4.3 A NEW ANALYSIS ON VARIABILITY AND PREDICTABILITY OF SEASONAL RAINFALL OF CENTRAL SOUTHERN AFRICA

Davison Mwale[a], Thian Yew Gan[a]* and Samuel S.P. 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

With a dynamic relationship between rainfall in southern Africa and Sea Surface Temperature (SST) in the surrounding oceans in recent years (Landman and Mason, 1999) and a consistent pattern of declining runoff after 1970 (Fanta et al, 2002), one of the major challenges for meteorologists and hydrologists has been to understand and to predict the nature of this variability over different spatial and temporal scales. Recent research (e.g., Reason and Mulenga, 1999) has made progress in providing evidence of relationships between rainfall variability and a number of atmospheric and oceanic variables for southern Africa. From the developments in long range forecast models utilizing SST and other predictors one would have hoped that with these developments countries within the southern African region are better equipped to manage climate variability rather than always being surprised victims of unexpected extreme events, such as droughts (e.g., Landman et al., 2001). Unfortunately, the ability to predict rainfall has not progressed much in the last 20 years and in recent years has even deteriorated (Mason, 1997). The drop in statistical and dynamical seasonal forecasting skill since 1980 has also been reported for the Indian Monsoon and in the 1990's for the North Atlantic storm frequencies (Mason, 1997). Decreased predictability has been associated with non-stationarity of the ocean-atmosphere system (Webster et al., 2002), strong non-linearities demonstrated by important climate processes (Landman et al., 2001) and general problems associated with model inadequacies (Allan et al., 1995). To avoid ambiguous forecast signals (Jury and Engert, 1999) or compromised skill in general (Shen et al., 2001), new methods of analysis, non-linear prediction models and identification of robust predictor fields are emphasized (Shen et al., 2001, Mason, 1997).

This study presents results on the use of wavelet analysis and wavelet-based empirical orthogonal analysis to:

(1) Identify and analyze the spatial, temporal and frequency variability, and the dominant oscillations in the rainfall fields of central southern Africa (CSA) and SST fields in the South Atlantic Ocean and the Indian Ocean.
(2) Explore teleconnection patterns between SST fields from the Indian and south Atlantic Oceans and rainfall in CSA to identify regions of predictor fields from the oceans relevant to statistical prediction of CSA rainfall and,
(3) The predictor fields identified from (1) and (2) are used to drive both a linear model, Canonical Correlation Analysis (CCA) and a non-linear model artificial neural network calibrated by a genetic algorithm (ANN-GA) to predict rainfall at seasonal time scales (3-month lead time) for CSA during the period 1986-1995.

A wavelet transformation is used in this study because they have been successfully applied to climate characteristics analysis, such as streamflow characterization (Smith et al., 1998) and interannual temperature events and shifts in global temperature (Park and Mann, 2000). Results from these and other studies have shown that wavelets are capable of locating irregularly distributed, multi-scale features in space and/or time (Smith et al., 1998) of climate elements and are thus suitable for the analysis of SST and rainfall.

2. DATA AND METHODS.

Monthly rainfall data (1950-1995) from 21 grid locations at a resolution of 2.5°×3.75° latitude and longitude was extracted for CSA (6ºS-18ºS, 12ºE-37ºE), 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, known to influence CSA rainfall. 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.

*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

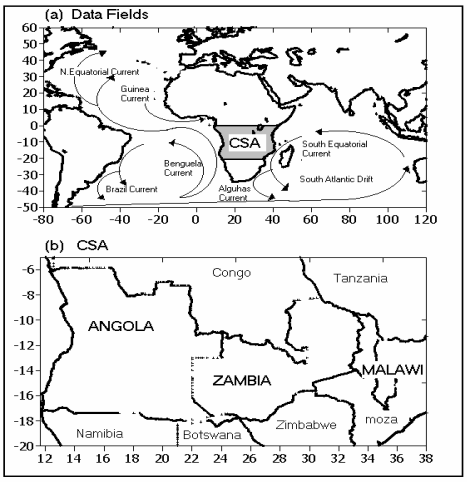


Fig. 1. (a) Major ocean currents of Atlantic and Indian oceans and (b) detailed location and description of CSA (Zambia, Malawi and Angola)

2.2 Wavelet Analysis

A brief outline of wavelet transformation is given here. More detailed information is given by Terrence and Compo (1998). 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 (i.e. 2 years), j =0,1,….20, and δj = 0.25, thus giving scales(periods) ranging from 2 to 64 years, that cover the length of both data periods. The wavelet transform of a time series contains a wealth of information, which needs to be condensed over a range of scales or time to be useful for multivariate analysis. Two ways suggested by Torrence and Compo (1998) are (1) Time integrated variance of energy coefficients at every scale to construct global wavelet power

$$\overline{W}_t^{\,2}(a) = \frac{1}{T}\sum_{t=0}^{T-1}|W_t(a)|^2 \qquad (6)$$

and (2) Scale (band limited) integrated variance of energy coefficients over time to construct the scale averaged wavelet power (SAWP)

$$W_t^2 = \frac{\delta_j\delta_t}{C_\delta}\sum_{j=j_1}^{j_2}\frac{|W_t(a_j)|^2}{a_j} \qquad (7)$$

