analysis was conducted based on the different temporal resolutions (i.e. monthly, seasonal and annual time scales) considering each region shown in **Figure 1**. It is known that daily data include both synoptic and climatic time scales. On the other hand, monthly data include seasonal, inter-annual and inter-decadal time scales. Therefore for clarity of variability results, daily time scale was not considered, and seasonal time variation which is a climate signature was first removed from the monthly series before conducting the EOF analyses. The procedure taken to deseasonalize the monthly series can be obtained illustrated in **Appendix A**. The EOF Analysis was

also performed considering rainfall over the entire Nile Basin as the spatial domain of the gridded series.

A small number of linearly uncorrelated series can be derived from the given inter-correlated datasets so as to explain the maximum amount of variance. The structures explaining the maximum amount of variance in the two dimensional series (i.e. gridded rainfall in this case) are analyzed using the EOF. In the EOF analyses, space and time are used as the structure and sampling dimensions respectively. The set of structures produced in the first (i.e. the structure) dimension are called the EOFs. In the sampling (i.e. time) dimension, Principal Components (PCs) are produced as the complementary set of structures. The PCs and EOF's are orthogonal in their own dimension.

Consider that m variables Y_1, Y_2,... Y_m are correlated with each other and Q is to be predicted given the Y_j's. First and foremost, it would be valuable to establish a new predictor Z as linear combinations of the Y_j's using:

$$\left.\begin{aligned} Z_1 &= a_{11}Y_1 + a_{12}Y_2 + \ldots + a_{1j}Y_j \ldots + a_{1m}Y_m \\ Z_2 &= a_{21}Y_1 + a_{22}Y_2 + \ldots + a_{2j}Y_j \ldots + a_{2m}Y_m \\ &\ldots\ldots\ldots \\ Z_i &= a_{i1}Y_1 + a_{i2}Y_2 + \ldots + a_{ij}Y_j \ldots + a_{im}Y_m \\ &\ldots\ldots\ldots \\ Z_n &= a_{n1}Y_1 + a_{n2}Y_2 + \ldots + a_{nj}Y_j \ldots + a_{nm}Y_m \end{aligned}\right\} \quad (1)$$

In this study, Y_j is the rainfall series at the j^{th} grid point, n is the possible maximum number of dimensions to explain the desired amount of variance in the data, m equals to the number of grid points. In summary, $Z_i = a_{ij}Y_j$ such that the matrix of coefficients a_{ij} (for the i^{th} observation at the j^{th} grid point) rotates the original set of series into a second one. The values of the a_{ij} can be determined such that the Z_i's are uncorrelated and orthogonal variables. Eventually, Z_1 explains the maximum amount of variance of the Y_j's, the Z_2 explains the maximum amount of the remaining variance of the Y_j's, etc. The values of the coefficients a_{ij} are the EOFs onto which when the series Y_j are projected, the PCs i.e. Z_i (the original series in the new coordinate system) are obtained. In other words, the coefficients a_{ij} are the "loadings" and represent the weights of the original variables in the PCs. It is the orthogonality (i.e. lack of correlation in time) which makes the PCs very efficient and suitable for analyses of variability.

The extraction of the PCs in this study was based on the correlation matrix such that the n order eigenvector (a_{11} $a_{12,}$ …, a_{1n}) included the coefficients of n PCs. Other bases of PC extraction include variance, covariance, etc. The amount of variability explained by the first, second, .., n^{th} EOF is referred to as the first (η_1), second (η_2), n^{th} (η_n) eigenvalues respectively such that $\eta_1 \geq \eta_2 \geq \ldots \geq \eta_n$. The proportion of the variance explained by the i^{th} component is expressed as $\eta_{i/\sum}\eta_i$.

Stable patterns in the form of localized structures are expected to emerge over the spatial domain of analysis. However, this expectation is affected by the significant amplitude which tends to occur due to effect of the orthogonality constraint. Eventually, to reduce the effect of orthogonality constraint and

achieve more stable spatial patterns, rotation of the eigenvectors can be considered (Horel, 1984; Richman, 1986). Based on the rotation, regions with similar temporal variations are isolated thereby improving the identification of regions with maximum correlation between the variables and the components (Santos et al., 2010). Several methods exist for rotation of the eigenvectors including the Varimax, Qartimax, Equamax, Parsimax, Quartimin, Promax, Orthomax, Oblimin, etc. In this study, the Varimax procedure which is the most commonly used was adopted because according to Richman (1986) it preserves the orthogonality, and gives more physically explainable variability patterns than other approaches. Varimax is also known to exhibit less sensitivity to the number of variables in the EOF analysis.

Consider that the rotation of a given factor-loading matrix, say, a_{ij} into a new matrix c_{ij} is required; the Varimax method employs the approximation of simple structure of individual factors i.e. the variance s_j^2 of its squared loadings (**Eq. 2**). It is important to note that the number of the factors retained is characterized by the new index limit t which should be less than n i.e. the length of the state vector or number of original variables. The sum of the s_j^2 values (from i = 1 to n) is also referred to as the 'row varimax criterion' in which large eigenvalues (i.e. those with large factor loadings) contribute more to the criterion than those with small factor loadings. Eventually, the bias especially on the first eigenvector can be reduced using the weights of the variables based on their communalities g_j^2. The normalized Varimax criterion (R) is given by **Eq. (3)**,

$$s_j^2 = \frac{1}{n}\sum_{i=1}^{n}\left\{a_{ij}^2\right\}^2 - \frac{1}{n^2}\left\{\sum_{i=1}^{n}a_{ij}^2\right\}^2 , \qquad j = 1, 2, ..., t \tag{2}$$

$$R = t\sum_{j=1}^{t}\sum_{i=1}^{n}\left\{\frac{a_{ij}}{g_i}\right\}^4 - \sum_{j=1}^{t}\left\{\sum_{i=1}^{n}\frac{a_{ij}^2}{g_i^2}\right\}^2 \tag{3}$$

where, $$g_i^2 \infty \sum_{j=1}^{t} a_{ij}^2 \tag{4}$$

3.2 Correlation Analyses

Under the null hypothesis H_0 that the variation of a given climate index is uncorrelated to the variability of the rainfall series, correlation analyses was conducted at the significance level of 5%. **Figure 5** shows, for illustration, the co-occurrence of the anomalies of the climate indices and those of the EOF factor scores for $10^0 \times 10^0$ gridded annual rainfall in Kenya and Ethiopia. Since the magnitudes of the correlation coefficients of 0.55 (for Kenya, **Figure 5 a**) and -0.63 (for Ethiopia, **Figure 5 b**) are greater than the critical value of 0.25, it means that the H_0 is rejected at the significance level of 5% in both regions. The correlation analyses were performed in two steps. Firstly, considering the gridded rainfall over the entire Nile Basin, secondly, based on various sizes of the rainfall grids. In each of the two steps, correlation analyses were conducted for monthly, seasonal and annual time scales.

4 Results and Discussions

4.1 Explained Variance versus Spatial Scale

Figure 6 shows, for the annual rainfall, variation of the amount of explained variability with PCs. For all the time scales (though only shown for annual rainfall), the explained variability was found to be small for higher than lower PCs. Considering the 1°×1° spatial domain, the variability explained by only the first PC ranged from about 50% (**Figure 6 a, c, e**) to nearly 60% (**Figure 6 b**, **d, f**). However, by combining the first and the second PCs, the cumulative variability explained in the annual rainfall went above 90% for the 1°×1° gridded rainfall of all the selected regions. Generally, as the size of the spatial domain was increased, the amount of explained rainfall variability noticeably reduced. This, which is clearly evident for the most important (i.e. first two) PCs, can be explained in terms of the increase in rainfall heterogeneity as the size of the spatial domain increases. As the distance between any two grid points increases, the magnitude of the correlation between the data at the selected locations (or grid points) reduces. Eventually, the average of the correlation is higher between data at the various grid points within a small- than that for large-sized spatial domain. This follows the influence of the micro-climate or features e.g., water bodies, topography, etc which become more evident on the rainfall variation as the size spatial domain increases. Of course it is known that the amount of variability that can be explained in the EOF depends on how highly correlated the variables for the analyses are. As the size of the spatial domain increases, the heterogeneity in the rainfall also increases. In the same line, as the rainfall series at the various grid points within a selected spatial domain become less correlated, the amount of variability explained in the EOF also reduces. Eventually, from 1°×1° to 10°×10° size of the spatial domain, the cumulative variability which can be explained from the first two PCs dropped by 43.8, 53.2, 55.1, 36.7, 56.8, 69.1% for regions 1 to 6 as shown in **Figure 6 a-f** respectively.

4.2 Explained Variance versus Temporal Scale

Figure 7 shows the influence of the variation of temporal resolution on rainfall variability considering a given spatial scale. Generally, as the rainfall size of the spatial domain increases, the amount of explained variability reduces as already highlighted in **Figure 6**. For a selected spatial domain, the amount of explained variability reduces as the temporal resolution of the rainfall increases i.e. from annual to seasonal, and monthly time scale (**Figure 7 a-f)**. This is because of the temporal smoothing of series through aggregation e.g. from monthly to annual which removes the short-term fluctuations to obtain a more representative data and allows the study of its general behavior. This realization was expected based on **Figure 3** in which the measure of variability was demonstrated to vary from the highest to the lowest in monthly, seasonal and annual time scales respectively.

