A Potential Problem With Extended EOF Analysis of Standing Wave Fields

Adam H. Monahan, Fredolin T. Tangang and William W. Hsieh
Oceanography, Department of Earth and Ocean Sciences
University of British Columbia
Vancouver, B.C. V6T 1Z4

ABSTRACT *The extended empirical orthogonal function (EEOF) analysis of signals containing a standing wave component is considered. It is shown using a simple analytical model that for a single time-lag EEOF analysis, two EEOF patterns describe the standing wave, and that their associated eigenvalues become degenerate if the lag chosen is close to a zero of the sample autocovariance function of the standing wave time series. This degeneracy strongly affects the relative magnitudes of the two sectors of each EEOF pattern and the structure of the associated spatial pattern. A set of EEOF analysis of varying lags is performed on a tropical Pacific sea level pressure (SLP) dataset, known to be dominated by standing variance, and the predictions of the analytical model are verified. As well, it is seen that at degeneracy there is a reduction in the quality of the spatial patterns produced by the EEOF analysis.*

RÉSUMÉ *Nous présentons une analyse en Fonctions orthogonales empiriques élargies (FOEE) de signaux contenant une composante ondulatoire stationnaire. À l'aide d'un modèle analytique simple, nous montrons que, pour une FOEE à un seul décalage temporel, deux patrons de FOEE décrivent l'onde stationnaire. De plus, les valeurs propres qui leur sont associées dégénèrent si le décalage choisi est près d'un zéro de la fonction d'autocovariance de la série temporelle de l'onde stationnaire. Cette dégénérescence affecte fortement les grandeurs relatives des deux secteurs de chacun des patrons de FOEE ainsi que la structure des patrons spatiaux associés. Un ensemble d'analyses en FOEE pour des décalages variables est effectué sur un jeu de données de pression (au niveau de la mer) recueillies dans le Pacifique tropical et dont nous savons que la variance est dominée par une onde stationnaire. Ces analyses confirment les prévisions du modèle analytiqe. On constate également qu'au moment de la dégénérescence, il y a une diminution de la qualité des patrons spatiaux produits par l'analyse FOEE.*

1 Introduction

Time-lagged extended empirical orthogonal function (EEOF) analysis has been used in a number of recent studies (e.g., Barnston, 1994; Shabbar and Barnston,

1996; Black et al., 1996; Tangang et al., 1998). In such an analysis, an empirical orthogonal function (EOF) analysis is performed on a heterogeneous dataset composed of observations of the same physical fields at two or more different times. The method has shown some promise for both the purposes of data reduction and of understanding the nature of the dominant patterns of spatially and temporally coherent variability in a dataset. However, other than the work by Chen and Harr (1993), little has been written about the idiosyncrasies of EEOF analysis.

We demonstrate that if the dataset subjected to a single time lag EEOF analysis contains a standing wave component, then the two EEOF patterns describing this standing variance become degenerate when the time lag is equal to a zero of the sample autocovariance function of the standing wave's time series. This degeneracy can complicate the interpretation of the two affected EEOF patterns.

A brief overview of EEOF analysis and an analytical demonstration of the above-mentioned degeneracy are presented in Section 2 of this paper. In Section 3, we perform a series of EEOF analyses with varying lags of tropical Pacific SLP data, a field dominated by standing variance (Tangang et al., 1998), and it is shown that this degeneracy and its consequences occur as predicted.

2 Analytical Model

Consider the noisy standing wave field $x_i(t_n)$ described by

$$x_i(t_n) = \psi_i f(t_n) + N_i(t_n), \tag{1}$$

where $f(t_n)$ is the standing wave time series, the index i labels station number, $i \in \{1, \ldots, N\}$, and n labels the observation time, $t_n = n\Delta$, $n \in \{1, \ldots, T\}$. By definition the standing wave spatial pattern ψ is of unit norm

$$|\psi|^2 = 1, \tag{2}$$

and $N_i(t_n)$ is a spatially uncorrelated noise field of arbitrary covariance structure in time,

$$\mathcal{E}\{N_i(t)N_j(t+\tau)\} = \delta_{ij}\eta(\tau). \tag{3}$$

We assume the time series f and N_i are centred in time,

$$\langle N_i \rangle = \langle f \rangle = 0 \tag{4}$$

and T is sufficiently large that

$$\langle f(t)N_i(t) \rangle \simeq 0 \tag{5}$$

and

$$\langle N_i(t)N_j(t+\tau) \rangle \simeq \delta_{ij}\eta(\tau), \tag{6}$$

where the angle brackets denote the sample mean, i.e.

$$\langle z \rangle = \frac{1}{T} \sum_{n=1}^{T} z(t_n). \tag{7}$$

Assuming the above approximate equalities hold exactly, the sample covariance matrix Γ of the field is given by

$$\Gamma_{ij} = \Lambda_{ij} a_f(0) + \delta_{ij} \eta(0), \tag{8}$$

where we have defined the matrix Λ as the dyadic product

$$\Lambda_{ij} = \psi_i \psi_j, \tag{9}$$

and where $a_f(\tau)$ is the sample autocovariance function of f, i.e.

