15-year simulation of the December to March rainfall season of the 1980s and 1990s using canonical correlation analysis (CCA)

WA Landman* and E Klopper
Research Group for Seasonal Climate Studies, South African Weather Bureau, Private Bag X097, Pretoria, 0001

Abstract

South African summer rainfall is associated with sea-surface temperature variations in the global oceans. The South African Weather Bureau uses a canonical correlation analysis (CCA) model to make operational seasonal rainfall predictions. Evolutionary features of global-scale sea-surface temperatures, such as a warming equatorial Pacific Ocean, are used as predictors. During the Workshop on the Use and Benefits of Seasonal Outlooks (WUBSO) held on 27 August 1996, it was suggested by the participants that an assessment be made on how the model would have performed during previous El Niño/Southern Oscillation (ENSO) years in order to establish the confidence level that could be employed by end-users of these forecasts. In this study the operational model used by the Research Group for Seasonal Climate Studies (RGSCS) is validated over a 15-year period from 1981/82 to 1995/96 in order to determine how it would have performed if this prediction model had been implemented from the beginning of the 1980s, predicting December to March aggregate rainfall. It is found that the model is highly successful during ENSO years, and has a success hit rate of 40% for the non-ENSO years. Possible explanations for the model's successes and failures are presented.

Introduction

Rainfall over Southern Africa is highly seasonal. Except for the south-western Cape, the southern coastal regions and adjacent interior, more than 80% of the annual rainfall occurs between October and March (Tyson, 1986). Most of the austral summer rainfall of South Africa is of convective origin. The intensity of convection is regulated, apart from the diurnal heating of the surface, by tropical easterly disturbances to the north and westerly disturbances to the south, and their interactions.

Sea-surface temperature variations are responsible for a large portion of the rainfall variability of the austral summer rainfall over Southern Africa. On a seasonal time-scale, the El Niño/Southern Oscillation (ENSO) phenomenon (Zhang et al., 1997) affects the atmospheric circulation outside the tropics (Philander, 1990), and Southern Africa tends to experience dry conditions during warm ENSO events (Ropelewski and Halpert, 1987).

Sea-surface temperature anomalies of the oceans adjacent to Southern Africa are also related to South African seasonal rainfall (Nicholson and Entekhabi, 1987; Walker, 1989; 1990; Walker and Lindesay, 1989; Mason, 1990; 1992; 1995; Mason et al., 1994; Jury and Pathack, 1991; Jury et al., 1993; Landman, 1997). Sea-surface temperature gradient intensity in both the far south-western and south-eastern Atlantic Ocean varies closely in phase with the annual rainfall totals of the summer rainfall region of South Africa (Mason, 1990). However, the strongest rainfall/sea-surface temperature association is in the central south Atlantic (Mason, 1990; 1992). Areas of strong association between sea-surface temperature and South African rainfall are located distant from land in the tropical Indian Ocean east of 50°E (Walker, 1990), but with the strongest rainfall/sea-surface temperature relation in the western equatorial Indian Ocean (Mason, 1990).

Rainfall/sea-surface temperature associations vary over the summer rainfall season. The Arabian sea area, as well as the equatorial Pacific Ocean present potential for prediction of December rainfall. January rainfall has a poor association with sea-surface temperatures, while the central equatorial Indian Ocean shows very significant associations with February to March rainfall (Pathack et al., 1993). In this study the sea-surface temperature tendencies are incorporated in a statistical model to make operational seasonal rainfall predictions for South Africa.

Data and methods

The sea-surface temperatures (4 consecutive 3-month seasonal means), which are used as predictors in a statistically based model are a combination of different data sets. Global Ocean Global Atmosphere (GOGA) sea-surface temperature data (Pan and Oort, 1990; Lau and Nath, 1994) of the global oceans between about 45°N and 45°S (704 grid-points), are used as predictors of rainfall over South Africa. The sea-surface temperature data set has a resolution of 4.5°x7.5° latitude-longitude, and was obtained for the period 1950 to 1985. Blended sea-surface temperature data (Reynolds, 1988) have been obtained for the period 1985 to 1995. The blended (2° latitude-longitude) data are interpolated to the GOGA grid, using cubic interpolation.

