Chapter 16. EMPIRICAL MODAL DECOMPOSITION IN COASTAL OCEANOGRAPHY

HANS VON STORCH

GKSS Research Center

CLAUDE FRANKIGNOUL

Université Pierre et Marie Curie

Contents

1. Introduction: needles in haystacks
2. Modal decomposition as an eigenproblem
3. Modal decomposition I: EOF analysis and generalizations
4. Model decomposition II: Canonical correlation analysis and related techniques
5. Summary
Bibliography

1. Introduction: Needles in Haystacks

In 1983 a campaign was launched to monitor the circulation in the coastal waters off the Californian coast in the Santa Barbara Channel (Brink and Muench, 1986). For that purpose various observing platforms were installed (see Fig. 16.1): two surface moorings, labeled C1 and C2, which reported horizontal velocity and temperature at six and five different depths every 7.5 minutes; and 13 subsurface moorings, labeled P1 to P13, which measured horizontal velocity and temperature at 60 m and below—a total of 32 locations and depths every 30 minutes. Furthermore, surface winds were recorded by two floating buoys, named NS and NC, at hourly intervals.

In the end, a data set covering about 60 days in April to June 1983 was available. For some parameters longer time series were available. As an example of the raw data, "stick diagrams" of vector time series of wind stress, currents and times series of subsurface temperature are shown in Fig. 16.2.

This is obviously a large amount of data, which exhibit a wide mixture of "signals" and "noise." The purpose of statistical analysis is to disentangle this mixture to find the needle (signal) in the haystack (noise). (The allegory with the needle in the

The Sea, Volume 10, edited by Kenneth H. Brink and Allan R. Robinson
ISBN 0-471-11544-4

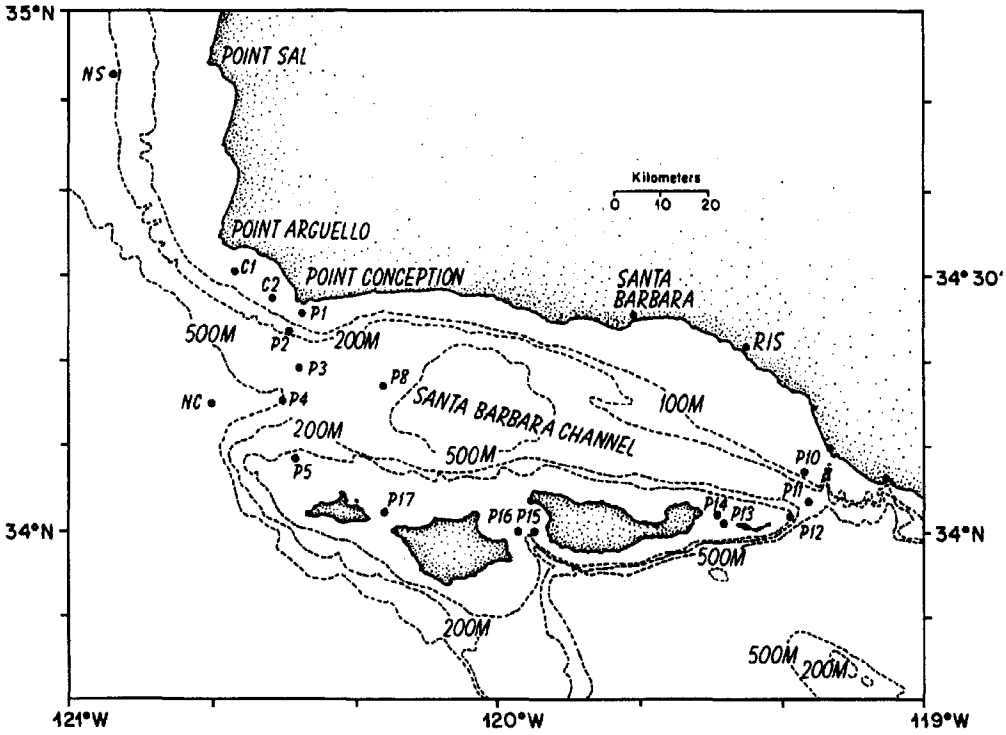

Fig. 16.1. Location of observational platforms in the Santa Barbara Channel off the Californian coast. (From Brink and Muench, 1986.)

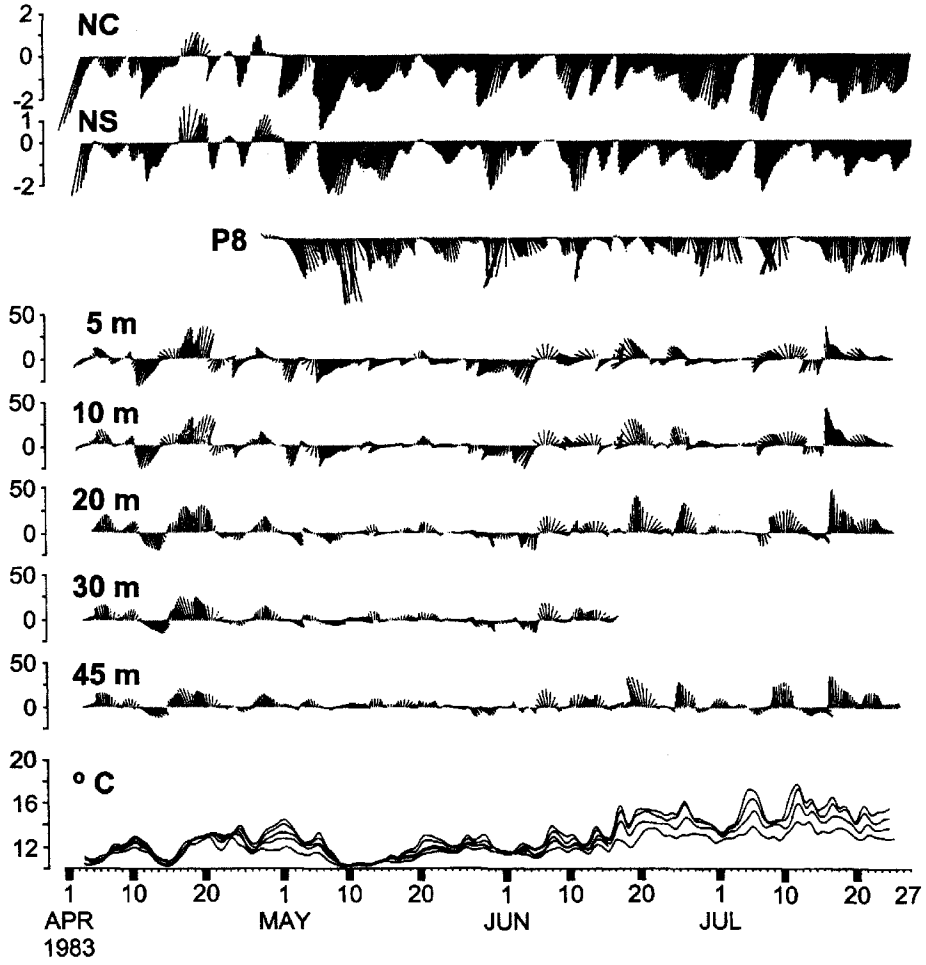

Fig. 16.2. Plots of vector wind stress (top three panels; for NC, NS and P8), vector currents (middle five panels; C1; at 5, 10, 20, 30 and 45 m) and temperatures (bottom panel; C1; same depth as for the currents; warmest water is at shallowest depth). (From Brink and Muench, 1986.)

haystack has two sides: First, it is difficult to *find* the needle in the haystack, and second, *after* it has been found, it is easily recognizable as a needle simply by looking at it. To identify a climatic signal, advanced techniques may be required, but after its identification the signal usually may be described by means of simple techniques such as composites, correlations, and the like.)

The definitions of *signal* and *noise* are somewhat arbitrary and depend on the interest of the researcher. In most general terms, a signal is a pattern in space, or time, or space and time, which is determined by the system dynamics. Noise, on the other hand, can be physical or instrumental and can often be considered to be, to first approximation, unrelated to the main signal. More liberally defined, noise, or random contributions, comprise all those features that are considered irrelevant for the chosen signal.

A straightforward choice of a signal is the time mean state (e.g., the mean currents for C1 displayed in Fig. 16.7), so that none of the time variations (trends, deterministic and random fluctuations) are considered interesting. Often, however, the space/time variability is thought to reveal something about the underlying dynamics, and regular features in the space/time variability are considered to be the "signal".

The concept of the separation of the full data field into a "signal" and a "noise" can be formalized. Assume that the full data field at any time t is given by the vector $\vec{\mathbf{X}}(t)$. Then the separation may be written as

$$\vec{\mathbf{X}}(t) = \vec{\mathbf{X}}^s(t) + \vec{\mathbf{X}}^n(t) \tag{1}$$

with the two contributions on the right side representing the signal and the noise. About the noise nothing specific is known, but it is assumed that it is characterized by certain well-behaving statistics, such as spatial and temporal correlation scales which are distinct from that of the signal. The signal, on the other hand, is assumed to have only a few degrees of freedom; otherwise, the signal would be as untractable as the full data set and little would have been gained by the separation (1). More specifically, it is often assumed that the signal may be expressed as a sum of only a few characteristic patterns $\vec{p}^k$:

$$\vec{\mathbf{X}}^s(t) = \sum_{k=1}^{K} \alpha_k(t)\vec{p}^k \tag{2}$$

The K patterns p^k and time coefficients $\alpha_k(t)$ are supposed to be determined by the dynamics of the signal. The time coefficients are defined uniquely by the scalar product

$$\alpha_k = \langle \vec{p}_A^k, \vec{\mathbf{X}} \rangle \tag{3}$$

of the vector of state $\vec{\mathbf{X}}$ and the "adjoint patterns" $\vec{p}_A^k$, which are formally given by the columns of the matrix

$$(\vec{p}_A^1 | \cdots | \vec{p}_A^K) = (\vec{p}^1 | \cdots | \vec{p}^K)\mathcal{P}^{-1} \tag{4}$$

with the matrix $\mathcal{P} = (\langle \vec{p}^{\,k}, \vec{p}^{\,i} \rangle)_{ki}$. A more robust and, in practice, simpler approach is to determine the time coefficients by a least-squares fit such that the difference

$$\sum_t \left[\vec{\mathbf{X}}(t) - \sum_{k=1}^{K} \alpha_k(t) \vec{p}^{\,k} \right]^2 \tag{5}$$

is minimum. [For a more complete explanation, see, e.g., von Storch (1995b).]

The patterns can be constructed in many different ways. The conceptually most appealing approach is to define the pattern dynamically by manipulating the equations of motion (Section 2). However, such a dynamical approach is impossible when the dynamics are too complex to derive the appropriate patterns or when the nonhomogeneity and the anisotropy are too large, as in many coastal problems. In that case, statistically defined patterns are often a useful and wholesale alternative. One approach is to prescribe the type of dynamics of the system and then to derive the patterns and the free parameters which describe their time evolution from the data available. A dynamical model with certain prescribed functional elements is thus fitted to the observations. In most cases the dynamics are supposed to be linear. A well-established prototype of this approach are the principal oscillation patterns (POPs), which we also discuss in Section 2.

The characteristic patterns can also be derived without making any dynamical assumption about the signal, by maximizing instead a certain functional of the observations. Principal component analysis, often referred to as empirical orthogonal function (EOF) analysis, is one such technique which identifies patterns that maximize the variance in a field, while canonical correlation analysis (CCA) determines patterns that maximize the correlation between two fields. These kinds of techniques, which disregard the temporal order of the events, are discussed in Sections 3 and 4. We also discuss two extensions of EOF analysis that identify the dominant space/time patterns in a field without making any assumptions on the time behavior. One technique is the extended EOF analysis (EEOF), which is also known as multichannel singular system analysis (MSSA); the other is the EOF analysis in the frequency domain (FDEOF).

In Sections 2 to 4 we deal mainly with what one could call *exploratory analysis:* how to summarize certain dominant properties of a field, such as its dominant space/space patterns, and how to discriminate between a signal of interest and unrelated processes. We are not discussing the uncertainty of the derived properties which one should expect from the limited sampling of the field data. To be more complete, we should distinguish between the true dominant space/time patterns and their estimates from the available sample, and we should discuss the expected errors of the latter, as in classical textbooks such as Anderson (1984) or Seber (1984). We would then discuss *confirmatory analysis:* how to test whether a signal that has been identified is consistent with the true signal at a given level of confidence, taking into account the randomness of a limited set of observations. As reviewed by Frankignoul (1995), the problem of hypothesis testing in multidimensional fields has received increasing attention since the work of Hasselmann (1979) on sensitivity studies with atmospheric general circulation models, and its generalization to space/space behaviors has been used for ocean model testing and intercomparisons (Frankignoul et al., 1989, 1996),

as well as for climate change detection (Hasselmann, 1993). These concepts are and should be used in coastal oceanography, but lack of time and space prevents us from discussing what one could call in the present context *pattern uncertainty and testing.*

Nonetheless, we want to touch upon two aspects that are commonly encountered in the analysis of coastal data set. The first is that of *serial correlation,* which occurs when a variable has a longer correlation time than the interval between samples, so that the latter are not independent. Since the independence assumption is required to use the Student t distribution in describing the distribution of univariate statistics such as the *mean* and the *correlation coefficient,* one must either subsample the time series to get independent observations, or take into account their finite correlation time. This is commonly done by defining an *equivalent sample size* based on lag correlations or parametric modeling of the time series (e.g., Thiébeaux and Zwiers, 1984) or derived from spectral analysis (e.g., Jones, 1976). As discussed in Zwiers and von Storch (1995), this is applicable only when the equivalent sample size n_e is large enough $n_e > 30$) for the t variable to become approximately normal. Otherwise, tests based on n_e perform poorly and establishing statistical significance may require Monte Carlo simulations.

The second problem is that of interpreting an *ensemble* of univariate tests, which requires taking into account the multiplicity of local tests. Even in the simple case where all the univariate tests are independent, the overall rate of rejection of the null hypothesis at the a % level is often larger than a (global rejection of the null hypothesis) if the number of local tests is finite. The critical rejection rate can be inferred from the binomial distribution (von Storch, 1982; Livezey and Chen, 1983), and for a small number of tests, the threshold for field significance can be large. In the more common case where the observations, hence the local tests, are not independent but spatially correlated, one observes that the tests tends to be rejected in "pools" of points, rather than at randomly distributed points, and one expects the critical rejection rate to be larger since the *equivalent* number of independent tests is smaller. This number is difficult to estimate, however, because the tests have poorly known spatial correlations. Livezey and Chen (1983) have suggested establishing field significance by using permutation techniques. An alternative is to use the methods of multivariate statistical analysis, although they also have limitations in the (usual) small sample case (see Frankignoul, 1995).

2. Modal Decomposition as an Eigenproblem

The general form of the equations of motion in the ocean is

$$\frac{\partial \mathbf{X}}{\partial t} = \mathcal{L}(\mathbf{X}) + \mathbf{F} \tag{6}$$

where $\mathbf{X}$ is a field variable (velocity, pressure and density), t the time, $\mathcal{L}$ a (nonlinear) operator, and $\mathbf{F}$ the forcing field. In many cases, one can study the system dynamics by linearizing (6) and then expressing the solution of the forced linear problem in term of the solutions of the homogeneous one. This is often facilitated by assuming a separation of variables

$$\mathbf{X}(x,t) = g(t)\mathbf{h}(x) \tag{7}$$

so that the time evolution is described by the differential equation

$$\frac{dg}{dt} = \lambda g \tag{8}$$

and the spatial variations by the eigenvalue problem

$$\mathcal{A}\mathbf{h} = \lambda\mathbf{h} \tag{9}$$

where λ is the separation constant and $\mathcal{A}$ a linear operator. The solutions of (9) that satisfy the boundary conditions are the eigenfunctions or normal modes $\mathbf{h}^k(x)$ corresponding to the eigenvalues λ_k. For each λ_k equation 7 is solved by $g_k(t) = a_k \exp(\lambda_k t)$, where the constant a_k is determined by the initial conditions. The nature of these solutions depends on the eigenvalue λ_k. If it is real and positive, the solution grows unboundedly; if it is real and negative, the solution decays exponentially. If the eigenvalue is complex, the eigenfunction is also complex, but as the system is real valued, the complex conjugate (denoted by an askerisk) λ_k^* is an eigenvalue corresponding to $\mathbf{h}^{k*}$, with coefficient $a_k^*(t)$. Thus the overall contribution of this mode to $\mathbf{X}$ is given by

$$2\,\mathrm{Re}(g_k(t)\mathbf{h}^k) = 2e^{t/\tau_k}(\cos(\omega_k t)\,\mathrm{Re}(\mathbf{h}^k) + \sin(\omega_k t)\,\mathrm{Im}(\mathbf{h}^k)) \tag{10}$$

with $\lambda_k = 1/\tau_k + i\omega_k$. The real numbers τ_k and ω_k represent the growth or decay time, depending on the sign, and the frequency of the mode, respectively.

