STATISTICAL PREDICTION OF SEA-SURFACE TEMPERATURE OVER THE TROPICAL ATLANTIC

CARLOS ALBERTO REPPELLI* and PAULO NOBRE
Centro de Previsão de Tempo e Estudos Climáticos (CPTEC), Instituto Nacional de Pesquisas Espaciais (INPE), Cachoeira Paulista, SP, Brazil

ABSTRACT
A statistical system to predict sea-surface temperature anomalies (SSTAs) over the tropical oceans, with emphasis on the tropical Atlantic, is described. Canonical correlation analysis is used to identify critical sequences of predictor patterns, which tend to evolve into subsequent patterns and which can be used to form a forecast. The results indicate that SST fields over the equatorial Pacific and tropical Atlantic can be a potential predictor of the SSTAs over the tropical Atlantic 3–4 months in advance. The spatial structures of the SSTAs over the tropical Atlantic for the period March–May are well captured by the predictions done with initial conditions from September to February. Model performance is better over the northern tropical Atlantic than over the southern tropical Atlantic, where persistence is hardly beaten. Results of this work can contribute to improve seasonal climate predictions of rainfall anomalies over the northeast Brazil region.

KEY WORDS: Atlantic; canonical correlation analysis; climate modelling; sea-surface temperature; statistical prediction

1. INTRODUCTION

Seasonal climate predictions have become a reality over the last decade, with operational seasonal forecasts appearing even more recently (Graham, 1994). However, present approaches for climate prediction, either statistical or dynamical, are critically dependent on sea-surface temperature (SST) forecasts (Shukla, 1984; Ward and Folland, 1991).

The dynamical models use SST fields as a forcing for the interactive processes (heat flux, for example) of the ocean–atmosphere system (Cavalcanti et al., 1998). The statistical models use SST anomalies (SSTAs) as a predictor parameter of the regression equations (Hastenrath, 1990; Ward and Folland, 1991). The inter-hemispheric SST gradient of the tropical Atlantic, often referred to as the Atlantic SST dipole pattern, is a mode of variability apparently unique to the tropical Atlantic. Studies suggest that this mode of variability is associated with the rainfall variability over northern South America and the African Sahel (Moura and Shukla, 1981; Parker et al., 1988; Servain, 1991; Nobre and Shukla, 1996).

The most recent and promising class of methods for seasonal climate forecasting employs general circulation models of the atmosphere (GCMA) to forecast seasonal rainfall anomalies over some regions of the globe with 1–3 months lead time. The great advantage of predicting the climate using dynamical methods relative to statistical models is the possibility of forecasting the spatial distribution of precipitation anomalies over a region (Ward and Folland, 1991; Hastenrath and Greischar, 1993; Graham, 1994).

* Correspondence to: Carlos Alberto Repelli, DB Live/INFO Weather, 5539 Hitcher Bend, Austin, Texas 78749, USA; e-mail: carlos@repelli.com

Copyright © 2004 Royal Meteorological Society
Statistical and dynamical models have been developed to forecast SSTA fields over the equatorial Pacific (e.g. Cane et al., 1986; Barnston and Ropelewski, 1992; Penland and Magorian, 1993; Chen et al., 1997). Studies about the possibility of forecasting SSTAs over the tropical Atlantic through statistical methods are recent (Penland and Mastrosova, 1998) and useful owing to the great importance of the tropical Atlantic on climate variability over some areas of the Americas and the Caribbean (Hastenrath, 1984).

Penland and Mastrosova (1998) developed a model to predict the tropical Atlantic SSTA index for different areas of the basin (north, equatorial, south and Caribbean) using a statistical technique known as linear inverse modelling. They used global tropical SSTA fields as predictor parameters and concluded that, on time scales of several months to a year, predictions of northern tropical Atlantic and Caribbean SSTAs are much more skilful than equatorial and southern tropical Atlantic SSTAs. That model did not beat persistence over the southern area.

Using different methods, various authors found that, during the development of an El Niño event, there is a tendency for the northern tropical Atlantic to be warmer than normal (Hastenrath et al., 1987; Curtis and Hastenrath, 1995; Enfield and Mayer, 1997). Uvo et al. (1998) used a statistical technique known as singular value decomposition (Bretherton et al., 1992) and found that the SSTAs are well correlated in both tropical Pacific and Atlantic Oceans. The highest correlations were found between the Pacific SSTA (over the El Niño–southern oscillation (ENSO) region especially) in January and the Atlantic SSTA (tropical northern sector) in March.