4.3 Spatial and Temporal Variability

Figure 8 shows, for the entire Nile Basin, the spatial variation of the leading EOF factor loadings. Whereas the grid points in the equatorial region and those in the northern part of the study area loaded somewhat negatively, the central portion had positive EOF loadings for all the time scales (**Figure 8 a-d**). The EOF loadings appear to divide the rainfall of the Nile Basin into three parts i.e. that in the southern, central and northern areas. This is highly consistent with the grouping of meteorological or

rainfall stations in the study area from the recent study by Onyutha and Willems (2015). The variation in the EOF loadings across the study area suggests the regional differences in the rainfall variability driving forces. Such regional differences in the rainfall variation can be influenced by the micro-climate or features such as topography, water bodies, changes in land-use cover, etc. Moreover, the drivers of the rainfall variation may also differ from one time scale to another. Further investigation of the regional difference was made in a spatio-temporal way by considering the various regions of the study area. The overall distance between the equator and the areas with high positive factor loadings is longer for the JJAS season (**Figure 8 b**) than that of the MAM season (**Figure 8 c**). This is because of the latitudinal migration of the inter-tropical convergence zone as already mentioned in **Section 1**.

Figure 9 and **Figure 10** respectively show the spatial and temporal view of the leading EOF factor loadings. Locations with positive/negative loadings (**Figure 9**) show periods when the rainfall was above/below the reference or the mean of the long-term series as presented in **Figure 10**. For regions 1 and 2, the rainfall was below the reference in the 1950s, 1970s and the early 1990s; however, wet conditions occurred in the 1960s, 1980s and the late 1990s (**Figure 10 a**, **d**). These results agree with those of Onyutha and Willems (2015a) who found that the rainfall in the equatorial region was below (above) reference from the late 1940s to 1950s (1960s to mid-1980s). For regions 3 and 5, dry conditions occurred from the late 1970s to the early 1990s, and rainfall was above the reference from the early 1950s to late 1960s, and around 2000 (**Figure 10 b**). Using the maximum rainfall in each year, Taye and Willems (2012) obtained similar patterns of variability for the upper Blue Nile basin in Ethiopia. For regions 4 and 6 (arid condition), rainfall totals dropped below the reference from the late 1960s to the mid 1970s, and again from the early 1990s to mid 2000s. Somewhat wet periods occurred from the early 1950s to early 1960s, and again around 1980 (**Figure 10 c**). For the Sahel (i.e. region 3), the decline in rainfall from the late 1960s to the mid 1970s was also found by Hulme (1992).

4.4 Correlation Analyses

Figure 11 and **Figure 12** show the correlation between the the anomalies in rainfall and those of the climate indices. The regional difference in the rainfall variability across the study area as highlighted from **Figure 8** is also demonstrated in **Figure 11** and **Figure 12**. It is noticeable that some difference exist among the spatial correlation maps of the various time scales (**Figure 11 a-h** and **Figure 12 a-h**). Instead of the basin-wide spatial scale, further correlation analyses were also conducted using rainfall from $10^{o}\times10^{o}$ spatial domain. The spatial correlation between the annual rainfall variability anomalies and the variation in the climate indices is presented in **Figure 13**. The difference in the magnitude of the correlation from one grid point to the next is evident for the various regions (**Figure 13 a-f**). Because of the tendency of the correlation to differ from one time scale to another, the climate indices IOD (for region 1 and 2), PDO (for region 3 and 5), and AMO (for region 4 and 6) were selected based on the spatial map of annual rainfall shown in **Figure 11 d**, **h** and **Figure 12 d**, **h**.

Table 3 shows the linkage between rainfall variability and the variation in selected climate indices considering various spatio-temporal scales. The magnitudes of the correlation are generally shown to increase as the size of the spatial domain also increases. This shows that the variation in the large-scale ocean-atmosphere interactions can be used to explain the rainfall variability more at a regional than location-specific spatial scale. Although the correlation would be expected to systematically increase with the reduction in the series temporal resolution (e.g. from monthly to annual time scale), it is noticeable that the results tended vary from one time scale to another. In other words, the drivers of the changes in annual rainfall of the study area may suggestively be different from those of other time scales. This implies that the attribution of rainfall variability can be made in a temporal-scale-specific way. The results of such attribution based on various temporal scales can be relevant for different water resources applications. For instance, the consideration of point rainfall of high temporal resolution (e.g. daily time scale) can be relevant for risk-based applications related to drought and flood conditions which occur over various locations of study area. Rainfall of seasonal and annual time scales are vital for planning and management of agricultural practices given that agriculture is the main occupation of majority of the population in the study area.

5 Conclusions

In this study, the influence of variation in spatial and temporal scales on rainfall variability in the Nile Basin was analyzed using gridded (0.5°×0.5°) daily rainfall long-term data of the Princeton Global Forcings. Apart from point rainfall, series from grids of various sizes including 1°×1°, 3°×3°, 5°×5°, and 10°×10° were considered. The grids were selected over various regions of the Nile Basin. The daily rainfall was converted to monthly, seasonal and annual time scales. Seasonal time variation which is a climate signature was first removed from the monthly series before conducting the variability analyses. Empirical orthogonal function was applied to extract the structures explaining the maximum amount of variance in space and time of the gridded data. The attributes of the rainfall variability across the study area were assessed with respect to some climate indices characterizing variation in the sea surface temperature and/or sea level pressures in the Indian, Atlantic, and Pacific Oceans.

The variability in annual rainfall of 1°×1° spatial domain explained by only the first principal component (FPC) ranges from about 50% to nearly 60%. As the size of the spatial domain for the rainfall increases from 1°×1° to 10°×10°, the amount of explained variability reduces to about 40% on average across the study area. For various temporal resolutions, on average, the amounts of variance explained by the FPC for annual, seasonal, and monthly rainfall series were respectively 53.1, 48.8, 26.2% (for 1°×1° spatial domain), 46.6, 40.7, 20.7% (for 3°×3° spatial domain), 38.0, 32.2, 14.5% (for 5°×5° spatial domain), and 29.5, 26.0, 10.2% (for 10°×10°). The variation in the climate indices was found to explain the rainfall variability more suitably at a regional than location-specific spatial scale. The magnitudes and sometimes signs of the coefficients of correlation between rainfall variability and the variation in climate indices tended to vary one time scale to another.

These findings are vital in the selection of spatial and temporal scales for more considered attributions of rainfall variability across the study area. The various spatial scales and temporal resolutions are relevant for incorporating the effect of climate variability into planning, design and/or management of different water resources applications e.g. agricultural practices, risk-based hydraulic structures, etc.

Acknowledgement

The gridded rainfall data used were based on the Princeton Global Forcings obtained online from http://hydrology.princeton.edu/data/pgf/ [acessed: 12th February, 2016]. This research was financially supported by an IRO Ph.D. scholarship of KU Leuven.

Appendix A: Removal of seasonal component of monthly rainfall

Illustration is made using monthly data from Station 8 of **Table 2**. According to Onyutha (2016d), the following steps can be taken in removing the seasonal component of monthly rainfall series:

1) For the given data (**Figure A1 a**), separately extract the series of each month. This will yield 12 datasets considering all the months from January to December (**Figure A1 b**);
2) Compute the average of each dataset or the extracted series for each month (**Figure A1 c**);
3) For each year, take the seasonal component of series as the long-term monthly mean values from January to December as obtained in Step 2 (**Figure A1 d**). In other words, replicate the long-term monthly mean values from January to December for all the years of the data record. It is expected that the seasonality follows the variation of the full time series (**Figure A1 e**).
4) To deseasonalize the data, the seasonal component of each month is subtracted from the original sub-series for the corresponding month in the year under consideration. This yields the residual rainfall (**Figure A1 f**) which can be used for analysis of, e.g., variability.

References

As-syakur, A.R., Osawa, T., Miura, F., Nuarsa, I.W., Ekayanti, N.W., Dharma, I.G.B.S., Adnyana, I.W.S., Arthana, I.W., and Tanaka, T., 2016. Maritime Continent rainfall variability during the TRMM era: The role of monsoon, topography and El Niño Modoki. Dyn. Atmospheres Oceans 75:58–77.

Bretherton, C.S., Smith, C., and Wallace, J.M., 1992. An intercomparison of methods for finding coupled patterns in climate data. J. Clim., 5:541–560.

Cherry, S., 1996. Singular value decomposition analysis and canonical correlation analysis. J. Clim., 9:2003–2009.

Donohue, K.A., Watts, D.R., Hamilton, P., Leben, R., Kennelly, M., and Lugo-Fernández, A., 2015. Gulf of Mexico Loop Current path variability. Dyn. Atmospheres Oceans., http://dx.doi.org/10.1016/j.dynatmoce.2015.12.003

Hoell, A., Shukla, S., Barlow, M., Cannon, F., Kelley, C., and Funk, C., 2015. The forcing of monthly precipitation variability over Southwest Asia during the Boreal cold season. J. Clim., 28:7038–7056.

Horel, J.D., 1984. Complex principal component analysis: Theory and examples. J. Clim. Appl. Meteorol., 23:1660–1673.

Hulme, M., 1992. Rainfall changes in Africa: 1931–1960 to 1961–1990. Int. J. Climatol., 12:685–699.

Indeje, M., Semazzi, F.H.M., and Ogallo, L.J., 2000. ENSO signals in East African rainfall seasons. Int. J. Climatol., 20:19–46.

Jones, P.D., Jonsson, T., and Wheeler, D., 1997. Extension to the North Atlantic oscillation using early instrumental pressure observations from Gibraltar and south-west Iceland. Int. J. Climatol., 17:1433–1450.

Jury, M.R., 2010. Ethiopian decadal climate variability. Theor. Appl. Climatol., 101:29–40.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., et al., 1996. The NCEP/NCAR 40-Year Reanalysis Project. Bull. Am. Meteorol. Soc., 77:437–471.

