8-1-2014

Precipitation Forecasting With Wavelet-Based Empirical Orthogonal Function And Artificial Neural Network (WEOF-ANN) Model

Sanaz Imen

Ni-Bin Chang

Y. Jeffrey Yang

11th International Conference on Hydroinformatics
HIC 2014, New York City, USA

Short-term Precipitation forecasting with Wavelet-based Empirical Orthogonal Function and Artificial Neural Network (WEOF-ANN) Model

SANAZ IMEN (1), NI-BIN CHANG (2), Y.JEFFREY YANG (3)
(1), (2): Department of civil, environmental, and construction engineering, University of Central Florida, Orlando, FL, USA
(3): Water Supply and Water Resources Division, Water Quality Management Branch, ORD/NRMRL, U.S. EPA, Cincinnati, OH, USA

Since 2000, western drought caused sharp drop by about 100 feet in the largest reservoir of North America, Lake Mead. About 97% of inflow into Lake Mead is supplied by Colorado River Basin which is extremely sensitive to changes in precipitation and temperature. Oceans play an important role on earth's climate via oceanic-atmospheric interactions known as climate teleconnections, which deeply affect the terrestrial precipitation patterns. This issue signifies the necessity of developing a modern hydroinformatics tool - precipitation forecasting model - to account for teleconnection signals from climate change and mitigate drought hazards impact on lake water, quantitatively and qualitatively, which cannot be achieved by using traditional Global Circulation Model. Therefore, understanding the relationship between precipitation and teleconnection patterns could be the first step for precipitation forecasting. However, highly non-linear and non-stationary nature of teleconnection patterns result in large uncertainties in estimates, since simple linear analyses failed to capture underlying trends at sub-continental scales. For this purpose, high-resolution remote sensing imagery, spectral analysis techniques, and wavelet analysis were integrated to explore the nonstationary and nonlinear behavior of teleconnection signals between the Pacific and Atlantic sea surface temperature (SST) on a short-term basis (10 years) from which the precipitation pattern shift in the Upper Colorado River Basin can be elucidated. These processes lead to the creation of correlation maps which specify index regions within the Atlantic and Pacific Oceans where SST anomaly can be statistically significant in correlation with terrestrial precipitation. These indexed regions delivering some kind of memory effects of SST were extracted to be inputs into an Artificial Neural Network (ANN). Advances in Wavelet-based Empirical Orthogonal Function and Artificial Neural Network (WEOF-ANN) model for rainfall prediction assists the local water management agencies to mitigate the drought impacts and obtain sustainable development strategies a month ahead of the time in urban drinking water infrastructure assessment plan around Lake Mead area.

Keywords: Climate Teleconnection, Wavelet Analysis, Remote Sensing, Artificial Neural Network, Lake Mead.

INTRODUCTION

One of the leading sources of uncertainties in forecasting streamflow is poor understanding of precipitation in future. This issue is of particular concern in Colorado River Basin which provides 96.9% of inflow into Lake Mead. Streamflow of this basin is significantly sensitive to precipitation and temperature changes, Wolock and McCabe [1]. For instance, 70% of annual runoff in Colorado River Basin is provided by Rocky Mountains' high elevation snow pack and spring melt, Christensen et al. [2].

The characteristics of climate processes (e.g. precipitation) change over time as a result of external forcing and internal feedbacks. One of the external forcing which plays a key role in controlling the weather and climate change is oceans [3]. Knowing the effects of sea surface anomalies on earth's weather patterns via oceanic-atmospheric interactions, which is known as climate teleconnection, provides us with knowledge of terrestrial responses to changing climate patterns.

The most prominent teleconnection patterns in the North Pacific and Atlantic Oceans are North Atlantic Oscillation (NAO), El Nino Southern Oscillation (ENSO), and Pacific Decadal Oscillation (PDO). Earlier study showed the occurrence of wetter than normal cold season precipitation during ENSO warm phase (El Nino), and drier than normal during ENSO cold phase (La Nina) in southwestern United States, Hidalgo and Dracup [4]. In addition, the correlation between precipitation and PDO/ENSO across the southwestern United States was presented by Gutzler et al. [5].

