Wavelet Empirical Orthogonal Functions of Space-Time-Frequency Regimes and Predictability of Southern Africa Summer Rainfall

Davison Mwale[1]; Thian Yew Gan[2]; Samuel S. P. Shen[3]; Ting Ting Shu[4]; and Kyu-Myong Kim[5]

Abstract: Wavelet-based empirical orthogonal function (WEOF) analysis was used to analyze the nonstationary spatial, temporal, and frequency regimes of the regional variability in southern African summer (October–March) rainfall. The leading modes of rainfall variability were then used to establish associations with gridded scale-averaged wavelet power of the sea surface temperature (SST) for the Indian and Atlantic Oceans. The WEOF revealed that southern African rainfall is out of phase between areas north and south of 25°S and that areas north of 25°S and northern South Africa experienced decreased rainfall between 1970 and 1997. The decrease in rainfall was modulated by periods of between 2 and 8 years. Using judiciously selected windows of April-May-June SST data for the Atlantic and Indian Oceans as predictors in the artificial neural network-genetic algorithm (ANN-GA), high prediction skill of standardized summer rainfall of southern Africa was achieved. For the validation period 1988–97, Pearson correlation between 0.83 and 0.98 (i.e., 69–96% of observed rainfall variability), Hanssen Kuipers skill scores of between 0.2 and 1.0, and root-mean-square errors of between 0.25 and 0.72 mm of standardized rainfall were found between observed and predicted summer rainfall at a 3-month lead time.

DOI: 10.1061/(ASCE)1084-0699(2007)12:5(513)

CE Database subject headings: Africa; Rainfall frequency; Algorithms; Neural networks.

Introduction

Patterns of spatial, temporal, and frequency regimes of southern African summer rainfall (October–March) and their predictability have been of interest to climatologists in southern Africa (10°S–35°S, 10°E–40°E; see Fig. 1) since 1888 (Tyson et al. 1975; Mason and Tyson 2001). These patterns have been used for long-range summer rainfall forecasting since the 1970s (e.g., Dyer and Tyson 1977; Currie 1993). Since the droughts of 1991–92, 1994, 1995, and 1996 in southern Africa, there has been a resurgence of extensive research to analyze and predict the nature of these patterns and their association with the variability of external forces, such as the sea surface temperature (SST) of the surrounding Indian and Atlantic Oceans and the more remote Pacific Ocean (e.g., Mason 1995; Jury 1996; Reason and Mulenga 1999; Jury and Engert 1999; Richard et al. 2000). The majority of these studies, except for Richard et al. (2000), have three major shortcomings:

1. Data analysis was performed using Fourier-based techniques, thereby ignoring the nonstationary characteristics of climate processes (e.g., Mason 1995);
2. Rainfall only on small geographical areas was analyzed and the regional-scale, summer rainfall variability of southern Africa was neglected (e.g., Jury 1996); and
3. Linear statistical schemes for predicting seasonal rainfall, such as linear regressions or canonical correlation analysis (CCA), were applied thereby ignoring the nonlinear aspects of ocean-atmosphere interaction (e.g., Jury and Engert 1999).

Rainfall time series are nonstationary and their interaction with SST variability is nonlinear (Landman et al. 2001). Adopting stationary and linear assumptions when analyzing rainfall could therefore give misleading results (Huang et al. 1998). For example, interannual oscillations of between 2 and 7 years and interdecadal oscillations of between 18 and 20 years have been identified in the southern Africa rainfall data (Mason 1995; Mason and Tyson 2001). However, even though we are aware of the general spectral characteristics of the southern Africa rainfall data, we do not have information on how the spectral information is distributed over the southern African region and how it has changed over time. Hence, it is not surprising that following the 18–20-year cycles in the rainfall of southern Africa, i.e., the dry 1920s, 1940s, 1960s, and 1980s (Mason and Tyson 2001), and the wet 1950s, 1970s (Dlamini 1995), the 1990s were expected to be wet. However, contrary to the expectation, droughts occurred in 1991–92, 1994, 1995, and 1996 and caused widespread famine in the southern African region (e.g., BBC 2002a,b,c). The abrupt and irregular changes observed in the 1990s, from what appeared to be a regular 18–20-year cycle is a good example of rainfall's nonstationarity. The nonstationary characteristics exhibited by the southern Africa climate processes suggest that analysis of the re-

[1]Dept. of Civil and Environmental Engineering, Univ. of Alberta, Edmonton, AB, T6G 2W2 Canada.

[2]Dept. of Civil and Environmental Engineering, 3-033 Markin/CNRL Natural Resources Engineering Faculty, Univ. of Alberta, Edmonton, AB, T6G 2W2 Canada (corresponding author). E-mail: TGan@ualberta.ca

[3]Dept. of Mathematics and Statistics, San Diego State Univ., San Diego, CA 92182.

[4]Dept. of Mathematical and Statistical Sciences, Univ. of Alberta, Edmonton, AB, T6G 2G7 Canada.

[5]Climate and Radiation Branch, NASA Goddard Space Flight Center, Greenbelt, MD 20771.

Note. Discussion open until February 1, 2008. Separate discussions must be submitted for individual papers. To extend the closing date by one month, a written request must be filed with the ASCE Managing Editor. The manuscript for this paper was submitted for review and possible publication on February 1, 2005; approved on August 31, 2006. This paper is part of the ***Journal of Hydrologic Engineering***, Vol. 12, No. 5, September 1, 2007. ©ASCE, ISSN 1084-0699/2007/5-513–523/$25.00.

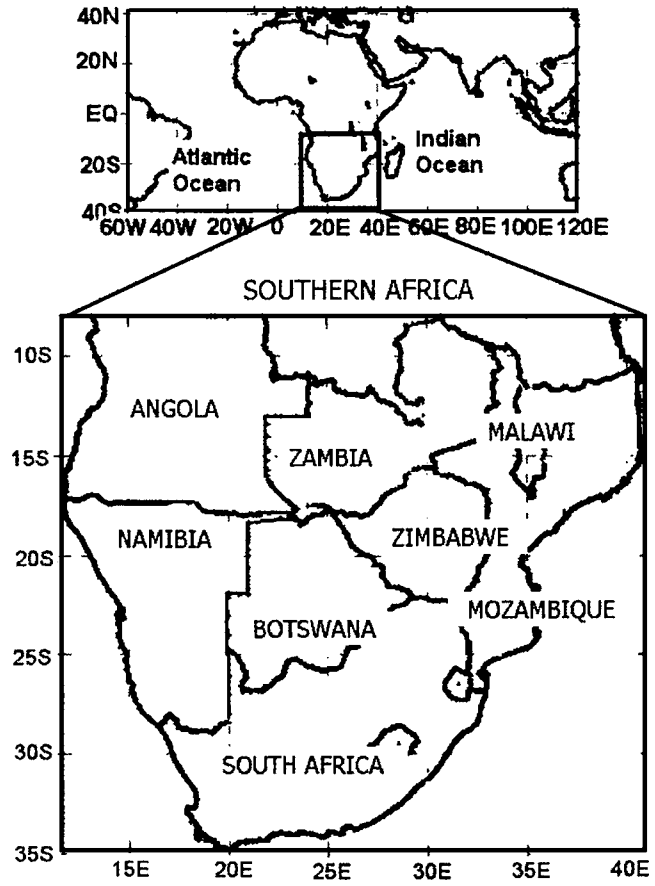

Fig. 1. Map showing study location, southern Africa, and Indian and Atlantic Oceans

gion's climate variability should employ methodologies that account for the nonstationarity found in the climate.

Since accurate prediction of rainfall has direct economic benefits (Shen et al. 2001), we also identified sectors of the surrounding oceans that are partly responsible for the southern Africa rainfall variability (Jury 1996; Reason and Mulenga 1999; Jury and Engert 1999; Landman and Tennant 2000; Landman et al. 2001). Then we used a nonlinear artificial neural network trained by a genetic algorithm (ANN-GA) (Mwale et al. 2004) and the SSTs of the identified sectors of the oceans to predict the summer rainfall of southern Africa.