Li, Z., Yu, W., Li, K., Liu, B., and Wang, G., 2015. Modulation of interannual variability of tropical cyclone activity over Southeast Indian Ocean by negative IOD phase. Dyn. Atmospheres Oceans, 72:62–69.

Mantua, N.J., Hare, S.R., Zhang, Y., Wallace, J.M., and Francis, R.C., 1997. A Pacific interdecadal climate oscillation with impacts on Salmon production. Bull. Am. Meteorol. Soc., 78:1069–1079.

Menne, J.M., Durre, Imke, Korzeniewski, B., McNeill, S., Thomas, K., Yin, X., Anthony, S., Ray, R., Vose, R.S., Gleason, B.E., et al., 2012a. Global Historical Climatology Network - Daily (GHCN-Daily), Version 3 (NOAA National Climatic Data Center).

Menne, M.J., Durre, I., Vose, R.S., Gleason, B.E., and Houston, T.G., 2012b. An Overview of the Global Historical Climatology Network-Daily Database. J. Atmospheric Ocean. Technol., 29:897–910.

Nicholson, S.E., 1996. A review of climate dynamics and climate variability in Eastern Africa. In The Limnology, Climatology and Paleoclimatology of the East African Lakes; Edited by: Johnson, T.C. and Odada, E.O., Gordon and Breach, (Amsterdam, the Netherlands: Gordon and Breach), pp. 25–56.

Nicholson, S.E., and Entekhabi, D., 1986. The quasi-periodic behavior of rainfall variability in Africa and its relationship to the southern oscillation. Arch. Meteorol. Geophys. Bioclimatol. Ser. A, 34:311–348.

Nyeko-Ogiramoi, P., Willems, P., and Ngirane-Katashaya, G., 2013. Trend and variability in observed hydrometeorological extremes in the Lake Victoria basin. J. Hydrol., 489:56–73.

Ogallo, L.J., 1988. Relationships between seasonal rainfall in East Africa and the Southern Oscillation. J. Climatol., 8:31–43.

van Oldenborgh, G.J., te Raa, L.A., Dijkstra, H.A., and Philip, S.Y., 2009. Frequency- or amplitude-dependent effects of the Atlantic meridional overturning on the tropical Pacific Ocean. Ocean Sci., 5:293–301.

Onyutha, C., 2016a. Statistical analyses of potential evapotranspiration changes over the period 1930-2012 in the Nile River riparian countries. Agric. For. Meteorol., 226–227C:80–95.

Onyutha, C., 2016b. Variability of seasonal and annual rainfall in the River Nile riparian countries and possible linkages to ocean–atmosphere interactions. Hydrol. Res., 47:171–184.

Onyutha, C., 2016c. Influence of hydrological model selection on simulation of moderate and extreme flow events: A case study of the Blue Nile Basin. Adv. Meteorol., 2016, 1–28:Article ID 7148326.

Onyutha, C., 2016d. Statistical uncertainty in hydro-meteorological trend analyses. Adv. Meteorol., in press, first online on 8 September 2016 at https://www.hindawi.com/journals/amete/aip/8701617/

Onyutha, C., and Willems, P., 2015a. Spatial and temporal variability of rainfall in the Nile Basin. Hydrol. Earth Syst. Sci., 19:2227–2246.

Onyutha, C., and Willems, P., 2015b. Uncertainty in calibrating generalised Pareto distribution to rainfall extremes in Lake Victoria basin. Hydrol. Res., 46:356–376.

Phillips, J., and McIntyre, B., 2000. ENSO and interannual rainfall variability in Uganda: implications for agricultural management. Int. J. Climatol., 20:171–182.

Richman, M.B., 1986. Rotation of principal components. J. Climatol., 6:293–335.

Santos, J.F., Pulido-Calvo, I., and Portela, M.M., 2010. Spatial and temporal variability of droughts in Portugal. Water Resour. Res., 46:W03503.

Sheffield, J., Goteti, G., and Wood, E.F., 2006. Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. J. Clim., 19:3088–3111.

Taye, M.T., and Willems, P., 2012. Temporal variability of hydroclimatic extremes in the Blue Nile basin. Water Resour. Res., 48:W03513.

Tierney, J.E., Smerdon, J.E., Anchukaitis, K.J., and Seager, R., 2013. Multidecadal variability in East African hydroclimate controlled by the Indian Ocean. Nature, 493:389–392.

Wallace, J.M., Smith, C., and Bretherton, C.S., 1992. Singular value decomposition of wintertime sea-surface-temperature and 500-mb height anomalies. J. Clim., 5:561–576.

Yeshanew, A., and Jury, M.R., 2007. North African climate variability. Part 1: Tropical thermocline coupling. Theor. Appl. Climatol., 89:25–36.

Zeng, R., and Cai, X., 2016. Climatic and terrestrial storage control on evapotranspiration temporal variability: Analysis of river basins around the world. Geophys. Res. Lett., 43:185–195.

Zuo, J., Ren, H.-L., Wu, J., Nie, Y., and Li, Q., 2016. Subseasonal variability and predictability of the Arctic Oscillation/North Atlantic Oscillation in BCC_AGCM2.2. Dyn. Atmospheres Oceans, 75:33–45.

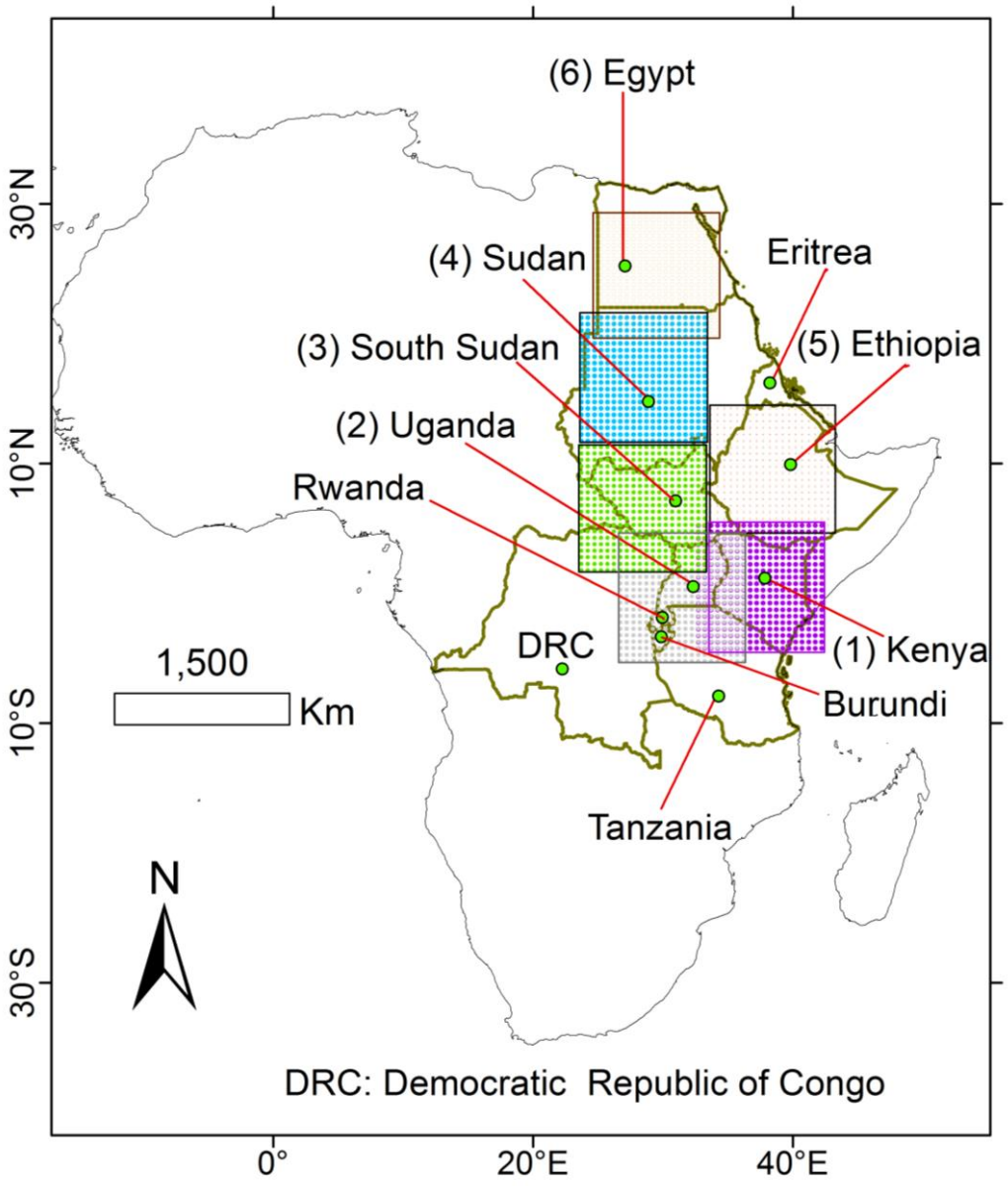


Figure 1 The River Nile riparian countries. The number in "( )" denotes this paper's ID of the country in which the center of the 10°×10° rainfall grids is located.

