Simulation of Decadal Precipitation over Nairobi in Kenya

Muthama Nzioka John*, Manene M.Moses** and Ndetei Cornelius Joseph*

*Department of Meteorology, University of Nairobi, Kenya; **School of Mathematics, University of Nairobi, Kenya, P.O.Box 30197-00100 Nairobi, Kenya, Email: jmuthama @uonbi.ac.ke.

(MAM)
. (OND)
OND MAM
(10)

ABSTRACT: In investigating Kenya rainfall variability and its relationship to other climatic elements it has become imperative to analyze the irregularly distributed rainfall events in time. To meet this requirement, this study used a stepwise regression technique. The study seeks to improve existing rainfall monitoring and prediction in Nairobi. Monthly rainfall data was fitted to several mathematical functions. The best mathematical model which best simulated the March-May (MAM) and October -December (OND) seasonal rainfall over the three stations of analysis was chosen using a stepwise regression technique. The value of R-squared for the best fit was computed to show the percentage of rainfall information that is explained by the variation in the independent (time) variable. From the results obtained, the stepwise regression technique selected the fourth degree polynomial as the best fit for analyzing the March-May (MAM) and October -December (OND) seasonal rainfall data set. Solar cycle period of ten (10) years was employed to get the fourth degree polynomial variables. Hence from the study, it can be deducted that the 4th degree polynomial function can be used to predict the peak and the general pattern of seasonal rainfall over Nairobi, with acceptable error values. This information

can be used in the planning and management of water resources over Nairobi. The same information can be extended to other areas.

KEYWORDS: Stepwise regression; Rainfall variability; Polynomial function; Solar cycle period; Nairobi.

1. Introduction

There is an ongoing effort to better understand the coherent multidecadal fluctuations in the global climate system that result from interactions between the various components of the climate system itself, such as the oceans, land, and atmosphere. It is generally assumed that the slowly varying sea surface temperatures (SSTs) are the dominant source of this variability (e.g., Kawamura *et al.* 1995a,b; Livezey and Smith 1999; Mo *et al.* 2001; Hoerling *et al.* 2001), which has prompted several extended climate simulations using general circulation models (GCMs) forced by the observed near-global SSTs (Graham 1994; Lau and Nath 1994). Chelliah and Bell, 2004 studied the coherent decadal fluctuations in tropical convection and surface temperatures over the central Pacific, the West African monsoon region, the Amazon basin, and the Indian Ocean, which are shown to constitute the dominant decadal-scale mode of variability in the global Tropics.

The East African region is characterized by a complex topography. This topographical inhomogeneity includes large lakes, Rift Valleys, flat plains and snow-capped mountains on the Equator. Almost all facets of societal and economic activities in the region are critically dependent on the variability of seasonal rainfall, which occurs during boreal Spring (Long Rains, March-May) and Autumn (Short Rains, October-December). However, the societies are often unprepared to adjust quickly to dramatic deviations from normal rainfall regimes. Extremes in weather and climate may result in loss of lives and damage to property including massive disruptions of existing infrastructures. These negative impacts have been detrimental to the economy.

The forecasting of severe weather and extreme climate events is one of the major challenges facing meteorological services worldwide and more so in the tropics. However, researchers have made enormous efforts in addressing the issue of accurate rainfall predictability. Amongst some of the techniques that have been employed to predict rainfall in East Africa include the use of numerical and statistical methods. Mathematical functions especially polynomial regressions have also been used to characterize weather elements including rainfall in different parts of the world. For instance, Christine *et al* (1998) used Polynomial regression to derive a simple model for each monthly climate variable to relate climate to position and elevation on Digital Elevation Model (DEM) in Ireland. Accuracy assessments with subsets of each climate data set showed that polynomial regression can predict average monthly climate in Ireland with mean absolute errors of 5 to 15 mm for monthly precipitation, 0.2 to 0.5°C for monthly averaged maximum and minimum temperature, and 6 to 15 min for monthly averaged sunshine hours.