Data

Summer Rainfall

Monthly summer rainfall data for 1950–97 from 58 grids located within the window 12.5°S–35°S, 15°E–42.653°E at a resolution of 2.5° latitude ×3.75° longitude were extracted for southern Africa (see Fig. 1). The rainfall data are part of a monthly precipitation data set constructed using the Thiessen polygon for global land areas from 1900 to 1998 provided by the UK Meteorological Office. This data set was constructed from station data and interpolated onto 2.5° × 3.75° grids. Thiessen polygon weights were used to average the gauge data within each gridbox. The quality control of the data is described in Hulme (1994). The grid data were also transformed to standardized rainfall data, whose quality control is described in Hulme (1994).

The 1950–97 data were extensively used to analyze the rainfall and SST variability and for prediction purposes because this data is of relatively high quality for this part of Africa, unlike the data collected before 1950.

Sea Surface Temperature

Monthly sea surface temperature (SST) anomaly grid data at 5° × 5° latitude and longitude resolution was extracted from the Indian Ocean (20°N–40°S, 40°E–105°E) and the Atlantic Ocean (10°N–40°S, 55°W–5°E) (Fig. 1). The SST data sets for 1950–98 were transformed into seasonal data by computing 3-month averages, i.e., October–December (OND), January–March (JFM), April–June (AMJ), and July–September (JAS). The SST data set is part of MOHSST6, a historical global data set of mean monthly SST anomalies with respect to the 1961–90 normals provided by the UK Meteorological office.

Research Methodology

Wavelet Transformation

The wavelet transformation of a real signal $x(t)$ with respect to the mother wavelet, ψ, is a convolution given as

$$W(b,a) = \frac{1}{\sqrt{a}} \int_0^T x(t) \psi^* \left(\frac{t-b}{a} \right) dt \tag{1}$$

where ψ^*=complex conjugate of ψ; and T=total length of the signal $x(t)$. In this equation, $W(b,a)$ is a wavelet spectrum, a matrix of energy coefficients of the decomposed time series at each scale a and time b. The magnitude of the wavelet spectrum coefficients show how well the wavelet matches with the time series and at each scale, spectrum coefficients also depict the amplitude of a time series. In a rainfall time series, power at each scale will be a good indicator of the frequency and also the magnitude of the rainfall events. To examine fluctuations in power over scales that have significant energy (as seen in the global wavelet spectrum) in a time series, the scale-averaged wavelet power (SAWP) of the wavelet spectrum is computed:

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

where C_δ=0.776 for the Morlet wavelet (used in this paper); δ_j=factor for scale averaging; δ_t=sampling period; and $j_1 \ldots j_2$ represent scales over which SAWP is computed. Many wavelets could be used as the mother wavelet. The Morlet wavelet is just one of them, but it is one of the most commonly used wavelets because it resembles the rainfall and SST temperature time series. Since SAWP is a time series of averaged variance in a certain band, the SAWP can be used to examine the modulation of one time series by another or modulation of one frequency by another within the same time series (Torrence and Compo 1998). The wavelet-based empirical orthogonal function (WEOF) was used to extract the joint modes of the SAWP's spatial and temporal variability and individual scale power.

Wavelet-Based Empirical Orthogonal Function

WEOF, also known as the wavelet principal component analysis (WPCA), has recently been adopted by Mwale et al. (2004) for extracting nonstationary uncorrelated modes of spatial and tem-

poral variability of rainfall and SST time series. The approach is called WEOF analysis because the empirical orthogonal function (EOF) method (Wilks 1995) is used to find eigenvectors of SAWP and individual scale power that describe the joint modes of spatial and temporal variability of climate data. Since the SAWP largely retains variations of climate with statistically significant energy, it is expected that the intrinsic nonstationarity of the climate process will be isolated. When actual data is used to extract the EOFs, the EOFs may have ambiguous meanings (when mode mixing occurs), since no frequency content is suggested by the modes of the EOFs of raw data (Huang et al. 1998).

Measures of Prediction Skill

To evaluate the prediction skill between southern Africa's observed and predicted summer rainfall, the Hansen Kuipers (HK), the Pearson correlation, and root-mean-square error (RMSE) were used.

The HK skill score, which is also known as the Kuipers' performance index (KPI) or the true skill score (TSS), is based on categorical forecasts, where the prediction consists of a flat statement that one and only one of a set of possible events will occur (Wilks 1995). For rainfall, the categories "dry," "near normal," and "wet," with tercile percentages of below 33%, 33–66%, and above 66%, respectively, have been used (e.g., Ntale et al. 2003; Mwale et al. 2004). In this paper, the HK score was applied to the rainfall grid boxes, each $2.75° \times 3.75°$. The reference point of the HK skill score is the number of correct observation hits expected by chance. With this reference point, the HK skill score is computed as follows (Ntale et al. 2003):

$$HK = \frac{C - C_c}{P - C_{oc}} \tag{3}$$

where C=total number of correct forecasts; P=total number of forecasts obtainable with a perfect forecast model; C_c=number of correct hits expected by chance; and C_{oc}=reference point. HK skill score can also be expressed using probabilities. For example, for the "dry," "near normal," and "wet" categorical forecasts, the HK score is computed as follows (Wilks 1995; Ntale et al. 2003):

$$HK = \frac{\sum_{i=1}^{3} P(O_i, P_i) - \sum_{i=1}^{3} P(O_i) \times p(P_i)}{1 - \sum_{j=1}^{3} [p(O_j)]^2} \tag{4}$$

where O_i and P_i=ith observed and predicted values in a square contingency table. To take into account the "dry," "near normal," and "wet" categories, a 3×3 contingency table is used. The maximum and minimum HK scores are between −1 and 1. HK scores of 1 represent perfect forecast, while HK scores of zero represent forecasts that are as good as random and HK scores of less that zero represent forecasts that are inferior to random forecasts.

The Pearson correlation coefficient (correlation coefficient), ρ, is a single-valued measure of linear association between two variables. The correlation coefficient is computed as follows:

$$\rho = \frac{\sum_{k=1}^{n} (O_k - \bar{O})(P_k - \bar{P})}{\left(\sum_{k=1}^{n} (O_k - \bar{O})^2 \right)^{1/2} \left(\sum_{k=1}^{n} (P_k - \bar{P})^2 \right)^{1/2}} \tag{5}$$

where O_k and P_k=observed and predicted values; $\bar{O}$ and $\bar{P}$=their respective means; and n=sample size. The correlation coefficient ranges between 1 and −1. The extreme positive and negative values indicate perfect positive and negative linear associations, respectively.

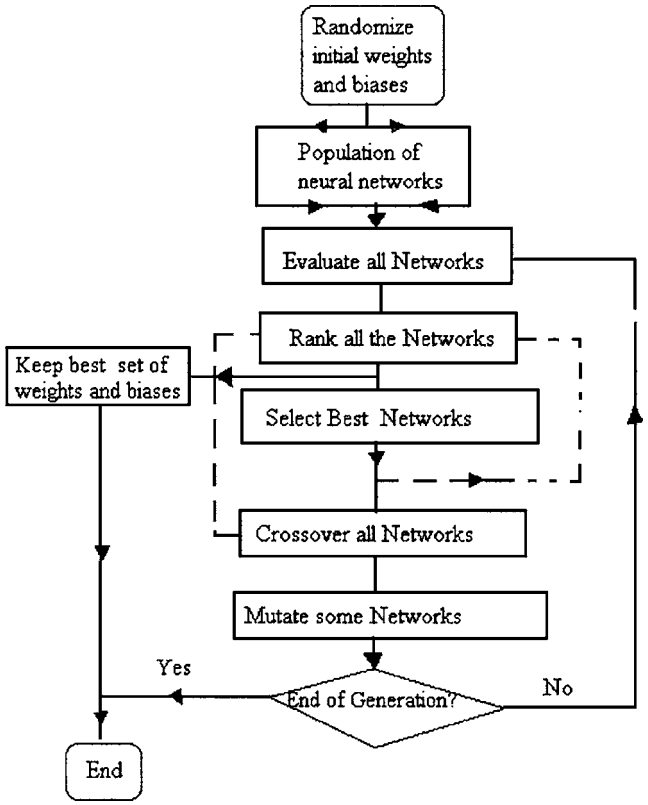

Fig. 2. ANN-GA model

The RMSE is computed as

$$\text{RMSE} = \left(\frac{1}{n} \sum_{k=1}^{n} (O_k - P_k)^2 \right)^{1/2} \tag{6}$$

The RMSE for a perfect prediction is zero and increasing RMSE indicates decreasing the predictions' accuracy.