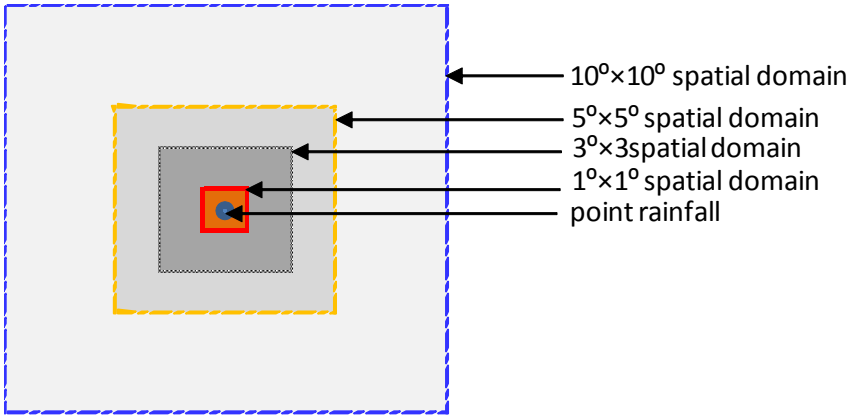


Figure 2 Spatial domains of the rainfall series considered for analyses.

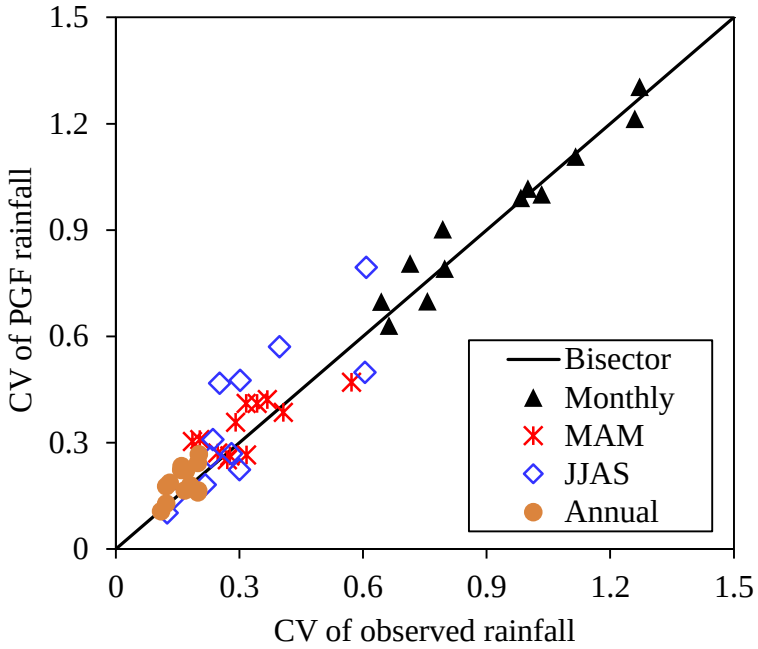


Figure 3 Coefficient of variation (CV) of observed rainfall *versus* that of PGF gridded series.

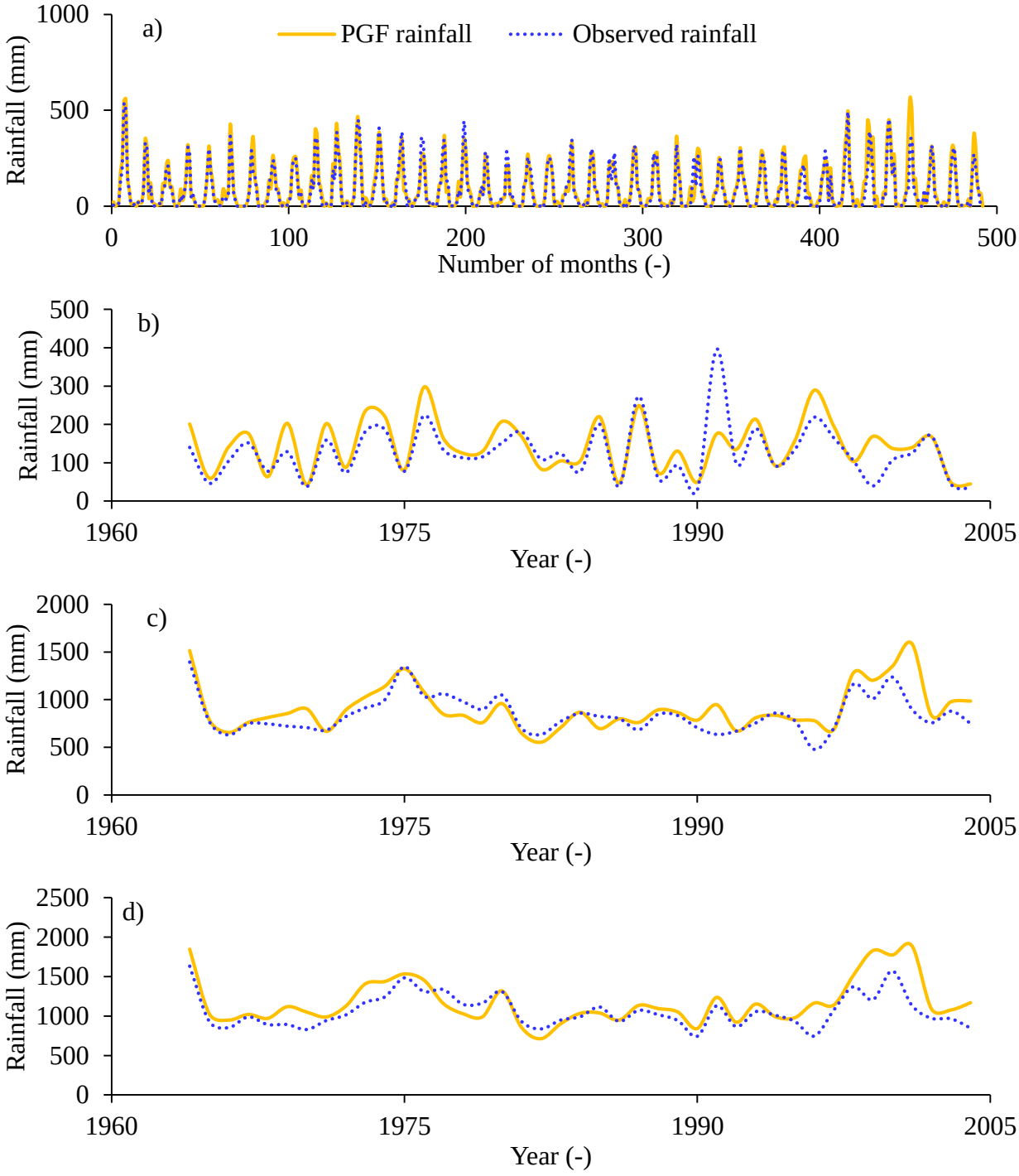


Figure 4 Comparison of observed and PGF rainfall based on a) monthly, b) MAM, c) JJAS and d) annual series at Station 8. All the charts share the same legend as in a).

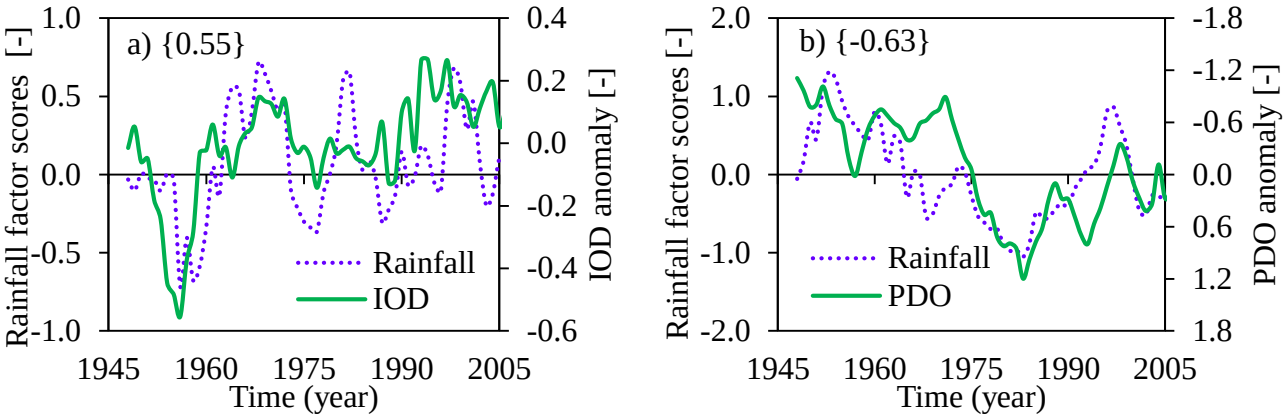


Figure 5 Variability in climate indices (IOD or PDO) and the EOF factor scores obtained from the annual rainfall based on $10^0 \times 10^0$ spatial domain with its center located in a) Kenya and b) Ethiopia. The label in "{ }" shows the correlation between the anomalies of rainfall and those obtained using the climate indices.