$$a_f(\tau) = \sum_{n=1}^{T-\tau/\Delta} f(t_n) f(t_n + \tau) \quad \text{if} \quad \tau > 0, \tag{10}$$

$$= \sum_{n=1+\tau/\Delta}^{T} f(t_n - \tau) f(t_n) \quad \text{if} \quad \tau < 0. \tag{11}$$

Note that $a_f(\tau)$ so defined is symmetric:

$$a_f(\tau) = a_f(-\tau). \tag{12}$$

The eigenvectors of Γ are the EOFs of (1). It can easily be shown that the spatial pattern ψ_i is an EOF of (1) with associated eigenvalue

$$a_f(0) + \eta(0). \tag{13}$$

As the total sample variance in the field (1) is given by

$$\sigma^2 = a_f(0) + N\eta(0), \tag{14}$$

we see that the fraction of the sample variance explained by the standing wave mode is

$$\frac{a_f(0) + \eta(0)}{a_f(0) + N\eta(0)}, \tag{15}$$

which becomes 1 in the limit of vanishing noise.

In a time-lagged EEOF analysis involving a single time lag τ, we construct the stacked vector time series

$$\mathbf{X}(t_n) = \begin{pmatrix} \mathbf{x}(t_n) \\ \mathbf{x}(t_n + \tau) \end{pmatrix}, \tag{16}$$

and perform an EOF analysis of this time series. The sample covariance matrix of the stacked dataset (16) is denoted $\mathbf{G}$ and has the block-diagonal form

$$\mathbf{G} = \begin{pmatrix} a_f(0)\Lambda + \eta(0)I & a_f(\tau)\Lambda + \eta(\tau)I \\ a_f(\tau)\Lambda + \eta(\tau)I & a_f(0)\Lambda + \eta(0)I \end{pmatrix}, \tag{17}$$

with I the identity matrix. Then the vectors $\mathbf{\Psi}_+$ and $\mathbf{\Psi}_-$ defined by

$$\mathbf{\Psi}_{\pm} = \frac{1}{\sqrt{2}} \begin{pmatrix} \mathbf{\Psi}_{\pm}^{(0)} \\ \mathbf{\Psi}_{\pm}^{(\tau)} \end{pmatrix} \tag{18}$$

$$= \frac{1}{\sqrt{2}} \begin{pmatrix} \psi \\ \pm\psi \end{pmatrix} \tag{19}$$

are mutually orthonormal eigenvectors of $\mathbf{G}$ with corresponding eigenvalues

$$g_{\pm} = (a_f(0) + \eta(0)) \pm (a_f(\tau) + \eta(\tau)). \tag{20}$$

The spatial maps $\mathbf{\Psi}_{\pm}^{(0)}$ and $\mathbf{\Psi}_{\pm}^{(\tau)}$ are denoted the (0) and (τ) sectors of the EEOF pattern, respectively.

The total sample variance in the stacked dataset (16) is simply twice that of the field (1), so the fractions of variance explained by $\mathbf{\Psi}_{\pm}$ are

$$r_{\pm} = \frac{(a_f(0) + \eta(0)) \pm (a_f(\tau) + \eta(\tau))}{2(a_f(0) + N\eta(0))}. \tag{21}$$

Note that the sum $r_+ + r_-$ is equal to the fraction of sample variance in (1) explained by the EOF ψ; the fraction of sample variance explained by the standing wave has thus been split between the $\mathbf{\Psi}_+$ and $\mathbf{\Psi}_-$ modes. In general, then, each of these modes will individually explain a smaller fraction of the sample variance than does the standing wave EOF ψ.

Associated with the EEOFs $\mathbf{\Psi}_{\pm}$ are the time series

$$F_{\pm}(t_n) = \mathbf{X}(t_n) \cdot \mathbf{\Psi}_{\pm} \tag{22}$$

$$= \frac{1}{\sqrt{2}} [(f(t_n) + \nu(t_n)) \pm (f(t_n + \tau) + \nu(t_n + \tau))]. \tag{23}$$

where we have defined the time series

$$\nu(t_n) = \mathbf{N}(t_n) \cdot \psi \tag{24}$$

of the projection of the noise on the standing wave spatial pattern.

In general, the $\boldsymbol{\Psi}_+$ and $\boldsymbol{\Psi}_-$ modes correspond to different eigenvalues of G and are thus distinct. However, for those values of τ at which $a_f(\tau) + \eta(\tau)$ is nearly zero, the eigenvalues g_+ and g_- become close and may not be well-separated. When this occurs, these eigenvalues may become degenerate such that the pair of vectors $\boldsymbol{\Xi}_\pm$

$$\boldsymbol{\Xi}_+ = \alpha\boldsymbol{\Psi}_+ + \beta\boldsymbol{\Psi}_- \tag{25}$$

$$= \frac{1}{\sqrt{2}} \begin{pmatrix} (\alpha + \beta)\psi \\ (\alpha - \beta)\psi \end{pmatrix}, \tag{26}$$

$$\boldsymbol{\Xi}_- = -\beta\boldsymbol{\Psi}_+ + \alpha\boldsymbol{\Psi}_- \tag{27}$$

$$= \frac{1}{\sqrt{2}} \begin{pmatrix} (\alpha - \beta)\psi \\ -(\alpha + \beta)\psi \end{pmatrix} \tag{28}$$

will be an orthonormal basis set on the space of eigenvectors of G describing the standing wave, where α and β satisfy

$$\alpha^2 + \beta^2 = 1, \tag{29}$$

but are otherwise arbitrary. This degeneracy is a generalization of that found by Chen and Harr (1993) in their two-station model.