Prediction Model

The prediction model used in this paper was the ANN-GA of Mwale et al. (2004) (see Fig. 2). The ANN-GA model uses several thousand neural networks during each iteration, or generation. The weights of the ANN-GA model are updated at each generation by the genetic algorithm, which is based on principles of natural biological evolution. This model has previously been shown to perform better than linear canonical correlation analysis (CCA) for predicting rainfall (Mwale et al. 2004). Each neural network in the ANN-GA model had three layers (input, hidden, and output). In the current application, the ANN-GA model was applied twice during each prediction exercise, once with 2 neurons in the hidden layer and another with 10 hidden neurons. The two solutions were then averaged.

The ANN-GA functions as follows. An initial set of weights, $\mathbf{W}^1$ and $\mathbf{W}^2$, and biases, $\mathbf{B}^1$ and $\mathbf{B}^2$ (bold capital letters indicate a population, and superscripts indicate the first and second layers of the ANNs), is randomly created for a large number of neural networks, called a population. These weights are assigned to each link of all the neural networks in the population. The output of each neural network in the population is evaluated against a

known target called the predictand. Based on how they perform against the predictand, all the neural networks in the population are ranked, the best network at the top and the worst network at the bottom.

The best 85% of the ranked neural networks in the population are selected at random to comprise offspring of the next generation. The population is kept constant, hence certain neural networks are selected more than once. Selecting from the best 85% of the ranked population ensures that on the average the new generation of neural networks has comparatively better fitness than the original population. Therefore, the selection procedure shifts the search toward better solution spaces of the problem.

Next, pairs of neural networks are selected in the same order in which they were selected from the previous population, and their weights and biases are exchanged. This procedure is called crossover. To effect exchange of weights and biases, a location is randomly chosen in the hidden layer and weights on either side of the location are exchanged between the two neural networks. This procedure is repeated between all other pairs of neural networks in the population.

At each generation, the weights and biases of a small percentage (1% was used in this study) of the population is randomly selected to be replaced. The process is called mutation and is designed to restore good weights and biases eliminated during selection process. Mutation is a purely random procedure, and it is always kept to a minimum to prevent the search from becoming a random process. If mutation results in a better neural network, that network will likely survive in the next selection; but if mutation results in a more inferior network, that network will likely perish in the next selection. The neural networks are once again evaluated against the same known predictand.

This procedure is repeated through several generations. At each generation, the best network is stored away and replaced only when a better solution is found. After many generations, the selection pressure reduces the variation among the neural networks as similar but better neural networks drive out inferior ones. Eventually, the whole population of neural networks has the same weights and biases, i.e., the so-called global solution. There is no way to tell whether the final neural network is a global solution or not. Since mutation is always active, it is not possible to achieve 100% convergence of neural networks. Hence, convergence is assumed to be reached when at least 95% of the neural networks have the same weights and biases. At the end of the run the weights and biases of the best surviving network are kept to be used for making predictions with new input data.

To evaluate each neural network, the output, y, is obtained as a nonlinear translation of the weighted average of the input data, x, which has been normalized; i.e., $x=x-\bar{x}/\sigma_x$, where $\bar{x}$ and σ_x=mean and standard deviation of x, respectively. The first step is to compute the weighted input to the jth-hidden unit, $hidunit_{pj}$

$$hidunit_{pj} = \sum_{i=1}^{N} W_{ji} x_{pi} + B_{jo} \tag{7}$$

where N=total number of input nodes; W_{ji}=weights from input unit i to the hidden unit j; B_{jo}=biases for hidden neuron j; and x_{pi}=ith input of pattern p (in this case SST PCs are used). The hidden layer undergoes a nonlinear translation

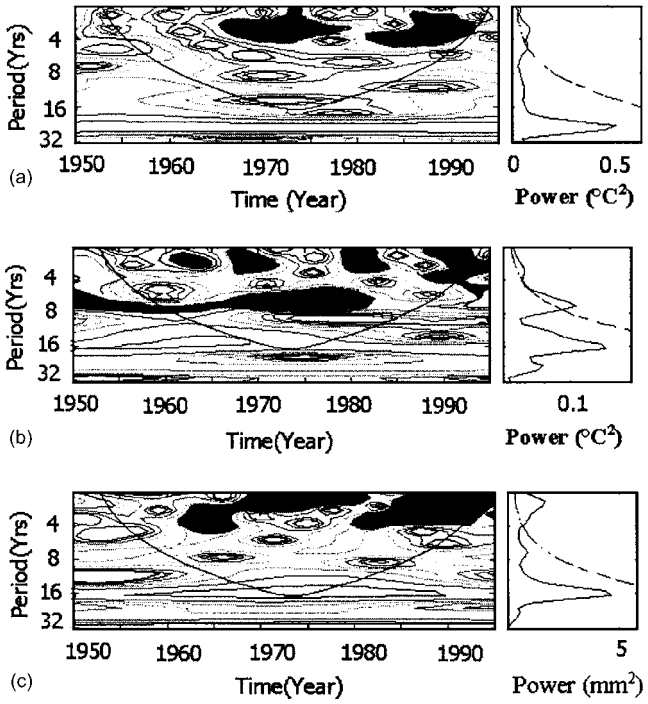

Fig. 3. Examples of (left) local and (right) global wavelet spectra for sea surface temperature in: (a) Atlantic Ocean; (b) Indian Ocean and rainfall in (c) southern Africa. The solid line in the local wavelet spectra is cone of influence and dark regions have power that is statistically significant. In global wavelet spectrum, dotted line represents 95% confidence level and peaks above line are statistically significant.

$$f_1(hidunit_{pj}) = \frac{1}{1 + e^{-hidunit_{pj}}} \tag{8}$$

where $f_1(hidnunit_{pj})$=jth neuron nonlinear activation function.

The output, y_{pk}, is computed as a weighted average of the hidden units

$$y_{pk} = \sum_{j=1}^{M} W_{kj} f_1(hidunit_{pj}) + B_{ko} \tag{9}$$

where M=number of hidden units; W_{kj}=weight connecting the hidden node j to the output k; B_{ko}=bias for output neuron k.

Results and Discussion

Dominant Periods of Rainfall and SST via Wavelet Spectra

The local and global wavelet spectra were constructed for some locations in southern Africa, the Indian Ocean and the Atlantic Ocean. Fig. 3 shows that appreciable power in the wavelet spectra of the SST and rainfall data was found to exist between 2 and 8 years, i.e., within the El Niño–southern Oscillation (ENSO) band. Power within this period range was used for the analysis.

Spatial Regimes of Summer Rainfall

Theoretically 58 WPCs were found using the WEOF analysis, but only the first two leading modes revealed regional patterns of