The eigenfunctions form a complete set, hence the solution of the linear forced problem (6) can be obtained by superposition of the solution of separable equations for each eigenfunction,

$$\mathbf{X}(x) = \sum_k b_k(t)\mathbf{h}^k(x) \tag{11}$$

where $b_k(t)$ is the solution of the forced version of equation 8. This has the form (2). Note that when the problem is discretized in space, the operator $\mathcal{A}$ becomes a matrix, the function $\mathbf{X}$ a time-dependent vector $\vec{\mathbf{X}}(t)$, and the eigenfunctions $\mathbf{h}^k(x)$ the eigenvectors of $\mathcal{A}$, say $\vec{p}^{\,k}$.

In some coastal problems, (9) can be reduced by separating the alongshore y-dependence, yielding an eigenvalue problem for the across-shore and vertical x- and z-dependence only, which is usually solved numerically, as reviewed by Brink (1991). Under strongly simplifying assumptions, the x- and z-dependence can also be separated, as illustrated in Section 2.1. However, there are complex settings where the theoretical-analytical approach is not tractable unless oversimplifications are made, in which case the chosen theoretical framework may not cover the most relevant part of the phase space and the modes may become irrelevant.

If the modes cannot be derived from the equations of motion, a purely empirical

approach to estimating the modal solutions can still be used in the discrete case (Section 2.2): One simply assumes that the vector $\vec{\mathbf{X}}(t)$ is controlled by a linear equation of the form (6), which has separable solutions of the form (11). If it is, furthermore, assumed that the forcing can be represented as a white noise, the system can be written

$$\vec{\mathbf{X}}(t + 1) = \mathcal{A}\vec{\mathbf{X}}(t) + \text{noise} \tag{12}$$

which describes a discrete multivariate first-order autogressive (Markov) process. As described below, the $\vec{\mathbf{X}}$-times series can then be used to estimate the system matrix $\mathcal{A}$ and its eigenvectors. These empirical modes are called the principal oscillation patterns (POPs) (Hasselmann, 1988; von Storch et al., 1988, 1995). It should be remarked that this approach makes sense only if the system has, at least approximately, a Markovian behavior. This is often the case, however, in particular for wind-driven coastal currents, which have been described successfully by stochastic models [see the review in Brink (1991)].

The relation between empirical and dynamical modes has been investigated by Schnur et al. (1993), who calculated from quasi-geostrophic theory the dynamical modes describing the extratropical atmospheric variability and also used the POP approach on a long sequence of analyzed geopotential height data. The spatial and temporal characteristics of the most significant POPs were very similar to the most unstable waves in the stability analysis, but the POPs also identified modes representative of the evolution of finite-amplitude waves. Thus the POPs appear to be useful descriptors of the variability in cases where the dynamics were complex. A generalization of the POP concept to nonlinear dynamics, the principal interaction patterns, has been proposed by Hasselmann (1988); first partial implementations are offered by Selten (1995), Kwasniok (1996), Achatz et al. (1995) and others.

2.1. *Dynamically Defined Eigenproblems*

In this section we illustrate the use of dynamical modes by means of a simple example provided by Kundu et al. (1975). They considered the problem of the time-variable flow along a coast forced by an alongshore wind stress. Under a number of simplifying assumptions, such as that the Coriolis parameter f is constant, the time scale is longer than the inertial time scale (f^{-1}), and the spatial scale parallel to the coast (y-direction) is much longer than that perpendicular to the coast ($x = 0$), the linearized system (6) can be reduced to the two-dimensional potential vorticity equation

$$\frac{\partial}{\partial t}\frac{\partial^2 p}{\partial x^2} + \frac{\partial}{\partial t}\frac{\partial}{\partial z}\left[\frac{f^2}{N(z)^2}\frac{\partial p}{\partial z}\right] = \text{forcing} \tag{13}$$

where $N(z)$ is the Brunt–Väisälä frequency, plus the boundary conditions. When the bottom slope is small, an approximate solution of (13) may be obtained by a separation of variables

$$p(x,z,t) = \sum_k \beta_k(t) F^k(x) E^k(z) \tag{14}$$

where $F^k(x)$ is given by $\exp(\gamma_k x)$ and $E^k(z)$ satisfies

$$\frac{\partial}{\partial z}\left[\frac{f^2}{N(z)^2}\frac{\partial E^k(z)}{\partial z}\right] = \gamma_k^2 E^k(z) \tag{15}$$

with $\partial E/\partial z = 0$ at $z = 0$ (rigid lid) and $z = -h(x_m)$, where $h(x_m)$ is the depth at the mooring distance x_m from the coast, and γ_k^2 the separation constant. An analytical solution of the eigenproblem (15) cannot be given for the observed $N(z)$-profile, so Kundu et al. (1975) solved it numerically. The first two modes with the smallest eigenvalues γ_k are shown in Figure 16.3. The mode associated with the smallest

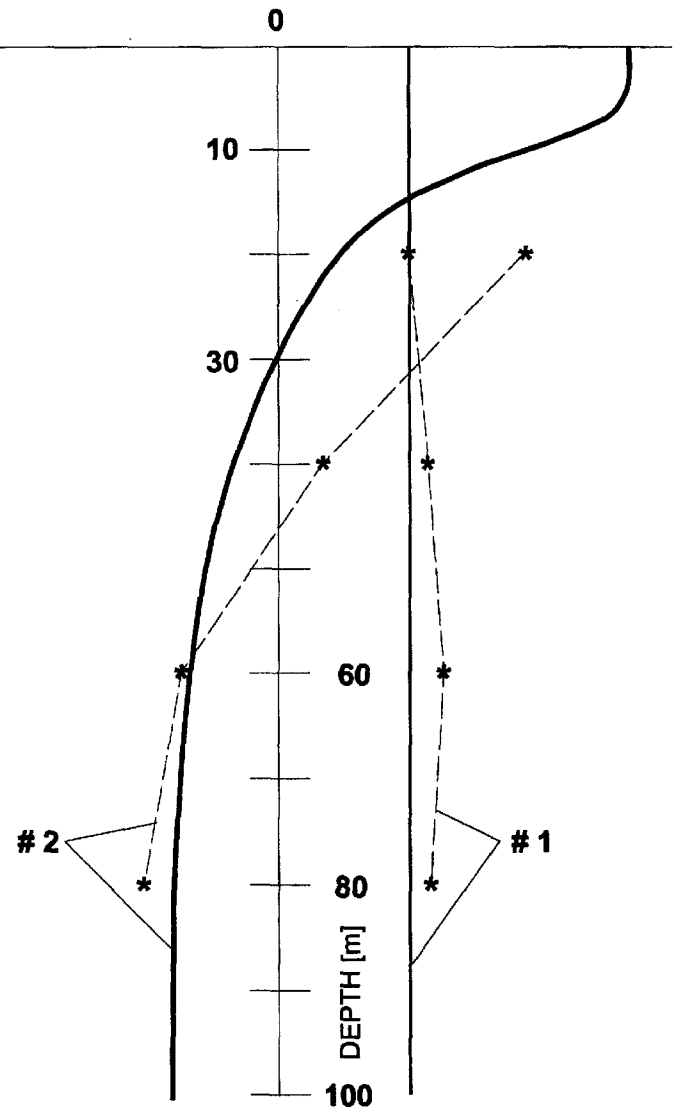

Fig. 16.3. Vertical eigenfunctions $E^k(z)$ of the eigenproblem (15) for the two smallest eigenvalues (continuous lines; see Section 2.1). First two EOFs derived from observations of the alongshore current 20, 40, 60 and 80 m. The points are connected by dashed lines to improve the clarity of the pattern (see Section 3.1). (Adapted from Kundu et al., 1975.)

eigenvalue ($\gamma_1 = 0$) is barotropic and has no horizontal nor vertical structure. The next mode is the lowest "baroclinic" mode, with a sign reversal at a depth of about 30 m. Because of geostrophy, the modes also represent the structure of the alongshore horizontal current.

Kundu et al. (1975) used these two orthogonal modes to analyze time series of horizontal currents measured at a mooring off the coast of Oregon (Fig. 16.4*a*). For

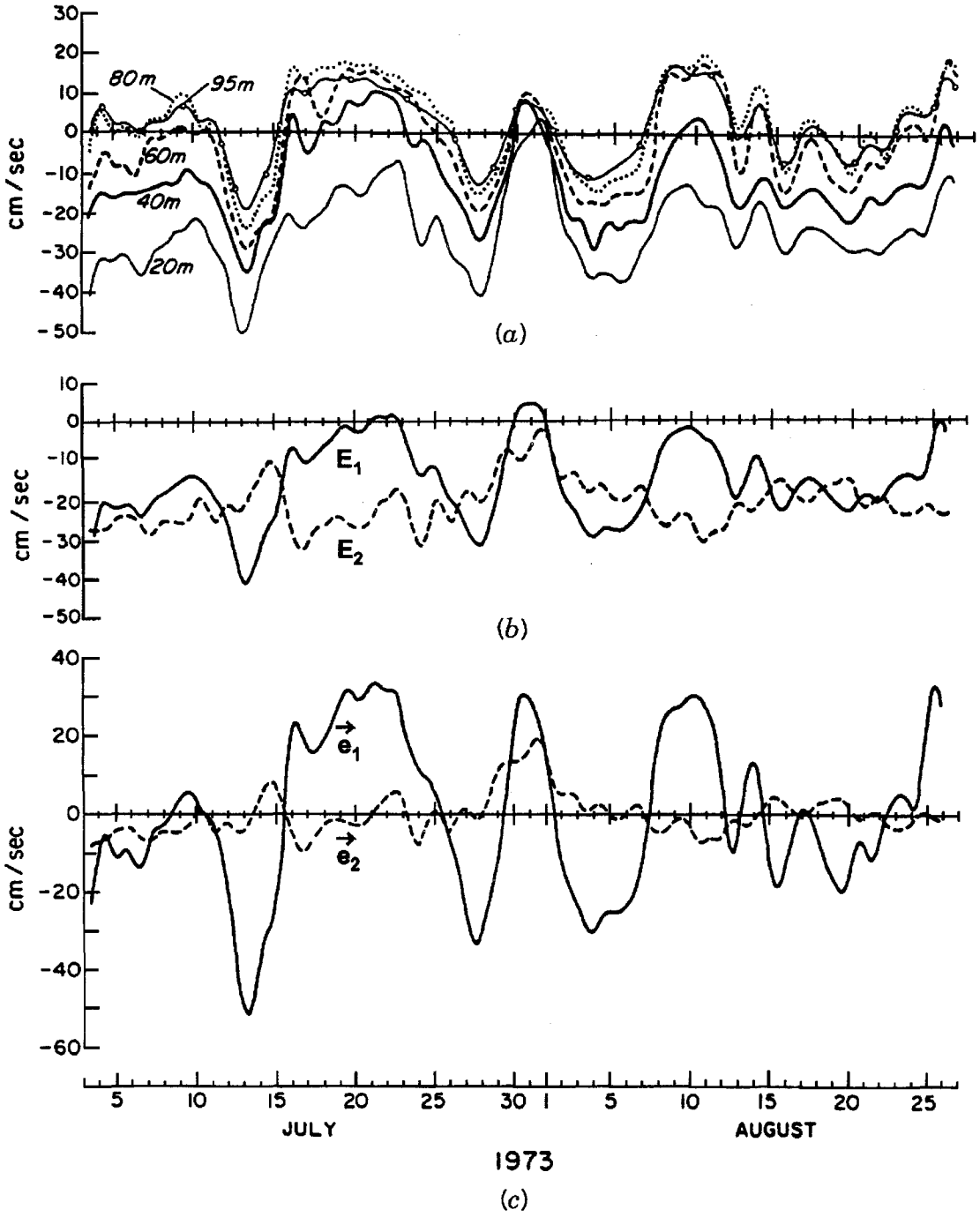

Fig. 16.4. (*a*) Time series of the alongshore current at different depths of a mooring; (*b*) time series of the fitted time series $\alpha_1(t)$ and $\alpha_2(t)$ of the barotropic and first baroclinic mode shown in Fig. 16.3 (see Section 3.1); (*c*) time series of the first two EOF coefficients $\alpha_1(t)$ and $\alpha_2(t)$. The EOFs are displayed with dashed lines in Fig. 16.3. [The EOF analysis has been done with anomalies so that the time mean coefficients are zero in this case. In (*b*) the analysis was done without a priori subtraction of the mean. see Section 3.1.] (From Kundu et al., 1975.)

that purpose the continuous functions $E^1(z)$ and $E^2(z)$ were discretized to form new vectors $\vec{p}^1$ and $\vec{p}^2$ representative of the currents at the four current meter depths 20, 40, 60 and 80 m. Then, using the least-squares fit (5) since the discretized modes are no longer orthogonal, the coefficients $\alpha_1(t) \approx \beta_1(t) \cdot F^1(x_m)$ and $\alpha_2(t) \approx \beta_2(t) \cdot F^2(x_m)$ were determined. In that way, the full data field $\vec{\mathbf{X}}(t)$, consisting of the time series at the four monitored depths (Fig. 16.4*a*), is decomposed into a two-dimensional signal, made up of a barotropic and a baroclinic component, and noise:

$$\vec{\mathbf{X}}(t) = \alpha_1(t)\vec{p}^1 + \alpha_2(t)\vec{p}^2 + \text{noise} \tag{16}$$

The fitted time series account for 85% of the variance (Fig. 16.4*b*). A major part of the overall variability is controlled by barotropic variations, whereas the baroclinic variations are smaller by a factor of 2.

As discussed by Kundu et al. (1975), however, the requirement for the existence of a separable solution (that the slope is small) is not really satisfied, illustrating a common drawback of the theoretical-dynamical approach, namely the need of simplification.

A simplification is sometimes an oversimplification and may have a significant effect on the eventual outcome of the dynamical argument. *In principle it can always be that the chosen theoretical framework does not cover the relevant part of the phase space, so that the resulting modes are irrelevant for the data under investigation.* Thus in many cases the theoretical-dynamical approach cannot be pursued, and purely empirical approaches, unrelated to specific dynamical arguments, are often the only avaiable diagnostic tools. They may also be valuable additional tools to evaluate the significance of theoretically derived structures.

2.2. *Empirically Defined Eigenproblems: Principal Oscillation Pattern Analysis*

In the *principal oscillation pattern* analysis (POP) [for a review, see von Storch et al. (1995)] equation 12 is assumed to hold and the system matrix $\mathcal{A}$ is estimated from data and given by

$$\mathcal{A} = \Sigma_1 \Sigma^{-1} \tag{17}$$

where Σ and Σ_1 denote the lag 0 and lag 1 covariance matrices of $\vec{\mathbf{X}}$, which are easily calculated from the data. All the eigenvalues of (17) have a negative real component. [In the standard literature on POPs, such as in von Storch et al. (1995), the time-dependent problem is expressed as a time-difference problem [i.e., by $\vec{\mathbf{X}}(t) - \vec{\mathbf{X}}(t-1) = \mathcal{B}\vec{\mathbf{X}}_{t-1}$]. This is equivalent to $\vec{\mathbf{X}}(t) = \mathcal{A}\vec{\mathbf{X}}_{t-1}$ with $\mathcal{A} = \mathcal{B} - I$ (I is the unity matrix). The matrices $\mathcal{A}$ and $\mathcal{B}$ have the same eigenvectors, and the eigenvalues λ_k of $\mathcal{A}$ and η_k of $\mathcal{B}$ are related by $\lambda_k = \eta_k - 1$. Therefore, the statement $\mathrm{Re}(\lambda_k) < 0$ is equivalent to $\mathrm{Re}(\eta_k) < 1$.] Missing values should not create any problem as long as there are not too many of them.

The interpretation of the eigenvectors and eigenvalues in this empirically defined eigenproblem is the same as in the dynamical one above. In particular, pairs of complex eigenvalues correspond to damped oscillatory modes if the corresponding time series (the POP coefficients) are coherent and approximately 90° out of phase.

