

A NEW ANALYSIS OF VARIABILITY AND PREDICTABILITY OF SEASONAL RAINFALL OF CENTRAL SOUTHERN AFRICA FOR 1950–94

DAVISON MWALE,^a THIAN YEW GAN^{a,*} and SAMUEL S. P. SHEN^b

^a *Department of Civil Engineering, University of Alberta, Edmonton, Alberta, Canada*

^b *Department of Mathematics, University of Alberta, Edmonton, Alberta, Canada*

Received 15 June 2003

Revised 28 April 2004

Accepted 28 April 2004

ABSTRACT

Using wavelet analysis and wavelet-based empirical orthogonal function analysis on scale-averaged-wavelet power and individual scale power, we identified the non-stationary sea-surface temperature (SST) fields of the South Atlantic and Indian Oceans that are associated with coherent regions of rainfall variability in central southern Africa (CSA). The dominant mode of CSA rainfall is out of phase between the coastal areas and the centre of CSA and has been decreasing consistently since 1970. The frequencies associated with this mode are between 2–2.4 and 5.6–8 years. The Benguela ocean current SSTs form the dominant spatial pattern of the South Atlantic Ocean, and the Brazil and Guinea ocean current SSTs form the second leading mode. The Benguela spatial patterns were found to migrate seasonally between Africa's west coast and South America's east coast. The northern Indian Ocean SST forms the leading mode of variability, followed by the south Indian Ocean SST. Using predictor fields identified from both oceans, we achieved encouraging results of predicted CSA rainfall using a non-linear statistical teleconnection artificial neural network–genetic algorithm model. At 3 month lead time, correlations of between 0.8 and 0.9, root-mean-square errors of between 0.4 and 0.9 and Hansen Kuipers skill scores of between 0.4 and 0.8 were obtained between observed and predicted CSA rainfall. Copyright © 2004 Royal Meteorological Society.

KEY WORDS: wavelet empirical orthogonal function analysis; scale-averaged wavelet power; seasonal rainfall; central southern Africa; sea surface temperature; Indian Ocean; Atlantic Ocean; genetic algorithm–neural network

1. INTRODUCTION

Wavelet transformation, a process of decomposing a time series into frequency and energy components over time, is now an established analysis tool for non-stationary processes. Wavelet transformations have been successfully applied to climate characteristics analysis, such as streamflow characterization (Smith *et al.*, 1998), temporal structure of the southern oscillation (Wang and Wang, 1996) 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 of climate elements in space and time (Smith *et al.*, 1998).

Among the physical elements of our climate system, rainfall is very complex and exhibits considerable spatial, temporal and frequency variability (Basalirwa, 1995; Shen *et al.*, 2001). Problems of quantifying rainfall variability with respect to climatic elements, such as sea-surface temperature (SST), have been a subject of numerous studies (e.g. Mason, 1995; Mutai *et al.*, 1998; Reason and Mulenga, 1999). Many past studies have applied a combination of time domain multivariate statistics, such as harmonic analysis, Fourier analysis, linear regressions and cross-correlations, to climate data to determine relationships between and

* Correspondence to: Thian Yew Gan, Department of Civil and Environmental Engineering, University of Alberta, Edmonton, Alberta T6G 2G7, Canada; e-mail: tgan@ualberta.ca

among the climate elements. A common and well-recognized shortcoming of these approaches has been the assumption that climate data are stationary and linear, criteria that few data sets from natural phenomena satisfy (Huang *et al.*, 1998).

With an unstable relationship between rainfall in southern Africa and SST in the surrounding oceans in recent years (Landman and Mason, 1999) and a consistent pattern of declining runoff of most rivers in southern Africa after 1970 (Fanta *et al.*, 2001), one of the major challenges for climatologists and hydrologists has been to predict the nature of the variability in space and time.

Recent research has made progress in providing evidence of relationships between rainfall variability and a number of atmospheric and oceanic variables for parts of southern Africa (e.g. Mason, 1995; Jury, 1996; Reason and Mulenga, 1999). From developments in long-range forecast models utilizing SST and other predictors, one would have hoped that these developments would offer countries within the southern African region the prospect of managing climate variability rather than always being surprised victims of unexpected extreme events, such as droughts (e.g. Barnston *et al.*, 1996; Landman and Tennant, 2000; Landman *et al.*, 2001). However, there are still difficulties associated with rainfall prediction at seasonal lead time (Mason, 1997), and this seems more so in recent times. For example, the Indian Meteorological Department used models based on empirical relationships between monsoon and worldwide climate predictors with moderate success in the past (Webster *et al.*, 2002), but forecast skill after the 1980s has been low (Mason, 1997). Similarly, it seems that the skill for forecasting the frequencies of North Atlantic storms also faced difficulties after 1989 (Mason, 1997).

Decreased predictability has been associated with a combination of non-stationarity (interannual and interdecadal scales) of the ocean–atmosphere system, non-linearity of the ocean–atmosphere interaction and model inadequacies (Allan *et al.*, 1995; Webster *et al.*, 2002). In other words, empirical forecasts have to contend with the spectre of statistical non-stationarity (Webster *et al.*, 2002) in understanding climate variability and identifying relationships between climate elements to avoid driving predictor models with ambiguous input data that result in compromised prediction skill (Jury and Engert, 1999; Shen *et al.*, 2001).

Like the rest of southern Africa, rainfall in central southern Africa (CSA; 10–20°S, 12–42°E), Figure 1, occurs during Southern Hemisphere summer, which commences in October and ends in March. CSA has also suffered from repeated droughts in recent years (e.g. BBC, 2002a,b). The sudden outburst of droughts between 1991 and 2002 compared with periods prior to 1991 is a good example of the non-stationarity of the CSA climate.

A few possible ways to overcome the problem of non-stationarity include wavelet analysis and wavelet-based empirical orthogonal function (WEOF) analysis, empirical mode decomposition with Hilbert spectral analysis (EMD-HSA), and EMD-HSA-based EOF analysis. Here, we use wavelet analysis and WEOF analysis to study the spatial, temporal and frequency regimes of CSA rainfall, Indian and South Atlantic Ocean SST variability, and their associations. WEOF analysis has previously been used in chemical process monitoring (e.g. Bakshi, 1989), but, as far as we know, this is the first time WEOF analysis has been used to analyse the variability of rainfall and SST in the surrounding oceans. The specific research objectives of this study are as follows:

1. To identify and analyse the dominant, non-stationary, spatial, temporal and frequency regimes of the CSA rainfall and SST fields of the South Atlantic and the Indian Ocean SST, and, by identifying SST fields of both oceans relevant to CSA rainfall, their possible association.
2. Use the relevant predictor fields obtained from the foregoing to drive a non-linear, artificial neural network calibrated by a genetic algorithm (ANN–GA) to predict CSA rainfall at seasonal time scales (i.e. 3 month lead time).

By combining the analysis techniques of wavelet, WEOF, and the statistical forecasting method, ANN–GA, we expect to understand the non-stationarity of CSA rainfall and SST better, and also to improve the seasonal prediction of CSA rainfall using statistical methods. The paper is organized as follows: Section 2 discusses the data and the methodology of analysis used in this paper. Section 3 discusses SST and CSA rainfall variability,

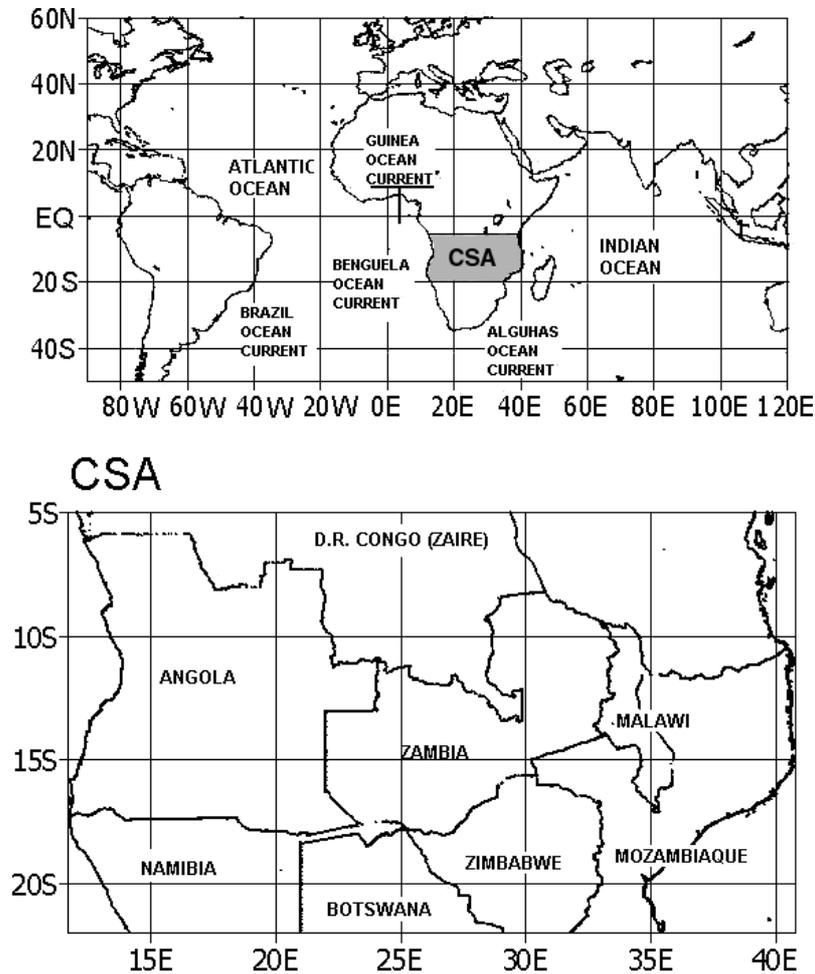


Figure 1. Map showing the locations and countries of CSA, the Indian and Atlantic Oceans, and prevalent ocean currents

and Section 4 establishes the associations between rainfall and SST. Section 5 examines the influence of El Niño–southern oscillation (ENSO) on CSA rainfall, and Section 6 presents the ANN–GA model. Section 7 presents the performance of the ANN–GA model in predicting CSA rainfall. The summary and conclusion are presented in Section 8.

2. DATA AND RESEARCH METHODOLOGY

2.1. CSA rainfall

Monthly rainfall data (1950–94) from 31 grid locations at a resolution of $2.5^\circ \times 3.75^\circ$ latitude and longitude was extracted for CSA. The rainfall data are part of a monthly precipitation dataset for global land areas from 1900 to 1998 provided by the UK Meteorological Office. The dataset was constructed using the Thiessen polygon approach. The quality control of the data is described in Hulme (1994). Locations (15°S , 18.75°E and 10°S , 18.75°E) in Angola had missing data and were excluded from the analysis.

