MAUSAM, **66**, 3 (July 2015), 569-578

551.509.5 : 551.577 (630)

Empirical statistical modeling of March-May rainfall prediction over southern nations, nationalities and people's region of Ethiopia

WONDIMU TADIWOS HAILESILASSIE and GIZAW MENGISTU TSIDU*

National Meteorological Agency, Addis Ababa, Ethiopia

**Dept. of Earth and Environmental Sciences, College of Science, Botswana International University of Technology and Science (BIUST), Priv. Bag 16, Palapye, Botswana*

e mail : wonde721@gmail.com

सार – इस शोध पत्र में दक्षिणी राष्ट्रों, इथोपिया के नैशेनिलिटीस और पीपल्स रीजन (SNNPR) में मार्च - मई (MAM) वर्षा से संबद्ध भूमंडलीय वर्षा के पूर्वसूचकों के बारे में पता लगाने के लिए सांख्यिकीय प्रागुक्त मॉडल विकसित किए गए। इस अध्ययन में प्रयोग में लाए गए आँकड़ों में स्टेशन के वर्षा के आँकड़े, समुद्री और वायुमंडलीय सूचनाएं शामिल है। अन्तः वार्षिक परिवर्तिताओं और वर्षा के वार्षिक चक्र में स्थानिक भिन्नताओं के कारण वर्षा संबंधी सूचनाओं को प्राप्त करने के लिए अध्ययन वाले क्षेत्र में 20 स्टेशनों के संजाल को तीन एक समान वर्षा वाले क्षेत्रों में निरूपित करने के लिए संचयी पदानुक्रमी समूह के विश्लेषण प्रयोग में लाए गए। निरूपित किए गए क्षेत्रों से उत्पन्न समय श्रृंखलाओं का उपयोग बाद में वर्षा/दूर सम्पर्क संबंधी सूचनाओं के विश्लेषणों के लिए किया गया। इसमें सहसंबंध विश्लेषण और बहुल रैखिक समाश्रयण वाली पद्धतियां लगाई गई। समाश्रयण की विधि 1987 – 2007 की प्रशिक्षण अवधि पर आधरित थी और 2008–2012 की मुक्त सत्यापित अवधि के लिए प्रेक्षणों हेतु इन मॉडलों को वैध माना गया। विश्लेषण से प्राप्त किए गए परिणामों से पता चला है कि समुद्र सतह तापमान (SST) की भिन्नताओं से मौसमी वर्षा में भिन्नताएं उत्पन्न हुई। तथापि मौसमी वर्षा में भिन्नता की बहुलता के लिए SST के विवरण से प्रागुक्ति मॉडलों में वायुमंडलीय सूचनाओं के समावेशन से वर्षा की प्रागुक्ति में सामान्य रूप से सुधार हुआ। इन तकनीकों से स्पष्ट रूप से यह पता चला है कि मॉडल पुनः तैयार किए गए थे और इनमें हितों वाले स्थानों के लिए वर्षा के पैटर्न बताए गए। प्रचालनात्मक स्तर पर पूर्वानुमान को उपयोगी बनाने के लिए कौशल में सुधार लाने और पूर्वानुमान के त्रुटिपूर्ण सीमाओं को निर्धारित करने के लिए मॉडल को और अधिक विकसित करने की आवश्यकता है।

ABSTRACT. Statistical predictive models were developed to investigate how global rainfall predictors relate to the March-May (MAM) rainfall over Southern Nations, Nationalities and People's Region (SNNPR) of Ethiopia. Data utilized in this study include station rainfall data, oceanic and atmospheric indices. Because of the spatial variations in the interannual variability and the annual cycle of rainfall, an agglomerative hierarchical cluster analyses were used to delineate a network of 20 stations over study area into three homogeneous rainfall regions in order to derive rainfall indices. Time series generated from the delineated regions were later used in the rainfall/teleconnection indices analyses. The methods employed were correlation analysis and multiple linear regressions. The regression modes were based on the training period from 1987-2007 and the models were validated against observation for the independent verification period of 2008-2012. Results obtained from the analysis revealed that sea surface temperature (SST) variations were the main drivers of seasonal rainfall variability. Although SSTs account for the majority of variance in seasonal rainfall, a moderate improvement of rainfall prediction was achieved with the inclusion of atmospheric indices in prediction models. The techniques clearly indicate that the models were reproducing and describing the pattern of the rainfall for the sites of interest. For the forecast to become useful at an operational level, further development of the model will be necessary to improve skill and to determine the error bounds of the forecast.

Key words – Seasonal rainfall, Statistical, Hierarchical clustering, Teleconnection, Linear regression.