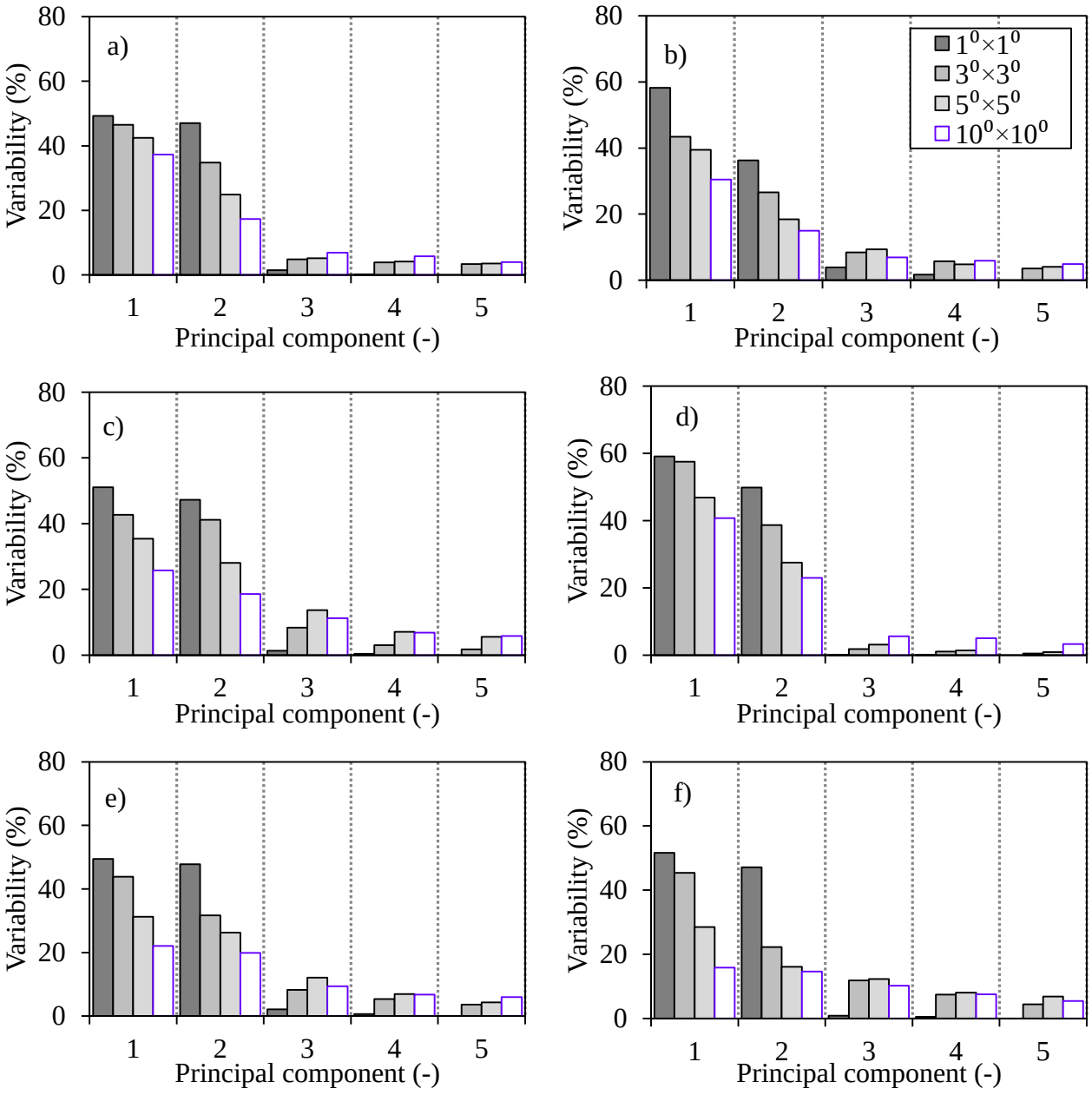


Figure 6 Percentage of variability explained in the annual rainfall from region a) 1, b) 2, c) 3, d) 4, e) 5, and f) 6. The charts share the same legend as in b). The numbering of the regions is consistent with the IDs assigned on **Figure 1**.

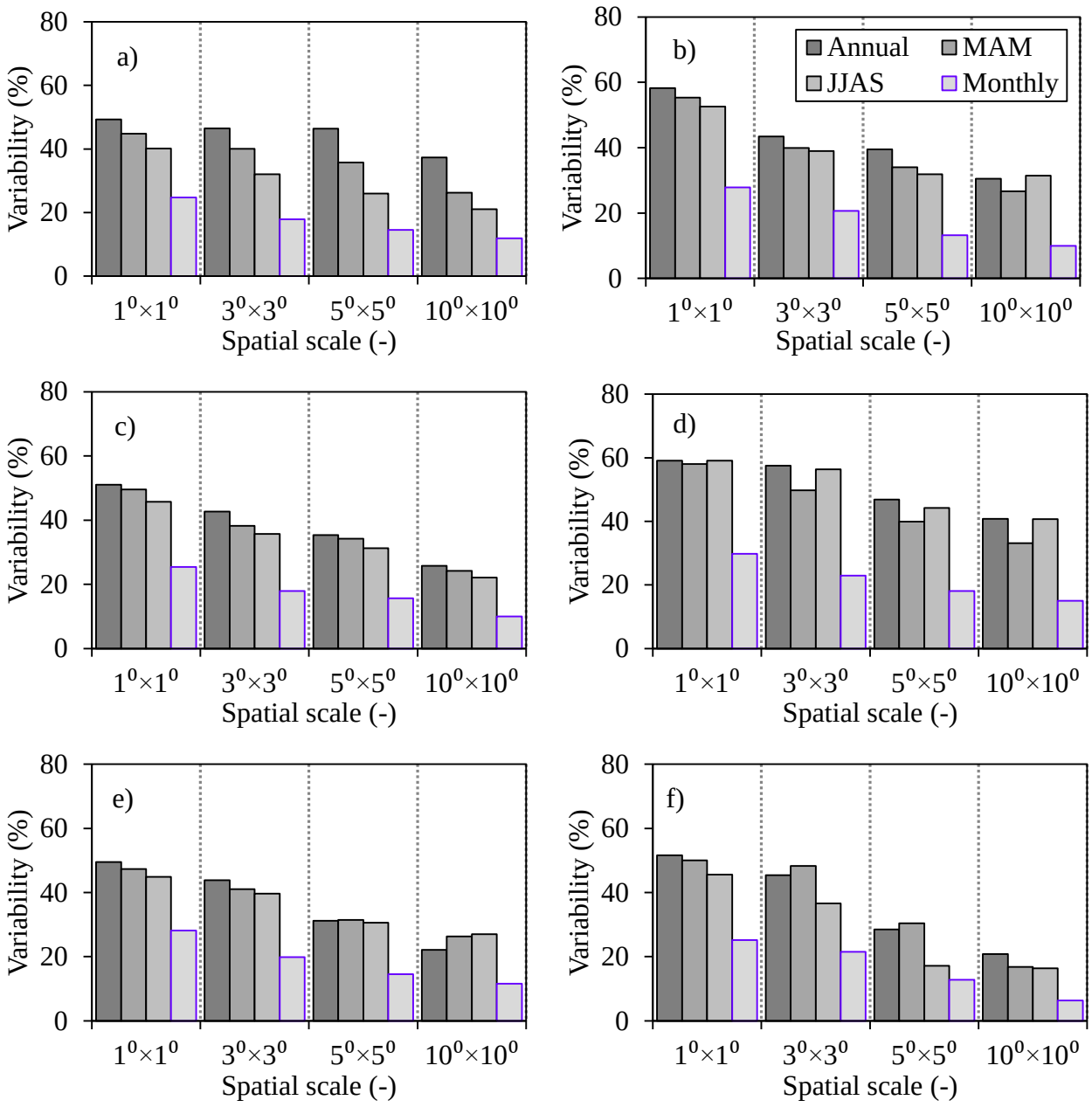


Figure 7 Percentage of variability explained by the first principal component based on the rainfall for region a) 1, b) 2, c) 3, d) 4, e) 5, and f) 6. The charts share the same legend as in b).

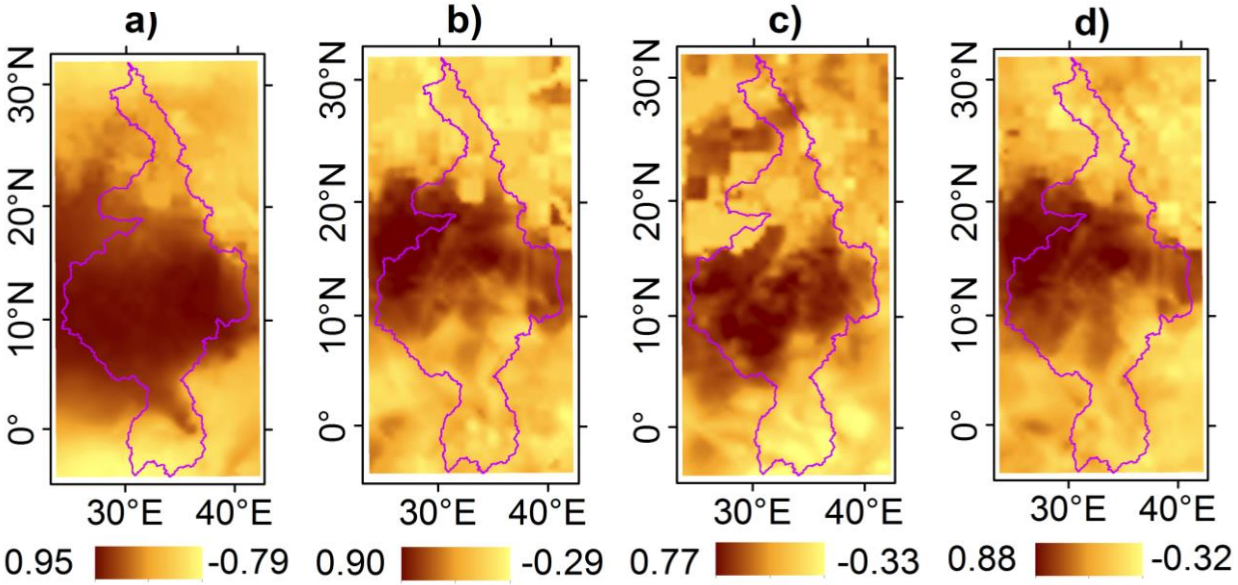


Figure 8 Leading EOF factor loadings after Varimax rotation for 0.5°×0.5° gridded series extracted over the spatial domain covering the entire Nile basin based on a) monthly, b) JJAS, c) MAM, and d) annual rainfall. The shape of the Nile Basin is superimposed on the spatial map.