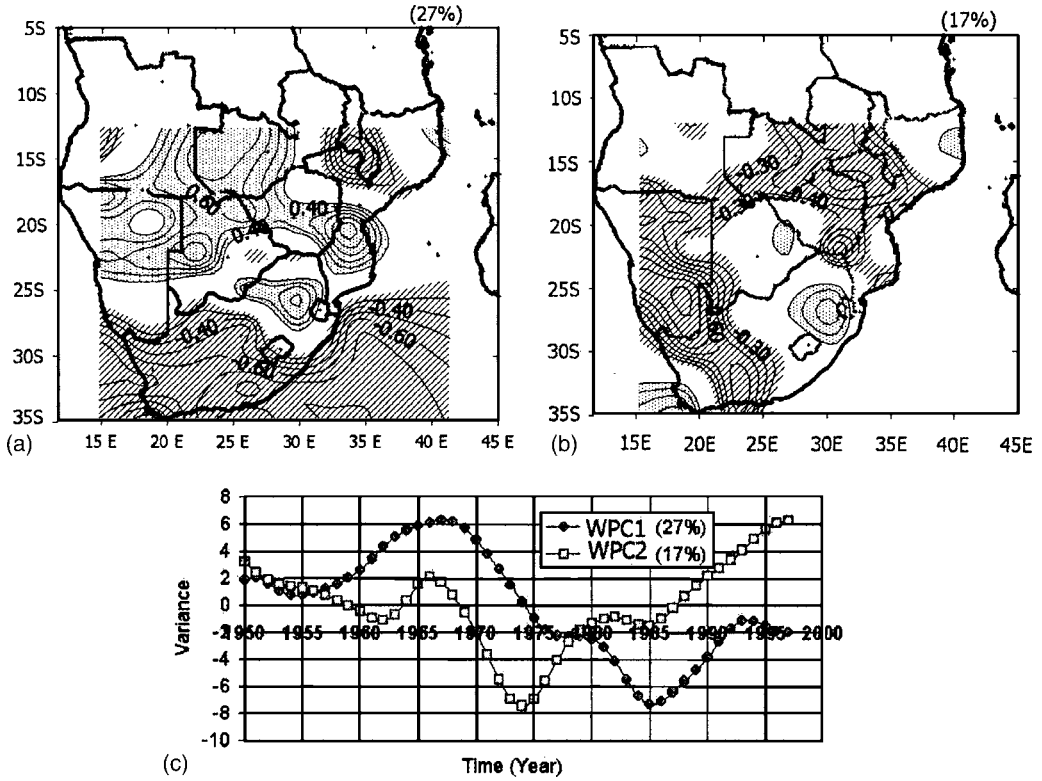

Fig. 4. Spatial displays of correlation between: (a) WPC1; (b) WPC2 of summer rainfall of southern Africa and SAWP of individual grids. Dotted areas indicate positive correlations, while hatched areas indicate negative correlations. Percentage values above figures represent proportions of total variance explained by each WPC. Corresponding time domain WPCs are shown in (c).

spatial variability of the southern Africa summer rainfall. The two modes explained a combined variance of 44% (i.e., 27 and 17% for the first and second modes, respectively). Their spatial distribution patterns are shown in the form of correlation coefficients between the rainfall WPCs and each of the $2.5° \times 3.75°$ gridded rainfall SAWP (Fig. 4). Because of the strong persistence in the WPCs and SAWP time series, the effective length used to compute the 95% level of the correlation was found using the method given by Wilks (1995)

$$N_{eff} = N\frac{1-\rho_1}{1+\rho_1} \tag{10}$$

where ρ_1=lag1 autocorrelation; and N_{eff}=effective or revised length of the time series. As expected, N_{eff} is much shorter than N.

The leading spatial mode (WPC1) of the southern Africa summer rainfall suggests that the variation of summer rainfall between countries in northern interior of southern Africa and northern South Africa is out of phase to the regions located south of 25°S and areas in southern Malawi/northern Mozambique [Fig. 4(a)]. WPC1 is positively correlated to the rainfall SAWP in the northern parts of southern Africa and northern South Africa and negatively correlated to the regions south of 25°S and southern Malawi/northern Mozambique. This mode explains variances of between 16 and 49% (i.e., correlations of 0.4 and 0.7, Houghton and Tourre 1992) of the rainfall SAWP variance in Zambia, northern Namibia, northern South Africa, central Mozambique and southern South Africa and correlations of between −0.3 to −0.7 are found south of 25°S. The largest local variations are accounted for in northern-central Zambia (49%), showing that the source of atmospheric variability is from the Congo/Atlantic Ocean region. Fig. 5 shows the prevailing moisture transport by winds at the 850 hpa heights during the summer rainfall season, OND and JFM. Figs. 5(a and b) show that the rainfall season of the northern sections of southern Africa is dominated by strong inflows of moisture from the Congo/Atlantic Ocean basin. The variability associated with this moisture inflow accounts for 49% of the rainfall SAWP variability. In addition, Fig. 5 also shows that during early summer, the moisture transport for the region is from the southwest and northern Indian Ocean and the southeast Atlantic Ocean. During the dry season, which begins in April, anticyclones are prevalent in southern Africa. These centers of high pressure are responsible for dry winds that descend into the region for the period April–September.

WPC2 is positively correlated to the rainfall SAWP of northern South Africa, the Western Cape, eastern Botswana, and the northern-most sections of Mozambique and is negatively correlated to Namibia, southeastern Angola, Zambia, Zimbabwe, and most of central and southern Mozambique. Large correlations between WPC1 and rainfall SAWP are found in Namibia, with the largest correlation in the southern desert region of Namibia. Elsewhere in the region, this mode accounts for between 0 and 9% of the rainfall SAWP variance. WPC2 accounts for the variation of rainfall SAWP that is not accounted for by WPC1 in these countries. The strong correlations along the Namibian coast suggest the variation of rainfall SAWP may be influenced by the Benguela Ocean current of the Atlantic Ocean, while correlations on the east coast and in Mozambique suggest that rainfall SAWP variation in this region is also influenced by the variations of SST in the Indian Ocean.

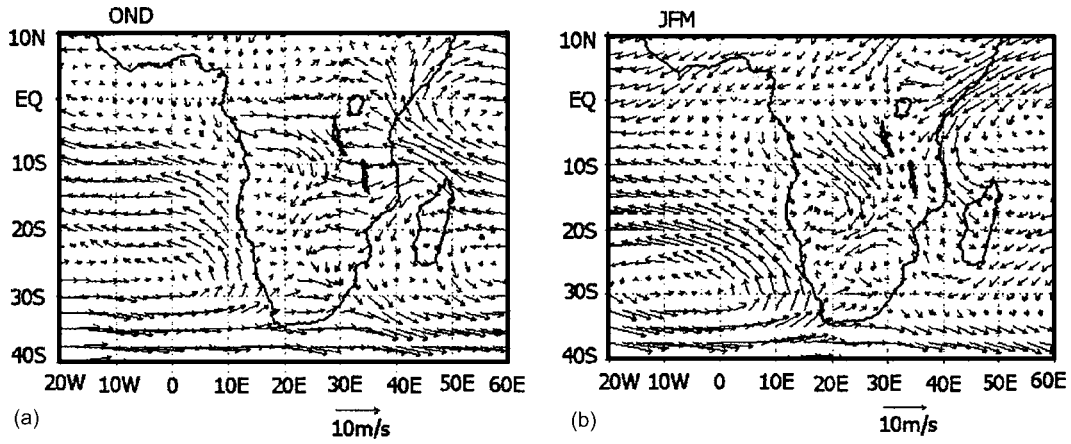

Fig. 5. 850-hPa winds showing prevailing atmospheric circulation and moisture transport in terms of wind speed (m/s) and direction throughout year averaged for 1959–70: (a) OND; (b) JFM

Temporal Regimes of Summer Rainfall

The temporal variabilities of the WPCs corresponding to the leading spatial modes are shown in Figs. 4(c and d). WPC1 shows increased energy in 1955–68 and 1986–93, and decreased energy in 1969–85. Since WPC1 is positively correlated to the northern interior of southern Africa and northern South Africa, these regions experienced increased rainfall for 1955–68 and 1986–90. However, between 1969 and 1985 rainfall decreased in these regions. Although the 1969–85 decrease in rainfall was followed by an increase of rainfall between 1986 and about 1990, the 1986–90 rainfall increase remained largely small to offset the rainfall decrease that in 1969–85.

The WPC2 time series shows a decrease in energy between 1950 and 1973, followed by an increase in energy between 1974 and 1997. Since WPC2 is out of phase with rainfall SAWP in most of southern Africa (i.e., southern Namibia, Zambia, Zimbabwe, and most of Mozambique south of 15° S), these areas experienced an increase in rainfall between 1950 and 1973 but experienced a consistent decrease in rainfall between 1974 and 1997. Since WPC2 explains the proportion of variance not accounted for by WPC1, it emphasizes the decrease of summer rainfall that occurred in Namibia, Zambia, Zimbabwe, and Mozambique between 1974 and 1997 and explains the widespread droughts that occurred in the northern sections of the southern Africa region, including northern South Africa, observed in 1991–92 and 1994–96. As will be seen when SST variations in the surrounding oceans are examined, the decrease in the southern Africa summer rainfall, shown by WPC1 and WPC2, coincides with similar decreases in the power of SST WPC1 and 2 of the surrounding oceans. As will be seen in the next section, WPC1 and WPC2 account for the reduction of rainfall as modulated by different frequency regimes and explain the decrease in streamflow and dam levels between 1974 and 1997 observed in Namibia, Zambia, Zimbabwe, and parts of Mozambique and the Veld regions in northern South Africa (Fanta et al. 2001; Jury 1996).

