JP1.8 **WAVELET ANALYSIS ON VARIABILITY, TELECONNECTIVITY AND PREDICTABILITY OF EAST AFRICA RAINFALL**

Davison Mwale[a], Thian Yew Gan[a]* and Samuel Shen[b]
[a] Department of Civil and Environment Engineering, University of Alberta, Alberta, Canada
[b] Department of Mathematics, University of Alberta, Alberta, Canada

1. INTRODUCTION

Problems of quantifying rainfall variability in space with respect to climatic elements such as SST, winds, air temperature and specific humidity have been a subject of numerous studies for Eastern Africa, (EA) (e.g., Mutai et al. 1998; 2002; Ntale et al. 2003), and other studies as summarized in Beltrando (1990). Following some of these studies, EA has been divided into between 6 and 20 zones of homogenous rainfall variability, found to teleconnect to SST variability in the Atlantic, Indian and Pacific oceans, (Mutai et al. 1998; Ntale et al. 2003) and atmospheric signals, Philippon et al. (2002). Many of these studies applied a combination of time domain multivariate statistics such as harmonic analysis, Fourier analysis, linear regressions and cross correlations to climate data to determine relationships between and among the climate elements. A common and well-recognized shortcoming of such approaches has been the assumption that climate data is stationary and linear, a criterion few data sets from natural phenomena satisfy (Huang et al., 1998).

With unstable relationships between rainfall in EA and SST in the oceans and atmospheric circulation fields highlighted by erratic rainfall between 1965 and 1997, one of the major challenges for meteorologists has been to predict the nature of this variability in the rainfall of EA over different spatial and temporal scales. 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 EA rainfall. However, it has never been clear as to when these oscillations occurred in the rainfall.

*Corresponding Author address: Thian Yew Gan, Univ. of Alberta, Dept of Civil and Environmental Engineering, Edmonton, Alberta AB T6G 2G7, Canada; Email: TGan@civil.ualberta.ca

In investigating EA rainfall variability and its relationship to other climatic elements it has become imperative to analyze the irregularly distributed events in time that depict non-stationary power at many different frequencies.

To meet this requirement, this study used wavelet analysis and wavelet based empirical orthogonal function analysis (WEOF), also known as wavelet principal component analysis (WLPCA) and wavelet independent component analysis (WLICA) to study the variability, teleconnectivity and predictability of EA rainfall. In this study, two sets of information from the wavelet spectra are used. These are (1) energy coefficients (also known as power) of individual scales of the wavelet spectra, and (2) the average wavelet spectra energy over some significant scales (called scale averaged wavelet power, SAWP). Combinations of empirical orthogonal function analysis and wavelets analysis have previously been applied to multivariate statistical process monitoring (e.g., Bakshi 1998).

2. DATA AND METHODS.

EA has two rainy seasons, the September-November (SON) and March-May (MAM). Monthly rainfall data (1950-1995) from 21 grid locations at a resolution of 2.5°×3.75° latitude and longitude was extracted for EA (4ºS-12ºS, 28ºE-44ºE), see Fig 1.The rainfall data was provided by the UK meteorological office.

The monthly rainfall data was transformed into seasonal data by summing the monthly values for each grid location. Monthly SST anomaly grid data (1950-1995) at 5^{o} x 5^{o} latitude and longitude resolution was extracted from the Indian (20ºN-40ºS, 40ºE-105ºE) and Atlantic (10ºN-30ºS, 50ºW-10ºE) Ocean. This data was transformed into seasonal and annual data by computing 3-month averages for JFM, AMJ, JAS and SON and annual averages, respectively. The SST dataset is part of MOHSST6 and

was provided by the UK meteorological office.

