ELSEVIER

Dynamics of Atmospheres and Oceans 22 (1995) 91–114

dynamics of atmospheres and oceans

Application of statistical techniques to the analysis and prediction of ENSO: Bayesian oscillation patterns as a prediction scheme

A. RuizdeElvira *, M.J. OrtizBeviá

Departamento de Física, Facultad de Ciencias, Universidad de Alcalá, Campus, 28880 Alcalá de Henares, Madrid, Spain

Received 31 August 1993; revised 17 June 1994; accepted 5 July 1994

Abstract

Here we study the low-frequency variability of the tropical Indian and Pacific basins with a new statistical technique, Bayesian oscillation patterns (BOP). To describe the climatic system in this region, zonal wind and sea surface temperature (SST) are the selected variables. Their variability can be explained in terms of a reduced number of frequencies and spatial patterns. These are identified for each field by a statistical procedure. With the help of the patterns and the frequencies a predictive scheme is devised and applied in two forecast experiments. In the first, zonal wind anomalies are predicted using patterns and frequencies identified in the wind field. A more sophisticated scheme, a linear model which includes non-harmonic oscillations and interactions between patterns, is used when forecasting SST seasonal anomalies in the Niño3 region. In this case, the predictors include the values of the frequencies identified in the BOP analysis of both wind and SST fields, and the corresponding patterns.

1. Introduction

In the anomalous climatic variability that takes place in the tropical Pacific basin known as ENSO (El Niño–Southern Oscillation), two features are relevant. One is the magnitude of the atmospheric and oceanic anomalies, and the other is the quasi-periodicity of the phenomenon. The first determines the importance of the climatic signal on the global scale, whereas the second points to the possibility of predicting it. ENSO occurs at time intervals between 2 and 8 years. Such a

* Corresponding author.

SSDI 0377-0265(94)00395-5

considerable time span points to the need for more than one time scale to describe and predict the phenomenon.

The need for several time scales to explain ENSO variability had been noted already in the pioneering work of Kidson (1975). When anomalies of sea level pressure (SLP) and sea surface temperature (SST) in the region were analysed in terms of empirical orthogonal functions (EOFs), the time coefficients or principal components of the first few EOFs did show important spectral peaks for a limited number of frequencies. Owing to the short length of the records, the significance of these peaks could not be easily assessed. Furthermore, there was the problem of the ambiguity of the response when statistical tests (Preisendorfer, 1988) aimed at determining the number of relevant EOFs are applied to observed climatic records. As has been pointed out (Barnett, 1991), conventional spectral analysis may not be strictly appropriate for studying ENSO variability.

Principal interaction patterns (PIP) and principal oscillation patterns (POP), the statistical techniques proposed by Hasselmann (1988), are an alternative to the classical identification of relevant frequencies by means of Fourier analysis. In the PIP (or POP) technique a first selection of a subspace of interest in the space of field variables is performed by means of EOF analysis. This is a separation based on an energy criterion. Although the first few EOFs may pick the most energetic phenomena, it is difficult to separate noise and oscillations in the next EOFs. To determine the relevant frequencies and patterns of variability via the POP analysis, the time series must satisfy the hypothesis of being a realization of an auto-regressive process of order n (AR(n)) driven by white noise.

The POP scheme has been applied to the prediction of the Southern Oscillation (Xu and Von Storch, 1990) using observed SLP data, and of ENSO (Latif and Flügel, 1991), using the output of a general circulation model (GCM) for the region. Predictive schemes based on the POP analysis retain usually only one of these frequencies and a pair of patterns. When dealing with observed climate time series (very noisy), this reduction of the number of degrees of freedom of the system is arbitrary.

Other statistical techniques used for the identification of the relevant time scales present in ENSO variability give a more complex view. For instance, Rasmusson et al. (1990) used singular spectrum analysis (SSA) to study ENSO variability in the wind field of the tropical Indo-Pacific region. They found that the spectra contained both a biennial oscillation and also a signal in the band of 48–60 months. The latter signal, however, was more difficult to identify than the former. Similar results were obtained by Lau and Sheu (1988). SSA does not assumes an AR(n) form, but implies an infinite and ergodic time series and needs several strong assumptions on the nature of the noise. Moreover, as with any method that extracts information from eigenvalue analysis of matrices obtained as samples of some hypothetical ideal multivariate time series, there are problems with the assigning of the largest eigenvalues to the signal and with the degeneracy of eigenvalues.

Although SSA performs better than the traditional Fourier analysis at separating closely spaced relevant spectral peaks, SSA does suffer some of the problems

connected with this analysis, such as the requirement of stationarity and the limitation to situations of high signal-to-noise ratios. Besides, techniques related to spectrum analysis share a problem when applied to short time series with noise. The problem is that it is rather difficult to decide if a peak in the spectrum is real or if it is a consequence of the short length of the series and the presence of noise. Furthermore, interesting signals may occur with less energy than the noise, so a separation into EOFs is, at the very least, doubtful.

In the problem we address here, the question is not so much to determine the spectrum of a time series, but to locate some low-frequency signals possibly embedded in the data. The problem is then no longer one of spectral estimation but one of what Marple (1987) called 'parameter estimation method'.

The Bayesian signal analysis (BSA) we use here is not restricted to time series of special characteristics or to white noise. BSA is designed as an optimal method to use for analysing short time series and can work with signal-to-noise ratios as low as 0.6. In BSA no hypotheses are made about the actual series belonging to any hypothetical ensemble or infinite series. Only the data at hand are used, and the question addressed is: What is the probability of some a priori signals being contained in the data? As we are asking about probabilities, we have the advantage that we can obtain a measure of the accuracy of our estimates, which the least-squares method does not give at all, and which maximum likelihood gives via a complex procedure not related to the specific data set at hand.

BSA is a development of Jaynes' Bayesian methods (Jaynes, 1985), which have been applied to statistical mechanics, quantum electrodynamics and statistics. Their main use has been in image reconstruction problems and astrophysics (see Justice (1986) and references therein), modelling systems whose structure changes in time (dynamical modelling), and in applications to econometrics and Kalman filtering (Spall, 1988). In particular, BSA is an application of Bayesian methods to spectrum analysis (Bretthorst, 1988). The determination of Bayesian oscillating patterns (BOP) by using BSA (Bayesian spectrum analysis) was proposed by RuizdeElvira (1993), who applied it to an analysis of the chaotic solution of the nonlinear equation of Franceschini and Tebaldi (1989). It was shown by RuizdeElvira that although Fourier analysis of the solution indicated the presence of a fundamental frequency and its three odd harmonics, a reconstruction of the solution using these four frequencies failed to reproduce its characteristic behaviour. On the other hand, BSA could fit this behaviour very well.

In the present paper, the BOP analysis is applied to real climate data. Time scales and patterns relevant for the evolution of ENSO are determined. We obtain a measure of the precision of the time scales, and an improvement in the hindcast skill for the analysis of wind and SST anomalies in the Indo-Pacific region with respect to other techniques. Other relevant questions tackled here include propagation of the signal between basins and correlation between wind and SST. An indication of the performance of the method is inferred from the success in the forecast.

Details of the statistical method are given in Section 2 and completed in the Appendix. The results of the application of this technique to the analysis of

observed data (zonal wind and SST) appear in Section 3. In Section 4, we describe the forecast procedure, with details of the first forecast experiment (zonal wind) in 4.1, and of the second experiment (SST in the Niño3 region) in 4.2. Conclusions are presented in Section 5.

2. Bayesian oscillation patterns

As we have already mentioned, for short time series including noise, the procedure that is optimum in the sense that it does not throw away any information present in the series, and does not require exclusively harmonic oscillations, is BSA (Jaynes, 1987). This method can be used without any previous reduction of the degrees of freedom of the system, and, for an a priori chosen model, determines both the number of significant signals and the parameters of the model that maximize the probability of the model fitting the data.

Bayesian methods include both maximun likelihood estimation, (MLE), when using uniform priors, and least squares estimation, when we can assume that the noise is Gaussian. However, the Bayesian methods are superior to these two techniques, as the Bayesian methods allow for non-uniform and informative priors when these two are needed (as in the case of the prior for the variance σ), their treatment of non-interesting parameters (nuisance parameters) is very easy, and in their framework it is straightforward to obtain the precision or confidence intervals in the determination of the parameters. Furthermore, in Bayesian analysis one uses exclusively the data at hand, and there is no need to make complicated arguments about the plausibility that a model has high probability for producing the data, as in the MLE. Indeed, if we propose a certain model and we obtain a probability of, say, 0.98 that a parameter b has the value b_0, as compared with probabilities lower that 0.4, say, for all other values, we can easily accept that this value b_0 is the correct one for the parameter.