1. Introduction

Rainfall is the climatic factor of maximum significance for the East African countries, with extreme occurrences resulting in droughts and floods, which are often associated with food, energy and water shortages, loss of life and property, and many other socio-economic disruptions. The economies of East African countries largely depend on agriculture, which is highly vulnerable to the amounts and distribution of rainfall (Indeje *et al*., 2000). Successful development of predictive capability for the seasonal rainfall anomalies is beneficial to planners and policy makers in making more constructive socioeconomic decisions by taking into account the detailed spatial and temporal climate variations (Mamo, 2005). The increasing concern about the socio-economic

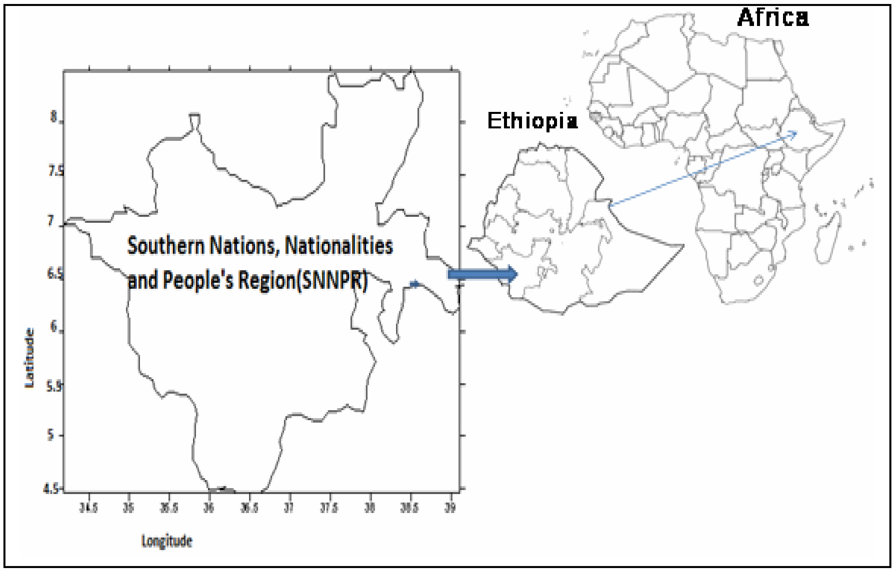


Fig. 1. Location of southern nations, nationalities and people's region of Ethiopia

impacts of climate risks, both due to natural and anthropogenic reasons has led to a rapid model development (Kassahun, 1990**).** The March-May (MAM) rains over East Africa are more influenced by the north/south movement of the zonal arm of the Intertropical Convergence Zone (ITCZ) in the region and other weather systems such as the meso-scale systems, interactions between extra-tropical and tropical weather systems among others (Ogallo, 1989).

Sea surface temperature (SST) is a reliable variable to be used as a forecast tool. The high inertia of the sea makes the analysis of SST anomalies useful in monthly and seasonal rainfall prediction in some areas of the globe (Barnston, 1994). In the last two decades, a large number of studies showed that the significant part of interannual rainfall variance over several African regions is induced by, or at least related to SST variability, especially in tropical areas (Camberlin *et al*., 2001). A number of studies have documented increased precipitation in East Africa at times when the eastern Indian Ocean is cold and the western Indian Ocean is anomalously warm. This basin-wide east west SST-gradient, often referred to as the positive Indian Ocean dipole (Saji *et al*., 1999).

Schott *et al*. (2007) indicated that Indian Ocean Dipole (IOD) affects rainfall in remote regions, such as the east Africa short rains, and there is evidence that a co-occurring IOD event may affect both the growth and termination of El Nino Southern Oscillation (ENSO). Emily (2005) concluded that the observed teleconnection between East African rainfall and ENSO is a manifestation of a link between ENSO and IOD. Tadewos (2007) concluded that not only during the El Niño condition does the southern Ethiopia experience large amount of rainfall but also during the positive IOD. Yeshanew and Jury (2006) suggested that using ENSO information and SST over the Atlantic and Indian Ocean, climate prediction over Africa has been achieved at one season lead time. Ogallo (1988) indicated that ENSO explains about 50 per cent of the East African rainfall variance with other factors explaining the remaining variance.

There have been some studies over Ethiopia (Gissila *et al*., 2004; Segele and Lamb, 2005; Diro *et al*., 2011) examining connections between observed rainfall and a number of large-scale climate signals. The aims of the previous studies have generally been to develop, where possible, methods for long-range forecasting at national scale based on oceanic indices as predictors, however, Gissila *et al*. (2004) indicated that the possibility of other climate parameters (other than SST) would improve the skill of the forecast model. In addition to most of the past studies focused on summer (JJAS) rainfall. In JJAS, convective activity typically develops over Ethiopia highlands, while southern and southern eastern Ethiopia receives little rain (Korecha and Barnston, 2007). This study focuses on the predictability of MAM rainfalls which have significant economic importance for the country because the 'long cycle (March-October) high yield' crops such as sorghum and maize are planted in March over the southern region of Ethiopia.

The main objective of this work is to investigate the potential for statistical precipitation forecasts based on

TABLE 1

Predictors and their locations

S. No.	Predictor	Code	Location	Index
1.	South Atlantic SST	SATL	0-20° South, 30° West-10° East	Oceanic
2.	Global Tropics SST	GTROP	10° South - 10° North, 0-360	Oceanic
3.	Nino 1+2 SST	Nino1+2	0 - 10° South, 90° West - 80° West	Oceanic
4.	Sothern Oscillation Index	SOI	Air pressure difference at Darwin and Tahiti	Atmospheric