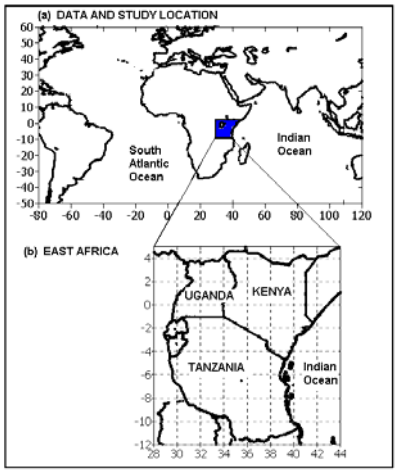

Fig. 1. (a) Study location (b) detailed location and description of EA (Uganda, Kenya and Tanzania)

2.2 Wavelet Analysis

Wavelets are a set of limited duration waves, also called daughter wavelets, because they are formed by dilations and translations of a single prototype wavelet function $\psi(t)$, where t is real valued, called the basic or mother wavelet (Castleman, 1996). The mother wavelet designed to oscillate like a wave, is required to span an area that sums to zero, and die out rapidly to zero as t tends to infinity to satisfy the so called "admissibility" condition.

$$\int \Psi(t)dt = 0 \qquad (1)$$

A set of wavelets can be generated by translating and scaling the basic wavelet as follows

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{a}}\Psi\left(\frac{t-b}{a}\right) \qquad (2)$$

where the scale (width) of the wavelet and translated position along the t-axis are a and b respectively. By continuously varying a long b, a picture is constructed depicting how the energy over various frequencies varies with time. The parameters a and b are real and a, always positive, may take continuous or a discrete values. The quantity $a^{-1/2}$ in Equation (2) is an energy normalization term, which ensures that energy of the mother, and daughter wavelets remain the same over all scales and making it possible to directly compare wavelet transforms of one time series with another (Torrence and Compo, 1998).

The wavelet transform of a real signal X (t) with respect to the mother wavelet is a convolution integral given as

$$W(b,a) = \frac{1}{\sqrt{a}}\int_0^T X(t)\Psi^*\left(\frac{t-b}{a}\right)dt \qquad (3)$$

where ψ^* is the complex conjugate of ψ. In this equation, $W(b,a)$ is a wavelet spectrum, a matrix of energy coefficients of the decomposed time series X (t). A faster and much more efficient way to compute the wavelet transform is done in the Fourier space using the Fourier transform of a discrete time series, X (t), as

$$W_t(a) = \sum_{k=0}^{T} \hat{X}\,\hat{\Psi}^*(s\omega_k)e^{i\omega_k n\delta t} \qquad (4)$$

where the caret symbolizes Fourier Transform, k is the frequency index (0,…T) and $\Psi(s\omega_k)$ is the Fourier transform of the wavelet function. The wavelet spectrum was computed using a discrete set of 20 scales starting at 2 years in fractional power of two using

$$s_j = s_o 2^{j\delta j} \qquad (5)$$

where s_0 is twice the sampling rate,j =0,1,….20, and δj = 0.25, thus giving scales(periods) ranging from 2 to 64 years. The wavelet transform of a time series contains a wealth of information, which can be condensed over a range of scales to construct the scale averaged wavelet power (SAWP) and used in multivariate analysis.

$$W_t^2 = \frac{\delta_j\delta_t}{C_\delta}\sum_{j=j_1}^{j_2}\frac{\left|W_t(a_j)\right|^2}{a_j} \qquad (6)$$

where Cδ is the reconstruction factor that takes on values depending on the mother

wavelet used, δj is a factor for scale averaging and δt is the sampling period. The global wavelet spectrum shows dominant oscillations present in a time series. The local wavelet power shows how the dominant oscillations vary with time. To examine whether two remotely located multivariate time series are related to each other or if one modulates the other over a number of scales, the individual scale power or the SAWP can be constructed by varying the scale, *a*, and computing the weighted sum of the wavelet power over those scales. Using the results of global spectra computed for some selected SST fields located in the Indian and Atlantic Oceans and rainfall fields located in EA, we extracted energy of some individual scale and SAWP from the 2-8 year range. To compute the wavelet power for this study, the Morlet wavelet (k = 6), was used because its structure resembles that of a rainfall time series.

$$\Psi(t) = \pi^{-1/4} e^{i6t} e^{-t2/2} \quad (7)$$

2.3 Wavelet Empirical Orthogonal Function Analysis (WEOF).

Empirical Orthogonal Functions analysis (EOF) has been widely used (e.g. Kutzbach, 1967) for analyzing spatial and temporal variability of physical fields to objectively identify the spatially uncorrelated modes of variability of a given field. In this study EOF is used on individual scale power or SAWP. Since the WLPCs are obtained from SAWP, they are interpreted as 'frequency compacted' energy variability. To identify and delineate temporally and spatially uncorrelated patterns at regional scale, we applied the WEOF analysis on SAWP of the SST of the South Atlantic and Indian Ocean and rainfall time series of EA.