Relationship between teleconnection patterns and precipitation is highly nonlinear which leads to difficulties in conducting precipitation forecasting. Hence, intelligent methods (e.g. genetic algorithm (GA), neuro-fuzzy systems, and artificial neural network (ANN)) have been developed to bridge the gap of traditional numerical and statistical methods which are based on finding linear correlation between variables. In 1997, Tsintikidis [6] showed the advantages of ANN to linear regression models in estimating precipitation. Later, along with ground-truth data, different types of data, such as infrared satellite imagery and combinations of TRMM radar and GEOS multispectral satellite data were utilized as inputs into ANN model (Hsu et al . [7]; Bellerby et al. [8]). In 20 century, scholars focused on developing ANN model. For instance, Grimes et al. [9] applied network-prunning technique analysis to identify redundant input data. In addition, principle component analysis was applied to reduce dimensionality of all input data (Grimes et al. [9]; Behrangi et al. [10]). Wang and Ding [11] developed a wavelet-based neural network model which encompasses abilities of both ANN and wavelet transform. Karamouz et al. [12] predicted precipitation using standard multi-layer perception (MLP) trained with back propagation algorithm.

The main objective of this study is short term precipitation forecasting at a local scale using wavelet empirical orthogonal function and artificial neural network to obtain sustainable development strategies a month ahead of the time. The study site for precipitation forecasting is upper Colorado River Basin. In addition, Weminuche wilderness area was selected as a pristine terrestrial site to find the significant oceanic indices on precipitation.

Methodology

The analytical framework of this study is shown in Figure 1. This framework includes the following five primary components: (1) site selection, (2) data collection and processing, (3) wavelet analysis of variability, (4) mapping oceanic indexed regions, and (5) WNN modeling.

1. Site Selection

This study focus on identification of teleconnection patterns at a local scale. In order to minimize natural and anthropogenic effects "noise" on the climate teleconnection signals, a pristine site was selected. From Figure 1, it can be seen that the screening factors reject areas with anthropogenic effects, forest fires, and disease within at least the past 10 years. In addition, the site should be >190,000 acres (76,890 ha) to include at least one of precipitation grids. Based on these criteria, Weminuche Wilderness area in Colorado in U.S. was selected.

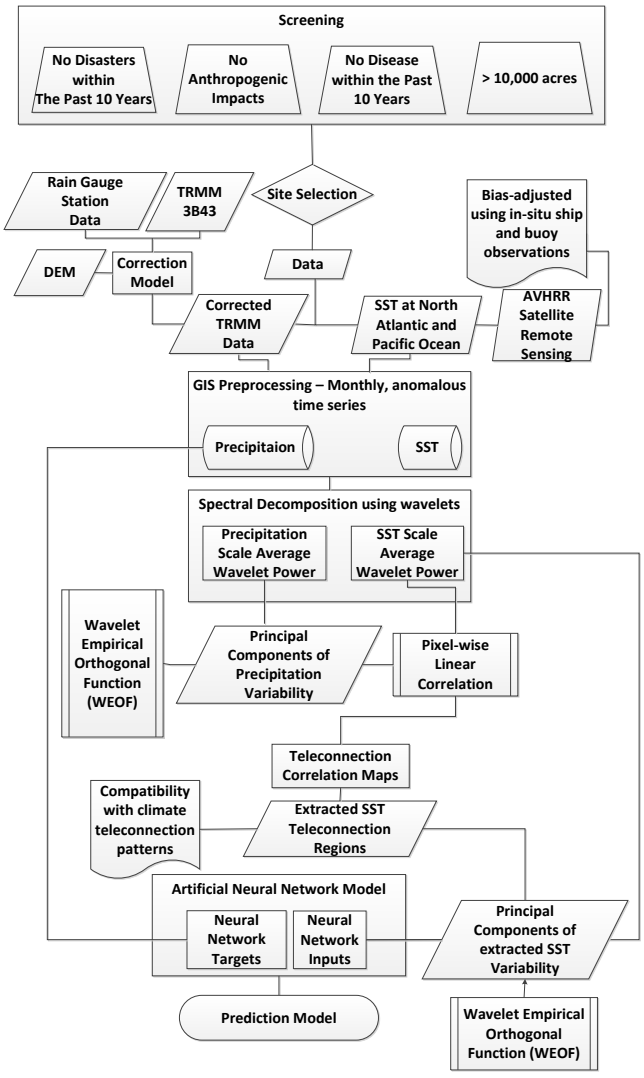

Figure 1. Analytical Framework of Study

In 1975, the Weminuche Wilderness, the largest wilderness in Colorado, was designated and expanded to 488,340 acres by the Colorado Wilderness Acts of 1980 and 1993. It has elevation between 8000-14,000 feet, and located in southwest Colorado (Figure 3). There have not been any reported forest fires during the study period (2000-2010), and the only reported one is West Fork Fire which happened in 2013. On average, precipitation ranges from 0.65 to 2.62 inch/month, and temperature varies between -2^{o}F to 77^{o}F.