The original monthly rainfall data were obtained from the Computing Centre for Water Research (CCWR), but have been updated since 1996 using data supplied by the South African Weather Bureau. Monthly data for 418 stations for the period 1961 to 1990 were used. Clustering was performed using Euclidian distances and Ward's minimum variance to obtain eight regions. Figure 1 shows the geographical distribution of the rainfall stations in the eight homogeneous rainfall regions, and is an update of regions described by Mason (1998).

The prediction technique used here is called canonical correlation analysis (CCA) (Glahn, 1968; Anderson, 1984; Barnett and Preisendorfer, 1987; Graham et al., 1987a; b; Jackson, 1991; Barnston, 1994; Barnston and Smith, 1996; Wilks, 1995). CCA is a regression-based technique and at the top of the regression modelling hierarchy (Barnett and Preisendorfer, 1987). CCA has the ability to seek relationships between two sets of variables

* To whom all correspondence should be addressed.
☎ (012) 309-3717; fax (012) 323-4518; e-mail willem@cirrus.sawb.gov.za
Received 28 August1997; accepted in revised form 26 June 1998

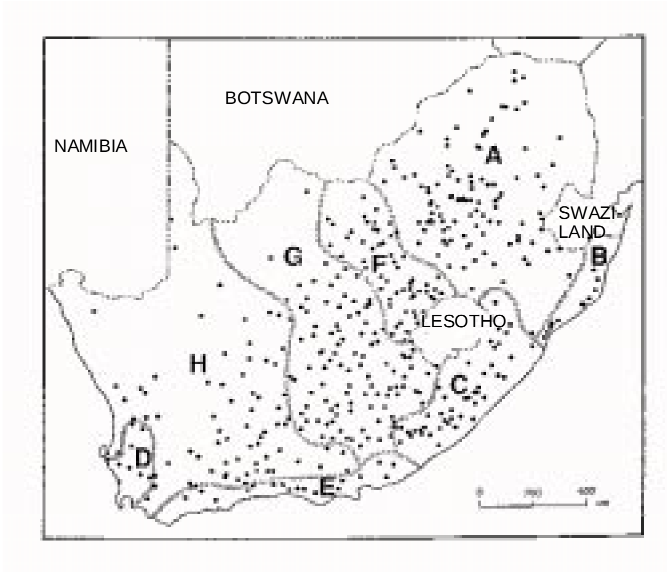

Figure 1
Geographical distributions of the rainfall stations and the eight homogeneous regions of South Africa (after Mason,1998)

which vary in both time and space by identifying the optimum linear combination between the two sets with maximum correlation being produced. Model skill of a statistical method can be considered to be a base-line skill level (Barnston et al., 1994; Carson, 1998) that has to be outscored by more elaborate techniques, such as general circulation models. If the ocean-atmosphere system contains adequate inherent predictability, dynamic models should be able to outscore models that do not accommodate physical processes or non-linearity (Barnston et al., 1994), such as the one evaluated here.

For this linear model, empirical orthogonal function analysis (Anderson, 1984; Preisendorfer, 1988; Jolliffe, 1990; 1993; Jackson, 1991; Johnston, 1992; Peixoto and Oort, 1992; Von Storch and Navarra, 1995) is first performed on the rainfall and sea-surface temperature sets. The number of principal components retained account for about 70% and 80% of the variance respectively. Only two CCA modes are retained in the prediction equation.

The prediction scheme

The rainfall for December to March (DJFM) is predicted using sea-surface temperature anomalies for each of the four preceding three-month seasons:

$$D_{(i-1)}(\text{JF MAM JJA SON})_i \Rightarrow D_i(\text{JFM})_{(i+1)}$$

where i represents the current year. Similar schemes are used to predict the same rainfall season for different leads and seasons. Forecasts are made for three equi-probable categories of below-normal (B), near-normal (N) and above-normal (A) rainfall. A forecast is variance-adjusted by increasing its standard deviation in order to account for extreme events (Ward and Folland, 1991).