Since POP analysis has not yet been applied to coastal problems, it is illustrated by an oceanic application to equatorial variability (von Storch, 1993). The goal was to investigate the modes of intraseasonal variability in 7 years of moored measurement in the upper tropical Pacific Ocean (Hayes et al., 1991). Oscillatory modes were searched for by using a POP analysis of daily averages of horizontal current and temperature at three equatorial locations, 165°E, 140°W and 110°W, for various depths; their three-dimensional structure was then estimated by regression analysis of a much larger data set onto the POP time series. Here we discuss only the first stage of the analysis.

The data were first filtered by an EOF analysis to suppress small-scale noise and the EOF coefficients filtered in the time domain to eliminate the variability on periods larger than about half a year. The POP analysis then yielded two oscillatory modes (complex pairs of eigenvectors with the foregoing properties). The normalized amplitude time series are displayed in Fig. 16.5. [In the notation of (11) the eigenmodes are given by $\mathbf{h}^k$ and the amplitude time series by $|b_k(t)|$. Normalization of the amplitude time series means that $\mathrm{Var}(b_k(t)) = 1$.]

One oscillatory mode has a period of $T = 2\pi/\omega = 65$ days and a damping time of $\tau = 73$ days [compare with (10)]. The amplitude time series reveals a annual cycle with a semiannual component. The intraseasonal mode activity is strongest during solstice conditions and weakest during equinoctial conditions, and it is enhanced during warm ENSO conditions (1986–1987 and 1990).

The other oscillatory mode, operating at a perid of about 120 days and a damping time of about 105 days, is affected by the state of the southern oscillation as well with enhanced activity during warm episodes and reduced activity during the cold 1988 event. The spatial amplitudes and phases of the two modes, in terms of zonal currents, are displayed in Fig. 16.6. Both modes represent eastward-traveling signals.

The 120-day mode has its largest amplitude, with typical maximum values of about 16 cm s^{-1}, at 50 m depth at 65°E and 160 m depth at 140°W. In contrast, the 65-day mode has maximum zonal current anomalies at upper levels (50 m and above) in the eastern part of the basin, with a typical maximum of 12 cm s^{-1} at 140°W and 19 cm s^{-1} at 110°W. The zonal current 120-day signal propagates in about 60 days from 165°E to 110°W, so that the phase speed is about 1.8 m s^{-1}. The phase lines are vertically tilted, with the upper levels lagging the lower levels by about 15 days. The phase speed for the 65-day mode is estimated to be 2.1 m s^{-1}. At the two eastern positions, the phase lines are tilted, with the lower levels leading the upper levels by about 8 days.

Both modes feature a temperature signal of the order of 1° along the thermocline (not shown). The temperature signal of the 120-day mode is stronger than that of the 65-day mode. The phase speed of the 120-day temperature signal is faster than that for the zonal current signal, namely about 2.7 m s^{-1}. The propagation of the 65-day temperature signal parallels that of the current signal, but with a lag of about 10 days. The two modes are not correlated; their time coefficients share a correlation of about −0.25. The two modes have, however, a similar pattern and are *not* orthogonal. Indeed, the POP analysis does not require that the modes be orthogonal.

3. Modal Decomposition I: EOF Analysis and Generalizations

In this and the following section, we deal with two analyses that are designed to identify the simultaneous occurrence of characteristic patterns in one or several vector (field) time series. These techniques, *empirical orthogonal function analysis* (Sec-

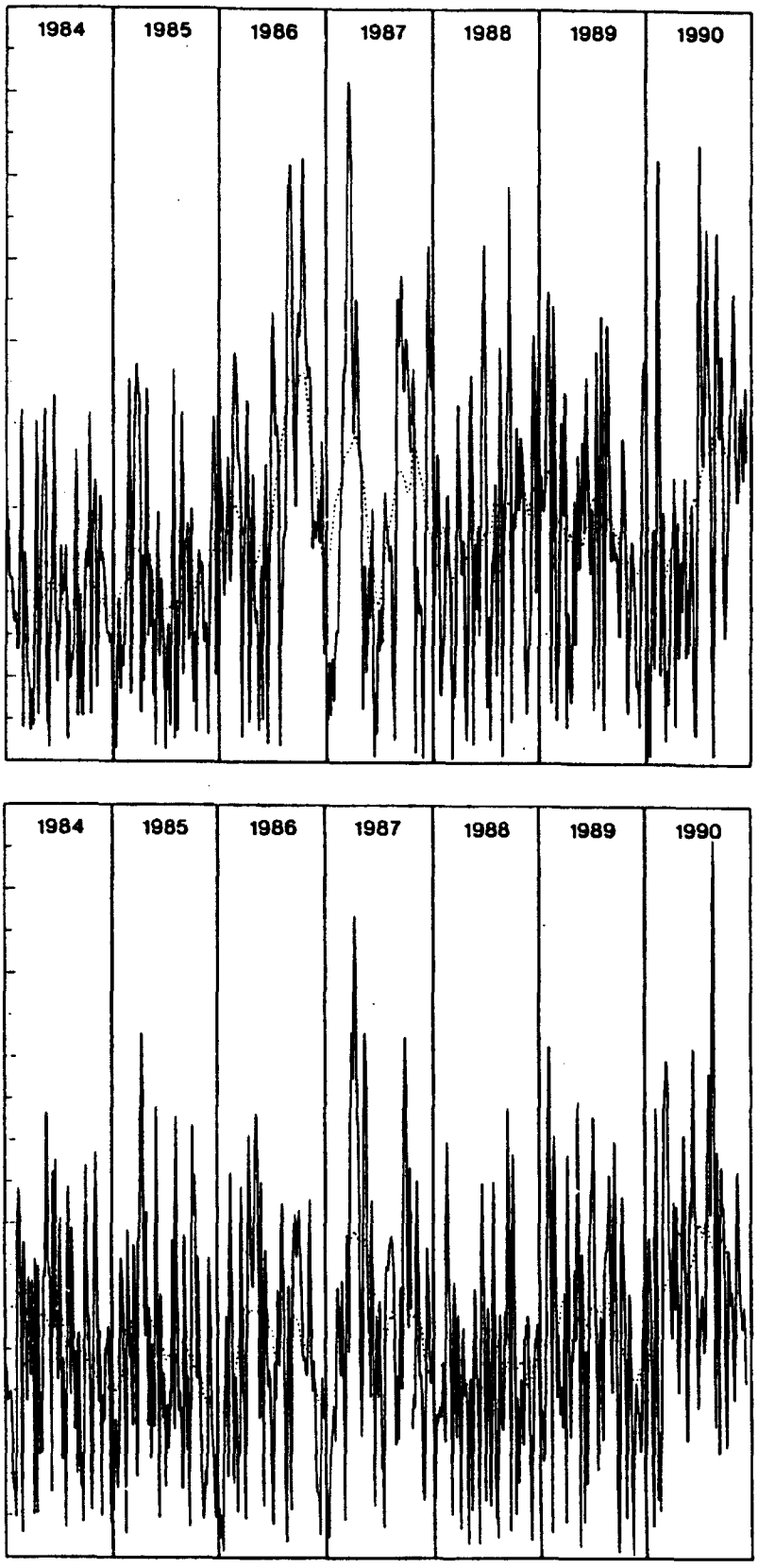

Fig. 16.5. Amplitude time series of two POP modes identified in the daily intraseasonal variability monitored by three equatorial moorings at 165°E, 140°W and 110°W. The top curve refers to the 120-day mode, and the bottom one to the 65-day mode. The continuous line represents the coefficients of the filtered data, after retaining all variability on time scales shorter than half a year; the dashed line is a smoothed version of the continuous line. The amplitudes are normalized to standard deviation 1. The years are given as May to April intervals (thus "1984" represents the time from May 1984 until April 1985). (From von Storch, 1993.)

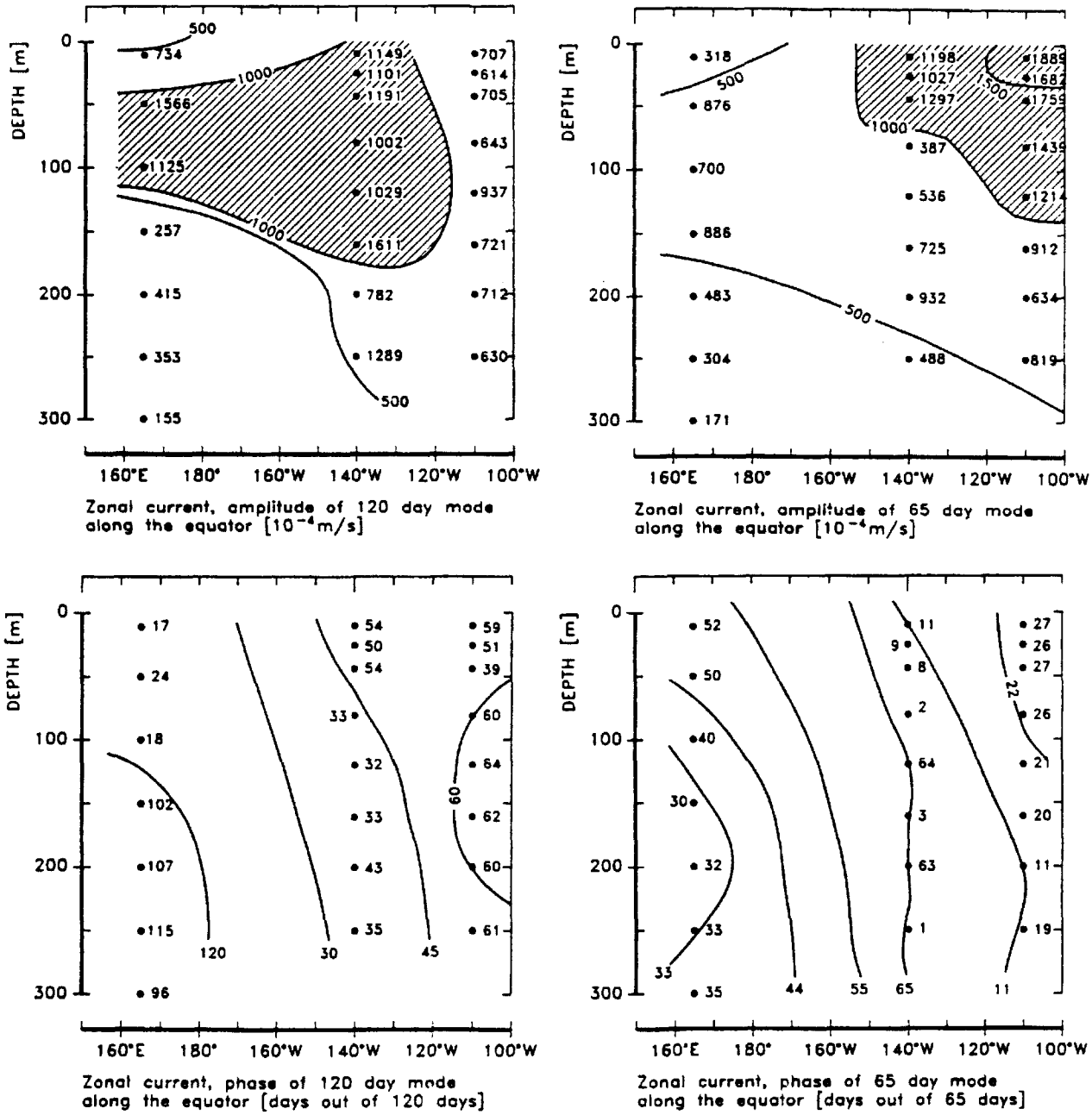

Fig. 16.6. Amplitude and phase distributions of the two eastward-propagating oscillatory POPs of the zonal currents at three equatorial moorings at 165°E, 140°W and 110°W. The coefficient time series are normalized to unity so that the amplitude pattern represents typical distributions in 10^{-4} m s^{-1}. The phases are given in days relative to the base period of 120 and 65 days. (From von Storch, 1993.)

tions 3.1 to 3.3) and *canonical correlation analysis* (Section 4), do not exploit the temporal sequence of the events; instead, any inference about temporal statistics must be done a posteriori by an analysis of the expansion coefficient time series $\alpha_k(t)$ in (2). However, generalizations of empirical orthogonal functions that describe both the spatial and temporal behavior are discussed in Section 3.4.

3.1. Conventional Empirical Orthogonal Functions

Empirical orthogonal functions (EOFs) are a widely used tool in meteorology, oceanography and climatology. This technique, first proposed by Pearson (1902) and extended by Hotelling (1936), has long been used in other disciplines under the

name of *principal component analysis*. It was introduced to earth sciences by Lorenz (1956), who seems to be responsible for the name *Empirical Orthogonal Functions*. [For an extensive description, see Jollife (1986) or von Storch (1995a). The algebraic and geometric aspects are presented in great detail by Preisendorfer (1988).]

The basic idea of principal component analysis is to identify a series of orthogonal patterns $\vec{e}^k$, called in our discipline EOFs, which for any number K minimize the mean-squared error

$$\sum_t \left[\vec{\mathbf{X}}(t) - \sum_{k=1}^{K} \alpha_k(t) \vec{e}^k \right]^2 \tag{18}$$

The index t usually counts time, but it could also be any index counting statistically equivalent realizations of the m-dimensional random vector $\vec{\mathbf{X}}$. In many cases the components of the random vector will be the same variable at different locations, such as the alongshore current at different depths. However, in other applications the dimension of the vector $\vec{\mathbf{X}}(t)$ can be increased to include different variables, for instance by considering simultaneously alongshore and across-shore currents, wind stress and temperature. In such cases some normalization may be needed for dimensional consistency.

In most cases the time mean is subtracted from the data, so that (18) represents a variance. An EOF analysis can also be done for uncentered data; then the expression *variance* has to be replaced by *second moment*. In the following we tacitly assume that the time mean has been subtracted.

Because of the orthogonality of the EOFs the optimal coefficients $\alpha_k(t)$ are given as the scalar product of the state vector $\vec{\mathbf{X}}$ with the EOF $\vec{e}^k$:

$$\alpha_k(t) = \langle \vec{\mathbf{X}}(t), \vec{e}^k \rangle \tag{19}$$

These coefficients are called *EOF coefficients,* or *principal components* in the statistical literature. They are pairwise uncorrelated.

It can be shown that the EOFS are the eigenvectors of the sample covariance matrix

$$\Sigma = \frac{1}{T} \sum_t \vec{\mathbf{X}}(t) \vec{\mathbf{X}}(t)^T \tag{20}$$

(with T denoting the total number of observations and the superscript ^{T} denoting transposition) and that the EOF coefficients are statistically independent if the involved distributions are Gaussian.

If the EOFs are normalized, the total variance of the vector time series $\vec{\mathbf{X}}$ may be decomposed into independent contributions from the EOFs:

$$\sum_t |\vec{\mathbf{X}}(t)|^2 = \sum_{t,k} \alpha_k(t)^2 \tag{21}$$

Thus if no truncation is made, the EOFs reproduce the original fields.

The theoretically more consistent way to introduce the EOFs is to use the expectation operator in (18, 20) instead of sample moments. The general philosophy in statistical analysis is first to define certain parameters of a random variable to derive their properties and their informational value and to understand their significance for the problem at hand. (A *parameter* of a random variable is any characteristic of it; it may be a number, such as the variance, a series, such as the covariance function, a matrix, such as the covariance matrix, or a continuous function, such as the spectrum.) At that stage there is no uncertainty from limited sampling. In practical applications, however, only a limited amount of evidence is available. Then the "true" parameter (e.g., the covariance matrix) is replaced by its estimate [e.g., the sample covariance matrix (20)], and it is hoped that the estimate has the same informational value as the true parameter. Statistical theory tells us under which circumstances this hope is warranted. Confirmatory analysis is the tool to assess the probabilities of deriving correctly the true ensemble properties from the randomly drawn sample. In the present context, this need not be discussed, as EOFs are used primarily as descriptors of a given set of observations. For more details, see, for example, von Storch (1995a) and von Storch and Zwiers (in press).

An important aspect of this sampling uncertainty refers to what is called *significance* in EOF analysis. This aspect is a standard ritual in our discipline and is often confused and merely used as a black box, without real understanding of the statistical problem. Again, we do not go into detail. We outline briefly the problem, the standard approach and its limitation. For further reading we again recommend von Storch (1995a) and von Storch and Zwiers (in press).

The basic question is to what extent the estimated EOFs (i.e., the eigenvectors of the sample covariance matrix) resemble the true EOFs (i.e., the eigenvectors of the true covariance matrix). The true and sample EOFs differ for two reasons.