Camberlin and Philippon (2001) presented a forecast of the March-May 2001 rain season in East Africa, using complementary statistical techniques: the Multiple Linear Regression (MLR) and the Linear Discriminant Analysis (LDA) as in Folland *et al.* (1991). Their prediction scheme involved the use of sea Surface Temperatures (SST) indexes and atmospheric predictors. The March to May period corresponds to the main rain season in Equatorial Eastern Africa (so-called "long rains"). Although this rain season exhibits lesser interannual variability than that of October-December ("short rains"), it is the main agricultural period in East Africa, and droughts severely affect the entire economy (e.g., in 2000 in Kenya). Yet, these interannual variations are much less known than those of the "short rains". In particular, the teleconnections with sea surface temperature anomalies (both regional and global) are much weaker (Ogallo *et al.*, 1988; Rowell *et al.*, 1994; Mutai and Ward, 2000), and this precludes the definition of acute seasonal rainfall prediction models based on that sole variable.

The rainfall over East Africa is of oscillatory character in time, and as shown by Ogallo (1977), the result of the spectral analysis indicated that there were some oscillations in the annual rainfall series. Potts (1971), and Rodhe and Virji (1996) showed that oscillatory peaks of 2-2.5, 3.5 and 5.6 years in addition to other peaks

associated with EL Nino (e.g., Ropelewski and Halpert 1987) exist in East Africa rainfall. However, it has never been clear as to when these oscillations occurred in the rainfall.

In investigating Kenya rainfall variability and its relationship to other climatic elements, it has become imperative to analyze the irregularly distributed rainfall events in time. To meet this requirement, this study used stepwise regression technique. The study seeks to improve existing rainfall monitoring and prediction in Nairobi.

2. Data and methods

Monthly totals rainfall data obtained from the Kenya Meteorological Department (KMD) was used in the study. The data was for three stations in Nairobi city, namely Dagoretti (1^o 18'S, 36^o 45'E), JKIA (1^o 19'S, 36^o 55'E) and Wilson (1^o 19'S, 36^o 49'E) which are on the same homogeneous climatic zone. The data covers more than 30 years which was considered as sufficient climatic period.

2.1 Data quality control and estimation of missing records

Data quality control is important in order to detect any discontinuities in the data that may have occurred from non-natural influences like changes in observational schedules and methods, instrumental and other human processes (WMO, 1966). Heterogeneity makes records not strictly comparable over long time periods and between different stations.

It is therefore important to ascertain the homogeneity of any meteorological data before using it in any study. The single mass curve was used to check on homogeneity of the rainfall data. The mass curve was obtained by plotting cumulative records of rainfall against time. Basically, from plot of single mass curve, a straight line indicates a homogenous record whereas heterogeneity can be indicated by significant deviations, of some of the plots from the straight line.

The missing rainfall records were less than 10% of the total data. The different methods of estimating missing data (WMO, 1962) include; the arithmetic mean method, the isopleths method, Thiessen polygon method, isohyetal method, finite differencing method, Correlation and regression method. In this study the correlation method was used to estimate the missing records.

The correlation method of estimating the missing data is computed as;

$$X_{\text{missing period}} = \frac{\overline{x}}{\overline{y}} \cdot y_{\text{period}}$$
 (1)

where y_{period} refers to the period of the station with complete data set, \bar{x} is the mean of the available data set for the station with missing records, \bar{y} is the long term mean of the station with complete data set and r_{xy} is the correlation coefficient, and is given by;

$$r_{xy} = \frac{\frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\left[\frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2 \cdot \frac{1}{n} \sum_{i=1}^{n} (y_i - \overline{y})^2\right]^{\frac{1}{2}}}$$
(2)

2.2 Choosing the Best Fit Model

Monthly rainfall data for the three stations (i.e Dagoretti, JKIA and Wilson) was fitted to several mathematical functions. These mathematical functions were grouped into five families based on their characteristic behaviour as given by Muthama (2002). These families include Exponential models which have the exponential or logarithmic functions involved, Growth models characterized by a monotonic growth from some fixed value to an asymptote, the yield density models which have the 'asymptotic' and 'parabolic' yield density relations, Processes producing sigmoidal or 'S'—shaped growth curves, and a miscellaneous family

described to include the sinusoidal fit, Gaussian model, Hyperbolic fit, Heat - capacity model, Rational function, etc. The best mathematical model which best simulated the March-May (MAM) and October -December (OND) seasonal rainfall over the three stations of analysis was chosen.