2. Data collection and processing

The precipitation dataset applied in this study is the Tropical Rainfall Measuring Mission (TRMM3B43). TRMM3B43 products provide us with monthly precipitation data with a resolution of $0.25^0 \times 0.25^0$. To cope with underestimation/overestimation issues of TRMM dataset especially in mountainous regions, the satellite-gauge-biased-adjusted method developed by Ji and Chen [7] was applied in this study.

The SST dataset was obtained from the Group of High Resolution Sea Surface Temperature (GHRSST) global Level 4 SST analysis. This product provides us with a daily 0.25 degree grid. In this study the spatial resolution of SST dataset was resampled to $0.5^o \times 0.5^o$ to facilitate computation. Period of all applied datasets were from Jan 2000 to Dec 2010 and they were converted to monthly anomaly values.

3. Wavelet analysis of variability

Spectral decomposition using wavelet has some advantages compare to traditional Fourier transform, as it decomposes a time series into time and frequency domain, Romacy-Pereira [15]. This ability provides us with information about the dominant localized variation of power, Keener et al. [17]. After assessing the wavelet power spectra for each SST and precipitation grids, the time period band in which both SST and precipitation grids have dominant wavelet power was selected. To extract the wavelet power over the identified time period band, the scale average wavelet power (SAWP) is calculated based on Eq. 1, Torrence and Compo [17]. Where, C_δ is 0.776 for the Morlet wavelet (the applied wavelet in this study), δ_j is a factor for scale averaging, δ_t is the sampling period, j_1,…..,j_2 indicates the scale over which the SAWP is calculated.

$$W^2(b) = \frac{\delta_j \delta_t}{C_\delta} \sum_{j=j_1}^{j_2} \frac{|W(b,a_j)|^2}{a_j} \quad (1)$$

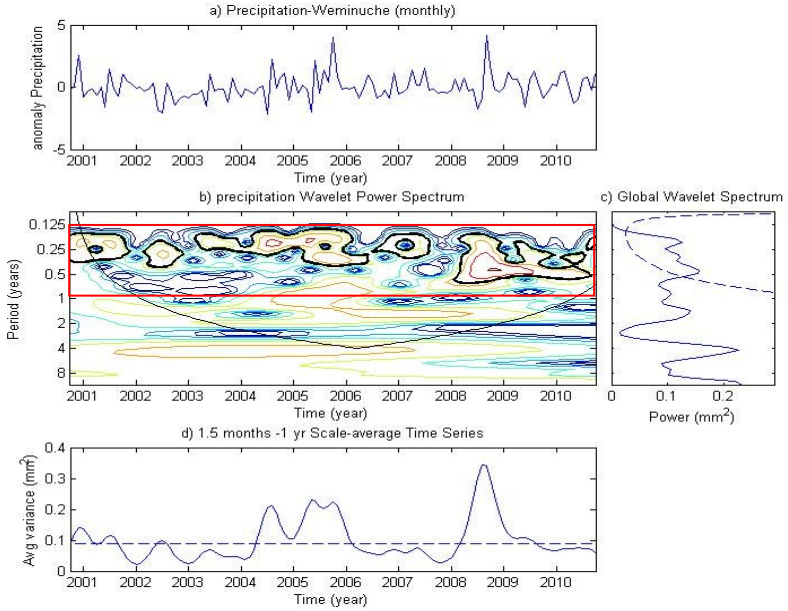

Figure 2. Wavelet decomposition example of a single precipitation grid; (a) anomaly time series of one of precipitation grids calculated by subtracting the pixel-based monthly median values for each month; (b) precipitation wavelet power spectrum; (c) global wavelet spectrum. The dashed red lines show the boundary of cone of influence; (d) the blue solid line represents the SAWP, which was calculated based on 0.125 month- 1 year period band.

