Q. J. R. Meteorol. Soc. (2000), **126**, pp. 1387–1429

On the mapping of multivariate geophysical fields: Error and variability subspace estimates

By P. F. J. LERMUSIAUX*, D. G. M. ANDERSON and C. J. LOZANO
Harvard University, USA

(Received 29 June 1998; revised 16 September 1999)

SUMMARY

A basis is outlined for the first-guess spatial mapping of three-dimensional multivariate and multiscale geophysical fields and their dominant errors. The a priori error statistics are characterized by covariance matrices and the mapping obtained by solving a minimum-error-variance estimation problem. The size of the problem is reduced efficiently by focusing on the error subspace, here the dominant eigendecomposition of the a priori error covariance. The first estimate of this a priori error subspace is constructed in two parts. For the 'observed' portions of the subspace, the covariance of the a priori missing variability is directly specified and eigendecomposed. For the 'non-observed' portions, an ensemble of adjustment dynamical integrations is utilized, building the non-observed covariances in statistical accord with the observed ones. This error subspace construction is exemplified and studied in a Middle Atlantic Bight simulation and in the eastern Mediterranean. Its use allows an accurate, global, multiscale and multivariate, three-dimensional analysis of primitive-equation fields and their errors, in real time. The a posteriori error covariance is computed and indicates complex data-variability influences. The error and variability subspaces obtained can also confirm or reveal the features of dominant variability, such as the Ierapetra Eddy in the Levantine basin.

KEYWORDS: Data and dynamics nonlinear balance Ensemble of adjustment dynamical integrations Error subspace Multiscale Multivariate

1. INTRODUCTION

Even though the computation of gridded fields for realistic meteorology and oceanography has been considered in a probabilistic sense for some time (e.g. Gandin 1965), carrying out the probabilistic approach is complex for several reasons. Ideally, based on estimation theory (Jazwinski 1970), the statistics of three-dimensional (3D) and multivariate fields are determined at a given time, taking into account all available knowledge of the geophysical system studied, including estimates of imperfections. The main sources of knowledge consist of data and dynamics. The main sources of uncertainties include the difficulty of comprehensive and accurate data acquisition, the challenges of modelling complex, multiscale and multidisciplinary geophysical phenomena (Robinson 1996; Glimm and Sharp 1997) and the sensitivity to initial conditions (Lorenz 1965; Houghton 1991; Molteni and Palmer 1993). In the light of these challenges, an issue addressed here is the rational truncation of the probability density of realistic fields and its subsequent estimation.

For today's operational methods, as for this study, a primary goal is to obtain a gridded estimate of the real state and its error: in other words, its probability of being right. Common techniques include statistical interpolation or objective analysis (Bretherton *et al.* 1976; McWilliams *et al.* 1986; Lorenc 1981, 1986; Daley 1991; Wunsch 1996), successive corrections (Cressman 1959; Barnes 1964, 1994; Lorenc 1992), functional fitting (Wahba and Wendelberger 1980; McIntosh 1990; Brankart and Brasseur 1996), multi-resolution interpolation and spatial filtering (Fieguth *et al.* 1996), feature and structured data models (Lozano *et al.* 1996; Gangopadhyay *et al.* 1997), and time filtering or smoothing data assimilation schemes (Robinson *et al.* 1998a). For an overview of the three-dimensional variational (3D-Var) scheme of the European Centre for Medium-Range Weather Forecasts, presenting the specifics of different

* Corresponding author: Harvard University, DEAS, Pierce Hall G2A, 29 Oxford Street, Cambridge, MA 02318, USA.

developments in several meteorological centres (e.g. Parrish and Derber 1992; Gauthier *et al.* 1996), we refer to (Courtier *et al.* 1998; Rabier *et al.* 1998; Andersson *et al.* 1998). The majority of mapping methodologies utilize a priori weights or covariance fields to interpolate and extrapolate the data at hand into field estimates. These weights are commonly based on empirical or modelled variability properties. Even though the multiscale geophysical variability is observed to be 3D, non-homogeneous, anisotropic and multivariate, several schemes simplify these characteristics, for example to 2D or univariate covariances. The mapping is then reduced to a succession of uncorrelated estimation problems.

Within the Bayesian framework (e.g. Jazwinski 1970; Lorenc 1986), accounting for the complete a priori probability or error statistics is today too expensive, mainly because of the cost of modelling and observing systems. Truncating the error statistics is thus necessary. Nonetheless, the truncation should be rational, reflecting the dominant properties of data and dynamics at the time of the mapping, as measured by the estimation criterion. The term 'dominant' should in fact be defined in accord with the error measure of this criterion (e.g. dominant in the sense of a convex metric for a convex error measure). The neglected statistics are then insignificant for the estimation at hand, and the retained dominant statistics determine the error subspace.

Using these ideas, a basis for a first estimate of multivariate geophysical fields and their dominant uncertainty is outlined. The time for the estimation is fixed; for time evolutions we refer to Lermusiaux (1997; henceforth, LER97). The statistics of the dominant a priori error are assumed to be variability from the a priori state. The mapping therefore focuses on the significant subspace of the a priori missing variability, referred to hereafter as the error subspace. The possible 3D, multiscale or multivariate properties are not removed *ab initio*; as will be shown, their dominant components powerfully interpolate the sparse data, simultaneously across all variables and scales that matter. With the premise that all knowledge is useful, the error subspace is constructed combining a priori data and dynamics, in accord with their respective deficiencies. For the variables, regions and regimes with good historical (synoptic) data, the a priori missing variability is specified, either directly from these data or via an analytical model fit to this data. These portions of the error subspace are referred to as the 'observed' portions. The most advanced nonlinear dynamical model available, possibly including a stochastic component, is then used to build the 'non-observed' portions in accord with the observed ones, via an ensemble of adjustment dynamical integrations.

The remainder of the paper is organized as follows. Section 2 introduces the choices, issues and objectives specific to this study. Section 3 outlines the specification of the observed portions of the error subspace, focusing on a univariate formulation for convenience. Section 4 constructs the complete multivariate error subspace. The methodology is exemplified in an oceanographic context. The examples include the estimation of the mesoscale variability subspace of a realistic Levantine Sea experiment and of an idealized, primitive-equation (PE) simulation of the Middle Atlantic Bight 'shelfbreak' front. Section 5 presents a global multivariate 3D analysis of the state and uncertainty of PE fields in the Levantine Sea. The results are briefly compared with those of the operational, univariate 2D objective analysis scheme of the Harvard Ocean Prediction System (HOPS, e.g. Lozano *et al.* 1996). Section 6 contains a summary and the conclusions. Wherever possible, the notation of Ide *et al.* (1997) is used. Appendix A describes specifics of the notation and outlines the machinery of the mapping. Appendix B outlines the algorithm employed for computing the vertical multivariate empirical orthogonal functions (EOFs) of a priori tracer residuals, which are encountered in sections 3 to 5.

The emphasis of the present paper is mainly methodological. We refer to a companion paper (Lermusiaux *et al.* 1998; henceforth, LLA98) for more on the fundamental and applied possibilities raised by the global and multivariate dominant variability estimates briefly explored in sections 4 to 6. The sensitivity of the error subspace estimate to the data and dynamical model parameters, the sensitivity of the state estimate to the properties of the error subspace and further intercomparisons with the univariate 2D scheme are also presented in LLA98.

2. SPECIFIC CHOICES, ISSUES AND OBJECTIVES

The approach introduced in section 1 is particularized by choosing the conditional mean for the optimal state and by estimating its gridded values using a minimum-error-variance criterion. The a priori error statistics are also assumed to be described by error covariance matrices, denoted here by **B**. For consistency with the variance measure, a subspace is then determined by the dominant eigendecomposition of such a covariance*. The corresponding estimation criterion and scheme are given in appendix A.

The main objective is thus to construct algorithms for a first guess† at the dominant eigendecomposition of error covariances (sections 3 and 4). Most 3D, multivariate covariance matrices are large, and the knowledge of geophysical covariance functions remains limited. The data available are usually not sufficient for their complete estimation (e.g. Dee 1995; Kaplan *et al.* 1997). This is especially true for biological or chemical fields (Flierl and Davis, private communication). Covariance functions are therefore specified using simple statistical or dynamical models' fit to data (e.g. Thiebaux 1976; McWilliams *et al.* 1986; Daley 1991). Determining the dominant error balances is also not immediately obvious; but, in analogy with the appreciation of dominant dynamical balances, it is important for deeper understanding. The specification of background errors in meteorology is usually based on the studies of Balgovind *et al.* (1983); Hollingsworth and Lönnberg (1986); Lönnberg and Hollingsworth (1986); Phillips (1986); Parrish and Derber (1992); Bartello and Mitchell (1992); Daley (1991, 1992b, 1996). In oceanography some of these studies are useful, but for several reasons, results are often based on simplified models and progress is more limited (e.g. Jiang and Ghil 1993; Tsaoussi and Koblinsky 1994). For the observed portions of error covariances, the primary goal here is to illustrate the efficacy of the subspace approach based on known specifications, increasing complexity up to non-homogeneous and multiscale subspace computations. For the non-observed portions, the goal is to validate the use of an ensemble of adjustment dynamical integrations. The combination of both portions into complete dominant error eigendecompositions yields progress toward more advanced covariance formulations.

Another objective is to illustrate these algorithms and to evaluate their performance. The main evaluation is carried out by examining the results of mapping applications (section 5 and LLA98). The error-subspace approach (i) allows a *global* analysis; (ii) estimates all fields at once, based on the 3D multivariate correlations that matter; (iii) directly specifies the dominant field (**B**) and data (**R**) error covariances which, as in 3D-Var schemes, facilitates the use of nonlinear-measurement models frequently found with acoustic or biological data (e.g. Munk *et al.* 1995; Evans and Fasham 1993); and

* In this paper, the term covariance relates to a matrix quantity unless otherwise mentioned. Of course, for the eigenvectors of a covariance to possibly be physically meaningful, a normalization and adequate grid (appendix A and section 4(b)) are necessary prior to the eigendecomposition (Lorenc 1992; LER97).

† First-guess error subspaces can be improved: for example, subsequent filtering or smoothing (e.g. adjoint) schemes would yield errors balanced in time and space, in accord with data and dynamics. The resulting error estimates can then be used as historical data in future error subspace computations.

(iv) permits the use of all data available via sequential data processing (Parrish and Cohn 1985). A specific goal here is to exemplify these properties (i)–(iv) of efficiently reducing **B**. Each of them can be contrasted with simplifications made in many ocean mappings which consist of (i) limiting the domain of influence of data, both in space and time, leading to a succession of so-called *local* analyses; (ii) estimating parts of the fields one at a time and zeroing certain correlations a priori; (iii) only specifying the data-to-data points ($\mathbf{HBH}^{\mathrm{T}}$) and field-to-data points ($\mathbf{BH}^{\mathrm{T}}$) covariances (appendix A), which is well-suited to linear-measurement models; and (iv) subsampling the observations in 'data-rich' regions to reduce the size of $\mathbf{HBH}^{\mathrm{T}}$ and thus the cost of the inversion. For examples of such simplifications, we refer to Carter and Robinson (1987), Watts *et al.* (1989), Carton and Hackert (1989), Mariano and Brown (1992), and Lozano *et al.* (1996). For additional evaluation, the significant components of **B** are compared to what is known about the region and regime considered (section 4(c)). The error subspace can in fact be used for validating its construction; if adequate, it should explain any significant sample of the a priori missing variability. Such global covariance evaluations complement the local 'single-observation analyses' utilized by Parrish and Derber (1992) and Thépaut *et al.* (1996).

Since field mapping is often used to initialize a predictive model, the last specific objective is a method for initialization. In real-time oceanography today, an accurate initial error can in fact be as valuable as an error forecast. The present error-subspace construction has been critical in initializing data-driven predictions and smoothings for sea exercises in North Atlantic Treaty Organization (NATO) operations. Such real-time experiments occurred for example in the Strait of Sicily, Ionian Sea and Gulf of Cadiz (Robinson *et al.* 1998b, 1999). Aspects of this first-guess field and error initialization are also discussed in LLA98.

As an aside, while the foregoing three objectives were the forces driving the present research, it is the impact of the field and error mapping scheme in an operational oceanographic context that created the need to consider the mapping problem in detail.

3. Error subspace: observed portions

The portions of the error subspace for which sufficient data are available are specified directly from data or via an analytical model fit to data. The presentation is ordered according to complexity: the possible separability (e.g. horizontal/vertical) or scale independence (e.g. mesoscale/large-scale) is used before addressing correlated multiscale issues. Section 3(a) deals with horizontal error covariances and briefly exemplifies advantages of their truncation. Section 3(b) considers vertical error covariances. Section 3(c) extends these results to the estimation of the dominant decomposition of 3D and multiscale error covariances without storing these matrices.

(*a*) *Horizontal error subspace*

For the horizontal error covariances that are observed, we employ statistical models fitted to data. For practical reasons, covariance functions (denoted here by C) are usually simplified (e.g. Lorenc 1986): for example, they are often chosen to be stationary, isotropic, homogeneous or partitioned assuming scale separations. Successive-correction techniques (Daley 1991; Lorenc 1992) provide approximate iterative methods, usually mapping one scale at a time. The horizontal analysis of HOPS makes such a scale-separation assumption: its two-scale version is employed here as a benchmark. The non-dimensional (*) covariance function used is a scalable, negated, second

derivative of a Gaussian or 'Mexican hat'. The parameters are the zonal l_a^x and meridional l_a^y zero-crossings, the zonal l_b^x and meridional l_b^y e-folding decay-scales and the decorrelation time τ. For each scale these parameters are fitted to scale-filtered data residuals; and the elements of $\mathbf{B}$, after non-dimensionalization (appendix A), are of the form $C^*(\mathbf{r}_1, \mathbf{r}_2) = (1 - a^2)\exp(-b^2/2)$, where $a^2 = (\mathbf{r}_1 - \mathbf{r}_2)^{\mathrm{T}}\mathbf{L}_a^{-2}(\mathbf{r}_1 - \mathbf{r}_2)$ and $b^2 = (\mathbf{r}_1 - \mathbf{r}_2)^{\mathrm{T}}\mathbf{L}_b^{-2}(\mathbf{r}_1 - \mathbf{r}_2) + (\Delta t_i/\tau)^2$ are scalars, and $\mathbf{r}_1$ and $\mathbf{r}_2$ horizontal position vectors, with $\mathbf{r} = (x, y)$. The [2 × 2] anisotropic scaling matrices contain the local zero-crossings, $\mathbf{L}_a = \mathbf{diag}(l_a^x, l_a^y)$, and length-scales, $\mathbf{L}_b = \mathbf{diag}(l_b^x, l_b^y)$. The scalar Δt_i accounts for the intervals between the data times t_i and the fixed estimation time t_0 (appendix A).

Today, typical horizontal grids contain 30 to 200 zonal/meridional nodes (e.g. Robinson 1996). The number of horizontal grid points ℓ_h is thus of order 10^3 to 10^4. For such dimensions, the numerical eigendecomposition of $\mathbf{B}$ is feasible and the usefulness of the subspace truncation can be evaluated. As an example, we consider horizontal sub-basin-scale and mesoscale error fields in the north-western Levantine. The numerical domain is 500 km in the x and 520 km in the y direction, with a horizontal resolution of 10 km. The size of $\mathbf{B}$ is [2073 × 2073]. The results regarding efficiency (LLA98) can be summarized as follows: for sub-basin-scale errors ($l_a^x = l_a^y = 200$ km, $l_b^x = l_b^y = 100$ km and $\tau = \infty$, horizontally constant for convenience), 10 eigenvectors represent 71% of the variance and 20 vectors 92%; for mesoscale errors ($l_a^x = l_a^y = 60$ km, $l_b^x = l_b^y =$ 30 km and $\tau = \infty$), a larger number is necessary, 20 eigenvectors represent 26% of the variance, 100 vectors 83% and 240 vectors 99%. A few eigenvectors corresponding to this mesoscale covariance are plotted in Fig. 1. One first observes that the patterns tend to be global, in part because of the orthogonality constraint. Secondly, the dominant vectors have a nearly uniform oscillatory character (e.g. panels 1–8). Thirdly the scale of the features explained by a given vector decreases as the vector number increases. These last two facts are associated with the lack of information on the recent dynamical evolution. In reality, every scale of motion has some memory which imposes dynamical constraints on the variations of an evolving field. For example, the eigenbase illustrated by Fig. 1 allows eddies to be anywhere, but in fact they are not. Dynamic eigenvalue spectra tend to decay more quickly than purely statistical ones which are often built assuming ergodicity. Such statistical models thus usually give upper bounds on the size of the error subspace. For other examples with anisotropic, non-homogeneous covariances or correlations, and their use in real-time ocean forecasts, we refer to LER97.