To apply Bayesian techniques we need to define the possible models we would like to test for the presence of some signals in the data. These models should be oscillations (harmonic, anharmonic, elliptic or Bessel functions, etc.), as pertains to solutions of the Navier–Stokes equations.

If we have data at the n points of a (possibly irregular) grid, depending on time, $\boldsymbol{d}(t_i)$, a common supposition in statistical modelling is that these data can be modelled as a regular part and noise:

$$\boldsymbol{d}(t_i) = \boldsymbol{m}(t_i) + \text{noise}; \qquad i = 1, \ldots, m \tag{1}$$

$$\boldsymbol{m}(t_i) = \sum_{\nu=1}^{N} \boldsymbol{a}_\nu G_\nu(t_i; \{\omega\}) \tag{2}$$

where $\{\omega\}$ is the set of parameters of the model, $\boldsymbol{d}$ is the n-dimensional data vector, and $\boldsymbol{m}$ the n-dimensional model vector. The N spatial coefficients $\boldsymbol{a}_\nu$ are n-dimensional vectors, which will be called patterns, the N functions G_ν are the 'a

priori' model functions, and the 'noise' is all that part of the variability of the data that we are not interested in modelling.

To focus on a particular case, we would like to understand a set of n time series representing some multivariate climate data as a superposition of a (small) number of spatial patterns, each pattern oscillating with a fixed time scale (RuizdeElvira, 1993). In this case, the parameters $\{\omega\}$ are frequencies. As an example, the model that we will propose for the zonal wind is

$$\boldsymbol{m}(t_i) = \sum_{\nu=1}^{N} \boldsymbol{a}_\nu \cos(\omega_\nu t_i) + \boldsymbol{b}_\nu \sin(\omega_\nu t_i) + \cos(\omega_a t_i) \sum_{\nu=1}^{N} (\boldsymbol{c}_\nu \cos(\omega_\nu t_i) + \boldsymbol{f}_\nu \sin(\omega_\nu t_i)) \tag{3}$$

that is, harmonic oscillations modulated by the annual cycle. Here ω_a denotes the annual angular frequency and the n-dimensional vectors $\boldsymbol{a}_\nu$, $\boldsymbol{b}_\nu$, $\boldsymbol{c}_\nu$ and $\boldsymbol{f}_\nu$ are the Bayesian oscillation patterns (BOP). Patterns $\boldsymbol{c}_\nu$ and $\boldsymbol{f}_\nu$ are the cyclostationary corrections to patterns $\boldsymbol{a}_\nu$ and $\boldsymbol{b}_\nu$. As we have already indicated, what we understand by noise is not only statistical noise (in physical terms, turbulence, as our data are wind velocities, i.e. fluid motion), but also some other terms that we choose to disregard for the time being.

As is shown in the Appendix, the probability for the set of parameters $\{\omega\}$, given the data, is

$$P(\{\omega\} \mid d) \propto \left(1 - \frac{n\overline{\boldsymbol{h}^2}}{m\overline{\boldsymbol{d}^2}}\right)^{(N-m)/2} \tag{4}$$

where

$$\overline{\boldsymbol{h}^2} = \frac{1}{n}\frac{1}{N}\sum_{k=1}^{n}\sum_{\nu=1}^{N} h_{k\nu}^2 \tag{5}$$

$$\overline{\boldsymbol{d}^2} = \frac{1}{n}\frac{1}{m}\sum_{k=1}^{n}\sum_{t_i=1}^{m} d_{kt_i}^2 \tag{6}$$

and $h_{k\nu}$ is the projection of $\boldsymbol{d}$ onto a linear combination of the model functions $\boldsymbol{G}$.

For different values of $\{\omega\}$ (4) gives different numbers. The optimal parameters $\{\omega_0\}$ are chosen as those for which probability (4) is maximum.

Less straightforward is the criterion to decide the order N of the model. We have used an empirical iterative procedure analogous to the Monte Carlo method. First we created l ($l = 1000$) time series as a superposition of harmonic signals with r frequencies and noise of different statistical characteristics. The signal-to-noise ratios of these series varied between 2 and 0.3.

To determine the number N of harmonic signals contained in the time series (N should equal r), we calculate $\ln P$ for frequencies between (2 month)$^{-1}$ and (222 month)$^{-1}$, using Model (2) with only one harmonic signal: $G_1(\{\omega\},t) = \cos(\omega t)$. These frequencies correspond to our 444 month long time series of monthly averaged data. We identify the frequency that maximizes $\ln P$, hereafter called

ω_{max}. We look for all frequencies ω for which $\ln P(\omega) \geqslant \beta \ln P(\omega_{max})$. For the above synthetic series we find that for signal-to-noise ratios greater than 0.6, we can recover the original r frequencies when β is greater than 0.5.

Now we apply the criterion to the time series of real data for each grid point. First we select N' relevant frequencies with Model (2) and $N = 1$. Then we use Model (3) with N' harmonic signals corresponding to these frequencies, and determine the signal-to-noise ratio following (A16). For all grid points of our analysis the value of the signal-to-noise ratio exceeds 0.7. Therefore we feel confident that the a priori model can be accepted with N' signals.

3. Data and results of the analysis

The data used here consist of 98 time series of surface zonal wind seasonal anomalies over the tropical Indian and Pacific Oceans and 98 time series of SST seasonal anomalies. Both series comprise 37 years of monthly averages from 1950 to 1986. The wind data set was constructed by T.P. Barnett from several sources and has been described by Barnett (1983). The SST data are from the Comprehensive Ocean–Atmosphere Data Sets (COADS) (Fletcher et al., 1983) from 1950 to 1986. Both data sets were kindly provided by T.P. Barnett.

The original grid points of zonal wind and SST do not coincide. Therefore, an interpolation had to be performed to match both wind and SST data to the grid points represented in Fig. 1. No other manipulation or filtering was performed on the data.

3.1. Zonal wind

As the seasonal dependence of wind anomalies is important, we used Model (3) including a cyclo-stationary dependence. The optimal model is obtained for six frequencies corresponding to time scales $T = \{19,23,26,43,59,73\}$ months.

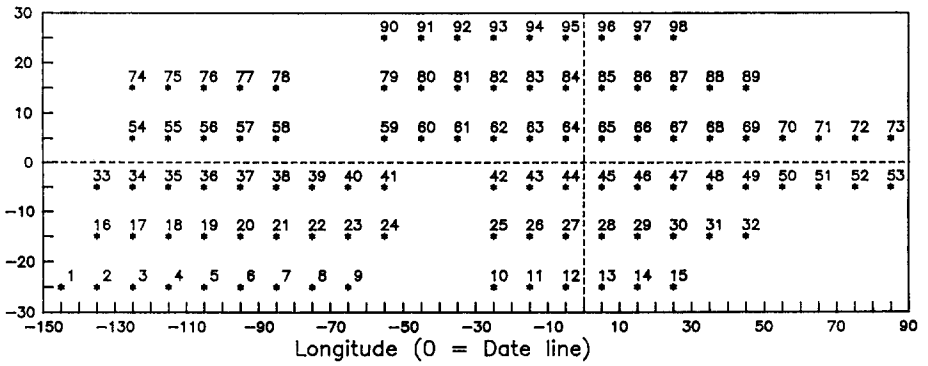

Fig. 1. Grid points for data and model in the Indo-Pacific basin.

Among these are a quasi-biennial oscillation and a 5 year oscillation. The patterns corresponding to the 5 year time scale show the maximum probability $P(\{\omega\} \mid d)$ (Eq. (4)), and carry the maximum variability.

Some of the patterns corresponding to these time scales can be understood in terms of stationary oscillations, whereas others, as for instance the patterns corresponding to the 5 year time scale, can be understood as propagating oscillations. The six double pairs of patterns account for 38% of the field variance. The signal-to-noise ratio is 0.8.