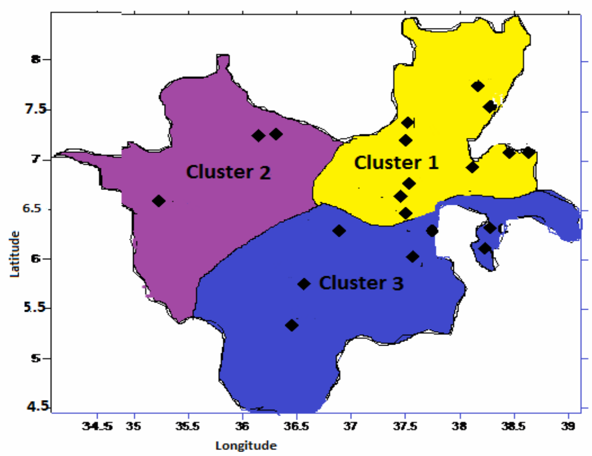


Fig. 2. Location of the rain gauge stations used in this study and the homogeneous rainfall zones. The methodology for deriving the clusters is described in Section 3.2

large-scale climate signals; and thereby develop predicting capability of its MAM rainfall at one or more monthly lead times with good prediction skill. Additionally, forecast regression models are proposed. The paper is organized as follows. Details regarding the study area and general precipitation features are provided in Section 2. Data and methodology used are described in Section 3. Results are presented and discussed in Section 4 and finally, conclusions are drawn in Section 5.

2. Study area

2.1. *Geographical location*

The Southern Nations, Nationalities and People's Region (SNNPR) as shown in Fig. 1, is located in the southern and south-western region of Ethiopia. It lies between 04°43' - 08°58' North latitude and 34°08' - 39°14' East longitude.

2.2. *General precipitation features*

The annual cycle in rainfall is heavily influenced by the Intertropical Convergence Zone (ITCZ). In Africa, the ITCZ oscillates annually between an extreme northward location of 15° N in July and an extreme southward location of 15° S in January (Diro *et al.*, 2008 and referenced therein). The amount, duration and intensity of rainfall in the region vary considerably. It generally decreases from west and northwest to south-eastwards. The mean annual rainfall of the region ranges from 400 mm to 2,200 mm.

As indicated in Figs. 3(a-c), the passage of the ITCZ gives rise to the bimodal rainfall pattern in southern portion of the study area (cluster 3) with the rainfall peaks observed in April/May and November and the monomodal pattern (cluster 1) in the northern portion of the area with one rainfall maximum during July or August. However, Figs. 3(a-c) show that western portion of the study area (cluster 2) gets rain for a long period at a stretch usually from eight to nine months. In general, the inter-annual oscillation of the surface position of the Inter-Tropical Convergence Zone (ITCZ) causes a variation in the wind flow patterns over Ethiopia. This periodical anomaly of winds causes rainfall to be variable and seasonal in Ethiopia.

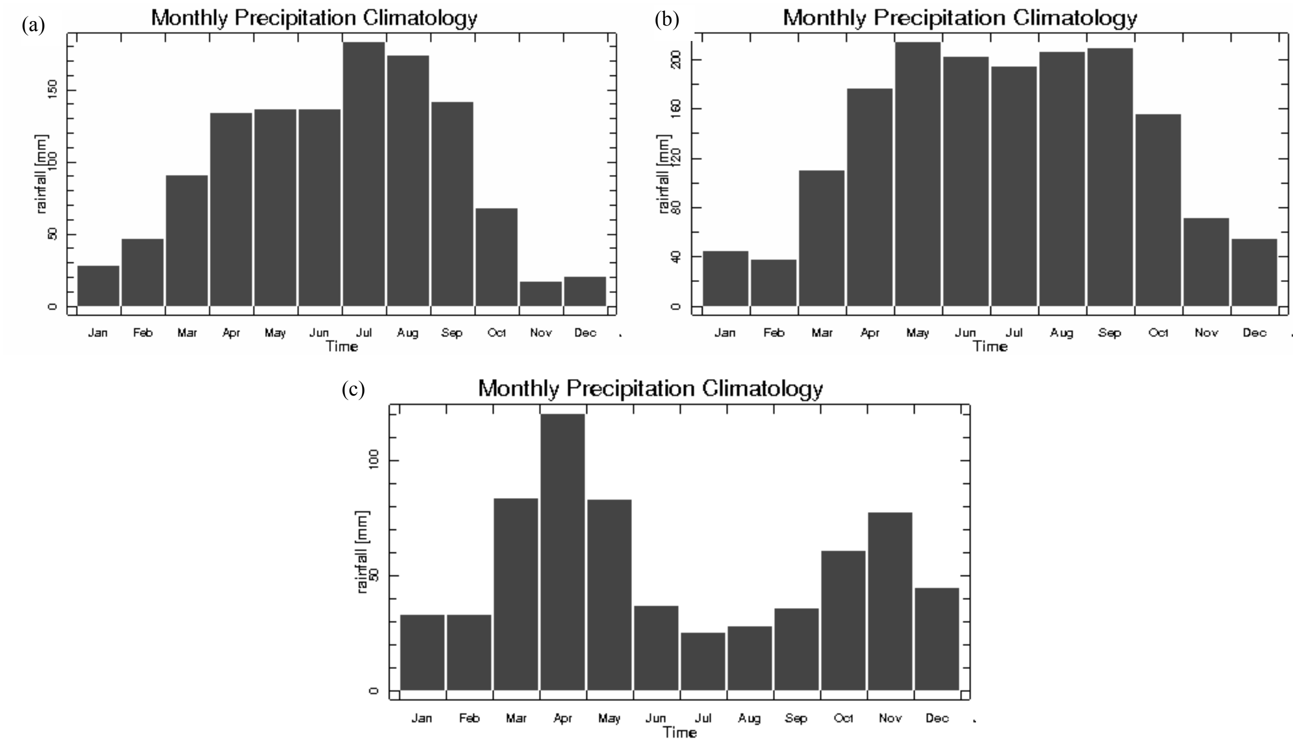
Rainfall information is important for food production plan, water resource management and all activity plans in the nature. The occurrence of prolonged dry period or heavy rain at the critical stages of the crop growth and development may lead to significant reduce crop yield (Win and Thinn, 2008). Thus rainfall prediction becomes a significant factor in agricultural countries like Ethiopia.