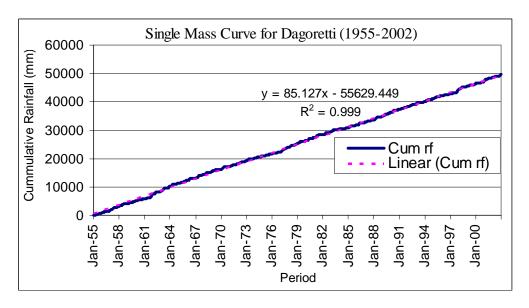


Figure 1. Single mass curve for Dagoretti.

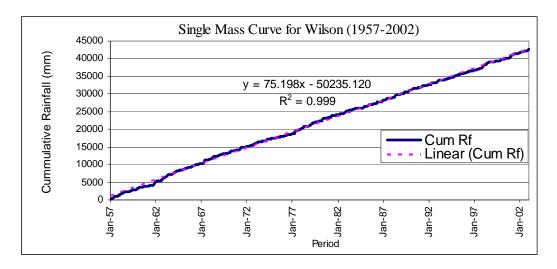


Figure 2. Single mass curve for Wilson.

Through the use of stepwise regression technique, the fourth degree polynomial was selected as the best fit for analyzing the data set. This technique involves choosing the variables, i.e., terms, to include in a multiple regression model. Forward stepwise regression starts with no model terms. At each step it adds the most statistically significant term (the one with the highest F statistic or lowest p-value) until there are none left. Backward stepwise regression starts with all the terms in the model and removes the least significant terms until

all the remaining terms are statistically significant. It is also possible to start with a subset of all the terms and then add significant terms or remove insignificant terms.

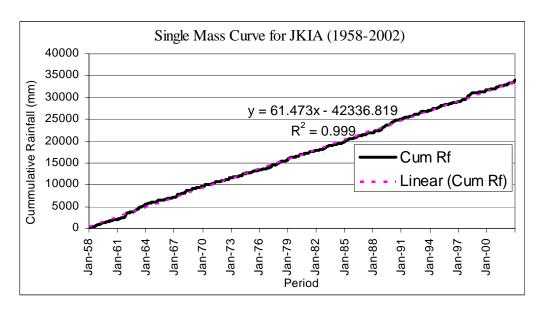


Figure 3. Single mass curve for JKIA.

The oscillations of the MAM and OND rainfalls followed a unique repetitive pattern which may be associated with solar cycle. Recent observation shows a strengthening of tropical circulation associated with the decadal variability in the tropical mean radiative energy budget (Chen *et al.*, 2002; Wielicki *et al.*, 2002). Hence, solar cycle period of ten (10) years was employed to get the fourth degree polynomial variables. The fourth degree polynomial function is of the form;

$$Y = a + bx + cx^{2} + dx^{3} + ex^{4}$$
(3)

where a,b,c, d and e are constants and x refers to the independent variable (also called the predictor) and Y is the dependent variable (also called the predictand). The error sum of squares (SSE) was computed for 5, 10 and 11 years periods. The sum of squared prediction errors (SSPE) is given by;

$$SSPE = \sum_{t=1}^{n} (y_t - \hat{y}_t)^2 \tag{4}$$

where \hat{y}_t is the one-step predicted value and y_t is the observed/recorded value.

3. Results and discussion

The results for data quality control tests and 4th degree polynomial fit of the MAM and OND rainfalls is as depicted below.

3.1 Data quality control results

The single mass curve was used to ascertain the quality of the data. Cumulative rainfall was plotted against time and the plot showed that data for the three stations used in the study were homogeneous. The rainfall records were therefore declared of good quality and hence suitable for climatological analysis. The single mass

curves are shown in Figures 1, 2 and 3. R-squared values (also called the coefficient of determination) have also been calculated by assuming a linear trendline. R-squared value is a number from 0 to 1 that reveals how closely the estimated values for the trendline correspond to the actual data. A trendline is most reliable when its R-squared value is at or near 1.

Figure 1 shows the single mass curve for Dagoretti, and it is clear from the plot that the data record over this station was of good quality as depicted in the almost straight line. The R^2 value for the mass curve fit was very close to 1 and the linear trendline equation was of the form y = mx + b; 'm' being the slope and 'b' the intercept.