When separate oceans carry unique climatic signals, significant improvements in forecast skill should be attained by combining the oceans in a single model (Barnett and Preisendorfer, 1987). Considering the oceans adjacent to Southern Africa and the equatorial Pacific Ocean separately instead of the global set of sea-surface temperatures, the model skill seems to be less when only the first few empirical orthogonal functions (EOFs) are included in the CCA process as opposed to using lower order EOFs. Skill improves significantly when lower order EOFs are included, suggesting that the climate signal from the individual oceans are not included in the first few EOFs of global sea-surface temperatures (SSTs), and are subsequently excluded from the CCA prediction process described here.

Retroactive real-time validation is performed to make the hindcast predictions, for example, to predict 1981/82 to 1983/84 the model is trained over the 1951 to 1980 period (30 years); to predict 1984/85 to 1986/87 the model is trained over the 1951 to 1983 period (33 years), and so on, until the 42-year climate of 1951 to 1992 is used to train the model to predict the 1993/94 to 1995/96 seasons.

TABLE 1
SUMMARY OF THE CATEGORICAL FORECASTS ("FOR" - LEFT) AND OBSERVATIONS ("OBS" - RIGHT) FOR DJFM FOR THE 15-YEAR PERIOD OVER THE SUMMER RAINFALL REGIONS (A, B, C, F, G, H). EN = EL NIÑO YEARS; LN = LA NIÑA YEARS; A = ABOVE-NORMAL; N = NEAR-NORMAL; B = BELOW-NORMAL.

	81		82 EN		83		84		85		86 EN		87		88 LN		89		90		91 EN		92		93		94		95 LN	
	FOR	OBS	FOR	OBS	FOR	OBS	FOR	OBS	FOR	OBS	FOR	OBS	FOR	OBS	FOR	OBS	FOR	OBS	FOR	OBS	FOR	OBS	FOR	OBS	FOR	OBS	FOR	OBS	FOR	OBS
A	N	B	B	B	N	N	N	N	N	N	B	N	B	A	N	N	N	N	B	A	B	B	B	N	A	N	B	N	N	A
B	N	B	B	B	B	A	A	A	A	N	B	N	B	A	N	A	A	A	B	A	B	B	B	B	N	N	B	B	N	A
C	A	B	B	B	N	N	A	A	A	A	B	B	B	A	A	A	N	N	B	N	B	B	B	B	N	A	B	N	N	A
F	N	B	B	B	N	B	A	B	N	N	B	B	N	A	A	A	N	N	B	A	B	B	B	B	A	A	B	N	N	A
G	A	B	B	B	N	B	A	N	A	N	B	B	B	A	A	A	N	B	B	A	B	B	B	B	N	A	B	N	N	N
H	N	B	B	B	N	N	A	A	A	A	B	B	B	N	A	A	N	B	B	N	B	B	B	B	N	A	B	N	N	A

Results

Table 1 gives a summary of the categorical forecasts for the DJFM seasons for the 15 years of 1981/82 to 1995/96. Only Regions A, B, C, F, G and H (Fig. 1) which receive their maximum rainfall over the DJFM period are considered. Only the results for the 0-month lead predictions are shown, but lag simulations of 4 to 6 months lead show that the forecasts are still statistically significant (95%) over the main maize-growing regions. This substantial lead time confirms the utility of the forecasts over this particular region.

Table 2 shows the number of hits and one- and two-category misses. In order for the predictions to outscore chance the percentage of perfect forecasts has to be at least 33%, less than 44% for one category missed, and less than 22% for two categories missed. These percentages are calculated from a contingency table representing three equi-probable categories. Except for region H which missed by one category, the forecasts are outscoring chance.

TABLE 2
NUMBER OF HITS AND MISSES DURING THE 15-YEAR PERIOD. PERCENTAGE OF HITS AND MISSES IS SHOWN IN BRACKETS.