Instead of using a statistical model for the covariance function and eigendecomposing the covariance, one could directly decompose the historical data into horizontal EOFs (e.g. Preisendorfer 1988; Thacker and Lewandowicz 1997; Kaplan *et al.* 1997). In the applications of sections 4 and 5, the horizontal resolution of the data is not sufficient to do so. One could also decompose a simplified dynamics fit to data into horizontal normal modes (e.g. Monin 1990; Lozano and Candela 1995; Miller *et al.* 1996), but the open domain and multiscale regime of the examples considered are not suited to such decompositions.

(*b*) *Vertical error subspace*

In sections 4 and 5, to decompose the observed vertical error covariances, either a vertical function is fitted to the differences between the data and an a priori state $\mathbf{x}^{\mathrm{b}}$ and the resulting covariance eigendecomposed, or the vertical EOFs of these a priori data residuals are directly computed. Since the efficacy of reducing vertical covariances to

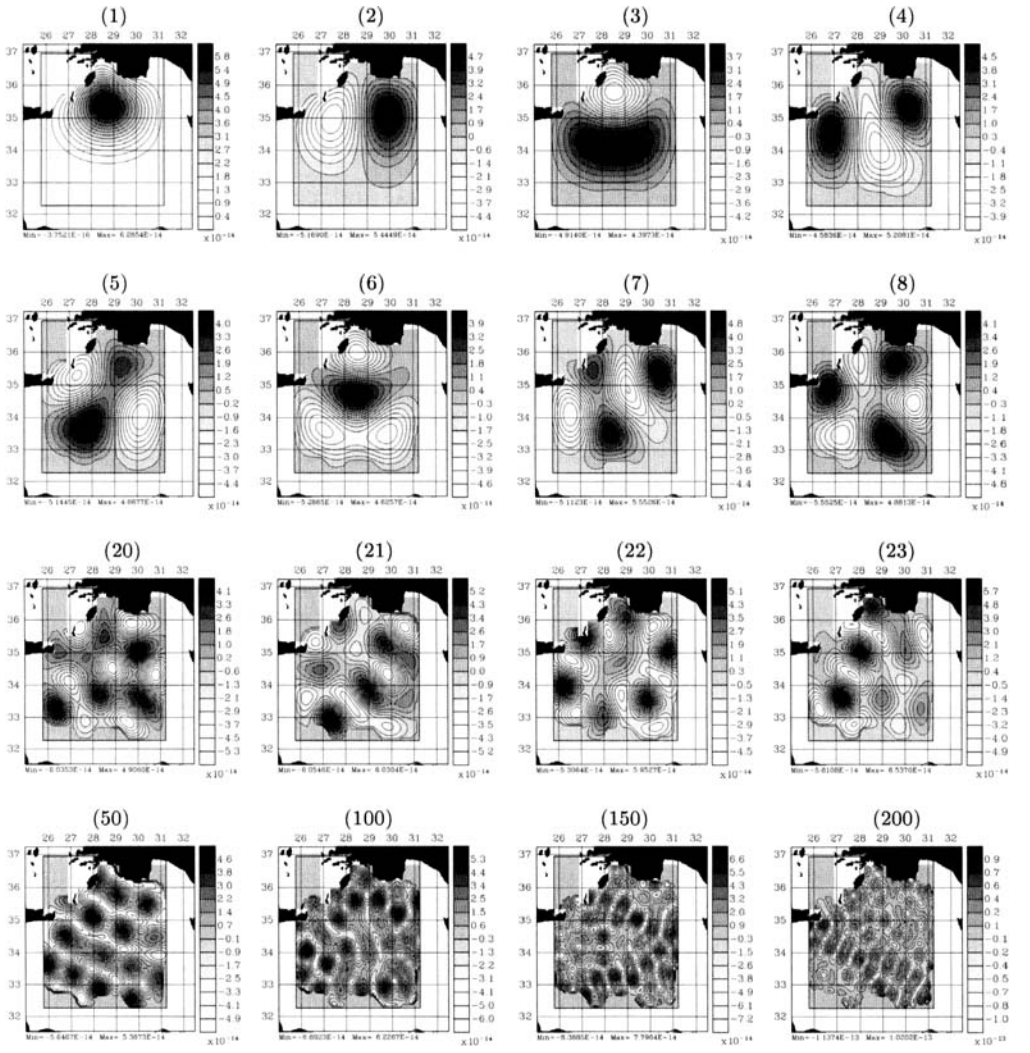

Figure 1. Eigenvectors of a tracer covariance for mesoscale 2D fields. The non-dimensional covariance function is a Mexican hat, with $l_a^x = l_a^y = 60$ km, $l_b^x = l_b^y = 30$ km and $\tau = \infty$. The panel numbers are the vector numbers. The dominant 8 vectors explain 11% of the variance, the dominant 200 vectors 98%. Higher vectors are useless in most computations. Scale decreases as the vector number increases, down to wavelengths of $\mathcal{O}(60$ km), e.g. vectors 100 and above. Due to the Mercator projection, some vectors appear distorted. All vectors are non-dimensional; the scale values have no absolute meaning. Error covariance values are presently weighted by a function proportional to the quantity of mesoscale observations in the vicinity of each grid point. Within the data array, this function here is one, and outside of the array, it decays to zero with a Gaussian scale of 30 km. In multiscale mappings, such horizontal weight functions lead to 'compactly supported' correlations and non-homogeneous error length-scales (e.g. Bouttier 1994; LER97; Gaspari and Cohn 1999). See text for further details.

their dominant components is established in oceanography and meteorology, it is not evaluated here.

Considering two depths z_1 and z_2, the vertical covariance functions we have explored are: $\epsilon^2/(1+\alpha\frac{|z_1-z_2|}{z_1+z_2})$, $\epsilon^2\exp(-\alpha\frac{|z_1-z_2|}{z_1+z_2})$, $\epsilon^2\exp\{-\alpha(\frac{z_1-z_2}{z_1+z_2})^2\}$ and $\epsilon^2\exp\{-\alpha(\frac{z_1-z_2}{D})^2\}$. The dimensional variance ϵ^2 and parameter α depend on the variable considered and are fitted to data; D is a characteristic vertical length. The first

three functions were selected so as to increase the vertical length-scale with the average depth, and usually give better results than the last one. Vertical covariance functions have also been specified for topography-following coordinates ('sigma' levels). For example, Bennett *et al.* (1995) employed $\epsilon^2\{1+\beta \ln^2(\frac{\sigma_1}{\sigma_2})\}^{-1}$, where σ_1, σ_2 are two sigma levels and ϵ^2 and β can be fitted to data.

Vertical EOFs are classic in geophysical estimations (e.g. von Storch and Frankignoul 1998), in particular for sea surface data extension (e.g. Haney *et al.* 1995; Gavart and DeMey 1997). They are efficiently computed using the singular-value decomposition (SVD). The present multivariate vertical EOF decomposition is given in appendix B. Another common approach is to compute the vertical normal modes of a simplified dynamics fit to data (Gill 1982; Monin 1990; Woodgate and Killworth 1996; Wunsch 1997).

(c) *Three-dimensional error subspace*

Because of the thin-fluid properties of the ocean and atmosphere, 'horizontal', $\mathbf{r}=(x, y)$, and 'vertical', z, coordinates play a special role: we denote by $\phi_{\mathbf{r},z}$ the value of a field ϕ at location $(\mathbf{r}, z)$, with the superscripts b for a priori and t for true (see appendix A for the conventions). Using notation similar to McWilliams *et al.* (1986), $C(\mathbf{r}_1, \mathbf{r}_2, z_1, z_2)$ is the error covariance function between $(\mathbf{r}_1, z_1)$ and $(\mathbf{r}_2, z_2)$:

$$C(\mathbf{r}_1, \mathbf{r}_2, z_1, z_2) \doteq \mathcal{E}[\{\phi^{\mathrm{b}}(\mathbf{r}_1, z_1)-\phi^{\mathrm{t}}(\mathbf{r}_1, z_1)\}\{\phi^{\mathrm{b}}(\mathbf{r}_2, z_2)-\phi^{\mathrm{t}}(\mathbf{r}_2, z_2)\}]. \quad (1)$$

The two factors in (1) contain the same field ϕ only for ease of notation. Fields of different nature (ϕ and φ) and thus cross-covariance functions $C_{\phi\varphi}$ are considered; the results to follow are employed in a multivariate context in section 4(a). The dominant error being assumed to be variability from ϕ^{b}, the vertical dependence in (1) can often be expressed efficiently by expanding $\phi^{\mathrm{b}}-\phi^{\mathrm{t}}$ in the form:

$$\phi^{\mathrm{b}}(\mathbf{r}, z)-\phi^{\mathrm{t}}(\mathbf{r}, z)=\sum_{i=0}^{\infty}\phi_i(\mathbf{r})Z_i^*(z), \quad (2)$$

where the $Z_i^*(z)$ terms are normalized vertical functions, determined from section 3(b). The $Z_i^*(z)$ terms are already averages, therefore substituting (2) into (1) yields:

$$C(\mathbf{r}_1, \mathbf{r}_2, z_1, z_2)=\sum_{i,j=0}^{\infty,\infty}A_{i,j}C_{i,j}^*(\mathbf{r}_1, \mathbf{r}_2)Z_i^*(z_1)Z_j^*(z_2), \quad (3a)$$

where the normalization factor $A_{i,j}=A_{j,i}$ is the horizontal average of $\mathcal{E}\{\phi_i(\mathbf{r})\phi_j(\mathbf{r})\}$ and $C_{i,j}^*(\mathbf{r}_1, \mathbf{r}_2) \doteq \mathcal{E}\{\phi_i(\mathbf{r}_1)\phi_j(\mathbf{r}_2)\}^*$ is the non-dimensional horizontal cross-covariance function associated with i and j, and is in general non-homogeneous and anisotropic. The aim is to determine decompositions (2) that describe most of $C(\mathbf{r}_1, \mathbf{r}_2, z_1, z_2)$ in (3a) with a few terms, i.e. the $A_{i,j}$ decay rapidly with increasing i and j. The expansion (3a), in general non-separable, is then efficiently truncated to:

$$C(\mathbf{r}_1, \mathbf{r}_2, z_1, z_2) \simeq \sum_{i,j=0}^{I,J}A_{i,j}C_{i,j}^*(\mathbf{r}_1, \mathbf{r}_2)Z_i^*(z_1)Z_j^*(z_2). \quad (3b)$$

To this end, other decompositions than those of section 3(b) have been considered: e.g. Taylor polynomial, Fourier or wavelet fits (McWilliams *et al.* 1986; Gamage and Blumen 1993).

In HOPS experiments, a typical number of levels (ℓ_v) is between 10 and 50. Using the horizontal sizes ℓ_h of section 3(a), the total number of grid points $\ell = \ell_h \ell_v$ is thus of order 10^4 to 10^6. Because of computer time and memory constraints, the direct use of (3b) in a global 3D analysis is not feasible. This obstacle can be surmounted if estimates of the dominant eigendecomposition of covariances based on (3b) are found. To derive such estimates, section 3(c)(i) considers the mapping of one scale or spectral window (e.g. Nihoul and Djenidi 1998) for the special case where the vertical functions in (2) are associated with a common non-dimensional horizontal covariance in (3b). This leads to horizontal/vertical separability and exact practical eigendecomposition. Section 3(c)(ii) considers the mapping of multiple scales and non-separable covariance functions; efficient decomposition schemes are outlined for the limit of isolated spectral windows and for the general situation (3b).

(i) *One scale: common non-dimensional horizontal error covariance functions.* In this sub-section, the $C^*_{i,j}(\mathbf{r}_1, \mathbf{r}_2)$ are assumed to be independent of the vertical combination i, j considered. If the Wiener–Khinchine theorem applies, all $C^*_{i,j}(\mathbf{r}_1, \mathbf{r}_2)$'s then have a single horizontal spectrum. This simplification has observational and theoretical support. For some mesoscale ocean datasets (e.g. Hua *et al.* 1986), it has been regionally observed that, for small I and J, e.g. $0 \leqslant I, J \leqslant 1$ in (3b), the spectra of the $C^*_{i,j}(\mathbf{r}_1, \mathbf{r}_2)$ do not vary significantly with i, j. Theoretically, for the mesoscale, mid-latitude open ocean, a few modes dominate the vertical problem, usually the barotropic and first baroclinic mode (e.g. Pedlosky 1987; Wunsch 1996, 1997). The series (3b) then contains 4 terms. In forced open-ocean (quasi-geostrophic) turbulence close to statistical equilibrium, these first two vertical modes tend to have the same horizontal spectrum (Rhines 1977; Salmon 1980; Haidvogel 1983). The corresponding spectral window, with length scales of the order of the first internal Rossby radius of deformation, is usually used to define the mesoscale. Hence, with $C^*_{i,j}(\mathbf{r}_1, \mathbf{r}_2) = R^*(\mathbf{r}_1, \mathbf{r}_2)$ in (3b):

$$C(\mathbf{r}_1, \mathbf{r}_2, z_1, z_2) = R^*(\mathbf{r}_1, \mathbf{r}_2) Z(z_1, z_2), \tag{4}$$

where $Z(z_1, z_2) = \sum_{i,j=0}^{I,J} A_{i,j} Z^*_i(z_1) Z^*_j(z_2)$. Since (4) is separable, for a suitable ordering of elements (LER97), the corresponding covariance $\mathbf{B} \in \mathrm{R}^{\ell \times \ell}$ is the Kronecker product[†] of the vertical $\mathbf{C}^z \in \mathrm{R}^{\ell_v \times \ell_v}$ and non-dimensional horizontal $\mathbf{C}^{\mathbf{r}^*} \in \mathrm{R}^{\ell_h \times \ell_h}$ covariances:

$$\mathbf{B} = \mathbf{C}^z \otimes \mathbf{C}^{\mathbf{r}^*}. \tag{5}$$

For every pair of vertical $\mathbf{e}_z$ and horizontal $\mathbf{e}_r$ error eigenvectors, with eigenvalues λ_z and λ_r, such that $\mathbf{C}^z \mathbf{e}_z = \lambda_z \mathbf{e}_z$ and $\mathbf{C}^{\mathbf{r}^*} \mathbf{e}_r = \lambda_r \mathbf{e}_r$, this yields

$$\mathbf{B} \mathbf{e}_z \otimes \mathbf{e}_r = \lambda_z \lambda_r \mathbf{e}_z \otimes \mathbf{e}_r. \tag{6}$$

The dominant eigendecomposition of $\mathbf{B}$ can be computed easily from those of $\mathbf{C}^z$, $\mathbf{E}^z \mathbf{\Pi}^z \mathbf{E}^{z^{\mathrm{T}}}$, and $\mathbf{C}^{\mathbf{r}^*}$, $\mathbf{E}^{\mathbf{r}^*} \mathbf{\Pi}^{\mathbf{r}^*} \mathbf{E}^{\mathbf{r}^{*\mathrm{T}}}$, obtained in sections 3(a) and 3(b). Using (6), one sorts the products $\lambda_z \lambda_r$, truncates for significant variance to the largest p eigenvalues, and only evaluates the corresponding eigenvectors $\mathbf{e}_z \otimes \mathbf{e}_r$. These eigenvalues and eigenvectors, constituting the diagonal of $\mathbf{\Pi}$ and columns of $\mathbf{E}$, determine the dominant subspace of $\mathbf{B}$. Figure 2 outlines the procedure. The result is the dominant rank-p

[†] The Kronecker product or tensor product of matrices $\mathbf{A} \in \mathrm{R}^{r \times s}$ and $\mathbf{B} \in \mathrm{R}^{m \times n}$ defines a block matrix $\mathbf{G} = \mathbf{A} \otimes \mathbf{B} \in \mathrm{R}^{rm \times sn}$ (Graham 1981).

