International Journal of Statistics: Advances in Theory and Applications
Vol. 1, Issue 1, 2020, Pages 121-141
Published Online on September 14, 2019
© 2020 Jyoti Academic Press
http://jyotiacademicpress.org

EMPIRICAL ORTHOGONAL FUNCTION (EOF) ANALYSIS OF PRECIPITATION OVER GHANA

DORCAS A. ADDO[1], FRANCIS T. ODURO[1]
and RICHARD K. ANSAH[2]

[1]Department of Mathematics
Kwame Nkrumah University of Science and Technology
Kumasi
Ghana
e-mail: dorcas@aims.edu.gh
ftoduro@gmail.com

[2]Department of Mathematics and Statistics
University of Energy and Natural Resources
Sunyani
Ghana
e-mail: richard.ansah@uenr.edu.gh

Abstract

Climate change has serious implications on food security and livelihood in West Africa since it modulates the socio-economic activities in the sub region. The inter-decadal, seasonal, and inter-annual precipitation variability over Ghana has been studied and their spatial and temporal patterns analysed using empirical orthogonal function approach with a view to detecting climate change. A secondary monthly mean precipitation space-time series data from 1970-2010 from Climate Research Unit (CRU) at a high-resolution of 0.5×0.5 has been used in this analysis. Empirical Orthogonal Function (EOF) analysis

2020 Mathematics Subject Classification: 00A69.
Keywords and phrases: empirical orthogonal function, principal component, precipitation.
Received February 24, 2017; Revised July 12, 2017

was used to study the spatial and temporal trend of monthly mean precipitation and temperature over Ghana. A decadal and inter-annual variation of the spatial and temporal patterns were observed respectively. The temporal analysis showed a decrease of precipitation.

1. Introduction

Climate change in recent times has gained more attention and the discussion is on globally. This has been due a general belief of degradation in weather patterns due to the impact it has shown. The concept of global warming has also contributed to the debate of climate change over the years. Most continents especially Africa has become unsafe these effects of change in climate due to the over-dependence of rain.

The influence of various activities by humans on climate has become visible and calls for actionable, particular and lofty-resolution of zonal and information regionally concerning future climate. Decision-makers, scientist and policy-makers, as well as those assessing climate change impact require information for observations. It has therefore become eminent for these people to know more about the earth's happening to find remedy and prepare for the practical actions to undertake with climate change. The part of our planet has experienced simultaneous changes in the climate which has come along with rapid climate data growth. These changes in the climate has raised concerns of significant scientific, economic, and societal interest all over the world and the most discussed.

The change in climate over the years has given the need to have accurate information about the temporal and spatial variations of rainfall (precipitation), temperature, and humidity at the earth's surface. There are different ways of estimating these trends and variability, and also finding out if there has been any significant change over these decades. Researchers in mathematics, statistics and other field of study play a very important role in this area by developing very effective models in studying the climate. Ghana took part to the debate of climate change when Ghana agreed to sign the United Nations Framework Convention on Climate Change (UNFCCC) which was in June 1992, after the adoption in May 9, 1992 (Owusu et al. [11]) at the Rio de Earth Summit.

Climate variability is of interest to the scientific community in the study of the atmosphere involving its description and analysis of the variable (precipitation, temperature, humidity, and among others) and also how other artificial factors also contribute to the study. The focus on climate change has been to analyse certain trends in the regional and global average of climate variables and accessing of necessary trends or variability in the future and detecting climate change.

Hansen et al. [4] said, change in the conditions of the weather may have important repercussion for temperature and rainfall. Climate change on precipitation and temperature could be divided into spatial trends at regional scales (Nam et al. [8]). The regional climate change do require findings of correlated variations covering space (i.e., geographical variations) and time (period over which the change occurs). Owusu and Waylen [10] reported a general shift of in season thus decrease in rainfall as it disappears from the wet to the dry season.

It is fundamental to fully understand the environmental assessments at regional scales to fully comprehend the spatial and temporal patterns in the variability of climate. Different mathematical and statistical tools have been used in studying this spatial and temporal pattern. Gaussian processes are normally used in modelling of data in space and time. It is defined by specifying a covariance and mean function.

These variations in the climate are due to complicated non-linear reciprocal action of degrees of modes or freedom. The challenge has been to reduce the dimensionality and find patterns or trend in explaining this variability. The empirical orthogonal function (EOF) used by meteorologist has seen progress as a major technique in explaining the variability in the climate. EOF analysis is tool efficient for pattern recognition.