The significance of the model can be assessed using several techniques. A good measure is obtained from the consideration of the joint probability distribution of the estimated model coefficients (Barnett and Hasselmann, 1979). If, for each grid point, we collect the N estimated model coefficients into an N-dimensional vector $\tilde{\boldsymbol{a}}$ and the true coefficients into $\boldsymbol{a}^t$, then the probability distribution that the coefficients have a value $\tilde{\boldsymbol{a}} \pm \mathrm{d}\tilde{\boldsymbol{a}}$, given $\boldsymbol{a}^t$, is approximately Gaussian:

$$P(\tilde{\boldsymbol{a}} \mid \boldsymbol{a}^t)\mathrm{d}\tilde{\boldsymbol{a}} = (2\pi)^{-n/2} \mid M \mid^{-1/2} \exp(-\rho^2/2)\mathrm{d}\tilde{\boldsymbol{a}} \tag{7}$$

where the statistic ρ^2 is given by

$$\rho^2 \equiv \sum_{i,j=1}^{n} M_{ij}^{-1} \delta a_i \delta a_j \tag{8}$$

$$M_{ij} \equiv \langle \delta a_i \delta a_j \rangle \tag{9}$$

and

$$\delta a_i \equiv \tilde{a}_i - a_i^t \tag{10}$$

are the differences between the determined coefficients $\tilde{a}_i$ and the unknown true values a_i^t. We can test if there is a significant probability that the true values are zero. As in this first study we have assumed an uniform prior probability for the parameters a_i^t, in this case the probability $P(\boldsymbol{a}^t \mid \tilde{\boldsymbol{a}})\mathrm{d}\boldsymbol{a}^t$ is numerically equal to the likelihood $P(\tilde{\boldsymbol{a}} \mid \boldsymbol{a}^t)\mathrm{d}\tilde{\boldsymbol{a}}$ (Appendix A).

Then, given the estimated coefficients $\tilde{a}_i$, we substitute $a_i^t = 0$, $i = 1, \ldots, N$ into (10). In this case $\delta\tilde{a}_i$ are simply equal to $\tilde{a}_i$. We will obtain a high probability that $a_i^t = 0$ if ρ^2 is small. Indeed, the distribution of ρ^2 is really a Hotelling distribution, as $\boldsymbol{M}$, the covariance matrix for the coefficients, must be estimated from the data. However, for reasonably large N (as is the case here), this distribution can be approximated by a χ_N^2 distribution with N degrees of freedom. We are testing for $N = 24$ degrees of freedom (four coefficients for each of the six frequencies). The χ^2 95% confidence level for 24 degrees of freedom is 36.42.

In our case, we find an average across grid points of $\overline{\rho^2}$ of 206, with a minimum of $\rho^2 = 70$, a maximum $\rho^2 = 408$. We conclude that for all the grid points, the probablity that the parameters a_i are zero is negligible, and therefore that the model is significant for all points.

The errors for the estimated time scales (in months), determined using (A19) are:

T	19	23	26	43	59	73
ΔT	0.35	0.46	0.41	0.80	1.1	3.0

The patterns accounting for the maximum variance are the four corresponding to the time scale of 59 months (stationary and cyclostationary correction). The evolution of the stationary patterns of this set is represented in Fig. 2. Assuming that for $t = t_{in}$, the state of the system can be represented by Pattern 1 (lower left-hand corner), then at $t = t_{in} + T/4$, (15 months later) the field is represented by Pattern 2 (upper left-hand corner). At $t = ti_n + T/2$, (30 months later) the field is represented by minus Pattern 1 (upper right-hand corner), and for $t = t_{in} + 3T/4$ (44 months later), by minus Pattern 2 (lower right-hand corner).

This sequence represents the propagation of a signal (positive wind anomalies) from the south equatorial Indian Ocean to the north equatorial western Pacific Ocean, then to the equatorial central Pacific Ocean, where it becomes amplified, and finally disappears during the last 15 months of the 59 month cycle, after which it starts again.

A further indication of the model performance can be obtained through a hindcast of the data. Two graphical comparisons of model hindcast with data are noteworthy. The first is a series of snapshots between August 1982 and May 1983 (Fig. 3). Prominent features in the data (right panels) are the large positive anomalies in the equatorial western Pacific in August 1982, the eastward propagation of these anomalies during the following months and the damping of the anomalies in May 1983 (the anomalies disappear some months later, not shown).

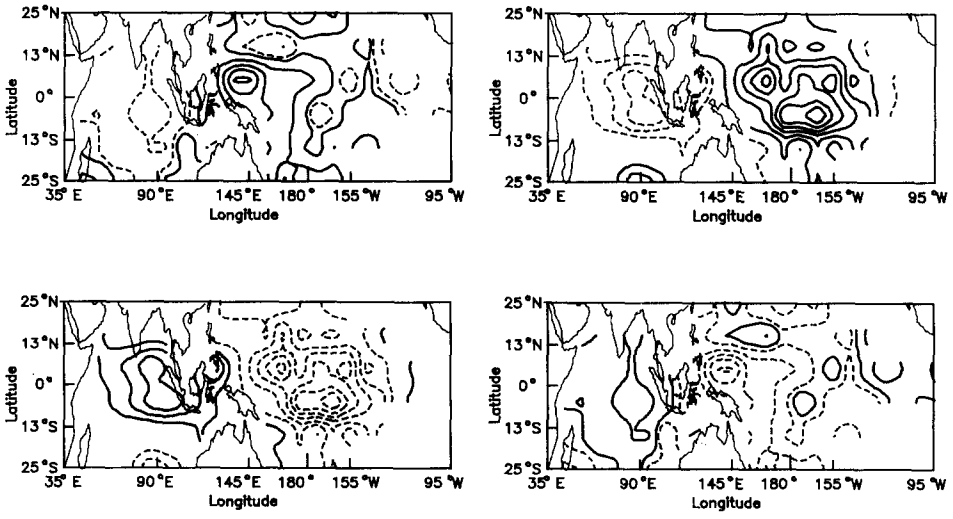

Fig. 2. Evolution sequence for the BOP stationary pattern of the zonal wind corresponding to the time scale of 59 months. Minimum value: -6.8 m s^{-1}, maximum: 4.25 ms^{-1}. Contour interval: 1 m s^{-1}. The sequence is lower left, upper left, upper right, lower right.

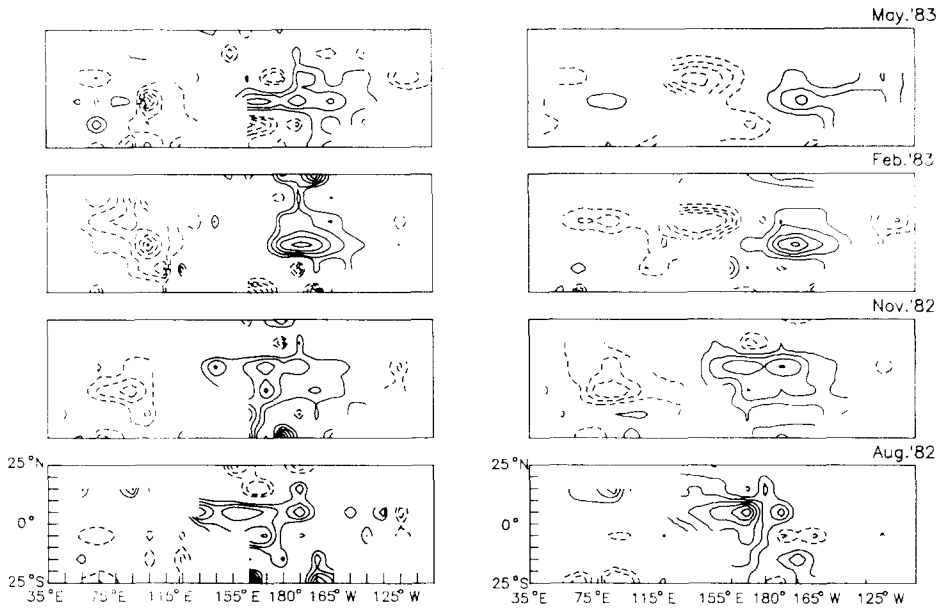

Fig. 3. Snapshots of the zonal wind for the interval summer 1982 to spring 1983. Right panels: data. Left panels: BOP model. Minimum contour: -3.5 m s^{-1}. Maximum contour: 4.5 m s^{-1}. Contour interval: 1 m s^{-1}.

These features are well known, and indeed have been modelled by a variety of methods, some statistical, some dynamical. In all of these methods, the hindcast starts with known conditions around January 1982 and follows the evolution for 24 months.

The model snapshots (Fig. 3, left panels) represent the model evolution with initial values in January 1950. We point out the appearence of a positive anomaly in the west equatorial Pacific in August 1982, its propagation to the central equatorial Pacific in February 1983, with the generation of negative anomalies in the Indian Ocean, and the final weakening of the anomalies in May 1983. The hindcast captures the most important features of the data evolution.