In the present study, the use of canonical correlation analysis (CCA; Bretherton et al., 1992; Cherry, 1996) pre-filtered with an empirical orthogonal function (EOF; Preisendorfer, 1988) for predicting fluctuations of monthly SSTAs over the tropical Atlantic with 1–12 months lead time is examined. Special emphasis is given for the season March–May (MAM), because of its impact on interannual variability of the northeast Brazil (Nordeste) rainy season (Hastenrath and Heller, 1977; Moura and Shukla, 1981). The CCA technique has previously been described in both meteorological and geophysical literature and it has been shown to be a useful statistical method for climate modelling (Barnett and Preisendorfer, 1987; Graham et al. 1987; Barnston and Ropelewski, 1992; He and Barnston, 1996; Repelli and Alves, 1996). Bretherton et al. (1992) compared the variant of the CCA pre-filtered with an EOF with other methods of finding coupled patterns in climate data and concluded that, although the method is more complex to implement, it has a very good performance and non-zero but practically insignificant systematic errors.

A discussion related to the CCA technique and the data used in the calculations are presented in Section 2. The results are presented in Section 3, and Section 4 summarizes the results and conclusions.

2. DATA AND METHODOLOGY

The datasets used in this work are monthly mean global 1° × 1° latitude–longitude gridded SSTs from the Comprehensive Oceanic and Atmospheric Data Set (COADS) for the 1945–93 period (Da Silva et al., 1994); and the optimal interpolation SST (OI-SST) for the period 1994–98 provided by the National Centers for Environmental Prediction (Reynolds and Smith, 1994).

A routine is applied to mask out the points over the continents and calculate the long-term monthly means and standard deviations at each grid point. Then, monthly anomaly time series are calculated by subtracting the long-term monthly means from the monthly values and normalizing by the local standard deviations. This process ensures that all grid points have equal opportunity to participate in the predictive patterns, regardless of their latitude- and longitude-dependent interannual variances (Barnston and Ropelewski, 1992).

Both predictand and predictor fields are SSTAs, but they are used in a lead–lag scheme, e.g. SSTA field for January is the predictor field of the SSTA for February, March, etc. Before applying the CCA calculations, the predictor and predictand anomaly datasets are filtered separately with the use of the EOF. The EOFs of each dataset are calculated independently. Then, the filtered data are reconstructed using a number of eigenvectors and eigenvalues, retaining 80% of the original variance of each field.
CCA is calculated in order to find the predictive equations that relate the predictand and predictor fields and a set of diagnostic and prognostic fields can be obtained. CCA is a multivariate statistical technique and is at the top of the hierarchy of regression modelling approaches. It is more complex than multiple regression or discriminant analysis, which treat only one predictand at a time. According to Barnett and Preisendorfer (1987), two sets of vectors, \( r \) and \( q \), can be found from two physical fields \( Y(nt, ny) \) and \( Z(nt, nz) \), where \( nt \) represents the number of observations in time and \( ny \) and \( nz \) are each field’s number of points in space, in a way that \( u(t) = Y_r \) and \( v(t) = Z_q \). Time series \( u(t) \) and \( v(t) \) are also called ‘canonical variables’. The CCA problem is solved under the condition that \( u(t) \) and \( v(t) \) have maximum correlation coefficient for each mode. As described in detail in Graham et al. (1987), the solution of this maximization problem consists in solving the eigenvalue–eigenvector problem \( (S_{yy}^{-1}S_{yz}S_{zz}^{-1}S_{zy} - \mu I) v = 0 \), where \( S \)-subscripts denote the covariance and cross-covariance matrices. This means that CCA uses eigenvalues and eigenvectors in a specialized way, such that the structure of the covariance fields between the predictor and predictand variables is defined under the constraint of maximization of cross-dataset correlation explained by each successive mode.

The solution of the CCA is subject to instabilities, especially when the number of observations in time is not so long, or when there is noise in the data. To avoid this problem, the original datasets are pre-filtered through the EOF technique. Through EOF, the predictor and predictand data sets are separately orthogonalized and then truncated. Then, the resulting predictor and predictand principal component time series (alpha and beta respectively) are analysed in the main portion of the CCA procedure and new canonical variables (\( u \) and \( v \)) are obtained. The principal components of the predictand field and the canonical variable of the predictor field are mutually orthogonal. Linear regression between them is performed and a coefficient of adjustment \( s \) for each mode can be found. Thus, the predictand field can be easily reconstructed through a linear matrix operation \( Z ^ \wedge = \hat{u}sf^* \), where \( f \) is the eigenvector matrix of the predictand field and \( ^\wedge \) represents estimate values. Details about the canonical equations can be seen in Graham et al. (1987).