The degeneracy manifests itself in the following manner. Away from degeneracy, the two sectors of the standing wave EEOFs have magnitudes

$$|\boldsymbol{\Psi}_\pm^{(0)}|^2 = |\boldsymbol{\Psi}_\pm^{(\tau)}|^2 = \frac{1}{2}, \tag{30}$$

whereas at degeneracy, these magnitudes are

$$|\boldsymbol{\Xi}_+^{(0)}|^2 = |\boldsymbol{\Xi}_-^{(\tau)}|^2 = \frac{(\alpha + \beta)^2}{2}, \tag{31}$$

$$|\boldsymbol{\Xi}_+^{(\tau)}|^2 = |\boldsymbol{\Xi}_-^{(0)}|^2 = \frac{(\alpha - \beta)^2}{2}. \tag{32}$$

Thus, a hallmark of degeneracy is predicted to be a marked change away from 1 in the value of the ratio of the magnitudes of the two sectors of the EEOF.

Finally, it is interesting to consider the Fourier spectra of the time series $F_\pm$. The standing wave EOF time series has Fourier decomposition

$$f(t_n) + v(t_n) = \sum_{k=0}^{T-1} (\tilde{f}_k + \tilde{v}_k) \exp\left(\frac{2\pi i n k}{T}\right), \tag{33}$$

and those of the EEOF time series $F_\pm(t_n)$ are

$$F_{\pm}(t_n) = \sum_{k=0}^{T-1} (\tilde{F}_{\pm})_k \exp\left(\frac{2\pi i n k}{T}\right). \tag{34}$$

From equation (23) it follows that

$$|(\tilde{F}_{\pm})_k|^2 = |\tilde{f}_k + \tilde{v}_k|^2 \left(1 \pm \cos\frac{2\pi m k}{T}\right), \tag{35}$$

where $m = \tau/\Delta$. Note that in the spectrum of $F_+(t_n)$, low frequencies are amplified and those near $k = T/2m$ suppressed relative to the spectrum of $f(t_n)$, while the low frequency components of $F_-(t_n)$ are suppressed and those near $k = T/2m$ amplified. Thus, the EEOF analysis splits the signal $f(t_n)$ into two parts: high- and low-pass filtered versions of the original time series.

These results can be generalized to EEOF analyses with more than one time lag. For example, it can be shown that an EEOF analysis using the heterogeneous data vector

$$\hat{\mathbf{X}}(t_n) = \begin{pmatrix} \mathbf{x}(t_n) \\ \mathbf{x}(t_n + \tau) \\ \mathbf{x}(t_n + 2\tau) \\ \mathbf{x}(t_n + 3\tau) \end{pmatrix} \tag{36}$$

will yield degenerate eigenvalues if $a_f(\tau) + \eta(\tau) = a_f(3\tau) + \eta(3\tau) = 0$.

3 EEOF Analysis of SLP Data

The SLP dataset used in this study was the same as that used in Tangang et al. (1998). Monthly-averaged SLP data from the Comprehensive Ocean-Atmosphere Data Set (COADS) spanning the period January 1952 to June 1997 was averaged to a 4° by 10° grid covering the region from 28°S to 20°N and from 52°E to 88°W. The annual cycle was removed by subtracting from each data point the monthly climatological value. The resulting anomaly field was further smoothed using a 3-point running average in time and a 1-2-1 filter in each spatial dimension.

As described in Tangang et al. (1998), the coherent spatial variability in the equatorial Pacific SLP is dominated by an east-west dipole field whose associated time series is the familiar low-frequency oscillation component of the El-Niño Southern Oscillation (ENSO) (Barnett, 1991). This dipole field displays no propagating features; it can, to a reasonable first approximation, be described by the model (1). The first EOF mode of the SLP data, explaining 36.2% of the variance, corresponds to this pattern of standing variance; Fig. 1 shows a plot of the sample autocovariance function of the time series of this mode. The sample autocovariance first becomes zero at a time very close to 13 months; if the sort of degeneracy predicted in the previous section appears in the EOF analysis of this data, it should occur for time lags τ near 13 months.