EOF examines a space-time data set observation as it reduces it to spatial patterns mostly known as EOFs, which explains most of the variance in the data and the temporal patterns also called PCs. In this work, EOFs can be performed either by diagonalizing the covariance grid, from the eigenvectors and eigenvalues which explains the spatial and temporoal patterns. The spatial examination from the EOF lessens the information to an imperative modes which clarifies the fragmentary rates of fluctuation, whiles that of the fleeting EOF investigation clarifies transient difference in the information set.

Research in the zone of remote detecting of the seas, in warm infrared (Lagerloef and Bernstein [5]) and in the unmistakable (Eslinger et al. [1]) utilized EOF investigation as a part of watching the spatial and temporal examples. Regardless of the intriguing results acquired in the utilization of EOFs, issues have emerged which is still in exchange to move forward. The vulnerabilities in the physical interpretations of EOF modes with the spatial pattern and ambiguities in distinguishing relationships amongst spatial and temporal EOFs are the issues encountered in understanding of EOF analysis.

EOF generally have been grouped according to their statistical significance which is useful in describing the significance of each mode but difficult in explaining the physics describing each mode. Most researchers and authors have not been able to provide an approach in describing the patterns developed by EOF. The selection of the modes have basically been based on statistical approach, although one has not explored their limitations. Distinguishing the relationship between the spatial and temporal difference (variance) is an issue in utilizing EOF analysis. The understanding of the patterns relies on upon how the covariance matrix is built. Gallaudet and Simpson [3], in their investigation of five years time series of SST images, for the spatial and temporal, performing both the spatial and temporal EOF of almost five years observed the patterns to be comparative. This is regular when an investigation of long time series of satellite images are seen with EOF decomposition which causes interpretation problems.

2. Main Objective

The main objective is to use empirical orthogonal function to extract the spatial and temporal pattern in the variable.

3. Empirical Orthogonal Function (EOF)

3.1. Structure of the data

The data considered for our study is a gridded data for precipitation which normally come as an array of three-dimensional, two-dimensional in space, defining the spatial structure (that is, values at individual grid points) and one-dimensional for each vertical level of a rectangular scalar field F as a function of time t, latitude θ, and longitude φ. Assuming our horizontal coordinates are discretised to yield latitudes θ_j, $j = 1, \cdots, p_1$, longitudes φ_k, $k = 1, \cdots, p_2$, and similarly for time, i.e., t_i, $i = 1, \cdots, n$. This yields a total number of grid points $p = p_1 p_2$. The discretised field reads:

$$F_{ijk} = F(t_i, \theta_j, \varphi_k),$$

with $1 \leq i \leq n$, $1 \leq j \leq p_1$, and $1 \leq k \leq p_2$. We transform the field into a two-dimensional array called the data matrix X where the two spatial dimensions (longitude and latitude) are linked together. Thus, the EOF is performed using the scalar field X as a data matrix.

If our gridded data set consist of a space-time field $\mathbf{X}(t, s)$ which represent the value of the field F, such as precipitation, at time t and spatial positions. Therefore, the value of the field at a discretised time t_i and a grid point s_j is expressed as x_{ij} for $i = 1, \cdots, n$ and $j = 1, \cdots, p = p_1 p_2$. The data matrix of the observed field can be represented as:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}.$$

Considering the average time $\bar{x}._j$ of the field at the j-th grid point, thus

$$\bar{x}._j = \frac{1}{n}\sum_{k=1} x_{kj}. \tag{1}$$

Climatology of the field is therefore defined by

$$\bar{x} = (\bar{x}._1, \cdots, \bar{x}._p). \tag{2}$$

In application, our interest will be detect the variability about the climatological mean. Thus find the difference between the original data and the climatological mean field, which is called the anomaly. Anomaly data often are indicated by a prime. Thus, the anomaly field of the defined climatology at (t, s) is

$$x'_{ts} = x_{ts} - \bar{x}._s, \tag{3}$$

and in a matrix form as:

$$X' = X - \mathbf{1}\bar{\mathbf{x}} = (I - \frac{1}{n}\mathbf{1}\mathbf{1}^T)X. \tag{4}$$

3.2. Creating weights in area

Due to the non-uniform distribution of data (either simulated or observed) over the surface of the earth, as it could be densely populated towards one end. The computed structure of the EOFs obtained could be influenced due to the non-uniformity. The data is therefore normally weighed before analyzing. The most efficient way has been to weigh the data point by its local location area. Thus each of the datum is weighed by the cosine of its latitude. Considering the latitude θ_n of the n-th grid point, $n = 1, \ldots, k$ and the diagonal matrix D_θ :

$$D_\theta = \mathrm{Diag}[\cos\theta_1, \cdots, \cos\theta_p]. \tag{5}$$

Thus the weighted anomaly matrix will be

$$X_w = X'D_\theta. \tag{6}$$

3.3. Derivation of EOF

The EOF analysis allows us to explain the variance-covariance of the data through the modes of variability. The maximum modes accounts for the largest percent of the original variability are considered significant. These modes are represented by orthogonal spatial patterns (eigenvectors) and corresponding time series (principal components). We have two modes which are spatially and temporally uncorrelated due to the orthogonal nature of the EOF.