2.2. SST of Indian and Atlantic Oceans

Monthly SST anomaly grid data at $5^\circ \times 5^\circ$ latitude and longitude resolution was extracted from 290 grid locations in the Indian Ocean (20°N – 40°S , 40 – 105°E) and Atlantic Ocean (10°N – 40°S , 55°W – 10°E); see Figure 1. The SST datasets were 48 years long (1950–97) and were transformed into seasonal and annual data by computing 3 month (i.e. January–March (JFM), April–June (AMJ), July–September (JAS) and October–December (OND)) and 12 month averages respectively. The SST dataset is part of MOHSST6, a historical global dataset of mean monthly global SST anomalies with respect to the 1961–90 normals provided by the UK Meteorological Office.

2.3. Wavelet analysis

A brief outline of wavelet transformation is given herein (see Torrence and Compo (1998) for details). 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 is designed to oscillate as a wave and required to span an area that sums to zero, and to die out rapidly to zero as t tends to infinity in order to satisfy the ‘admissibility’ condition:

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

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

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

where the scale (width) of the wavelet and translated position along the t -axis (usually the x -axis in the x - y plane) are a and b respectively. When a is increased, the wavelet width increases and a convolution of a time series with the wavelet isolates the low-frequency part of the time series. Conversely, if a is decreased, the wavelet width decreases and the high-frequency components of the time series can be isolated. This means that, if the scale is continuously varied along the translation b , a picture can be constructed depicting how the isolated components of the time series at each frequency vary with the time. Associated with each frequency are numerical coefficients referred to as the energy of the wavelet spectrum, which represent how well the wavelet matches with the time series. The parameters a and b in Equation (2) are real and a , always positive, may range over a continuous or a discrete set. The quantity $a^{-1/2}$ in Equation (2) is an energy normalization term, which ensures that the energies of the mother and daughter wavelets remain the same over all scales, making it possible to compare wavelet transforms of one time series with another directly (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

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

where Ψ^* is the complex conjugate of Ψ . In Equation (3), $W(b, a)$ is a wavelet spectrum, a matrix of energy coefficients of the decomposed time series $X(t)$. The energy coefficients also represent the magnitude of variance (variability) of coefficients at each scale a and location in time 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

$$W_I(a) = \sum_{k=0}^T \widehat{X} \widehat{\Psi}^*(s\omega_k) e^{i\omega_k n \delta_t} \quad (4)$$

where the caret symbolizes Fourier transform, k is the frequency index ($0, \dots, T$) and $\Psi(s\omega_k)$ is the Fourier transform of the wavelet function. The wavelet spectrum was computed using discrete scales in fractional powers of two:

$$s_j = s_0 2^{j\delta_j} \quad (5)$$

where s_0 is twice the sampling rate, $j = (0, 1, \dots, 20)$, and δ_j is the step size (e.g. 0.25). The wavelet transform of a time series contains a wealth of information at each time scale. This information can also be condensed over a range of scales or time in order to be conveniently used 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:

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

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

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

where C_δ is the reconstruction factor that takes on values depending on the mother wavelet used, δ_j is a factor for scale averaging, j_1 and j_2 are scales over which the averaging takes place, and δ_t is the sampling period (Torrence and Compo, 1998). The global wavelet spectrum depicts dominant oscillations present in a time series, and the local wavelet power shows how the dominant oscillations vary with time. 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. It is made up of a harmonic wave modulated by a Gaussian envelope:

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

2.4. WEOF analysis

Empirical orthogonal function (EOF) analysis of raw data fields has been widely used for the analysis of spatial and temporal variability of physical fields to identify objectively the spatially uncorrelated modes of variability of a given field (Kutzbach, 1967; Mason, 1995; Venegas *et al.*, 1997). In this study, EOF analysis is applied to the SAWP. To distinguish the EOF of raw data from that of the SAWP, the latter is called a WEOF and its corresponding principal components (PCs) are herein referred to as wavelet PCs (WPCs) to distinguish them from raw data PCs. The time-domain WPCs are obtained by projecting the SAWP onto the SAWP eigenvectors. Since the WPCs are obtained from the SAWP, they are interpreted as ‘frequency compacted’ energy variability (Park and Mann, 2000).

To identify and delineate temporal and spatially uncorrelated patterns in the time–frequency plane with coherent variations at regional scale, we applied WEOF analysis to the SAWP and the individual scale power (ISP) of the SST of the South Atlantic and Indian Ocean and CSA rainfall. Other techniques can be used in place of WEOF analysis to meet this objective, such as cluster analysis (Wilks, 1995), harmonic analysis (Ntale *et al.*, 2003), etc. However, the skill of the above methods is often difficult to evaluate statistically (Basalirwa, 1995).

3. VARIABILITY OF CSA RAINFALL AND SST IN THE INDIAN AND ATLANTIC OCEANS

3.1. Dominant modes of SST and rainfall variability

Local and global wavelet spectra were constructed for some SST locations in the Indian and South Atlantic Oceans and rainfall in CSA (see Figure 2). The dark shading in the wavelet spectra represents

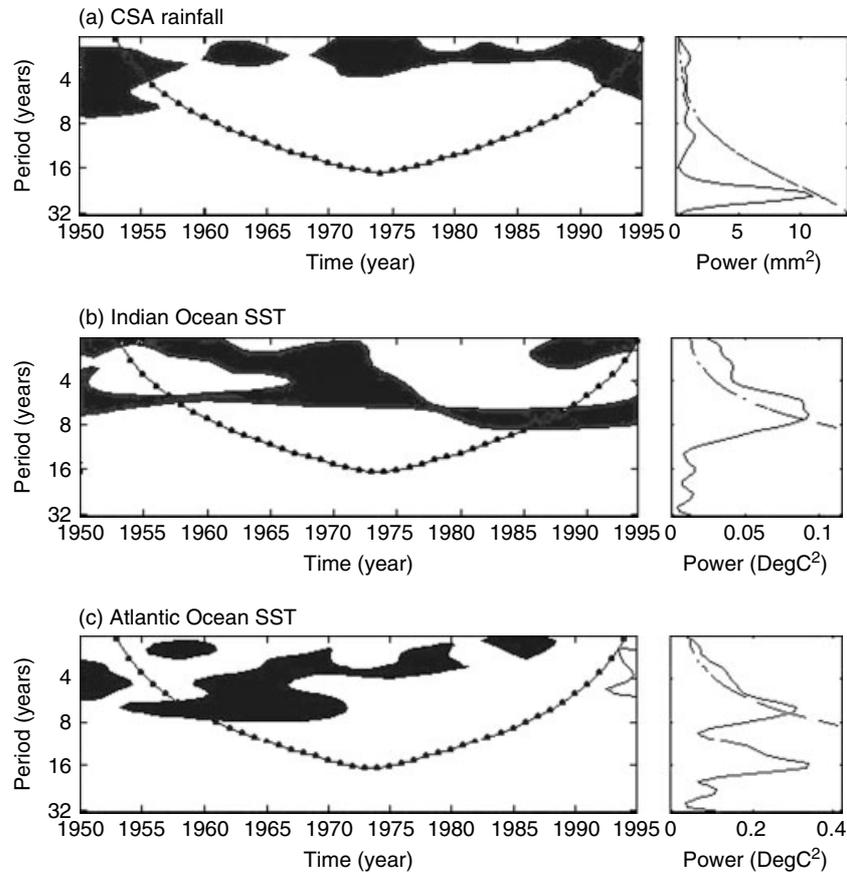


Figure 2. Examples of local and global wavelet spectra constructed for (a) the CSA rainfall, (b) the Indian Ocean, and (c) the Atlantic Ocean SST. The variation of power is suppressed at the beginning and end of each wavelet spectrum, since zeros used to pad both ends of the time series decrease the power. These affected areas are located outside the line with dots, which represents the COI. The dark shading in the wavelet spectra represents areas that contain power significant at the 95% confidence level of a red-noise process. In the global wavelet spectrum, the dotted curve represents the 95% confidence level, and peaks above the curve are considered statistically significant

areas that contain power (or energy) significant at the 95% confidence level of a red-noise process. The line with dots through the wavelet spectra delineates the cone of influence (COI). Since the length of the data used is finite, the ends of the time series were padded with zeros to bring the total length of the time series to the next-higher power of two, e.g. 512, 1024, etc. This facilitated the computation of wavelet power at larger scales and also speeded up the computation of wavelet transformation. However, padding the ends of time series with zeros introduces discontinuities at the endpoints of the time series and, as one goes towards larger scales, the amplitude near the edges decreases as more zeros enter the analysis. Therefore, at the beginning and end of the SAWP time series, the variation of power is generally suppressed.

There is appreciable power in the CSA rainfall and SST data at both high and low frequencies (Figure 2). However, the variation of rainfall and SST data is clearly dominated by peaks whose periods were between 2 and 8 years, i.e. the ENSO band. The concentration of energy within these periods is clearly visible as statistically significant peaks in the global wavelet spectra. Hence, for this study, the periods of between 2 and 8 years were chosen for the analysis of rainfall and SST variability.

3.2. CSA rainfall variability

We applied WEOF analysis to the SAWP of the 31 rainfall grid stations and retained two leading WPCs, which individually explained 24% and 17% of the total SAWP variance. The WEOF analysis was based on the correlation matrix, and the spatial distributions of the WPCs are shown in form of the correlation coefficient between WPCs and the 31 SAWP time series (see Figure 3). The leading spatial patterns show the regional variation of the SAWP. The discarded WPCs mainly described the SAWP variations of local features (e.g. the Zambezi River basin and the Lake Malawi basin). Since this paper analyses the regional variation of the SAWP and its relationship to SST, the third and higher WPCs are not discussed further.

WPC1 displays an out-of-phase relationship between the central CSA and the coastal regions along the east and west coasts. Statistically significant negative correlations between WPC1 and the SAWP occur along the coastal regions of Angola. Large positive correlations are evident over all of Zambia, Malawi, northern Zimbabwe and parts of Mozambique. WPC2 displays an out-of-phase relationship that extends diagonally from the northwest to the southeast of CSA. Statistically significant correlations occur in northeast Zimbabwe and central Mozambique.

The temporal variability of the WPCs is shown in Figure 3(c). The variance of WPC1 was high between 1950 and 1980, but decreased significantly between 1981 and 1994. Since WPC1 is positively correlated to central CSA (Zambia, Malawi, Zimbabwe and northern Mozambique), rainfall in these regions has been on the decline for over three decades (i.e. 1970–94). These results are consistent with findings of Fanta *et al.* (2001), who found that streamflow runoff consistently declined between the 1970s and 1997. Hence, the time-domain WPCs accurately represent the temporal variability of CSA rainfall. Unlike WPC1, WPC2 exhibited higher variations throughout the 1950–94 period and shows the diagonal north–south variation of CSA rainfall.