Frequency Regimes of Summer Rainfall

The wavelet spectra constructed using the rainfall data for northern southern Africa and the SST data of the southern Indian Ocean showed that the dominant oscillations of rainfall and SST variability generally lay in the 2–8 year band. By choosing various values of δ_j, the wavelet power of a number of periods between 2 and 8 years were initially analyzed. However, the periods of 2, 4.8, 5.6, and 6.7 were selected for presentation of results. These periods generally show representative spatial and temporal patterns for southern Africa rainfall and the surrounding oceans, SST. The wavelet power at the 2, 4.8, 5.6, and 6.7 years was extracted for each of the 58 grid locations and subjected to WEOF analysis.

The spatial distribution patterns of the leading modes of the rainfall wavelet power extracted for 2 (representing power in the 2 to 2.4 year band) and 6.7-year periods (representing power in the 4 to 8 year band) are shown in Fig. 6. Inspection of Fig. 6 shows that the leading modes of the rainfall wavelet power at 6.7 years explain more variance than the one at 2 years (i.e. 33 vs. 29%). Comparison of Figs. 6, 4(a, and b) shows that the spatial patterns of WPC1 for the 6.7-year period has similar spatial patterns as the WPC1 of the rainfall SAWP, while the spatial patterns of WPC1 periods at 2 years have spatial patterns similar to WPC2 of the rainfall SAWP. This shows that on the regional scale, the main mode of variation of summer rainfall was dominated by 4- to 8-year periods for most of southern Africa, followed by the ones around 2 years. The 4- and 8-year cycles are within the ENSO band and hence El Niño is partly responsible for rainfall decrease. However, since part of the SAWP variation is also accounted for by the periods around 2 years, especially in Namibia, Zambia, Angola, Zimbabwe, and Mozambique, especially areas along the coastal regions (i.e., areas where WPC2 was strongest), summer rain fall in these locations likely is influenced by SST variations from the surrounding oceans.

Temporal variations of WPC1 at the 2 and 6.7 years show that power variations at the 2-year period peaked in 1974 and declined between 1974 and 1997 (for most countries in the interior of southern Africa), while power in the 5.6- and 6.7-year periods peaked around 1965 and declined until around 1985 and remained low between 1985 and 1997.

In summary, the decline of rainfall in the interior of southern Africa was associated with the 4- to 8-year cycles from the mid 1960s until 1997, while rainfall in the southern parts of South Africa and Malawi has been on the increase. The 4- to 8-year cycles are within the ENSO band and hence El Niño is partly responsible for the decrease in rainfall. However, the cycles

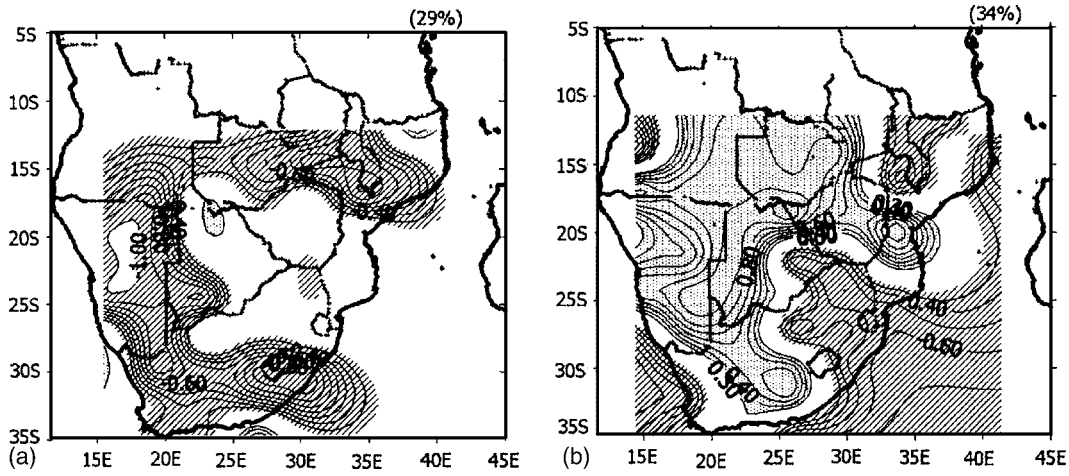

Fig. 6. Spatial displays of the correlation between: (a) the 2-year WPC1; (b) 6.7 year WPC1 and gridded summer rainfall SAWP. Percentage values above figures represent proportions of total variance explained by each WPC.

around 2 years also explained a large proportion of the 2 to 8 year SAWP variation in Angola, Zambia, Namibia, Zimbabwe, and Mozambique, especially in Namibia's coastal regions. This shows that the surrounding oceans are also responsible for the decrease observed in the summer rainfall. It is also not surprising that following almost 30 years of consistent decrease in the rainfall of the region, especially the northern sections and northern South Africa, the droughts caused widespread famines.

Associations between Rainfall and Indian and Atlantic Ocean SSTs

To understand the association between the SST of the surrounding oceans and southern Africa summer rainfall, the temporal and spatial patterns of the leading modes of SST (WPC1 and WPC2) of both the Atlantic and Indian Oceans were also established. Then the southern Africa rainfall WPCs were correlated to the

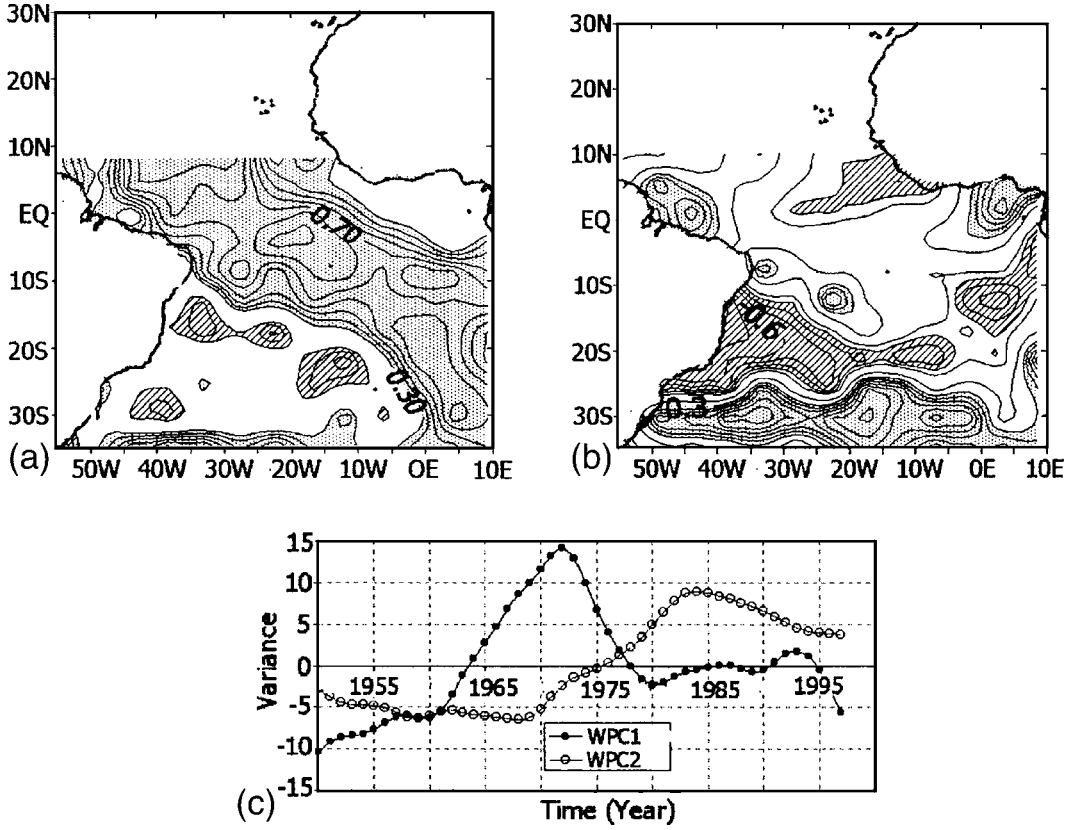

Fig. 7. Contour plots of spatial correlation patterns between: (a) WPC1; (b) WPC2 of annual Atlantic Ocean SST and SST SAWP of individual grids at 0.1 contour intervals. WPC1 and WPC2 explain 28 and 22%, each of total SAWP variance, respectively. The temporal variations of two WPCs of Atlantic Ocean SST are shown in (c).

gridded SST SAWP of both oceans. The resulting correlation patterns were analyzed against the observed spatial patterns of the SST WPCs. The association was assumed to exist if the temporal variation of the SST and rainfall WPCs were well correlated and the correlation patterns between rainfall WPCs and the SST SAWP appeared similar to the patterns of SST WPCs.