3. Data and methodology

3.1. *Data*

3.1.1. *Rainfall (Predictand)*

Due to the problem of significant missing data, monthly total rainfalls of 20 stations were used. These were obtained from NMA (National Meteorological Agency) of Ethiopia for the period of 1987-2012. Fig. 2 indicates the distribution of selected rain gauge stations. The predict and in this work was the MAM rainfall.



Figs. 3(a-c). Monthly rainfall variation in annual cycle for the homogenous rainfall zones in: (a) Cluster 1, (b) Cluster 2 and (c) Cluster 3; this rainfall time series reconstructed from station observation and remote sensing and other proxies for the time period of 1983-2010 (www.ethiometmaprooms.gov.et)

3.1.2. *Predictors*

Monthly datasets of Global Tropics Ocean SST, South Atlantic Ocean SST, NINO1 + 2 regions (one of the standard ENSO indices), and Southern Oscillation Index (SOI) were used. These data were taken from the websites: www.esrl.noaa.gov/psd/data/climateindices/list and www.cpc.ncep.noaa.gov/data/indices/. The IOD index was obtained from www.jamstec.go.jp/frcgc/research/d1/iod/box_sst.dat. These variables were considered for lag correlations analysis from 1 to 8 months. Predictors selected and listed in Table 1 have been identified by many authors to be related to large-scale eastern African climate.

3.2. *Methodology*

3.2.1. *Homogeneous rainfall zones*

Ethiopian rainfall is highly variable, both temporally and spatially, and so must be analyzed in clusters rather than as a whole (Gissila *et al*., 2004). Cluster analysis (CA) is one of the most useful tasks in the data mining process for discovering groups and identifying interesting patterns in the underlying data (Muñoz-Díaz and Rodrigo, 2004). Regionalization and averaging of rainfall over large but homogeneous regions have the advantages of reducing meteorological noise in the data as well as minimizing the number of variables, which describe the regional climate variability. Gillian *et al*. (2000) concluded that Euclidean distance-based cluster analysis is frequently used to define climate regions. Most commonly implemented CA procedures are hierarchical (Wilks, 1995). Ward (1963) proposed a very general hierarchical cluster method known as “Ward’s method” or the “minimum variance method.” Hierarchical clustering is a way to investigate grouping the data by creating a cluster tree. The tree is a multi-level hierarchy, where clusters at one level are joined as clusters at the next higher level. In data mining hierarchical clustering is a method of cluster analysis which seeks to build a hierarchy of clusters. Strategies for hierarchical clustering generally fall into two types. Agglomerative clustering is a “bottom up” approach: each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy. Divisive clustering is a “top down” approach: all observations start in one cluster, and splits are performed recursively as one moves down the hierarchy (Muñoz-Díaz and Rodrigo, 2004; Fionn and Pedro, 2011).

TABLE 2

Mean cross-correlations between stations within the clusters and with stations in the other clusters

MAM	Cluster 1	Cluster 2	Cluster 3
Cluster 1	**0.50**	0.39	0.44
Cluster 2	0.39	**0.43**	0.40
Cluster 3	0.44	0.40	**0.53**

In this paper, an agglomerative hierarchical technique of clustering was used to divide meteorological variable of MAM rainfall for 20 stations. Variables were usually standardized before applying CA, to eliminate possible scale effects. This type of clustering used a data array, or matrix, to classify (or cluster) the cases of data into groups according to values of their attributes. In this case, MAM rainfalls of 21 years (1987-2007) sample were arranged in columns and station names were in rows. In addition to this, the interannual correlations (correlation between all stations) of monthly rainfalls of 21 years sample were used to confirm and further subdivide the clusters. Table 2 displays mean cross-correlations (Intra-zonal correlation matrix) between stations within the clusters and with stations in the other clusters. Thus, for example, the cross-correlation between cluster 1 and cluster 2 would be the mean of the cross-correlations between every station in cluster 1 and every station in cluster 2; and the cross-correlation for cluster 1 would be the mean of the cross-correlations between all stations in cluster 1.

There can be evidence from Table 2 (shown in bold), for all of the clusters the interannual variability of monthly rainfall within a given cluster is reasonably coherent (the stations whose monthly rainfall pattern is well correlated). Overall these interstation correlations suggest that the interannual rainfall pattern is reasonably homogeneous. As described previously (Section 2.2) monthly rainfall variation in annual cycle was also used for further confirmation of the cluster analysis. Based on the above criteria, the study area was delineated into three homogeneous rainfall zones (Fig. 2) in order to derive rainfall indices.

According to South African Development Community (SADC, 2000) the time series for individual stations contain a lot of variability that is either local resulting from local storm that was not attributed to large scale climate process or attributed to errors of instrumentation. One way to get a salient picture of rainfall variability is to calculate area-averaged rainfall indices, and it is important to average the indices and not the actual values. This index accounts for local climatological intensity and variability and is a relative value rather than an absolute value. The rainfall index used in this study was calculated using the CRC (Centre de Recherche de Climatologie) monthly station-rainfall database (Philippon *et al*., 2002).