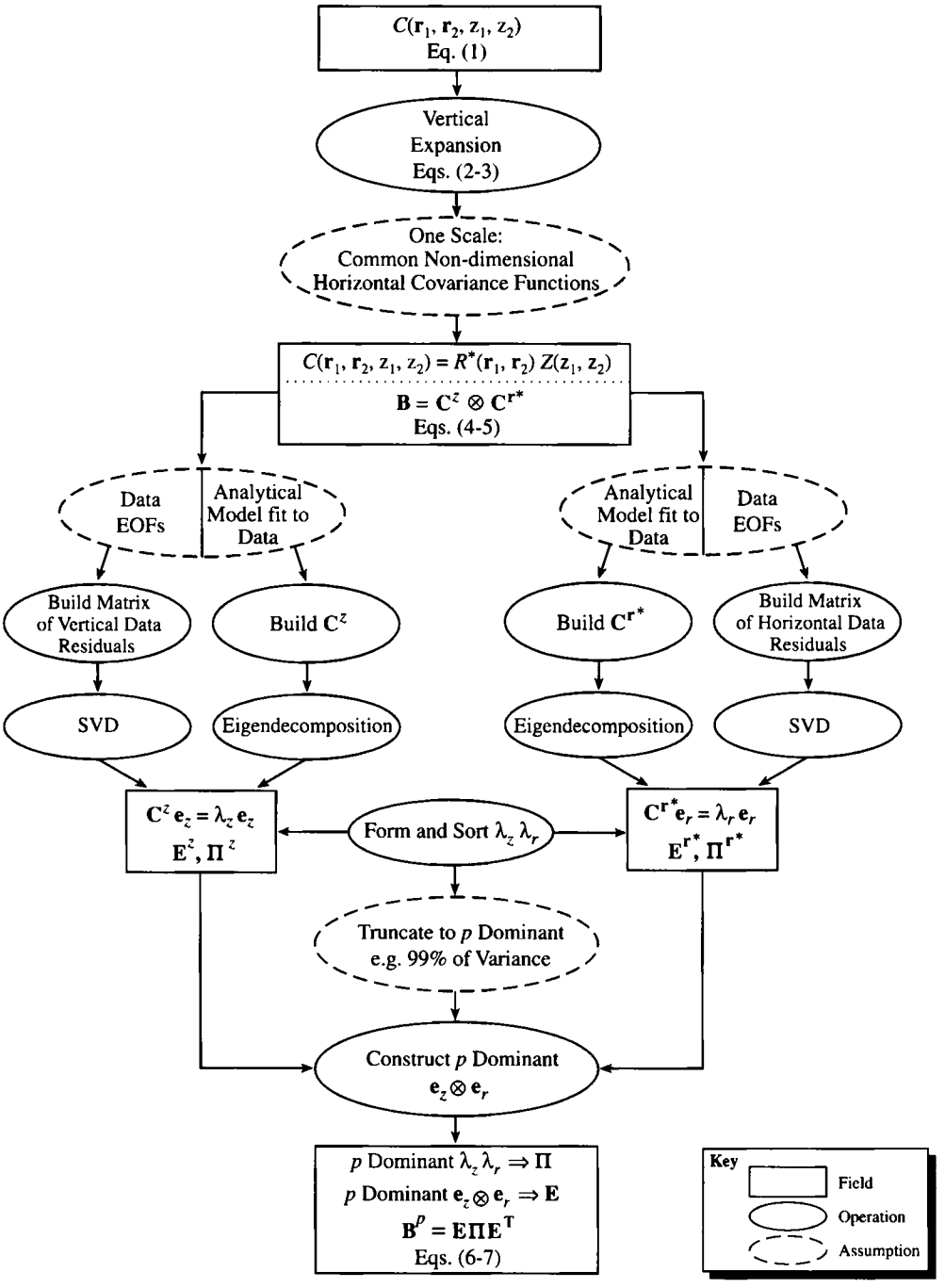

Figure 2. Flow diagram for constructing the rank-p eigendecomposition of a 3D error covariance or error subspace, for the assumptions of section 3(c)(i) of the text. The equation numbers refer to those in the text, which also gives definitions of variables.

eigendecomposition $\mathbf{B}^p$ (see appendix A for notation), under the assumptions of this sub-section,

$$\mathbf{B}^p = \mathbf{E}\mathbf{\Pi}\mathbf{E}^{\mathrm{T}}. \tag{7}$$

(ii) *Multiple scales of multiple horizontal error covariances.* In multiscale realistic flows, a common $C^*_{i,j}(\mathbf{r}_1, \mathbf{r}_2)$ is often not adequate. Nonetheless, considering the terms in (3b) one at a time gives, $C^*_{i,j}(\mathbf{r}_1, \mathbf{r}_2)A_{i,j}Z^*_i(z_1)Z^*_j(z_2) = C^*_{i,j}(\mathbf{r}_1, \mathbf{r}_2)Z_{i,j}(z_1, z_2)$, which is separable. Hence, using (4) and (5) for each term, (3b) becomes:

$$\mathbf{B} = \sum_{i,j=0}^{I,J} \mathbf{C}^z_{i,j} \otimes \mathbf{C}^{\mathbf{r}^*}_{i,j} = \sum_{i,j=0}^{I,J} \mathbf{B}_{i,j}, \tag{8}$$

where each $\mathbf{B}_{i,j} \doteq \mathbf{C}^z_{i,j} \otimes \mathbf{C}^{\mathbf{r}^*}_{i,j}$ is symmetric and positive semi-definite by construction ($\mathbf{C}^z_{i,j} = \mathbf{C}^{z^{\mathrm{T}}}_{i,j}$ and $\mathbf{C}^{\mathbf{r}^*}_{i,j} = \mathbf{C}^{\mathbf{r}^{*\mathrm{T}}}_{i,j}$). Estimating the dominant eigendecompositions of each $\mathbf{B}_{i,j}$ is feasible, using (5) and (6). Based on these facts, two schemes for the dominant decomposition of $\mathbf{B}$ in (8) are developed.

A practical iterative approach invokes independence of scales: the vertical components of distinct $\mathbf{C}^{\mathbf{r}^*}_{i,j}$ are hypothesized to be uncorrelated. Using the Wiener–Khinchine theorem, this amounts to separated spectral peaks. Within an isolated spectral window containing a given peak, the $\mathbf{C}^{\mathbf{r}^*}_{i,j}$ are identical. The scheme (4) to (6) of section 3(c)(i) then holds within each of these windows, indexed here by w; in practice with a small number of windows or scales:

$$\mathbf{B} = \sum_w \mathbf{B}_w = \sum_w \mathbf{C}^z_w \otimes \mathbf{C}^{\mathbf{r}^*}_w. \tag{9}$$

The successive-correction mapping technique is then attractive; because of the lack of correlation between scales, the $\mathbf{B}_w$ in (9) and scale-filtered data $\mathbf{y}^o_w$ (appendix B) are used successively to correct the corresponding scale in $\mathbf{x}^b$. Reducing each term to its subspace,

$$\mathbf{B} \simeq \sum_w \mathbf{B}^p_w = \sum_w \mathbf{E}_w \mathbf{\Pi}_w \mathbf{E}^{\mathrm{T}}_w. \tag{10}$$

The cost of the corrections is thus reduced by using successively the dominant eigendecomposition $\mathbf{B}^p_w$ of each $\mathbf{B}_w$ (appendix A, (A.2a) to (A.2c)). For each w, $\mathbf{E}_w$ and $\mathbf{\Pi}_w$ are estimated following Fig. 2, as for (7). The extension of (10) to multivariate fields (section 4(a)) is exemplified in section 5, in a two-scale analysis.

A scheme for the general case (3b) of correlated vertical components with multiple horizontal covariances is now outlined. Considering the eigendecompositions of each $\mathbf{B}_{i,j}$, the covariance $\mathbf{B}$ in (8) is a sum of outer-products of eigenvectors multiplied by their eigenvalues. The number of such outer-products, $n(I+1)(J+1)$, is much larger than the size, n, of $\mathbf{B}$. For efficiency, each $\mathbf{B}_{i,j}$ in (8) is first truncated to its dominant eigendecomposition, $\mathbf{E}_{i,j}\mathbf{\Pi}_{i,j}\mathbf{E}^{\mathrm{T}}_{i,j}$, using (5) and (6). The resulting approximation of $\mathbf{B}$ is the outer-product $\mathbf{L}\mathbf{L}^{\mathrm{T}} = \sum_{i,j=0}^{I,J} \mathbf{E}_{i,j}\mathbf{\Pi}_{i,j}\mathbf{E}^{\mathrm{T}}_{i,j}$ where

$$\mathbf{L} \doteq [\mathbf{E}_{0,0}\mathbf{\Pi}^{\frac{1}{2}}_{0,0}; \mathbf{E}_{0,1}\mathbf{\Pi}^{\frac{1}{2}}_{0,1}; \ldots; \mathbf{E}_{I,J}\mathbf{\Pi}^{\frac{1}{2}}_{I,J}], \tag{11}$$

with the semi-colon indicating a side-by-side concatenation. This approximation is dominant since it accounts for most of the variance of $\mathbf{B}$; the fraction of tr[$\mathbf{B}$] explained by $\mathbf{L}\mathbf{L}^{\mathrm{T}}$ is the sum of the tr[$\mathbf{\Pi}_{i,j}$] divided by tr[$\mathbf{B}$], which is close to one by construction†. Following section 2, $\mathbf{B} \simeq \mathbf{L}\mathbf{L}^{\mathrm{T}}$ is further optimally approximated by its dominant rank-p decomposition. This is efficiently obtained by SVD of $\mathbf{L}$,

$$\mathrm{SVD}_p(\mathbf{L}) = \mathbf{E}\mathbf{\Sigma}\mathbf{V}^{\mathrm{T}},\ \mathbf{\Pi} = \mathbf{\Sigma}^2 \text{ and } \mathbf{B}^p = \mathbf{E}\mathbf{\Pi}\mathbf{E}^{\mathrm{T}}, \tag{12}$$

where the operator $\mathrm{SVD}_p(\cdot)$ selects the dominant rank-p SVD. The flow diagram of Fig. 3 summarizes the procedure: a series of decomposition problems of type (4) to (7) shown in Fig. 2 are solved and, using (8), concatenated into (11). The rank-p SVD of $\mathbf{L}$ (11) estimates the multiscale a priori error subspace (12). It can then be used in a global analysis (appendix A, (A.2a) to (A.2c)) to compute the gridded fields and a posteriori error covariance. The main advantage of (11) and (12) over the 'window' approximation (9) and (10) is that (11) and (12) keep all multiscale interactions, up to these associated with the vertical vectors of number I and J. For efficiency ($\mathbf{L}$ of few columns), the key, of course, is to determine decompositions of type (3b) which explain most of the multiscale $C(\mathbf{r}_1, \mathbf{r}_2, z_1, z_2)$ for small numbers I and J. This scheme (11) and (12) is not exemplified in sections 4 and 5.

4. Error subspace: complete multivariate formulation

The dominant decomposition of multivariate 3D error covariances will now be completed. In the spirit of sections 1 and 2, to estimate their non-observed portions in accord with the observed ones (section 3), the most complete dynamics available are utilized. The genesis is first described, then particularized to 'oceanic' PE dynamics in sections 4(a) and 4(b), and exemplified in 4(c). The PE model utilized is that of HOPS. The state vector $\mathbf{x} \doteq (\widehat{\mathbf{u}}, \widehat{\mathbf{v}}, \mathbf{T}, \mathbf{S}, \mathbf{p}) \in \mathrm{R}^n$ contains values of the 3D (dynamical) tracers, the temperature T and salinity S, of the 3D zonal $\widehat{u}$ and meridional $\widehat{v}$ internal baroclinic velocities, and of the 2D barotropic transport stream function ψ ($\mathbf{p}$ in vector form).

An a priori state estimate $\mathbf{x}^{\mathrm{b}}$ is assumed to be known. For any pair of fields (ϕ, φ), the multivariate error covariance function $C_{\phi\varphi} = C_{\phi\varphi}(\mathbf{r}_1, \mathbf{r}_2, z_1, z_2)$ depends on this a priori state: for example, if $\mathbf{x}^{\mathrm{b}}$ is large-scale, $C_{\phi\varphi}$ is the covariance function of the variability from this scale and so on. Non-homogeneous $C_{\phi\varphi}$ are considered: e.g. $C_{\phi\varphi}$ may account for shelf and deep ocean processes, boundary layers or biological patchiness. Since the time is fixed, non-stationarity (e.g. seasonal forcing) is addressed by modifying the $C_{\phi\varphi}$. The error covariance $\mathbf{B}$ is large, $n \sim \mathcal{O}(10^5 \text{ to } 10^7)$. Its eigendecomposition is challenging, in part because of the multivariate character.

Assuming for a moment that data are sufficient to use section 3(c) for each pair (ϕ, φ), the functions $C_{\phi\varphi}$ can be decomposed similarly to (3b). Under section 3(c)(i), $C_{\phi\varphi}$ is separable in the horizontal/vertical (4). Under section 3(c)(ii), it is equal to a sum of such functions. Restricting the discussion to a single term (multiple terms can be treated as in section 3(c)(ii), and lead to the same general conclusion), the assumption is:

$$C_{\phi\varphi}(\mathbf{r}_1, \mathbf{r}_2, z_1, z_2) = R^*_{\phi\varphi}(\mathbf{r}_1, \mathbf{r}_2) Z_{\phi\varphi}(z_1, z_2). \tag{13}$$

† If $\mathbf{\Pi}_{i,j}$ accounts for fraction $\delta_{i,j}$ of the tr[$\mathbf{B}_{i,j}$], $\mathbf{L}\mathbf{L}^{\mathrm{T}}$ accounts for fraction $\sum_{i,j=0}^{I,J} \delta_{i,j}\mathrm{tr}[\mathbf{B}_{i,j}]/\mathrm{tr}[\mathbf{B}]$ of the tr[$\mathbf{B}$]. If all $\delta_{i,j}$ are equal to δ (e.g. 0.99), this sum is always δ, regardless of the properties of the $\mathbf{B}_{i,j}$. In practice, a constant $\delta_{i,j}$ cannot be exactly achieved but, since the $\mathbf{B}_{i,j}$ are very large and have rapidly decaying eigenvalues, it can be very close.

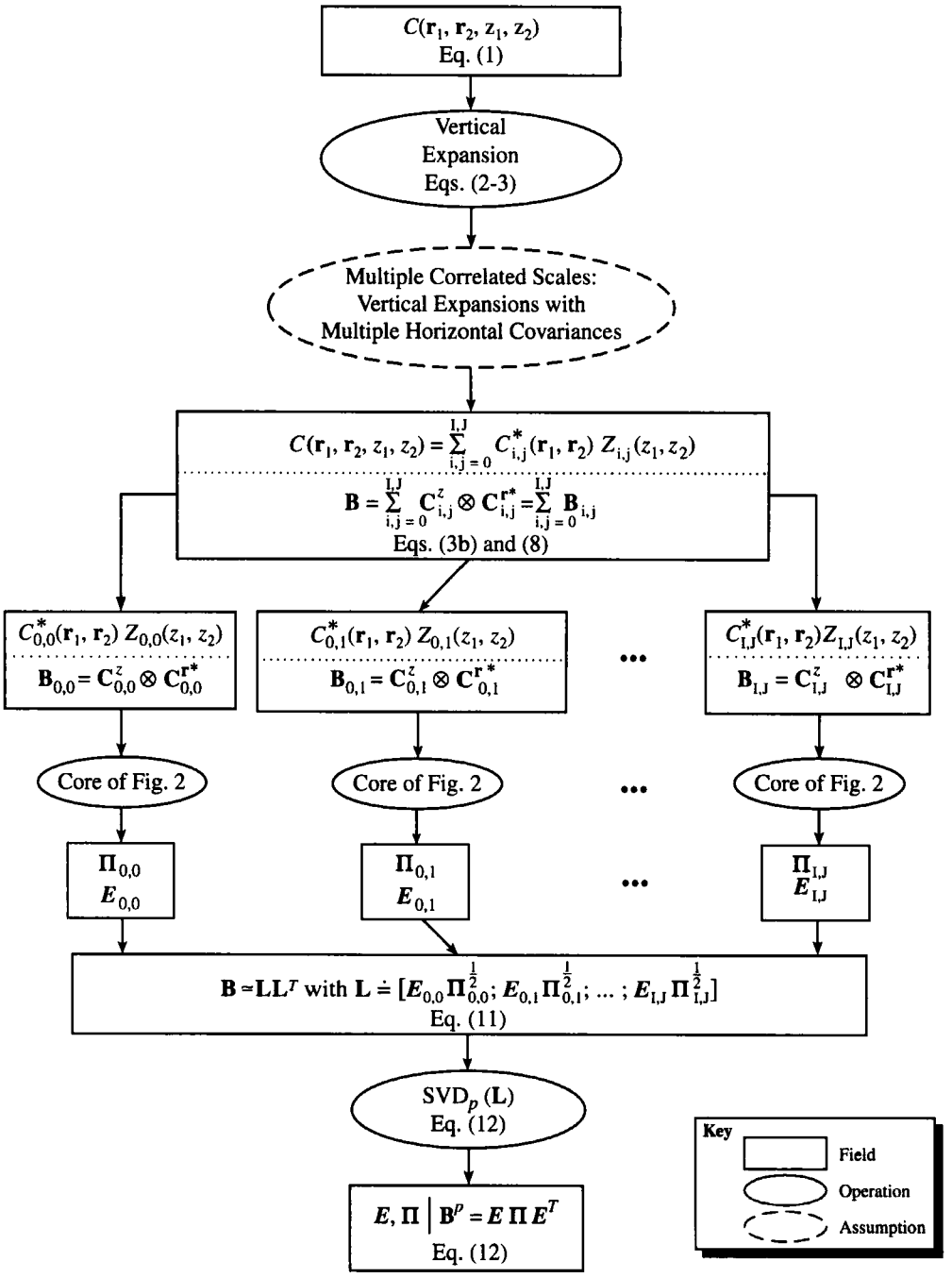

Figure 3. As Fig. 2, but for the assumptions of section 3(c)(ii).