The second graphical comparison is a time–longitude diagram of the equatorial band of the basin. As before, the model evolves from 1950. We have plotted (Fig. 4, left panel) the time interval 1969–1987. If we compare the model evolution with the data (Fig. 4, right panel), we can see that the appearance and subsequent eastward propagation of positive wind anomalies in the model in years 1971, 1976, 1982 and 1986 closely follows the data. An interesting feature to notice here concerns the propagation of the anomalies from the Indian Ocean to the Western Pacific prior to the start of the events. This propagation seems to be connected with the pattern associated with the 59 month time scale, and appears in the analysis as a common feature for all events, whereas it appears only in some of the events given by the data. We conclude that the model captures some significant features of the evolution of wind anomalies in the Indo-Pacific basin, and that

therefore these features can be understood as a sequence of coupled oscillations, as is proposed in this model.

3.2. SST anomalies

The analysis proceeds along the same path as for the U-wind case but, as tropical SST are less sensitive to the seasonal cycle than the winds, the a priori model does not include the annual cycle modulation:

$$m_2(t) = \sum_{\nu=1}^{n} a_\nu \cos(\omega_\nu t) + b_\nu \sin(\omega_\nu t) \tag{11}$$

Errors and statistical significance have similar values as in the U-wind field case and the model is validated in all the region. The time scales found are: $T = \{21,25,34,42,59,73\}$ months. If we take into account the margins of error in their determination, the only significant difference with the U-wind case is in the time scale of 34 months.

We conclude that the coupled ocean–atmosphere system in the Indo-Pacific basins exhibits fluctuations on time scales of both 2 and 5 years, as was expected from earlier studies. Such long-time scales can be physically explained only if thermodynamic processes involving energy exchanges between air and sea are present.

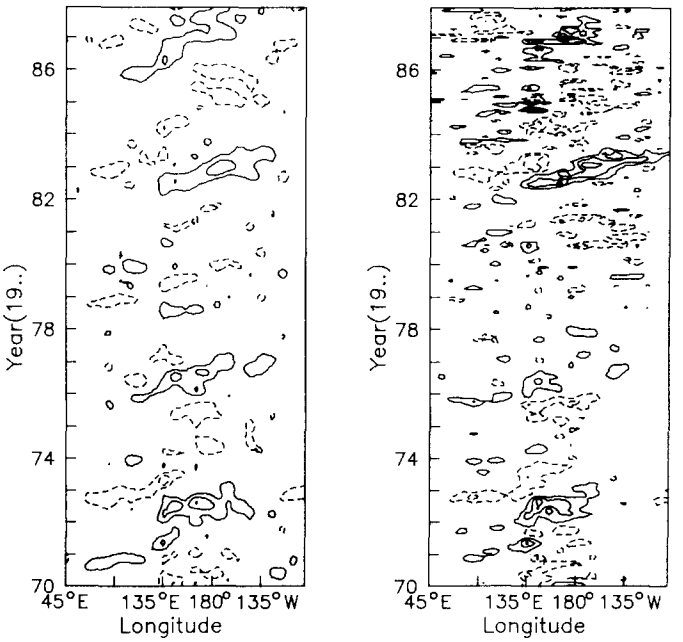

Fig. 4. Time–longitude diagram for the zonal wind in a band around the Equator. Right panel: data. Left panel: BOP model. Minimum contour: -3.6 m s^{-1}. Maximum contour: 4.5 m s^{-1}. Contour interval: 1.2 m s^{-1}.

3.3. Relationships between fields

To study the relationship between SST and zonal wind fields, time lag cross-correlations are computed between the modelled time series. Two regions, one with 36 grid points for the SST and another with 28 points for the zonal wind, show an absolute value of the correlation greater than 0.4 for some time lags (Fig. 5). The maximum correlation is obtained when the wind lags the SST by 1 month; there is appreciable correlation until 9 months, and between 28 and 36 months, which is not surprising owing to the periodic nature of the model.

The existence of several time scales with nearly equal values points to the presence of nonlinear oscillations in the system. The possibility of nonlinear interactions between oscillations with different time scales has been explored before (Barnett, 1991). To look at this possibility, four additional functions have been included into the model m_2. These functions are the products of the harmonic functions whose patterns account for the greatest variance for each field (those of 59 and 42 and 23 months, i.e. $G_{59,42} \propto [\cos(\omega_{59}t)\cos(\omega_{42}t)]$, etc.).

The peaks in the time correlations between patterns are -0.57 for the patterns in zonal wind and SST corresponding to $G_{59,42}$, and -0.20 for the $G_{59,23}$ pattern. The regions for which the patterns show correlation in the case $G_{59,42}$ are indicated in Fig. 6. This correlation is obtained between grid points displaced 10° with the U-wind to the east.

4. Forecast

By 'forecast' we understand the prediction of the evolution of a field in a region of the basin for the time interval t_i to t_f using an a priori model, whose

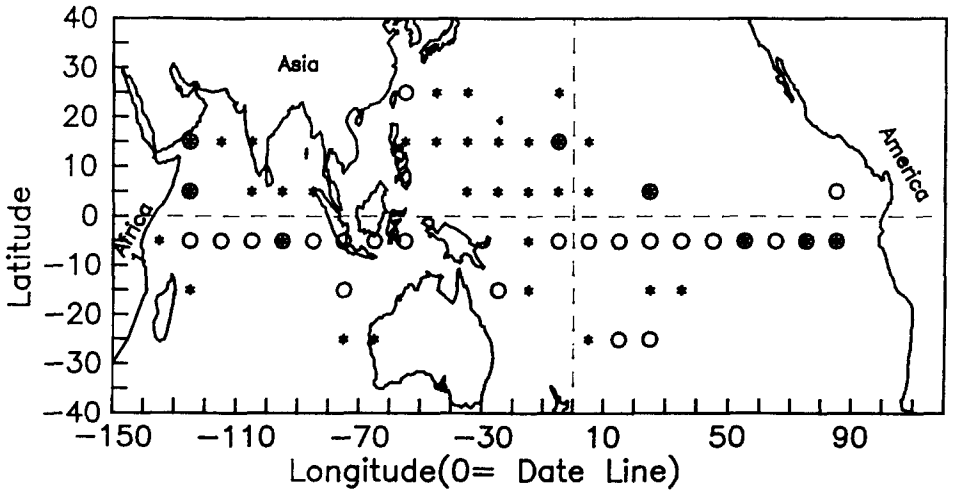

Fig. 5. Regions of significant correlations (linear model) between SST seasonal anomalies (asterisks) and zonal wind seasonal anomalies (circles).

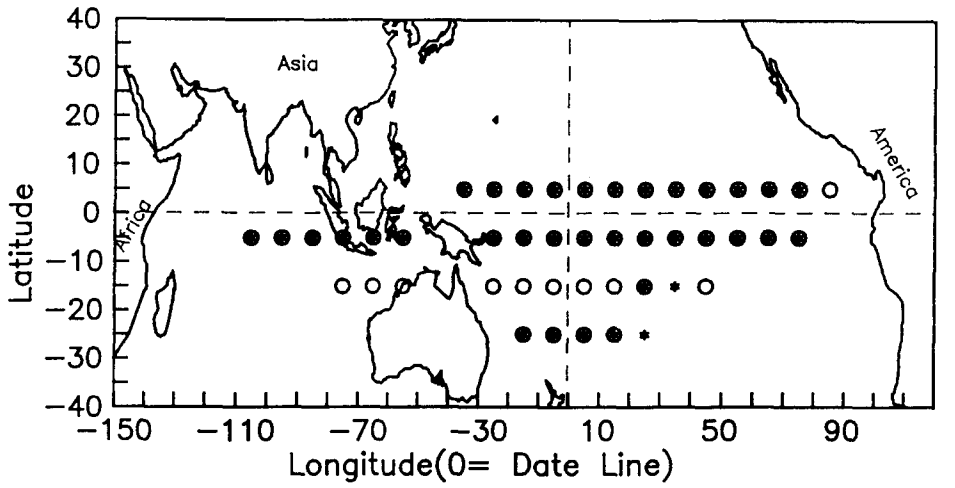

Fig. 6. Regions of significant correlations (nonlinear model) between SST seasonal anomalies and zonal wind seasonal anomalies. Circles denote the G1 function, asterisks the G2 function (see text).

parameters are determined using the data from the first point of the series, t_0, to $t_i - 1$. Once the model parameters are determined and the patterns are calculated, the model is left to evolve in time from t_i to t_f. A final test of the goodness of any statistical model is its ability to forecast. The dynamics of the system force oscillatory solutions, but random forcings must produce unexpected phase displacements. Can our model capture these displacements?