The areal average rain fall index (RFI) calculated as:

$$\text{RFI} = \frac{1}{N}\sum_{i=1}^{n} \frac{r_i - R_i}{\sigma_i} \tag{1}$$

where, N is the number of stations, r_i, R_i and σ_i are cumulative rainfall, the mean cumulative rainfall and the standard deviation of the cumulative rainfall at station i, respectively. The mean and the standard deviation were based on 21-year sample.

3.2.2. *Formulation of statistical models*

Regression analyses are a set of statistical techniques which allow one to assess the relationship between one dependent variable (DV) and several independent variables (IVs). The multivariate regression equation is of the form:

$$\begin{aligned} Y &= A + B_1 X_1 + B_2 X_2 + \ldots + B_k X_k + E \\ &= A + \sum_{i=1}^{k} B_k X_k + E \end{aligned} \tag{2}$$

where, y is the predicted value on the DV, A is the Y intercept, X is the various IVs, B are the various coefficients assigned to the IVs during the regression, and E is an error term. The goal of the regression is then to derive the B values, the regression coefficients, or beta coefficients. To select the most useful predictors for MLR (Multiple Linear Regression) model to preserve it against collinearity, we applied a stepwise regression procedure with a significance level for predictor to enter or leave the model fixed at 90% and 95%. In a stepwise regression, the models were assessed by the residual sum of squares (Wilks, 1995). Tolerances and variance inflation factors are indicators of collinearity (multi-collinearity) for multiple X variables. Tolerance is calculated as (1 - R^2) where R^2 is the correlation of determination (variance explained) for X_i against all of the other X variables.

The variance inflation factor (VIF) is calculated as:

$$VIF_j = \frac{1}{1 - R_j^2} \tag{3}$$

where, R_j^2 is correlation of determination (variance explained) from the regression of the j^{th} predictor on all

TABLE 3

The correlation coefficient between each cluster and potential predictors

Cluster 1	Cluster 2	Cluster 3
NOV (SOI) = -0.356	DEC (SOI) = 0.316	AUG (NINO1+2) = -0.339
OCT (SATL) = -0.377	SEP (IOD) = -0.390	FEB (SATL) = 0.336
JUL (GTROP) = -0.337	JAN (SATL) = -0.448	AUG (GTROP) = -0.324
	SEP (SATL) = 0.412	

TABLE 4

Regression forecast models and their goodness of fit measurement for the training period (1987-2007)

Cluster	Multiple linear regression equation for RFI (Rainfall Index)	R	MAE	RMSE
One	RFI = 0.119-0.220NOV(SOI)-0.509OCT(SATL)-1.474JUL(SATL)	0.65	0.37	0.47
Two	RFI = 0.018-0.014DEC(SOI)-0.126SEP(IOD)-1.192JAN(SATL)+1.820SEP(SATL)	0.70	0.44	0.56
Three	RFI = 3.678-0.170AUG(NINO1+2)+1.013FEB(SATL)-0.911AUG(GTROP)	0.60	0.46	0.53

other predictors in the model. A high value of VIF, typically greater than 10, indicates that the predictor is collinear with the others (Diro *et al*., 2008 and referenced therein).

3.2.3. *Model validation or verification*

Model validation is usually an important component of model construction. Wilks (1995) suggested a method for model validation that has been adopted in many studies. The method involves constructing the model using a subset of data available and then testing the constructed model over the remaining periods. Unfortunately, too little attention is paid to the assessment of error. The amount of error can be a useful criterion to identify the 'best' model. The mechanisms that can be used for model validation and testing depend on how the model was developed but most involve applying the model to independent data, *i.e.*, data that were not used to produce the model. In this study, the performances of the models were assessed for the independent period from 2008-2012 through correlation coefficients (R) between estimated and observed values. The models assessments were also made by calculating the mean absolute error (MAE) and root mean square error (RMSE). R and RMSE are computed as follows:

$$R = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}} \tag{4}$$

where, x_i is the i^{th} value of observed rainfall x, y_i is the i^{th} value of predicted rainfall Y, $\bar{x}$ is the mean of variable x, $\bar{y}$ is the mean of variable y, n is the number of observations.