For a suitable ordering of its elements, the multivariate **B** constructed from (13) is formed of submatrices $\mathbf{B}_{\phi\varphi}$ associated with a pair (ϕ, φ) which can be decomposed as in (5): e.g. $\mathbf{B}_{\widehat{uT}} = \mathbf{C}^{z}_{\widehat{uT}} \otimes \mathbf{C}^{\mathbf{r}^*}_{\widehat{uT}}$. However, for most dynamical models, the non-dimensional horizontal covariances in these submatrices are all different. For example, the dominant, mid-latitude geophysical flows are usually close to geostrophic balance, which involves horizontal gradients of pressure; e.g. $\mathbf{C}^{\mathbf{r}^*}_{\widehat{uT}}$, $\mathbf{C}^{\mathbf{r}^*}_{\widehat{uu}}$, $\mathbf{C}^{\mathbf{r}^*}_{TT}$ or $\mathbf{C}^{\mathbf{r}^*}_{T\psi}$ are thus different. This prevents the direct decomposition of **B** using Kronecker product properties as in sections 3(c)(i) and 3(c)(ii).

Nonetheless, the dominant eigendecompositions of the observed portions of **B**, a few submatrices $\mathbf{B}_{\phi\varphi}$, can be obtained from sections 3(c)(i) and 3(c)(ii). The remaining non-observed portions of **B** can then be built in dynamical accord with the observed ones, via an *ensemble of adjustment dynamical integrations*. To initialize such an integration, the dominant eigendecompositions of the observed $\mathbf{B}_{\phi\varphi}$ are used to perturb the corresponding variables (ϕ, φ) in the a priori state $\mathbf{x}^{\mathrm{b}}$. The dynamical equations governing the other variables (corresponding to the non-observed portions of **B**) are then integrated forward in time from the perturbed $\mathbf{x}^{\mathrm{b}}$, keeping the perturbed fields (ϕ, φ) fixed. Stationary stochastic and external forcings can be used to account for model uncertainties and boundary effects, respectively. The integration is usually continued until the volume-average kinetic and/or potential energy stabilize around a plateau, without rapid time-rate-of-change. The final state is said to be *dynamically adjusted*. The ensemble of differences between the dynamically adjusted states and the unperturbed state, $\mathbf{x}^{\mathrm{b}}$, are samples of a priori missing variability from $\mathbf{x}^{\mathrm{b}}$. They are normalized, organized by SVDs and the ensemble integrations ultimately stopped when a convergence criterion is satisfied. The variability amplitudes can then be scaled to appropriate a priori error values by scalar multiplications (section 4(b)).

This approach has several properties. (i) The portions of the error covariance directly specified from data (section 3) account for the properties of the real variability (e.g. non-homogeneous or multiscale). (ii) The ensemble of adjustment integrations constructs the non-observed portions of **B** by dynamical cross-covariances with the dominant decompositions of the observed portions. (iii) The nonlinearities and possible forcings excite strict auto-covariances among the non-observed portions of **B**, by dynamical interactions in the range of scales of interest; the uncontrolled components of this non-observed, but dynamically adjusted, variability are in fact pure uncertainties. (iv) The resulting error subspace estimate is in accord with the complexities of the historical data and dynamical model.

(*a*) *Dominant decomposition of the observed portions of the error subspace*

In oceanography, hydrographic data (T and S) are the most common and the observed portions of **B** are thus assumed here to be the tracer covariances. For fair comparisons with the multiscale univariate 2D scheme of HOPS, the successive-correction approach (9) and (10) of section 3(c)(ii) is used instead of the general one (11) and (12). For both T and S the assumptions are therefore: a series of independent scales and, for each scale, vertical eigenvectors in (2) of common non-dimensional horizontal error covariance functions (the normalization factor is the horizontally averaged variance for this scale). For ease of notation, we consider only one scale or spectral window w at a time, i.e. one term $\mathbf{B}_w$ in (9) and (10), so that we can omit for now the summation

over w. This yields, as in (4) of section 3(c)(i):

$$C_{TT}(\mathbf{r}_1, \mathbf{r}_2, z_1, z_2) = R^*_{TT}(\mathbf{r}_1, \mathbf{r}_2) Z_{TT}(z_1, z_2), \tag{14a}$$

$$C_{SS}(\mathbf{r}_1, \mathbf{r}_2, z_1, z_2) = R^*_{SS}(\mathbf{r}_1, \mathbf{r}_2) Z_{SS}(z_1, z_2), \tag{14b}$$

$$C_{TS}(\mathbf{r}_1, \mathbf{r}_2, z_1, z_2) = R^*_{TS}(\mathbf{r}_1, \mathbf{r}_2) Z_{TS}(z_1, z_2). \tag{14c}$$

The functions R^*_{TT}, R^*_{SS} and R^*_{TS} are also assumed to be equal, which implies

$$\mathbf{C}^{\mathbf{r}^*}_{TT} = \mathbf{C}^{\mathbf{r}^*}_{TS} = \mathbf{C}^{\mathbf{r}^*}_{SS} = \mathbf{C}^{\mathbf{r}^*}_{\text{trc}}, \tag{15}$$

where the subscript trc stands for tracers and $\mathbf{C}^{\mathbf{r}^*}_{\text{trc}} \in \mathrm{R}^{\ell_h \times \ell_h}$. Using (14) and (15) and section 3(c) for each scale, the 3D tracer covariance $\mathbf{B}_{\text{trc}} \in \mathrm{R}^{2\ell \times 2\ell}$ is, for a suitable ordering of its elements, the Kronecker product:

$$\mathbf{B}_{\text{trc}} = \mathbf{C}^{z}_{\text{trc}} \otimes \mathbf{C}^{\mathbf{r}^*}_{\text{trc}}, \tag{16}$$

where $\mathbf{C}^{z}_{\text{trc}} \in \mathrm{R}^{2\ell_v \times 2\ell_v}$. For each scale w, the vertical and horizontal eigendecompositions are obtained from sections 3(a) and (b), respectively:

$$\mathbf{C}^{z}_{\text{trc}} = \mathbf{E}^{z}_{\text{trc}} \mathbf{\Pi}^{z}_{\text{trc}} \mathbf{E}^{z^{\mathrm{T}}}_{\text{trc}}, \tag{17a}$$

$$\mathbf{C}^{\mathbf{r}^*}_{\text{trc}} = \mathbf{E}^{\mathbf{r}^*}_{\text{trc}} \mathbf{\Pi}^{\mathbf{r}}_{\text{trc}} \mathbf{E}^{\mathbf{r}^{*\mathrm{T}}}_{\text{trc}}. \tag{17b}$$

The dominant eigendecomposition of (16) then simply follows from relations as (6) and (7):

$$\mathbf{B}_{\text{trc}} \simeq \mathbf{B}^{p}_{\text{trc}} = \mathbf{E}_{\text{trc}} \mathbf{\Pi}_{\text{trc}} \mathbf{E}^{\mathrm{T}}_{\text{trc}}. \tag{18}$$

The sum over w of decompositions (16) to (18) extends (9) and (10) to the multivariate case. It is illustrated in section 4(c) and employed for a two-scale (i.e. $w = 1, 2$) 3D mapping in section 5. In passing, if (14a) to (14c) but not (15) hold for each scale w, the submatrices $\mathbf{B}_{TT}$, $\mathbf{B}_{TS}$, $\mathbf{B}_{SS}$ can be expressed first, using (4) and (5), as $\mathbf{C}^{z}_{TT} \otimes \mathbf{C}^{\mathbf{r}^*}_{TT}$, $\mathbf{C}^{z}_{TT} \otimes \mathbf{C}^{\mathbf{r}^*}_{SS}$ and $\mathbf{C}^{z}_{SS} \otimes \mathbf{C}^{\mathbf{r}^*}_{SS}$, respectively. The dominant eigendecompositions of these matrices can then be computed and employed to perturb $\mathbf{x}^{\mathrm{b}}$, to create an ensemble of initial conditions for adjustment dynamical integrations (section 4(b)). The hypothesis (15) is only used here in order to compare with the univariate 2D scheme.

(*b*) *Ensemble of adjustment PE integrations for the complete error subspace*

For most ocean phenomena, the dominant physical variability involves velocities. Under these circumstances, an ensemble of adjustment momentum integrations can construct the non-observed velocity variability in accord with the sum over w of dominant tracer decompositions (18). For ease of notation, as in section 4(a), only one scale w is considered. The first step is to sample the observed tracer portions. An ensemble of perturbed tracer fields, $\mathbf{x}^{\mathrm{b}^j}_{\text{trc}}$, whose covariance from $\mathbf{x}^{\mathrm{b}}_{\text{trc}}$ tends to (18) is defined by:

$$\mathbf{x}^{\mathrm{b}^j}_{\text{trc}} = \mathbf{x}^{\mathrm{b}}_{\text{trc}} + \mathbf{E}_{\text{trc}} \mathbf{\Pi}^{\frac{1}{2}}_{\text{trc}} \mathbf{u}^j, \quad j = 1, \ldots, q, \tag{19}$$

where $\mathbf{u}^j \in \mathrm{R}^p$ is a realization of the random vector $\mathbf{u} \in \mathrm{R}^p$ of zero mean, identity covariance $\mathbf{I}_p$, and, most often, Gaussian distribution. To cover extreme cases with a limited number of realizations, one can replace the $\mathbf{u}^j$'s by the $q = p$ base vectors

$\mathbf{e}^j$ multiplied by $\sqrt{p}$ so that $\frac{1}{q}\sum_{j=1}^{q}\mathbf{u}^j\mathbf{u}^{j^{\mathrm{T}}} = \mathbf{I}_p$. For flows with possibly strong nonlinearities, the $q = 2p$ vectors $\mathbf{u}^j = \pm\sqrt{2p}\mathbf{e}^j$ may also be an advantageous choice. In these last two cases, the ensemble tracer covariance is exactly (18); if the $\mathbf{u}^j$ are random, the ensemble covariance converges towards (18) with a standard-deviation error decay of $\mathcal{O}(1/\sqrt{q})$. Our experience suggests that $\mathbf{u}^j = \sqrt{q}\mathbf{e}^j$ and random $\mathbf{u}^j$ are often the most appropriate choices for use in linear and nonlinear models, respectively. Using (19), the resulting unbalanced a priori state vectors are:

$$\widetilde{\mathbf{x}}^{\mathrm{b}j} \doteq (\widehat{\mathbf{u}}^{\mathrm{b}}, \widehat{\mathbf{v}}^{\mathrm{b}}, \mathbf{T}^{\mathrm{b}j}, \mathbf{S}^{\mathrm{b}j}; \mathbf{p}^{\mathrm{b}}), \tag{20}$$

where $\mathbf{x}_{\mathrm{trc}}^{\mathrm{b}j} \doteq (\mathbf{T}^{\mathrm{b}j}; \mathbf{S}^{\mathrm{b}j})$ is given by (19) and the vectors $\widehat{\mathbf{u}}^{\mathrm{b}}$, $\widehat{\mathbf{v}}^{\mathrm{b}}$ and $\mathbf{p}^{\mathrm{b}}$ define the total velocity fields of $\mathbf{x}^{\mathrm{b}}$. The states $\widetilde{\mathbf{x}}^{\mathrm{b}j}$ (20) are brought to statistical PE equilibrium by adjustment dynamical integrations. Keeping $\mathbf{x}_{\mathrm{trc}}^{\mathrm{b}j}$ fixed, the nonlinear (and stochastic) PE momentum equations are integrated forward, usually until the volume-averaged kinetic energy stabilizes around a plateau, without rapid time-rates-of-change. The resulting final total velocities $(\widehat{\mathbf{u}}^{\mathrm{b}j}, \widehat{\mathbf{v}}^{\mathrm{b}j}, \mathbf{p}^{\mathrm{b}j})$ are dynamically adjusted to the perturbed tracers $\mathbf{x}_{\mathrm{trc}}^{\mathrm{b}j}$. Denoting the corresponding q adjusted states by $\mathbf{x}^{\mathrm{b}j}$, the matrix $\mathbf{S}^{\mathrm{b}}$ whose columns are samples of the error subspace is computed:

$$\mathbf{S}^{\mathrm{b}} \doteq [\mathbf{x}^{\mathrm{b}1} - \mathbf{x}^{\mathrm{b}}; \ldots; \mathbf{x}^{\mathrm{b}q} - \mathbf{x}^{\mathrm{b}}]. \tag{21}$$

In (21), the rows of $\mathbf{S}^{\mathrm{b}}$ corresponding to the differences $(\widehat{\mathbf{u}}^{\mathrm{b}j} - \widehat{\mathbf{u}}^{\mathrm{b}}, \widehat{\mathbf{v}}^{\mathrm{b}j} - \widehat{\mathbf{v}}^{\mathrm{b}}, \mathbf{p}^{\mathrm{b}j} - \mathbf{p}^{\mathrm{b}})$ are the non-observed portions of these samples. The matrix $\mathbf{S}^{\mathrm{b}}$ is then normalized and the SVD† of $\mathbf{S}^{\mathrm{b}*}$ evaluated. The left singular vectors $\mathbf{E}$ estimate the a priori error subspace eigenbase. New integrations are carried out in parallel and columns added to $\mathbf{S}^{\mathrm{b}}$ (increasing q), until the added value of new samples is determined to be small enough. In the examples (section 4(c)), the convergence criterion used is (LER97):

$$\rho = \frac{\sum_{i=1}^{\widetilde{p}} \sigma_i(\widetilde{\mathbf{\Pi}}^{\frac{1}{2}}\widetilde{\mathbf{E}}^{*^{\mathrm{T}}}\mathbf{E}^*\mathbf{\Pi}^{\frac{1}{2}})}{\sum_{i=1}^{p} \sigma_i(\mathbf{\Pi})} \geqslant \alpha, \tag{22}$$

where α is a chosen convergence limit $(1 - \epsilon \leqslant \alpha \leqslant 1)$; $\sigma_i(\cdot)$ selects the singular value number i; and $(\mathbf{E}^*, \mathbf{\Pi} = \frac{1}{q}\mathbf{\Sigma}^2)$ of rank p and $(\widetilde{\mathbf{E}}^*, \widetilde{\mathbf{\Pi}} = \frac{1}{q}\widetilde{\mathbf{\Sigma}}^2)$ of rank $\widetilde{p} \leqslant p$ correspond to the 'new' and 'previous' estimate of $\mathbf{B}^p$, respectively. When such a criterion is satisfied (i.e. ρ is close enough to one), the significant $\mathrm{SVD}_p(\mathbf{S}^{\mathrm{b}}) = \mathbf{E}\mathbf{\Sigma}\mathbf{V}^{\mathrm{T}}$, where $p \leqslant q$, yields the first-guess estimate of the error subspace covariance of the 3D multivariate $\mathbf{x}^{\mathrm{b}}$:

$$\mathbf{B}^p = \mathbf{\Gamma}\mathbf{E}\mathbf{\Pi}\mathbf{E}^{\mathrm{T}}\mathbf{\Gamma}^{\mathrm{T}}. \tag{23}$$

In (23), $\mathbf{\Gamma}$ is a block diagonal matrix, scaling the variability variances of each PE field to appropriate error variances. In practice, we estimate $\mathbf{\Gamma}$ from past experience or from the misfits between $\mathbf{x}^{\mathrm{b}}$ and the synoptic data. We have also adapted $\mathbf{\Gamma}$ in real time, such that forecast-error variances are in accord with the variances of the misfits between new data and the state forecast (LER97). In all regions of the Mediterranean and North

† For multivariate matrices, the SVDs are carried out on normalized matrices $\mathbf{S}^*$ such that $\mathbf{S} = \mathbf{N}\mathbf{S}^* = \mathbf{N}\mathbf{E}^*\mathbf{\Sigma}\mathbf{V}^{\mathrm{T}}$ and $\mathbf{E} = \mathbf{N}\mathbf{E}^*$, where the normalization matrix $\mathbf{N} \in \mathrm{R}^{n\times n}$ is diagonal. For each field, the associated diagonal element of $\mathbf{N}$ is the volume and sample averaged variance. In general, $\mathrm{SVD}_p(\cdot)$ denotes the operator that carries out this normalization, selects the dominant rank-p SVD and renormalizes the output, to yield $\mathbf{E}\mathbf{\Sigma}\mathbf{V}^{\mathrm{T}}$.

