Stochastic modelling of regional annual rainfall anomalies in East Africa

LABAN A. J. OGALLO, *Department of Meteorology, University of Nairobi, P.O. Box 30197, Nairobi, Kenya*

SUMMARY *ARIMA* (p, d, q) *models were fitted to areal annual rainfall of two homogeneous regions in East Africa with rainfall records extending between the period 1922–80. The areal estimates of the regional rainfall were derived from the time series of the first eigenvector, which was significantly dominant at each of the two regions. The first eigenvector accounted for about 80% of the total rainfall variance in each region.*

The class of ARIMA (p, d, q) *models which best fitted the areal indices of relative wetness/dryness were the ARMA (3, 1) models. Tests of forecasting skill however indicated low skill in the forecasts given by these models. In all cases the models accounted for less than 50% of the total variance.*

Spectral analysis of the indices time series indicated dominant quasi-periodic fluctuations around 2.2–2.8 years, 3–3.7 years, 5–6 years and 10–13 years. These spectral bands however accounted for very low proportion of the total rainfall variance.

Introduction

Many attempts have been made to use statistical methods in the forecasting of time series. Such time series have included agricultural, hydrological and meteorological series. Most of these statistical techniques are based on the statistics derived from the past and present data samples. Some of these statistical methods are based on the probability of occurrence of certain events, while many have been developed through correlation methods. The correlation methods often express the predicted values in terms of some chosen predictors. The predictors for the meteorological time series have included empirical orthogonal functions (Lorenz, 1959).

In forecasting hydrological time series, autocorrelation and response functions are sometimes used to describe the functional relationship between the input and output series (Kato, 1983). In case of agricultural time series (e.g. seasonal agricultural output), attempts have been made to forecast future agricultural output using the statistics derived from the major influencing factors (WMO, 1977).

During the last few years many sophisticated statistical forecasting models have been developed due to the availability of advanced computers. Such models include the Autoregressive Integrated Moving Average (ARIMA) models (Box & Jenkins, 1976). These ARIMA models have been widely used in attempts to forecast agricultural, hydrological and meteorological time series.

In this study an attempt was made to examine the class of ARIMA models that may best fit the annual areal rainfall anomalies of some two chosen regions in East Africa. The forecasting skill of the best fitting model was also investigated.

Rainfall was chosen as the climatic variable in this study because it is the climatic element of maximum significance, especially in the tropical regions where the inter-annual variations of the other climatic elements are not so large. Precipitation also controls the rain-fed agricultural and hydrological activities which are still dominant in many developing nations. The inter-annual variations in precipitation will therefore be generally reflected in all rain dependent activities.

1 Methodology

In order to determine the annual rainfall anomalies, composite indices were used to represent the inter-annual areal wetness/dryness. These composite indices were developed through empirical orthogonal analysis (Ogallo, 1981).

2 Composite Indices of areal wetness/dryness

In regions with high spatial rainfall variability many traditional methods of estimating areal rainfall have been noted to be generally unreliable. The use of empirical orthogonal functions have been suggested as one of the alternative approaches (Johnson, 1980). Similar approach was adopted by Ogallo (1981), in estimating areal rainfall anomalies for some homogeneous regions of East Africa. The homogeneous rainfall regions were delineated through principal component analysis (Ogallo, 1980). These homogeneous groups are marked by letters A to L in Fig. 1. The two regions which are marked B and E in Fig. 1 had a single dominant eigenvector, and have been chosen for this study.

Empirical approach of estimating areal rainfall anomalies were included in this study due to uneven distribution of the rainfall stations within the two regions. The topographical features were also varying. Indentical temporal rainfall patterns were however evident at the stations within these two regions (Ogallo, 1981, 1982).

In order to determine the areal composite indices of relative wetness/dryness for regions B and E, the annual rainfall data for the two regions were independently subjected to principal component analysis. Regions B and E had 17 and 25 stations respectively with annual rainfall records extending from 1922 to 80.