An example of local and global wavelet spectra constructed for a precipitation grid is shown in Figure 2. The solid black line drawn through the wavelet spectrum indicates the cone of influence (COI) in figure 2-b. This area is shown with dashed lines in Figure 2-c and 2-d. The solid black lines inside the COI in Figure 2-b shows the area in which the power is significant at 95% confidence level. As it is shown in Figure 2-b, the wavelet power is concentrated from 1.5 month to 1 year (red box). Several dominant frequencies within this band impact both SST and precipitation variability. To identify the significant oceanic indexed regions, correlation between isolate dominant modes of Atlantic and Pacific SST variability and dominant modes of precipitation were calculated. Since there are four precipitation grids within Weminuche Wilderness area (Figure 3), the empirical orthogonal function (EOF) were applied to extract the common spatial and temporal variability of SAWPs and consolidate them to a single time series which is named wavelet empirical orthogonal function (WEOF), Mwale and Gan [18]. The first WEOF strongly reflects a high percentage of energy variability. For instance, the first WEOF of precipitation in this study explained 96% of the total SAWP variance.

1. Mapping oceanic indexed regions

To prepare correlation maps, correlation between precipitation WEOF1 and each SST SAWP grid were calculated. Based on this procedure, five correlation maps for different time lags (i.e. 1,3,6,12, 24) were prepared. According to these maps, the significant oceanic indexed regions were extracted (Figure 3).

Based on Figure 3, in the Atlantic region, the extracted oceanic indexed regions are consistent with the known teleconnection patterns such as: North Atlantic Oscillation (Index J), Atlantic Multi-Decadal Oscillation (Index K), and Arctic Oscillation (Index I and H). In the Pacific Ocean, other known teleconnection regions overlapped the identified oceanic indexed regions includes ENSO (Index G and D), PDO (Index E, F, C), and PNA (Index A, B). In addition, at the eastern of Japan, another significant oceanic region (Index L) exists which is not associated with known leading teleconnection patterns. This region justifies the necessity of including oceanic regions which are not considered as traditional leading teleconnection patterns. To apply these identified oceanic regions as inputs into the WNN model, the first WEOF of each extracted oceanic regions were calculated.

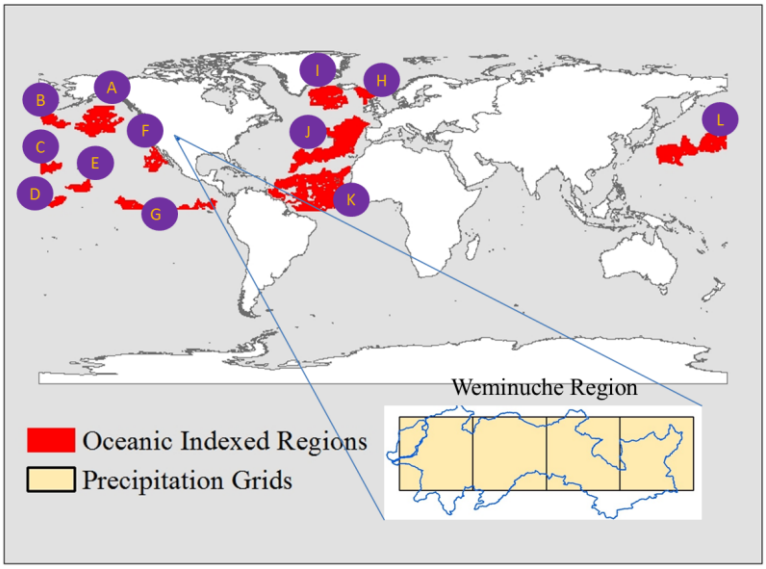

Figure 3. Identified oceanic indexed regions

2. Wavelet based artificial neural network

MATLAB ANN toolbox applied for this study includes two layers feed- forward network trained with the Levenberg-Marquardt back propagation algorithm (Figure 4). This toolbox was applied to find regression between precipitation and identified oceanic indexed regions. The input variables into ANN model are the raw precipitation time series with twelve WEOF's of identified indexed regions. Four different scenarios discussed in Table 1 were defined as inputs into ANN model. 70% of input data were used to train model, 15% were used to validate model, and 15% were applied as an unseen data set for testing the model. The number of hidden neurons in this study is 35.