A single-lag EEOF analysis was performed on the SLP dataset, with lags varying
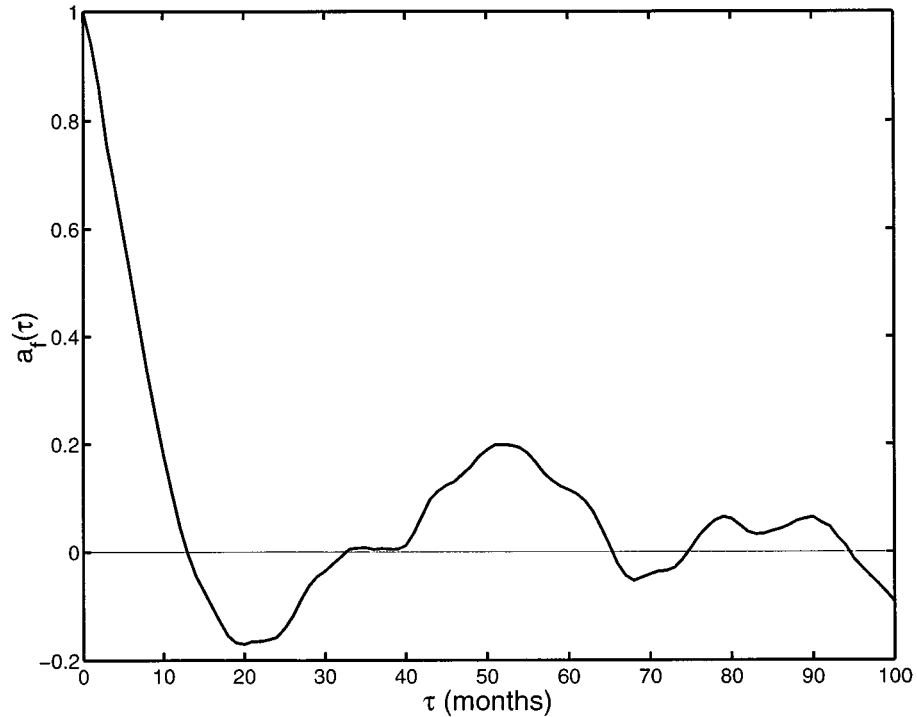
Fig. 1   Sample autocovariance function of the leading SLP EOF time series (normalized to unit variance).

between 9 and 17 months. For this range of lags, the leading two EEOF modes were very close to the $\Psi_{\pm}$ structure. Figure 2 displays the percentages of variance explained by Ψ_{+} and Ψ_{-} as a function of the lag τ. Note that these eigenvalues cross at a lag of slightly less than $\tau = 13$ months, a lag corresponding to the first zero of the sample autocovariance function of the standing wave signal, as expected from (20). The sum of the percentage of variance explained by the two leading modes varies with τ between 38.1% and 39.0%; it is thus slightly, but not much, higher than the percentage explained by the corresponding EOF mode. As well, this sum is remarkably invariant over a broad range of lags. We then see that the percentage of variance explained by the EOF standing wave pattern is indeed split between the percentages explained by the two single time-lag EEOF standing wave patterns, as predicted in the previous section.

Although the leading two eigenvalues are not identical at $\tau = 13$ months, they are not well separated and degeneracy of the kind described in the previous section does occur at this lag. Figure 3 shows a plot of the magnitudes of the two sectors of Ψ_{+} and Ψ_{-} as a function of τ. Of course, at degeneracy the EEOF patterns are not Ψ_{+} and Ψ_{-} but linear combinations of these; however, at $\tau = 13$ months,
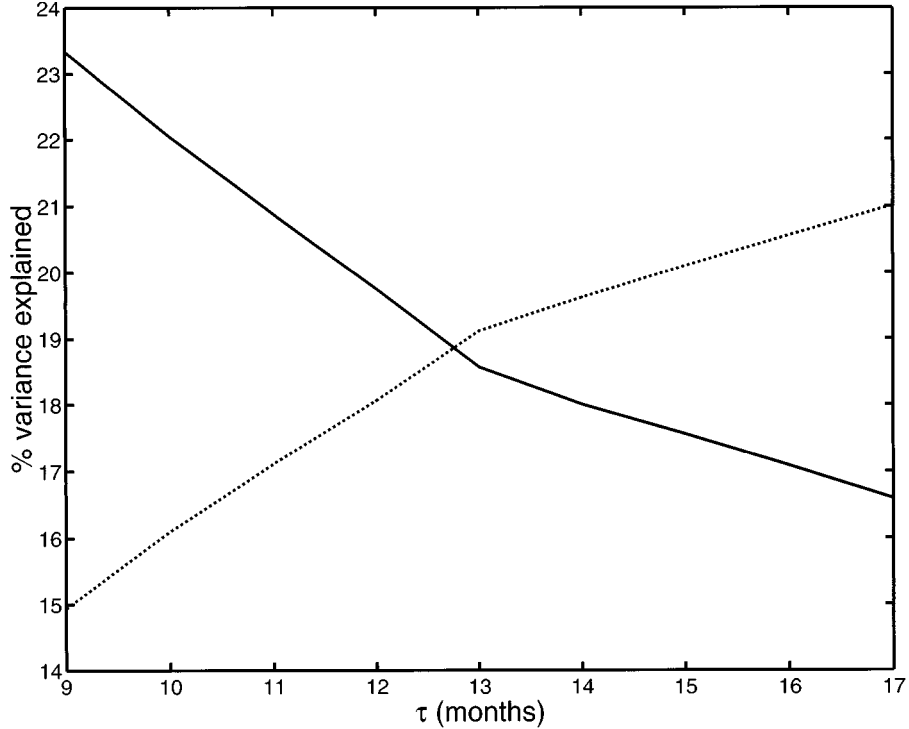
Fig. 2   Percentages of variance explained by SLP EEOF modes Ψ_+ (solid line) and Ψ_- (dashed line), as a function of lag time, τ.

mode 1 projects most strongly on Ψ_- and mode 2 most strongly on Ψ_+, so we continue to use these labels for convenience. Note that for τ well away from 13 months, the magnitudes are close to 0.5, as predicted by equation (30). However, for τ close to 13 months, these magnitudes become markedly different from 0.5, and the difference in magnitudes is most pronounced at the first zero of the sample autocovariance function, $\tau = 13$ months.