The first reason is the standard problem in statistical inference: namely, the *uncertainty* due to the limited sampling. A basic empirical rule is that the low-indexed EOFs, with the largest eigenvalues, are best estimated, while the high-index EOFs are poorly determined. The signals associated with low-index EOFs have large amplitudes and are in most cases sampled several times in the data set analyzed. However, the EOFs derived from a sample represent the total sample variance. Therefore, the events in the sample, which are *not typical* for the random variable as a whole but *specific to the sample,* are also represented by some, usually high-indexed, sample EOFs. On the other hand, the structures of small-amplitude and rare true signals, which are not well sampled by the available data set, are hardly captured by the high-index sample EOFs. Therefore, the high-indexed EOFs are prone to severe sampling errors. Unfortunately, there is no objective way to assess how severe the sampling uncertainty is.

The second reason has to do with the buzzword *degeneracy.* When an eigenvalue is multiple, or degenerate, its eigenvector directions are no longer uniquely determined. If, for instance, an eigenvalue is double, any linear combination of its two eigenvectors is an eigenvector as well. Obviously, the spatial structure of these linearly combined vectors may vary widely. The problem also occurs with sample covariance matrices: There is an *indeterminacy* due to degeneracy if the (sample) eigenvalues are close to each other. North et al. (1982) have proposed a rule of thumb to diagnose when such an indeterminacy prevails and the EOFs patterns are not uniquely

determined, using the slope of the (estimated) eigenvalue spectrum. Preisendorfer et al. (1981) (see also Preisendorfer, 1988) have produced many rules of the sort aiming at identifying significant EOFs [i.e., those that can be distinguished from those of white noise (a white noise process has identical true eigenvalues, hence is an extreme case of degeneracy)]. Many practitioners misunderstand the meaning of these tests by assuming that the selected patterns represent a reliably estimated signal. In fact, the eigenvalues can be different from those of white noise but the patterns poorly estimated. Note that the degeneracy of the eigenvectors needs not be an artifact of the limited sampling. Indeed, a single EOF can only represent a spatially standing pattern, and two EOFs or more are needed to represent a propagating pattern. When a spatially propagating oscillatory mode is in the data, this mode is represented by two EOFs with similar eigenvalues. They are thus degenerate, but physically significant, and their principal components are correlated for lags different from zero (see also Section 3.4).

In Section 2.1 we have discussed the analysis by Kundu et al. (1975) of alongshore currents monitored at four depths in terms of dynamical normal modes. These authors also made an EOF analysis of this data. The first two EOFs are shown in Fig. 16.3 (dashed line). They represent 91% and 6% of the total variance, hence explain almost the entire variability of the data set (note that four EOFs suffice to represent the entire variance). The first EOF is almost constant and has no sign reversal. It is similar in pattern to the theoretically derived barotropic mode (solid line). As the first barolinic mode, the second EOF has one sign reversal. However, the zero-crossing depths are rather dissimilar, with 30 m for the baroclinic mode and somewhere between 40 and 60 m for the second EOF. The time series $\alpha_k(t)$ fitted to the theoretical modes and to the EOFs shows very high correlations (Fig. 16.4*c*): 0.98 for the barotropic mode and the first EOF, and 0.97 for the first baroclinic mode and the second EOF.

There are several reasons why the empirical modes account for a considerably larger amount of variance than the dynamical ones. The EOF are *constructed* to represent a maximum of variance at the four depths, while the normal mode describe the simplified dynamics in the whole water column. Moreover, the EOFs are optimal for the representation of the variability within the available sample. If a larger sample had been considered, the EOFs would have been fitted to an environment with more diverse variability, with the probable result that the main EOFs would account for less variance. Indeed, it can be shown that the variance attributed to the first few EOFs is always overestimated, in particular when the sample size is small (von Storch and Hannoschöck, 1986).

In this example, the empirically derived modes compared favorably with the normal modes so that the dynamical argument and the simplifying assumptions leading to (13)–(15) were confirmed. This shows that empirical modes can also be valuable tools to evaluate the significance of theoretically derived structures.

3.2. *Complex EOFs*

The conventional EOF analysis is designed to expand scalar fields. To deal with two-dimensional vectors such as horizontal currents, say $(\vec{\mathbf{u}}, \vec{\mathbf{v}})$, one can, as noted before, form an extended vector $\vec{\mathbf{X}}$ which contains in the first half of its components the contributions from $\vec{\mathbf{u}}$ and in the second half that of $\vec{\mathbf{v}}$, and then use standard EOF

analysis, the dimension of the space being larger. This is easy to do and can be generalized to more dimensions.

An alternative in the two-dimensional case is to combine the two components into one complex vector, as done by Kundu and Allen (1976):

$$\vec{\mathbf{X}}(t) = \vec{\mathbf{u}}(t) + i\vec{\mathbf{v}}(t) \tag{22}$$

Then the approach outlined above may be retained, except that the metric is changed. The covariance matrix is given by the cross products of state $\mathbf{X}_l$ at location l with the complex conjugate state $\mathbf{X}_j^*$ at location j:

$$(\sigma_{jl}) = \left(\frac{1}{T} \sum_t \mathbf{X}_j(t)^* \mathbf{X}_l(t) \right) \tag{23}$$

(the asterisk indicates complex conjugation). The eigenvectors of a Hermitian matrix are orthogonal, so that

$$\langle \vec{e}^j, \vec{e}^l \rangle = \sum_r \vec{e}^{j*}_r \vec{e}^l_r = \delta_{jl} \tag{24}$$

and its eigenvalues are real. The EOF coefficients $\alpha_k(t)$ are complex numbers determined by (19).

The interpretation of these complex vectors is eased when they are expressed in polar coordinates:

$$\vec{e}^k(r) = A^k(r) e^{iB^k(r)} \tag{25}$$

so that a complex EOF is characterized by two patterns: an amplitude pattern $A^k(r)$ and a directional pattern $B^k(r)$. Therefore, a complex EOF may be displayed by a vector field, with vectors of length $A^k(r)$ and direction $B^k(r)$.

In the formulation (2) the full signal is expressed as a sum of products of a (complex) coefficients $\alpha_k(t)$ times a (complex) patterns $\vec{e}^k$. When we use the notation (25) and also write the amplitude in polar coordinates, $\alpha_k(t) = \eta_k(t) \exp[i\psi_k(t)]$, we see that the product $\eta_k(t) \exp[i\psi_k(t)] A^k(r) e^{iB^k(r)} = \eta_k(t) A^k(r) \exp[i(B^k(r) + \psi_k(t)]$ has everywhere the same amplitude distribution as the EOF, apart from the common time-varying factor $\eta_k(t)$ and the same directional distribution, apart from the common rotation $\exp[i\psi_k(t)]$. The latter points to the superiority of the complex EOF representation over the combined traditional one when the motions are polarized, as in near-inertial waves. Note that the complex EOFs are invariant to rotation.

Brink and Muench (1986) have applied complex EOF analysis extensively to the data set in Fig. 16.1. We show the first complex EOF for the current, at the mooring C1 at 5, 10, 20, 30 and 45 m, and for the wind-stress factor at the two buoys NS and NC (Fig. 16.7). The most important pattern of current variability is characterized by a maximum at the surface and by an anticlockwise veering with increasing depth. Whatever the anomalous (i.e., the deviation from the time mean) current is near the

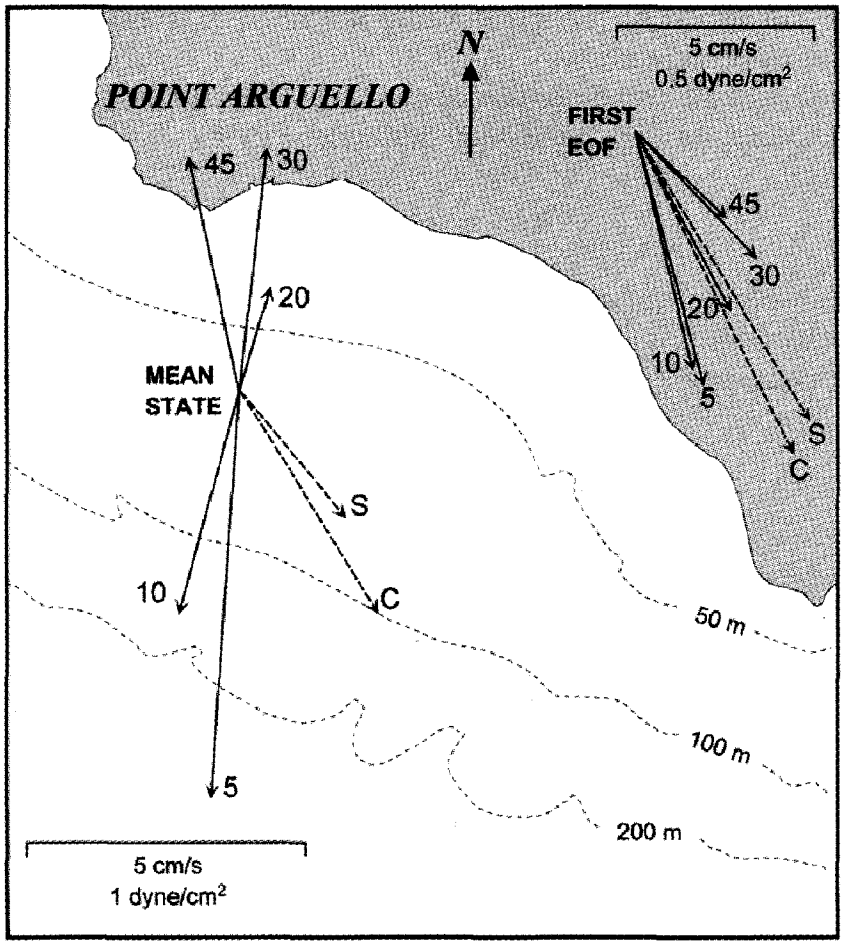

Fig. 16.7. Mean and first complex EOF of currents (solid arrows; depth in meters given by numbers) at mooring C1 (see Fig. 16.1) and wind stress at buoys NS and NC (labelled S and C). The mean state is the time average of the currents and the wind stress, and the first complex EOF was calculated separately for the wind stress and for the currents. (Adapted from Brink and Muench, 1986.)

surface, that at larger depths is smaller and more to the left. On the other hand, the first EOF of the wind stress indicates that it varies similarly at the two locations.

The first two complex EOFs of the wind stress and of the currents are well connected with a complex correlation of 0.62 between the EOF coefficients. The directional sense of this correlation is shown in Fig. 16.7 by the relative orientation of the two sets of vectors. Close to the surface, the current changes are slightly to the right of the wind-stress changes; at greater depths the veering is to the left.

Finally, there are two things to be noted: (1) Complex EOFs give no information about preferred directions of the vectors; this information can only be obtained by an a posteriori statistical analysis of the complex EOF-coefficient time series. (2) The complex EOF analysis as defined in this section must not be confused with the frequency-dependent EOF analysis, which is also called complex EOF analysis (CEOF). We discuss the latter briefly in Section 3.4.

3.3. Computational Aspects of EOF Analysis

The EOFs are the eigenvectors of the (sample) covariance matrix (20). There are two main approaches to the computation. In the direct approach the sample covariance

matrix is determined and the eigenproblem solved directly. In the other approach, a singular value decomposition (SVD) is made (e.g., Rasmussen et al., 1981; Seber, 1984; Kelly, 1985). The final product is the same in both cases, but the SVD approach is in general computationally cheaper (Kelly, 1988) and under certain circumstances numerically more robust.

In the direct approach the sample covariance matrix is expressed as the product $\Sigma = (1/T)\mathcal{X}\mathcal{X}^T$, with the *data matrix*

$$\mathcal{X} = \begin{pmatrix} X_1(1) & X_1(2) & \cdots & X_1(T) \\ X_2(1) & X_2(2) & \cdots & X_2(T) \\ \vdots & \vdots & \ddots & \vdots \\ X_m(1) & X_m(2) & \cdots & X_m(T) \end{pmatrix} \tag{26}$$

The T columns of the $T \times m$ data matrix $\mathcal{X}$ are the vectors of observations taken at the time $t = 1, \cdots, T$; the rows mark the $r = 1 \cdots m$ locations or variables. The matrix product $\mathcal{X}\mathcal{X}^T$ is a quadratic $m \times m$ matrix even if $\mathcal{X}$ itself is not quadratic. Then the eigenproblem $\mathcal{X}\mathcal{X}^T\vec{e}^k = \lambda_k\vec{e}^k$ is solved. Sometimes the numerical problem can be considerably eased by the following algebraic trick (von Storch and Hannoschöck, 1984): *If $\mathcal{Y}$ is a $n \times m$ matrix, then $\mathcal{A} = \mathcal{Y}\mathcal{Y}^T$ and $\mathcal{A}^T = \mathcal{Y}^T\mathcal{Y}$ are $n \times n$ and $m \times m$ matrices which share the same nonzero eigenvalues. If $\mathcal{Y}\vec{q}$ (or $\vec{r}$) is an eigenvector of $\mathcal{A}$ to the eigenvalue $\lambda \neq 0$, then $\vec{q}$ (or $\mathcal{Y}^T\vec{r}$) is an eigenvector of $\mathcal{A}^T$ to the same eigenvalue* λ. Thus one should calculate the eigenvalues and eigenvectors of the smaller of the two matrices $\mathcal{X}^T\mathcal{X}$ and $\mathcal{X}\mathcal{X}^T$. In practical situations the number of samples T is often much smaller than the dimension m of the fields, whence only estimates of the first T EOFs can be derived from the data set.

The alternative approach uses the singular value decomposition (SVD) of $\mathcal{X}$,

$$\mathcal{X}^T = \begin{pmatrix} \alpha_1(1) & \alpha_2(1) & \cdots & \alpha_T(1) \\ \alpha_1(2) & \alpha_2(2) & \cdots & \alpha_T(2) \\ \vdots & \vdots & & \vdots \\ \alpha_1(T) & \alpha_2(T) & \cdots & \alpha_T(T) \end{pmatrix} \mathcal{D}(\vec{e}^1|\cdots|\vec{e}^m)^T \tag{27}$$

where $\mathcal{D}$ is a rectangular $T \times m$ matrix with zero elements outside the diagonal and positive elements on the diagonal: $d_{ij} = s_i\delta_{ij} \geq 0$. The quadratic $T \times T$ and $m \times m$ matrices to the right and left of $\mathcal{D}$ are orthogonal and contain the EOF coeffiients and the EOFs, respectively. Note that if $T > m$, the last EOF coefficients $\alpha_j, j = m + 1, \cdots, T$, are zero. Similarly, when $m < T$, the last EOFs $\vec{e}^k, k = m + 1, \cdots, T$, are zero. The eigenvalues of the covariance matrix Σ are s_i^2.

Sometimes the data field is not complete (i.e., observations for some times and locations are lacking). One solution is to fill the gaps (i.e., to "repair" the data field by spatial and temporal interpolation). The drawback of this frequently used approach is that it introduces additional artificial information into the analysis, which gives the same weight to the real and interpolated data. A simple and methodically sound alternative is to estimate the elements σ_{jl} of the covariance matrix directly by forming sums only over those T' times t' when simultaneous observations exist at the locations j and l:

$$\sigma_{jl} = \frac{1}{T'} \sum_{t'} \vec{\mathbf{X}}_j(t')\vec{\mathbf{X}}_l(t') \tag{28}$$

However, this estimated covariance matrix is no longer positive definite, so that some eigenvalues may be negative, and the eigenvectors are no longer strictly orthogonal. Thus the EOF coefficients should be fitted by least squares:

$$\alpha_j(t) = \frac{\sum_{r'} X_{r'}(t)\vec{e}^{\,j}_{r'}}{\sum_{r'} (\vec{e}^{\,j}_{r'})^2} \tag{29}$$

where the indices r' represent those components of $\vec{\mathbf{X}}$ for which observations exist at time t.

3.4. *Generalizations of the EOF Technique*

In the zoo of pattern-related techniques a number of other animals may be found with the family name EOFs. Two of them, the frequency-domain EOFs and the extended EOFs are indeed related to the conventional EOF approach, and they are designed to identify spatially propagating oscillatory behavior. The third technique, the rotated EOFs, represent a different sort of species, which is intended to identify robust, and often compact, patterns. In the following we briefly outline the basic concepts and let the reader get the details from the cited literature.