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 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 CSA, we extracted energy in the 2-8 year range containing El Nino Southern Oscillation (ENSO), known to influence rainfall variability in CSA (Mason and Tyson, 2000). 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.

$$\Psi(t) = \pi^{-1/4} e^{i6t} e^{-t2/2} \qquad (8)$$

2.3 Wavelet Empirical Orthogonal Function Analysis.

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 SAWP and hence called wavelet empirical orthogonal function (WEOF) and its corresponding principal components as wavelet principal components (WLPCs). Since the WLPCs are obtained from SAWP, they are interpreted as 'frequency compacted' energy variability (Park and Mann, 2000). The number of WEOF retained for analysis is found using the scree diagram. 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 CSA. Techniques such as cluster analysis and among many others can be used to meet this objective. However, the skill of the above methods is often difficult to evaluate statistically (Basalirwa, 1995).

2.4 Models for Predicting Rainfall.

2.4.1 Artificial Neural Network Calibrated by Genetic Algorithm (ANN-GA)

The Genetic algorithm calibrated neural network used in this study consists of a population of feed forward neural networks embedded in a Genetic algorithm routine, Fig 2. 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

$$y = f_2 \sum w^2 \left(f_1 \sum \left(w^1 x \right) + b^1 \right) + b^2 \qquad (9)$$

where $x = \dfrac{X - \overline{X}}{\sigma_X}$ (10)

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.

The neural networks are trained and the best weights are retained for predicting the CSA rainfall using input data independent of the training.

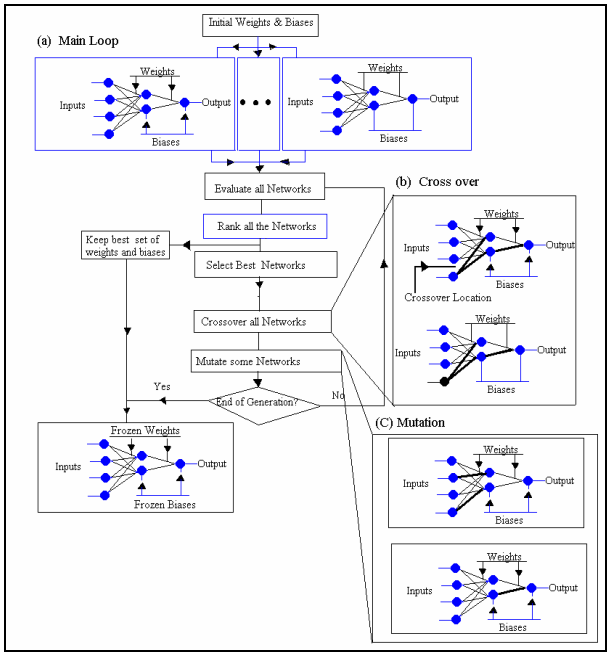


Fig 2. The ANN-GA flowchart.

2.4.2 Canonical Correlation Analysis (CCA) Principal Component Analysis (PCA).

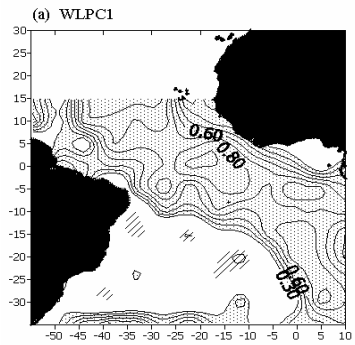
CCA is an established statistical forecasting scheme. Readers interested in details of CCA can refer to Barnett and Preisendorfer 1987 and Glahn (1968). The size of raw input data from the Indian and South Atlantic Ocean that was identified through WEOF was reduced to a few dominant EOF or PCA modes 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.