Region	Correctly forecast	1 Category missed	2 Categories missed
A	7 (47%)	6 (40%)	2 (13%)
B	7 (47%)	5 (33%)	3 (20%)
C	9 (60%)	4 (27%)	2 (13%)
F	8 (53%)	5 (33%)	2 (13%)
G	6 (40%)	6 (40%)	3 (20%)
H	8 (53%)	7 (47%)	0 (0 %)

Discussion

Table 2 suggests that the rainfall of Regions C, F and H is predicted the most accurately, because they have the highest number of correctly predicted categories (8 or 9) *and* the least number of misses by two categories (less than 3). The predictions during El Niño years are always below-normal for the 15-year period. For the three El Niño years (1982/83, 1986/87 and 1991/92) the observed category was below-normal throughout, except for Regions A and B of 1986/87, where near-normal rainfall was observed. For La Niña years (1988/89 and 1995/96) predictions are near-normal to above-normal with near-normal to above-normal rainfall observed, but the forecast for 1995/96 underestimated the rainfall. About 40% of the forecasts made for non-ENSO years are predicted correctly. False alarms (forecasts for below-normal were made, but near-normal to above-normal rainfall actually occurred) were made for the 1987/88 and 1990/91 seasons. These two seasons were investigated in more detail to find possible factors contributing to the inaccurate forecasts.

TABLE 3
RAINFALL AT KIMBERLEY AND BLOEMFONTEIN DURING THE DJFM SEASON OF 1987/88

	DJFM normal	Predicted category	Observed rainfall	Excluding flood
Kimberley	146 to 255 mm	Below-normal	494 mm (Above-normal)	242 mm (Near-normal)
Bloemfontein	212 to 353 mm	Near-normal	693 mm (Above-normal)	358 mm (Above-normal)

During February 1988 flooding occurred (Triegaardt et al., 1991) over the central interior (Fig. 2; mainly Regions F and G in Fig. 1). Short-term flooding events (2 to 5 d) cannot be predicted on a seasonal time-scale of 2 to 6 months. If the rainfall measured at Kimberley and Bloemfontein during 19 to 23 February 1988 is removed from the seasonal total, then the observed category would have been near-normal (Table 3). This indicates that the forecasts would still have been out by one category even if no flooding occurred, suggesting that sea-surface temperatures were not the main control mechanism for this particular season's rainfall.

By investigating the Niño3.4 (mean sea-surface temperature of the area between 170°W and 120°W, and 5°N to 5°S) anomaly time series (Fig. 3) over the 15-year period it is indicated that during these two seasons (1987/88 and 1990/91), although the sea-surface temperatures were anomalously warm, the absence of a distinct upward or downward trend is evident. Trends (indicated by the arrows in Fig. 3) were evident during seasons that had been predicted adequately. This suggests that anoma-

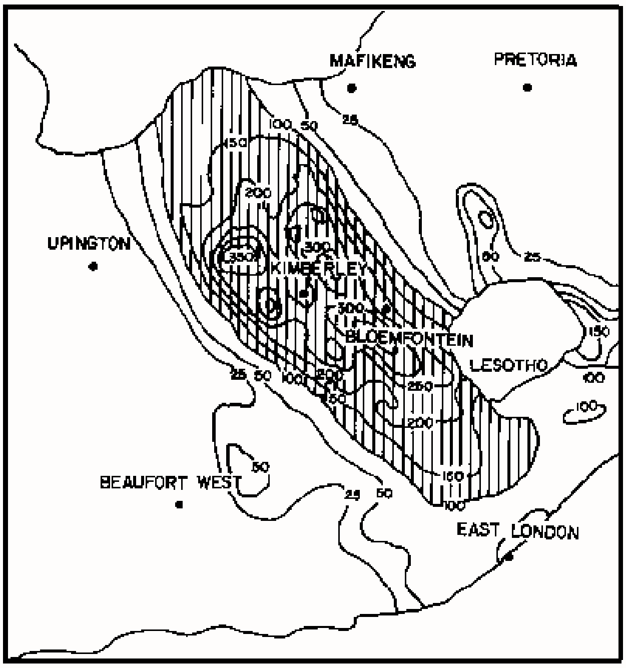

Figure 2
Rainfall (mm) during the Free State floods (19 to 23 February 1988)