It was observed from the results of principal component analysis that whereas only about 34% of the total variance was accounted for by the first eigenvector when about 100 rainfall stations spread all over East Africa were considered (Ogallo, 1980), more than 77% of the total rainfall variance could be accounted for by this eigenvector when only rainfall stations within regions B and E were independently used. The high dominance of a single eigenvector at each of the two homogeneous rainfall regions confirms the expected spatial similarity in the temporal patterns of rainfall within the two regions. The dominance of only one eigenvector also indicates that only one empirical orthogonal could be used to adequately describe spatial variability of rainfall within the two regions (Johnson, 1980). This principle

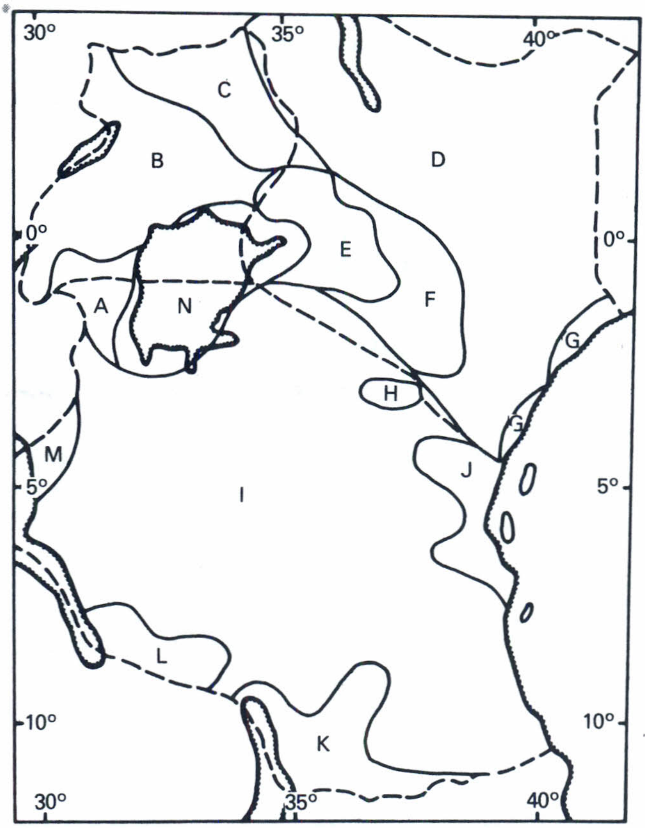

FIG. 1. Homogeneous rainfall divisions from principal component analysis (Ogallo, 1980).

was used in deriving the areal indices of relative wetness/dryness for regions B and E during the period 1922–80 (Ogallo, 1981).

From principle component analysis, the regression weights on the eigenvectors are unique for each station, and they represent the degree of correlation between rainfall at any station and that particular eigenvector. These weights may therefore be used to describe mean areal rainfall conditions in any homogeneous rainfall region with a single dominant eigenvector. In this study the composite index of the areal relative wetness/dryness (F_t) was defined as:

$$F_t = \frac{1}{m} \sum_{j=1}^{m} a_j Z_{tj} \tag{1}$$

where Z_{tj} represents standardised annual rainfall at station j in year t, a_j regression weight on the first eigenvector at station j, and m the number of stations within the homogeneous rainfall division. The annual records were standardised by subtracting the mean and dividing with the standard deviation.

Since the yearly data are standardised, and since regression weights are unique for each station, the composite index will have high positive values for wet years and high negative values for dry years, thus giving a measure of spatially weighted regional rainfall deviations (Ogallo, 1981). Similar empirical methods have been used as indicators of climatic fluctuations by many scientists (Murakami, 1980; Akiyama, 1981; Gregory, 1979).

When all values of a_j in equation (1) are units, this expression reduces to simple areal average of rainfall deviations. It was observed from the computed values of the indices, simple areal averages and Z_{ij} values that the composite indices gave satisfactory description of the temporal variations of regional annual rainfall series at regions B and E (Ogallo, 1981).

In this study ARIMA models were fitted to the composite indices of relative wetness/dryness in regions B and E. ARIMA models were also fitted the corresponding simple areal averages.

ARIMA models

In order to fit Autoregressive Integrated Moving Average (ARIMA) models to the regional series, an appropriate model was searched for within the extensive class of ARIMA models. This approach involves the construction of linear combination of autoregressive (AR) and Moving Average (MA) processes. The AR process expresses an observation of the time series in terms of the previous observations plus a residual term, while the MA process expresses the observations as a linear combination of independent residual terms.

Since the theory of ARIMA models have been fully discussed by many authors (Box & Jenkins, 1976; Parzen, 1974; Kendall, 1976; Anderson, 1977), only a brief mathematical account will be included here.

ARIMA models of order p, d, q generally abbreviated as ARIMA (p, d, q), may be represented as:

$$W_t = \nabla^d Z_t \tag{2}$$

Where W_t and Z_t are the stationary and non-stationary time series, ∇ the backward difference operator and d the order of differencing.