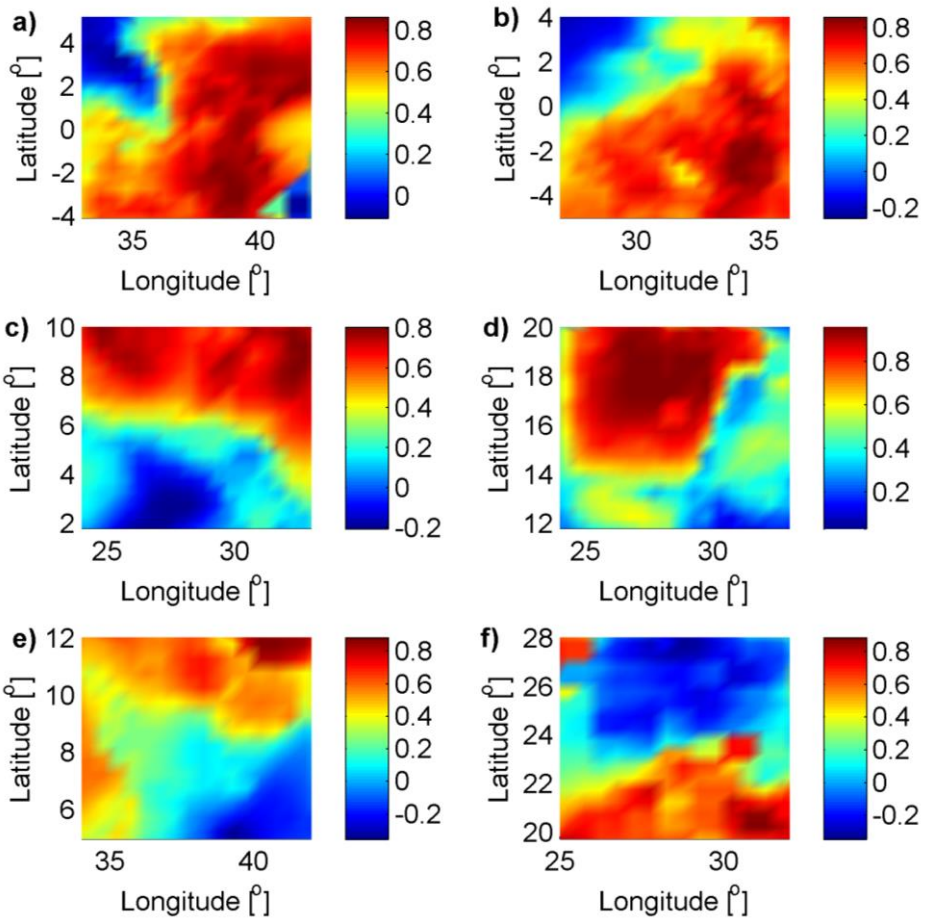


Figure 9 Leading EOF factor loadings after Varimax rotation for annual rainfall from the various grid points of the 10°×10° spatial domain from region a) 1, b) 2, c) 3, d) 4, e) 5, and f) 6.

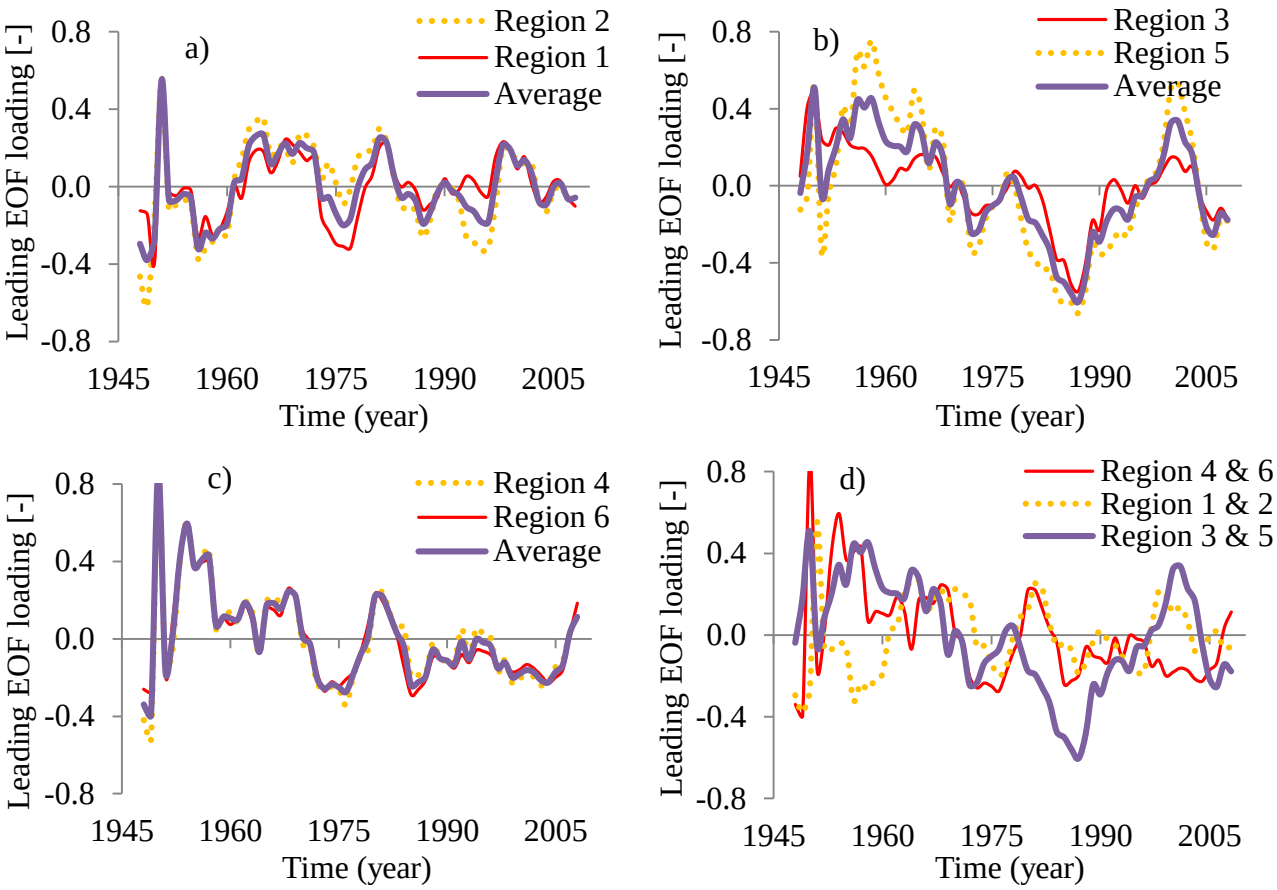


Figure 10 Temporal leading EOF factor loadings based on 10°×10° gridded annual rainfall from regions a) 1 and 2, b) 3 and 4, and c) 4 and 6. Chart d) is for the average loadings of the various regions.

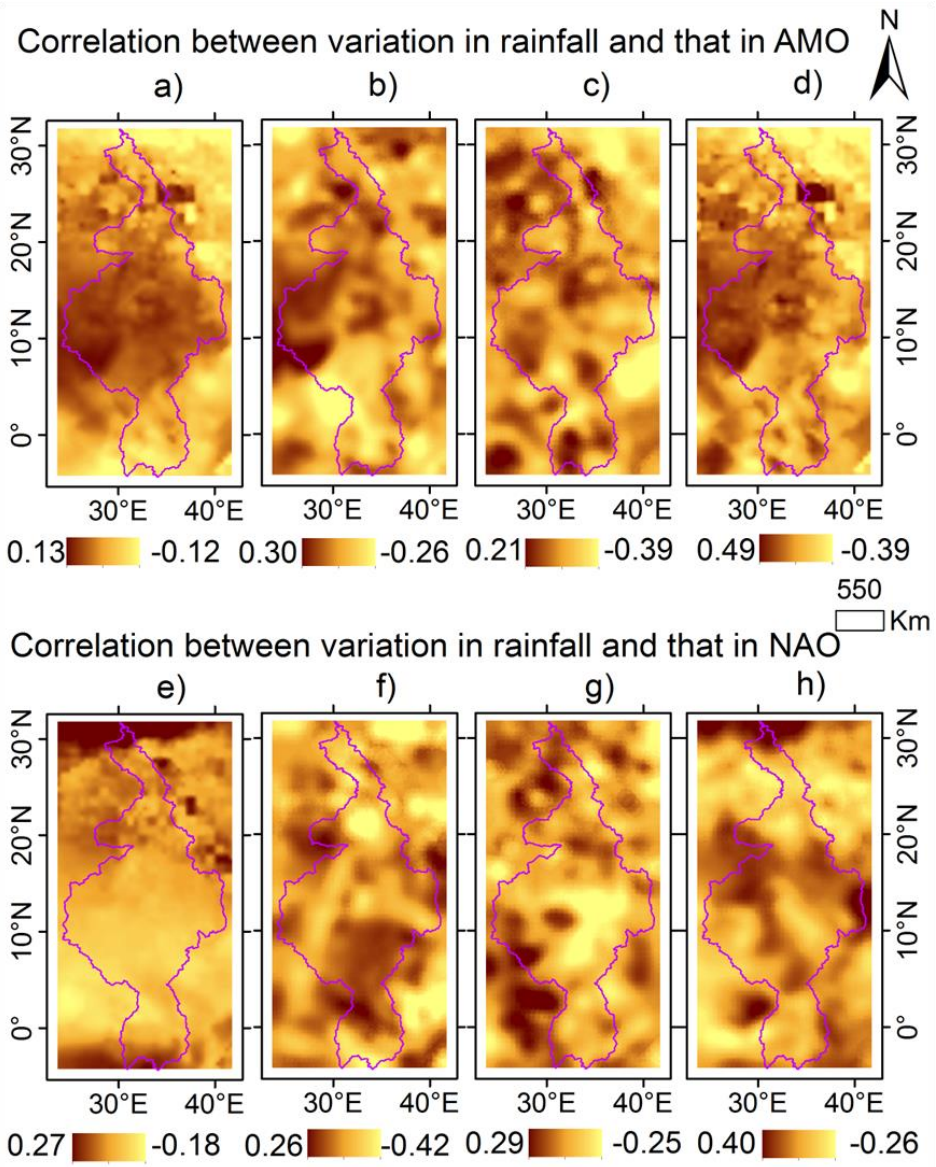


Figure 11 Co-occurrence of the variability in rainfall and those of (a-d) AMO, and (e-h) NAO based on (a, e) monthly, (b, f) JJAS, (c, g) MAM, and (d, h) annual time scales.