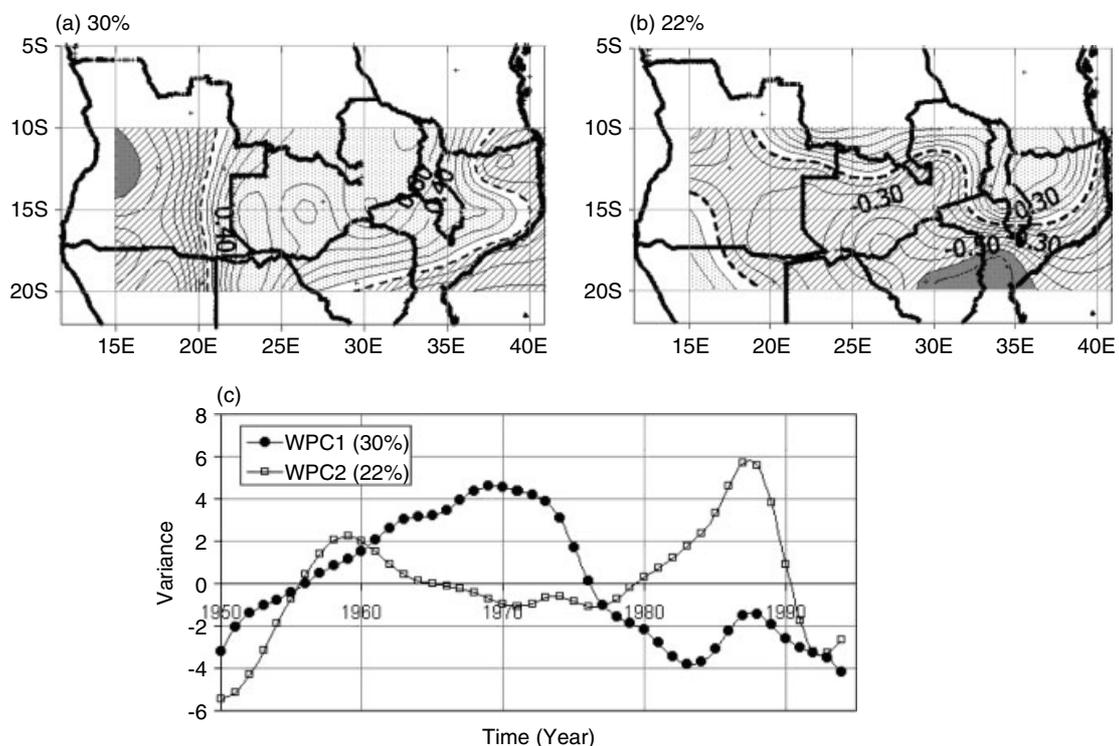


Figure 3. Contour plots of the spatial correlation patterns between (a) WPC1 and (b) WPC2 of the CSA rainfall and the SAWP of individual grids at 0.1 contour intervals. The numbers shown above represent the percentage of the total variance explained by each WPC. The dark areas correspond to correlations significant at the 95% confidence level. The temporal variations of the two WPCs of CSA rainfall are shown in (c)

3.3. South Atlantic Ocean SST variability

At the annual time scale, two WPCs accounting for 28% and 22% each of the total SAWP variance were retained, whereas at the seasonal time scale only WPC1 was retained for analysis. Each seasonal WPC1 explained 27% for JFM, 28% for AMJ, 27% for JAS and 28% for OND. The variability patterns of these WPCs are shown in Figures 4 and 5. At the annual time scale, the largest WPC variations (WPC1) are associated with the Benguela ocean SST, followed by variations of Brazil ocean SST and Guinea ocean current SST (WPC2). The annual WPC1 SST spatial variability patterns are similar to the patterns found by Venegas *et al.* (1997) and Houghton and Troure (1992).

At the seasonal time scale, the SST SAWP spatial variation patterns are non-stationary in space throughout the year. In the OND season, the largest SAWP variations (with explained variance of up to 64%) are located along Africa's west coast. The Benguela ocean current SST and parts of Brazil ocean SST form the dominant features of spatial variability. During the JFM season, the spatial variability patterns shift westwards. Maximum variations also shift and intensify along the equator and at about 20°S and 10°W. During this period, the Benguela ocean current SST appears as the dominant feature of spatial variability. By the AMJ season, the variations cover the entire east coast of South America from 10°N to about 40°S, with maximum loadings (up to 64%) along the northern Brazil coast. All the three ocean current SSTs form the dominant spatial

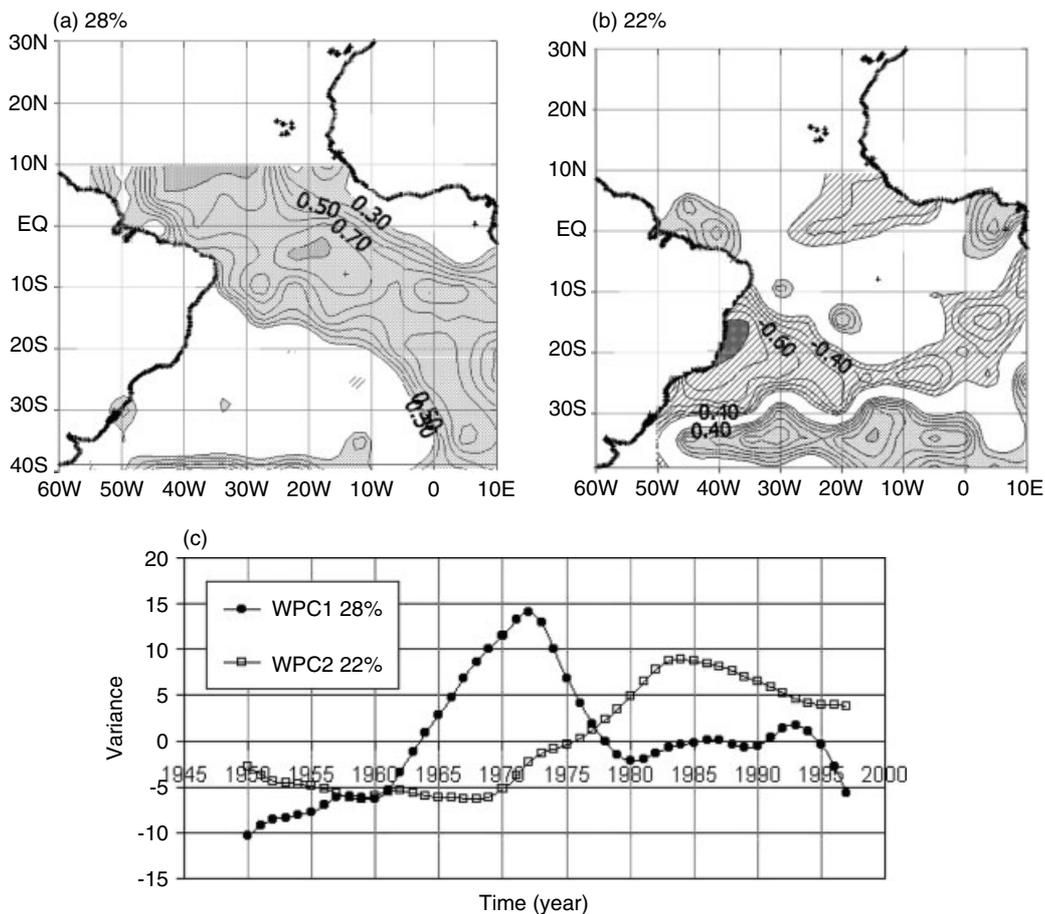


Figure 4. Contour plots of the spatial correlation patterns between (a) WPC1 and (b) WPC2 of annual Atlantic Ocean SST and the SAWP of individual grids at 0.1 contour intervals. The numbers shown above represent the percentage of the total variance explained by each WPC. The dark areas correspond to correlations significant at the 95% confidence level. The temporal variations of the two WPCs of Atlantic Ocean SST are shown in (c)