Atlantic investigated, we have found that the estimated elements of $\mathbf{\Gamma}$ are $\mathcal{O}(1)$. Note that in section 3, which mainly considered univariate fields, this matrix multiplication was a scalar multiplication and, for ease of notation, it was omitted. The flow diagram of Fig. 4 summarizes the steps involved in the (parallel) computation of (23).

In a context frequent in oceanography, velocity covariances are built from the observed tracer covariances, in accord with the complete PE dynamics and parametrizations (boundary layers, Reynolds stresses, etc.). With stochastic forcings, the statistical effects of model errors or stochastically modelled processes are accounted for. During such momentum integrations, for most tracer perturbations, at first-order and away from boundaries, the velocities usually tend to thermal-wind equilibrium while their vertical average undergoes an adjustment as in Gill (1982). After an ensemble of integrations, the significant decomposition of the normalized differences of PE adjusted fields (21) to (23) reflect these properties. In particular, the 3D tracer-variability subspace (18) is usually reorganized. Even though the adjustments include a joint effect of baroclinicity and relief (e.g. Cane *et al.* 1998), it is not a search for steady states. The aim is to obtain statistics of time-rates-of-change corresponding to PE dynamics, in a high-Reynolds-number regime. This procedure for generating the dominant decomposition of global error covariances for large, nonlinear and stochastic numerical models extends the direct geostrophic and thermal-wind balance approach. Such reduced-physics hypotheses have, for example, been used in meteorology by Daley (1991, 1992a,b,c) and by Jiang and Ghil (1993) to determine model error covariances, and in oceanography by McWilliams *et al.* (1986) for objective analysis of the POLYMODE data. In meteorology, these hypotheses were recently extended for use in global 3D-Var algorithms by applying the linear balance equation or a Rossby–Hough expansion (Daley 1991). However, these two global constraints appear incorrect in the Tropics and, for a remedy, Daley (1996) derived a procedure based on the SVD of their matrix representations.

(c) *Examples of multivariate error (a priori missing variability) subspace*

(i) *Middle Atlantic Bight shelfbreak front.* To exemplify the construction (14) to (23) summarized in Fig. 4, the case of the mesoscale error subspace of an idealized feature model of Sloan (1996) for the Middle Atlantic Bight shelfbreak front is employed. Figure 5 portrays this construction. The main feature at the shelfbreak is a tilted temperature and salinity front, separating the cold and fresh shelf water from the warm and salty slope water. In the idealized model, the background tracers $\mathbf{x}^{\mathrm{b}}_{\mathrm{trc}}$ consist of a zonally uniform front. The shelf water is to the north, the slope water to the south. The geometry is a periodic channel. The topography is also zonally uniform, with a constant slope of sign opposite to that of the front. The PE background state $\mathbf{x}^{\mathrm{b}}$ is in thermal-wind balance; the flow is east–west, with zero velocity at the bottom. The dynamics is that of the f-plane PEs. The mesoscale 3D tracer covariance (sections 3(c)(i) and 4(a)) is specified, based on statistical models, for both the horizontal and vertical components. Panel (a) of Fig. 5 shows the corresponding normalized cumulative variance of T as a function of the 3D eigenvalue number. The horizontal covariance function was a Mexican hat, with zero-crossings $l^x_a = l^y_a = 27$ km, and anisotropic decay-scales $l^x_b = 7$ km, $l^y_b = 5$ km (section 3(a)). The non-dimensional vertical covariance function was $\exp\{-\alpha(\frac{z_1-z_2}{z_1+z_2})^2\}$, with $\alpha = 4$ (section 3(b)). These horizontal and vertical parameters were fitted to real and simulated data. Panel (b) is a cross-section in the 3D non-dimensional T eigenvector number 100, computed using (5) and (6). Panel (c) is the horizontal map of the ψ response, dynamically adjusted to T and S perturbations of

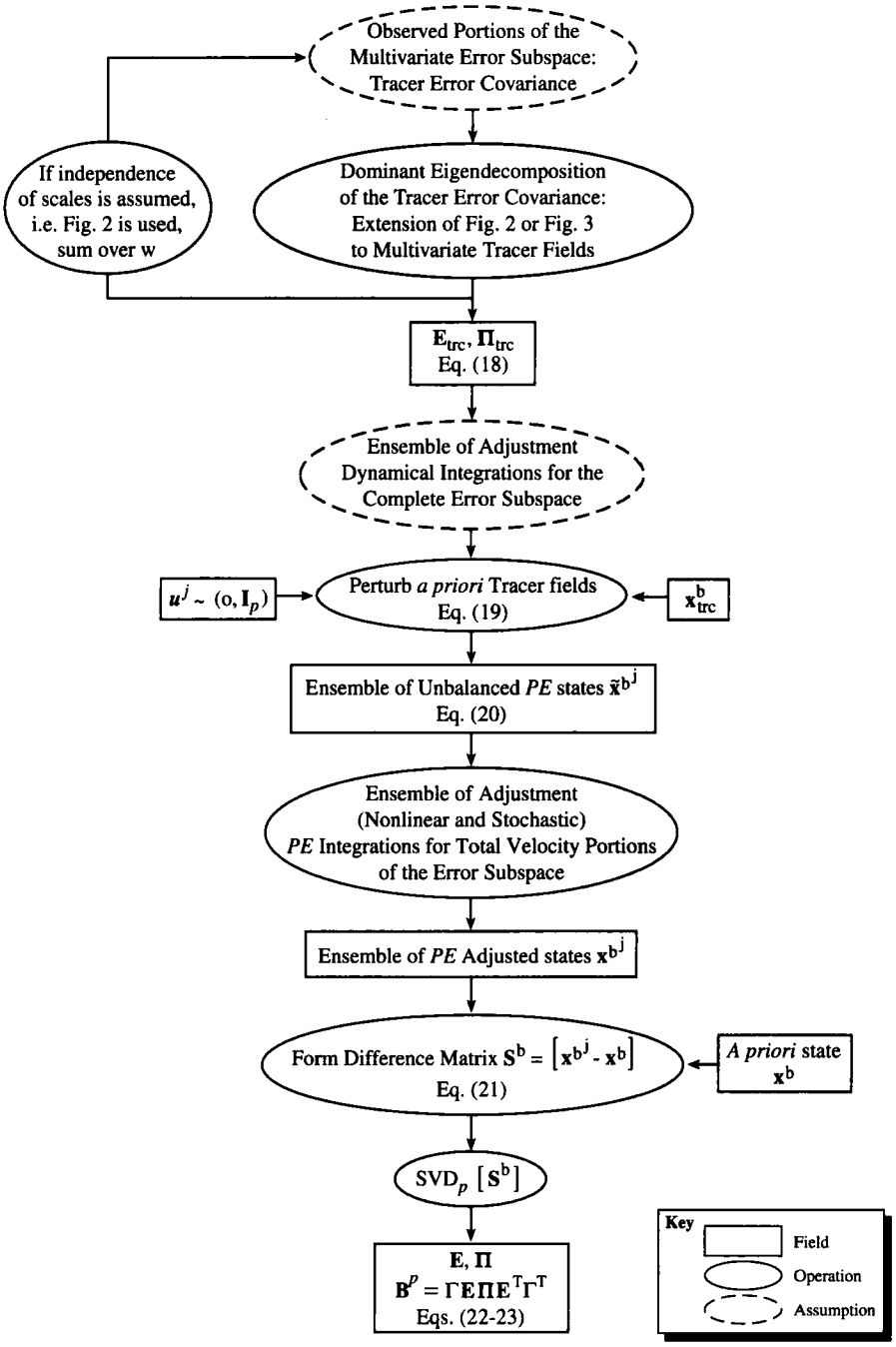

Figure 4. Flow diagram for constructing the rank-p eigendecomposition $\mathbf{B}^p$ (23) of 3D PE multivariate error covariances $\mathbf{B}$. The equation numbers refer to those in the text. The construction consists of two main stages (sections 4(a) and 4(b) in the text). The first stage (top) specifies the observed portions of $\mathbf{B}$: the dominant eigendecomposition of the tracer (T, S) 3D error covariance, $\mathbf{B}^p_{trc} = \mathbf{E}_{trc} \mathbf{\Pi}_{trc} \mathbf{E}^T_{trc}$ (18), is constructed. The second stage is the ensemble of (nonlinear and stochastic) adjustment PE integrations to compute the non-observed portions of $\mathbf{B}$ based on the knowledge of $\mathbf{B}^p_{trc}$: the result is an estimate of $\mathbf{B}^p$. Parallel computation is utilized in this second stage. See text for additional details.

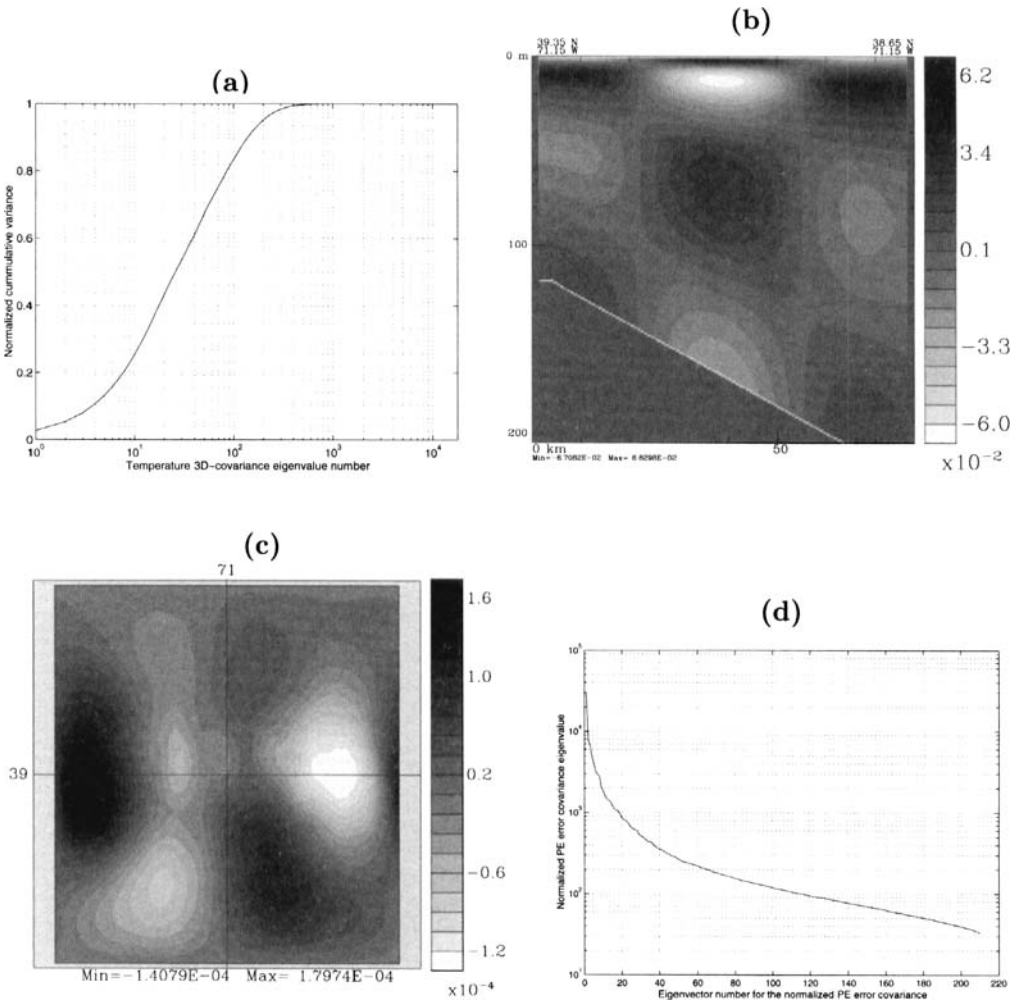

Figure 5. Construction of a mesoscale error subspace for an idealized feature model (Sloan 1996) of the Middle Atlantic Bight shelfbreak front. (a) The normalized cumulative variance as a function of the eigenvalue number for the 3D temperature mesoscale variability covariance [17794 × 17794] estimate. The dominant 100 eigenvectors explain 83.4% of the variance. (b) A vertical north-south cross-section (north is left) in the 3D non-dimensional T eigenvector number 100. (c) The horizontal map (north at the top) of the non-dimensional barotropic transport stream function response dynamically adjusted to a T and S perturbation of shape shown on (b). The response agrees with the tracers, f-plane and topography, as is for instance shown by its north–south varying intensity. (d) The eigenvalue spectrum of the normalized PE error covariance, i.e. diagonal of $\mathbf{\Pi}$ in Eq. (23), obtained after 210 adjustment dynamical integrations. The normalization divides all 3D error fields by their respective sample and volume averaged variance. See text for further details.

shape shown in panel (b). The ensemble of adjustment PE integrations (section 4(b)) for every T and S eigenvector built the matrix (21). As (21) grew in size, SVD_p's were evaluated and, using the criterion (22) with $\alpha = 0.97$, the adjustments were stopped after 210 integrations to yield (23). The eigenvalue spectrum, diagonal of $\mathbf{\Pi}$ in (23), is shown in panel (d). Notice the span of three orders of magnitude. The dominant 100 vectors explain more than 90% of the variance explained by the 210 vectors. The matrices $\mathbf{E}$ and $\mathbf{\Pi}$ can then be used for PE-based mapping (appendix A, (A.1) and (A.2)).

(ii) *Levantine basin.* The second illustration computes the dominant decomposition of a mesoscale PE variability covariance in the north-west, open-ocean Levantine Sea, and briefly evaluates the results with respect to what was previously known about the region. We refer to Fig. 2(b) of Robinson and Golnaraghi (1993) and section 5(a) hereafter for the name and location of common upper thermocline features in the region.

Presently, the a priori state $\mathbf{x}^b$ is a field estimate for 16 March 1995 (not shown). The observed, a priori missing, mesoscale 3D tracer covariance (sections 3(c)(i) and 4(a)) is specified based on EOF computations in the vertical and on a statistical model in the horizontal. Vertically (section 3(b) and appendix B), the EOFs of the tracer residuals are computed based on a dataset of 289 CTD (conductivity–temperature–depth) profiles gathered by the Physical Oceanography of the Eastern Mediterranean (POEM) group between 10 February and 16 April 1995 (Malanotte-Rizzoli *et al.* 1996). Horizontally (section 3(a)), the covariance function is a Mexican hat, with zero-crossings $l_a^x = l_a^y =$ 60 km and decay-scales $l_b^x = l_b^y = 30$ km, fitted to the residuals of the CTD data. These vertical (17a) and horizontal (17b) decompositions are combined using (16) and, by truncation, yield the dominant 3D tracer covariance (18). The complete error subspace is then obtained by nonlinear adjustment PE integrations (19) to (23), in the present case, with the choice $\mathbf{u}^j = \sqrt{q}\mathbf{e}^j$ in (19). Each nonlinear adjustment lasted for 2 model days (for model parameters see LER97). The ensemble of runs was stopped after $q = 400$ parallel integrations, based on the criterion (22) with $\tilde{q} = 360$, which led to $\rho = 0.935$. The aim was to capture 93% of the error variance.