When the magnitude of one of the two sectors of the EEOF becomes very small, some degradation in signal can occur. This is because the model (1) does not exactly describe the dataset, and the leading EEOF modes, while dominated by the standing wave signal, include other information. Figure 4 displays a plot of the absolute value of the spatial correlations between the two sectors of Ψ_+ and of Ψ_-, i.e.

$$\frac{\left| \Psi_+^{(0)} \cdot \Psi_+^{(\tau)} \right|}{\left| \Psi_+^{(0)} \right| \left| \Psi_+^{(\tau)} \right|} \tag{37}$$
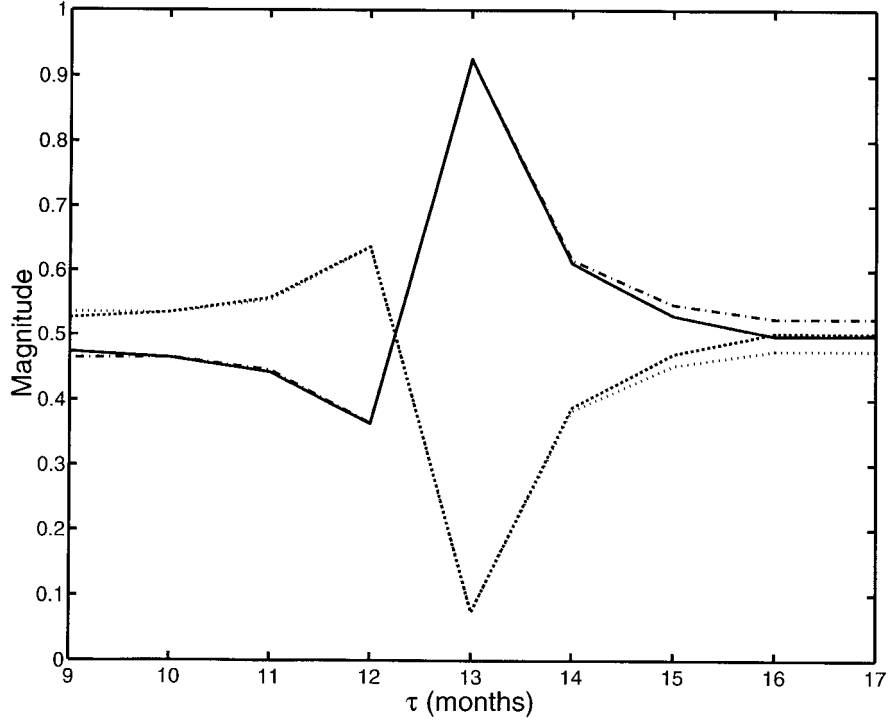
and

Fig. 3 Magnitudes of the two sectors of the leading two SLP EEOF modes, as a function of lag time, τ: $\Psi_+^{(0)}$ (solid line); $\Psi_+^{(\tau)}$ (dashed line); $\Psi_-^{(0)}$ (dotted line); $\Psi_-^{(\tau)}$ (dot-dashed line).

$$\frac{\left|\boldsymbol{\Psi}_-^{(0)} \cdot \boldsymbol{\Psi}_-^{(\tau)}\right|}{\left|\boldsymbol{\Psi}_-^{(0)}\right|\left|\boldsymbol{\Psi}_-^{(\tau)}\right|} \tag{38}$$