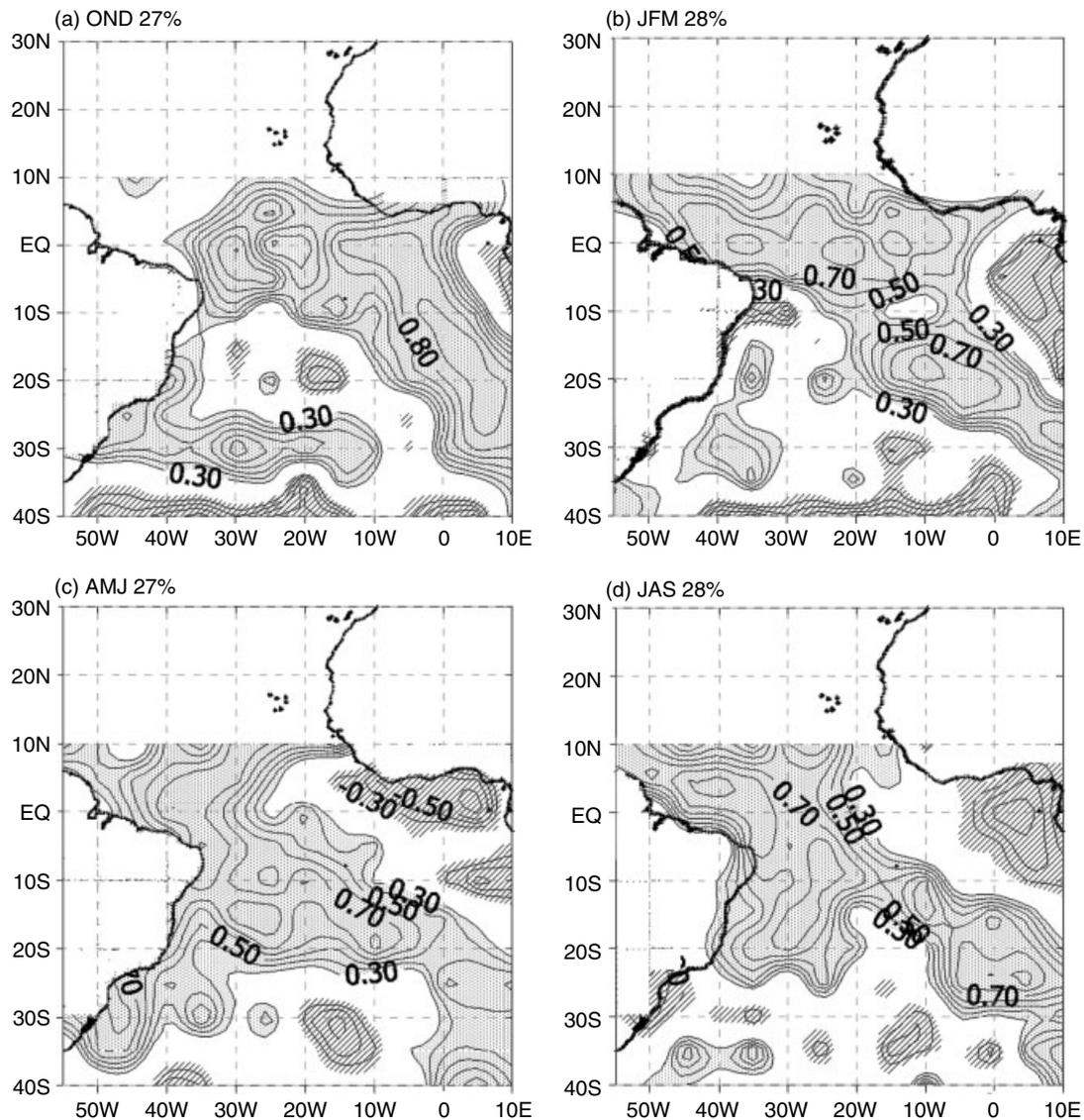


Figure 5. Contour plots of the spatial correlation patterns between WPC1 of seasonal Atlantic Ocean SST and the SAWP of individual grids at 0.1 contour intervals for (a) OND, (b) JFM, (c) AMJ, and (d) JAS seasons. The numbers shown above represent the percentage of the total variance explained by each WPC1. The seasonal migration of spatial variability patterns of SST can be seen from (a) to (d)

feature of variability. However, the variability of the Guinea ocean current SST is out of phase with the rest of the Atlantic Ocean SST. In the JAS season, the spatial variability patterns begin to shift eastwards towards the African coastal areas, with the Benguela as the sole pattern of spatial variability.

The migration of the SAWP spatial variation patterns between South America and Africa is a new finding that has important implications for understanding both the lagged and simultaneous relationships between SST in the South Atlantic Ocean and rainfall in South America and CSA, and probably most of sub-Saharan Africa.

The time-domain WPC for the annual and seasonal SAWP shows large variation between 1950 and 1980, followed by a relatively quiet period between 1980 and 1994, similar to the CSA WPC1 (see Figures 4 and 6). SST energy increased (temperature rise) between 1950 and 1972 and was followed by a decrease between 1972 and 1980. WPC2 (less variable than WPC1) shows increasing power from 1969 to 1984, with a decrease that started in 1985.

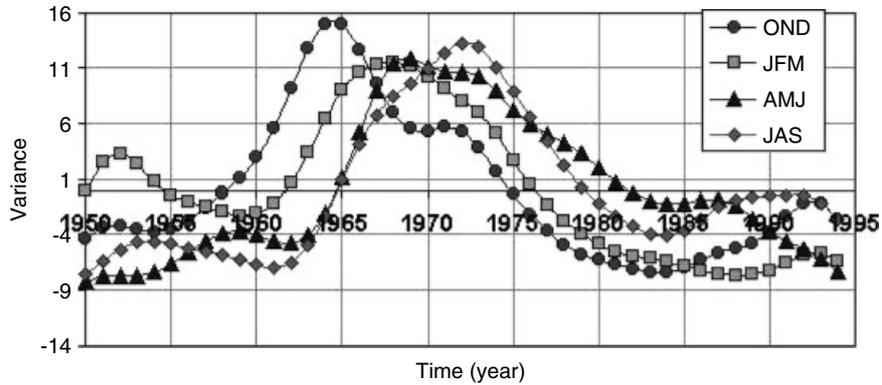


Figure 6. The temporal variations of the WPC1 of the four seasonal Atlantic Ocean SSTs shown in Figure 5

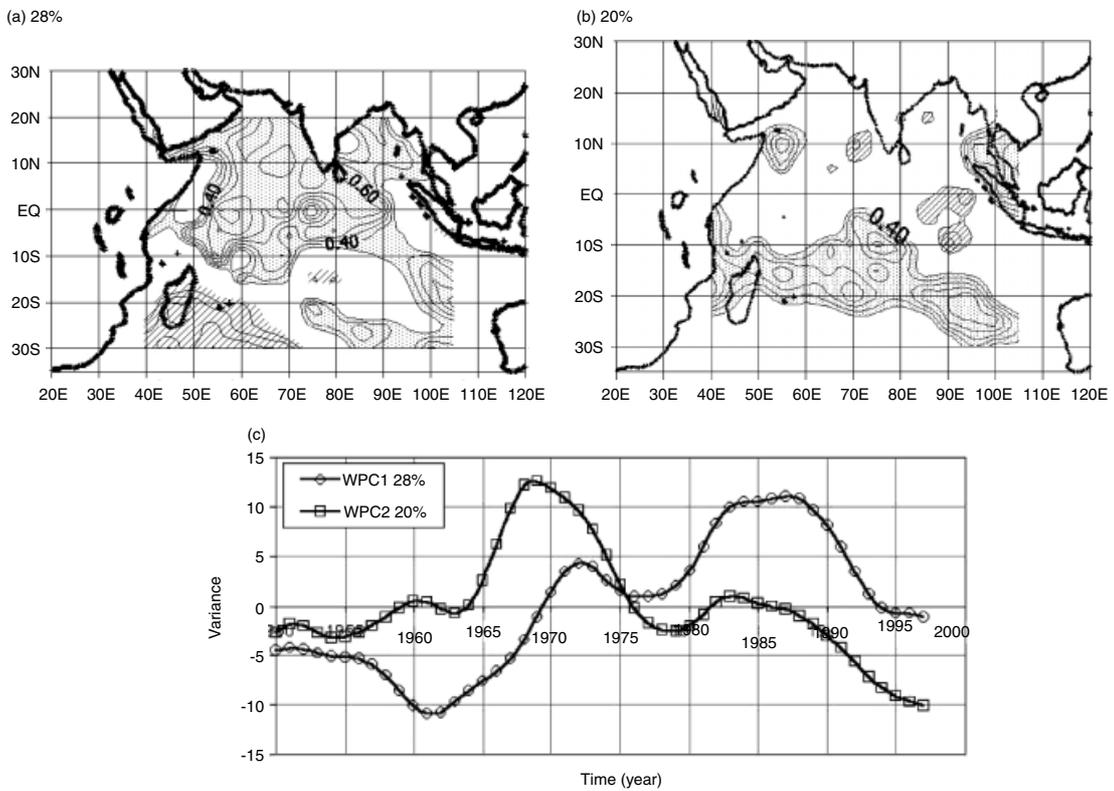


Figure 7. Contour plots of the spatial correlation patterns between (a) WPC1 and (b) WPC2 of annual Indian Ocean SST and the SAWP of individual grids at 0.1 contour intervals. The numbers shown above represent the percentage of the total variance explained by each WPC. The temporal variations of the two WPCs of the Indian Ocean SST are shown in (c)

3.4. Indian Ocean SST variability

WEOF analysis of the annual SAWP of the Indian Ocean SST revealed two leading WPCs, which explained 28% and 20% of the total SAWP variance. Their spatial variability patterns are shown in Figure 7. At the seasonal time scale, WPC1 explained 26% for the OND season, 35% for the JFM season, 23% for the AMJ season and 29% for the JAS season. Unlike for the Atlantic Ocean, the Indian Ocean seasonal patterns are stationary. However, as for the Atlantic Ocean SST, the WPC2 for each season was not spatially extensive

and its spatial correlation patterns explained very little variance. Thus, the WPC2 at the seasonal time scale is not discussed.

WPC1 describes variation of SAWP of the northern Indian Ocean SST. It is positively correlated to the SAWP of the northern Indian Ocean and negatively, but weakly, correlated to the southern Indian Ocean. Maximum variations of between 50 and 64% are accounted for by this WPC in the central and western Indian Ocean. WPC2 describes the remainder of the variance in the south Indian Ocean not accounted for by WPC1. It is positively correlated to the south Indian Ocean (extending from 40 to 105°E) and negatively correlated but weakly to the northern Indian Ocean. This WPC appears as a part of Mason's (1995) PC2 (13%), which he attributed to warming and cooling of the Mozambique Channel current and Agulhas system.

At seasonal time scales, except during the AMJ season, the northern Indian Ocean also appeared as the main mode of spatial variability throughout the year (see Figure 8). In addition, only the OND and JFM time-domain WPC1 appear similar to their Atlantic Ocean counterparts (see Figure 9), but with a 5 year lag.

4. ASSOCIATIONS BETWEEN CSA RAINFALL AND THE INDIAN AND SOUTH ATLANTIC OCEAN SST

The objective of this section is to establish the associations between rainfall variability in CSA and SST variability in the Indian and Atlantic Oceans between 1950 and 1994, by linearly correlating WPCs of CSA rainfall and the SAWP of the Indian and Atlantic Ocean SSTs. Instead of using untransformed (raw) rainfall or SST time series, which contain a wide range of frequencies, many of which are unknown, we used the

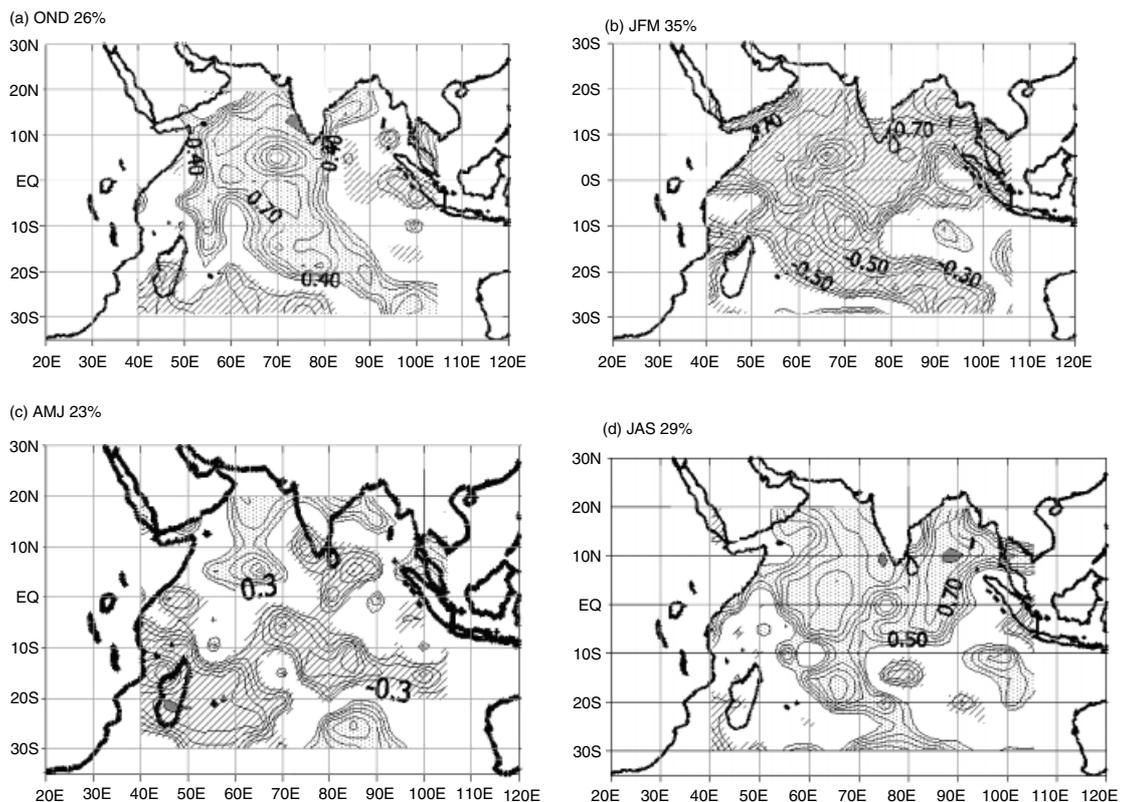


Figure 8. Contour plots of the spatial correlation patterns between WPC1 of seasonal Indian Ocean SST and the SAWP of individual grids at 0.1 contour intervals for (a) OND, (b) JFM, (c) AMJ, and (d) JAS seasons. The numbers shown above represent the percentage of the total variance explained by each WPC1