For a stationary time series the order of differencing d is zero, and the ARIMA (p, d, q) model reduces to the Autoregressive Moving Average (ARMA(p,q)) model. The ARMA (p,d) model may be expressed as:

$$W_t = \mu + \sum_{i=1}^{p} \phi_j W_{t-j} + E_t - \sum_{j=1}^{q} \theta_j E_{t-j} \tag{3}$$

Where ϕ and θ_j are regression constants, p and q respectively represent the orders of AR and MA processes, while μ is a constant and E_k represents moving average terms.

ARIMA models have been applied to climatological series by many authors including Gray (1976), Jakobsson (1977), Dyer (1977), Davis & Rappoport (1974), Rao (1980), Katz & Skaggs (1981), Delleur & Kavvas (1978) and Peagle & Haslam (1982).

In this study ARIMA models were fitted to the annual areal rainfall time series discussed in the previous sections.

3 Results and discussion

The results from the patterns of the autocorrelation and partial autocorrelation functions indicated that the variance of the first eigenvector time series increased considerably when any differencing scheme was applied. Minimum residual variance was achieved with no differencing or transformation indicating the stationary characteristics of the original time series. Stationarity and invertibility conditions for ARIMA models have been discussed by many authors. Although no significant

AKIYAMA, T. (1981) Time and spatial variations of heavy snowfalls in the Japan Sea coastal region, *Journal of Meteorology Society Japan*, 59(4), pp. 578–590.
ANDERSON, O.D. (1977) Time series and forecasting. Another look at the Box-Jenkins Approach, *Statistician*, 26, pp. 285–303.
BOX, G.E.P. & JENKINS, G.M. (1976) *Time series Analysis Forecasting and Control*, (Holden-Day Press).
BROWN, L.H. & COCHÉME, J. (1973) A study of agroclimatology of the highlands of Eastern Africa, *WMO Technical Note 125*.
CHANDER, S., KUMAR, A. & GOYAL, S.K. (1979) Comments on stochastic models from monthly rainfall forecasting and synthetic generation, *Journal of Applied Meteorology*, 18, p. 1380.
DAS, P.K. (1983) Droughts and famines in India—a historical perspective, *Mausam*, 34, pp. 123–130.
DAVIS, J.M. & RAPPOPORT, P.N. (1974) The use of time series analysis techniques in forecasting meteorological droughts, *Monthly Weather Review*, 102, pp. 176–180.
DELLEUR, J.M. & KAVVAS, M.L. (1978) Stochastic models for monthly rainfall forecasting and synthetic generation, *Journal of Applied Meteorology*, 17, pp. 1528–1536.
DYER, T.G.J. (1977) On application of some stochastic models to precipitation forecasting, *Quarterly Journal of Royal Meteorological Society*, 103, pp. 177–189.
GRAY, B.M. (1976) Medium term fluctuations of rainfall in south-eastern England, *Quarterly Journal of the Royal Meteorological Society*, 102, pp. 627–638.
GREGORY, S. (1979) The definition of wet and dry periods for discrete regional units, *Weather*, 34, pp. 363–369.
HOFFMAN, R.N. & KALNARY, E. (1983) Lagged average forecasting, an alternative to Monte Carlo Forecasting, *Tellus*, 35A, pp. 100–118.
JAKOBSSON, T. (1977) Long Range Forecasting by ARIMA models, *5th Conference on Probability and Statistics on Atmospheric Science, American Meteorological Society*, pp. 22–30.
JOHNSON, D.M. (1980) An Index of Arizona Summer Developed through eigenvector analysis, *Journal of Applied Meteorology*, 19, pp. 849–856.
KATO, P. (1983) Rainfall-Run-off relation in Ruvu Basin of Tanzania, *MSc Thesis* (University of Nairobi).
KATZ, R.W. & SKAGGS, R.H. (1981) On the use of autoregressive moving average processes to model meteorological time series, *Monthly Weather Review*, 109, pp. 479–484.
KENDALL, M.G. (1976) *Time Series* (London, Charles Griffin).
LORENZ, E.N. (1959) Prospects for statistical weather forecasting. *Final report, Statistical Forecasting Project, Massachusetts Institute of Technology*, pp. 1–26.
MUKARAMI, T. (1980) Temporal variation of satellite-observed outgoing long-wave radiation over the winter monsoon region, Part I. *Monthly Weather Review*, 108, pp. 408–426.
OGALLO, L.J. (1980) Regional classification of East Africa rainfall stations into homogenous groups using the method of principal component analysis, *Statistical Climatology, Developments Atmospheric Science*, 13, pp. 255–266.
OGALLO, L.J. (1981) The use of eigenvector analysis in the study of drought and wetness conditions in East Africa, *Tropical Drought Conference*, 7–11 September, New Delhi, India.
OGALLO, L.J. (1982) Quasi-periodic patterns in the East African rainfall records, *Kenya Journal of Science and Technology*, A3, pp. 43–54.
OZAKI, T. (1977) On the order determination of ARIMA models, *Applied Statistics*, 26, pp. 290–301.
PARZEN, E. (1974) Some recent advances in time series modelling, *IEEE*, AC19, pp. 723–730.
PEAGLE, J.N. & HASLAM, R.B. (1982) Statistical prediction of 500 mb height using eigenvectors, *Journal of Applied Meteorology*, 21, pp. 127–138.
RAO, A.R. (1980) Stochastic analysis of annual rainfall affected by urbanization, *Journal of Applied Metereology*, 19, pp. 41–52.
SCHWARZ, G. (1978) Estimating the dimension of a model, *Annals of Statistics*, 6, pp. 461–464.
THAPLILAL, V. (1982) Stochastic dynamic model for long-range prediction of monsoon rainfall in Peninsula India, *Mausam*, 33, pp. 399–404.
TUNNICLIFFE-WILSON, G. (1973) Discussion on 'Box-Jenkins' seasonal forecasting, *Journal of the Royal Statistical Society*, A136, pp. 337–352.
WMO (1977) Crop-weather models and their use in yield assessment, *WMO tech. No. 151*.