2.4 Wavelet Independent Component Analysis (WICA)

Because EOF technique uses second order statistics information only, the PCs extracted are uncorrelated but sometimes not really independent. This difficulty can be resolved if Independent Component Analysis *(*ICA) also called blind source separation (BSS), (Hyvärinen and Oja, 2000) is used to extract the independent modes from the data. ICA makes the extracted components as independent as possible. To obtain the initial most important independent components, pre-whitening, a process of projecting the original data into a subspace spanned by the first few PCs is employed. Using these projections optimization algorithms are used to estimate the independent components. ICA is considered an extension to PCA and used in this study as a complementary tool to EOF analysis, to allow the underlying structure of the data to be more readily observed.

2.5 Artificial Neural Network driven by Genetic Algorithm (ANN-GA) for rainfall prediction

The Genetic algorithm calibrated neural network used in this study (Fig not shown) consists of a population of feed forward neural networks embedded in a Genetic algorithm routine. The ANN parameters are iteratively improved via genetic evolution (selection, crossover and mutation) to more accurately model the joint SST-rainfall variability. The objective function used is a combination of the Pearson correlation and the root mean square error (RMSE). For each neural network of the population, the predictand, y, is obtained as a nonlinear translation of the weighted average of the PCs of raw data, x

$$y = f_2 \sum w^2 \left(f_1 \sum \left(w^1 x \right) + b^1 \right) + b^2 \quad (8)$$

where x is the standardized anomaly and w^1 *and* w^2 are weights for each solution to the hidden and the output layers respectively, and b^1 *and* b^2 are the bias vectors associated with the hidden and output layers. The transfer function, f_1, maps the PCs to the hidden layer, containing two or more neurons, whose number is dictated by the complexity of the problem. The biases, b^1 and b^2 are added to stabilize the solutions. Next the hidden layer is mapped to the third (output) layer containing a single neuron, through another transfer function f_2. The non-linear mapping function, f_1, used here is the hyperbolic tangent function, also called the squashing function (Haykin, 1994), used to squash or limit the output of a neuron to permissible amplitude.

2.6.1 Canonical Correlation Analysis (CCA) and Principal Component Analysis (PCA).

CCA is an established statistical forecasting scheme. Readers interested in details of CCA can refer to Barnett and Preisendorfer 1987 and Glahn (1968). The size of raw input data from the Indian and South Atlantic Ocean that was identified through WEOF was reduced to a few dominant EOF or PCA modes as input to CCA. In applying CCA model, 37 years of data was used to calibrate at each iteration. For example 1950-1986 AMJ data was used to predict the 1986/87 ONDJFM rainfall, 1951-1987 for predicting the 1987/88 rainfall, etc.

3. RESULTS

3.1 Variability of the rainfall in EA and SST in the Indian and South Atlantic Ocean.

Applying the WEOF and the WICA analyses techniques to individual scale power and SAWP for the SON season, one major WLPC, which explained a variance of over 60%, was retained (see Fig 1). This WLPC also explained the largest variance within the 2-2.4-year cycle. The spatial distribution of WLPC1 is shown as a strong SON signal over much of EA with R^2 of about 0.8 between WLPC1 and the individual 2.5ºx3.75º latitude, longitude grid station SAWP.

The MAM season found to have a weak internal signal (WLPC1 of SAWP, 26% using WEOF) was best diagnosed using WICA, where WLPC1 accounted for over 30% for both SAWP and most of the scales in the 4-8 year range. The MAM season was found to be out of phase between the east Tanzania and southern Uganda with the rest of the EA (see WLPC1 of MAM in Fig 1).

The temporal evolution of the SON and MAM variabilities shown in Fig 2 are reminiscent of warming and cooling of the Indian and Atlantic oceans (Figs not shown). The decrease and increase of energy in the figures are also consistent with below and above normal rainfall, respectively, in the last 20 years.

The SON season was found to be strongly linked to the south Indian Ocean but was also found to immediately respond to SST variabilities in both the north and southern portions of the ocean. The MAM was found to be strongly teleconnected to variabilities in the Atlantic Ocean.