The representation of fields of climate variables in terms of its eigenvectors and eigenvalues has been seen to be formulated in various ways (Freiberger and Grenander [2], Loeve [6], and Lorenz [7]).

From the definition of the anomaly data matrix (6) and its weighted anomaly matrix (4), the covariance matrix can therefore be expressed as

$$\sum = \frac{1}{n-1} X'^T X'. \tag{7}$$

The aim is to express our matrix as linear combination of the M eigenvector that is the grid points explaining the maximum variance. The mathematical approach used in this work is related to Freiberger and Grenander [2].

If X_n represent an F component vector with n-th observation of $n = 1$ to N, of F variables.

X is an $N \times M$ matrix whose column $(n\text{-th})$ is the observation vector X_n. Thus, an n-th observation with m-th variable is denoted by X_{mn}. To determine which eigenvector $\mathbf{e}_i$ has the highest eigenvalue λ_i to all the observation vectors X, implies

$$\frac{(\mathbf{e}'X)^2 N^{-1}}{\mathbf{e}'\mathbf{e}}. \tag{8}$$

The eigenvectors define directions in the initial coordinate space along which the maximum possible variance can be expounded. And it is equivalent to maximizing

$$\mathbf{e}'R\mathbf{e}, \tag{9}$$

where the prime means transpose, and on condition

$$\mathbf{e}'\mathbf{e} = 1. \tag{10}$$

R is an $M \times M$ symmetric matrix and therefore diagonalizable of elements r_{ij}-th term which is expressed as

$$r_{ij} = N^{-1} \sum_{n=1}^{N} X_{in} X_{jn}, \tag{11}$$

or

$$R = N^{-1}(XX'). \tag{12}$$

Thus, the set of its eigenvectors will form an orthogonal basis defined by the natural scalar product of a p-dimensional Euclidean space. The empirical orthogonal functions will therefore be considered orthogonal and its corresponding PCs will be uncorrelated.

If we maximize Equation (9) with the condition in Equation (10) leads to the eigenvalue problem with the equation

$$R\mathbf{e} = \mathbf{e}\lambda, \tag{13}$$

λ is the eigenvalue associated with the symmetric matrix.

To show that $\mathbf{e}_i$ are orthogonal and λ_i are positive and real, Equation (13) can be written as

$$R\Upsilon = \Upsilon D, \tag{14}$$

Υ is an $M \times M$ orthogonal matrix where $\mathbf{e}_i$ the eigenvectors represents the columns. D is a diagonal $M \times M$ matrix with each element λ_i associated with $\mathbf{e}_i$.

This implies that

$$\Upsilon'\Upsilon = I. \tag{15}$$

From Equations (12), (14), and (15), the orthogonal matrices and transpose is equal to the inverse which results in

$$\Upsilon' XX' \Upsilon = DN. \tag{16}$$

Take

$$C = \Upsilon' D, \tag{17}$$

C is $M \times M$ matrix. Thus

$$X = \Upsilon C, \tag{18}$$

therefore

$$X_n = \sum_{i=1}^{M} C_{in}\mathbf{i}, \quad n = 1, \cdots, M, \tag{19}$$

expressed as a linear combination and C_{in} represents the coefficient associated with the i-th eigenvector for the n-th observation.

Substituting Equation (17) into Equation (16) gives

$$CC' = DN. \tag{20}$$

The row vectors of C are orthogonal. This means that its coefficients associated with them are also orthogonal.

The maximum variance associated with the highest eigenvalues of R can therefore be expressed in fraction as

$$V_p = \frac{\sum_{i=1}^{p} \lambda_i}{\sum_{i=1}^{M} \lambda_i}. \tag{21}$$

3.4. Significance test

Since EOF analysis is basically based on calculating the eigenvectors and eigenvalues of the covariance matrix, a significance test is needed. A selection rule is used on deciding which of the eigenvectors to maintain and which to discard.

There are different types of selection rules. Our work focuses on the dominance variance rule from North et al. [9] of Equation (24) used to calculate the significance physical modes of EOF. This is based on the size of eigenvalues (λ_i) or the amount of variance obtained since it physically calculates the most important eigenvalues.

$$\Delta\lambda_k \approx \lambda_k \sqrt{\frac{2}{n}}, \tag{22}$$

$$\Delta\mathbf{a}_k \approx \frac{\lambda_k}{\lambda_j - \lambda_k} \mathbf{a}_j, \tag{23}$$

where λ_j represent the closest eigenvalue to λ_k and n represent the number of independent data.