Extended EOFs and Singular Systems Analysis

A straightforward way to identify systematic space/time variations in a vector field $\vec{\mathbf{Y}}(t)$ is to link together the field at m consecutive times into a larger vector, say $\vec{\mathbf{X}}(t) = (\vec{\mathbf{Y}}(t), \vec{\mathbf{Y}}(t+1), \cdots, \vec{\mathbf{Y}}(t+m-1))^T$, and apply the EOF algorithm to this concatenated vector. Since the field is considered simultaneously at m successive times by sliding a "window" of width m down the time series, the temporal evolution during m time steps becomes part of the eigenvectors $\vec{e}^{\,k}$ called extended EOFs (EEOFs) pattern by Weare and Nasstrom (1982), who first introduced the method. The principal components of EOF coefficients are given by

$$\alpha_k(t) = \langle \vec{\mathbf{X}}(t), \vec{e}^{\,k} \rangle = \sum_{j=0}^{m-1} \vec{\mathbf{Y}}(t+j)^T \vec{e}^{\,k}_{j+1} \tag{30}$$

where $\vec{e}^{\,k}_j$ is the eigenvector pattern for time j, with $j = 1, \cdots, m$, so they represent a space/time filtered version of the original vector time series $\vec{\mathbf{Y}}(t)$. [A time filter is the operator $X(t) \rightarrow \tilde{X}(t)$ with time series $X(t)$ and $\tilde{X}(t)$ such that $\tilde{X}(t) = \Sigma_{j=-m+1}^{m-1} a_j X(t+j)$ with a set of numbers, or "weights" $\{\alpha_{-m=1}, \cdots, a_{-1}, a_0, a_1, \cdots, a_{m-1}\}$. In case of (30), one has $a_j = 0$ for $j < 0$ and $a_j = \vec{e}^{\,k}_{j+1}$ for $j \geq 0$.] In contrast to time filters that are designed to emphasize certain frequency bands, or to POP anlysis, both of which

constrain the time evolution of the spatial patterns, the EEOF analysis is an adaptive filtering technique (i.e., its filtering characteristics are determined by the data). The original vector time series may be reconstructed (except for the endpoints) by any of the following $j = 1, \cdots, m$ representations:

$$\mathbf{Y}(t) = \sum_{k=1}^{m} \alpha_k(t - j + 1)\vec{e}^{\,k}_{j+1} \tag{31}$$

for $j = 1, \cdots, m$. Thus the original vector time series is decomposed into m series, each of which is a space/time-filtered version of the original series.

Just as EOFs identify the main patterns of variability that are coherent in space, EEOFs identify those that are coherent in space and time. The EEOF analysis emphasizes cyclical behaviors in the evolution. If the data field $\vec{\mathbf{Y}}(t)$ features a propagating wavelike structure, then this signal will appear as a pair of EEOFs explaining a similar amount of variance, having the same frequency, and dephased by a quarter period. In this case the EEOFs need to be considered as a pair. However, the presence of such an EEOF pair does not warrant that it corresponds to an oscillatory mode, as discussed below. The value of the window width m is a function of the expected time scales in the problem: It should be large enough to encompass all the stages of the expected oscillating structures, and small enough to retain sufficient independent degrees of freedom in the time series. As discussed in Vautard and Ghil (1989) and Allen and Smith (1994), the EEOF pairs are very effective narrow-bandpass filters, and increasing m increases the spectral resolution.

The EEOF analysis was introduced more theoretically in the context of dynamical systems by Broomhead and King (1986), who used it to visualize qualitative dynamics in noisy data and developed the method for the one-dimensional case under the name of singular system analysis (SSA), which is technically a univariate EEOF. In this case, $\vec{\mathbf{X}}(t)$ is composed of m values of the scalar time series $\mathbf{Y}(t), \cdots, \mathbf{Y}(t+m-1)$, and the EEOF patterns are temporal sequences of length m, so in (30) the $\vec{e}^{\,k}_j$ are adaptively determined filter weights. The method was expanded and thoroughly discussed by Vautard and Ghil (1989), who renamed it singular spectrum analysis and, in particular, emphasized its power for searching for physical oscillations and to analyze, filter and predict time series. However, a recent in-depth analysis by Allen and Smith (1996), who have made extensive Monte Carlo simulations, suggests that much care is needed when using SSA (or EEOF analysis) to detect quasi-periodicities in time series. Indeed, when applied to limited time series generated by a univariate AR(1) process that features no oscillatory modes, SSA often identifies pairs of the orthogonal patterns with similar eigenvalues, which may then be misinterpreted as oscillatory modes. [A univariate AR(1) process is the discretization of a first-order linear differential equation driven by white noise: $\mathbf{X}(t + 1) = a\mathbf{X}(t)$ + noise with some constant $|a| < 1$. The multivariate version has been given in (12).] Tests with Monte Carlo simulations are thus needed to distinguish such spurious signals from a genuine oscillatory mode.

Forecast exercises relating $\alpha(t)$ to $\alpha(t + L)$ usually give high forecast skills if $L < m$, which is part due to the fact that the numbers

$$\alpha_k(t) = \sum_{j=0}^{m-1} \mathbf{Y}(t + j)e_{j+1}^k$$

$$\alpha_k(t + L) = \sum_{j=L}^{m+L-1} \mathbf{Y}(t + j)e_{j-L+1}^k$$

are in part derived from the *same* observations, namely $\{\mathbf{Y}(t + L) \cdots \mathbf{Y}(t + m - 1)\}$. [For measuring forecast skill, refer to Livezey (1995).] Mathematically, the high forecast skills are correct, but the informational value of a correctly forecasted $\alpha_k(t+L)$ is considerably less than the informational value of a correctly forecasted $\mathbf{Y}(t+L)$. The questionable value of such "forecasts" is further demonstrated by the sudden drop of skill when the lead time L becomes longer than the window length m.

The generalization of SSA to multivariate time series (e.g., Plaut and Vautard, 1994), often referred to as multichannel SSA (MSSA), is identical to EEOF analysis, notwithstanding the claim (Vautard, 1995) that MSSA applies to the case where the temporal dimension is greater than the spatial one, since in practice the latter is usually artificially reduced by EOF truncation. In any case, a careful reading of the SSA literature is recommended to EEOF/MSSA users, for its attention to the power and pitfalls of the method.

Since we know of no application to coastal problems, we illustrate the EEOF analysis by an application to the North Atlantic decadal variability in a coupled ocean–atmosphere model (Zorita and Frankignoul, 1997). The method was applied to a 325-year model integration, using yearly averages of three EOF-truncated North Atlantic fields: sea level air pressure, sea-surface temperature and temperature at 450 m depth, which were considered simultaneously after suitable normalization. Using a window width $m = 20$ (the results are not very sensitive to m), they detected two eigenvector pairs among the combined EEOFs that stand out above the noise level (note that the three dominant EEOFs have long, unresolved time scales that correspond primarily to the initial model adjustment). Here we discuss the first EEOF pair, which suggests the presence of a quasi-oscillatory mode. Figure 16.8 shows the principal component of one of the eigenvectors (top) together with its autocorrelation function (that of the other is similar) (bottom, solid line) and the cross-correlation function between the principal component pair (bottom, dashed line). The correlations suggests that the combined eigenvector pair represents a damped oscillatory mode with a period of about 20 years. Its decay time is less than the period, but it cannot be estimated from the correlations if it is smaller than the window width, because of the bandpass filtering inherent to the technique. As independent but spatially correlated AR(1) processes may also lead to EEOF pairs in finite samples, although they do not correspond to oscillations (see Allen and Smith, 1996), Monte Carlo simulations were used as in Robertson (1996) to test this hypothesis, which was rejected. Thus the mode differs from (nonoscillating) POPs with real eigenvalues. Whether it represents a true oscillation (e.g., POPs with complex eigenvalues, see Section 2.2) or a propagating mode cannot be established in this way, however. Thus an interpretation of a pair of EEOFs as an oscillation must be based on physical grounds and/or other evidence.

Figure 16.9 represents the reconstructed anomaly patterns at 3-year intervals dur-

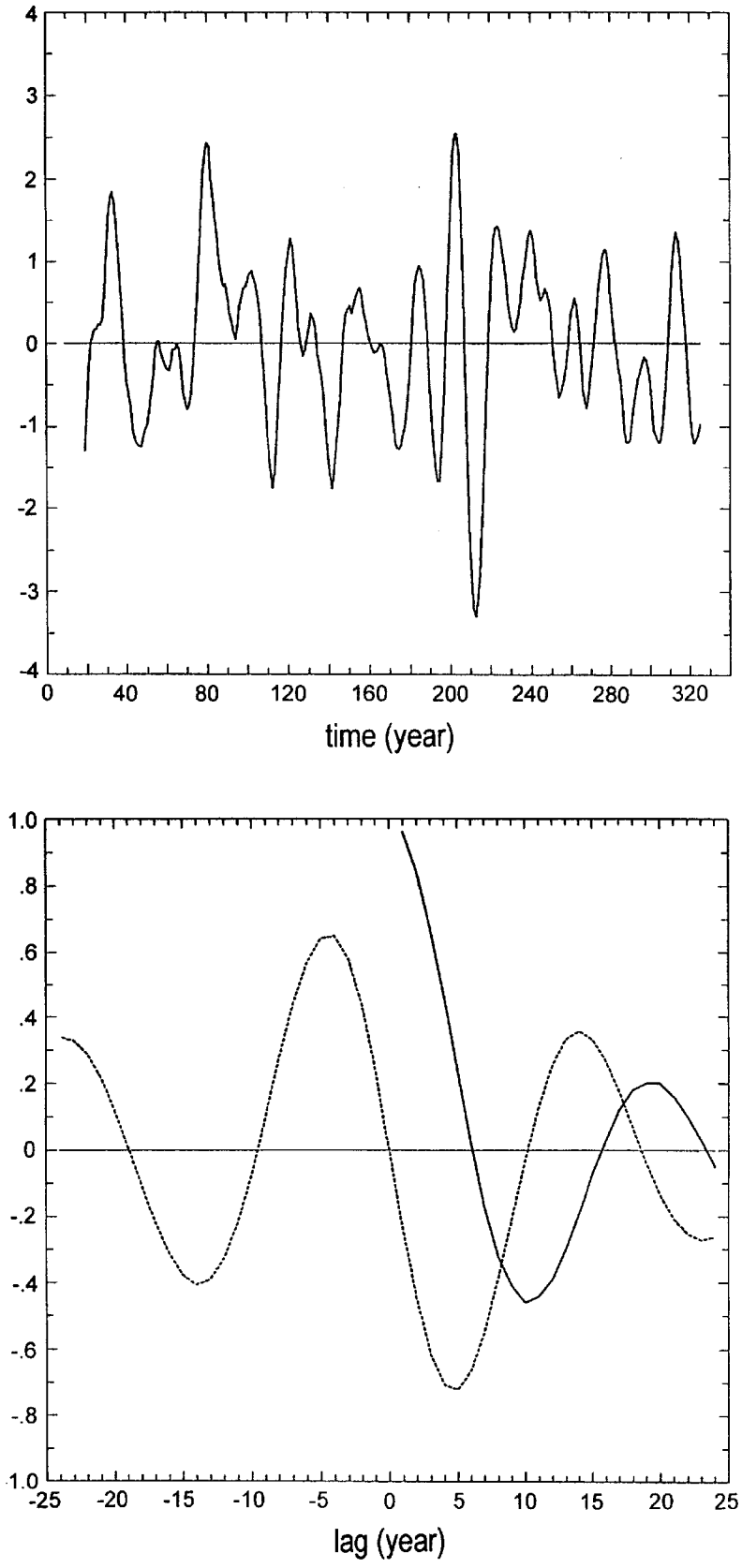

Fig. 16.8. Simultaneous EEOF analysis of air pressure, sea-surface temperature and subsurface temperature at 450 m simulated in a 325-year simulation with a coupled ocean–atmosphere GCM. Top: time series of the EEOF coefficient of one of the eigenvectors of the 20-year EEOF pair; bottom: autocorrelation function of this time series (solid) and cross-correlation function between the principal components of the EEOF pair (dashed). (From Zorita and Frankignoul, 1997.)

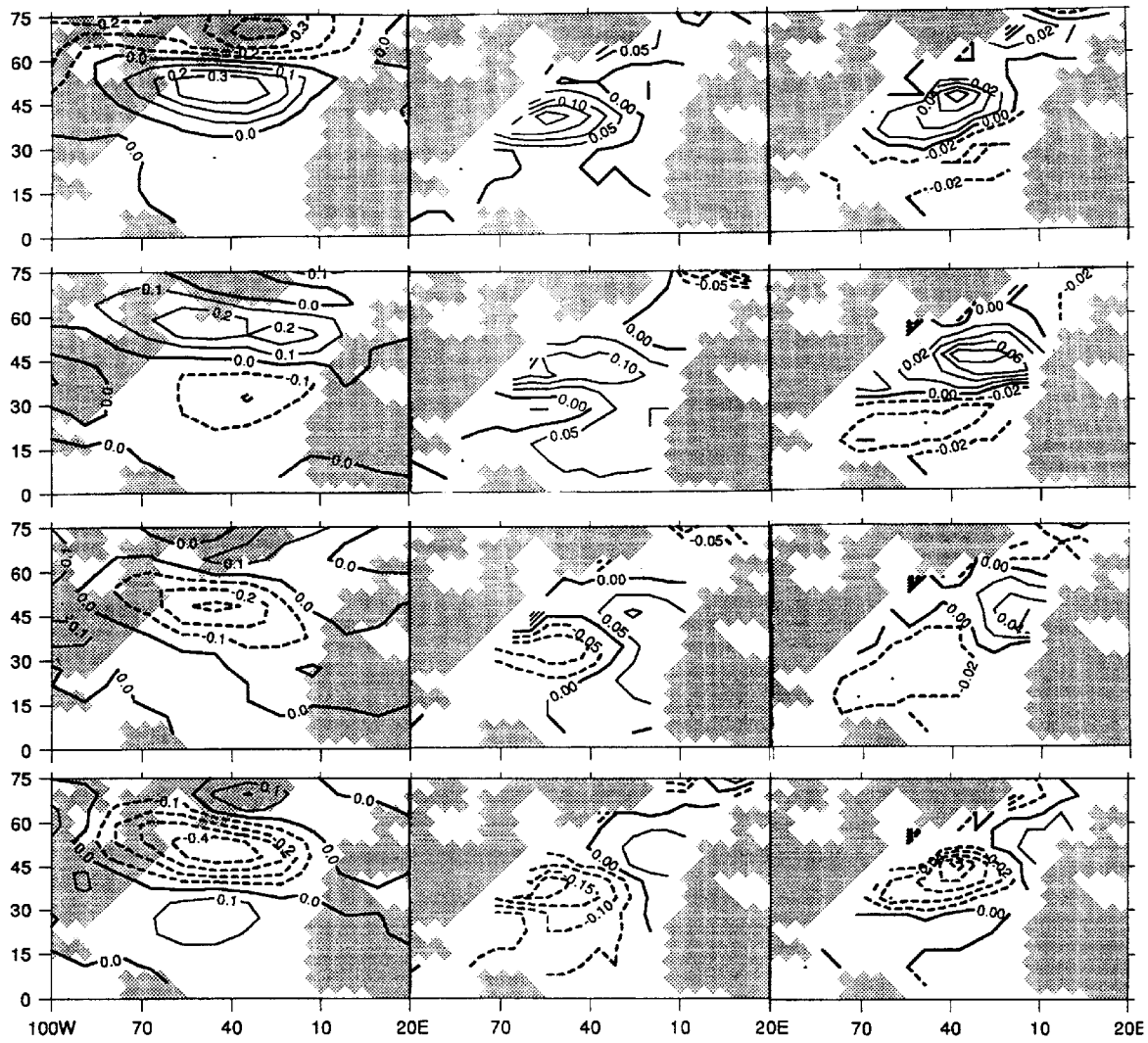

Fig. 16.9. As in Fig. 16.8. Reconstructed pattern associated with the 20-year mode at, from top to bottom, year 1, 4, 7 and 10 for, from left to right, sea level (mb), sea-surface temperature (K) and potential temperature at 450 m depth. The amplitudes represent rms values, and only half of the cycle has been represented. (From Zorita and Frankignoul, 1997.)

ing one-half of the oscillation. The starting year was chosen arbitrarily, and the other half cycle is similar but with the reversed sign. The air-pressure pattern starts as a positive anomaly centered around 50°N, flanked by small negative anomalies northward and southward. The positive anomaly migrates northward and is replaced by an anomaly of the reversed sign coming from the south. The mid-Atlantic anomalies reach their strongest amplitudes at years 1 (positive) and 10 (negative) and their minimum amplitude at year 5.