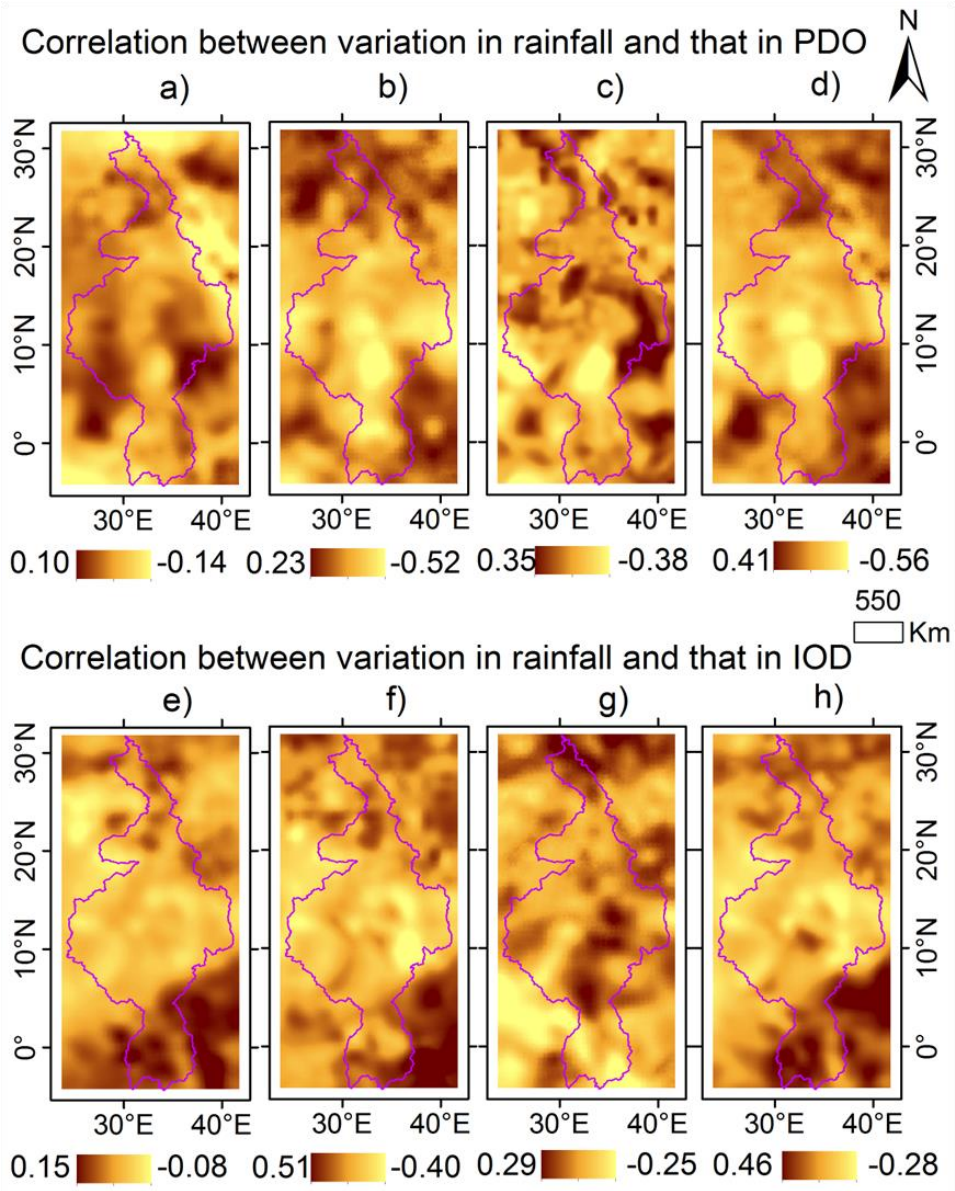


Figure 12 Co-occurrence of the variability in rainfall and those of (a-d) PDO and (e-h) IOD based on (a, e) monthly, (b, f) JJAS, (c, g) MAM, and (d, h) annual time scales.

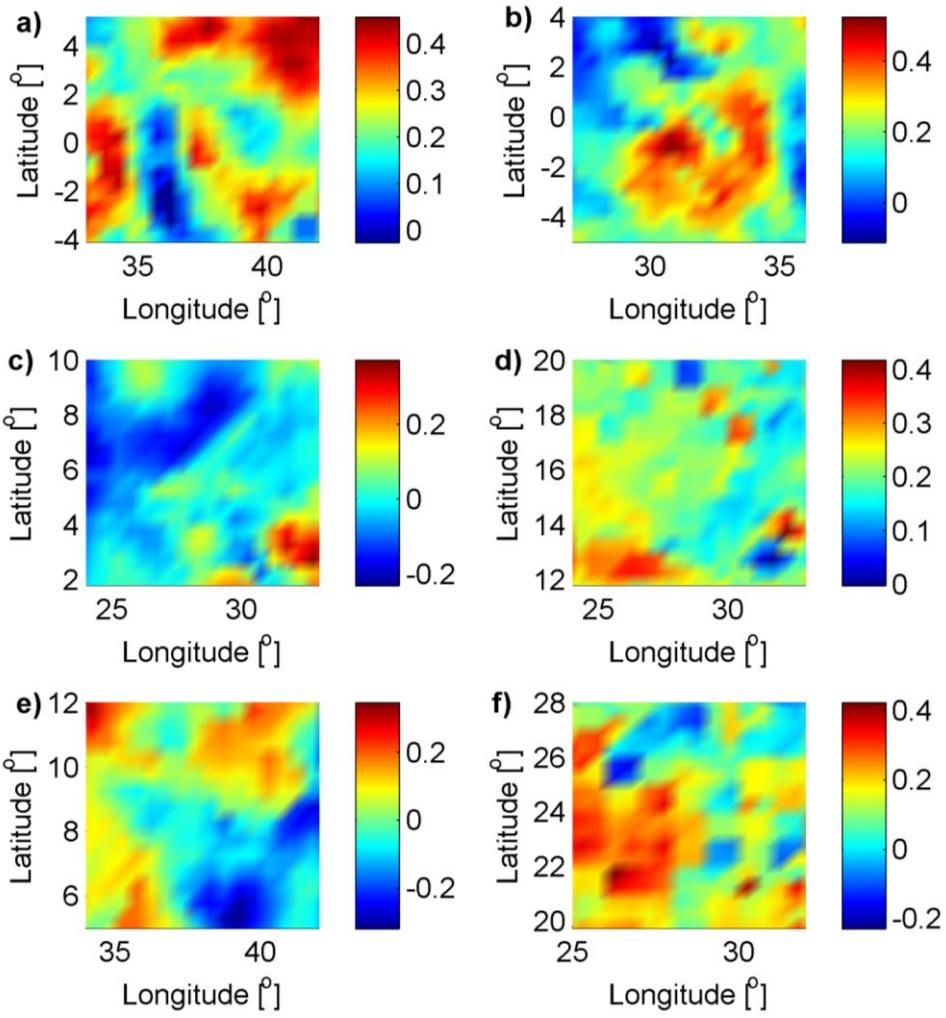
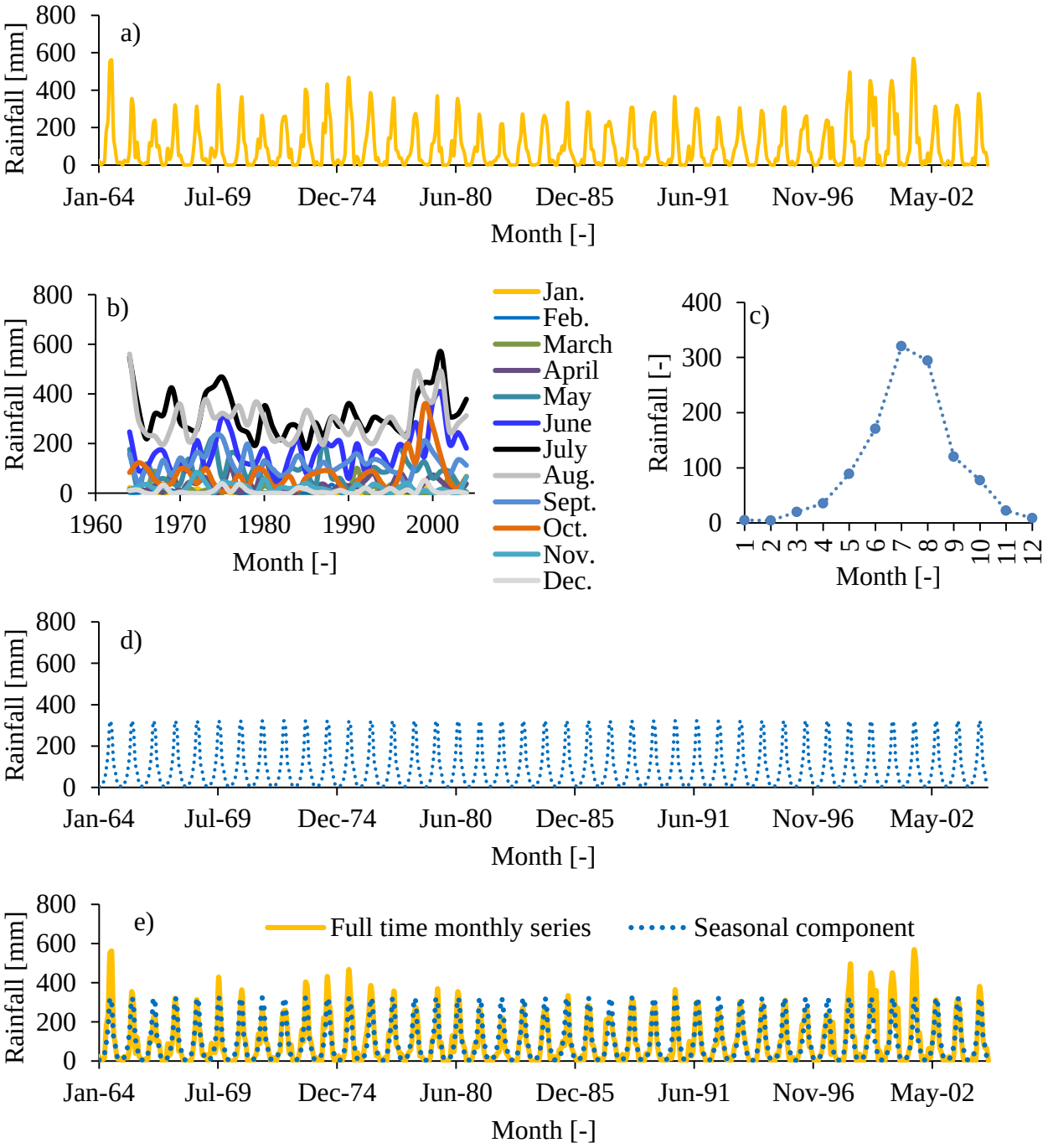