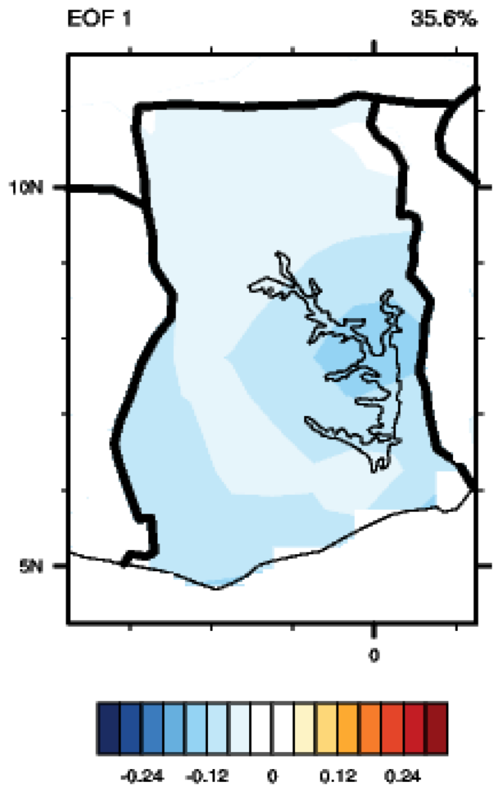
4. Results and Discussion

4.1. Introduction

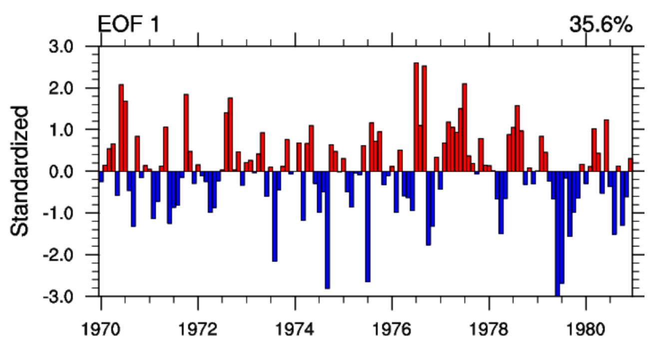
EOF was used to explore the structure of variability within a data set of monthly mean precipitation and to analyze the relationship within a set of variables. The eigenvector in EOF analysis shows the spatial variability as well as the variance. This method isolates the dominant mode of variability and decomposes it into spatial and temporal patterns. One major advantage in the use of EOF is that, it helps to identify the cause of variation in each component of data set. A major constraint of the EOF analysis is, the principal components are orthogonal in time and therefore there is no simultaneous temporal correlation between any two principal components. EOFs are orthogonal in space and show no spatial correlations between any two EOFs. The empirical orthogonal function (EOF) analysis was done on precipitation data to isolate the dominant modes of variability considering the spatial shift of the agro-ecological zones.

5. EOF Analysis of Decadal Variability

A decadal variability, 1970-1980, 1981-1990, 1991-2000, and 2001-2010 was analyzed in seeking the structures that explains the maximum amount of variance in a two-dimensional data set for precipitation.