There are some advantages in using the CCA technique in geophysical and climate studies. First, the technique allows field-to-field comparisons, which makes the analysis of the connection between the variables much more powerful and informative than that performed using just an index. Second, if more than one distinct mode of covariability is present, then CCA will (ideally) distinguish between them and the structure of each mode can be compared (for consistency) with physical relationships or numerical results. Finally, the true strength (in the linear least-squares sense) of the cross-field relation can be estimated, both on a mode-by-mode basis and as a whole. In addition to the facts mentioned above, CCA can be used in diagnostic studies and as a basis to develop a forecast method.

The reconstructed fields from the EOF (predictor and predictand) can be projected into the canonical temporal variables, \( u(t) \) and \( v(t) \) respectively, generating a set of spatial amplitudes of the predictor and predictand fields of each canonical mode. These spatial patterns, known as \( g \)-map and \( h \)-map, are very useful because they allow us to investigate what regions of variability over the predictor field have more influence on the variability of the predictand. At the same time, it is possible to verify what regions over the predictand are being more influenced by the predictor for several different lags. Cherry (1996) pointed out that there is a chance of such patterns being spurious, i.e. they have no physical relationships (purely mathematical) when sample sizes are small, and this potential can be increased when the observations are not independent of one another (e.g. the observations represent a temporally autocorrelated time series) (Kendall, 1975).

Different from the temporal canonical variables \( u \) and \( v \), the spatial patterns \( g \) and \( h \) are non-orthogonal. The non-orthogonality in this case must be taken into account when CCA is used purely to investigate teleconnection patterns in climate, for example, because different modes could eventually contain repeated information. However, neither the reconstructed fields are affected nor are artificial skills generated by this fact, because the patterns used as input to solve the optimization problem during CCA performance, i.e. the principal components of the predictor and predictand fields, are orthogonal.

An oceanic modelling system (named as SIMOC) was built to test the method, and the area chosen for the predictor field is the Pacific and Atlantic Oceans between 20°S and 30°N (Figure 1(a)). The area of the predictand is set over the same latitude band of the predictor field, but only over the tropical Atlantic, since the major interest is to investigate the potential predictability of the tropical Atlantic SSTAn (Figure 1(b)).

Copyright © 2004 Royal Meteorological Society

A total of 49 years in the period 1945–93 are used to validate the model; 34 years for the training period (1945–78) and the following 15 years (1979–93) for the hindcast period. Lead times from 1 to 12 months are tested, starting each month of the year.

The hindcasts are compared with the observed field through the correlation coefficients for every grid point of the predictand field. The average correlation coefficients for some areas of the ocean basin are analysed for all lead/lag experiments. Also, some experiments were done in order to test the ability of the model in predicting the SSTA of the particular period MAM. The focus of the predictions over this period is particularly useful for climatic predictions over Brazil’s Nordeste, since MAM is the period of its rainy season.

Once the hindcasts are calculated for different leads for the period 1979–93 (independent period), the performance of the model is investigated through the calculation of the correlation coefficients between the predicted and the observed SSTA. These are then compared with correlations obtained by persisting the SSTA. Also, a statistical test of the null hypothesis is applied (Student’s $t$-test) in order to calculate the significance level of the correlation coefficients found (Bendat and Persol, 1986).

3. RESULTS

As mentioned at the end of Section 2, the period 1979–93 is used to do the hindcast and all statistics and results obtained when testing the performance of the model. Area-averaged monthly correlation coefficients between observed SSTA and CCA model forecasts and persistence are shown for the southern tropical Atlantic (20–0°S) and northern tropical Atlantic (0–30°N) for up to 12 months lead time for predictor sets taken from September to August (figure not shown). For the southern tropical Atlantic as a whole, only the CCA forecasts with initial conditions in September outperformed persistence. For all the other cases, the model forecasts were essentially as skillful as persistence, or worse. Also, correlation coefficients were modest, with peak values ranging between 0.4 and 0.6. For the northern tropical Atlantic, the area-averaged correlation coefficients are generally small for both the persisted and the forecast SSTAs. However, it could be concluded that SSTA forecasts by the CCA model with lead times up to 6 months have a higher skill than persistence for the period March–June.

The average correlation coefficients over the whole basins as discussed above is a stringent test, since it encompasses areas where the models have high forecast skill with areas where the models perform poorly. In order to evaluate the model’s skill over some key regions over the tropical Atlantic, the same correlation calculations are averaged over the areas known as the northern and southern ‘centres of action’ of the tropical Atlantic SSTA, referred to as the ‘dipole pattern’ (Moura and Shukla, 1981). These areas are 20–0°W, 15–0°S (the southern branch of the dipole) and 60–20°W, 5–20°N (the northern branch of the dipole) (Moura and Shukla, 1981).
Figures 2 and 3 show the monthly correlation coefficients between the observed SSTA versus persistence and versus predicted SSTA by the model, for initial conditions between September and August. Although the numerical values of the correlation coefficients for the southern part of the dipole (Figure 2) are higher than those of the whole southern basin, the results are essentially the same as those for the whole basin, i.e. CCA beats persistence only for the forecasts done with initial conditions from September, and to some extent in November.