(iii) *Levantine covariance eigenvectors.* Figure 6 shows the first and second non-dimensional singular vectors, columns of $\mathbf{E}^*$ in (23). They are the 3D features associated with the dominant nonlinear, mesoscale PE variability during the early spring of 1995. Interestingly, they mainly explain variations of the anticyclonic Ierapetra Eddy (centre near 34°N, 27°E), a feature not always observed in the area during recent years (Robinson and Malanotte-Rizzoli (Eds.) 1993). The two vectors explain 1.49% and 1.18%, respectively, of the total variance explained by the 400 vectors. The first vector identifies the main structure of the vortex (panels (1a) and (1b)). The variability variance induced by squeezing/shallowing, stretching/deepening, tilting, horizontal displacements and internal variations of the vortex are, on average, best explained by this PE pattern. The barotropic transport (ψ) component (panel (1a)) is slightly asymmetric, because of relief, density, diffusion and nonlinear effects. It is, for example, larger where the topography is deeper (see Fig. 8(a) hereafter) and has a shape which follows that of the local tracer fields (see Figs. 9 and 10 hereafter). The T structure is surface intensified (panel (1b)), with an uniform extremum from 50 to 250 m and a significant extension down to about 600 m. The S structure (panel (1b)), however, has a strong subsurface extremum within 250 m to 340 m depths, and one significant zero-crossing around 80 m. Near 1000 m the sign of T and S reverses, but the amplitudes below 1000 m are quite small (one tenth to one thirtieth the size of the upper-thermocline values). Where they are both large (panel (1b)), T and S are in phase, compensating each other in density. Horizontally (panel (1a)), when significant, both structures are almost axisymetric, with horizontal scales near 30 km and shapes close to Mexican hats. This is what they should be by construction, even though only 400 PE vectors are employed. Caution should nonetheless be exercised regarding the effects of this truncation, which are mainly limited to the smallest amplitudes. For example, consider the S component at 5 m (panel (1a)), which is relatively small (less than 15% of the S extrema). Logically, the largest amplitudes are near the vortex, but there are also lobes far from the vortex. These lobes are in part spurious, and have very low significance since their amplitudes

are at about 7% or less than the S extrema (1.7×10^{-2}) and the aim is 93% of the variance. The horizontal structure of the $\widehat{u}$ component is shown at 5 m in panel (1a) and cross-sections in the $\widehat{u}$ and $\widehat{v}$ components are given in panel (1b). If the vortex was a perfect cone or cylinder, if thermal-wind balance held and if the topography was flat, the horizontal structures of the internal velocities $\widehat{u}$ and $\widehat{v}$ would be lobes antisymmetric with respect to the x and y axes, respectively (e.g. Daley 1991). However, because of inertia (e.g. centrifugal force in the gradient-wind balance), vertical and horizontal viscosities, bottom stress and regional topography, this does not hold exactly, and the internal velocity patterns at 5 m are non-classic and tilted. For the Ierapetra Eddy, the local Rossby number Ro is in fact about 0.22 ± 0.1 (LLA98). Two nonlinear effects on the vectors of Fig. 6 are the tightening of the horizontal velocity gradients and the dissipation of the internal oscillations by mixing. Vertically (panel (1b)) $\widehat{u}$ and $\widehat{v}$ reverse sign at about 400 m. From ψ (panel (1a)), a vertical minimum in total velocity amplitude is expected near 600 m, below this reversal. The second vector (panels (2a) and (2b)) relates to eastward–westward displacements of the vortex (with the plotted, arbitrary sign, it is an eastward displacement). In the vertical (panel (2b)), the tracer and internal velocity patterns have similar properties to those of the first vector (surface intensified T, mid-depth S extrema, T and S in phase where they are both important, T and S changing sign near 1000 m and of small amplitudes below 1000 m, and $\widehat{u}$ and $\widehat{v}$ surface intensified, with zero-crossings near 400 m). The ψ component (panel (1a)) is again slightly asymmetric. It also contains a small-amplitude radiation pattern, in accord with the properties of moving vortices (e.g. Cushman-Roisin 1994). This radiation pattern has corresponding structures in the T, S, $\widehat{u}$ and $\widehat{v}$ components (panels (2a) and (2b)). At the surface (e.g. 5 m) the prevailing horizontal shapes of $\widehat{u}$ and $\widehat{v}$ are, nonetheless, close to these given by thermal-wind balance: a pattern in '45 degree shamrock' dominates for $\widehat{u}$, and a triple lobe, meridionally oriented pattern dominates for $\widehat{v}$ (not shown at 5 m).

Overall, the above properties agree with what is commonly observed about the vertical locations of the Modified Atlantic Water and Levantine Intermediate Water. They also show that warm and salty anticyclonic eddies in the eastern Mediterranean (Brenner 1993) can have properties (LER97) similar to those of Gulf Stream warm core rings (e.g. Robinson 1983). Finally, as exemplified above, even though the inertial, nonlinear and diffusion terms are smaller overall than the Coriolis and pressure gradient terms, they do matter, at least locally (presently, eddy viscosities only matter in coastal and bottom boundary layers, and for scales smaller than 20 km). For instance, if the nonlinear terms are turned off in the adjustment momentum integrations (LLA98), the resulting two dominant vectors differ from the present ones: they correspond to the deeper and much wider Rhodes Gyre (centre near 35.5°N, 28.5°E, above the Rhodes basin, see Fig. 8(a)), impinging on the Cretan–Rhodes Ridge, and thus radiating and interacting with topographic wave patterns along the northern escarpment (e.g. Smith 1983). Within the other vectors (LER97), several groups correspond to the Rhodes Gyre and again to the Ierapetra but also to the eastward Mid-Mediterranean Jet (meandering between 32°N and 34°N), westward Asia Minor Current (along the northern coastline), Mersa Matruh Gyre (centre near 32°N, 27.5°E) and West Cyprus Gyre (centre near 35.5°N, 31.5°E). The 50 dominant vectors explain 38.4% of the variance explained by the 400 vectors; the 100 dominant vectors 62.7%.

(iv) *Levantine covariance functions.* An estimate of 3D error covariance functions is illustrated by Fig. 7, considering a row of $\mathbf{B}^p$ (23), with $\mathbf{\Gamma} = \mathbf{I}$. The mesoscale variability covariance between the 5 m temperature at (33.8°N, 27.85°E) and the other state variables is shown. Looking at the tracer auto-covariances (panels (a) to (b)), the

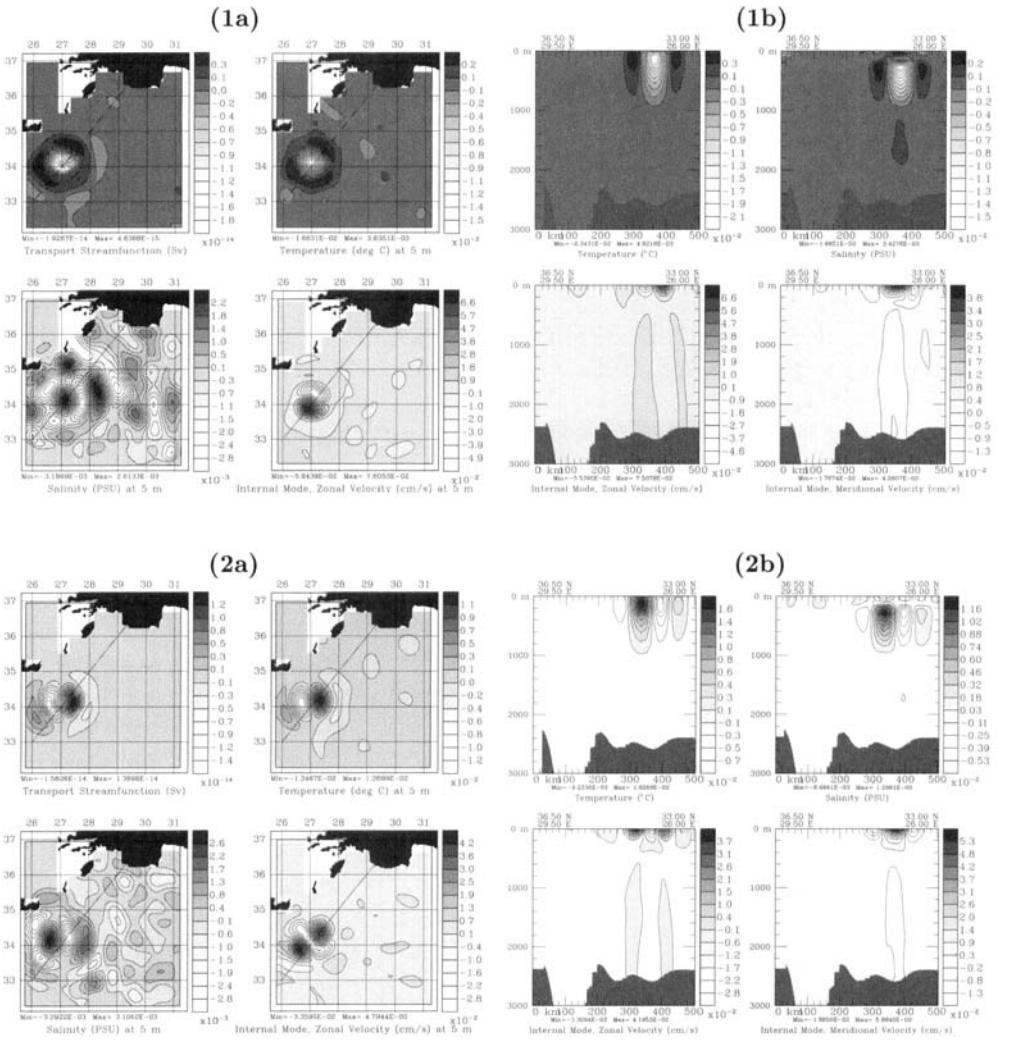

Figure 6. Multivariate eigenvectors of a mesoscale error covariance for 16 March 1995, in the northwest Levantine, as estimated by singular-value decomposition (SVD) of a normalized ensemble of PE adjusted variability samples. The panel number is the eigenvector number, with index (a) for the surface (5 m) level, and (b) for a vertical cross-section parallel to the Cretan Arc, above the Cretan–Rhodes Ridge (the section position is drawn on panel (a)). The estimation is based on Eqs. (14)–(23), as summarized by the flow diagram of Fig. 4. The size of the covariance is [218943 × 218943]. The grid has a horizontal spacing of 10 km and 20 vertical levels. For further discussion see the text.

horizontal Mexican hat and vertical EOF structures are well represented. As expected for the region and period, the surface T is positively correlated to deeper T and to S, with partially compensating effects in density (panel (b)). Magnitudes are almost constant with depth in the shallow mixed layer (0 to 30 m), the T–S cross-covariance has a subsurface extremum at about 300 m, and significant correlations exist down to about 400 to 600 m. Deeper, T–T and T–S decay and have a zero-crossing near 1200 m. Beyond this depth, their horizontal patterns are negative Mexican hats, but of amplitudes smaller than 5% of the volume extrema. The T–ψ cross-covariance (panel (a)) is also close to a Mexican hat, even though the rank of $\mathbf{B}^p$ is more than 500 times smaller

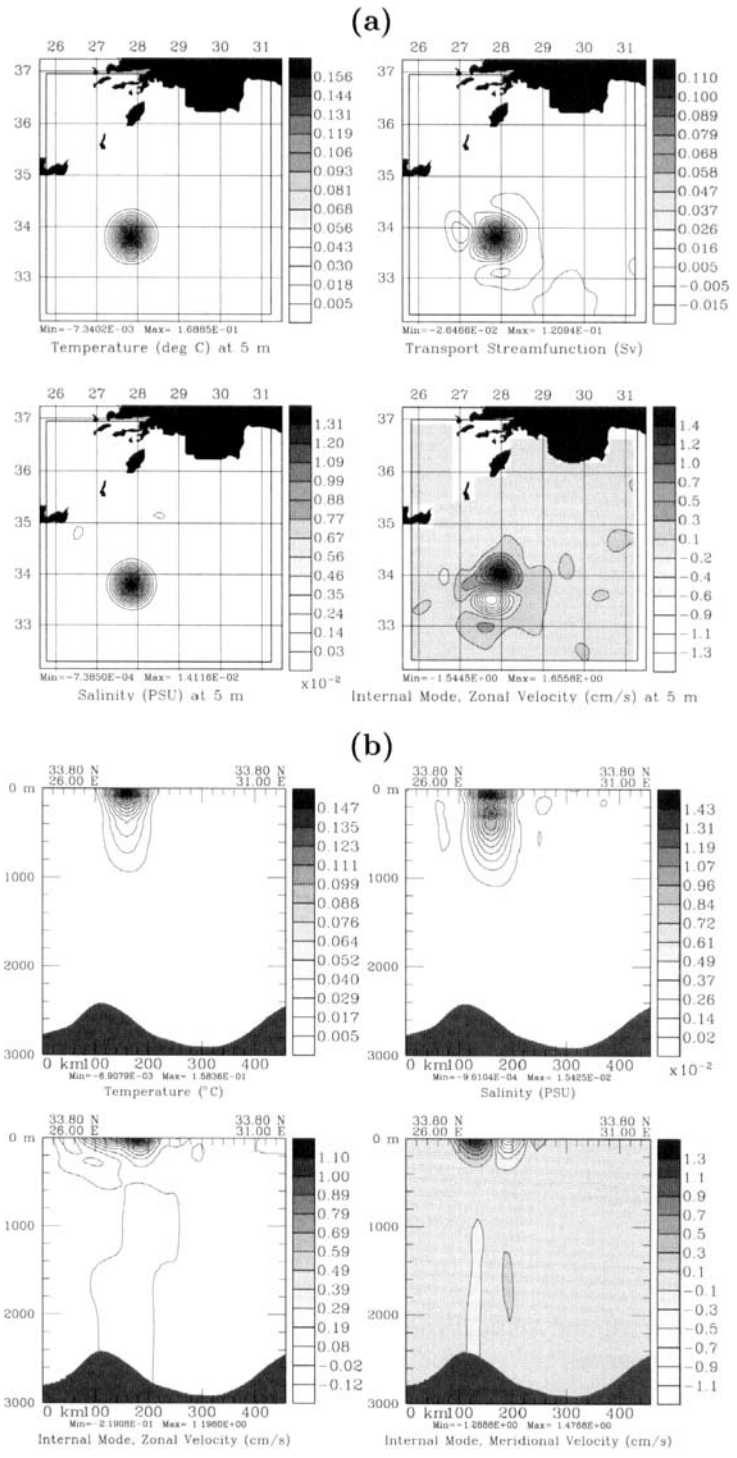

Figure 7. Estimate of the dimensional mesoscale PE covariance function between the 5 m temperature at (33.80°N, 27.85°E) and the other state variables (shown beneath each frame). Panel (a) shows values at 5 m, and panel (b) cross-sections along 33.80°N. Elements of a row of $\mathbf{B}^p$ for 16 March 1995 are shown. See text for further details.