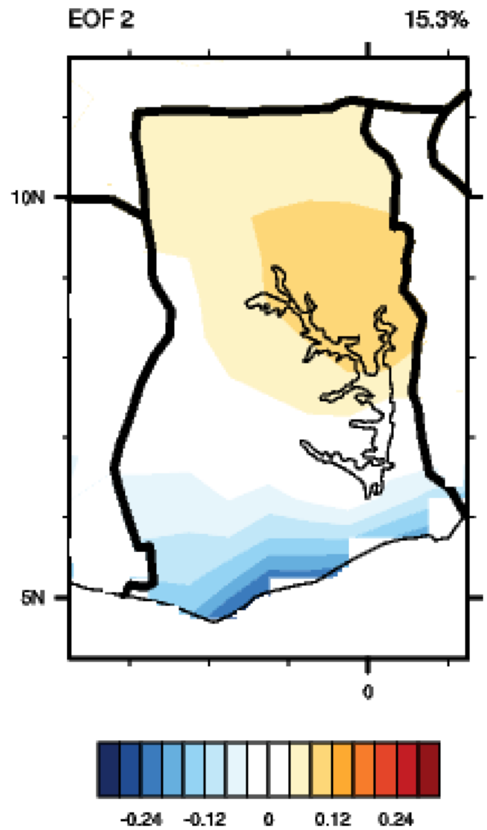


(a) EOF1

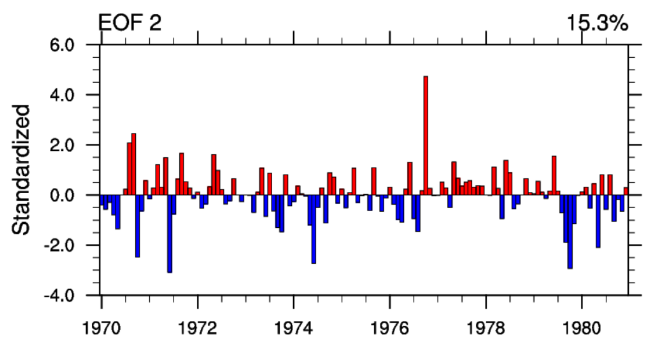


(b) PC1

Figure 1. Leading EOF and PC of the monthly mean precipitation from 1970-1980.

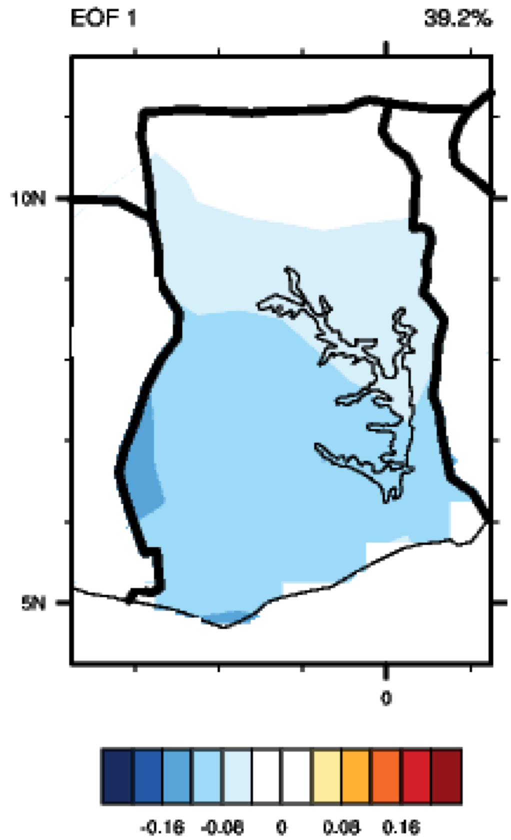


(a) EOF2

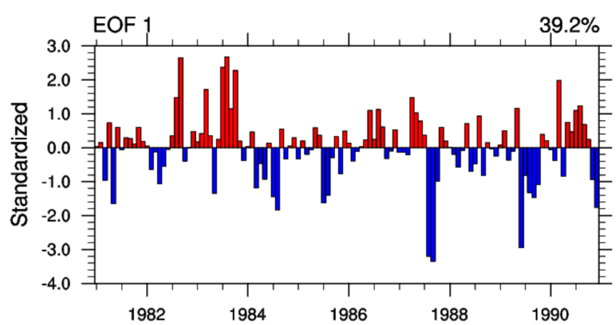


(b) PC2

Figure 2. The second EOF and PC of the monthly mean precipitation from 1970-1980.

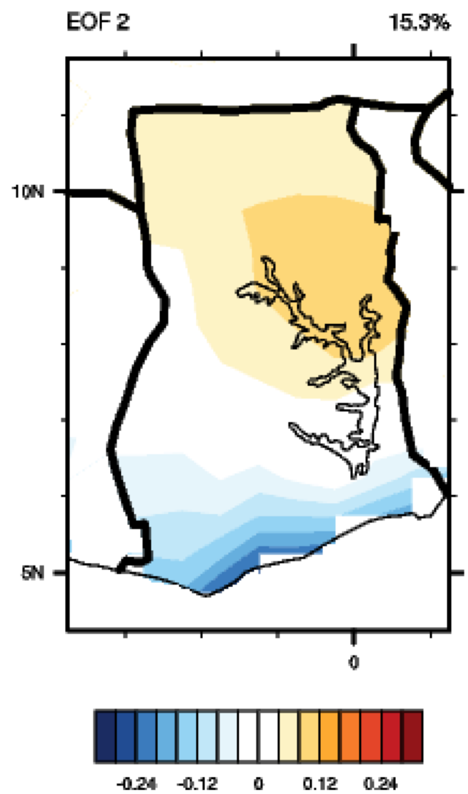


(a) EOF1

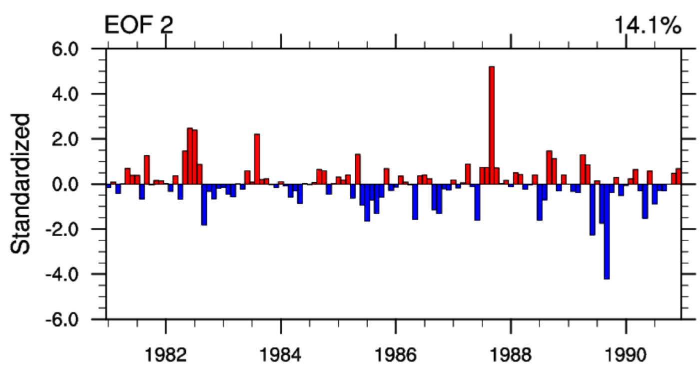


(b) PC1

Figure 3. Leading EOF and PC of the monthly mean precipitation from 1981-1990.