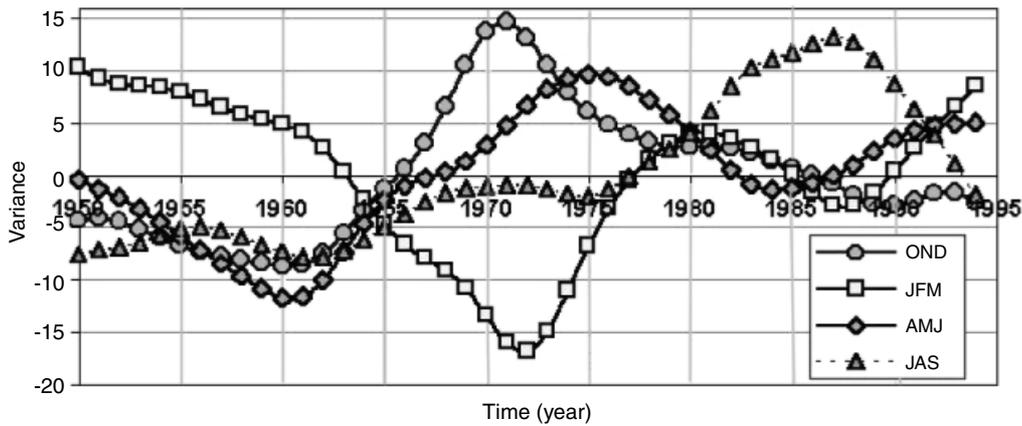


Figure 9. The temporal variations of the WPC1 of the four seasonal Indian Ocean SSTs shown in Figure 8

SAWP, which predominantly contains frequencies significant at the 95% level, to establish the presence of the associations. Owing to strong persistence in the SAWP and WPC time series, the effective length of the SAWP and WPCs N_{eff} was estimated using the approximation (Wilks, 1995)

$$N_{\text{eff}} = N \frac{1 - \rho_1}{1 + \rho_1} \quad (9)$$

where ρ_1 is the lag-1 autocorrelation and N is the total length of the time series. N_{eff} , which was much shorter than N , was then used to determine the 95% significance level of the correlation between WPCs and the SAWP.

Both lagged and simultaneous associations between CSA WPC1 and SST SAWP were examined. However, the SAWP of AMJ SST with a 3 month lead time was used to establish the associations with CSA rainfall that occurs in the summer, October to March. Figure 10 shows the correlations between CSA rainfall WPC1 and the individual $5^\circ \times 5^\circ$ SST SAWPs of the AMJ season in the South Atlantic and Indian Oceans. Strong positive correlations were found between the rainfall WPC1 and the Benguela and the Brazil ocean SST SAWPs, and negative correlations were found between the Guinea ocean current SST and the South Atlantic Ocean SST along 40°S , similar to WPC1 of the AMJ season (Figure 5(c)). Weak spatial correlation patterns were found between rainfall WPC1 and the AMJ Indian Ocean SST SAWP (Figure 10(b)). However, strong positive simultaneous correlations were found between WPC1 of CSA rainfall and the OND and JFM Indian Ocean SST SAWPs (Figure 11). The northern Indian Ocean has been found to be an important source of atmospheric moisture throughout the summer rainfall season and becomes the dominant source in the second half of the summer for most of southern Africa (Mason and Tyson, 2000). Figure 11(b) confirms these findings. An increase of SST in the same region has been observed to enhance rather than reduce rainfall over southern Africa (Hulme *et al.*, 1996; Mason and Tyson, 2001). Rainfall WPC2 and the seasonal Atlantic Ocean SST SAWP along the Benguela ocean current were found negatively but weakly correlated and not discussed further.

The associations between rainfall WPC1 and the SST variations in the Brazil, Benguela and Guinea ocean currents shows that variabilities in these ocean current SSTs have important climatological implications for CSA rainfall. Although, geographically, these ocean basins are located too far away to affect CSA rainfall, they nonetheless teleconnect to southern African rainfall through the displacements of westerly waves whose position is determined by shifts in the centre of the South American convective activity (Mason and Tyson, 2000). When the wave weakens and withdraws westwards, moist winds from the Indian and South Atlantic Oceans converge on the subcontinent, causing uplift and widespread rains in the southern African region. When the wave is displaced eastwards, convergence of winds takes place over Madagascar, causing below-normal rainfall or drought in southern Africa. Since the rainfall variability in central CSA is positively

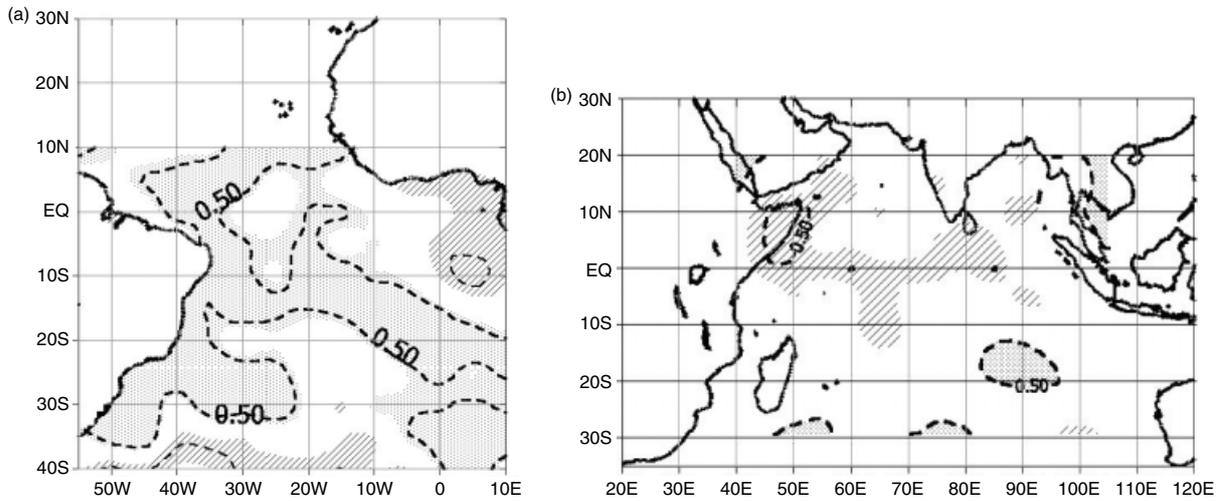


Figure 10. Plots showing the spatial correlation patterns between WPC1 of CSA rainfall and individual $5^{\circ} \times 5^{\circ}$ AMJ SST SAWP time series of the (a) Atlantic Ocean and (b) Indian Ocean. The areas inside the dotted line correspond to ocean zones with correlations greater than 0.5. Data from these delineated zones were later used to predict the CSA rainfall

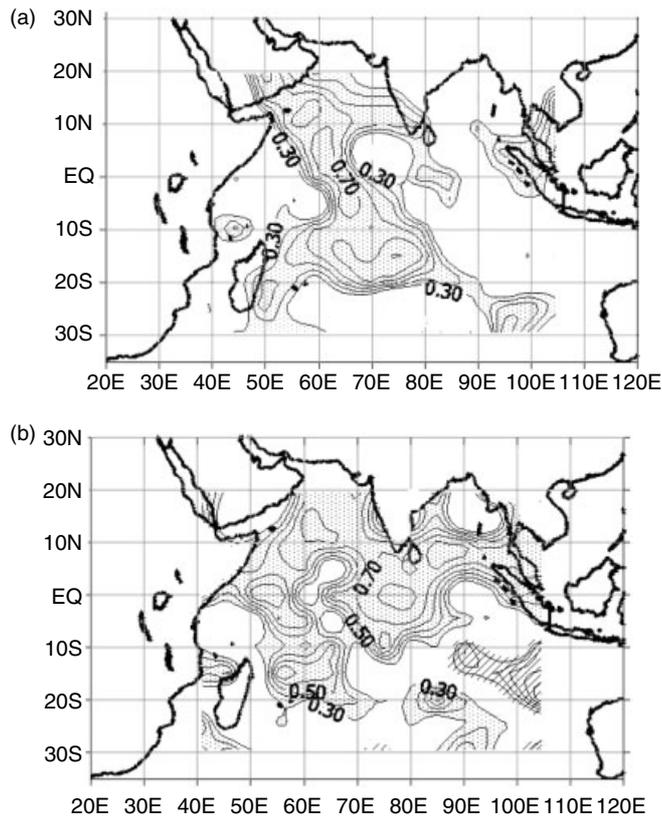


Figure 11. Contour plots at 0.1 contour intervals of the spatial correlation patterns between WPC1 of CSA rainfall and the SAWP of the Indian Ocean for (a) OND and (b) JFM seasons. During JFM, the latter half of the CSA rainy season, it seems that the dominant source of moisture supply to CSA's rainfall comes from the central Indian Ocean

associated with SST in the Brazil and Benguela ocean currents, warming in these ocean basins results in increased rainfall in central CSA and *vice versa* for the coastal areas.