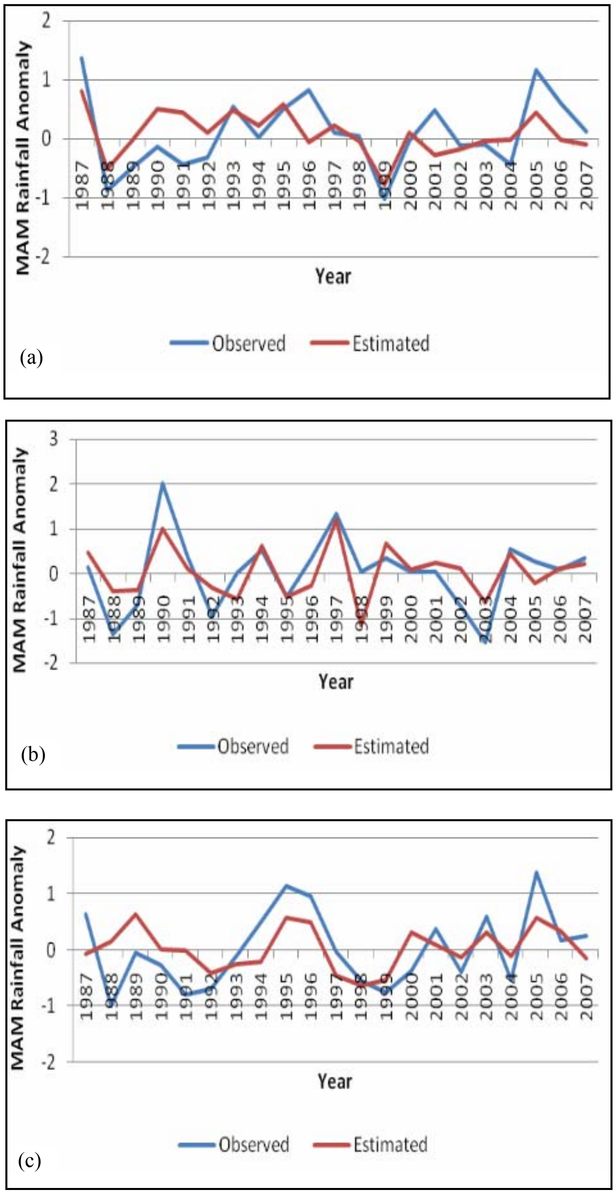
$$RMSE = \sqrt{\sum_{i=1}^{n}\left(x_i - y_i\right)^2 / n} \tag{5}$$

4. Results and discussion

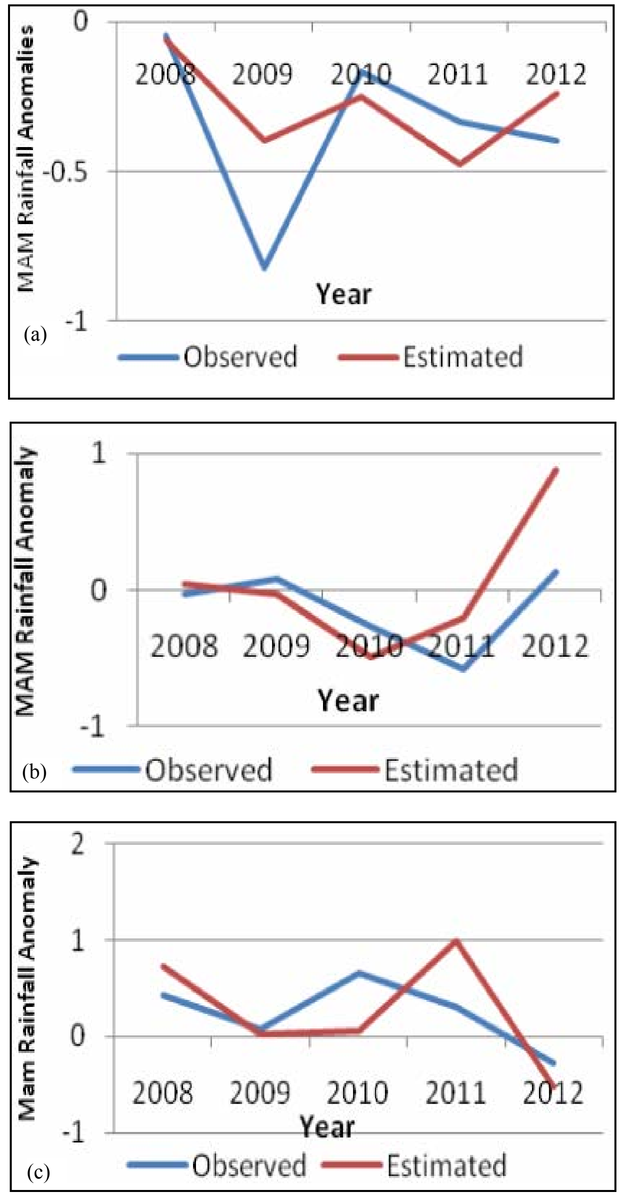
4.1. *Identification of predictors and development of the models*

To select the best predictors to be used in the multiple linear regression methodology, the correlations between the indices described in Table 1 and MAM rainfall were calculated considering 1-8 months of lags. However, as the predictors must be independent among each other, some of them were disregarded due to the criteria of VIF "Equation (3)", if the computed value of VIF>10, then the predictors were excluded to remove the multicollinearity. In addition to this, in Table 3 the input variables were also selected considering the statistically significant correlations correspond to the 90% and 95% significance levels (thresholds) by the method of a stepwise regression.

In order to validate our proposed models, we calculated the correlation coefficients, the root mean square error (RMSE) and mean absolute error (MAE)



Figs. 4(a-c). Observed and forecast rainfall for (a) cluster one, (b) cluster two and (c) cluster three, using the regression method



Figs. 5(a-c). Observed and forecast rainfall with the test data of 2008-2012 for (a) cluster one, (b) cluster two and (c) cluster three

presented in Table 4. These statistics for the model fitting based on the training data of 1987-2007. In general, as indicated in Table 4 the correlation coefficients (R values) between observed and forecast series of each cluster in descending order as follows: R = 0.70 for cluster 2, R = 0.65 for cluster 1 and R = 0.60 for cluster 3. The RMSE for the regression model in Table 4 shows that cluster one has the smallest while cluster two has the largest. In general, a single response variable (rainfall) could be predicted from a sum of the weighted predictors.