4.1. Forecasting the zonal wind

We forecast the zonal wind using Model (3). The predictand is in this case the zonal wind at a grid point, and the predictors are the zonal winds at all grid points in the basin. Our forecast experiment for the zonal wind consisted of predicting 24 months of wind at every grid point, with three starting times t_i: January 1983, 1984 and 1985. Using the 98 time series between 1950 and each of these starting times, we determined the time scales of oscillation of the wind in that period. To deal with the phase problem, we determined the patterns by fitting the model with these time scales to the data in the 4 years previous to t_i. We then used patterns and model functions for the 24 months following t_i. In Fig. 7 we show the result of the forecast for two grid points, and for the forecast periods 1985–1987 and 1986–1988.

The criterion used to calculate the skill of the forecast is based on the relative error between prediction and data weighted by a cut-off function. The inclusion of this weight function is justified by our interest in forecasting large departures from

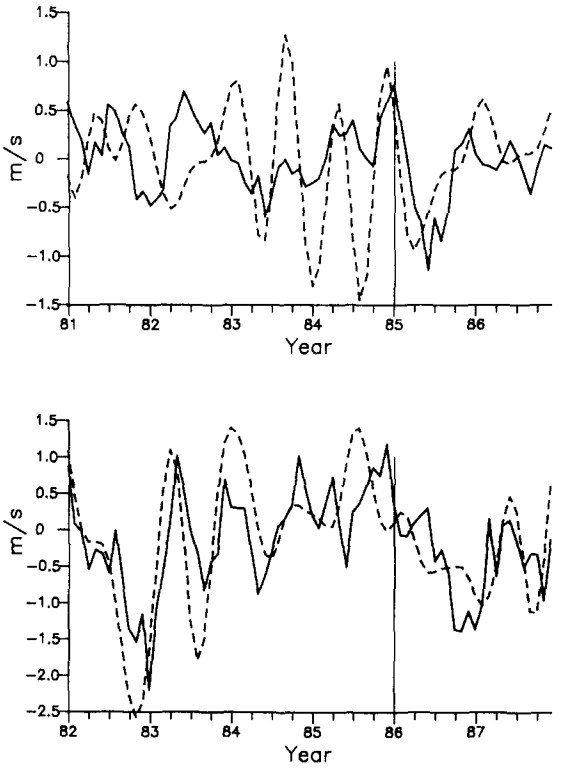

Fig. 7. Two instances of zonal wind forecast. Upper panel: grid point at 25°S,105°E; lower panel: grid point at 5°S,85°E. Forecast starts at the vertical line.

the norm rather than small values of the anomalies. Therefore the relative error ϵ^r_{ik} for each lag i and each grid point k is estimated by

$$\epsilon^r_{ik} = \left[1 - \exp\left(-4z_i^2\right)\right] \frac{(m_{ik} - d_{ik})^2}{d_{ik}^2} \tag{12}$$

where

$$z_i = d_i / D_i, \qquad D_i = \max_k (d_{ik}) \tag{13}$$

The weighted relative error, averaged across the 98 grid points and the three forecast intervals, is shown in Fig. 8(b), together with the more usual correlation skill (Fig. 8(a)). To estimate the significance of this measure we will assume that neighbouring grid points are correlated, and therefore we can consider a sample of 98/4 independent points and three independent time intervals, giving us a value of 0.21 above which correlation skill can be considered significant at the 95% level. In this case, though low, the skill of the forecast is significant for 24 months whereas persistence is not significant after 3 months.

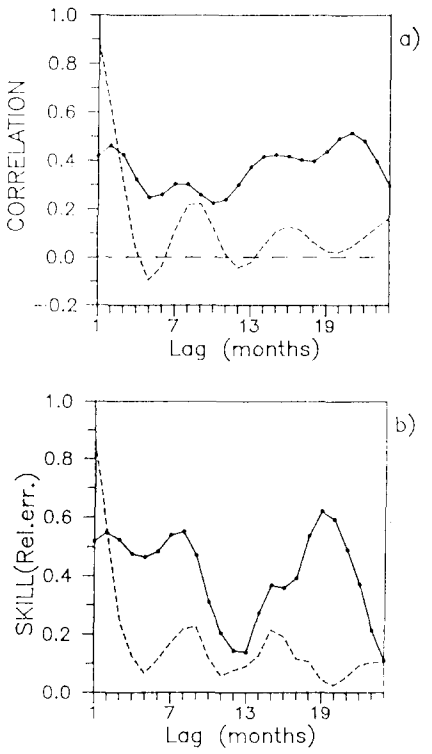

Fig. 8. (a) Average correlation skill for the zonal wind forecast. Continuous line: BOP model; dashed line: persistence. (b) Average relative mean square skill. Continuous line: BOP model; dashed line: persistence.

Another indication of the ability to forecast is given by the timelag longitude diagram for some of the latitude bands (Fig. 9). Here we can see that at some points the correlation skill reaches rather high values for considerable time lags. The dynamical causes of this are unknown, but its implications seem worthy of further research.

4.2. *Forecasting the Niño3 anomalies*

We have performed an experiment of forecasting the Niño3 SST anomalies for 36 months starting at each season from January 1963 to October 1983, and including the year 1956. Only in this case, and to help to solve the phase problem, the predictor time series were bandpass filtered retaining time scales between 10 and 84 months. Both zonal wind and SST were used as predictors and we choose the following a priori model:

$$m_3(t) = \sum_{\nu} a_\nu \cos(\omega_\nu t + \phi_\nu) + b_i \cos^3(\omega_i t + \psi_i) \tag{14}$$

where the parameters ω_ν were determined as above.

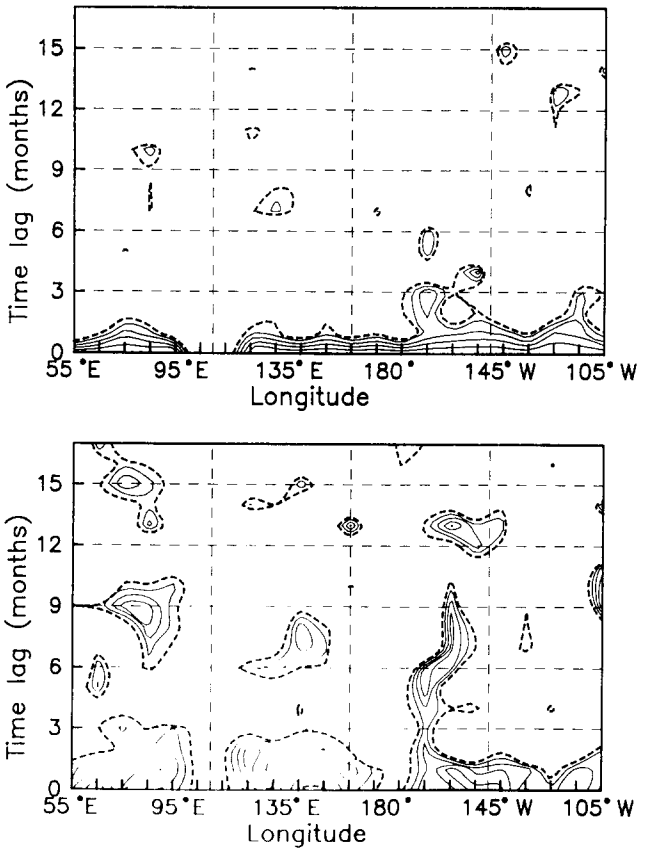

Fig. 9. Time lag–longitude diagram of the correlation skill for the zonal wind forecast using the BOP model in the 5°N band. Upper panel: persistence. Lower panel: BOP model. Minimum contour (dashed): 0.5. Contour interval: 0.1. Maximum correlation skill obtains with a value of 0.8 on the 165°W meridian for lags of 9 and 13 months.

The model was chosen as the simplest one that uses not only pure harmonic signals, but also includes signals steeper than harmonics, to improve the modelling of the sharp rises at some intervals in the SST series and improve the forecast over that obtained using pure harmonics. This model appeared in an analysis of the solutions to the equations of the dynamical system proposed by Franceschini and Tebaldi as mentioned in the Introduction. The patterns selected following the criterion of Section 2 and determined using the data prior to the forecast starting time t_i are the ones used as predictors. The time scales selected for each forecast varied in number between one and six, and the signal-to-noise ratio in the model determination varied between 1.3 and 0.71. Therefore, and for each forecast, the model used to obtain the predictors was validated.