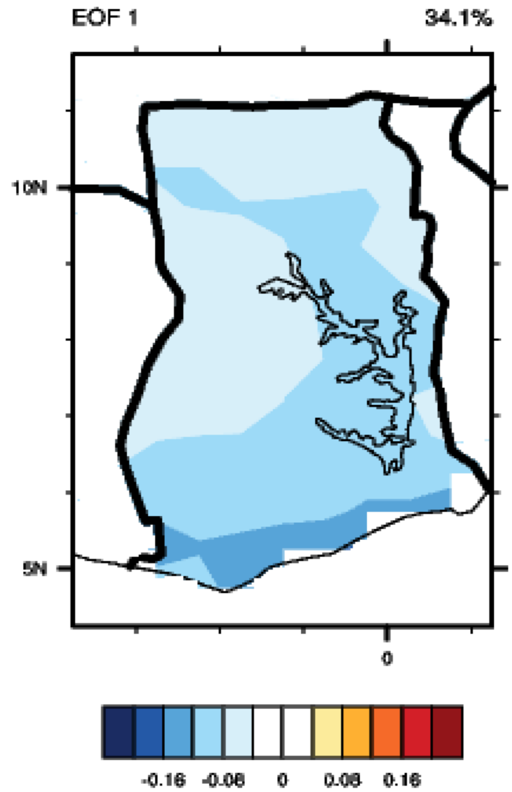


(a) EOF2

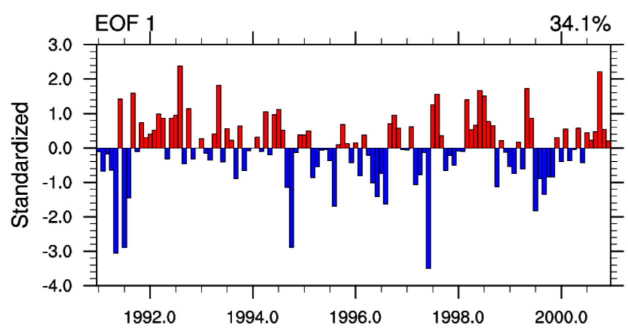


(b) PC2

Figure 4. The second EOF and PC of the monthly mean precipitation from 1981-1990.

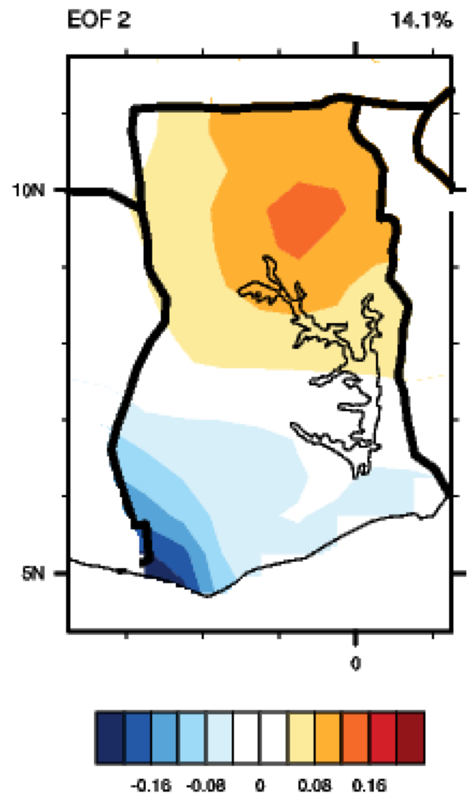


(a) EOF1

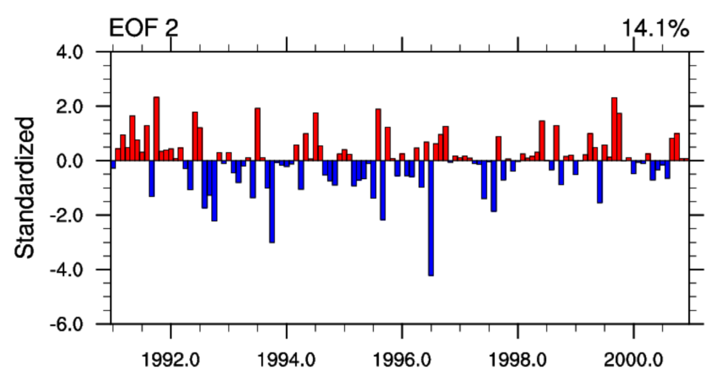


(b) PC1

Figure 5. Leading EOF and PC of the monthly mean precipitation from 1991-2000.