The sea-surface temperature is characterized by positive anomalies centered near 40°N, which slowly wander northward while decaying and being replaced by similar anomalies of the reversed sign; there are also weaker anomalies of the opposite sign in the eastern side of the basin. The modulation of the amplitude of the sea-surface temperature variations is in phase with that of air pressure, and its space/time structure is consistent with an upper ocean response to the atmospheric forcing (e.g., the temperature is colder than normal in areas where the anomalous wind comes from colder regions in winter and vertical mixing is enhanced).

The 450-m temperature is dominated by a dipole structure that remains stable until year 4, then weakens while rotating clockwise. By year 10, the dipole has reversed sign. The amplitude modulation at 450 m follows that of air pressure and sea-surface temperature by a few years, and the pattern is consistent with a gyre adjustment to the changes in the wind-stress curl that are associated with the SLP pattern.

The mode is strongest for 450-m temperature (it is detected in SST alone only if the data are low-passed to increase the signal-to-noise ratio), so it is difficult from this analysis alone to differentiate between a true oscillatory mode sustained by some positive air–sea feedback and a propagating or advected ocean interior response to the atmospheric forcing, which also forces the surface mixed layer [see the discussion in Zorita and Frankignoul (1997)].

Frequency-Domain EOFs

A single EOF can represent variations in time that are in phase or out of phase along a data array (i.e., standing oscillations), but not features that have variable phase relationships, such as propagating waves or moving structures. However, a single EOF can represent propagating features if one works either with the lagged covariance matrix, as in EEOF analysis, or with the cross-spectrum matrix, as in EOF analysis in the frequency domain, in place of the (unlagged) covariance matrix as in standard EOF analysis.

The idea of using as EOFs the eigenvectors of the cross-spectrum matrix was introduced in the ocean–atmosphere context by Wallace and Dickinson (1972). It was extended by Rasmusson et al. (1981), who used the Hilbert transform to present the method, and by Barnett (1983), who related the two approaches and thoroughly discussed their properties. See Brillinger (1975) for earlier references and a more theoretical presentation in terms of multivariate time-series filtering.

In frequency-domain EOF (FDEOF) analysis, the data vector $\vec{\mathbf{X}}_t$ is augmented by its Hilbert transform $\vec{\mathbf{X}}_t^H X(t)$ to form a complex vector $\mathbf{Y}(t)$:

$$\vec{\mathbf{Y}}(t) = \vec{\mathbf{X}}(t) + i\vec{\mathbf{X}}^H(t) \tag{32}$$

The Hilbert transform is the raw time series phase shifted in the frequency domain by 90° for each of its Fourier components, so it is characteristic of the time derivative of the series. For example, a cosine time series is transformed into a sine one, and (32) into a cosine series.

The covariance matrix of the complex series $\vec{\mathbf{Y}}(t)$ is estimated as in (20) by

$$\Sigma = \frac{1}{T} \sum_t \vec{\mathbf{Y}}(t)^* \vec{\mathbf{Y}}(t)^T \tag{33}$$

and it follows from spectral analysis theory that it represents the cross-spectrum matrix averaged over some (wide) frequency band. The frequency bandwidth can be controlled by suitable filtering operation, however. When $\vec{\mathbf{X}}(t)$ is filtered so that only a narrow band around a frequency ω has been retained, the covariance matrix approaches the cross-spectrum matrix in this frequency band. In practice, Σ can thus be estimated using the Hilbert transform [see Barnett (1983) for details], or taken as the cross-spectrum matrix averaged over a specified frequency band.

The matrix Σ is Hermitian, hence its eigenvalues are real and its eigenvectors $\vec{e}^k$ complex, so the latter are best represented by the spatial distribution of their amplitude and phase. The eigenvectors are not determined with respect to rotation, so only the phase differences between different locations is determined. The principal components are also complex, providing useful measures of the stationarity of the dominant features.

The FDEOFs allow to decompose a field into its largest spatially coherent modes in a specific frequency band, thus retaining the phase relationships between the different locations in a data array. In the simple case where only one or two dominant propagating modes are present, the method is powerful and leads to easy interpretation. However, problems may arise if the propagating signal is nondispersive, as in some coastal wave problems and a large frequency bandwidth (with respect to array length) has to be used to reduce the statistical uncertainty in Σ (see Merrifield and Guza, 1990; Johnson and McPhaden, 1993). The interpretation in more complex fields is much more difficult, and the results very sensitive to the bandwidth.

Although FDEOF analysis has been used in coastal oceanography to search for long coastal waves (Wang and Mooers, 1977), we prefer to describe briefly the analysis of equatorial Pacific temperature and current measurements of Johnson and McPhaden (1993), which correspond to the (more extensive) data set analyzed by von Storch with the POP method (Section 2.2). Moored data from 1984 to 1987 were considered at 110° and 140°W along the equator, and the FDEOF analysis based on the cross-spectrum matrix using the frequency band 0.008–0.017 cycles day^{-1} (13 frequencies). All variables and depths were considered simultaneously (28 time series normalized to have the same variance in the frequency band selected). The variance-weighted average in the frequency band corresponds to a period of 83 days. The analysis yielded a first FDEOF representing 68% of the variance. For comparison with Fig. 16.6, we give only the amplitude and phase of the zonal velocity, as well as the percentage of explained variance at each depth (Fig. 16.10, top) and the normalized principal component for a hypothetical location having zero phase (Fig. 16.10, bottom). Although the data set is smaller than that in von Storch (1993), the mode strongly resemble the 65-day POP, especially at 140°W, where the amplitude and vertical distribution are pretty similar, and also for the temperature distribution (not shown). Note in particular that in both analyses, the upper levels lag the lower ones in a similar way and there is enhanced activity during the 1986–1987 ENSO. See Johnson and McPhaden (1993) for an interpretation in terms of equatorial Kelvin waves modulated by their interaction with the equatorial undercurrent system. It should be noted that the 120-day POP mode was not found in the FDEOF analysis. Presumably, this is due to the substantial similarities in the spatial structure of the two POP modes, which conflicts with the orthogonality constraint (in both space and time) inherent to EOF analysis.

Although the basic assumptions behind the FDEOF analysis and the POP method are different, the results are similar in cases where one propagating mode is dominant, and either method can be chosen. One advantage of the FDEOF analysis is that it optimizes the explained variance, but the orthogonality of the EEOFs may be too constraining when there are two or more similar modes at neighboring frequencies (see, however, the section "Rotated EOFs"). The relationship between FDEOFs and POPs is discussed further in Hasselmann (1988) and von Storch (1995b). Presumably, both methods will be less successful when the signal-to-noise ratio is smaller, and in

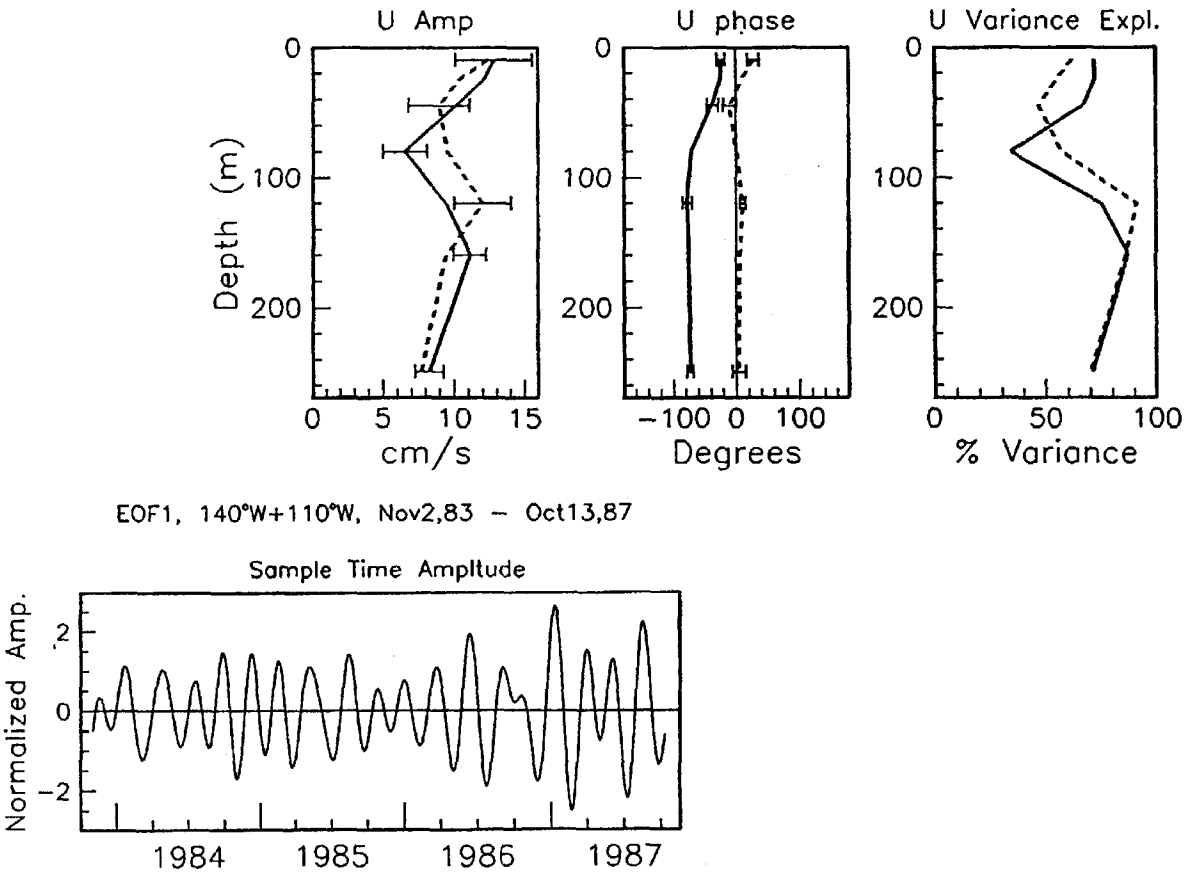

Fig. 16.10. Top: amplitude (left), phase (middle) and percentage of represented variance (right) for the first FDEOF of zonal velocity at 140° (solid line) and 110°W (dashed line) during 1984–1987. Error bars represent an estimate of the standard error. The amplitudes are rms amplitude for the mode. Bottom: principal component time series for a hypothetical location having zero phase. (From Johnson and McPhaden, 1993.)

such cases an EEOF analysis may prove more efficient. The idea of adding to a real vector an imaginary part representative of its rate of change has also been applied to the POP analysis, resulting in a complex POP analysis (Bürger, 1993).

Rotated EOFs

Although EOFs are the most efficient descriptors of the variability of a data set, they may not always be those that lead to the clearest physical interpretation since different modes of variability need not be orthogonal in both space and time, as required in the EOF analysis. Thus one sometimes relaxes some of these constraints by performing a rotation that replaces the EOF patterns $\vec{p}^{\,j}$ in the expansion (2) by nicer patterns $\vec{p}^{\,j}_R$:

$$\sum_{j=1}^{K} \alpha_j \vec{p}^{\,j} = \sum_{j=1}^{K} \alpha_j^R \vec{p}^{\,j}_R \tag{34}$$

with a matrix relationship

$$(\vec{p}^{\,1}_R | \vec{p}^{\,2}_R | \cdots | \vec{p}^{\,K}_R) = (\vec{p}^{\,1} | \vec{p}^{\,2} | \cdots | \vec{p}^{\,K}) \mathcal{R} \tag{35}$$

with some $K \times K$ matrix $\mathcal{R}$. This matrix is chosen from a class of matrices (e.g., orthonormal matrices), with the constraint that the resulting patterns $\vec{p}_R^{\,j}$ maximize a certain (nonlinear) functional of simplicity F_R. Richman (1986) lists five vague criteria for patterns being simple and there are many proposals of simplicity functionals. If the matrices are orthonormal, the rotation is named *orthogonal;* otherwise, the adjective *oblique* is used. Thus, rotated EOFs have none of the properties of EOFs, as they are not the most efficient at explaining the variance, the EOF coefficients are no longer orthogonal, and if the rotation is oblique, the patterns themselves are not orthogonal.

A widely used method is the varimax, which features orthonormal matrices $\mathcal{R}$ and the simplicity measure

$$F_R(\vec{p}_R^{\,1} \cdots \vec{p}_R^{\,K}) = \sum_{j=1}^{K} f_R(\vec{p}_R^{\,j}) \tag{36}$$

with the function f_R defined for a vector $\vec{q} = (q_i)$:

$$f_R(\vec{q}) = \frac{1}{m} \sum_{i=1}^{m} \left[\left(\frac{q_i}{s_i} \right)^2 \right] - \frac{1}{m^2} \left[\sum_{i=1}^{m} \left(\frac{q_i}{s_i} \right)^2 \right]^2 \tag{37}$$

s_i is a specified number: $s_i = 1$ in the raw varimax method; $s_i^2 = \Sigma_{j=1}^{K} (\vec{p}_i^{\,j})^2$ in the normal varimax or being the variance of the ith component of $\vec{\mathbf{X}}^{(K)}$, which is the projection of the original full random vector $\vec{\mathbf{X}}$ in the signal subspace spanned by the K vectors $\{ \vec{p}^{\,1} \cdots \vec{p}^{\,K} \}$.

The minimization of a functional like (36) is in general nontrivial since the functionals are nonlinear. Numerical algorithms to approximate the solutions robustly are readily available for truncations up to $K = 80$. However, the use of large K's is usually not meaningful since the noise will normally dominate the results.

Definition (37) has the form of a variance: In the raw varimax setup it is the (spatial) variance of the squares of the components of the pattern $\vec{p}^{\,j}$, and in the normal varimax it is the same variance of a normalized version $\vec{p}^{\,\prime} = (p_i/s_i)$. Minimizing (36) implies therefore finding a set of K patterns $\vec{p}_R^{\,j}$ such that their squared patterns have (absolute or relative) maximum spatial variance. The results of a rotation exercise depends on the number K, the lengths of the vectors $\vec{p}^{\,j}$ and on the choice of the measure of simplicity.

The opinion in the community is divided on the subject of rotation. Part of the community advocates the use of rotation fervently as a means to define physically meaningful, statistically stable patterns, and indeed Cheng et al. (1995) found rotated EOFs to be statistically more stable (less sensitive to sampling fluctuations) than conventional EOFs. However, others are less convinced because of the handwaving in specifying the simplicity functions and the implications of this specification and its implication for interpretation of the result. Successful application of the rotation techniques needs some experience, and it might be a good idea for the novice to have a look into Richman's (1986) review paper on that topic. Interesting examples are offered by, among many others, Barnston and Livezey (1987) and Chelliah and Arkin (1992).

4. Modal Decomposition II: Canonical Correlation Analysis and Related Techniques

In certain problems it is useful to identify *pairs* of patterns in two fields observed simultaneously. When these pairs appear at the same time, the spatial characteristics of the two patterns may permit assessment of the dynamical link between the two fields. *Canonical correlation analysis* is one such technique that is based on optimizing the correlation between patterns (Hotelling, 1936). An alternative is to optimize the covariance, which is accomplished by SVD [therefore, the approach as a whole is named SVD, which is misleading since it blends together an algebraic solution of a problem (the SVD) with the problem itself [maximization of the covariance; (Bretherton et al., 1992)] and could be called *maximum covariance analysis* [for an oceanic application, see Frankignoul et al. (1996)].