Results of five forecasts are presented in Fig. 10. Three of these forecasts correspond to the canonical cases proposed by Latif et al. (1994). The forecasts

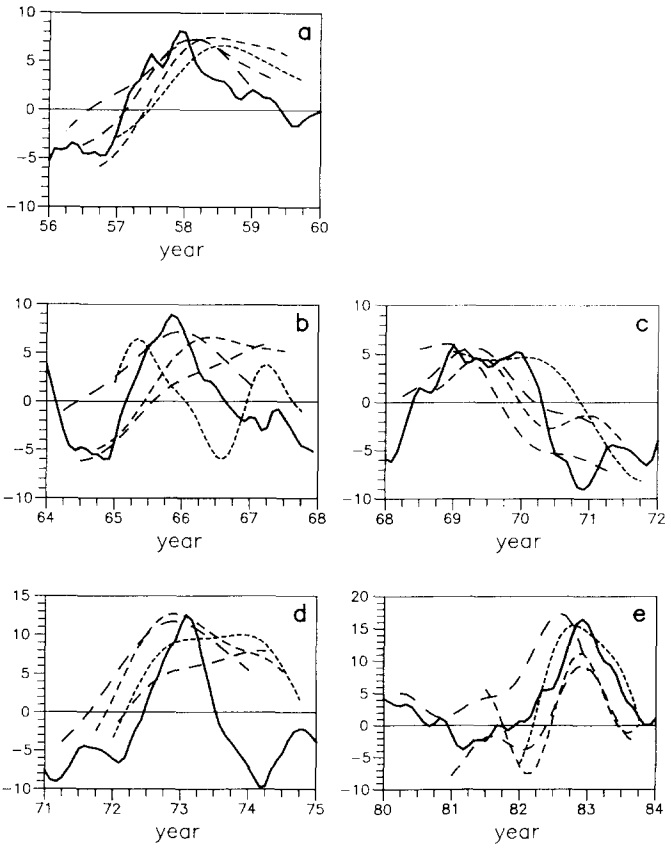

Fig. 10. SST anomalies forecast for the Niño3 region. Forecast periods: (a) April 1956–December 1959; (b) April 1964–December 1967; (c) April 1968–December 1971; (d) January 1971–October 1974; (e) April 1980–January 1984. Long to short dashes indicate forecasts starting in spring, summer, autumn and winter.

started at the first month of each of the four seasons, spring, summer, autumn and winter. We can observe how after an interval of 3 months the forecast follows well the observed data in its oscillatory behaviour as well in its order of magnitude, with evident degradation at the end of the forecast period.

We assess the scheme performance by plotting the curve of forecasts for different values of lead time τ. That is, for $\tau = 9$ months, for instance, we plot, for each time t, the value predicted by the model at time t when we use all the data until time $t - 9$ months. In Fig. 11 we show three plots corresponding to lead times of 9, 21 and 36 months. We can observe that the 1972–1973 El Niño is rather well represented for lead times until 21 months, and the representation degenerates for longer lead times in the sense that it is displaced by around 9 months, though no large negative value obtains for this time period. For El Niño 1982–1983 the worst prediction occurs between $\tau = 15$ months and $\tau = 24$ months (not shown), improv-

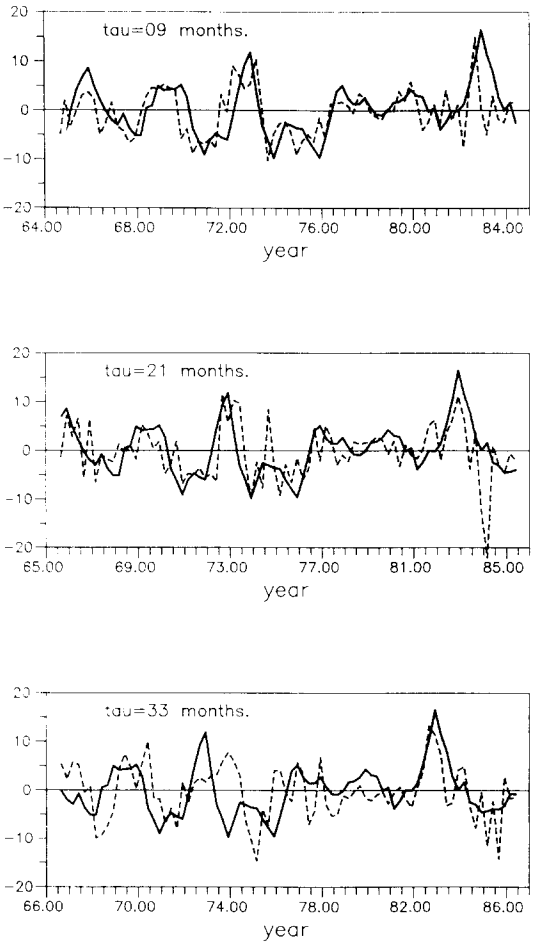

Fig. 11. SST anomalies forecast for the Niño3 region for three different lead times, tau.

ing again for lead times between 2 and 3 years, but for all lead times we obtain peaks of the right magnitude within the interval October 1982–April 1983.

We can conclude that the complex mechanisms that control the evolution of ENSO, or the solutions to the equations that represent them, are well modelled by a superposition of oscillations with different time scales, and that these scales can be determined more precisely for the 1983 El Niño (with 120 more data in the sample) than for the 1973 event.

In Fig. 12(a) we can see the correlation skill compared with the persistence skill. The latter degrades rapidly to non-significant values (less than 0.22 for a sample with $n = 88$), whereas the forecast skill is significant and very important for lead times up to 19 months, and can be accepted also for lead times up to 34 months

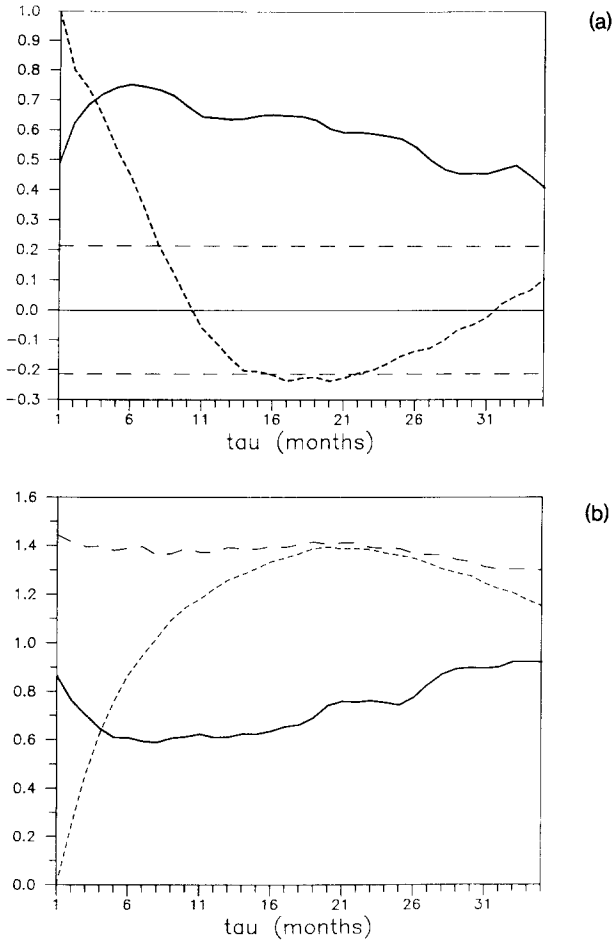

Fig. 12. (a) Average across 80 experiments of the correlation skill for the SST anomalies in the Niño3 region. Continuous line: BOP model. Dashed line: persistence. Dashed horizontal lines mark the statistical non-significant region for the correlation. (b) Average across 80 experiments of the relative mean square error. Continuous line: BOP model. Short dashes: persistence. Long dashes: the worst model, where at every time lag, this model would forecast the negative value of the actual data.

(for 34 months it is still larger than the critical value of 0.22). The relative RMSD error exhibits a similar behaviour (Fig. 12(b)).