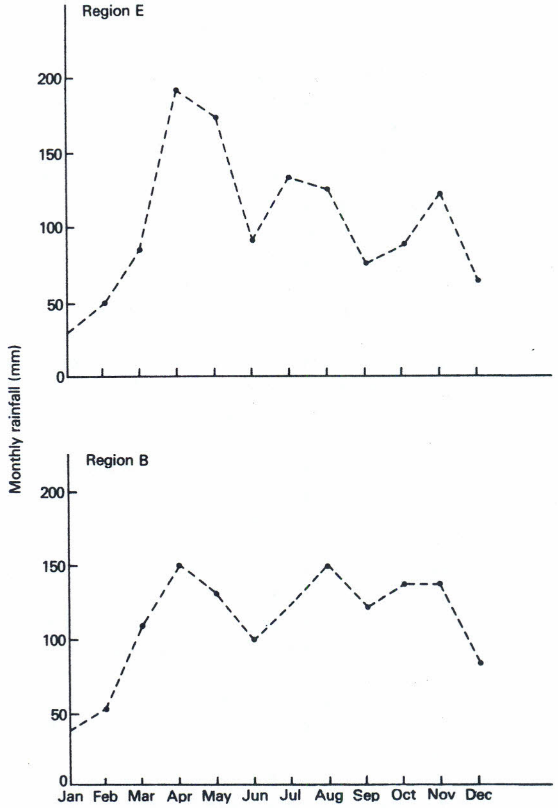

FIG. 3. Mean seasonal rainfall patterns.

Conclusion

In this study ARMA (3,1) models were observed to adequately fit regional rainfall anomalies of some two choosen regions of East Africa. The models however underestimated the regional rainfall anomalies. In all cases the models accounted for less than 50% of the total rainfall variance. The spectral analysis of the areal composite indices indicated dominant quasi-periodic fluctuations around 2.2–2.8 years, 3–3.7 years, 5–6 years and 10–13 years. This large range in the spectral bands indicate instability in the periods of the observed quasi-periodic fluctuations. This instability, together with the low variance accounted for by the spectral peaks, may give some explanations to the low forecasting skill of the fitted ARMA (3,1) models. A method of improving the autocorrelation characteristics of the rainfall series has been suggested through the use of seasonal rainfall data.

The fitting of a single class of ARIMA models (ARMA (3,1)) to rainfall records from two different regions, together with the similarity in the patterns of the spectral peaks indicate some commoness in the major rain generating functions within the two regions. Rainfall belts in East Africa generally migrate with the ITCZ. In some areas the seasonal patterns are modified by local factors.

REFERENCES

AKAIKE, H. (1974) A new look at the statistical model identification, *IEEE*, 19, pp. 716–723.

AKIYAMA, T. (1981) Time and spatial variations of heavy snowfalls in the Japan Sea coastal region, *Journal of Meteorology Society Japan*, 59(4), pp. 578–590.

ANDERSON, O.D. (1977) Time series and forecasting. Another look at the Box-Jenkins Approach, *Statistician*, 26, pp. 285–303.