From the single mass curve for Wilson, it is also evident that the data record used was of good quality as shown in Figure 2 above. The R^2 value was also very close to 1, a clear indication that the trendline was most reliable.

Figure 3 shows the single mass curve for JKIA. Similarly, the data record over this station was homogeneous as illustrated by the near straight line. The computed R^2 value was also very close to 1 and the linear trendline equation was of the form y = mx + b, where 'm' was the slope and 'b' was the intercept. Hence the data records used in this study were homogeneous for the three stations of interest, and hence the data was declared fit for use in the analysis. The linear trendlines for the various mass curve fits were reliable since they were very close to 1.

3.2 The error sum of squares (SSE)

The error sum of squares (SSE) was used to assess the deviation of predicted rainfall values from the observed values. The computed SSE values for 10 years were lower in comparison to the SSE values for 5 or 11 years. Hence, the sum of squared prediction errors (SSPE) was lower for 10 years as compared to 5 and 11 years. Table 1 below summarizes the SSPE value results obtained for 10 and 11 years respectively.

It is evident from Table 1 that the SSPE values for 10 years cycle were lower as compared to the SSPE values for 11 years cycle.

10 Years Cycle								
Station	Data period	SSPE for MAM	SSPE for OND					
Dagoretti	1957-2003	1139570.1	1371506					
JKIA	1958-2002	703804.56	1111207					
Wilson	1957-2002	1047649.8	1034849					
11 Years Cycle								
Station	Data period	SSPE for MAM	SSPE for OND					
Dagoretti	1957-2003	1199634.3	1426785					
JKIA	1958-2002	764238.11	1164118					
Wilson	1957-2002	1158147.2	1091048					

Table 1. SSPE values for 10 and 11 years cycles.

3.3 Inter-seasonal rainfall variability

Table 2 illustrates the rainfall variability during the March-May (MAM) and October-December (OND) seasons over Nairobi.

Table 2. MAM and OND seasonal rainfall variability. Range here refers to the difference between maximum and minimum rainfall whereas Std dev is the standard deviation.

March-May Season									
	Dagoretti		J	JKIA		Wilson			
	Range	Std dev	Range	Std dev	Range	Std dev			
Sum of 5 years	625.9	204.1	692.2	248.0	1485.5	528.6			
Sum of 10 years	312.4	142.7	188.6	78.4	458.2	210.7			
Sum of 11 years	362.8	161.1	714.1	354.5	805.0	371.3			
		October-De	ecember Sea	son					
	Dagoretti		J.	JKIA		Wilson			
	Range	Std dev	Range	Std dev	Range	Std dev			
Sum of 5 years	1320.7	380.8	947.0	344.6	834.8	315.9			
Sum of 10 years	1327.1	553.8	1156.6	560.6	1468.9	664.4			
Sum of 11 years	1539.2	638.9	1303.8	600.2	1237.1	518.2			

As observed from Table 2, the sum of MAM rainfall had minimal variability over a decade (10 years) as compared to 5 and 11 years where the variability was more pronounced, as depicted in the range and standard deviation values. Information on the minimal variability in MAM seasonal rainfall over a ten year period can be used in water planning and management over Nairobi.

The sum of October-December rainfall had minimal variability over Nairobi for 5 years period as compared to 10 and 11 years periods. Hence it is recommended that water planning and management over Nairobi for the OND season should be for a period of 5 years.

Observed and Modelled MAM Rainfall in Dagoretti

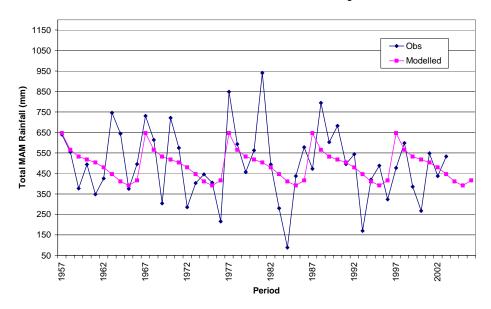


Figure 4. Recorded and simulated MAM rainfall at Dagoretti.