The concept of canonical correlation analysis has been introduced by Hotelling (1936). In the following we present the idea behind the CCA [for more details, see Anderson (1984) or von Storch (1995a) and further references therein]. Two simultaneously observed fields $\vec{\mathbf{X}}(t)$ and $\vec{\mathbf{Y}}(t)$ are decomposed into K patterns:

$$\vec{\mathbf{X}}(t) = \sum_{k=1}^{K} \alpha_k^X(t)\vec{p}_X^k \quad \text{and} \quad \vec{\mathbf{Y}}(t) = \sum_{k=1}^{K} \alpha_k^Y(t)\vec{p}_Y^k \tag{38}$$

where the considered fields are *anomalies* (i.e., the time means have been subtracted prior to the analysis). The dimensions m_X and m_Y of the fields $\vec{\mathbf{X}}(t)$ and $\vec{\mathbf{Y}}(t)$ and thus of the "canonical correlation patterns" $\vec{p}_X^k$ and $\vec{p}_Y^k$ are in general different. The expansion is done in such a manner that:

1. The coefficients $\alpha_k^X(t)$ and $\alpha_k^Y(t)$ in (38) are *optimal* in a least-squares sense [i.e., for given patterns $\vec{p}_X^k$ and $\vec{p}_Y^k$ the squared differences $\Sigma_t\ (\vec{\mathbf{X}}(t) - \Sigma_{k=1}^K \alpha_k^X(t)\vec{p}_X^k)^2$ and $\Sigma_t\ (\vec{\mathbf{Y}}(t) - \Sigma_{k=1}^K \alpha_k^Y(t)\vec{p}_Y^k)^2$ are minimized]. Therefore,

$$\alpha_k^X = \langle (\vec{p}_X^k)_A, \vec{\mathbf{X}} \rangle \quad \text{and} \quad \alpha_k^Y = \langle (\vec{p}_Y^k)_A, \vec{\mathbf{Y}} \rangle \tag{39}$$

   with certain *adjoint patterns* $(\vec{p}_X^k)_A$ and $(\vec{p}_Y^k)_A$.
2. The correlations between α_k^X and α_l^X, between α_k^Y and α_l^Y, and between α_k^X and α_l^Y are zero for all $k \neq l$.
3. The correlation between α_1^X and α_1^Y is maximum.
4. The correlation between α_2^X and α_2^Y is the maximum under the constraints of items 2 and 3. The correlations for the higher indexed pairs of coefficients satisfy similar constraints (namely, of being maximum while being independent of all previously determined coefficients).

It can be shown that the adjoint patterns are the eigenvectors of somewhat complicated-looking matrices, namely:

$$\mathcal{A}_X = \Sigma_X^{-1}\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{XY}^T \quad \text{and} \quad \mathcal{A}_Y = \Sigma_Y^{-1}\Sigma_{XY}^T\Sigma_X^{-1}\Sigma_{XY} \tag{40}$$

where Σ_X and Σ_Y are the covariance matrices of $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$. Σ_{XY} is the cross-covariance matrix of $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$. The two matrices $\mathcal{A}_X$ and $\mathcal{A}_Y$ have the same nonzero eigenvalues. The kth adjoint pattern $(\vec{p}_X^{\,k})_A$ is given by the eigenvector with the kth largest eigenvalue of $\mathcal{A}_X$ and the kth adjoint pattern of $\vec{\mathbf{Y}}$ is the kth eigenvector of $\mathcal{A}_Y$. The correlation between α_k^X and α_k^Y is given by the kth largest nonzero eigenvalue of $\mathcal{A}_X$ or $\mathcal{A}_Y$.

The covariance between the canonical correlation coefficients α_k^X and the original vector $\vec{\mathbf{X}}$ is given by

$$\sum_t \alpha_k^X(t)\vec{\mathbf{X}}(t) = \sum_t \alpha_k^X(t) \sum_i \alpha_i^X(t)\vec{p}_X^{\,i} = \vec{p}_X^{\,k} \tag{41}$$

so that, because of $\alpha_k^X = \vec{\mathbf{X}}^T(\vec{p}_X^{\,k})_A$,

$$\vec{p}_X^{\,k} = \Sigma_X(\vec{p}_X^{\,k})_A \quad \text{and} \quad \vec{p}_Y^{\,k} = \Sigma_Y(\vec{p}_Y^{\,k})_A \tag{42}$$

Thus, to determine the canonical correlation patterns and the canonical correlation coefficients, one has first to calculate the covariance matrices and cross-covariance matrices. From products of these matrices (40) the adjoint patterns are derived as eigenvectors. With the adjoint patterns the canonical correlation patterns are calculated via (42) and the coefficients through (39).

For the actual computation, it should be noted that:

1. The matrix $\mathcal{A}_X$ is a $m_X \times m_X$ matrix and $\mathcal{A}_Y$ is a $m_Y \times m_Y$ matrix. The two matrices $\mathcal{A}_X$ and $\mathcal{A}_Y$ may be written as products $\mathcal{B}_1\mathcal{B}_2$ and $\mathcal{B}_2\mathcal{B}_1$ of two matrices $\mathcal{B}_1$ and $\mathcal{B}_2$. Therefore, the two matrices share the same nonzero eigenvalues, and if $\vec{p}_X^{\,k}$ is an eigenvector of $\mathcal{A}_X$ with an eigenvalue $\lambda \neq 0$, then $\Sigma_Y^{-1}\Sigma_{XY}^T\vec{p}_X^{\,k}$ is an eigenvector of $\mathcal{A}_Y$ with the same eigenvalue. Because of the specific form of the matrices, it is advisable to solve the eigenvector problem for the smaller matrix.

2. Numerical experiments have shown that it is highly advisable to compress the data prior to a CCA (Barnett and Preisendorfer, 1987; Bretherton et al., 1992). Convenient tools for that purpose are conventional EOFs. The rationale for this need is the following: When executing a CCA we are looking for true pairs of patterns that reflect the real underlying dynamical structure of the problem under consideration. In practice, however, only a limited sample is available and we have to guard ourselves against the danger of misinterpreting the random details of the common variability within the sample as indications of true correlations. In particular, when the dimension of the random vectors $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$ is large and the number of samples is small, it is likely that in the many badly sampled noise contributions spuriously high sample correlations appear. Then the CCA emphasizes these sample correlations.

An important caveat to keep in mind is the method's intrinsic tendency to return overestimated correlation coefficients from a finite sample of observed fields (Glynn and Muirhead, 1978). Also, the results may depend on the a priori EOF truncation of the data, so their sensitivity should be investigated. Note that maximum covariance analysis does not require a priori data compression.

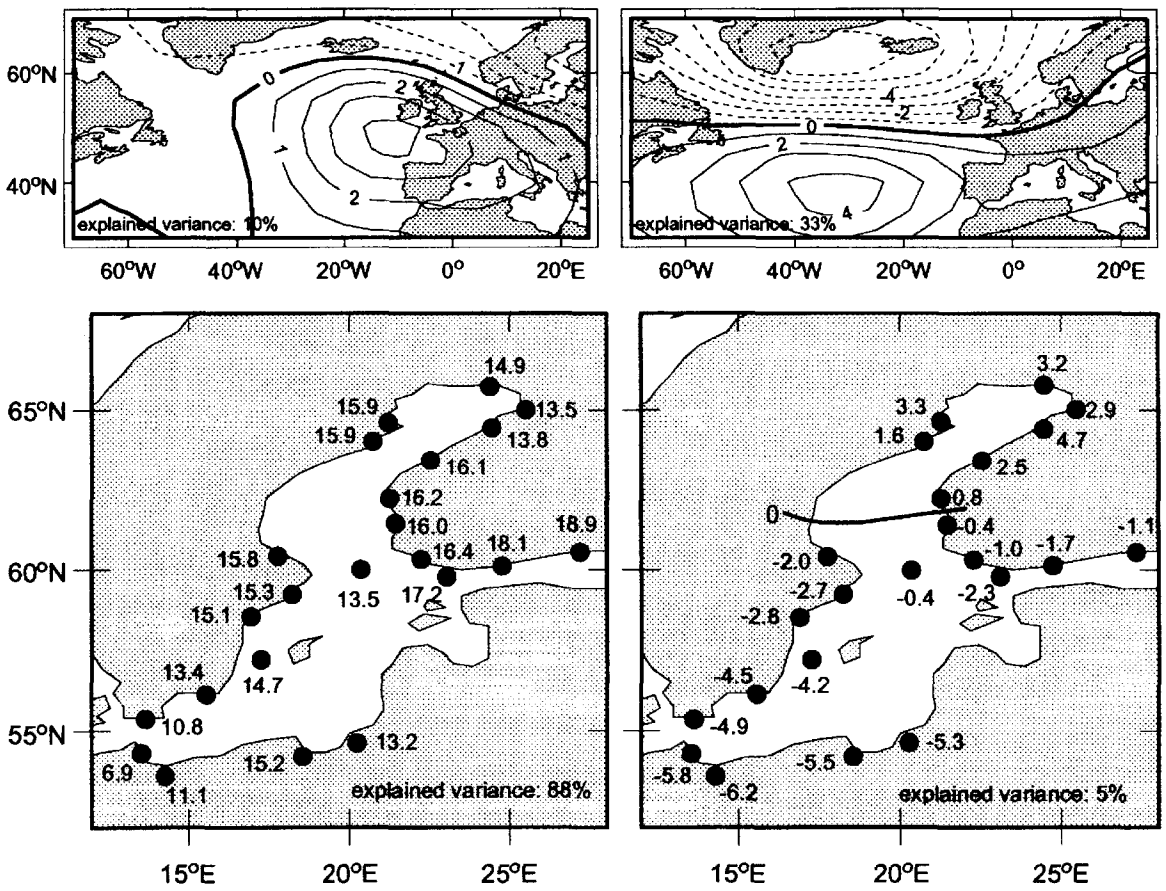

Fig. 16.11. First two pairs of CCA linking the winter mean air-pressure field over the North Atlantic with the winter mean sea level at various gauges in the Baltic Sea. The left pair of patterns is linked with a correlation coefficient of 0.80, and the right pair with a coefficient of 0.52. (From Heyen et al., 1996.)

Heyen et al. (1996) have examined the simultaneous variability of winter (DJF) air pressure in the North Atlantic/European region and an array of local sea level measurements in the Baltic Sea, using CCA. The vector $\vec{\mathbf{X}}(t)$ is formed from the gridded ($5° \times 5°$ lattitude $\times$ longitude) air pressure field in the North Atlantic, whereas the other vector $\vec{\mathbf{Y}}(t)$ is composed of the sea level at 23 stations along the coast of the Baltic Sea (for the area and locations, see Fig. 16.11). From all time series the long-term linear trend has been subtracted since the isostatic rebound of Scandinavia causes a long-term trend that is unrelated to climate variability. Also, an EOF truncation was done.

The two pairs of patterns with the largest correlations between the coefficients are shown in Fig. 16.11. The patterns are normalized such that the coefficients have unit standard deviation, so the patterns represent typical anomalies in hPa and cm, respectively.

The first sea level pattern (Fig. 16.11, left) represents 88% of the total winter-to-winter variability at the gauges; it describes an overall rise, or fall, of sea level everywhere in the Baltic. Typically, anomalies are of the order of 10–15 cm, with slightly larges values in the north and somewhat smaller anomalies in the southwest. The canonical correlation analysis matches this overall rise or fall of the Baltic with a characteristic pattern of winter-mean circulation with an anomalous high-pressure

center off the Bay of Biscay and a low-pressure center located over northern Finland. The time coefficients for the two patterns have a correlation coefficient of 0.80. The two patterns represent a physically plausible forcing/response relation. Anomalous westerly winds in the sea north of Denmark decelerate the outflow of waters from the Baltic, with the effect that the Baltic Sea holds more water.

The second sea level pattern (Fig. 16.11, right) represents 5% of the variance and is almost orthogonal to the first pattern. It is characterized by a zero line in the southern Bottenwyk. Maximum anomalies in the northern and southern parts are on the order of 5 cm. This pattern is connected, with a correlation coefficient of 0.52, to an air-pressure pattern with two centers of action in the central North Atlantic. These two centers of action represent an important climatological mode of atmospheric variability, the North Atlantic Oscillation. They are connected with anomalous southwesterlywind along the major axis of the Baltic Sea.

An application of a paired pattern analysis such as CCA is the downscaling problem in climate (change) research (von Storch et al., 1993). Most significant questions asked about the impact of expected climate changes concern changes of the abiotic environment, and their effects on the biosphere and society, on a regional or local scale. The primary tools for describing details of the expected climate change are *general circulation models* (GCMs) of the ocean and the atmosphere. Such models are powerful in reproducing the large-scale features of the ocean and the atmosphere, but they are inadequate to simulate facets with spatial scales at the lower end of the spatial resolution (Hewitson and Crane, 1992). For present-day models this means that regional and local aspects are not realistically reproduced. Therefore, it is sometimes required to build empirical models for relating large-scale features, which have long been observed and are reproduced reliably by GCMs, to the small-scale features that are of relevance for the assessment of climate change. In that case it is convenient to use the observations to relate the large-scale forcing field to the regional or local scale response. This is conveniently done by using CCA to define a few patterns, whose coefficients are pairwise highly correlated.

The two patterns in Fig. 16.11 are obviously capable of representing the bulk of winter-to-winter variations of (detrended) sea level variations in the Baltic. Therefore, the link between the large-scale air-pressure fields in Fig. 16.11 and the two sea level patterns may be used to build a linear regression model to downscale the large-scale information, encoded in the air-pressure field, to the local sea level information. The success of this downscaling model is displayed in Fig. 16.12 by the percentage of explained variance at all stations (upper left), the reconstruction of the (detrended) sea level averaged over all stations and for two individual stations (Ystad and Helsinki). By processing exclusively information about the state of the air-pressure field over the North Atlantic (dashed line), the empirical model reproduces between 40 and 75% of the winter-to-winter variance of the observed detrended coastal sea level (solid line). However, the extremes are not well reproduced, as might be expected from a regression technique.

5. Summary

Apart from the few cautionary remarks on statistical testing given in the introduction, this chapter has focused on various methods that can be used to summarize or detect the dominant space/time patterns in the complex data sets that are often collected

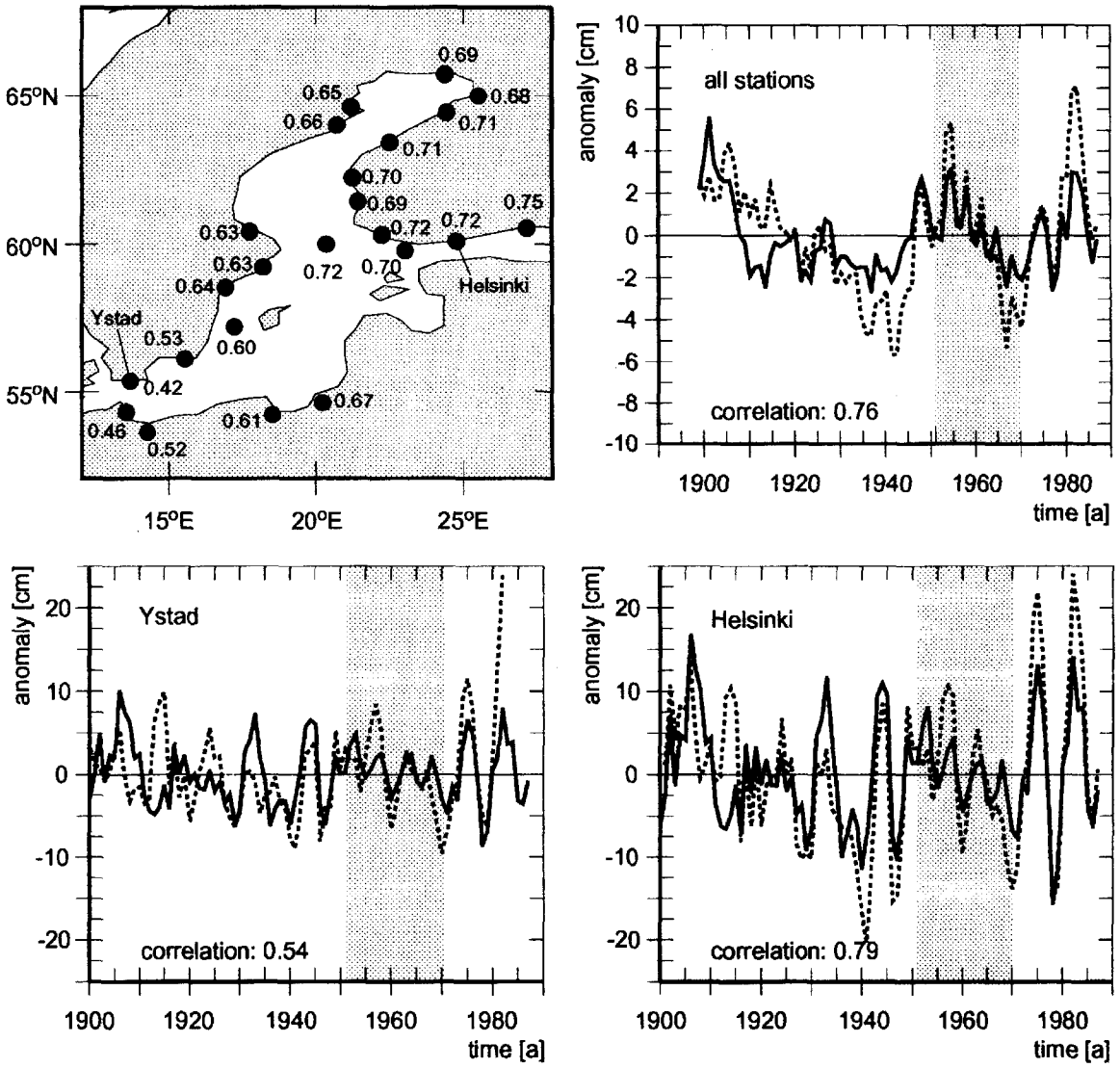

Fig. 16.12. Winter mean sea level variations in the Baltic Sea: observations versus reconstructions. The statistical downscaling model, which relates the air-pressure field in the North Atlantic with the local sea level, has been built with data from 1951 to 1979 (see stippling), so that all data after 1970 and prior to 1950 represent independent evidence. Top left: percentage of winter-to-winter variance explained by the empirical model. Top right: reconstructed (solid) and observed (dashed) detrended sea level variations averaged over all stations. A 3-year running mean filter has been applied. Bottom two panels: as the upper right panel, but for the two gauges Ystad and Helsinki. (From Heyen et al., 1996.)

for coastal studies. When the dynamics are simple and normal modes can be derived from the equations of motions, the most informative technique will rely on analytical modes, but the conditions are often too complex, or not well enough documented, to derive them without unacceptable assumptions. In such cases, one can estimate empirical modes using POP analysis, which is a powerful tool to identify oscillatory behaviors. However, the POPs require that the mode dynamics can be approximately represented by a multivariate first-order autoregressive process, so POPs are not well adapted to more complex dynamics, as when nonlinearities are important. Also, the conventional POPs can only describe propagating oscillations, not standing ones, but the latter may be detectable in the conventional EOF analysis which is generally performed prior to calculating the POPs.