For example the spatial and temporal patterns of annual SST WPC1 and WPC2 of the Atlantic Ocean are shown in Fig. 7. Seasonal spatial and temporal patterns (figures not shown) are similar to those shown in Fig. 7. Figs. 4(c and d and 7(c) show that decreasing rainfall in areas north of 25°S and northern South Africa are associated with decreasing SST in the Atlantic Ocean. The decreasing SST at the annual time scale was shown by WPC1 (28%) and WPC2 (22%). At the seasonal time scale, the SST is depicted by WPC1 only: i.e., JFM, 27%; AMJ, 28%; JAS, 27%; and OND, 28%. Fig. 7(c) shows that similar to rainfall, the overall SST increased between 1950 and 1971 and declined between 1972 and 1997. SST WPC2, as shown in Fig. 7(c), shows that SST increased between 1970 and 1983 and declined between 1984 and 1997 for areas of the Atlantic Ocean south of 30°S and vice versa for areas north of 30°S.

The association between summer rainfall WPCs and SST SAWP of both the Indian and Atlantic Oceans was examined for all the seasons. Although associations were found between summer rainfall WPCs and SST SAWP for all seasons, the relationship between rainfall WPCs and AMJ SST SAWP was chosen. As shown by Figs. 5(a and b), the relationship between southern Africa summer rainfall (OND-JFM) and SST variations of OND and JFM are synchronous. The JAS season offers no lead time. Hence the AMJ season was used.

The association between southern Africa summer rainfall WPC1 and 2 and AMJ Atlantic Ocean SST SAWP was best described by the variations in the Benguela Ocean current SST off the west coast of southern Africa, which also forms the dominant spatial pattern of the annual and AMJ SST SAWP of the Atlantic Ocean. This shows that southern Africa rainfall is associated with the AMJ Atlantic Ocean SST variations.

The correlation patterns formed between WPC1 of southern African summer rainfall and the individual 5°×5° AMJ SST SAWP time series of the Atlantic and Indian Oceans are shown in Fig. 8. Also shown in the figure is the relationship between WPC2 and the Atlantic Ocean. Summer rainfall WPC1 is positively correlated to the SAWP of the Southwest Indian Ocean and the Brazil Ocean current SST and negatively correlated to the SAWP of the northern Indian Ocean and the Guinea Ocean current SST. Since the rainfall WPC2 is negatively correlated to both the rainfall SAWP and the SST SAWP of the Benguela Ocean current of the Atlantic Ocean, WPC2 has a positive relationship with the Benguela Ocean current SST.

Since summer rainfall is out of phase with the northern Indian Ocean and in phase with the southwest Indian Ocean for the northern parts of southern Africa, warming of the northern Indian Ocean SST in AMJ decreases summer rainfall for Zambia, northern Namibia, Zimbabwe, parts of central Mozambique, Botswana, and northern South Africa. Conversely, cooling of the southwest Indian Ocean decreases rainfall in the northern parts of southern Africa and increases rainfall for areas south of 25°S. Positive correlation was found between northern Indian Ocean SST SAWP during OND and JFM and summer rainfall WPCs (figures not shown) meaning that the increase in the OND and JFM SST in the northern Indian Ocean enhances summer rainfall in Zambia, northern Namibia, Zimbabwe, parts of central Mozambique, Botswana, and northern South Africa. WPC1 of the Indian Ocean

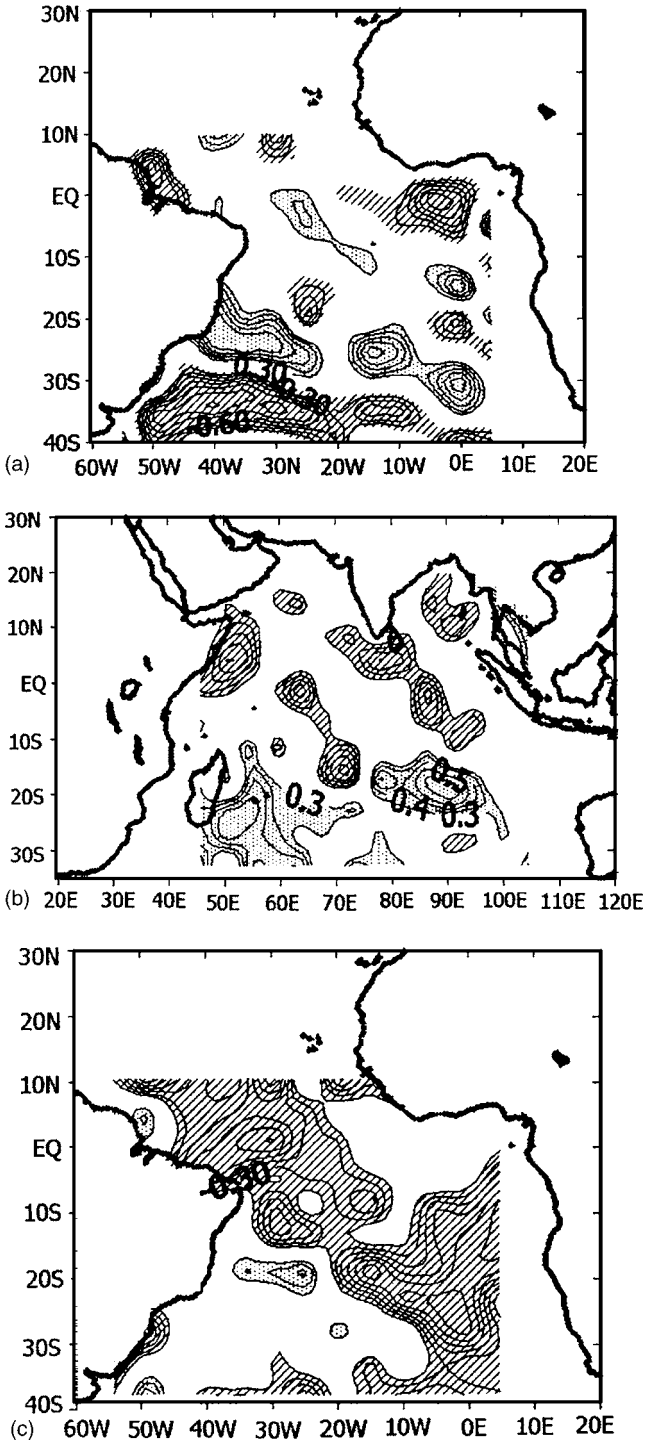

Fig. 8. Associations between southern Africa summer rainfall WPCs and SST SAWP: (a) rainfall WPC1 and Atlantic Ocean SST SAWP; (b) rainfall WPC1 and Indian Ocean SST SAWP; and (c) rainfall WPC2 and Atlantic Ocean SST

SST SAWP (figures not shown) indicates that the northern Indian Ocean SST began warming in 1960 and the southern Indian Ocean SST begun cooling around 1970. From the WPC2 time series in Fig. 4, it is clear that southern Africa responded almost immediately to SST variations in the southwest Indian Ocean. Other researchers have found the southwest Indian Ocean to influence rainfall variability in some parts of southern Africa (e.g., Mason 1995; Jury 1996; Landman et al. 2001).

Predicting Southern African Summer Rainfall

The seasonal relationships found between the WPCs of southern African summer rainfall and the SST SAWP of the Indian and Atlantic Oceans implies that SSTs in certain regions of the Indian and Atlantic Oceans may be good predictors of the summer rainfall in southern Africa. From Fig. 8, April, May, and June SST data from both the positively and negatively correlated regions of the Atlantic and Indian Oceans were extracted. Since the effective lengths (i.e., the revised lengths) of the WPCs and the SAWP reduced to 3 or 4 degrees of freedom, the significant level of correlation was above 0.997. Very few data at this level of statistical significance could be collected. Since the rainfall variability was partly transmitted from the SST variability, the explained variance in the rainfall could be used to extract SST data from the oceans. Hence, we collected the SST data from ocean areas where the correlation between rainfall WPCs and SST SAWP ranged between −0.3 and −1.0 or 0.3 and 1. The monthly data were averaged over a 3-month period (AMJ) for each $5° \times 5°$ ocean SST grid.