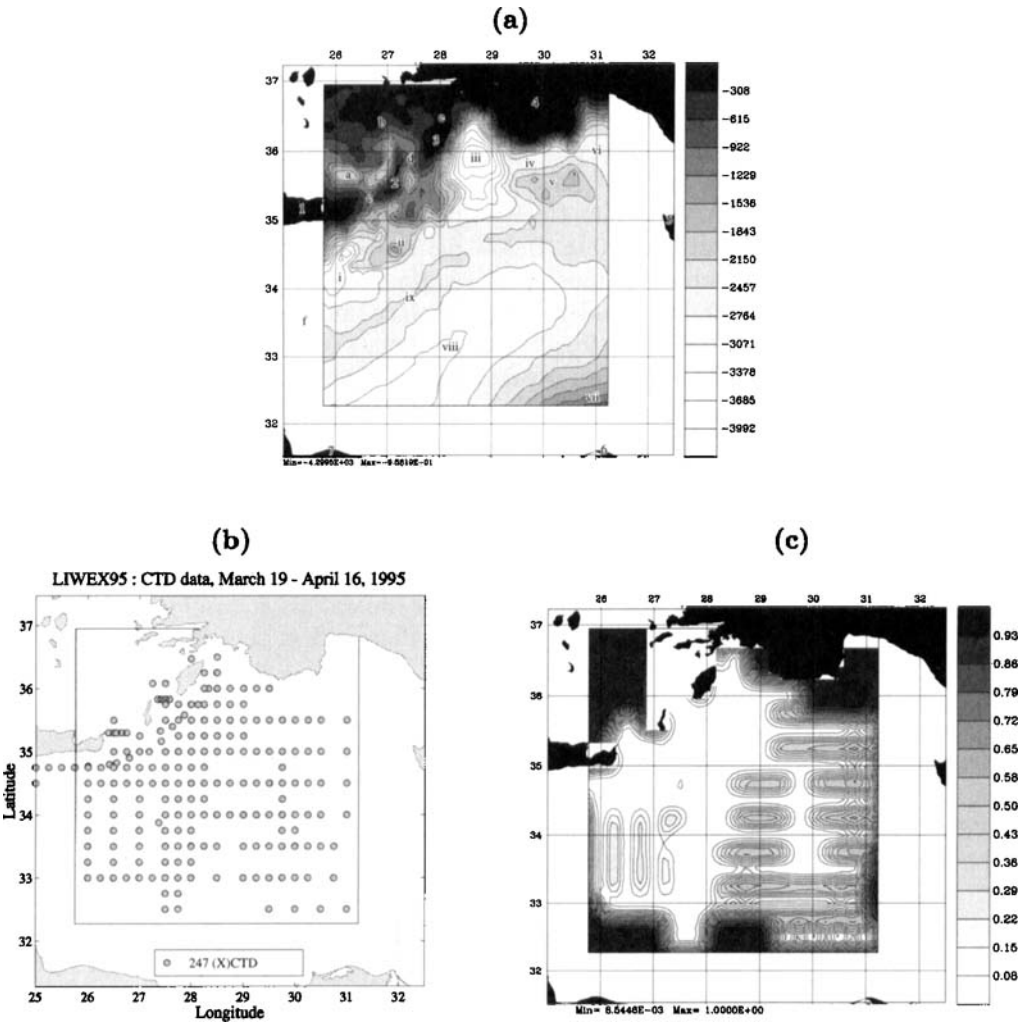

Figure 8. Panel (a) shows the geography and etopo5 bathymetry. The Arabic numbers indicate: 1, Crete; 2, Karpathos; 3, Rhodes; 4, Turkey; 5, Cyprus; 6, Egypt; and 7, Libya. The letters indicate: a, the Cretan Sea; b, Aegean Sea; c, Kasos Strait; d, Karpathos Strait; e, Rhodes Strait; and f, the Cretan Passage to the Ionian Sea. The Roman numbers indicate the sea topography: i, the Strabo Trench; ii, Cretan–Rhodes Ridge; iii, Rhodes basin; iv, Finike Trough; v, Anaximander Seamounts; vi, Antalya basin; vii, Nile cone; viii, Herodotus Abyssal Plain; and ix, Mid-Mediterranean Ridge. Panel (b) shows the coordinates of the 247 CTD (conductivity-temperature-depth) profiles collected during March 19–April 16, 236 of which are in the numerical domain and form $\mathbf{y}^o$ (appendix A) in the mesoscale multivariate three-dimensional (3D) analysis of section 5. Panel (c) shows the surface (5 m) values of the *a posteriori* mesoscale error variance of the corresponding temperature field, as estimated by the univariate 2D scheme. The values are normalized (0–1) and computed using a Mexican hat form covariance function, with $l_a^x = l_a^y = 60$ km and $l_b^x = l_b^y = 30$ km, and an *a priori* mesoscale error variance of uniform amplitude equal to 1.

than that of **B**. It is, however, slightly asymmetric: it is elongated in the north-east direction, presents radiative patterns of small amplitudes, and has an enhanced cyclonic correlation at the location of the Ierapetra (near 34°N, 27°E). These facts are in accord with the regional topography, density field and Mid-Mediterranean Jet which locally is a relatively broad flow to the north-east. The horizontal T–$\widehat{u}$ and T–$\widehat{v}$ cross-covariances (panels (a) and (b)) also differ from the classic double-lobe structures, antisymmetric

with respect to the x and y axis, respectively. One reason is that the inertial terms at (33.80°N, 27.85°E) locally matter in the surface layers, tilting the classic structures (see LLA98). Another is the nonlinear effects which tighten the horizontal gradients of the T-velocity cross-covariances. They also locally extend these cross-covariances to the Ierapetra and lead to small but non-negligible radiative lobes (see panels (a) and (b)). In the vertical (panel (b)) T–$\widehat{u}$ and T–$\widehat{v}$ have a zero-crossing within 500 to 600 m. Using the T–ψ pattern (panel (a)) the total velocity variability near (33.80°N, 27.85°E) thus has a vertical minimum at about 700 m. Below the zero-crossings (panel (b)), the T–$\widehat{u}$ and T–$\widehat{v}$ structures are opposite to these of the surface layers, but have much smaller amplitudes with extrema of, at most, 10% of the surface ones.

In relation to mapping, in particular to surface data extension (e.g. Rienecker and Ademec 1995; Stammer 1997), a 5 m observation of T at (33.80°N, 27.85°E) would influence the 3D fields of T, S, ψ, $\widehat{u}$ and $\widehat{v}$ according to the structures of Fig. 7. Such structures would be on the left-hand-side of the gain $\mathbf{K}^p$ (appendix A), extrapolating the data residuals onto $\mathbf{x}^b$. It is important to note that even though the eigenvectors are global (e.g. Figs. 1, 5, 6), if their number is sufficient, the spurious effects of the truncation, which could be manifest as remote influences of the local data, are by construction insignificant. The multivariate 3D covariance functions computed from $\mathbf{B}^p$ (23) propagate the significant data information over limited distances. For example, in Fig. 7 the possibly undesired remote effects of T data at 5 m (e.g. see panel (a) for $T-\psi$ and $T-\widehat{u}$, and panel (b) for effects at greater depths) would have covariance amplitudes that are less than about 7% of the horizontal covariance extrema. This result agrees with the 'more than 93%' convergence criterion (22) used.

As mentioned in section 2, the above examples briefly show that the decomposition of the variability is useful for verifying or exploring the dominant features and processes of a given region (e.g. Lermusiaux 1999a,b). The adequacy of error subspace estimates and corresponding covariance functions must, in fact, be established. Two verification approaches have been illustrated here: the global evaluation of the dominant error eigendecompositions (Figs. 5 and 6), and the local assessments of covariance function fields (Fig. 7), analogous to a 'single-observation analysis'. In our real-time ocean operations (e.g. Robinson *et al.* 1998b) both approaches are employed on a regular basis, as well as techniques similar to these of Wahba and Wendelberger (1980). Such sensitivity computations are also useful to estimate the dimension of the error subspace that is adequate for specific purposes. In practice, pre-exercise validation experiments can test the feasibility of an error subspace approach, design Observation System Simulation Experiments, and improve the error weights of cheaper, less-optimal estimation of methods (LER97, LLA98).

5. THREE-DIMENSIONAL MULTIVARIATE ANALYSIS

The algorithms for a first estimate of the a priori error subspace (sections 3 and 4) are now employed for a global, multiscale and multivariate 3D mapping (appendix A, (A.2a) and (A.2c)). Results are compared to those of the univariate 2D scheme of HOPS (section 3(a)). The example is a sub-basin-scale to mesoscale analysis of PE fields in the Levantine Sea (eastern Mediterranean). For this region, sensitivity and geophysical studies based on the algorithms of Figs. 2 to 4 are carried out in LLA98, and process studies in Roether *et al.* (1996), Malanotte-Rizzoli *et al.* (1996) and LER97. The model domain and topography are shown on Fig. 8(a). The data $\mathbf{y}^o$ (appendix A, (A.1b)) are 236 CTD profiles gathered during 19 March and 16 April by the POEM group (Fig. 8(b)). The univariate 2D scheme assumes that two independent scales are present. For each

scale it carries out horizontal analyses of the T and S data, and computes total velocity fields in geostrophic balance with the gridded tracers, assuming a level of no motion at 600 m. When compared to other levels, this choice led to the smallest horizontal and vertical adjustments in subsequent forecasts, in accord with results in section 4(c) and previous studies (e.g. Robinson and Golnaraghi 1993). To ensure fair comparisons with this analysis, the present multivariate 3D scheme follows Fig. 4, using the successive-correction approach (sections 4(a) and (b)), with two scales ($w = 1, 2$).

(*a*) *Fields*

For each scheme, the sub-basin-scale ($w = 1$) PE fields are first estimated for 16 March. The data $\mathbf{y}^o$ utilized in this first-stage are 512 gridded data of the MODB (Med5) winter climatology for the region (Brankart, personal communication), and 49 CTD and 96 salted AXBT (aircraft-deployed expendable bathythermograph) profiles which originated from the *in situ* CTD and AXBT measurements gathered during 10–18 February and 15 March 1995, respectively. The resulting two sets of sub-basin-scale PE estimates (not shown) constitute the a priori fields $\mathbf{x}^b$ for the subsequent univariate 2D and multivariate 3D mesoscale analyses ($w = 2$), respectively.

The 236 profiles to be gridded during this mesoscale stage lead to 9440 T, S scalar observations on hybrid levels, i.e. the number of data-points (dimension of $\mathbf{y}^o$). In the univariate 2D scheme, the a priori T and S covariances were specified using a Mexican-hat fit on basin average to zero-crossings $l_a^x = l_a^y = 60$ km, and decay-scales $l_b^x = l_b^y = 30$ km (section 3(a)). The data error covariance at data points was assumed to be diagonal, with horizontally uniform non-dimensional variance $r^* = 0.03$. The corresponding estimate of the surface (5 m) mesoscale coverage is shown on Fig. 8(c). In the multivariate 3D scheme, the construction of the a priori error covariance decomposition for this stage $w = 2$ was presented in section 4(c). For the mesoscale vertical tracer EOFs, it employed all 289 CTD profiles gathered during Feb. 10 and April 16; for the horizontal tracer covariance, the zero-crossings and decay-scales of the univariate 2D scheme; and, for the non-observed velocity portions, an ensemble of nonlinear adjustment PE integrations. Based on (22), $p = 400$ eigenvectors sufficed for 93% of the error variance, with $\mathbf{\Gamma} = \mathbf{I}$. The results of the two schemes are compared in Figs. 9 to 12, variable by variable.

(i) *Tracer fields.* Figure 9 shows the T and S mesoscale univariate 2D analyses at four horizontal levels. Figure 10 shows the same fields, but corresponding to the T, S components of $\mathbf{x}^a$ obtained by multivariate 3D analysis (appendix A (A.2a) to (A.2c)). For the large mesoscales to the sub-basin-scales, the two schemes yield tracer fields that are very similar, with differences within data error bounds. Without another intertwined data array, the better of the two estimates cannot be determined at those scales. The largest discrepancies are in the sub-mesoscales to small mesoscales. This is especially true within and just below the surface mixed layer, where the late winter atmospheric forcings and internal instabilities lead to an increase of small-scale potential and kinetic energies (see Figs. 9 and 10, at 5 and 105 m, e.g. near the Asia Minor Current, Ierapetra and Mersa Matruh Gyre). There are two main technical reasons for these discrepancies. First, for most of the domain, the data (Fig. 8(b)) do not resolve these small scales. The multivariate 3D scheme (Fig. 10), which takes into account possible vertical correlations as well as correlations between the T and S fields, is shown to be less sensitive to such data noise than the univariate 2D scheme (Fig. 9). Second, this property is accentuated because, even though the dominant 400 PE eigenvectors were estimated to explain 93% of the normalized PE error variance (section 4(c)), the

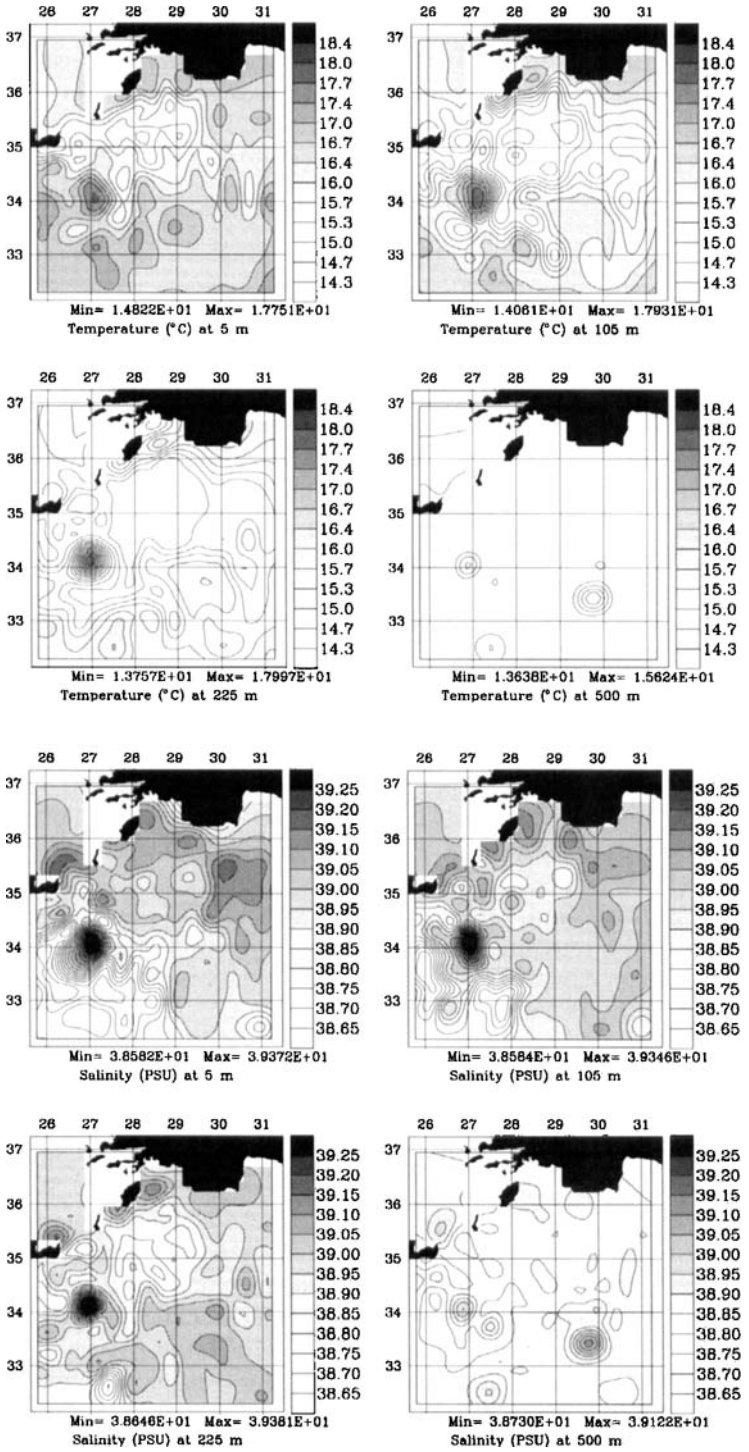

Figure 9. Harvard Ocean Prediction System (HOPS) mesoscale (second-stage) univariate two-dimensional temperature and salinity analyses, at 5, 105, 225 and 500 metres. The first-stage fields were sub-basin-scale univariate two-dimensional analyses.