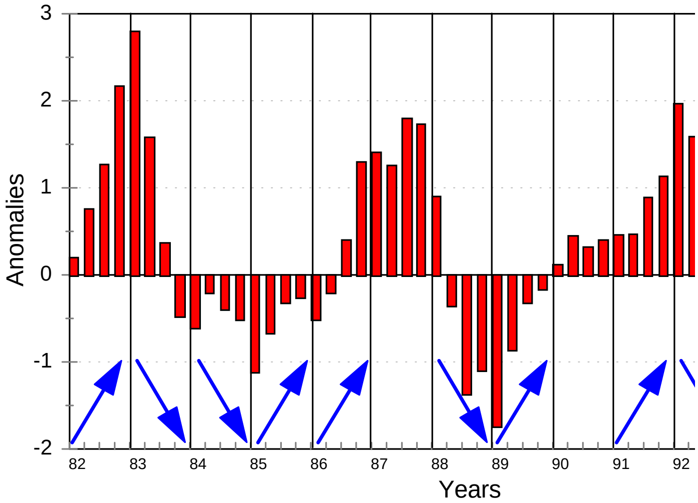

Figure 3
Three-month mean sea-surface temperature anomalies for the Niño3.4 region. Arrows indicate short-term trend

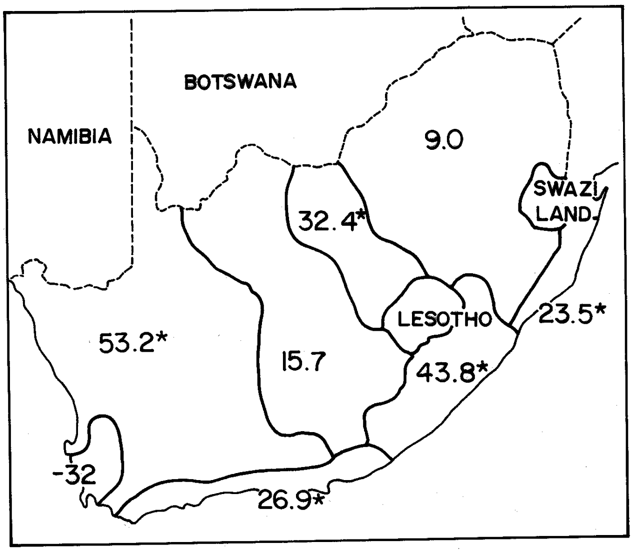

Figure 4
LEPS scores calculated over the 15-year period 1981 to 1995. Scores significant at the 95% level are marked with an asterisk

lously warm (cold) sea-surface temperatures in the equatorial Pacific Ocean may not be a sufficient condition for the model to predict below-normal (above-normal) rainfall over South Africa during DJFM. The model specifically uses short-term trends (such as a warming equatorial Pacific) as predictor and not only the current state (or amplitude) of the sea-surface temperature anomalies. This partly explains the successful forecasts during years of these observed trends.

From an analysis of the model skill, using *Linear Error in Probability Space* (LEPS) scores (Potts et al., 1996) over the period 1980 to 1995 (Fig. 4) it is found that skill scores over the north-eastern part of the country are very poor, while the highest skill scores are obtained over the western interior. The poor skill scores in Region A may mainly be due to the scarcity of data in the Lowveld regions.