BOX, G.E.P. & JENKINS, G.M. (1976) *Time series Analysis Forecasting and Control*, (Holden-Day Press).

BROWN, L.H. & COCHÈME, J. (1973) A study of agroclimatology of the highlands of Eastern Africa, *WMO Technical Note 125*.

CHANDER, S., KUMAR, A. & GOYAL, S.K. (1979) Comments on stochastic models from monthly rainfall forecasting and synthetic generation, *Journal of Applied Meteorology*, 18, p. 1380.

DAS, P.K. (1983) Droughts and famines in India—a historical perspective, *Mausam*, 34, pp. 123–130.

DAVIS, J.M. & RAPPOPORT, P.N. (1974) The use of time series analysis techniques in forecasting meteorological droughts, *Monthly Weather Review*, 102, pp. 176–180.

DELLEUR, J.M. & KAVVAS, M.L. (1978) Stochastic models for monthly rainfall forecasting and synthetic generation, *Journal of Applied Meteorology*, 17, pp. 1528–1536.

DYER, T.G.J. (1977) On application of some stochastic models to precipitation forecasting, *Quarterly Journal of Royal Meteorological Society*, 103, pp. 177–189.

GRAY, B.M. (1976) Medium term fluctuations of rainfall in south-eastern England, *Quarterly Journal of the Royal Meteorological Society*, 102, pp. 627–638.

GREGORY, S. (1979) The definition of wet and dry periods for discrete regional units, *Weather*, 34, pp. 363–369.

HOFFMAN, R.N. & KALNARY, E. (1983) Lagged average forecasting, an alternative to Monte Carlo Forecasting, *Tellus*, 35A, pp. 100–118.

JAKOBSSON, T. (1977) Long Range Forecasting by ARIMA models, *5th Conference on Probability and Statistics on Atmospheric Science, American Meteorological Society*, pp. 22–30.

JOHNSON, D.M. (1980) An Index of Arizona Summer Developed through eigenvector analysis, *Journal of Applied Meteorology*, 19, pp. 849–856.

KATO, P. (1983) Rainfall-Run-off relation in Ruvu Basin of Tanzania, *MSc Thesis* (University of Nairobi).

KATZ, R.W. & SKAGGS, R.H. (1981) On the use of autoregressive moving average processes to model meteorological time series, *Monthly Weather Review*, 109, pp. 479–484.

KENDALL, M.G. (1976) *Time Series* (London, Charles Griffin).

LORENZ, E.N. (1959) Prospects for statistical weather forecasting. *Final report, Statistical Forecasting Project, Massachusetts Institute of Technology*, pp. 1–26.

MUKARAMI, T. (1980) Temporal variation of satellite-observed outgoing long-wave radiation over the winter monsoon region, Part I. *Monthly Weather Review*, 108, pp. 408–426.

OGALLO, L.J. (1980) Regional classification of East Africa rainfall stations into homogenous groups using the method of principal component analysis, *Statistical Climatology, Developments Atmospheric Science*, 13, pp. 255–266.

OGALLO, L.J. (1981) The use of eigenvector analysis in the study of drought and wetness conditions in East Africa, *Tropical Drought Conference*, 7–11 September, New Delhi, India.

OGALLO, L.J. (1982) Quasi-periodic patterns in the East African rainfall records, *Kenya Journal of Science and Technology*, A3, pp. 43–54.

OZAKI, T. (1977) On the order determination of ARIMA models, *Applied Statistics*, 26, pp. 290–301.

PARZEN, E. (1974) Some recent advances in time series modelling, *IEEE*, AC19, pp. 723–730.

PEAGLE, J.N. & HASLAM, R.B. (1982) Statistical prediction of 500 mb height using eigenvectors, *Journal of Applied Meteorology*, 21, pp. 127–138.

RAO, A.R. (1980) Stochastic analysis of annual rainfall affected by urbanization, *Journal of Applied Metereology*, 19, pp. 41–52.

SCHWARZ, G. (1978) Estimating the dimension of a model, *Annals of Statistics*, 6, pp. 461–464.

THAPLILAL, V. (1982) Stochastic dynamic model for long-range prediction of monsoon rainfall in Peninsula India, *Mausam*, 33, pp. 399–404.

TUNNICLIFFE-WILSON, G. (1973) Discussion on 'Box-Jenkins' seasonal forecasting, *Journal of the Royal Statistical Society*, A136, pp. 337–352.

WMO (1977) Crop-weather models and their use in yield assessment, *WMO tech. No. 151*.