In Table 4 we can see that the models can show the importance of the SATL, IOD, GTROP, NINO 1+2 regions and SOI as good indicators of the MAM rainfall indices in defined homogenous areas (clusters). We also observed that the most important source of predictability comes from the interannual variability of SST in the SATL. This study is consistent with the works of, for instance, Indeje (2000) indicated that monthly time lag using SST anomalies were used to forecast the MAM rainfall season over East Africa; Camberlin and Philippon

TABLE 5

Evaluation of models performance for the verification periods (2008-2012)

Cluster	R	MAE	RMSE
One	0.64	0.16	0.22
Two	0.65	0.30	0.39
Three	0.57	0.37	0.44

(2002) used Nino 1+2 in predictability of MAM rainfall over East Africa in the development of MLR equation; Diro *et al*. (2008) also showed that for the southern Ethiopia, tropical Atlantic regions have robust correlations; Seleshi (1996) noticed that the Indian Ocean is another moisture source especially for southern Ethiopia in March and April; Gissila *et al*. (2004) also described links between Indian Ocean SST variability and rainfall in Ethiopia; and the Drought Monitoring Centre for Eastern Africa, in Nairobi, Kenya, uses the Southern Oscillation index(SOI), the Niño-1-4 SST indices, and SST over the Indian Ocean, among others, to predict the March-May seasonal rains (Camberlin and Philippon, 2002 and referenced therein).

4.2. *Observed and estimated rainfall for the training period (1987-2007)*

Data were separated into training period and test period. Essentially, 21 years of data (1987-2007) were used to develop forecast models. Figs. 4 (a-c) show the performance of the model between the original data and the regression estimates. From these, we can see that there are reasonable agreements or similar trends observed for all clusters.

4.3. *Model performance evaluation or verification*

The period from 2008-2012 was used for independent verification of the regression equation. The validation of the model consists of comparing the forecasts with observations. Figs. 5 (a-c) illustrate the observed and estimated rainfall by applying the model to independent data, *i.e*., data that were not used to produce the model. The results indicated that the predicted rainfall anomalies are close to the actual values.

Table 5 depicts the model performance evaluation based on the calculated values of correlations (R) between estimated and observed rainfall. It also describes the MAE and RMSE. In general, from this table, the correlation coefficients (R values) between observed and forecast series of each cluster in descending order as follows: 0.65, 0.64 and 0.57 for cluster two, one and three respectively. The RMSE for the regression model in this table shows that cluster one has the smallest while cluster three has the largest. Validation results indicated that the proposed methodology is a potential tool that may be used to predict the rainfall level in a given homogenous areas (clusters). However, in both the training set and independent verification cluster 3 seems the most difficult to predict accurately. This is due to the MAM rainfall of cluster 3 shows weak correlation coefficients in both training (Table 4) and verification period (Table 5). Besides, cluster 3 has the bimodal rainfall pattern [Figs. 3(a-c)] and MAM rain falls as the 'long rains' with similar features to the East Africa. This issue has been analyzed previously (Ogallo, 1988; Camberlin, 2001; Camberlin and Philippon, 2002). They found that the East Africa 'long rains' (MAM rainfall) show weak responses (low correlations) to global SST.

5. Conclusions

Knowledge of the processes controlling rainfall is essential to the development seasonal forecasting systems, which could be used to mitigate the effects of flood and drought. Rainfall over study area has considerable spatial variability on both interannual and seasonal time scales. For this reason, forecasts have developed and evaluated for three homogenous rainfall zones which have been identified by using an agglomerative hierarchical clustering approach.

This work has investigated links between large scale climate signals and MAM rainfall index. We have taken into account monthly correlations between large scale variables and MAM rainfall anomalies of different homogenous areas from one to eight months of lag. A stepwise regression methodology was used to develop a prediction model and independent data, *i.e*., data that were not used to produce the model was applied for verification. The analysis for the regression models (Table 4) show that important sources of predictability comes from the interannual variability of SST in SATL, IOD, Nino 1+2 regions and GTROP and the interannual pressure variation of SOI. The correlation coefficients between observed and estimated rainfall for clusters one, two and three were 0.65, 0.70 and 0.60 for the training and 0.64, 0.65 and 0.57 for the verification periods respectively. However, the correlation was weak for cluster 3 in both the training set and independent verification and so it looked difficult to forecast. It is possible that the addition of other climate parameters would improve the skill of the forecasting model.

In general from this study we concluded that the linkages that were observed from the statistical analyses had some physical reality with selected predictors. Such information will help to improve monitoring, prediction and early warning of extreme rainfall events, reduce the vulnerability and improve the resilience of the society of the region to negative impacts of extreme rainfall events that are common in the region. Seasonal climate forecasts have been issued and utilized in Ethiopia for national-scale famine and disaster planning and relief efforts. However, in order to have a significant impact on food production at the farm scale, and thus directly address the food security problem, seasonal forecasts need to be "tailored" and provided to the needs of small-scale producers.

Acknowledgements

We sincerely acknowledge the National Meteorological Agency (NMA) of Ethiopia for providing monthly rainfall data for this study. The authors are very much thankful to the anonymous reviewers for providing very useful suggestions, which helped a lot in improving the quality of the paper.

References

Barnston, A. G., 1994, "Linear statistical short-term climate predictive skill in the Northern Hemisphere", *Journal of Climate*, **7**, 1513-1564.