5. INFLUENCE OF ENSO ON CSA RAINFALL VARIABILITY

In this section we examine how the ENSO signal affects the spatial and temporal variability of CSA rainfall. The wavelet and global spectra showed that most of the power of the rainfall and SST lies within the ENSO band (i.e. 2–8 year cycles band). In addition, past studies of rainfall of some countries in CSA (e.g. northern Namibia, Zimbabwe and Botswana) found a number of cycles in the rainfall in the 2–8 year range (i.e. 2, 2.3, 2.8, 3.5, 6.7 and 7 years) and others at 11 years and 18–20 years (e.g. Nicholson, 1986; Ropelewski and Halpert, 1987; Mason and Tyson, 2000). In this paper, the spatial and temporal variability of the power (or energy) at each of these cycles is examined against the temporal evolution of ENSO episodes between 1950 and 1994.

The spatial distribution patterns of the correlation between the ISP and the ISP WPCs are shown in Figure 12 for some periods (2, 2.4, 3.4 and 5.6 years). Figure 12 suggests that the variation of energy with periods at either end of the ENSO band (2–2.4 years and 5.6–8 years) is dominant over CSA compared with that in the middle (i.e. 3.4 years). The 3.4 year period explains most of the variance in the northern Namibia–southern Angola region.

The time-domain ISP WPCs (Figure 13) show that the energy variations within the 2–2.4 and 5.6–6.7 year periods generally correspond to that of the rainfall SAWP WPC1 of Figure 3. Correlations of 0.5 and 0.8 were found between the WPC1 of Figure 3 and the WPCs of the 2–2.4 year and 5.6–6.7 year periods respectively. This suggests that, in terms of correlation, more variation in SAWP is explained by the 5.6–6.7 year periods (64%), followed by those between 2 and 2.4 years (25%).

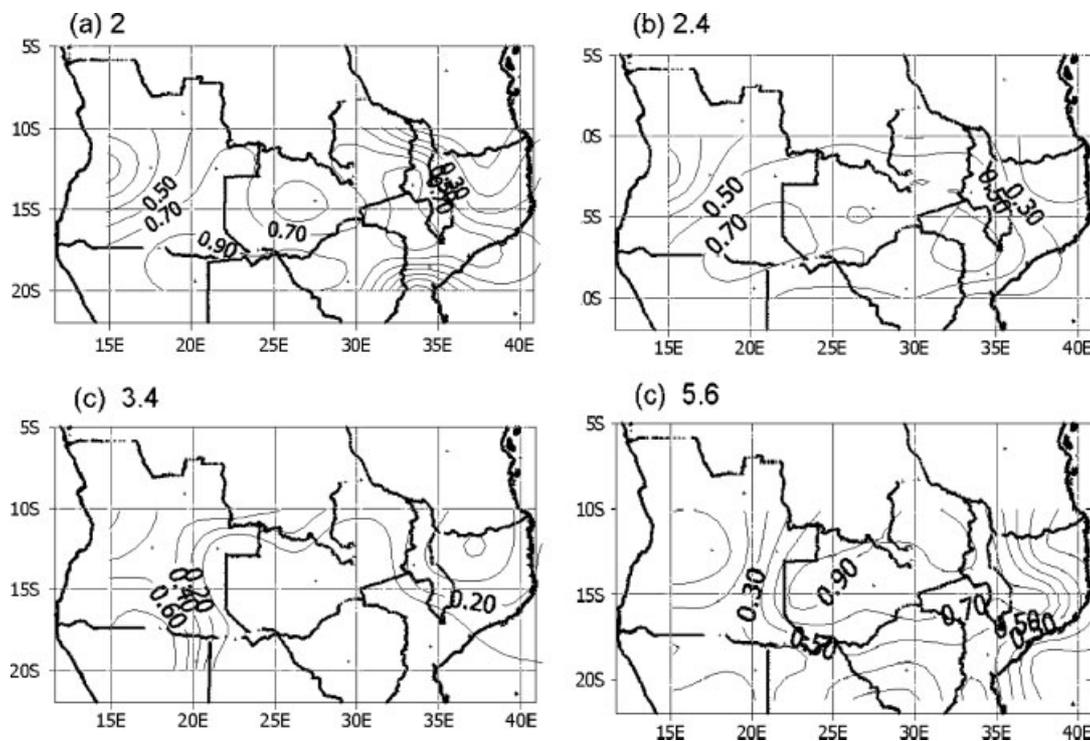


Figure 12. Contour plots of the correlation patterns at 0.2 contour intervals between WPC1 of ISP and ISP of CSA rainfall for some selected periods: (a) 2, (b) 2.4, (c) 3.4 and (d) 5.6 years

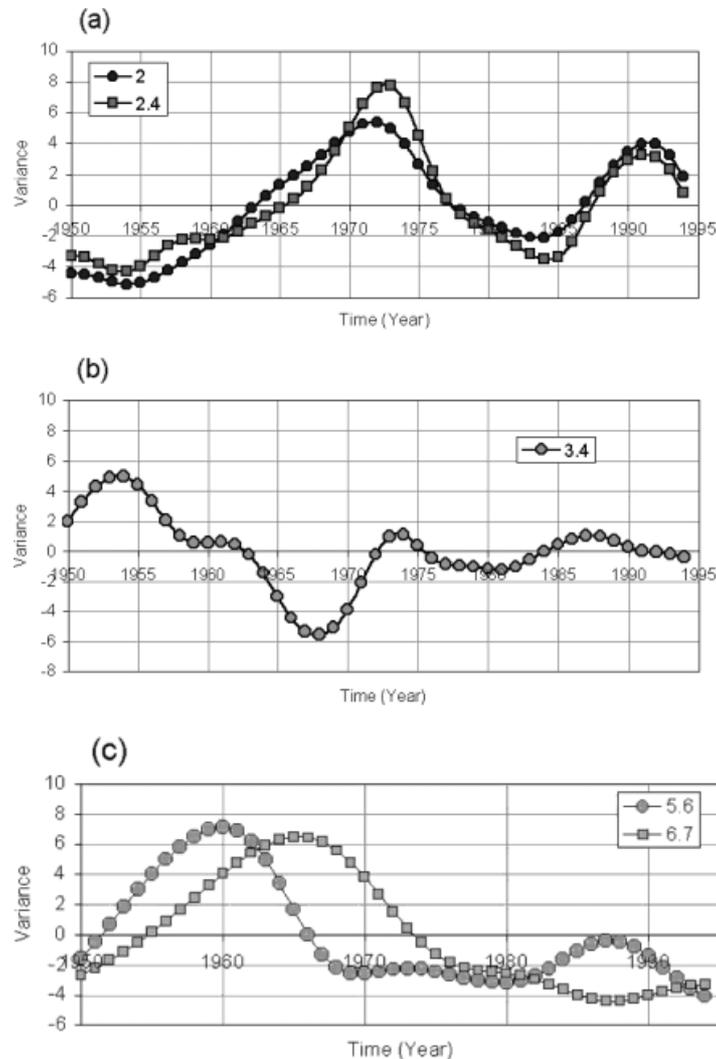


Figure 13. The temporal variations of ISP WPC1 of CSA rainfall at (a) periods of 2 and 2.4 years, (b) 3.4 years, and (c) 5.6 and 6.7 years. The effects of the six ENSO episodes between 1950 and 1970 were not felt in CSA rainfall as the energy at the 2 and 2.4 year periods increased, whereas between 1970 and 1985 and between 1991 and 1994 the ENSO episodes were accompanied by droughts in CSA

Of all the ENSO episodes between 1950 and 1994, six events occurred after 1970 (1972, 1977, 1982, 1986, 1992 and 1994) and six events occurred between 1950 and 1970 (1951, 1953, 1957, 1963, 1965 and 1969). A look at Figure 13(a) reveals that although the 2–2.4 year energy explains 25% of the SAWP, a consistent increase in the energy between 1955 and the early 1970s offset the effects of ENSO episodes during this period in CSA. Conversely, the dominance of the 2–2.4 year periods in the rainfall from the early 1970s to the mid 1980s exacerbated the effects of ENSO in 1972, 1977, 1982 and 1986. From the mid 1980s to 1990, the energy of the 2–2.4 year cycles increased again, and no drought was recorded during this period. However, from 1991 until the end of the data period in 1994, energy in the 2–2.4 year cycle decreased, amplifying the effects of the 1992 and 1994 ENSO events.

Richard *et al.* (2000) found that pre-1970 ENSO events had little effect on the southern Africa climate conditions and atmospheric circulation, whereas the ENSO events after 1970 were characterized by reduced rainfall. The time-domain ISP WPCs appear to show that when the 2–2.4 year energy is increasing, while

the 5.6–8 year energy is decreasing, ENSO has little effect on the CSA rainfall. However, it seems when the energy is decreasing at the above scales considered, ENSO causes drought in CSA. This preliminary information is important in predicting the effect of ENSO on CSA rainfall. However, longer datasets are desirable to investigate further the temporal variability and the interaction of energy at different frequencies to see how CSA rainfall responds to the effects of ENSO.

6. ANN-GA PREDICTION MODEL

The neural network used in this study has three layers (input, hidden and output layers) and three hidden neurons. A genetic algorithm (GA) was used to calibrate the parameters of the ANN-GA model (see Figure 14). GAs are global optimization techniques that create new solutions to a given problem by exploiting past performance of previous solutions based on a model of natural, biological evolution (Goldberg, 1989).