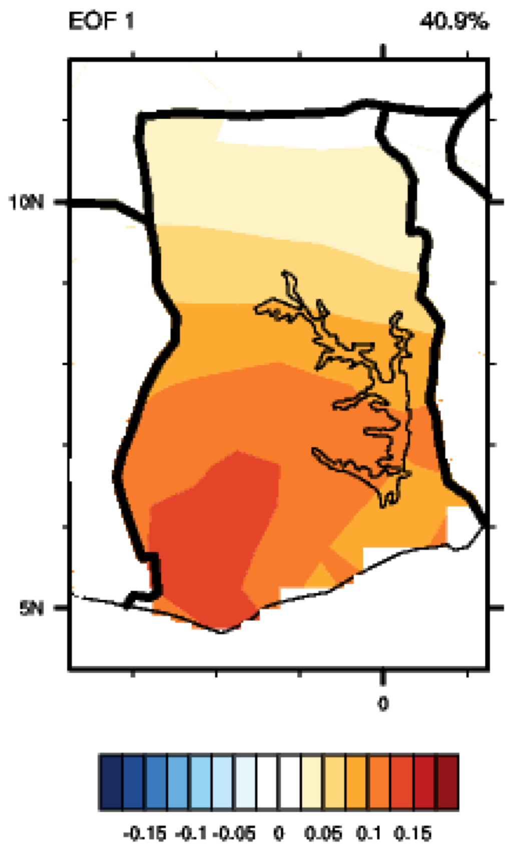


(a) EOF2

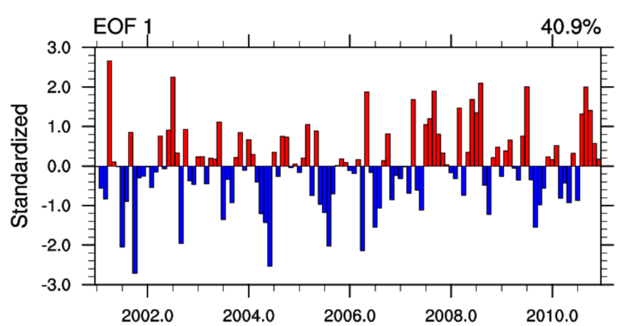


(b) PC2

Figure 6. The second EOF and PC of the monthly mean precipitation from 1991-2000.

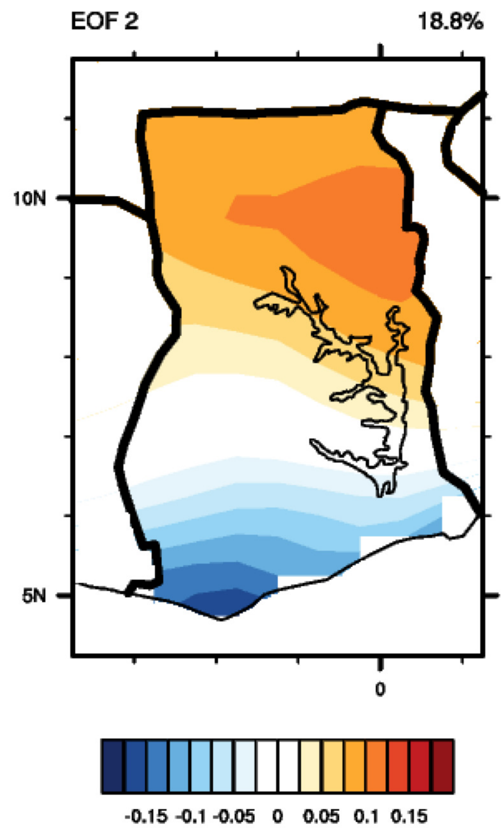


(a) EOF1

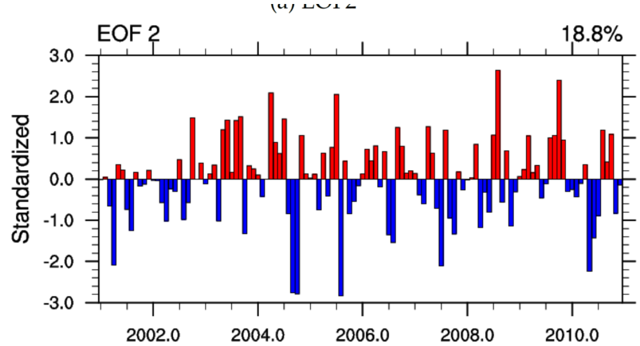


(b) PC1

Figure 7. Leading EOF and PC of the monthly mean precipitation from 2001-2010.