as a function of τ. For a system described exactly by (1), the magnitude of the spatial correlation would be 1 at all τ. For $\boldsymbol{\Psi}_+$, the value is near 1 for all τ except near 13 months, at which point the value drops to 0.50. Because the SLP data contains a large-scale coherent eastward-propagating pattern with a subannual timescale, the description of which is part of $\boldsymbol{\Psi}_-$ for τ less than 12 months, the spatial correlation of $\boldsymbol{\Psi}_-$ is smaller than that of $\boldsymbol{\Psi}_+$ for these lags. However, the spatial correlation between the two sectors of $\boldsymbol{\Psi}_-$ also has a sharp minimum at $\tau = 13$ months. Figures 5 and 6 display the spatial patterns of the standing wave EEOF modes at $\tau = 15$ and at $\tau = 13$ months, respectively. The patterns away from degeneracy have the predicted $\boldsymbol{\Psi}_+$ and $\boldsymbol{\Psi}_-$ forms. At degeneracy, the standing wave patterns depart markedly from this structure. The $\boldsymbol{\Psi}_-^{(\tau)}$ and $\boldsymbol{\Psi}_+^{(0)}$ patterns retain the familiar east-west dipole structure, while the other two sectors are corrupted. Those sectors with very small magnitudes experience a suppression of the standing wave signal and a consequent corruption by other small signals, because the model (1) does not
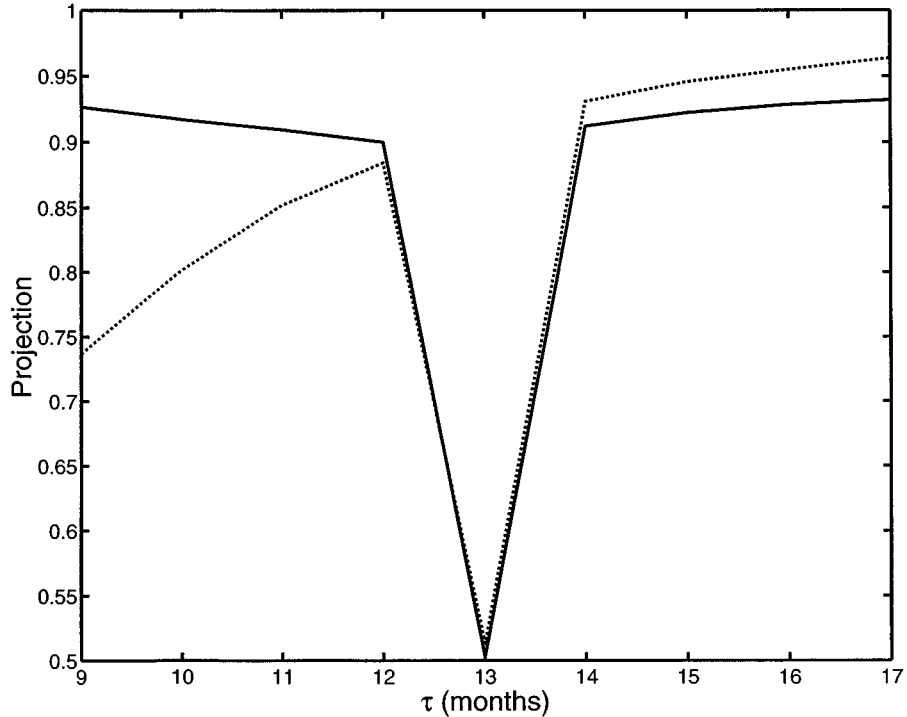
Fig. 4   Absolute value of the spatial correlation between the two sectors of Ψ_+ (solid line) and those of Ψ_- (dashed line).

exactly describe the SLP dataset. Conversely, the standing wave pattern is enhanced in those with large magnitude, and here the dipole pattern is robust. The degeneracy of these two modes has thus resulted in a marked degradation in the quality of the EEOF signal.

From the above results, it is clear that the degeneracy predicted in Section 2 does in fact occur in a dataset with a strong standing wave component. This degeneracy manifests itself in the structure of the EEOFs by greatly increasing the ratio of the magnitudes of the two sectors of the EEOF from near 1 and through a reduction of the ability of the EEOF pattern to diagnose spatial patterns of standing variability. This degeneracy also occurs for lags corresponding to other zeroes of the sample autocovariance function. An EEOF analysis with a lag of $\tau = 33$ months (not shown) displays the same degenerate mixing of modes observed with a lag of $\tau = 13$ months.