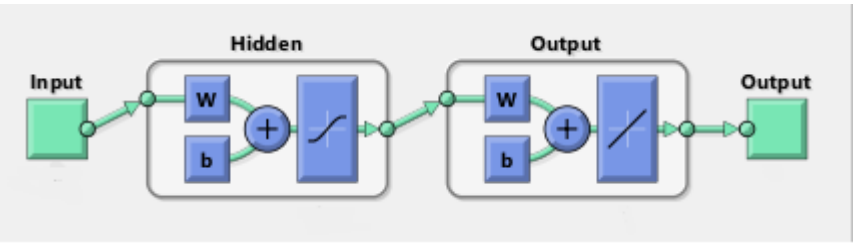

Figure 4. Network structure of ANN model

The root mean square error (RMSE) and the square of the Pearson correlation coefficient (coefficient of determination - R^2) of four different scenarios were shown in Table 1. R^2 and RMSE value closer to 1 and 0 indicates the least amount of difference between observed and predicted values, respectively. Therefore, the best scenario will be scenario 2 in which all different indices were included. In addition, comparing scenario 1 with scenarios 3 and 4 shows the effects of considering indices on reliability and predictability of the model. This result is in consistent with the results of earlier studies which showed the prominent effects of ENSO and PDO on precipitation in southwestern U.S., Hidalgo and Dracup [4]; Gutzler et al. [5].

Table 1. Neural network forecast results. RMSEs and square of Pearson correlation coefficients are used to compare the modeled monthly precipitation with one month lead forecast

Scenarios	R^2 (%)	RMSE (mm month^{-1})
Scenario 1. precipitation	0.40	0.36
Scenario 2. precipitation and all indices	0.95	0.10
Scenario 3. precipitation and indices A, B, C, D, E, F, J	0.82	0.19
Scenario 4. precipitation and indices D and E	0.84	0.18

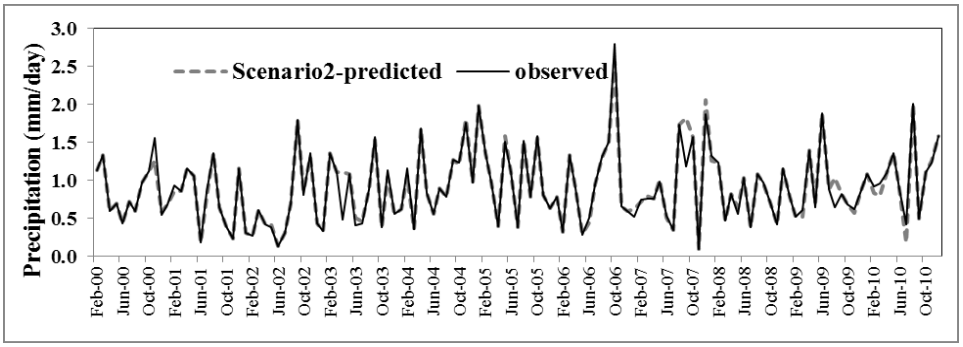

Figure 5. Comparing the results of modeled and observed precipitation in scenario 2 from Feb 2000 to Dec 2010.

Time series of modeled precipitation for the best scenario was compared to average monthly observed precipitation from Feb 2000 to Dec 2010 in Figure 5 to show the reliability of the developed model in modeling precipitation trend and peaks in this region.

Results and Discussion

The integrated wavelet-based empirical orthogonal function and artificial neural network (WEOF-ANN) model was applied in this study to predict precipitation a month ahead of the time over upper Colorado River basin which provides 97% of inflow into Lake Mead. As a result of non-stationary and non-linear relationships between climate teleconnections and terrestrial precipitation, the integration of high resolution imagery and wavelet analysis were applied to decompose time series into time-frequency domain to find the dominant localized variation of power, which is concentrated from 1.5 month to 1 year in this study. Correlation between the isolated dominant modes of Atlantic and Pacific SST variability and dominant modes of precipitation were calculated to identify the significant oceanic indexed regions. This correlation was calculated for different time lags (i.e. 1,3, 6, 12, 24). The identified oceanic indexed regions were consistent with NAO, AMDO, ENSO (Nino 3 and Nino 3.4), PDO, and PNA. In addition, one of the identified oceanic indexed regions is not considered as known leading teleconnection patterns. The first WEOF of the identified oceanic indexed region were extracted to use as inputs into ANN model. In addition, the historical precipitation data were applied as an input into this model. The results of this model were assessed for 4 different scenarios based on RMSE and R^2 values. The best result is related to scenario 2 which includes all indices. In addition, the result of scenario 4 shows the significant effects of Pacific Ocean on precipitation in southwestern United States. Moreover, comparing the results of scenario 1 with other scenarios shows the effect of including oceanic indexed regions on increasing the reliability and predictability of the developed model.