5. Conclusions

We have introduced Bayesian oscillation patterns (BOP) as a new tool for the detection of regular evolving patterns from climate fields with many degrees of

freedom, i.e. the direct application of Bayesian methods to signal detection. A noteworthy advantage over other methods is its ability to detect signals less energetic than the noise and to produce easily confidence intervals for the parameters it determines.

We cannot think of this method exactly as a dynamical tool, as no evolution equations have been proposed. On the other hand, it has a distinct dynamical flavour, as it uses as the a priori model functions the solutions of the most standard statistical equations (AR_n, for instance) or the solutions of simple dynamical system equations. This method can also employ the known solutions to any equation that could be proposed to model the evolution of climate fields.

The BOP method is able to resolve any type of oscillation, be it standing or travelling, including first- and higher-order powers of harmonic functions, and constructive and destructive interferences between different oscillations. As a Bayesian method, BOP uses only the data at hand, and in a similar way to another Bayesian technique, the Kalman filter, it is designed explicitly to incorporate any new piece of information provided by experimentation or observation.

A BOP analysis of zonal surface wind and SST anomalies in the Indo-Pacific basin shows the presence of six time scales, one of them corresponding to a quasi-biennial oscillation and another one of approximately 5 years. The fact that there is a number of different oscillations produces constructive interference effects, causing, at irregular time intervals, increases in the amplitude of the anomalies. Hindcast experiments for the zonal wind in the entire basin produced these interferences with similar propagating characteristics and at the same times as the ENSO events of 1973, 1977 and 1983.

The zonal wind pattern corresponding to the 5 year scale shows clearly a propagating signal originating in the Indian Ocean and propagating slowly to the central Pacific. This feature, which indicates a connection between the monsoon and ENSO, is present in data snapshots only at times prior to some of the ENSO events.

A measure of the success of the method is given by the performance of the BOP model in forecasting. One of the forecast experiments produces good results for the zonal wind when we use all the data previous to the start time of the forecast t_{st} for the determination of the relevant time scales; and then the 4 years previous to t_{st} for an optimization in the determination of the patterns. The correlation skill and relative error for the prediction are significantly better than for the persistence.

A simple model, which uses functions other than the usual first power of harmonic functions, and which is suggested by an analysis of the solutions of a dynamical system of the Lorenz class, performs well for the forecast of SST anomalies in El Niño3 region up to 36 months ahead. The model uses as predictors both the zonal wind and the SST anomalies in the entire basin of the Indian and Pacific Oceans. The good results obtained in the forecast experiments (compared with the results of other statistical or dynamical forecasts) demonstrate the ability of the method when dealing with time series of actual data. The potential benefits offered by the introduction of informative priors seem worth exploring.

Acknowledgements

This work has been financed in part by the European Community Climate Program under Grants EV4C/0051 and EPOC and by the CICYT (008/87). We would like to thank Dr. T.P. Barnett for providing the data.

Appendix

If we chose any model (consisting of a certain number of signals), the question is: what is the probability that the signals are embedded in the data? That is, we would like to know the probability

$$P(M \mid d,I) \tag{A1}$$

for Model M, given the data d and all the available information I. Here the data d are the time series we are actually analysing. I denotes all the previous information we have about the data and the experiment we are concerned with. For instance, the data could be time series of wind velocity anomalies. The previous information we have is that wind velocities follow the Navier–Stokes equations, that they are never stronger than 200 m s^{-1} at the surface, that they move in vortices and wave patterns, etc.

Bayes' theorem can be written as

$$P(M \mid d,I) = \frac{P(M \mid I)P(d \mid M,I)}{P(d \mid I)} \tag{A2}$$

(Jeffreys, 1961). Here $P(M \mid I)$ is the prior probability for the model, given the available information; i.e. given that the data must satisfy the Navier–Stokes equations, I, $P(M \mid I)$ denotes the probability of Model M consisting of oscillations. $P(d \mid I)$ is the probability of obtaining the data from the available information. We can include $P(d \mid I)$ in the normalization constant for the probability $P(M \mid d,I)$, as it does not change for any model we choose.

The posterior probability we are looking for, $P(M \mid d,I)$, is then the product of the prior probability of Model M given information I, and of the likelihood $P(d \mid M,I)$, that is, the probability of finding the data given a proposed model, M, and the a priori information (i.e. that the data follow the Navier–Stokes equations):

$$L(d) \equiv P(d \mid M,I) \tag{A3}$$

To apply Bayesian techniques we need to define the possible models we would like to test for the presence of some signals in the data. These models should be oscillations (harmonic, anharmonic, elliptic or Bessel functions, etc.), as a consequence of the Navier–Stokes equations. A crucial step in the procedure is to determine the likelihood. We can use the MaxEnt principle (Jaynes, 1978), for the application of which extensive literature is available (see Kapur and Kesavan (1992) and references therein).

The difference between the model (signal) and the data is defined as the noise w. If we knew the true signal, we could determine the characteristics of the noise. If we knew the noise, we could compute the likelihood. This not being the case, we are forced to assume that the noise has the most general attributes: we can assign a prior probability to it that is as uninformative as possible. If we knew nothing about the noise but its variance σ^2, the principle of maximum entropy would give the probability

$$P(w \mid \sigma^2, I) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{w^2}{2\sigma^2}\right) \tag{A4}$$

As we do not know the noise variance, we will have to integrate over this parameter in the applications.

The likelihood at time t_i is the probability of obtaining one data value d_{t_i} given the model m_{t_i}; it is proportional to the probability of the noise being zero:

$$P(d_{t_i} \mid m_{t_i}, I) \propto P(w \mid \sigma^2, I) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left[-\frac{(d_{t_i} - m_{t_i})^2}{2\sigma^2}\right] \tag{A5}$$

If we had some information about the noise (i.e. red or blue, its time correlation function, etc.), we could write the probability $P(w_{t_1}, w_{t_2}, \ldots, w_{t_m})$ as a function of the individual probabilities $P(w_{t_i})$. If we want to stick to the most uninformative probability for the sequence $\{w_{t_1}, w_{t_2}, \ldots, w_{t_m}\}$, we can suppose $P(w_{t_j} \mid w_{t_i}, \sigma, I) = P(w_{t_i} \mid \sigma, I), (t_i, t_j)$, $i, j = 1, \ldots, m$,

$$P(d \mid M, \sigma, I) = \prod \left\{ \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left[-\frac{(d_{t_i} - m_{t_i})^2}{2\sigma^2}\right] \right\} \tag{A6}$$

or

$$P(d \mid M, \sigma, I) \propto \sigma^{-m} \times \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{m} (d_{t_i} - m_{t_i})^2\right] \tag{A7}$$

Our model was proposed in Eq. (2) of the main text:

$$\boldsymbol{m}(t_i) = \sum_{\nu=1}^{N} \boldsymbol{a}_\nu G_\nu(t_i; \{\omega\})$$

where $\{\omega\}$ is the set of parameters of the model (frequencies in this case).

The likelihood for the parameters $\{\omega\}$, can be obtained by integrating over all possible values of the amplitudes $\boldsymbol{a}_\nu$ of the model. To do this, the calculations are simpler if we build functions H_ν, obtained from G_ν by orthonormalization

$$H_\nu(t_i) = \sum_{\mu=1}^{N} C_{\nu\mu} G_\mu(t_i) \tag{A8}$$

with

$$\sum_{i=1}^{m} H_\nu(t_i) H_\mu(t_i) = \delta_{\nu\mu} \tag{A9}$$

and then project the data into these functions:

$$h_{k\nu} \equiv \sum_{i=1}^{m} d_k(t_i) H_\nu(t_i), \quad k = 1,\ldots,n; \quad \nu = 1,\ldots,N \tag{A10}$$

For the time being, as we are interested only in the generalized frequencies $\{\omega\}$, we can substitute (A10) into (A7) and integrate over the amplitudes $\boldsymbol{a}_\nu$. This is where we could introduce previous knowledge about the amplitudes. However, in this first study we want to use the minimum number of assumptions. Thus we will take a most uninformative uniform prior for the amplitudes. Using the orthonormal variables H, this integration is rather simple, and we obtain,

$$L(\{\omega\} \mid \sigma, I) \propto \sigma^{-m+N} \times \exp\left(- \frac{m\overline{d^2} - N\overline{h^2}}{2\sigma^2} \right) \tag{A11}$$

We do not know the variance or equivalently, the standard deviation σ, of the noise, so we have to integrate over all its possible values. To do that, we can assign to this variance a prior $1/\sigma$, corresponding to a uniform probability distribution for its logarithm (Jeffreys, 1961). Then we can write the joint posterior probability for the set ω, given the data d, as

$$P(\{\omega\} \mid d, I) \propto \left(1 - \frac{N\overline{h^2}}{m\overline{d^2}} \right)^{(N-m)/2} \tag{A12}$$

where

$$\overline{h^2} = \frac{1}{n} \frac{1}{N} \sum_{k=1}^{n} \sum_{\nu=1}^{N} h_{k\nu}^2 \tag{A13}$$

$$\overline{d^2} = \frac{1}{n} \frac{1}{m} \sum_{k=1}^{n} \sum_{t_i=1}^{M} d_{kt_i}^2 \tag{A14}$$

The optimal parameters ω_o are chosen as those that maximize probability $P(\{\omega\} \mid d, I)$.