(1) SON season

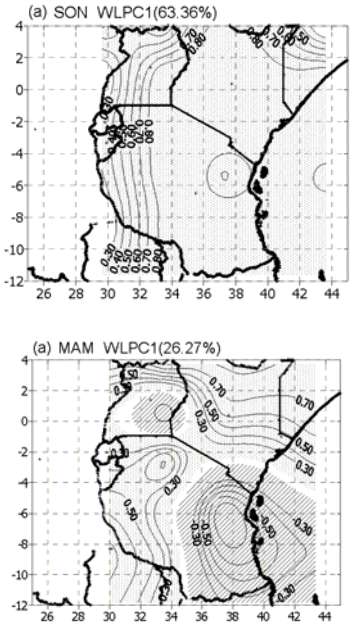

Fig 1. The WLPC1 of SON and MAM respectively.

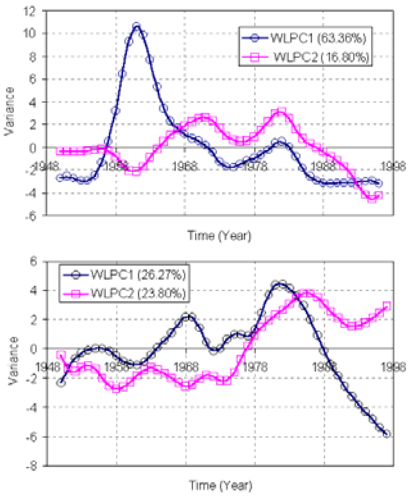

Fig 2. The temporal variabilities of the WLPCs of the SON and MAM WLPCs.

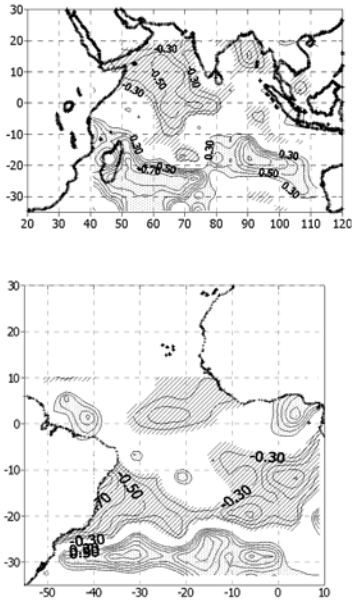

Fig. 3 The teleconnection between SON and Indian oceans and the MAM and South Atlantic Ocean.

3.2 Predicted EA Rainfall

The strong annual and seasonal relationships found above between the SST SAWP of the south Atlantic and Indian Oceans and the EA rainfall WLPCs shows that the predictability of rainfall in EA based on the SST of the two oceans is possible. Two predictor data sets were formed. The first data set (Data1) was extracted from the SW and the NW Indian Ocean and the second data set (Data2) was extracted from the Brazil and Guinea ocean currents in the South Atlantic Ocean. SST data from these fields was extracted for the months of April, May and June. The data was averaged over the three months for each grid station in the two oceans to give one AMJ data set. To evaluate the prediction skill, the Pearson correlation and RMSE are used.

The predicted SON standardized rainfall and the corresponding skills (shown for ANN-GA only) clearly indicate that the SON rainfall variability responds better to the Indian than the Atlantic Ocean SST variability and the MAM rainfall variability responds to the Atlantic Ocean SST variability better than the Indian Ocean SST variability. The skill also underscores the spatial distribution of the strength of the signals as found by WOEF and WICA, showing that only a single homogenous rainfall zone exist during SON and two zones exist during the MAM season.