Figure 13 Correlation between anomalies in climate indices and the variation in leading EOF factor scores based on 10°×10° gridded annual rainfall of region a) 1, b) 2, c) 3, d) 4, e) 5, and f) 6.



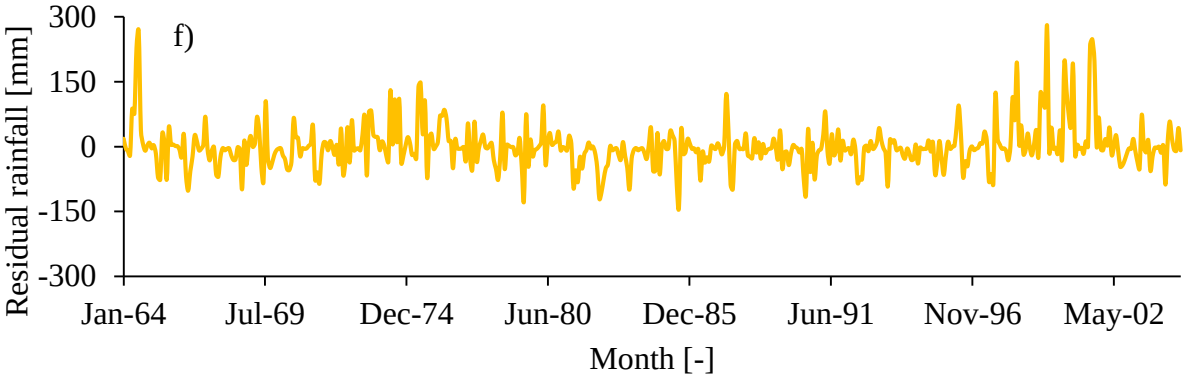


Figure A1 Removal of seasonal component of monthly rainfall series

Table 1 Extents of spatial domains for rainfall series

Spatial scale	Latitude [°]		Longitude [°]		Latitude [°]		Longitude [°]		Latitude [°]		Longitude [°]	
	From	To	From	To	From	To	From	To	From	To	From	To
	Region 1				Region 2				Region 3			
10°×10°	-4.5	5.5	32.5	42.5	-5.5	4.5	26.5	36.5	1.5	11.5	23.5	33.5
5°×5°	-2.0	3.0	35.0	40.0	-3.0	2.0	29.0	34.0	4.0	9.0	26.0	31.0
3°×3°	-1.0	2.0	36.0	39.0	-2.0	1.0	30.0	33.0	5.0	8.0	27.0	30.0
1°×1°	0.0	1.0	37.0	38.0	-1.0	0.0	31.0	32.0	6.0	7.0	28.0	29.0
Point	0.5	0.5	37.5	37.5	-0.5	-0.5	31.5	31.5	6.5	6.5	28.5	28.5
	Region 4				Region 5				Region 6			
10°×10°	11.5	21.5	23.5	33.5	4.5	14.5	33.5	43.5	19.5	29.5	24.5	34.5
5°×5°	14.0	19.0	26.0	31.0	7.0	12.0	36.0	41.0	22.0	27.0	27.0	32.0
3°×3°	15.0	18.0	27.0	30.0	8.0	11.0	37.0	40.0	23.0	26.0	28.0	31.0
1°×1°	16.0	17.0	28.0	29.0	9.0	10.0	38.0	39.0	24.0	25.0	29.0	30.0
Point	16.5	16.5	28.5	28.5	9.5	9.5	38.5	38.5	24.5	24.5	29.5	29.5

The numbers 1-6 after 'Region' are consistent with the country IDs in **Figure 1**.

Table 2 Correlation between PGF series and observed rainfall at selected stations

SNo.	Station name	Period		Location		Correlation for various time scales			
		From	To	Long.	Lat.	Monthly	MAM	JJAS	Annual
1	Kamenyamigo	1961	1990	31.67	-0.30	0.62	**0.21**	**0.15**	0.38
2	Maswa Hydromet.	1961	1990	33.77	-3.17	0.89	0.74	0.60	0.64
3	Nyabassi	1961	1990	34.57	-1.35	0.64	0.60	**0.04**	0.53
4	Rwoho Forest	1961	1990	30.55	-0.85	0.61	0.40	**0.27**	0.37
5	Mbarara Met.	1948	2008	30.60	0.56	0.59	**0.06**	0.47	0.31
6	Bahr Dar	1964	2004	37.42	11.60	0.91	0.68	0.44	0.51
7	Debre Marcos	1964	2004	37.67	10.33	0.96	0.85	0.67	0.73
8	Gonder	1964	2004	37.42	12.55	0.94	0.76	0.77	0.77
9	Addis Ababa	1964	2004	38.75	9.03	0.96	0.87	0.69	0.71
10	Kombolcha	1964	2004	39.73	11.12	0.91	0.75	0.79	0.75
Bold values indicate H_0 (no correlation) was rejected at the significance level of 5%.		Critical values of the correlation at the significance level of 5%							
		Rainfall at Stations 1 to 4				0.09	0.35	0.35	0.35
		Rainfall at Station 5				0.07	0.24	0.24	0.24
		Rainfall at Stations 6 to 10				0.08	0.31	0.31	0.31

Table 3 Correlation between rainfall variability and the variation in selected climate indices

Region	Country	Point	1°×1°	3°×3°	5°×5°	10°×10°	Point	1°×1°	3°×3°	5°×5°	10°×10°
		Monthly time scale {0.08}					MAM seasonal time scale {0.25}				
1	Kenya	0.29	0.30	0.33	0.39	0.46	**-0.14**	**-0.11**	**-0.22**	0.25	**0.08**
2	Uganda	0.24	0.25	0.29	0.35	0.42	-0.27	**-0.05**	-0.46	**-0.15**	**0.13**
3	South Sudan	-0.13	-0.16	-0.19	-0.23	-0.29	-0.19	0.48	-0.63	-0.61	-0.48
4	Sudan	-0.10	-0.14	-0.16	-0.18	-0.19	0.00	0.00	-0.27	**-0.20**	**-0.04**
5	Ethiopia	0.05	-0.15	-0.18	-0.27	-0.33	0.67	-0.18	0.37	0.58	0.48
6	Egypt	**-0.03**	0.09	**0.05**	0.16	0.27	**0.02**	**0.10**	**0.06**	0.25	0.29
		JJAS seasonal time scale {0.25}					Annual time scale {0.25}				
1	Kenya	0.57	0.63	0.29	0.73	0.79	0.25	0.29	0.43	0.48	0.55
2	Uganda	0.30	0.34	0.38	0.43	0.44	**0.13**	0.31	0.34	0.40	0.43
3	South Sudan	**-0.14**	**-0.15**	-0.28	-0.37	-0.43	-0.45	**-0.15**	**-0.05**	-0.58	-0.71
4	Sudan	**0.13**	**0.02**	**-0.01**	0.35	0.43	**0.13**	**0.19**	0.27	0.36	0.49
5	Ethiopia	-0.35	-0.33	-0.53	-0.68	-0.72	0.12	-0.19	0.33	0.36	-0.63
6	Egypt	**-0.12**	**-0.04**	**0.04**	0.25	0.32	**-0.10**	**-0.08**	**0.12**	0.26	0.40

- Point: refers to the point location at the centre of the 1°×1° grid.
- Enclosed in "{ }" is the critical correlation value at the significance level of 5%.
- Bold values denote that the H_0 (no correlation) was rejected at the significance level of 5%.
- The climate indices IOD, PDO and AMO were used for the correlation values of regions 1 and 2, 3 and 5, and 4 and 6 respectively.