TABLE 4
PERCENTAGE HITS DURING NON-ENSO YEARS

Region	Percentage of years correctly forecast
A	40 %
B	50 %
C	50 %
F	40 %
G	10 %
H	40 %

Conclusions

It is concluded that operational seasonal predictions using CCA have potential and this method would have been able to accurately predict rainfall anomalies during the EL Niño events of the 1980s and 1990s, especially during the 1982/83 and 1991/92 major events. During El Niño years, the forecasts are for below-normal rainfall to occur, and below-normal to near-normal rainfall was observed in all cases. Also, during La Niña years the model successfully predicted near-normal to above-normal rainfall. It was also shown that forecast skill did better than chance, thus indicating the potential of these forecasts if employed operationally to alter decisions in planning processes. Its value is further illustrated by non-ENSO years such as 1984/85, 1985/86, 1989/90 when rainfall for four out of the six regions was predicted correctly. For 1992/93, rainfall for five out of the six regions was predicted correctly, with the predicted rainfall for Region A only one category out. It is expected that the model should at least be successful in predicting the rainfall during ENSO years correctly because it uses SSTs as the only predictor. In addition to predicting rainfall during ENSO years successfully, the predictions during non-ENSO years outscored chance (higher than 33%), except for region G (Table 4).

The results of this study also suggest that a seasonal trend in sea-surface temperatures may be a necessary requirement for the model to successfully predict seasonal rainfall over South Africa. Also, the model is incapable of considering short-term rainfall variability, and cannot account for the role played by flooding events such as the Free State floods of 1988.

References

ANDERSON TW (1984) *An Introduction to Multivariate Statistical Analysis*. Wiley, New York. 675 pp.

BARNETT TP and PREISENDORFER R (1987) Origins and levels of monthly and seasonal forecast skill for United States air temperature determined by canonical correlation analysis. *Mon. Weather Rev.* **115** 1825-1850.

BARNSTON AG (1994) Linear statistical short-term climate predictive skill in the Northern Hemisphere. *J. Clim.* **7** 1513-1564.

BARNSTON AG, VAN DEN DOOL HM, ZEBIAK SE, BARNETT TP, JI M, RODENHUIS DR, CANE MA, LEETMAA A, GRAHAM NE, ROPELEWSKI CR, KOUSKY VE, O'LENIC EA and LIVEZEY RE (1994) Long-lead seasonal forecasts - Where do we stand? *Bull. Am. Met. Soc.* **66** 159-164.

BARNSTON AG and SMITH TM (1996) Specification and prediction of global surface temperature from global SST using CCA. *J. Clim.* **9** 2660-2697.

CARSON DJ (1998) Seasonal forecasting. *Q.J.R. Meteorol. Soc.* **124** 1-26.

GLAHN HR (1968) Canonical correlation and its relationship with discriminant analysis and multiple regression. *J. Atmos. Sci*. **25** 23-31.

GRAHAM ME, MICHAELSEN J and BARNETT T (1987a) An investigation of the El Niño-Southern Oscillation cycle with statistical models. 1. Predictor field characteristics. *J. Geophys. Res.* **92** 14251-14270.

GRAHAM ME, MICHAELSEN J and BARNETT T (1987b) An investigation of the El Niño-Southern Oscillation cycle with statistical models. 2. Model results. *J. Geophys. Res.* **92** 14271-14289.

JACKSON JE (1991) *A User's Guide to Principal Components.* John Wiley & Sons inc, New York. 569 pp.

JOHNSTON RJ (1992) *Multivariate Statistical Analysis in Geography*. Longman, New York. 280 pp.

JOLLIFFE IT (1990) Principal component analysis: A beginner's guide - I: Introduction and application. *Weather* **45** 375-382.

JOLLIFFE IT (1993) Principal component analysis: A beginner's guide - II: Pitfalls myths and extensions. *Weather* **48** 246-253.

JURY MR and PATHACK B (1991) A study of climate and weather variability over the tropical southwest Indian Ocean. *Meterol. Atmos. Phys.* **47** 37-48.

JURY MR, VALENTINE HR and LUTJEHARMS JRE (1993) Influence of the Agulhas current on summer rainfall along the southeast coast of South Africa. *J. Appl. Meteorol*. **32** 1282-1287.

LANDMAN WA (1997) A Study of Rainfall Variability of South Africa as Revealed by Multivariate Analysis. Unpublished M.Sc. Dissertation, University of Pretoria, South Africa. 134 pp.