Observed and Modelled OND Rainfall in Dagoretti

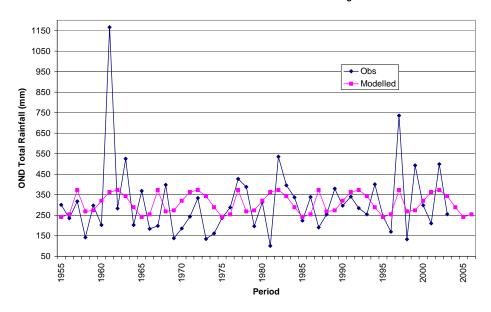


Figure 5. Recorded and simulated OND rainfall at Dagoretti.

Observed and Modelled MAM Rainfall in JKIA

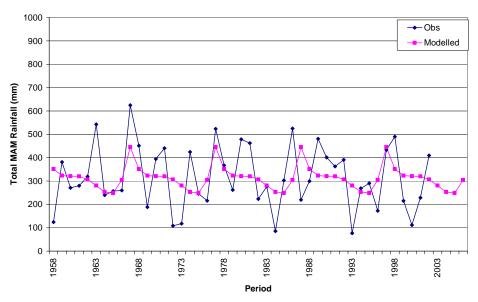


Figure 6. Recorded and simulated MAM rainfall at JKIA.

Observed and Modelled OND Rainfall in JKIA

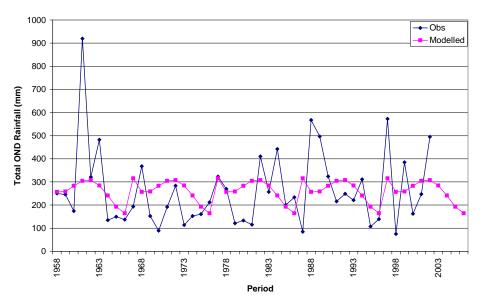


Figure 7. Recorded and simulated OND rainfall at JKIA.

Observed and Modelled MAm Rainfall in Wilson

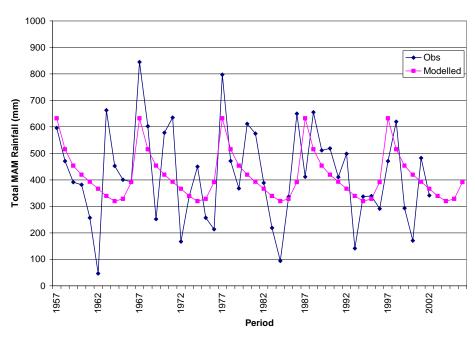


Figure 8. Recorded and simulated MAM rainfall at Wilson.

3.4 Fourth degree polynomial fit results

Through various curve fitting analyses carried out, it was evident that the 4th degree polynomial function was the best fit for the three stations of analysis selected (i.e Dagoretti, JKIA and Wilson). The simulated March-May (MAM) and October-December (OND) rainfalls using the 4th degree polynomial fit for the three stations of analysis is as depicted in Figures 4-9.

Observed and Modelled OND Rainfall in Wilson

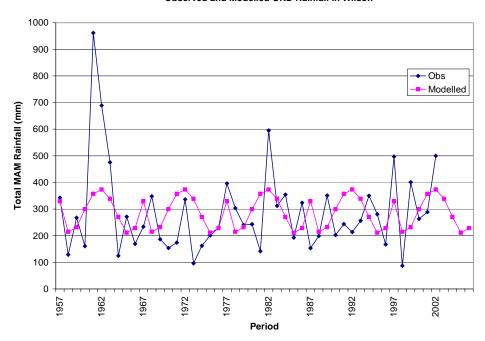


Figure 9. Recorded and simulated OND rainfall at Wilson.

The 4th degree polynomial function used to get the simulated MAM rainfall over Dagoretti was of the form

$$Y = a + bx + cx^2 + dx^3 + ex^4$$
,

where

$$a = 812.6$$
, $b = -223.5$, $c = 65.5$, $d = -8.6$, $e = 0.385$

From Figure 4, the March-May rainfall over Dagoretti seems to have a significant decadal cycle. This 10 years cycle is analogous to the solar cycle. Though some of the peaks are not well captured, but the overall rainfall trend is clearly illustrated.