A simple example can give us some indication about the usefulness of this sort of calculations. Let us suppose we have two small samples of two (possible different) populations: Sample 1 has $N = 9$, $\bar{x}_1 = 42$, $s_1 = 7.48$; Sample 2 has $N = 4$, $\bar{x}_2 = 50$, $s_2 = 6.48$. It seems obvious that Sample 2 comes from a population with greater mean. Can we assign a probability to this 'obvious fact'? We want to test two models for the two sets of observations. The first model, for Sample 1, is that the average is a. The second model, for Sample 2, is that the average is b. Further, we assume that $b > a$. We write the sample average and the sample standard deviations as $\overline{x_1}$ and s_1, $\overline{x_2}$ and s_2, respectively.

The likelihood, based on a model that specifies that the data come from a normal population of mean a, and standard deviation σ, is well known, and it is

$$P(x_i \mid a, \sigma, I) = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left\{ - \frac{n}{2\sigma^2} \left[(\bar{x}_1 - a)^2 + s_1^2 \right] \right\} \mathrm{d}x_i \tag{A15}$$

and similarly for mean b.

With an uniform prior $P(a \mid I)$ and a uniform prior for the logarithm of the scale parameter σ, the posterior probability for the mean a can be obtained by integrating over all the values of the standard deviation:

$$P(a \mid x_{1i}, I) \propto \left[(\bar{x}_1 - a)^2 + s_1^2\right]^{-n/2} \tag{A16}$$

and equivalently for $P(b \mid x_{2j}, I)$.Then the probability of model b being greater than model a is

$$P(b > a \mid x_{1i}, x_{2j}, I) = \int_{-\infty}^{\infty} da \int_{a}^{\infty} db P(a \mid x_{1i}, I) P(b \mid x_{2j}, I) \tag{A17}$$

We obtain $P(b > a) = 0.92$. We can see that even with small samples we can obtain a rather high probability, confirming what it is obvious at first sight: that Population 2 has a greater average value.

Turning again to the problem at hand, it is only now, when we have determined the probability of the parameters $P(\{\omega\} \mid d, I)$ of the model, and we have obtained the optimal parameters ω_o, that we can estimate the noise variance:

$$\langle \sigma^2 \rangle \approx \frac{1}{m - N - 2} \left[\sum_{t_i=1}^{m} \boldsymbol{d}^2(t_i) - \sum_{\nu=1}^{N} \boldsymbol{h}_\nu^2 \right] \tag{A18}$$

and the signal-to-noise ratio

$$\text{signal/noise} \approx \left[\frac{N}{m} \left(1 + \frac{\overline{\boldsymbol{h}^2}}{\overline{\boldsymbol{d}^2}} \right) \right]^{1/2} \tag{A19}$$

The patterns $\boldsymbol{a}_\nu$ are easily determined by inverting Eq. (2) using the orthonormal functions H_ν.

To estimate the error for the parameters, we build a $\mathscr{B}$ matrix, with elements $b_{\nu\mu}$ given by

$$b_{\nu\mu} \equiv -(N/2) \frac{\partial^2 \overline{\boldsymbol{h}^2}}{\partial \omega_\nu \partial \omega_\mu} \tag{A20}$$

Let λ_μ be the μ-eigenvalue of matrix $\mathscr{B}$, and $u_{\nu\mu}$ the corresponding eigenvector. Then

$$\gamma_\nu^2 \equiv \langle \sigma^2 \rangle \sum_{\mu=1}^{N} \frac{u_{\nu\mu}^2}{\lambda_\mu} \tag{A21}$$

and

$$\omega_\nu \approx \omega_{o\nu} \pm \gamma_\nu \tag{A22}$$

Equivalently, we could estimate the error for the patterns $\boldsymbol{a}_j$, as functions of $\langle \sigma^2 \rangle$ and coefficients $C_{\nu\mu}$ of Eq. (11).

References

Barnett, T.P., 1983. Interaction of the monsoon and Pacific trade wind system at interannual time scales. Part I. The equatorial zone. Mon. Weather Rev., 111: 756–773.

Barnett, T.P., 1991. The interaction of multiple time scales in the tropical climate system. J. Climate, 4: 269–285.

Barnett, T.P. and Hasselmann, K., 1979. Techniques of linear prediction, with application to oceanic and atmospheric fields in the tropical Pacific. Rev. Geophys. Space Phys., 17: 949–968.

Bretthorst, G.L, 1988. Bayesian Spectrum Analysis and Parameter Estimation. Springer, Berlin.

Fletcher, J.O., Slutz, R.J. and Woodruff, S.D., 1983. Towards a comprehensive ocean–atmosphere data set. Trop. Ocean Atmos. Newslett., 20(13–14).

Franceschini, E. and Tebaldi, C., 1989. Sequences of infinite bifurcations and turbulence in a five mode model truncation of Navier Stokes equations. J. Phys., 21: 707.

Hasselmann, K., 1988. POPs and PIPs. The reduction of complex dynamical systems using principal oscillation and interaction patterns. J. Geophys. Res., 93(D9): 10975–10988.

Jaynes, E.T., 1978. Where do we stand on maximum entropy? In: R. Levine and M. Tribus (Editors), The Maximum Entropy Formalism. MIT Press, Cambridge, MA.

Jaynes, E.T., 1985. Predictive statistical mechanics. In: Proceedings, NATO Advanced Study Institute, Santa Fe, NM, 3–16 June 1984.

Jaynes, E.T., 1987. Bayesian spectrum and chirp analysis. In: C.R. Smith and G.J. Erikson (Editors), Maximum Entropy and Bayesian Spectral Analysis and Estimation Problems. D. Reidel, Dordrecht, pp. 1–37.

Jeffreys, H., 1961. Theory of Probability. Oxford University Press, Oxford.

Justice, J.H., 1986. Maximum Entropy and Bayesian Methods in Applied Statistics. Cambridge University Press, Cambridge.

Kapur, J.N. and Kesavan, H.K., 1992. Entropy Optimization Principles with Applications. Academic Press, Boston, MA.

Kidson, J.W., 1975. Tropical eigenvector analysis and the Southern Oscillation. Mon. Weather Rev., 103: 187–196.

Latif, M. and Flügel, M., 1991. An investigation of short range climate predictability in the tropical Pacific. J. Geophys. Res., 96: 2661–2674.

Latif, M., Barnett, T.P., Cane, M.A., Flügel, M., Graham, N.E., von Storch, H., Xu, J.S. and Zebiak, S.E., 1994. A review of ENSO prediction studies. Climate Dyn., 9: 167–179.

Lau, K.M. and Sheu, P., 1988. Annual cycle, QBO and Southern Oscillation. In: Global Precipitation. J. Geophys. Res., 93(D9): 10975–10988.

Marple, S.L., 1987. Digital Spectral Analysis with Applications. Prentice–Hall, Englewood Cliffs, NJ.

Preisendorfer, R.W., 1988. Principal Component Analysis in Meteorology and Oceanography. Elsevier, Amsterdam.

Rasmusson, E.M., Wang, X. and Ropelewski, C., 1990. The biennial component of ENSO variability. Proc. 21st Int. Liège Colloquium on Ocean Hydrodynamics. Elsevier, Amsterdam.

RuizdeElvira, A., 1993. The climate problem as a complex physical system. Complexity, nonlinearity and reduction techniques. In: J.I. Diaz and J.L. Lions (Editors), Mathematics, Climate and Environment. Masson, Paris.

Spall, J.C., 1988. Bayesian Analysis of Time Series and Dynamic Models. Marcel Dekker, New York.

Xu, J. and von Storch, H., 1990. Predicting the state of the Southern Oscillation using principal oscillation pattern analysis. J. Climate, 3: 1316–1329.