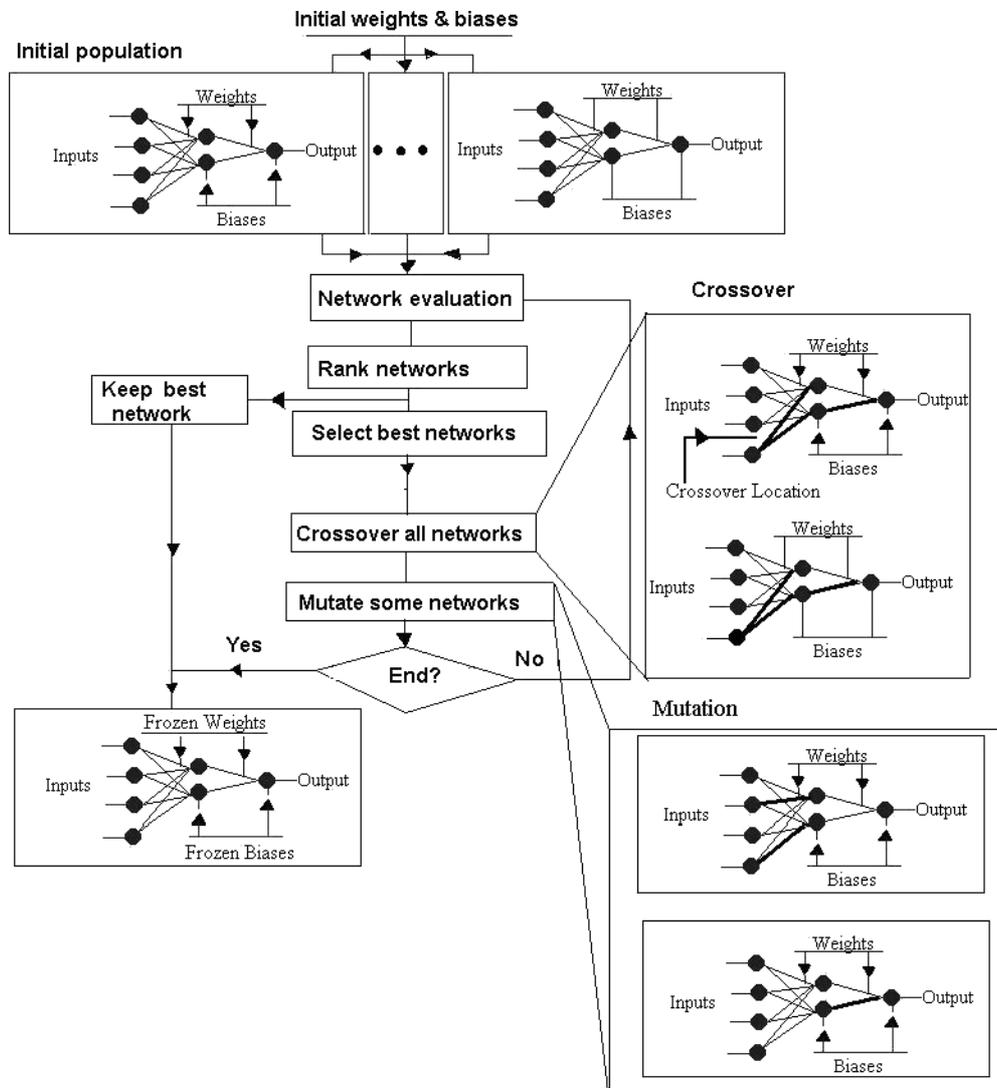


Figure 14. The flow chart of a three-layer (input, hidden, output), ANN-GA summarized in three key operations: selection, crossover and mutation

The solution space (or population) from which the individual solutions are drawn is represented in the form of finite lengths of strings called chromosomes. The adaptability of chromosomes is improved through a process of systematic modifications made up of crossover and mutation. For the ANN–GA model, the chromosomes were made of weights and biases assigned to each node. The weights and biases have traditionally been found using the back-propagation algorithm (Rumelhart *et al.*, 1986). In recent years, the back-propagation algorithm is less preferred compared with more efficient optimization algorithms, such as the conjugate gradient method, simulated annealing and GAs (Hsieh and Tang, 1998).

6.1. Population and initial evaluation

The ANN–GA creates 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) for a large number of neural networks (e.g. 2300 ANNs). Since the starting point of the initial weights and biases of an ANN is not known, the weights are created in random fashion to ensure that the initial population contains diverse information. In using an ANN–GA for prediction, each output of the neural network in the population is initially evaluated against a known predictand. The objective function (fitness) used was based on either the Pearson correlation or the root-mean-square error (RMSE). To evaluate each neural network, the predictand y is obtained as a nonlinear translation of the weighted average of the predictor data x , which has been normalized, i.e., $x = (x - \bar{x})/\sigma_x$:

$$\text{hidunit}_{pj} = \sum_{i=1}^N W_{ji}x_{pi} + B_{jo} \quad (10)$$

where hidunit_{pj} is the weighted input to the j th hidden unit, N is the total number of input nodes W_{ji} (weights from input unit i to j), B_{jo} are the biases for hidden neuron j , and x_{pi} is the i th input of pattern p (in our case SST PCs are used). The hidden layer undergoes a non-linear translation

$$f_1(\text{hidunit}_{pj}) = \frac{1}{1 + e^{-\text{hidunit}_{pj}}} \quad (11)$$

where $f_1(\text{hidunit}_{pj})$ is the j th neuron non-linear activation function, and

$$\text{out}_{pk} = \sum_{j=1}^M W_{kj} f_1(\text{hidunit}_{pj}) + B_{ko} \quad (12)$$

$$y_{pk} = f_2(\text{out}_{pk}) \quad (13)$$

where M is the number of hidden units, W_{kj} represents the weight connecting the hidden node j to the output k , B_{ko} is the bias for output neuron k , y_{pk} is the predicted output and f_2 is a linear function.

6.2. Ranking the neural networks

All the neural networks are ranked according to the computed fitness: the best network at the top and the worst at the bottom. The best 85% of the ranked population is randomly selected to comprise offspring of the next generation. Selecting from the best 85% of the ranked population ensures that, on average, the new generation has comparatively better fitness than the original population. Selection thus shifts the search space towards improved solution spaces of the problem. In this study, the population is kept constant; hence, some members of the old population are selected more than once.

6.3. Crossover

Pairs of neural networks are selected, either randomly or 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 one-point crossover scheme was used. In one-point crossover schemes, a single location is randomly chosen in the hidden layer and weights on either side of the location are exchanged between the two neural networks (shown as dark lines in the inset to Figure 14). This procedure is repeated between all other pairs of neural networks in the population.

6.4. Mutation

Next, a small percentage (e.g. between 0.1 and 1% of the total population) is randomly selected and then a handful of weights and biases are randomly replaced (see dark lines in the inset to Figure 14). The process is called mutation, and is designed to restore good weights and biases eliminated during selection. Since mutation is a purely random process, it is always kept to a minimum to prevent the search degenerating into a random process. If mutation results in a better neural network, then that network will likely survive in the next selection; but, if the mutation results in a more inferior individual, then that individual will likely perish in the next selection. In this paper, 1% of the population of neural networks is randomly chosen and a proportion of their weights is randomly mutated. The neural networks are once again evaluated against the same known predictand.

The above procedures are repeated through several epochs (or generations in GA terminology). At each generation, the best network is kept until a better solution is found in successive generations. Convergence is reached when at least 95% of the solutions 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.

7. PREDICTION OF CSA RAINFAL (1985–94)

7.1. Selection of SST predictor fields

The existence of 2–2.4 year periods found within the ENSO band (Section 5) and the strong seasonal associations between the SST SAWP of the South Atlantic and Indian Oceans and the CSA rainfall WPCs (Section 3) suggests that: (1) CSA rainfall is predictable at interannual scales, and (2) prediction of CSA rainfall based on the preceding AMJ SST of the two oceans is possible. From Figure 10(a) and (b), raw SST predictor data were extracted for the months of April, May and June from all regions of the oceans where the correlation between CSA WPC1 and the SAWP of the oceans exceeded 0.50. Since the autocorrelation of the SAWP and WPCs was high (average 0.9), from Equation (9), the effective lengths of the SAWP and the WPCs decreased to about 3–4 years and the significant correlation between CSA WPCs and SST SAWP at the 95% level was above 0.997. The ocean areas covered by this level of correlation were small, and very few data could be collected. Since the spatial correlation pattern between CSA WPC1 and CSA SAWP (Figure 3(a)) covering most of the region varied between 0.4 and 0.7, it was decided that actual SST data be collected from areas of the ocean where the correlation between CSA rainfall WPC1 and SST SAWP was 0.5 or more. The actual data were standardized and averaged over the 3 months for each grid station in the two oceans to give one AMJ data set. To speed up the computations, EOF analysis was applied to the raw AMJ SST dataset and six PCs accounting for 87% of the SST data variance were used as input data to the ANN–GA prediction model. Standardized rainfall data from the 31 CSA grid stations were used.

7.2. Evaluation of prediction skill

To assess the prediction skill of the ANN–GA for 11 years (i.e. between 1984 and 1994), the Pearson correlation, the Hansen Kuipers (HK) skill score and RMSE were used. The Pearson correlation coefficient

ρ is computed as

$$\rho = \frac{\sum_{k=1}^n (\text{obs}_k - \overline{\text{obs}})(\text{pred}_k - \overline{\text{pred}})}{\left[\left(\sum_{k=1}^n (\text{obs}_k - \overline{\text{obs}})^2 \right) \left(\sum_{k=1}^n (\text{pred}_k - \overline{\text{pred}})^2 \right) \right]^{1/2}} \quad (14)$$

where obs_k and pred_k are the observed and predicted values, $\overline{\text{obs}}$ and $\overline{\text{pred}}$ their respective means, and n is the sample size. This ρ varies between +1 and -1, with the maximum and minimum values indicating perfect positive and negative linear relationships respectively. To compute the HK skill score, the predicted and observed rainfall data are grouped into categories, say 'Dry', 'Near Normal' and 'Wet'. Tercile percentages of below 33%, 33–66% and above 66% were used to define the categories in a square contingency table.

$$\text{HK} = \frac{H - E_c}{T - E_m} \quad (15)$$

where H is the total number of correct forecasts, T is the total number of forecasts obtainable with a perfect forecast model, E_c is the number of correct hits expected by chance and E_m is the marginal number of correct (observation) hits expected by chance. For a $K \times K$ contingency table, the HK score may be expressed in terms of probabilities as

$$\text{HK} = \frac{\sum_{i=1}^K p(\text{obs}_i, \text{pred}_i) - \sum_{i=1}^K p(\text{obs}_i) \times p(\text{pred}_i)}{1 - \sum_{j=1}^K [p(\text{obs}_j)]^2} \quad (16)$$

The HK score values range from -1 to +1. Perfect forecasts receive a score of one, random forecast receive a score of zero and forecasts inferior to random forecasts receive negative scores. The RMSE is computed as

$$\text{RMSE} = \left(\frac{1}{n} \sum_{k=1}^n (\text{obs}_k - \text{pred}_k)^2 \right)^{1/2} \quad (17)$$

An RMSE of zero indicates a perfect prediction.

7.3. Skill of the predicted CSA rainfall

The spatial display of the correlation, HK skill scores and RMSEs between 10 years of predicted and observed CSA rainfall is shown in Figure 15. Correlations of between 0.8 and 0.9, RMSEs of 0.4 and 0.9 and HK skill scores of between 0.4 and 0.8 were achieved for most of CSA. The skill decreased slightly towards the coastal areas, following the strength of CSA rainfall WPC1.