Over the northern branch of the dipole, the correlation coefficients of the CCA forecasts starting in September–January (Figure 3(a)–(e)) are considerably higher than those obtained by the persisted SSTA fields. In particular, the analysis of Figure 3 suggests that CCA predictions of SSTA for March, April, May and June are considerably better than persistence, and could be forecast by the model with moderate skill from September. Based on these results, new experiments were done for these lead/lags, but with the predictand field taken as the average of MAM SSTA. According to a Student’s $t$-test applied to the correlation coefficients calculated so far (Bendat and Persol, 1986), the values higher than 0.6 are significant at the 99% level and values higher than 0.4 are significant at the 95% level.

Figure 4(a)–(f) shows the correlation coefficients between observed and forecast MAM SSTA fields using initial conditions from September through to February. It is readily noticeable in Figure 4 that the correlation coefficients increase both in magnitude and area as the lead time for the prediction decreases. The best predictions are those using the initial conditions of January and February. Correlation coefficients higher...
than 0.6 (significant at the 99% level) and 0.7 occur over both the northern and the southern parts of the basin.

The correlation coefficients between the persisted SSTA from September to February and observed MAM SSTA are shown in Figure 5(a)–(f). The persisted SSTA has some skill for predicting MAM SSTA only during February. Comparing the results shown in Figure 4 with those in Figure 5, it can be seen that the forecast skill of the CCA model outperforms persistence for all lead times.

Moreover, the spatial structure of the correlation field of the forecasted SSTA (Figure 4) resembles the spatial structure of the observed SSTA much closer than that obtained by persistence. This fact alone suggests the advantages of using CCA forecasts of MAM SSTA over the tropical Atlantic as boundary conditions for the use of atmospheric general circulation modelling for seasonal climate predictions (Cavalcanti et al., 1998).

Figure 6(a)–(c) shows the first mode of the canonical variables (g-map, h-map and temporal series) obtained from the lagged experiment January–MAM. The higher values of the g-map over the tropical Pacific and northern tropical Atlantic (Figure 6(a)) suggest that SSTAs over these areas represent the best potential predictors for this lag, with the area over the equatorial eastern Pacific presenting more influence as a predictor than the tropical Atlantic. Also, it is noteworthy that the g values over the southern and northern tropical Atlantic are of opposite sign, suggesting that an SSTA dipole configuration during January is a precursor of further development of an inter-hemispheric SSTA gradient during the following months. This is also indicated in the h-map shown in Figure 6(b), which shows a dipole pattern over the tropical Atlantic. Still, the h-map in Figure 6(b) shows that the region over the northern tropical Atlantic is the most influenced by the variability
Figure 4. Correlation coefficient maps between observed SSTA versus SSTA forecast by SIMOC for the period MAM, with SSTA initial conditions of: (a) September; (b) October; (c) November; (d) December; (e) January; (f) February. Regions with correlation coefficients higher than 0.6 are shaded and have significance higher than 99% of the predictor, agreeing with Uvo et al. (1998). The canonical time series $u$ and $v$ shown in Figure 6(c) depict the time evolution of the $g$ and $h$ patterns (Figure 6(a) and (b)), and indicate the high correlation (0.95) between $g$ and $h$. Both the high correlation between $u$ and $v$ and the high frequency variability apparent in Figure 6(c) are suggestive of the eastern equatorial Pacific SSTA influence for the January forecast of SSTA over the tropical Atlantic during MAM. It is also interesting to notice that the seasonal dependence of the model’s forecast skill may be linked to the time of the year when the Pacific ENSO normally peaks (i.e. during the Northern Hemisphere winter).

There are suggestions in the literature that SSTAs over the eastern equatorial Atlantic are potentially predictable from wind stress variability over the western equatorial Atlantic during July–September, from the May–June predictor period (Servain and Arnault, 1995). However, using SSTAs as the predictor fields, our model did not present a significant skill over the equatorial Atlantic (figures not shown) compared with persistence. Zonal and meridional components of the surface wind, and sea-level pressure were also tested in this work as predictors to construct CCA forecast models. However, the forecast skill scores (not shown), as measured by correlation coefficient maps for the region of study as a whole, were lower than the ones for SSTA used as a predictor alone, and are not discussed further here.