The polynomial equation for the simulated OND rainfall over Dagoretti was of the form

$$Y = a + bx + cx^2 + dx^3 + ex^4$$
,

where

$$a = 673.8$$
, $b = -433.1$, $c = 151.1$, $d = -19.6$, $e = 0.836$.

From Figure 5, the 4th degree polynomial fit depicts the mean rainfall over Dagoretti. The predicted rainfall seems also to have a unique oscillation pattern of 10 years. A comparison of Figures 4 and 5 shows that there is minimum variability in the March-May seasonal rainfall over Dagoretti as compared to the October-December seasonal rainfall. This variability may be associated with the systems that influence the short (OND) and the long (MAM) rainfalls over Kenya such as the inter-tropical convergence zone (ITCZ), easterly and westerly winds, monsoons and orographic features among others.

The polynomial equation for the simulated MAM rainfall over JKIA was of the form

$$Y = a + bx + cx^{2} + dx^{3} + ex^{4}$$
,

where

$$a = 654.4$$
, $b = -287.3$, $c = 89.1$, $d = -11.6$, $e = 0.523$.

From Figure 6, similar decadal cycle pattern as observed over Dagoretti for the same season was repeated. The rainfall peaks slowly decayed over a period of 10 years.

The polynomial equation for the simulated OND rainfall over JKIA was of the form

$$Y = a + bx + cx^{2} + dx^{3} + ex^{4}$$
.

where

$$a = 482.3$$
, $b = -238.6$, $c = 82.2$, $d = -10.5$, $e = 0.432$.

Figure 7 depicts the mean rainfall over JKIA, and has the same decadal cycle pattern as Figure 5.

The polynomial equation for the simulated MAM rainfall at Wilson was

$$Y = a + bx + cx^{2} + dx^{3} + ex^{4}$$
,

where

$$a = 843.3$$
, $b = -272.3$, $c = 69.8$, $d = -8.6$, $e = 0.392$.

Figure 8 above has the same decadal cycle pattern as Figures 6 and 4. By the use of the 4th degree polynomial fit, the prediction of the peak and general rainfall for the March-May Season seems to be well simulated over Wilson as compared to Dagoretti and JKIA.

The polynomial equation for OND rainfall over Wilson was

$$Y = a + bx + cx^{2} + dx^{3} + ex^{4}$$
.

where

$$a = 686.7$$
, $b = -520.7$, $c = 186.9$, $d = -24.5$, $e = 1.051$.

Figure 9 has the same decadal cycle pattern as Figures 7 and 5 above. Even though Figure 9 does not predict the peaks well, but the general pattern of the mean seasonal rainfall has been captured.

4. Conclusions

Amongst the numerous mathematical functions considered, the fourth degree polynomial was found to be the best fit for rainfall prediction purposes in Nairobi, for both the short rains (October-December) and the long rains (March-May). The ten years period cycle of rainfall which was used to generate the 4th degree polynomial function, corresponded to the sunspots cycle of approximately 11.3 years. The ten years period cycle had the least sum of squared prediction errors (SSPE), as compared to five and eleven years period cycles for both the long and the short rains.

The sum of the March-May rainfall had minimal variability over a ten-year period as compared to 5 and 11 years periods. Unlike the March-May rainfall, the sum of October-December rainfall had minimal variability over a 5-year period.

The 4th degree polynomial function was found to predict the peaks and the general pattern of March-May seasonal rainfall over Nairobi with minimal/acceptable error values. For the October-December season, the 4th degree polynomial function simulated the general pattern of the mean values. From this study, it can be concluded that planning for water use and management over Nairobi for the March-May season should be for a ten year period, since we have minimal variability during this period. For October-December season, it is recommended that planning and management of water over Nairobi should be for a period of 5 years.

5. References

CAMBERLIN, P. and PHILIPPON N. 2001. The east African March-May rainy season associated atmosphere dynamics and predictability over the 1968-97 period. *Am. Meterol. Soc.* **15:** 1002-1019.