Prediction Skill of Artificial Neural Network-Genetic Algorithm Model

To predict the summer rainfall of southern Africa, a population of 2,300 neural networks was used at each generation. Since the rainfall data were normalized, correlation was used as the objective function. Selection of neural networks to comprise the next population was made from 1,955 of the ranked neural networks. The one-point crossover scheme was applied to the entire population at each generation and the mutation rate was restricted to 23 neural networks. The ANN-GA model generally converged after 50 generations, thus only 120,000 solutions were evaluated for each run. The ANN-GA model kept the best neural network at each generation that was used for prediction.

The performance of the best neural networks in terms of the Pearson correlation, the categorical HK scores, and the RMSE between predicted and observed standardized southern Africa summer rainfall for the 1988–97 validation period (independent of the calibration experience of 1950–87) are shown in Fig. 9. The correlations generally ranged between 0.83 and 0.98. This means that ANN-GA predicted rainfall accounted for between 69 and 96% of the observed rainfall variability. The HK scores ranged between 0.2 and 1.0. Generally an HK score of 0.81 (see Table 1) means that 9 out of 10 years of forecasted seasonal rainfall fall in their correct categories while 2 years fall in wrong categories. The HK score gets poorer as either more predicted rainfall falls in the wrong category or the predicted rainfall falls in categories far away from the correct category (i.e., instead of wet, dry rainfall is predicted). Hence, some locations (e.g., eastern Angola) may have high Pearson correlation (i.e., 0.9) but HK scores around 0.5 and RMSE of 0.55. Table 2 shows that in such locations, 3 out of 10 years fall in the wrong category (i.e., 1 year falls in the near normal instead of wet and 2 years fall in the dry instead of the near normal category). The RMSE ranged between 0.25 and 0.72. They were generally higher in areas with poor predicted summer rainfall.

Fig. 10 shows the plot of observed versus predicted standardized rainfall for some locations throughout the southern African region. The neural networks were able to satisfactorily predict the summer rainfall using the previous AMJ SST from the surrounding oceans. Fig. 10 clearly shows two findings: (1) that the ANN-GA models captured the nonlinear features of the rainfall and SST interaction; and (2) that the SST data of the ocean that influences summer rainfall was correctly identified. The capability to identify robust predictors of southern Africa summer rainfall is encouraging.

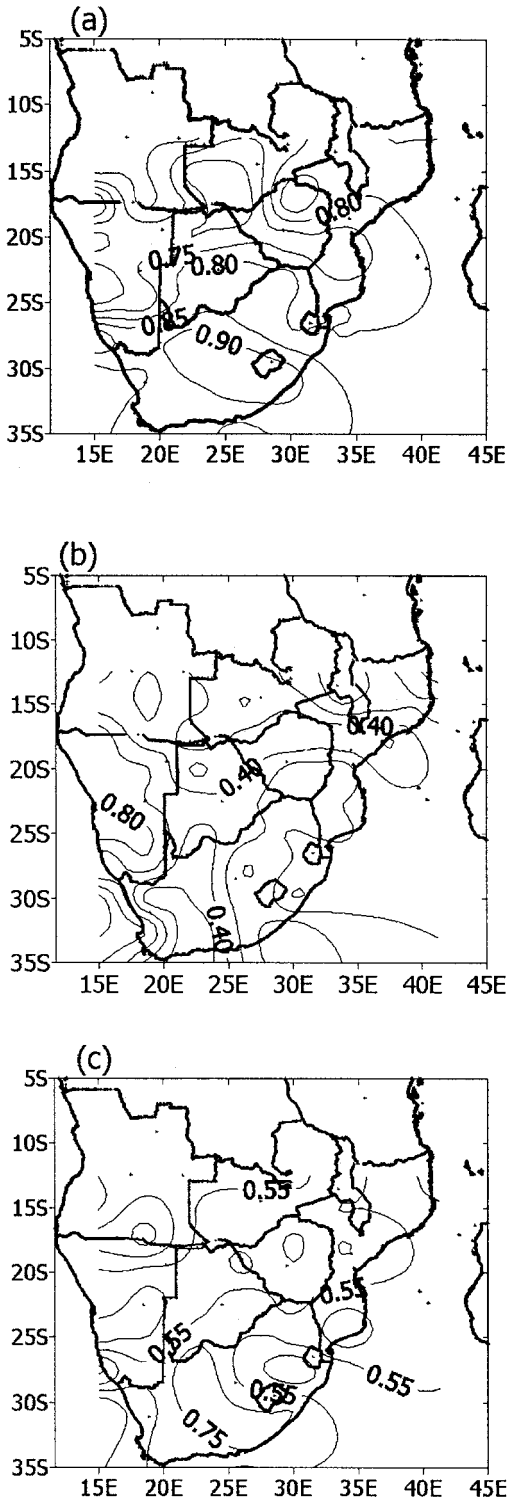

Fig. 9. Skill scores of ANN-GA prediction model of summer rainfall of southern Africa: (a) correlation; (b) RMSE; and (c) HK

The above results show that the summer rainfall variability is influenced by SST variations in the southwest and northwest Indian and the southern Atlantic Oceans. Using SSTs from these sectors of the oceans to predict rainfall in southern Africa, the

Table 1. Square Contingency Table for HK Score of 0.81

	Wet	Near normal	Dry	
Wet	**2**	1	0	
Near normal	0	**5**	0	
Dry	0	0	**2**	
		Correct scores (diagonal)		**9**
		Total score		10

nonlinear ANN-GA system could predict between 69 and 96% of the rainfall variability at 3-month lead time. Given that the prediction skill of ANN-GA was consistently high shows that the nonlinear relationships between southern African rainfall and SST variations of the Indian and Atlantic Ocean basins are captured by the ANN-GA model. It also demonstrates that when nonstationarity of the atmosphere-ocean processes is taken into account, robust climate predictors could be identified.

Summary and Conclusions

We have used wavelet empirical orthogonal function analysis (WEOF) to identify and analyze dominant spatial, temporal, and frequency regimes of southern African rainfall and their relationships with the SSTs of the Indian and Atlantic Oceans.

Using the 1950–97 rainfall SAWP data (which represents the average rainfall variance of 2–8 year cycles), the spatial pattern of the leading summer rainfall WPC1 showed that southern Africa can be delineated into two main regions, north (including northern South Africa) and south of 25°S. The summer in the two regions is out of phase. The northern region, consisting of eastern Angola, Zambia, Zimbabwe, Namibia, Botswana, Mozambique, and northern South Africa are out of phase with areas in the south of 25°S (which is mainly the southern half of South Africa, probably representing the cape region) and southern Malawi/northern Mozambique.

Using the 1950–97 SAWP data, the northern part of the region was found to have generally experienced an increase in rainfall between 1950 and 1965 and between 1985 and 1990 and a decrease in rainfall between 1965 and 1985. The increase in rainfall between 1985 and early 1990s was largely small and short-lived to offset the 1965–85 decrease. Hence, generally the northern half of southern Africa (including northern South Africa) experienced decreasing rainfall between 1965 and 1997, representing over 30 years of decreasing rainfall. The summer rainfall decrease over such a long period might explain the high number of widespread droughts during the 1990–97 period.