Figure 7(a) shows the time series corresponding to the first SLP EOF, and 7(b) shows that corresponding to the first SLP EEOF with a lag of $\tau = 3$ months. While these time series closely resemble each other, the one arising from the EEOF analysis is clearly smoother. This is precisely the result predicted by equation (35):
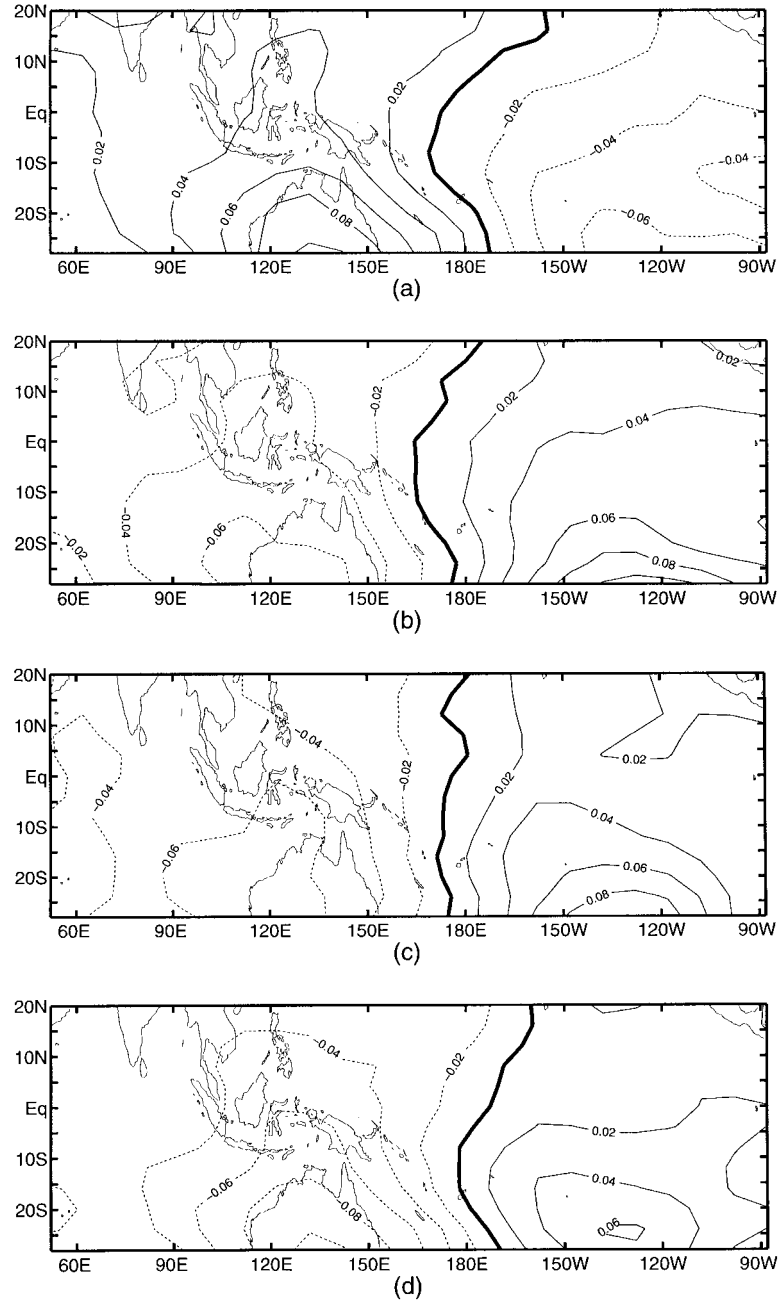
Fig. 5   Maps of EEOF patterns at $\tau = 15$ months: (a) $\Psi_{-}^{(0)}$, (b) $\Psi_{-}^{(\tau)}$, (c) $\Psi_{+}^{(0)}$, (d) $\Psi_{+}^{(\tau)}$.
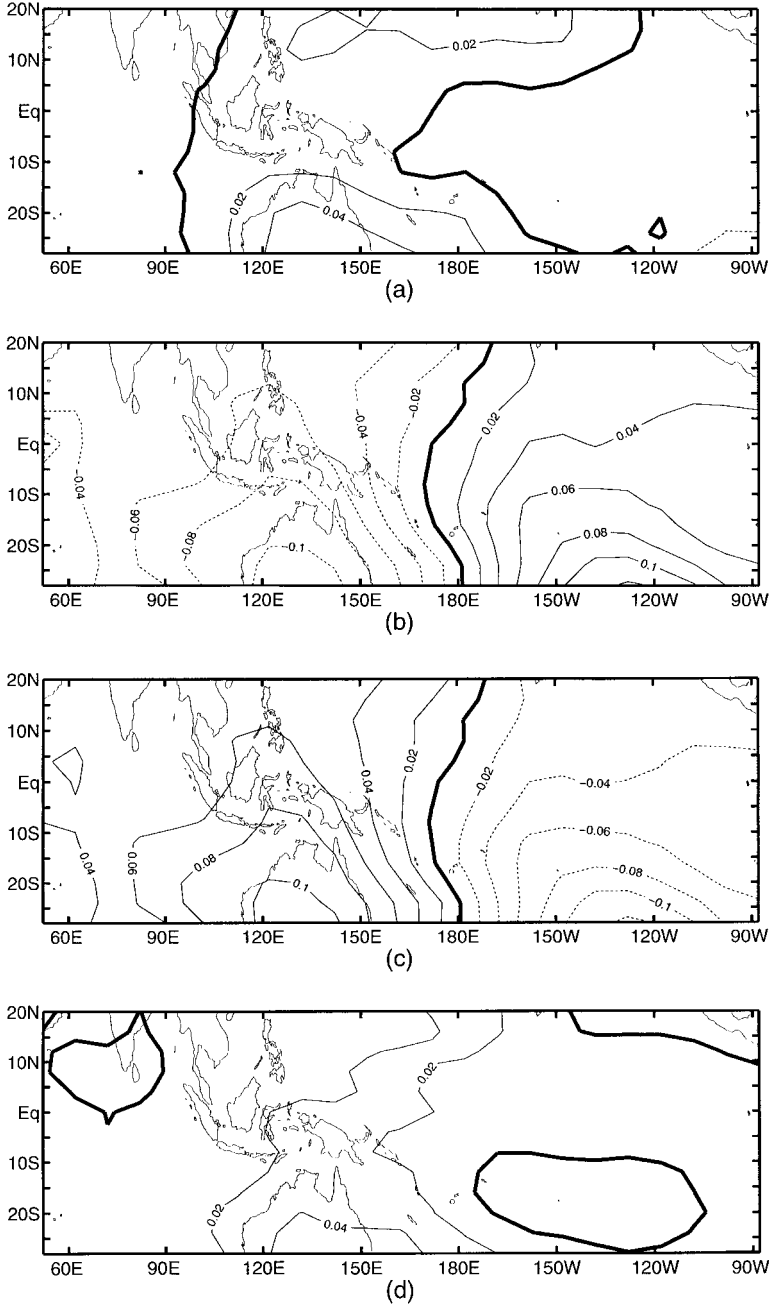
Fig. 6   Maps of EEOF patterns at degeneracy, $\tau = 13$ months: (a) $\Psi_-^{(0)}$, (b) $\Psi_-^{(\tau)}$, (c) $\Psi_+^{(0)}$, (d) $\Psi_+^{(\tau)}$.
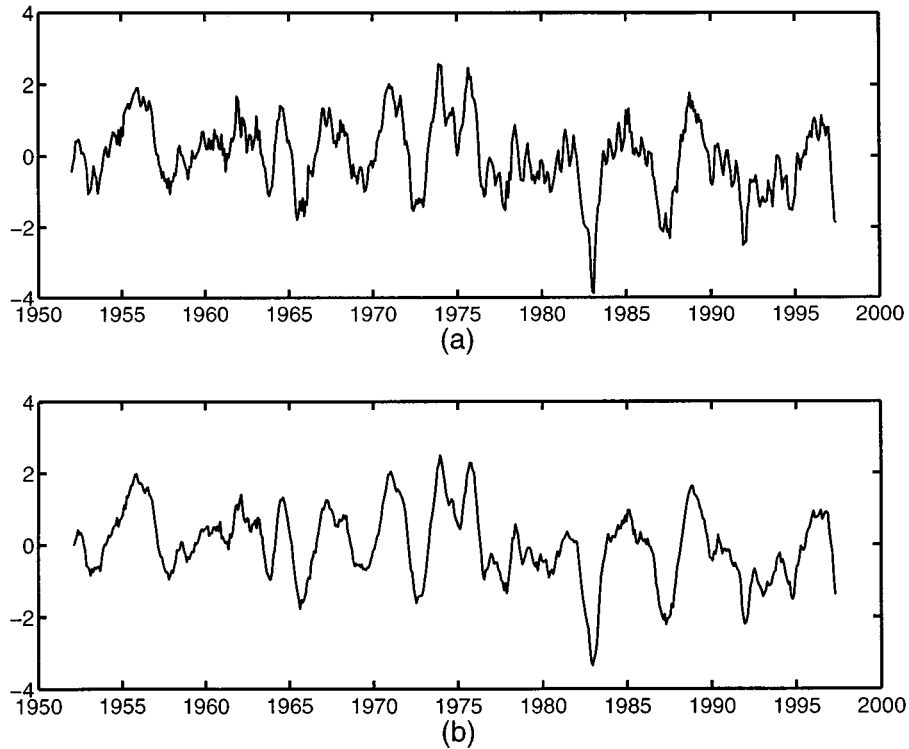
Fig. 7   (a) SLP EOF mode 1 time series, (b) SLP EEOF mode 1 time series for $\tau = 3$ months, each
normalized to unit variance.