- CHELLIAH, M. and BELL, G.D. 2004. Tropical multidecadal and interannual climate variability in the NCEP–NCAR Reanalysis. *J. Climate*, **17:** 1777-1803.
- CHEN, H.L., and RAO, A.R. 2002. Testing hydrologic time series for stationarity. J. Hydrol. Eng. 7: 129-136.
- CHRISTINE, L.G., JOHN, D.A and SCOTT, V.O. 1998. Mapping monthly precipitation, temperature, and solar radiation for Ireland with polynomial regression and a digital elevation model. *J. Climate*, **10**: 35–49.
- FOLLAND, C.K., OWEN, J., WARD, N. and COLMAN, A. 1991. Prediction of seasonal rainfall in the Sahel region using empirical and dynamical methods. *J. Forecasting*, **10:** 21-56.
- GRAHAM, N.E. 1994. Decadal scale climate variability in the tropical and North Pacific during the 1970s and 1980s: Observations and model results. *J. Climate Dyn.*, **10:** 135-162.
- HOERLING, M.P., HURRELL J.W. and XU T. 2001. Tropical origins for recent North Atlantic climate hange. *J. Science*, **292**: 90–92.
- KAWAMURA, R., SUGI M. and SATO N. 1995a: Interdecadal and interannual variability in the northern extratropical circulation simulated with the JMA global model. Part I: Wintertime leading mode. *J. Climate*, **8:** 3006–3019.
- KAWAMURA, R., SUGI, M. and SATO N. 1995b. Interdecadal and interannual variability in the northern extratropical circulation simulated with the JMA global model. Part II: Summertime leading mode. *J. Climate*, **8:** 3006–3019.
- LAU, N.C. and NATH M. 1994. A modeling study of the relative roles of tropical and extratropical SST anomalies in the variability of the global atmosphere—ocean system. *J. Climate*, **7:** 1184-1207.
- LIVEZEY, R.E. and SMITH T.M. 1999. Covariability of aspects of North American climate with global sea surface temperatures on interannual to interdecadal time scales. *J. Climate*, **12**: 289–302.
- MO, K.C., BELL G.D. and THAIW W. 2001. Impact of sea surface temperature anomalies on the Atlantic tropical storm activity and West African rainfall. *J. Atmos. Sci.*, **58:** 3477-3496.
- MUTAI, C.C. and WARD, M.N. 2000. East African rainfall and tropical circulation / convection on intraseasonal to interannual timescales. *J. Climate*, **13:** 3915-3939.
- MUTHAMA N.J. 2002. A simple atmospheric systems risk 'indicator' model: Application to fog. *International journal of BioChemiPhysics*, Vol 11&12(Nos. &2), 2003.
- OGALLO, L.J., JANOWIAK, J.E. and HALPERT M.S. 1988. Teleconnection between seasonal rainfall over East Africa and global sea-surface temperature anomalies. *J. Met. Society.* **66-6**: Ser. II, 807-822.
- OGALLO L.A.J. 1977. Perodicities and trends in the annual rainfall over Africa. Msc. thesis, Univ. of Nairobi.
- POTTS, A. 1971. Application of harmonic analysis to the study of east African rainfall data. *J. Trop. Geog.*, **34:** 31-42.
- RODHE, H. and VIRJI H. 1996. Trends and periodicities in East African rainfall data. *Mon. Wea. Rev.*, **104**: 307-315.
- ROPELEWSKI, C.F. and HALPERT, M.S. 1987. Global and regional scale precipitation patterns associated with EL Nino/southern oscillation. *Mon. Wea. Rev.*, **115**: 1601-1626.
- ROWELL, D.P., ININDA, J.M and WARD, M.N. 1994. The impact of global sea surface temperature patterns on seasonal rainfall in East Africa. Proc. Int. Conf. Monsoon Variability and Prediction, Trieste, Italie, 9-13 May 1994. *WMO/TD* n619. 666-672.
- WIELICKI, B.A. and COAUTHORS, 2002. Evidence for large decadal variability in the tropical mean radiative energy budget. *J.Science*, **295**: 841-843.
- W.M.O. 1962. Statistical analysis and prognosis in meteorology. Technical note no. 71. WMO no.178. TP.88.
- W.M.O. 1966. Climate change. Technical notes no., 79. WMO no.195 TP.100.

Received 9 March 2007 Accepted 22 September 2008