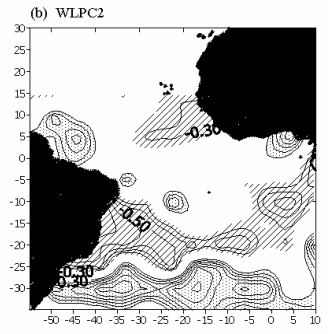
3. RESULTS

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

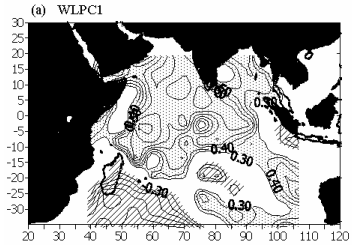
Independently, the first three WLPCs of SST in both oceans and CSA were retained for analysis (Fig 3). These explained 66.9%, 63.8% and 68.5% of the total energy variance for the Atlantic, the Indian Oceans and CSA respectively. The major ocean currents with climatological importance to CSA rainfall variability were the Benguela, Guinea and the Brazil in the South Atlantic and the Agulhas and south equatorial in the Indian Ocean. The northern half of Indian ocean was also observed as an important source of rainfall variability of CSA. Increasing power of wavelet energy of the Benguela current and northern half of Indian ocean coupled with a decrease in power of the Agulhas and south equatorial ocean currents were found to result in decrease (increase) of rainfall in eastern (western) CSA. The long-term wavelet power increases and decreases of the oceans and corresponding decrease (increases) in eastern (western) CSA rainfall was found to have begun about 1970 and went on till 1995.

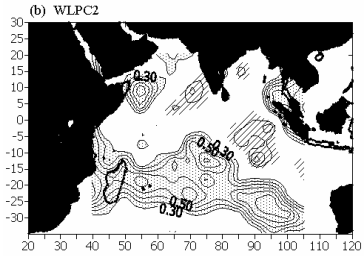
(1) South Atlantic Ocean



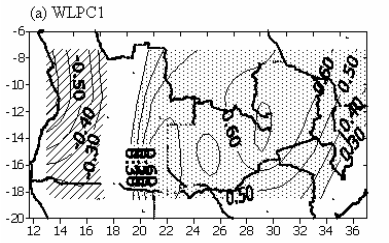


(2) Indian Ocean





(3) CSA



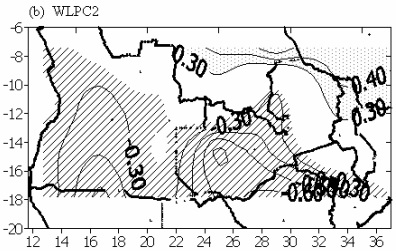


Fig 3. The first two WLPCs: (1)(a) and (b) South Atlantic Ocean (2) (a) and (b) Indian Ocean and (3) (a) and (b) for CSA. Dotted (cross-hatched) areas are positively (negatively) correlated to the WLPCs

3.2 Predictability of CSA Rainfall

Validation runs for both the calibrated models of ANN-GA and CCA were based on four PCs with a total variance of 72.6% (e.g., 38.6%, 15.3%, 10.5% and 8.2%) as input from the South Atlantic Ocean and five PCs explaining a total variance of 73.1% (e.g., 34.9%, 17.5%, 9.6%, 7.2% and 3.9%, respectively) from the Indian Ocean. Figure 4 presents the results. Excellent predictions were found for ANN-GA. The CCA had comparatively low prediction skill.

4. OBSERVATIONS AND CONCLUSIONS

We used wavelets and wavelet empirical orthogonal function analysis of scale averaged wavelet power to (1) identify and analyze the spatial, temporal and frequency variability and dominant oscillations in the CSA rainfall and SST of the South Atlantic Ocean and Indian Ocean, (2) Explore teleconnection patterns between SST fields from the Indian and south Atlantic Oceans and rainfall in CSA to identify relevant predictor fields to statistical prediction models and (3) used linear and non-linear statistical teleconnection models, CCA and ANN-GA for predicting seasonal rainfall at 3 months lead-time.

Results show that long term warming (cooling) of the Northern Indian Ocean and Benguela current coupled with cooling (warming) of the south Indian Ocean, the Guinea and Brazil ocean currents resulted in prolonged below (above) normal rainfall in eastern (western) CSA.

Using the identified predictor fields of both the oceans to predict rainfall in CSA, lead to higher prediction skills in both the linear and non-linear statistical teleconnection models, CCA and ANN-GA. However, ANN-GA achieved higher prediction skill that CCA.

Finally, the prediction skills of most statistical teleconnection models are affected by the non-stationary and non-linear characteristics of climate data. We consider the data non-stationariy by identifying regions of predictor fields (SST of Indian and south Atlantic Ocean) correlated to the predictand(CSA rainfall) via wavelet analysis and demonstrated that ANN-GA a nonlinear model achieved higher prediction skill that it linear counterpart , the CCA.