the variabilty on seasonal timescales has been eliminated from the EEOF time series. This smoothing of the standing wave time series was also noted in Tangang (1997), in which a three-lag EEOF analysis of the form (36) was performed on this SLP data.

Results similar to those obtained in the EEOF analysis of the SLP dataset were found in a single time lag EEOF analysis of the tropical Pacific sea surface temperature (SST) data used by Tangang et al. (1998), and degeneracy was again found to occur at a time lag of 13 months. These results were not markedly different from those described above, so they will not be discussed in any detail.

4 Summary

It has been shown above, both through analytical arguments and through the example of a tropical Pacific SLP dataset, that if a dataset contains a strong standing wave component, then the EEOFs describing this signal will be degenerate if the time lag chosen is sufficiently close to a zero of the sample autocovariance function of the associated time series. Plaut and Vautard (1994) also report the connection

between oscillatory behaviour and degeneracy of the eigenvalue spectrum in Multichannel Single Spectrum Analysis (M-SSA), which is formally identical to EEOF analysis. Their analytical demonstration of the effect, however, is restricted to signals that are exactly sinusoidal in time, and they do not discuss the problems that arise in interpretation of spatial patterns obtained at degeneracy. This degeneracy has a profound impact on the relative magnitudes of the two sectors of the standing wave EEOF and on the quality of the spatial patterns. The interpretation of the results of an EEOF analysis performed with a time lag near a zero of the standing wave's sample autocovariance function is complicated by these effects. In particular, the quite marked degradation of signal quality displayed in Fig. 6 could conceivably be as bad, or worse, in other datasets. In the example of Section 3, the lag time of 13 months which leads to degeneracy is somewhat unnatural; it is unlikely that such a lag time would be used in most conceivable applications with this particular dataset. However, other datasets with strong standing wave components may have sample autocovariance functions with zeros at lags more natural for use in an EEOF analysis, in which case this degeneracy could prove to be a problem, if not accounted for.

Acknowledgements

References

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

BARNSTON, A.G. 1994. Linear statistical short-term climate predictive skill in the northern hemisphere. *J. Clim.* **7**: 1513–1564.

BLACK, R.X.; D.A. SALSTEIN and R.D. ROSEN. 1996. Interannual modes of variability in atmospheric angular momentum. *J. Clim.* **9**: 2834–2849.

CHEN, J.M. and P.A. HART. 1993. Interpretation of extended empirical orthogonal function (EEOF) analysis. *Mon. Weather Rev.* **121**: 2631–2636.

PLAUT, G. and R. VAUTARD. 1994. Spells of low-frequency oscillations and weather regimes in the northern hemisphere. *J. Atmos. Sci.* **51**: 210–236.

SHABBAR, A. and A.G. BARNSTON. 1996. Skill of seasonal climate forecast in Canada using canonical correlation analysis. *Mon. Weather Rev.* **124**: 2370–2385.

TANGANG, F.T. 1997. Forecasting El Nino-Southern Oscillation (ENSO) events: a neural network approach. Ph.D. dissertation, UBC, 153 pp.

————; B. TANG, A.H. MONAHAN, W.W. HSIEH. 1998. Forecasting ENSO events: a neural network-extended EOF approach. *J. Clim.* **11**: 29–41.