Alternative techniques that are useful in detecting propagating patterns and oscillatory behaviors have also been discussed. Frequency-dependent EOFs can be used either by doing an EOF analysis in the frequency domain or by using the Hilbert transform, with the same advantages and, possibly, drawbacks as the standard EOFs (maximum explained variance, orthogonality). Contrarily to POPs, the frequency-dependent EOFs impose no dynamical constraint of the time dependence, but then the latter needs to be estimated a posteriori from the EOF coefficients. Although the two methods provide very comparable results in many cases, the POPs are more efficient in detecting oscillations in noisy environments and are easier to interpret if the conditions for their applicability are reasonably satisfied.

Another powerful tool to identify oscillatory modes is provided by EEOF analysis, which searches for the dominant space/time patterns without any constraint on the mode behaviors and is thus well adapted to complex dynamics. EEOF pairs are very efficient adaptive narrowband filters, but their interpretation is sometimes difficult and there is a risk that spurious pairs appear in limited-length records. Thus, as in spectral analysis, the most powerful methods are the most difficult to use wisely, and it is recommended that several different approaches be attempted when possible. More important, the interpretation should be validated by physical reasoning, and due caution used.

Most of these methods require some prior data compression, which is usually done by an EOF analysis where only the main patterns are considered, even though the selection criteria are not always obvious, as discussed briefly. Some users favor using rotated EOFs, which tend to give more localized and sometimes more robust patterns, but caution is needed in performing the rotation.

We have also discussed techniques that were designed to identify pairs of patterns appearing in two or more fields observed simultaneously. CCA is a powerful and commonly used technique that optimizes the correlations between fields, but it generally requires, and is sensitive to, an initial EOF truncation of the data. A variant that is easier to use, as no prior truncation is needed, is SVD or maximum covariance analysis, although only two fields can be considered simultaneously. Pairs of patterns can also be identified by considering the combined fields directly: EOF, POPs or EEOF analysis will then also extract the coupled patterns, with various degrees of efficiency. An interesting comparison between some of these techniques is given in Bretherton et al. (1992), but the most useful one depends on the problem at hand, and no general recipe should be offered.

Acknowledgments

Useful discussions with Myles Allen, Ken Brink, Hauke Heyen, Viacheslav Kharin and Eduardo Zorita are gratefully acknowledged.

Bibliography

Achatz, U., G. Schmitz and K.-M. Greisiger, 1995. Principal interaction patterns in baroclinic wave life cycles. *J. Atmos. Sci.*, **52,** 3201–3213.

Allen, M. R. and L. A. Smith, 1994. Investigating the origins and significance of low-frequency modes of climate variability. *Geophys. Res. Lett.*, **21,** 883–886.

Allen, M. R. and L. A. Smith, 1996. Monte Carlo SSA: detecting irregular oscillations in the presence of coloured noise. *J. Climate.*, **9,** 3373–3404.

Anderson, T. W., 1984. *An Introduction to Multivariate Statistical Analysis.* Wiley, New York, 374 pp.

Barnett, T. P., 1983. Interaction of the monsoon and Pacific trade wind system at interannual time scale. Part I. *Mon. Weather Rev.*, **111,** 756–773.

Barnett, T. P. and R. Preisendorfer, 1987. Origins and levels of monthly and seasonal forecast skill for United States surface air temperature determined by canonical correlation analysis. *Mon. Weather Rev.*, **115,** 1825–1850.

Barnston, A. G. and R. E. Livezey, 1987. Classification, seasonality and persistence of low frequency circulation patterns. *Mon. Weather Rev.*, **115,** 1083–1126.

Bretherton, C. S., C. Smith and J. M. Wallace, 1992. An intercomparison of methods for finding coupled patterns in climate data. *J. Climate,* **5,** 541–560.

Brillinger. D. R., 1975. *Time Series: Data Analysis and Theory.* Holt, Rinehart and Winston, Austin, Tex., 500 pp.

Brink, K. H., 1991. Coastal trapped waves and wind-driven currents over the continental shelf. *Ann. Rev. Fluid Mech.*, **23,** 389–412.

Brink, K. H. and R. D. Muench, 1986. Circulation in the Point Conception–Santa Barbara Channel region. *J. Geophys. Res.*, **91C,** 877–895.

Broomhead, D. S. and G. P. King, 1986. Extracting qualitative dynamics from experimental data. *Physica D,* **20,** 217–236.

Bürger, G., 1993. Complex principal oscillation patterns. *J. Climate,* **6,** 1972–1986.

Chelliah, M. and P. Arkin, 1992. Large-scale interannual variability of monthly outgoing longwave radiation anomalies over global tropics. *J. Climate,* **5,** 371–389.

Cheng, X., G. Nitsche and J. M. Wallace, 1995. Robustness of low-frequency circulation patterns derived from EOF and rotated EOF analysis. *J. Climate,* **8,** 1709–1713.

Frankignoul, C., 1995. Statistical analysis of GCM output. In *Analysis of Climate Variability: Applications of Statistical Techniques,* H. von Storch and A. Navarra, eds. Springer-Verlag, New York, pp. 139–158.

Frankignoul, C., C. Duchêne and M. Cane, 1989. A statistical approach to testing equatorial ocean models with observed data. *J. Phys. Oceanogr.*, **19,** 1191–1208.

Frankignoul, C., F. Bonjean and G. Reverdin, 1996. Interannual variability of surface currents in the tropical Pacific during 1987–1993. *J. Geophys. Res.*, **101,** 3629–3647.

Glynn, W. J. and R. J. Muirhead, 1978. Inference in canonical correlation analysis. *J. Multivar. Anal.*, **8,** 468–478.

Hasselmann, K., 1979. On the signal-to-noise problem in atmospheric response studies. In *Meteorology over the Tropical Oceans,* B. D. Shaw, ed. Royal Meteorological Society, Bracknell, Berkshire, England, pp. 251–259.

Hasselmann, K., 1988. PIPs and POPs: the reduction of complex dynamical systems using principal interaction and oscillation patterns. *J. Geophys. Res.*, **93,** 11015–11021.

Hasselmann, K., 1993. Optimal fingerprints for the detection of time dependent climate change. *J. Climate,* **6,** 1957–1971.

Hayes, S. P., L. J. Mangnum, J. Picaut, A. Sumi and K. Takeuchi, 1991. TOGA-TAO: a moored array for real time measurements in the tropical Pacific Ocean. *Bull. Am. Meteorol. Soc.*, **72,** 339–347.

Hewitson, B. C. and R. G. Crane, 1992. Regional-scale atmospheric controls on local precipitation in tropical Mexico. *Geophys. Res. Lett.*, **19,** 1835–1838.

Heyen, H., E. Zorita and H. von Storch, 1996. Statistical downscaling of winter monthly mean North Atlantic sea-level pressure to sea-level variations in the Baltic Sea. *Tellus,* **48A,** 312–323.

Hotelling, H., 1935. The most predictable criterion. *J. Ed. Psychol.*, **26,** 139–142.

Hotelling, H., 1936. Relations between two sets of variants. *Biometrika,* **28,** 321–377.

Johnson, E. S. and M. J. McPhaden, 1993. On the structure of intraseasonal Kelvin waves in the equatorial Pacific Ocean. *J. Phys. Oceanogr.*, **23,** 608–625.

Jolliffe, I. T., 1986. *Principal Component Analysis.* Springer-Verlag, New York, 271 pp.

Jones, R. H., 1976. On estimating the variance of time averages. *J. Appl. Meteorol.*, **15,** 514–515.

Kelly, K. A., 1985. The influence of winds and topography on the sea surface temperature patterns over the northern California slope. *J. Geophys. Res.*, **90C,** 11783–11798.

Kelly, K. A., 1988. Comment on "Empirical orthogonal function analysis of advanced very high resolution radiometer surface temperature patterns in Santa Barbara Channel" by G. S. E. Lagerloef and R. L. Bernstein. *J. Geophys. Res.*, **93C,** 15753–15754.

Kundu, P. K. and J. S. Allen, 1976. Some three-dimensional characteristics of low frequency current fluctuations near the Oregon coast. *J. Phys. Oceanogr.*, **6,** 181–199.

Kundu, P. K., J. S. Allen and R. L. Smith, 1975. Modal decomposition of the velocity field near the Oregon coast. *J. Phys. Oceanogr.*, **5,** 683–704.

Kwasniok, F., 1966. The reduction of complex dynamical systems using principal interaction patterns. *Physica D,* **92,** 28–60.

Livezey, R. E., 1995. The evaluation of forecasts. In *Analysis of Climate Variability; Applications of Statistical Techniques,* H. von Storch and A. Navarra, eds. Springer-Verlag, New York, pp. 177–196. (ISBN 3-540-58918-X).

Livezey, R. E. and W. Y. Chen, 1983. Statistical field significance and its determination by Monte Carlo techniques. *Mon. Weather Rev.*, **111,** 46–59.

Lorenz, E. N., 1956. Empirical orthogonal functions and statistical weather prediction. *Statistical Forecast Project Report 1.* Department of Meteorology, MIT, Cambridge, Mass., 49 pp.

Merrifield, M. A. and R. T. Guza, 1990. Detecting propagating signals with complex empirical-orthogonal functions: a cautionary note. *J. Phys. Oceanogr.*, **20,** 1628–1633.

North, G. R., T. L. Bell, R. F. Cahalan and F. J. Moeng, 1982. Sampling errors in the estimation of empirical orthogonal functions. *Mon. Wea. Rev.*, **110,** 699–706.

Pearson, K., 1902. On lines and planes of closest fit to systems of points in space. *Philos. Mag.*, **2,** 559–572.

Plaut, G. and R. Vautard, 1994. Spells of low-frequency oscillations and weather regimes over the Northern Hemisphere. *J. Atmos. Sci.*, **51,** 210–236.

Preisendorfer, R. W., 1988. *Principal Component Analysis in Meteorology and Oceanography.* Elsevier, Amsterdam, 426 pp.

Preisendorfer, R. W., F. W. Zwiers and T. P. Barnett, 1981. Foundation of principal component selection rules. *SIO References Series 81-4.* Scripps Institution of Oceanography, La Jolla, Calif., 192 pp.

Rasmusson, E. M., P. A. Arkin, W.-Y. Chen and J. B. Jalickee, 1981. Biennial variations in surface temperature over the United States as revealed by singular decomposition. *Mon. Weather Rev.*, **109,** 587–598.

Richman, M. B., 1986. Rotation of principal components. *Int. J. Climatol.*, **6,** 293–335.

Robertson, A. W., 1966. Interdecadal variability over the North Pacific in a coupled ocean-atmosphere general circulation model. *Climate Dyn.*, **12,** 227–241.

Schnur, R., G. Schmitz, N. Grieger and H. von Storch, 1993. Normal modes of the atmosphere as estimated by principal oscillation patterns and derived from quasigeostrophic theory. *J. Atmos. Sci.*, **50,** 2386–2400.

Seber, G. A. F., 1984. *Multivariate Observations.* Wiley, New York, 686 pp.

Selten, F. M., 1995. An efficient description of the dynamics of barotropic flow. *J. Atmos. Sci.*, **52,** 915–936.

Thiébeaux, H. J. and F. W. Zwiers, 1984. The interpretation and estimation of effective sample size. *J. Climate Appl. Meteorol.*, **23,** 800–811.

Vautard, R., 1995. Patterns in time: SSA and MSSA. In *Analysis of Climate Variability: Applications of Statistical Techniques,* H. von Storch and A. Navarra, eds. Springer-Verlag, New York, pp. 259–280.

Vautard, R., and M. Ghil, 1989. Singular spectrum analysis in nonlinear dynamics with applications to paleoclimatic time series. *Physica D,* **35,** 395–424.

von Storch, H., 1982. A remark on Chervin–Schneider's algorithm to test significance. *J. Atmos. Sci.*, **39,** 187–189.

von Storch, H., 1993. Principal oscillation pattern analysis of the intraseasonal variability in the equatorial

Pacific Ocean. *Proc. 'Aha Huliko'a Workshop 1993.* SOEST Special Publication. University of Hawaii, Manoa, Hawaii, pp. 201–227.

von Storch, H., 1995a. Spatial patterns: EOFs and CCA. In *Analysis of Climate Variability: Applications of Statistical Techniques,* H. von Storch and A. Navarra, eds. Springer-Verlag, New York, pp. 227–258.

von Storch, J., 1995b. Multivariate statistical modelling: POP model as a first order approximation. In *Analysis of Climate Variability: Applications of Statistical Techniques,* H. von Storch and A. Navarra, Springer-Verlag, New York, pp. 281–297.

von Storch, H. and G. Hannoschöck, 1984. Comment on "Empirical orthogonal function analysis of wind vectors over the tropical Pacific region," *Bull. Am. Meteorol. Soc.,* **65,** 162.

von Storch, H. and G. Hannoschöck, 1986. Statistical aspects of estimated principal vectors (EOFs) based on small sample sizes. *J. Climatol. Appl. Meteorol.,* **24,** 716–724.

von Storch, H. and F. W. Zwiers, in press. *Statistical Analysis in Climate Research.* Cambridge University Press, Cambridge.

von Storch, H., T. Bruns, I. Fischer-Bruns and K. Hasselmann, 1988. Principal oscillation pattern analysis of the 30- to 60-day oscillation in a general circulation model equatorial troposphere. *J. Geophys. Res.,* **93,** 11022–11036.

von Storch, H., E. Zorita and U. Cubasch, 1993. Downscaling of climate change estimates to regional scales: application to winter rainfall in the Iberian Peninsula. *J. Climate,* **6,** 1161–1171.

von Storch, H., G. Bürger, R. Schnur and J. von Storch, 1995. Principal oscillation patterns. *J. Climate,* **8.**

Wallace, J. M. and R. E. Dickinson, 1972. Empirical orthogonal representation of time series in the frequency domain. I. Theoretical considerations. *J. Appl. Meteorol.,* **11,** 887–892.

Wang, D. P. and C. N. K. Mooers, 1977. Long coastal-trapped waves off the west coast of the United States, Summer 1973. *J. Phys. Oceanogr.,* **7,** 856–864.

Weare, B. C. and J. N. Nasstrom, 1982. Examples of extended empirical orthogonal function analyses, *Mon. Weather Rev.,* **110,** 481–485.

Zorita, E. and C. Frankignoul, 1997. Modes of North Atlantic decadal variability in the ECHAM1/LSG coupled ocean–atmosphere general circulation model. *J. Climate,* **10,** 183–200.

Zwiers, F. W. and H. von Storch, 1995. Taking serial correlation into account in tests of the mean. *J. Climate,* **8,** 336–351.