In order to verify whether the SSTA forecast fields can reproduce the same patterns of spatial variability as observations, EOF patterns were calculated for observed and forecast MAM SSTAs over the tropical Atlantic. The first EOF pattern of the observed MAM SSTAs (Figure 7) shows a spatial configuration that is inter-hemispherically asymmetric, and is referred to in the literature as the Atlantic SSTA dipole pattern (Moura and Shukla, 1981; Servain and Legler, 1986; Servain, 1991), with opposite sign of variability over the northern and southern tropical Atlantic. The loading patterns of forecast MAM SSTA, which present the dipole pattern over the tropical Atlantic for the initial conditions from September to February, are shown in
Figure 5. The same as Figure 4 but for SSTA forecast by persistence of SSTA initial conditions

Figure 8(a)–(f). Each loading pattern is calculated based on the 15 maps (1979–93) obtained as a predicted SSTA field during the hindcast, i.e. 1 month of initial condition is taken for each year (1979–93), and the SSTA field is predicted. The set of results using one initial condition is used to calculate the EOF. As it can be seen clearly in Figure 8, the spatial structure of the model forecast variability resembles very well the observed pattern of variability on the northern and southern portions of the basin. This is consistent with the $g$-map shown in Figure 6(b). The southern pattern can be better represented by the forecasts made from November, January and February. These findings confirm the ability of the model to represent well the spatial structure of the observed SSTAs, a feature that is crucial to numerical climate predictions.

Examples of real predictions of the SSTAs for MAM 1998 are shown in Figure 9(a)–(c). During December 1997 (Figure 9(a)), SSTAs were slightly warmer than the mean over the northern tropical Atlantic, with anomalies ranging between 0 and 0.5°C over most of the area south of 20°N. Over the southern tropical Atlantic, SSTAs ranged around 1.0°C. Hence, the SSTA meridional gradient pointed southward during December 1997. On the other hand, the SSTA forecast by the model using December 1997 (Figure 9(b)) initial conditions shows a warmer northern tropical Atlantic, thus reversing the meridional SSTA gradient. This actual forecast for MAM 1998, based on the initial conditions of December 1997, was used as part of the boundary conditions for the general circulation model of the Center for Weather Forecast and Climatic Studies (CPTEC) to forecast the precipitation for the 1998 Nordeste rainy season (Cavalcanti et al., 1998). Pezzi et al. (1998) also show an SSTA forecast by the CCA model for MAM 1998 using different initial condition lead months (October 1997–January 1998).

4. CONCLUSIONS

The results of this work show that skilful SSTA forecasts over the tropical Atlantic using the CCA technique can be obtained. However, the forecasts’ skill score, as measured by correlation maps over the tropical
Atlantic presented a seasonality, with the best performance over the northern tropical Atlantic with a few months prior to MAM. It is suggested that the model skill score seasonality might be related to the influence of the tropical Pacific over the northern tropical Atlantic. Other seasons are potentially less predictable using the methodology followed in this work. The results have also indicated that the SSTA field is a better predictor of the whole tropical Atlantic than surface winds or pressure alone. Further work is needed to test the prediction potential of a combination of these variables.

Persistence of SSTAs over the southern tropical basin is hardly beaten by the statistical forecast scheme developed, but the spatial structure of the correlation fields between persisted and observed MAM SSTAs is
Figure 8. (a) The loading patterns of the SSTA forecasts for the period MAM which showed a dipole pattern, for the initial conditions of: (a) September; (b) October; (c) November; (d) December; (e) January; (f) February. The fraction of the variance explained for each EOF and the EOF number for each initial condition are indicated in the figure. Values greater than 0.04 and less than −0.04 are shaded.

Figure 9. SSTA anomaly (°C) (a) persisted from December 1997, (b) forecasted for MAM 1998 (initial conditions: December 1997), and (c) observed for MAM 1998.

very noisy. Meanwhile, the CCA forecast SSTA for MAM shows an inter-hemispheric asymmetric mode of variability, which is similar to the mode found on observed data. The results show that the spatial structure of SSTA fields over the tropical Atlantic for the average period of MAM can be predicted with reasonably good skill from September to February.

As this model forecasts the SSTA spatial structures for several lags better than persistence, and because the predicted field is available for every grid point of the domain, this represents a furthering of SSTA forecasting capabilities over the tropical Atlantic, and a step forward towards numerical climate forecasting. The SIMOC model has been run experimentally at CPTEC to generate SSTA forecasts over the tropical Atlantic since November 1996 (Cavalcanti et al., 1998).
REFERENCES