Acknowledgment

This study is funded through EPA contract WA 1-52. Support from the USEPA is gratefully acknowledged

REFERENCES

[1] Wolock D.M., McCabe G..J., "Estimates of Runoff Using Water-Balance and Atmospheric General Circulation Models", *J.Amer. Water Resour. Assoc.*, Vol. 35, (1999), pp 1341-1350.

[2] Christensen N.S., Wood A.W., Voisin N., Lettenmaier D.P., and Palmer R.N., "The Effects of Climate Change on the Hydrology and Water Resources of the Colorado River Basin", *Climate Change,* Vol. 62, No. (1-3), (2004), pp 337-363.

[3] NOAA Ocean Service Education,"The Ocean's Role in Weather and Climate". http://oceanservice.noaa.gov/education/pd/oceans_weather_climate/. Revised July 18, 2012

[4] Hidalgo H.G., Dracup J.A., "ENSO and PDO Effects on Hydroclimatic Variations of the Upper Colorado River Basin", *Journal of Hydrometeorology*, Vol. 4, (2003), pp 5-23.

[5] Gatzler D.S., Kann D.M., Thomburgh C., "Modulation of ENSO-Based Long-Lead Outlooks of Southwestern U.S. Winter Precipitation by the Pacific Decadal Oscillation Weather and Forecastin", *Wea. Forecasting* , Vol. 17, No. 6, (2002), pp. 1163-1173.

[6] Tsintikidis D., Haferman J.L., Anagnostou N., Krajewski W.F., Smith T.F., "A neural network approach to estimating rainfall from spaceborne microwave data", IEEE Trans. Geosci. Remote Sens., Vol.35, (1997), pp. 1079-1092.

[7] Hsu K.L., Gao X., Sorooshian S., Gupta H., "Precipitation estimation from remotely sensed information using artificial neural networks", Journal of Applied Meteorology, Vol.36, (1997), pp.1176-1190.

[8] Bellerby, T., Todd M., Kniveton D., Kidd C., "Rainfall estimation from a combination of TRMM precipitation radar and GOES multispectral satellite imagery through the use of an artificial neural network", Journal of Applied Meteorology" Vol. 39, (2000), pp. 2115-2128.

[9] Grimes D.I.F., Coppola E., Verdecchia M., Visconti G., " A neural network approach to real-time rainfall estimation for Africa using satellite data", Journal of Hydrometeorology, Vol. 4, (2003), pp. 1119-1133.

[10] Behrangi A., Hsu K.L., Imam B., Sorooshian S., "PERSIAN-MSA: a precipitation estimation method from satellite–base multispectral analysis", Journal of Hydromteorology, Vol. 10, (2009), pp. 1414-1429.

[11] Wang W., Ding J., "Wavelet Network Model and Its Application to the Prediction of Hydrology", *Nature and Science*, Vol. 1, No.1, (2003), pp. 67-71.

[12] Karamouz M., Razavi S., Araghinejad S., "Long-lead seasonal rainfall forecasting using time-delay recurrent neural networks: a case study", Hydrological Processes, Vol.22, (2008), pp. 224-241.

[15] Romacy-Pereira RN, de Araujo DB, Leite JP, Garcia-Cairasco N, "A semi-automated algorithm for studying neuronal oscillatory patterns: a wavelet-based time frequency and coherence analysis", *J Neurosci Methods*,Vol. 167, (2008), pp.384–392

[16] Keener VW, Feyereisen GW, Lall U, Jones JW, Bosch DD, Lowrance R., "El-Niño/Southern Oscillation (ENSO) influences on monthly NO3 load and concentration, stream flow and precipitation in the Little River Watershed, Tifton, Georgia (GA)", *Journal of Hydrology*, Vol. 381, No. 3-4, (2010), pp. 352-363

[17] Torrence, C., and Compo G.P., "A practical guide to wavelet analysis", *Bulletin of the American Meteorological Society*, Vol. 79, No.1, (1998), pp. 61-78.

[18] Mwale, Davison, Thian Yew Gan., "Wavelet Analysis of Variability, Teleconnectivity, and Predictability of the September–November East African Rainfall", *J. Appl. Meteor.*, vol. **44**, (2005), pp. 256–269.