(a) EOF2



(b) PC2

Figure 8. The second EOF and PC of the monthly mean precipitation from 2001-2010.

EOF analysis was performed on monthly mean precipitation data into spatial and temporal loadings explaining the maximum variance. The first two leading modes accounted for 35.6% and 15.3%, respectively, as in Figure 2. The spatial pattern in EOF1 indicated below average precipitation conditions over all the ecological zones and EOF2 showed above average conditions over the Savannah and other zones except for the Coastal that showed below average precipitation conditions (within lat. 4N to lat. 5.45N). It is observed that there is a shift downwards with the below average conditions of EOF1 in EOF2, and the northern part of the country recording the highest precipitation intensity in the spatial field maps. The temporal variability of EOF2 in Figure 2 showed a decadal-like variability. From the years 1981-1990, the spatial loadings accounted for 39.2% and 14.1% in Figure 4 in EOF1 and EOF2, respectively. The Coast and Forest zone showed below normal conditions and above normal conditions in the upper part of the Savannah zone. But above normal conditions were mostly observed in the Coastal of the EOF2. There was not much difference of variance in the years 1981-1990. Which shows that the decadal variability was small. EOF1 as compared to EOF2 in Figure 2 showed a downward shift of the Savannah zone and upward toward the transition zone from the coastal zone. This shows a decadal variability in the spatial pattern of precipitation over the years. The temporal patterns showed an inter-annual variability. The spatial pattern in Figure 8 also accounted for a variance of 40.9% and 18.8% in EOF1 and EOF2, respectively, with EOF1 showing above normal conditions over Ghana with the southern part of the country recording the highest. This means precipitation is higher in that area. EOF2 showed below normal conditions over the Coastal and Forest zones and above normal conditions over the Savannah and Transition belts. But EOF2 also recorded above normal conditions in the northern part of the country as compared with EOF2 in Figure 4 where the dominant mode was recorded in the coastal zone and south-west of Ghana. The temporal pattern shows an inter-annual variability with the years 2004, 2005, and 2007 recording the lowest.

6. Conclusion

The research has confirmed that, there is decadal spatial variability of precipitation (latitudinal shift) of the boundaries of the agro-ecological zones over the country but with mostly the northern part of the nation recording the highest mean precipitation values. This is due to the shifts of upward or downward movements of the northern and southern zone. The time series for the PCs showed an inter-annual variability.

References

[1] D. Eslinger, James J. O'Brien and R. Iverson, Empirical orthogonal function analysis of cloud-containing coastal zone color scanner images of northeastern North American coastal waters Journal of Geophysical Research 94(10) (1989), 884-910.

[2] W. Freiberger and U. Grenander, On the formulation of statistical meteorology, Revue de I'Institute International de Statistique (1965), 59-86.

[3] T. C. Gallaudet and J. J. Simpson, An empirical orthogonal function analysis of remotely sensed sea surface temperature variability and its relation to interior oceanic processes off Baja California Remote Sensing of Environment 47(3) (1994), 375-389.

[4] J. Hansen, R. Ruedy, M. Sato and K. Lo, Global surface temperature change, Reviews of Geophysics 48(4) (2010).

[5] G. S. Lagerloef and R. L. Bernstein, Empirical orthogonal function analysis of advanced very high resolution radiometer surface temperature patterns in Santa Barbara Channel, Journal of Geophysical Research: Oceans 93(C6) (1988), 6863-6873.

[6] M. Loeve, Probability theory, Vol. ii, Graduate Texts in Mathematics 46 (1978), 0-387.

[7] E. N. Lorenz, Empirical orthogonal functions and statistical weather prediction, (1956).

[8] W. H. Nam, E. M. Hong and G. A. Baigorria, How climate change has affected the spatio-temporal patterns of precipitation and temperature at various time scales in North Korea, International Journal of Climatology 36(2) (2016), 722-734.

[9] G. R. North, T. L. Bell, R. F. Cahalan and F. J. Moeng, Sampling errors in the estimation of empirical orthogonal functions, Monthly Weather Review 110(7) (1982), 699-706.

[10] K. Owusu and P. R. Waylen, The changing rainy season climatology of mid-Ghana, Theoretical and Applied Climatology 112 (2013), 419-430.

[11] K. Owusu, P. Waylen and Y. Qiu, Changing rainfall inputs in the Volta basin: Implications for water sharing in Ghana, Geo Journal 71 (2008), 201-210.

■