The high correlations show that the linear correlation between the predicted and observed rainfall was captured almost everywhere in CSA. The higher HK skill score also showed that most forecasts fell in their correct categories. For example, Table I shows a categorical forecast for an area in central CSA. Out of 10 years, 8 years were correctly predicted by the ANN-GA. Table I also shows that most of the years between 1985 and 1994 were wet (i.e. 6 years were wet, and 2 years had near-normal rainfall, and 2 years were below-normal or dry) consistent with Figure 13(a), which shows that the period 1985–91 (i.e. 7 years) was generally wet and 1992–94 experienced below-normal rainfall (or droughts). A look at Table I shows that 2 years were wrongly predicted, one as wet and the other as near normal. Generally, areas

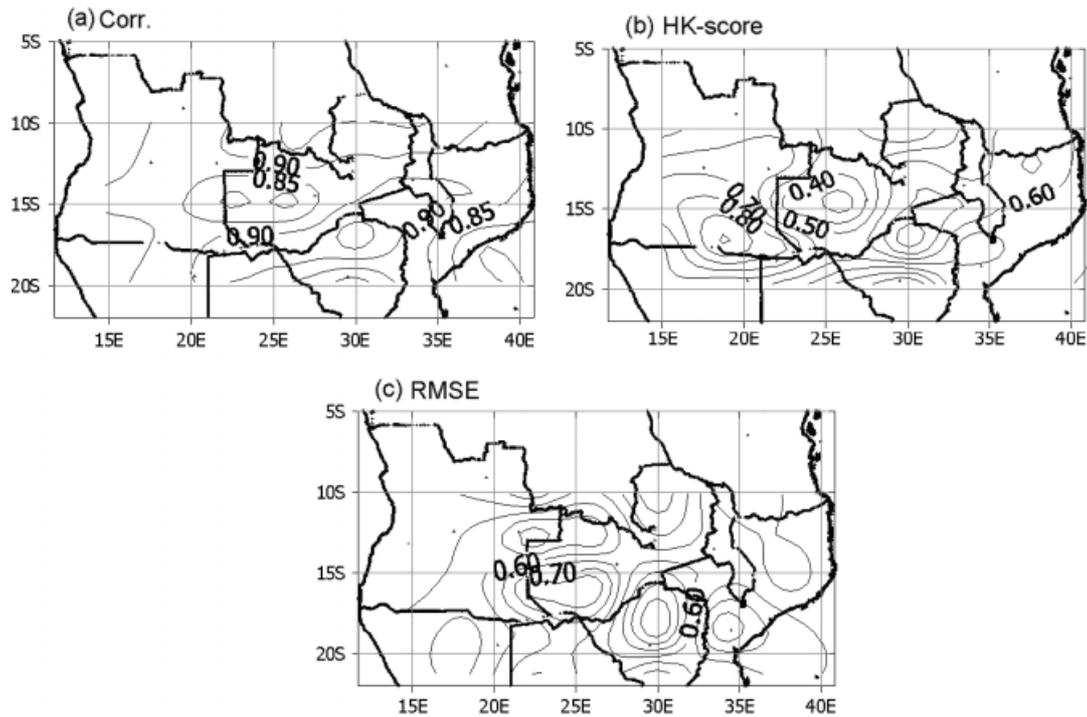


Figure 15. Contour plots of (a) correlation coefficient ρ , (b) HK skill scores and (c) RMSE between predicted and observed October–March CSA rainfall predicted by the ANN–GA model (at 3 month lead time) driven by selected windows of AMJ SST predictor fields taken from the Indian and Atlantic Oceans shown in Figure 10 where correlations were greater than 0.5

Table I. An example contingency table for a grid in central CSA with HK = 0.57

HK 0.57	Dry	Near normal	Wet
Dry	1	1	0
Near normal	0	1	1
Wet	0	0	6

with HK skill scores above 0.50 had between 7 and 8 years out of 10 years falling into correct categorical forecasts.

The results also show that the CSA rainfall variability is influenced by SST variations in the Atlantic and parts of the southern Indian Ocean. Overall, using sectors of predictor fields identified in the oceans, the non-linear ANN–GA system could predict between 64 and 81% of the rainfall variability at 3 month lead time. The consistently high prediction skill of the ANN–GA probably demonstrates that the non-linear relationships between CSA rainfall and SST variations of the Atlantic and Indian Ocean basins are well captured. These results demonstrate that selection of relevant SST predictor field constitutes an important step in seasonal climate prediction. By combining wavelet analysis, WEOF analysis and an ANN–GA prediction model, we have been able to achieve encouraging preliminary climate analysis and prediction results.

8. SUMMARY AND CONCLUSIONS

Using wavelet analysis and WEOF analysis on SAWP and ISP, we identified the non-stationary SST fields of the South Atlantic and Indian Oceans that are associated with coherent regions of rainfall variability in CSA. The Benguela ocean current SSTs form the dominant spatial pattern of the South Atlantic Ocean. The spatial patterns were found to migrate seasonally between Africa's west coast and South America's east coast. The Brazil and Guinea ocean current SSTs form the second leading spatial patterns of variability of the Atlantic Ocean. The northern Indian Ocean SST forms the leading spatial pattern, followed by the south Indian Ocean SST. The Atlantic Ocean SST variability was found to have stronger associations with CSA rainfall variability than the Indian Ocean SST.

The ISP WPCs show strong spatial and temporal variations at the 2–2.4 and 5.6–6.7 year cycles over CSA, whereas the time-domain WPCs accurately represent temporal variations of CSA rainfall. Further, it seems that CSA rainfall responded to ENSO only when simultaneous decreases of energy at 2–2.4 and 5.6–6.7 year periods occurred.

Using predictor fields identified from the Atlantic and Indian Oceans, the non-linear ANN–GA system could predict accurate CSA rainfall at 3 month lead times. At 3 month lead time, correlations between 0.8 and 0.9, RMSEs between 0.4 and 0.9, and HK skill scores between 0.4 and 0.8 were obtained between observed and predicted CSA rainfall.

ACKNOWLEDGEMENTS

This project was partly supported by NSERC of Canada. The first author was also supported by the Commonwealth scholarship of CIDA, Canada. The UK Meteorological Office provided both the SST grid data of the Indian and Atlantic Oceans (part of MOHSST6) and the rainfall data for CSA. The wavelet analysis was done using the software of Torrence and Compo (1998) downloaded from <http://www.paos.colorado.edu/research/wavelets>.

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.
- Bakshi B. 1998. Multiscale PCA with applications to multivariate statistical process monitoring. *AIChE Journal* **44**: 1596–1610.
- Barnston AG, Thiao W, Kumar V. 1996. Long-lead forecasts of seasonal precipitation in Africa using CCA. *Weather and Forecasting* **11**: 506–520.
- Basalirwa CPK. 1995. Delineation of Uganda into climatological rainfall zones using the method of principal component analysis. *International Journal of Climatology* **15**: 1161–1177.
- BBC 2002a. <http://news.bbc.co.uk/2/hi/africa/2027079.stm> [Accessed April 2004].
- BBC 2002b. <http://news.bbc.co.uk/2/hi/africa/2504661.stm> [Accessed April 2004].
- Castleman KR. 1996. *Digital Image Processing*. Prentice Hall: Englewood Cliffs, NJ.
- Fanta B, Zaake BT, Kachroo RK. 2001. A study of the variability of the river flow of the southern Africa region. *Hydrological Sciences* **4**: 513–524.
- Goldberg DE. 1989. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley: Reading, MA.
- Houghton RW, Tourre YM. 1992. Characteristics of low-frequency sea surface temperature fluctuations in the tropical Atlantic. *Journal of Climate* **5**: 765–771.
- Hsieh WW, Tang B. 1998. Applying neural network models to prediction and data analysis in meteorology and oceanography. *Bulletin of the American Meteorological Society* **79**: 1855–1868.
- Huang NE, Shen Z, Zheng Q, Yen N, Tung CC, Liu HH. 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.
- Hulme M. 1994. Validation of large-scale precipitation fields in general circulation models. In *Global Precipitations and Climate Change*, Desbois M, Desalmand F (eds). NATO ASI Series. Springer-Verlag: Berlin; 387–406.
- Hulme M, Conway DD, Joyce A, Mulenga H. 1996. A 1961–90 climatology of Africa south of the Equator and a comparison of potential evapotranspiration estimates. *South African Journal of Science* **92**: 334–343.
- Jury MR. 1996. Regional teleconnection pattern associated with summer rainfall over South Africa, Namibia and Zimbabwe. *International Journal of Climatology* **16**: 135–153.
- 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 WA, 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, Tennant WJ. 2000. Statistical downscaling of monthly forecasts. *International Journal of Climatology* **20**: 1521–1532.
- 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. 1995. Sea-surface temperature–south African rainfall associations, 1910–1989. *International Journal of Climatology* **15**: 119–135.
- 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; 113–134.
- 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.
- Nicholson SE. 1986. The nature of rainfall variability in Africa south of the equator. *Journal of Climatology* **6**: 515–530.
- Ntale HK, Gan TY, Mwale D. 2003. Prediction of East Africa seasonal rainfall using canonical correlation analysis. *Journal of Climate* **16**(12): 2105–2112.
- 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.
- Richard Y, Trzaska S, Roucou P, Rouault M. 2000. Modification of the southern African rainfall variability/ENSO relationship since the late 1960s. *Climate Dynamics* **16**: 883–895.
- Ropelewski CF, Halpert MS. 1987. Global and regional scale precipitation patterns associated with El Niño/southern oscillation. *Monthly Weather Review* **115**: 1601–1626.
- Rumelhart DE, Hinton GE, Williams RJ. 1986. Learning internal representations by error propagation. *Parallel Distributed Processing*, Vol. 1, Rumelhart DE, McClelland JL, Group PR (eds). MIT Press: 318–362.
- 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, MD.
- Smith LC, Turcotte DL, Isacks BL. 1998. Streamflow characterization and feature detection using a discrete wavelet transform. *Hydrological Processes* **12**: 233–249.
- Torrence C, Compo GP. 1998. A practical guide to wavelets analysis. *Bulletin of the American Meteorological Society* **1**: 61–78.
- United Nations Development Program/Food and Agriculture Organization of the United Nations report. Multipurpose Survey of the Kafue River Basin, Zambia Final Report (FAO/SF:35/ZAM), Volume III, Climatology and Hydrology.
- Wang B, Wang Y. 1996. Temporal structure of the southern oscillation as revealed by waveform and wavelet analysis. *Journal of Climate* **9**: 1586–1598.
- Venegas SA, Mysak LA, Straub DN. 1997. Atmosphere–ocean coupled variability in the South Atlantic. *Journal of Climate* **10**: 2904–2920.
- Webster PJ, Clark C, Cherikova G, Fasullo J, Han W, Loschnigg J, Sahami K. 2002. The monsoon as a self-regulating coupled ocean–atmosphere system. In *Meteorology at the Millennium*, Pearce RP (ed.). International Geophysics Series, vol. 83. Academic Press: San Diego; 198–219.
- Wilks DS. 1995. *Statistical Methods in the Atmospheric Sciences*. Academic Press: San Diego.