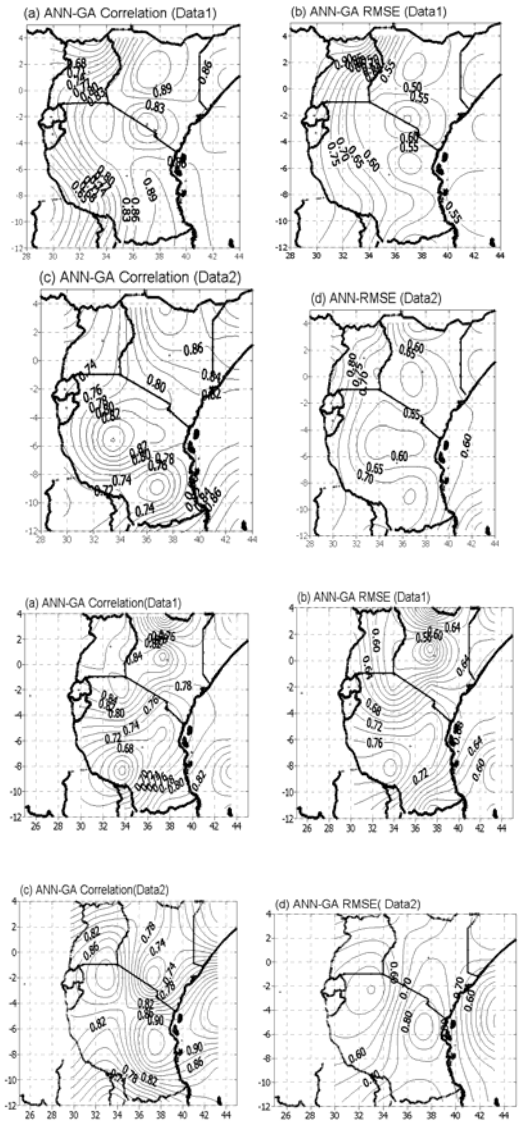

Fig. 4 Prediction statistics of the SON and MAM seasons.

4. OBSERVATIONS AND CONCLUSIONS

We used wavelets, wavelet empirical orthogonal function analysis and wavelet independent component analysis of individual scale power and scale averaged wavelet power to analyze the spatial, temporal and frequency variability and dominant oscillations of EA rainfall and SST of the South Atlantic Ocean and Indian Ocean and explore teleconnection patterns

between SST fields from the Indian and south Atlantic Oceans and rainfall in EA to identify relevant predictor fields to statistical prediction models. The prediction skill showed that non-linear statistical teleconnection models exploited the non-stationary and non-linear characteristics of climate data by using ingeniously identified data sets found between rainfall and SST via wavelet analysis.

References:

Barnett, T. P., and R. Preisendorfer, 1987: Origins and Levels of Monthly and seasonal forecast Skill for United States Surface Air Temperatures Determined by Canonical Correlation Analysis. *Mon. Wea. Rev*., **115**, 1825 – 1851.

Beltrando, G., 1990: "Space-time variability of rainfall in April and October-November over East Africa during the period 1932-1983. *J. Climatol*., **10**, 691-702.

Castleman KR. 1996. *Digital Image Processing. Prentice hall*, Englewood cliffs, New Jersey.

Glahn HR. 1968.Canonical correlation and its relationship to discriminant analysis and multiple regression. Journal of atmospheric sciences **25**: 23-31

Haykin S. 1994. Neural Networks: *A Comprehensive Foundation*, Macmillan/IEEE Press.

Hyvärinen, A., and E. Oja, 2000: Independent component analysis: algorithms and Applications. *Neural Networks* **13**, 411-430.

Huang NE, Shen Z., Zheng Q., Yen N., Tung, C.C., and Liu, H.H., 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London*. **454:** 903-995.

Kutzbach JE. 1967: Empirical eigenvectors of sea-level pressure, surface temperature and precipitation complexes over North America. *Journal of applied meteorology* **6:** 791-802.

Mutai CC, Ward MN, Colman A, 1998. Towards the prediction of the East Africa short rains based on sea surface temperature-atmosphere coupling. *International Journal of Climatology* **18:** 975-997

Terrence C, Compo GP. 1998 A practical guide to wavelets analysis. *Bulletin of the American Meteorological Society (*79) **1**: 61-78

Ntale, H.K., T. Y., Gan and D. Mwale, 2003: Prediction of East African Seasonal Rainfall Using Simplex Canonical Correlation Analysis. *J. Clim.,* **16** 2105–2112.

Philippon, N., P. Camberlin, and N. Fauchereau 2002: Empirical predictability study of October-December East African rainfall. *Q.J.R. Meteorol*. Soc **128** 2239-2256.

Potts, A., 1971: Application of harmonic analysis to the study of east African rainfall data. *J. Trop. Geog*., **34**, 31-42.

Rodhe, H., and H. Virji 1979: " 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.