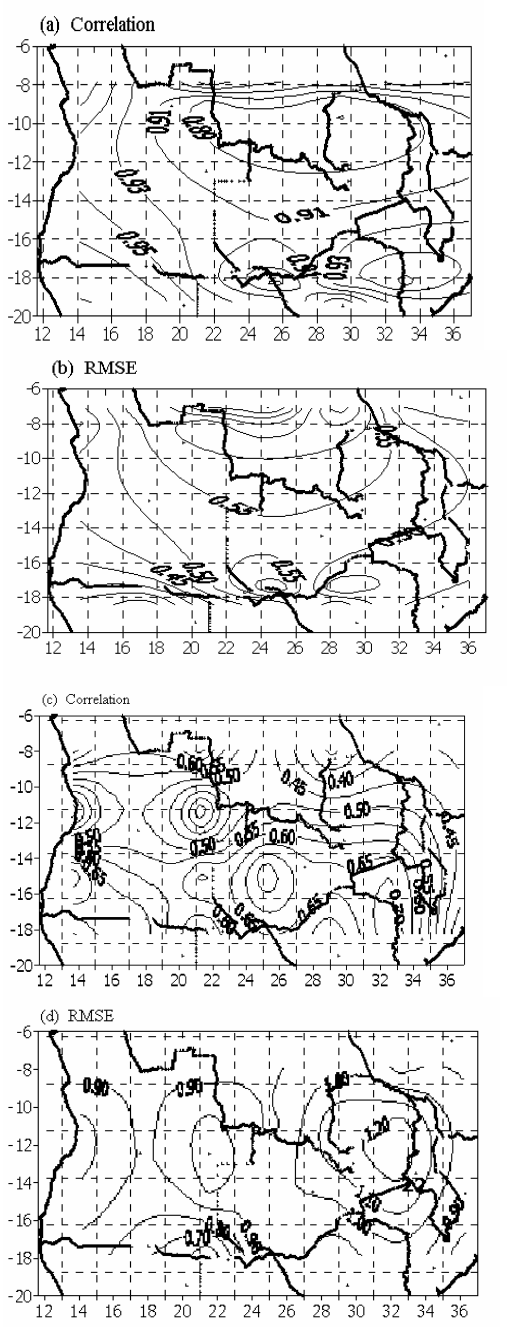


Fig. 4. The prediction skill of Models, (a) and (b) are for ANN-GA and (c) and (d) are for CCA using AMJ SST from the south Atlantic over the period 1986-1995.

References:

Allan RJ, Lindesay JA, Reason CJC. 1995. Multi-decadal variability in the climate system over the Indian Ocean region during the austral summer. *Journal of Climate* 8**:** 1853-1873

Basalirwa CPK. 1995. Delineation of Uganda into climatological rainfall zones using the method of principal component analysis. *International Journal of Climatology* **15:** 1161-1177.

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

Fanta B, Zaake BT, Kachroo RK. 2002. A study of the variability of the river flow of the southern Africa region. *Hydrological Sciences*. 46(**4**): 513-524.

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.

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.

Jury MR, Engert S. 1999. Teleconnections modulating inter-annual climate variability over northern Namibia. *International Journal of Climatology* **19,** 1459-1475.

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

Landman W A, Mason SJ. 1999. Change in the association between Indian Ocean sea-surface temperatures and summer rainfall over South Africa and Namibia. *International Journal of Climatology* 19: 1477-1492

Landman WE, Mason SJ, Tyson PD, Tennant WJ. 2001. Retroactive skill of multi-tiered forecasts of summer rainfall over southern Africa. *International Journal of Climatology* **21:** 1-19

Mason SJ. 1997. Review of recent developments in seasonal forecasting of rainfall. Water SA **23**: 57-61

Mason SJ, Tyson PD. 2000. *The occurrence and predictability of droughts over Southern Africa.* In drought volume 1: A global assessment Wilhite DA(ed). Routledge: New York

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

Park J, Mann ME. 2000. Inter-annual temperature events and shifts in global temperature: A "multiwavelet" correlation approach. *Earth Interactions*. **4**. 1-53

Reason CJC Mulenga H.1999. Relationships between South African rainfall and SST anomalies in the southwest Indian Ocean. *International Journal of Climatology* **19:** 1651-1673.

Shen SPS, Lau WKM, Kim KY, Li G. 2001. *A canonical ensemble correlation prediction model for seasonal precipitation anomaly*, Technical memorandum NASA/TM-2001-209989. Greenbelt, Maryland 20771. Pp53.

Smith LC, Turcotte DL, Isacks BL. 1998. Streamflow characterization and feature detection using a discrete wavelet transform. *Hydrological Processes* **12:** 233-249.

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

Webster PJ, Clark C, Cherikova G, Fasullo J, Han W, Loschnigg and Sahami K. (2002). The monsoon as a self-regulating coupled ocean-atmosphere system. In Meteorology at the Millennium (Edited by Pearce). *The Royal Meteorological Society* **83**: 198-219.

Williams J. 2000. *Drought Risk in Southern Africa*, In drought volume 1: A global assessment Wilhite DA(ed). Routledge: New York