LAU N-C and NATH MJ (1994) A modelling study of the relative roles of tropical and extratropical SST anomalies in the variability of the global atmosphere-ocean system. *J. Clim.* **7** 1184-1207.

MASON SJ (1990) Temporal variability of sea surface temperatures around Southern Africa: a possible forcing mechanism for the 18-year rainfall oscillation? *S. Afr. J. Sci.* **86** 243-252.

MASON SJ (1992) Sea Surface Temperature and South African Rainfall Variability. Unpublished Ph.D Thesis, University of the Witwatersrand, Johannesburg. 235 pp.

MASON SJ (1995) Sea-surface temperature - South African rainfall associations 1910-1989. *Int. J. Climatol.* **15** 119-135.

MASON SJ (1998) Seasonal forecasting of South African rainfall using a non-linear discriminant analysis model. *Int. J. Climatol.* **18** 147-164.

MASON SJ, LINDESAY JA and TYSON PD (1994) Simulating drought in Southern Africa using sea-surface temperature variations. *Water SA* **20** 15-22.

NICHOLSON SE and ENTEKHABI D (1987) Rainfall variability in equatorial and Southern Africa: relationships with sea surface temperatures along the southwestern coast of Africa. *J. Clim. Appl. Meteorol.* **26** 561-578.

PAN YH and OORT AH (1990) Correlation analysis between sea surface temperature anomalies in the eastern equatorial Pacific and the world ocean. *Climate Dyn.* **4** 191-205.

PATHACK BMR, JURY MR, SHILLINGTON FA and COURTNEY S (1993) South African Summer Rainfall Variability and its Association with the Marine Environment. Water Research Commission Report 278/1/94.

PEIXOTO JP and OORT AH (1992) *Physics of Climate*. American Institute of Physics. New York 520 pp.

PHILANDER SG (1990) *El Niño, La Niña, and the Southern Oscillation.* Academic Press. London 293 pp.

PREISENDORFER RW (1988) *Principal Component Analysis in Meteorology and Oceanography*. Elsevier. New York 425 pp.

ROPELEWSKI CF and HALPERT MS (1987) Global and region scale precipitation patterns associated with the El Niño/Southern Oscillation. *Mon. Wea. Rev*. **115** 1606-1626.

POTTS JM, FOLLAND CK, JOLLIFFE IT and SEXTON D (1996) Revised "LEPS" scores for assessing climate model simulations and long-range forecasts. *J. of Climate* **9** 34-53

REYNOLDS RW (1988) A real-time global sea surface temperature analysis. *J. Clim.* **1** 75-86.

TYSON PD (1986) *Climate Change and Variability in Southern Africa*. Oxford University Press Cape Town 220 pp.

TRIEGAARDT DO, VAN HEERDEN J and STEYN PCL (1991) Anomalous precipitation and floods during February 1988. *South African Weather Bureau Technical Paper No 23* 25pp.

VON STORCH H and NAVARRA A (eds.) (1995) *Analysis of Climate Variability. Application of Statistical Techniques.* Springer-Verlag Berlin 334 pp.

WALKER ND (1989) Sea Surface Temperature - Rainfall Relationships and Associated Ocean-atmosphere Coupling Mechanisms in the Southern African Region. CSIR Research Report Stellenbosch 173 pp.

WALKER ND (1990) Links between South African summer rainfall and temperature variability of the Agulhas and Benguela currents systems. *J. Geophys. Res.* **95** 3297-3319.

WALKER ND and LINDESAY JA (1989) Preliminary observations of oceanic influences on the February-March 1988 floods in central South Africa. *S. Afr. J. Sci.* **85** 164-169.

WARD MN and FOLLAND CK (1991) Prediction of seasonal rainfall in the north Nordeste of Brazil using eigenvectors of sea-surface temperature. *Int. J. Climatol.* **11** 711-743.

WILKS DS (1995) *Statistical Methods in the Atmospheric Sciences*. Academic Press San Diego 467 pp.

ZHANG Y, WALLACE JM and BATTISTI DS (1997) ENSO-like interdecadel variability: 1900-93. *J. of Climate* **10** 1004-1020.