Regionally the summer rainfall decrease was associated with periods of between 4 and 8 year cycles, showing ENSO was partly responsible for the decrease in rainfall. However, the cycles around 2 years also explained a large proportion of the 2 to 8 year rainfall SAWP variation in Angola, Zambia, Namibia, Zimbabwe,

Table 2. Square Contingency Table for HK Score of 0.50

	Wet	Near normal	Dry	
Wet	**4**	0	0	
Near normal	0	**0**	2	
Dry	0	1	**3**	
		Correct scores (diagonal)		**7**
		Total score		10

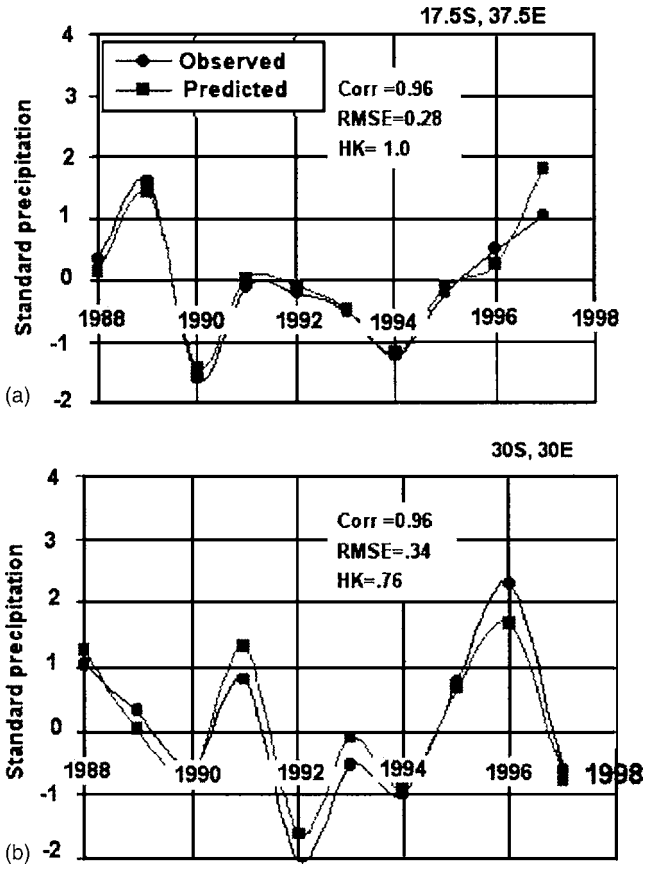

Fig. 10. Plots of observed versus predicted for some selected stations

and Mozambique, especially along the coastal regions of Namibia. This shows that the surrounding oceans are also partly responsible for the decrease in southern African summer rainfall. The decrease in SST in the Benguela Ocean current of the Atlantic Ocean and the southwest Indian Ocean between 1970 and 1997 were well correlated to the decrease in summer rainfall of areas north of 25°S of southern Africa. This shows that cooling of SST in the Benguela Ocean current and the southwest Indian Ocean are associated with decreasing southern African summer rainfall.

Using the southern African summer rainfall's WPC1 and WPC2 and gridded SST SAWP, the relevant SST fields from the Indian and southern Atlantic Oceans associated with southern African summer rainfall were identified. The ocean regions associated with the rainfall (i.e., with correlations above 0.3) were chosen as predictors and used to drive the nonlinear artificial neural network trained by the genetic algorithm (ANN-GA) prediction model to predict the summer rainfall of southern Africa at 3-month lead time.

As a nonlinear model, ANN-GA predicted the regional summer rainfall of southern Africa accurately, with correlations of between 0.83 and 0.98, HK scores of 0.2 and 1.0, and RMSE of between 0.25 and 0.72. Such kinds of preliminary prediction skill on a regional scale achieved by a statistical model at seasonal lead time is very encouraging.

Apparently by exploiting the nonstationary characteristics of climate data, we archived an understanding of the regional spatial and temporal variability of the southern African summer rainfall, the spatial and temporal variabilities of the periods associated with the southern African summer rainfall, and the spatial and

temporal variability of SST of the surrounding oceans. By exploiting the nonlinear characteristics of the ocean-atmosphere interaction, we managed to obtain an encouraging prediction skill of southern African summer rainfall using the SST data for the surrounding oceans.

Acknowledgments

This study is partly funded by Natural Science and Engineering Research Council (NSERC) of Canada, and the first author is supported by the commonwealth scholarship of CIDA, Canada. The U.K. meteorological office provided both the SST grid data for the Indian and Atlantic Oceans (part of MOHSST6) and the rainfall data for eastern Africa. The wavelet analysis was done using the software of Torrence and Compo (1998) downloaded from http://www.paos/colorado.edu/research/wavelets.

References

BBC. (2002a). "Africa famine: Country by country." ⟨http://news.bbc.co.uk/2/hi/africa/2027079.stm⟩ (July 24, 2004).

BBC. (2002b). "Hunger spectra over Angola." ⟨http://news.bbc.co.uk/2/hi/africa/2504661.stm⟩ (July 24, 2004).

BBC. (2002c). "Why famine stalks Africa." ⟨http://news.bbc.co.uk/2/hi/africa/2449527.stm⟩ (July 24, 2004).

Currie, R. G. (1993). "Luna-solar 18.6 and 10-11-year solar cycle signals in South African rainfall." *Int. J. Climatol.*, 13(3), 237–256.

Dlamini, E. M. (1995). "Modeling of an Africa basin using 4 rainfall-runoff models." MS thesis, Univ. of Alberta.

Dyer, T. G. J., and Tyson, P. D. (1977). "Estimating above and below normal rainfall periods over South Africa 1972–2000." *J. Appl. Meteorol.*, 16(12), 145–147.

Fanta, B., Zaake, B. T., and Kachroo, R. K. (2001). "A study of the variability of the river flow of the southern Africa region." *Hydrol. Sci. J.*, 46(4), 513–524.

Houghton, R., and Tourre, Y. (1992). "Characteristics of low frequency sea surface temperature fluctuations in the tropical Atlantic." *J. Clim.*, 5(7), 765–772.

Huang, N. E., 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 nonstationary time series analysis." *Proc. R. Soc. London, Ser. A*, 454(1971), 903–995.

Hulme, M. (1994). "Validation of large-scale precipitation fields in general circulation models." *Global precipitations and climate change*, M. Desbois and F. Desalmand, eds., NATO ASI Series, Springer, Berlin, 387–406.

Jury, M. R. (1996). "Regional teleconnection pattern associated with summer rainfall over South Africa, Namibia and Zimbabwe." *Int. J. Climatol.*, 16(2), 135–153.

Jury, M. R., and Engert, S. (1999). "Teleconnections modulating interannual climate variability over northern Namibia." *Int. J. Climatol.*, 19(13), 1459–1475.

Landman, W. E., Mason, S. J., Tyson, P. D., and Tennant, W. J. (2001). "Retroactive skill of multitiered forecasts of summer rainfall over southern Africa." *Int. J. Climatol.*, 21(1), 1–19.

Landman, W. E., and Tennant, W. J. (2000). "Statistical downscaling of monthly forecasts." *Int. J. Climatol.*, 20(13), 1521–1532.

Mason, S. J. (1995). "Seas-surface temperature-South African Rainfall Associations, 1910–1989." *Int. J. Climatol.*, 15(2), 119–135.

Mason, S. J., and Tyson, P. D. (2001). "The occurrence and predictability of droughts over southern Africa." *Drought, A global assessment*, D. A. Wilhite, ed., Vol. 1, Routledge, New York.

Mwale, D., Gan, T. Y., and Shen, S. P. P. (2004). "A new analysis of the variability, teleconnectivity, and predictability of central southern Africa rainfall for the period 1950–97." *Int. J. Climatol.*, 24(12), 1509–1530.

Ntale, H. K., Gan, T. Y., and Mwale, D. (2003). "Prediction of east African seasonal rainfall using simplex canonical correlation analysis." *J. Clim.*, 16(12), 2105–2112.

Reason, C. J. C., and Mulenga, H. (1999). "Relationships between south African rainfall and SST anomalies in the southwest Indian Ocean." *Int. J. Climatol.*, 19(15), 1651–1673.

Richard, Y., Trzaska, S., Roucou, P., and Rouault, M. (2000). "Modification of the southern African rainfall variability/ENSO relationship since the late 1960s." *Clim. Dyn.*, 16(12), 883–895.

Shen, S. P. S., Lau, W. K. M., Kim, K. Y., and Li, G. (2001). "A canonical ensemble correlation prediction model for seasonal precipitation anomaly." *Technical memorandum, NASA/TM-2001-209989*, Greenbelt, Md.

Torrence, C., and Compo, G. P. (1998). "A practical guide to wavelets analysis." *Bull. Am. Meteorol. Soc.*, 79(1), 61–78.

Tyson, P. D., Dyer, T. G. J., and Mametse, M. N. (1975). "Secular changes in south African rainfall: 1880 to 1972." *Q. J. R. Meteorol. Soc.*, 101(430), 817–833.

Wilks, D. S. (1995). *Statistical methods in the atmospheric sciences*, Academic, San Diego.