Camberlin, P. and Philippon, N., 2002, "The East African March-May Rainy Season Associated Atmospheric Dynamics and Predictability over the 1968-97 Periods", *Journal of Climate*, **15**, 1002-1019.

Camberlin, P., Janicot, S. and Poccord, I., 2001, "Seasonality and atmospheric dynamics of the teleconnection between African rainfall and tropical sea-surface temperature: Atlantic *vs*. ENSO", *International Journal of Climatology*, **21**, 973-1005.

Diro, G. T. Black, E. and Grimes, D. I. F., 2008, "Seasonal forecasting of Ethiopian Spring rains", *Meteorol. Appl.*, **15**, 73-83.

Diro, G. T., Grimes, D. I. F. and Black, E., 2011, "Teleconnection between Ethiopian Summer rainfall and sea surface temperature: Part I - observation and modeling", *Clim. Dyn.*, **37**, 103-119.

Emily, B., 2005, "The relationship between Indian Ocean Sea-surface temperature and East African rainfall", *Phil. Trans. R. Soc. A.*, **363**, 43-47.

Fionn, M. and Pedro, C., 2011, "Methods of Hierarchical Clustering", *Egham TW20 0EX, England*, 1-21.

Gillian, M. M., Simon, J. M. and Jacqueline, S. G., 2000, "Notes and Correspondence Choice of Distance Matrices in Cluster Analysis : Defining Regions", **14**, 2790-2790.

Gissila, T., Black, E., Grimes, D. I. F. and Slingo, J. M., 2004, "Seasonal forecasting of the Ethiopia Summer Rains", *Int. J. Climatol*., **24**, 1345-1358.

Indeje, M., 2000, "Prediction and numerical simulation of the regional climate of equatorial eastern Africa", Ph.D thesis, Marine, Earth and Atmospheric Science, North Carolina State University. Raleigh North Carolina, USA. p352.

Indeje, M., Semazzi, F. and Ogallo, L., 2000, "ENSO signals in East Africanrainfall and their prediction potential", *Int. J. Climatol*., **20**, 19-46.

Kassahun, B., 1990, "Prediction of Seasonal Rainfall in the Sahel Region. MSc thesis (Meteorology)", University of Reading, U. K., p148.

Korecha and Barnston, A. G., 2007, "Predictability of June-September Rainfall in Ethiopia", *Mon. Wea. Rev.*, **135**, 628-650.

Mamo, G., 2005, "Using seasonal climate outlook to advice on sorghum production in the centeral rift valley of Ethiopia. Ph.D Thesis (Agrometeorology)", University of the Free State, Bloemfontein, South Africa, p171.

Mũnoz-D´ıaz, D. and Rodrigo, F. S., 2004, "Spatio-temporal patterns of seasonal rainfall in Spain (1912-2000) using cluster and principal component analysis : comparison", *Annales Geophysicae*, **22**, 1435-1448.

Ogallo, L. J., 1988, "Relationships between seasonal rainfall in East Africa and the Southern Oscillation", *Int. J. Climatol*., **8**, 31-43.

Ogallo, L. J., 1989, "The spatial and temporal patterns of the East African seasonal rainfall derived from principal component analysis", *Int. J. Climatol*., **9**, 145-167.

Philippon, N., Camberlin, P. and Fauchereau, N., 2002, "Empirical predictability Study of October-December East Africa rainfall", *Quarterly J. of the Royal Met. Society*, **128**, 2239-2256.

SADC Drought Monitoring Centre., 2000, "Proceedings of the second capacity Buildin Workshop on Seasonal Forecasting for the Workshop for SADC Region", Harare, Zimbabwe.

Saji, N. H., Goswami, B. N., Vinayachandran, P. N. and Yamagata, T., 1999, "A Dipole Mode in the Tropical Indian Ocean", *Nature*, **401**, 360-363.

Schott, F. A., Xie, S. and McCreary, J. P., 2007, "Indian Ocean circulation and Climate variability", Leibniz Institute Universität Kiel, Kiel, Germany and International Pacific Research Center, University of Hawaii, Honolulu, Hawaii.

Segele, Z. T. and Lamb, P. J., 2005, "Charcterirization and variability of Kiremt rainy season over Ethiopia", *Meteorol. Atmos. Phys*., **89**,153-180.

Seleshi, Y., 1996, "Stochastic Predictions of Summer Rainfall Amounts over the North African Highlands and over India. PhD thesis (Water Resources Engineering)", Virgie University of Brussels, Belgium, p352.

Tadewos, M., 2007, "Tropical oceans thermodynamic and atmospheric modulations over southern Ethiopia climate, water resource and vegetation variability. MSc thesis (Physics)", Addis Abeba University, Addis Abeba, Ethiopia, p101.

Ward, J. H., 1963, "Hierarchical grouping to optimize an objective function", *Journal of the American Statistical Association, Alexandria*, **58**, 1, 236-244.

Wilks, D. S., 1995, "Statistical Methods in Atmospheric Sciences", Academic Press, San Diego, p467.

Wint, T. Z. and Thinn, T. N., 2008, "Empirical statistical modeling of rainfall prediction over Myanmar", *Proceedings of World Academy of Science: Engineering & Technology*, **48**, 569-572.

Yeshanew, A. and Jury, M. R., 2006, "North African climate Variability. Part3. Resource prediction", *Theor. Appl. Climatol*., **89**, 51-62.