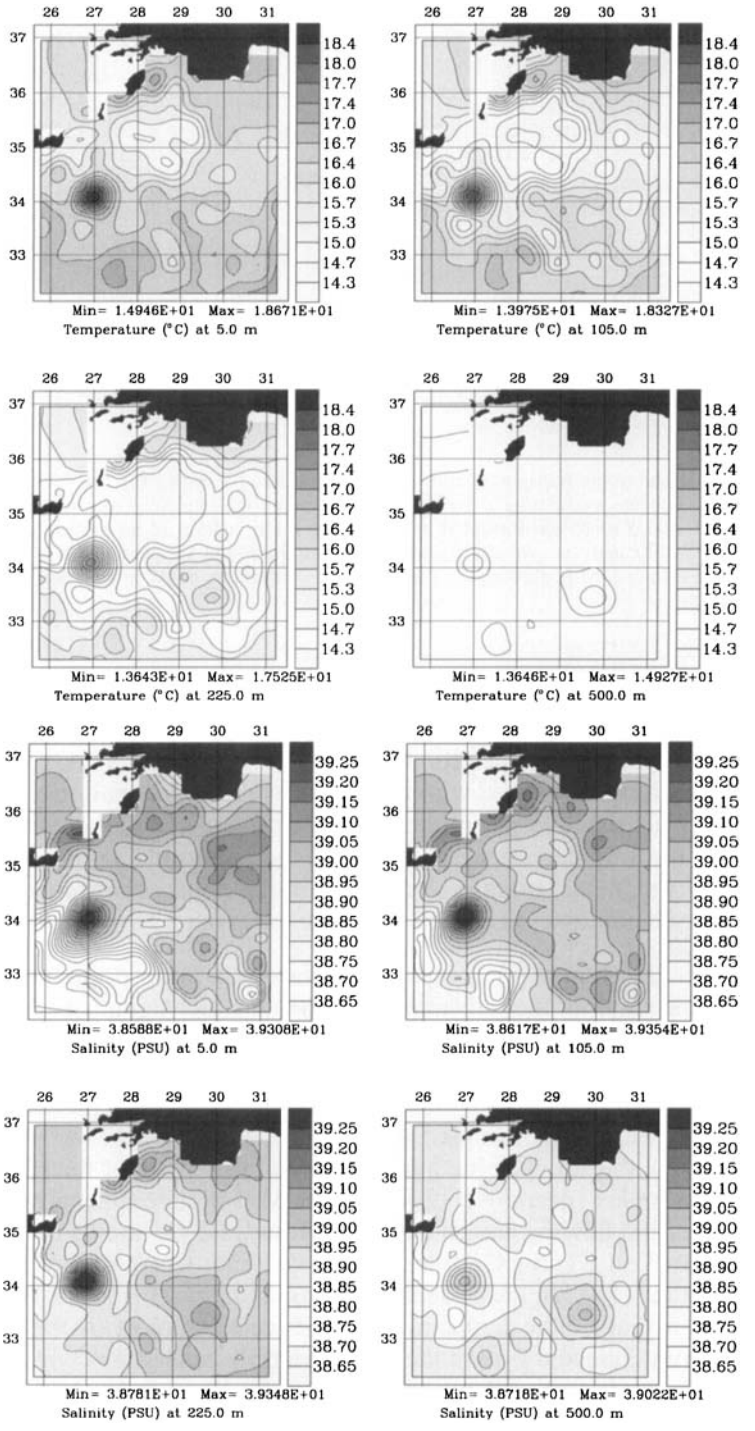

Figure 10. Mesoscale T and S fields resulting from the PE multivariate 3D analysis. The dominant 400 vectors of $\mathbf{B}^p$ in equation (23) are used. The levels shown and all scalings are as on Fig. 9. The first-stage fields resulted from a sub-basin-scale multivariate 3D analysis.

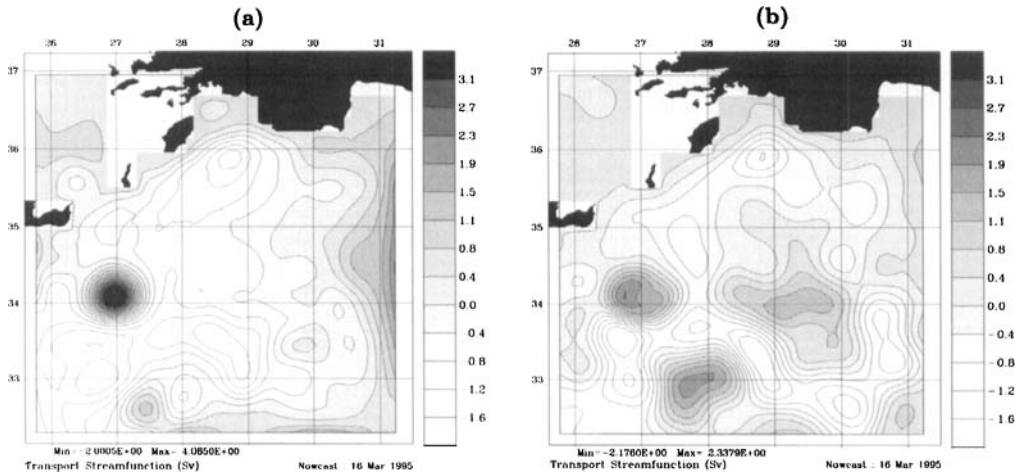

Figure 11. Mesoscale barotropic transport stream-function estimates (ψ in Sv). (a) The result of the univariate two-dimensional (2D) scheme, computing ψ by vertically averaging the total velocities in geostrophic balance with the horizontal T and S analyses illustrated by Fig. 9, assuming a level of no motion at 600 m. (b) The result of the multivariate 3D analysis, globally estimating ψ from the T and S data (Fig. 8(b)), in accord with the dominant SVD of the PE adjusted perturbations. The dominant 400 vectors of $\mathbf{B}^p$ in equation (23) are used. See text for details.

corresponding 400 tracer eigenvectors used in the adjustment PE integrations (19) to (22) only explained 72% of the normalized 3D tracer variance†, $\mathrm{tr}[\mathbf{\Pi}_{\mathrm{trc}}]$ in (18). Hence, using an $\mathbf{E}$ of 400 columns (appendix A, (A.2a) to (A.2c)) eliminates about 21% of the tracer variance, here associated with the sub-mesoscales to small mesoscales of the T and S data. For an example where the two reasons stated above apply, consider the Ierapetra, where the surface temperature signal is stronger on Fig. 9 than on Fig. 10 because the salinity signal is strong at all depths. Similar remarks hold for other features, e.g. small mesoscale processes near the surface mixed layer.

(ii) *Transport fields.* Figure 11 compares the barotropic transport stream functions. Panel (a) shows the result of the univariate 2D scheme which computes ψ by vertically averaging the total velocities in thermal-wind balance with the horizontal analyses of the tracer fields (Fig. 9), assuming a level of no motion at 600 m. Panel (b) shows the ψ component of $\mathbf{x}^a$ obtained using appendix A, (A.2a) to (A.2c). In Fig. 11(b) one directly distinguishes the sampled lobe of the anticyclonic Mersa Matruh Gyre (near 32°N, 27.5°E), the anticyclonic gyre forming at the centre of the Mersa Matruh–Shikmona Gyre complex (near 34°N, 29.5°E), the Mid-Mediterranean Jet meandering around these gyres and flowing east, the cyclonic Rhodes Gyre and its multiple centres, the weak cyclonic West Cyprus Gyre and the Asia Minor Current flowing west along the northern coastline. The flat level of reference computation (Fig. 11(a)) overestimates the barotropic transport of the Ierapetra, while it weakens most of the West Cyprus Gyre, Mersa Matruh Gyre and gyre of the Mersa Matruh–Shikmona Gyre complex. The horizontal transport is, in fact, difficult to estimate (e.g. Wunsch 1996). The simplest is to guess a reference velocity (e.g. level of no motion). Other approaches are the beta-spiral (Schott and Stommel 1978), box inverse (Wunsch 1978) and Bernouilli (Killworth

† PE convergence in (23) can occur without complete tracer convergence. This is accentuated here because numerical errors in the PE model utilized lead to overestimation of the covariances involving total velocities, and because the criterion (22) gives an upper bound of the true percentage of variance explained.

1986) methods. Killworth and Bigg (1988) compared the merits of these schemes using simulated data. Presently, it is the a priori missing PE variability and the properties of $\mathbf{y}^o$ (Fig. 8(b)) which determine what ψ is and what ψ uncertainty remains a posteriori (section 5(b)).

(iii) *Internal velocity fields.* Figure 12 compares the surface internal velocities. Panel (a) shows the results of the univariate 2D scheme which estimates $\hat{u}$ and $\hat{v}$ by computing the zero vertical-average velocities in thermal-wind balance with the horizontal analyses of the tracer fields (Fig. 9). Panel (b) shows the surface $\hat{u}$ and $\hat{v}$ components of $\mathbf{x}^a$ obtained using appendix A, (A.2a) to (A.2c). Panels (a) and (b) differ, mainly because of the differences between the T and S fields of Figs. 9 and 10 and because of ageostrophic motions. At scales larger than the small mesoscales, the T and S fields of Figs. 9 and 10 are almost identical, and the geostrophically balanced velocities of Fig. 12(a) are, for several features, similar to the globally adjusted velocities of Fig. 12(b), especially for the zonal component. However, even at these scales, there are regions for which geostrophy is not adequate because of: (i) surface, bottom and coastal non-adiabatic boundary layer processes; (ii) translating mesoscale meanders and waves; (iii) jets locally strongly anisotropic, for which inertia (mixing, stirring, etc.) becomes important; and (iv) sustained geostrophic adjustments. For example, comparing panels (a) and (b), the Asia Minor Current and its density front involve processes (i), (iii) and (iv); above the Anaximander Seamounts (about 35.5°N, 30.25°E), it is mainly (i) and (iv) that occur; and along the paths of the Mid-Mediterranean Jet around the Mersa Matruh–Shikmona Gyre complex, it is (ii) and (iii), e.g. see the $\hat{v}$ pattern. At sub-mesoscales and small mesoscales, nonlinear terms are important (Cushman-Roisin 1994): at the Ierapetra, the gradient-wind equation is necessary to explain surface velocities within 60 to 90 cm s^{-1}, as in panel (b). At these smaller scales, such effects and the differences in T and S of Fig. 9 and Fig. 10, explain the discrepancies between panels (a) and (b).

(*b*) *Errors*

(i) *A priori errors.* For conciseness, the discussion is limited to the stage $w = 2$. The univariate 2D scheme only provides error variance fields (see Fig. 8(c)). For the multivariate 3D scheme, the a priori error amplitude is that of the mesoscale variability, i.e. $\mathbf{\Gamma} = \mathbf{I}$ in (23). The eigenvectors (Fig. 6) and a row (Fig. 7) of the corresponding $\mathbf{B}^p$ were already shown in section 4(c). The a priori error variance is the diagonal of $\mathbf{B}^p$. By construction, for the tracers, error amplitudes are horizontally uniform in the data domain: for example, at 5 m, the $T(S)$ error standard deviation is 0.41 degC (0.083 PSU), while at 500 m, it is 0.19 degC (0.042 PSU). For the total velocities, amplitudes are non-homogeneous. At the surface, the largest error standard deviations are: near the Ierapetra, with a ψ error of 0.9 Sv and an internal velocity error of 23 cm s^{-1}; near the Mersa Matruh Gyre and inflow of the Mid-Mediterranean Jet, where the maxima of the ψ and $\hat{u}$, $\hat{v}$ errors are 0.84 Sv and 18 cm s^{-1}; and above the Rhodes basin, around which the largest ψ and $\hat{u}$, $\hat{v}$ errors are 0.9 Sv and 16 cm s^{-1}, respectively. The average surface error of the internal velocity is 14 cm s^{-1}. For ψ, the average error is 0.7 Sv. At 500 m depths, the largest $\hat{u}$, $\hat{v}$ errors are at the Kasos and Karpathos Straits (1.8 cm s^{-1}), above the Rhodes basin (1.5 cm s^{-1}), and at the deep levels of the Ierapetra and anticyclonic gyre of the Mersa Matruh–Shikmona Gyre complex (1.0 cm s^{-1}). The 500 m average is 0.7 cm s^{-1}. These results agree with previous studies (e.g. Robinson and Malanotte-Rizzoli 1993).

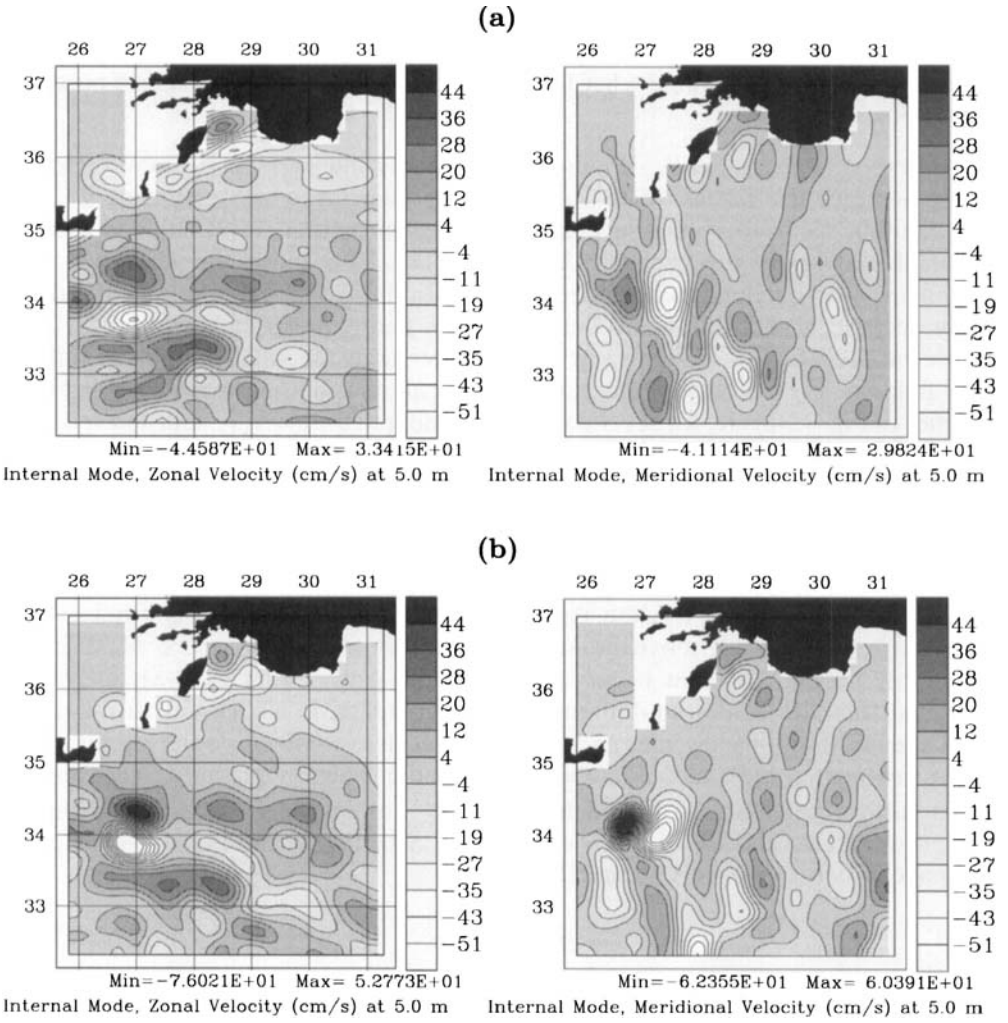

Figure 12. Mesoscale, zero-vertical mean, internal velocity estimates $(\widehat{u}, \widehat{v})$ at 5 m. Panel (a) shows the result of the univariate 2D scheme, computing $\widehat{u}, \widehat{v}$ in geostrophic balance with the horizontal 2D T and S analyses illustrated by Fig. 9. Panel (b) shows the result of the multivariate 3D analysis, globally estimating $\widehat{u}, \widehat{v}$ from the T and S data (Fig. 8(b)), in accord with the dominant SVD of the PE adjusted perturbations. The dominant 400 vectors of $\mathbf{B}^p$ in equation (23) are used. See text for details.

(ii) *A posteriori errors.* Figures 13 and 14 illustrate the a posteriori mesoscale error covariance (appendix A, (A.2b) and (A.2c)). The error standard deviations (Fig. 13) associate an uncertainty to each feature of the gridded PE fields shown in Fig. 10, Fig. 11(b) and Fig. 12(b). In the north-east corner where there is no data influence (Fig. 8(b)), a posteriori values are as a priori: this holds for all depths and variables. The only information in this corner is of sub-basin-scales. At the surface (panel (a)), the a posteriori T and S errors are overall similar in shape to Fig. 8(c). This confirms that, for error variance estimates, a few eigenvectors suffice[†]. Nonetheless, there are

[†] Here, 400 eigenvectors were used. For practical purposes, if an efficient error variance estimate (diagonal of $\mathbf{E}^a \mathbf{\Pi}^a \mathbf{E}^{a\mathsf{T}}$) had been our only desire, around 50 eigenvectors would have sufficed.

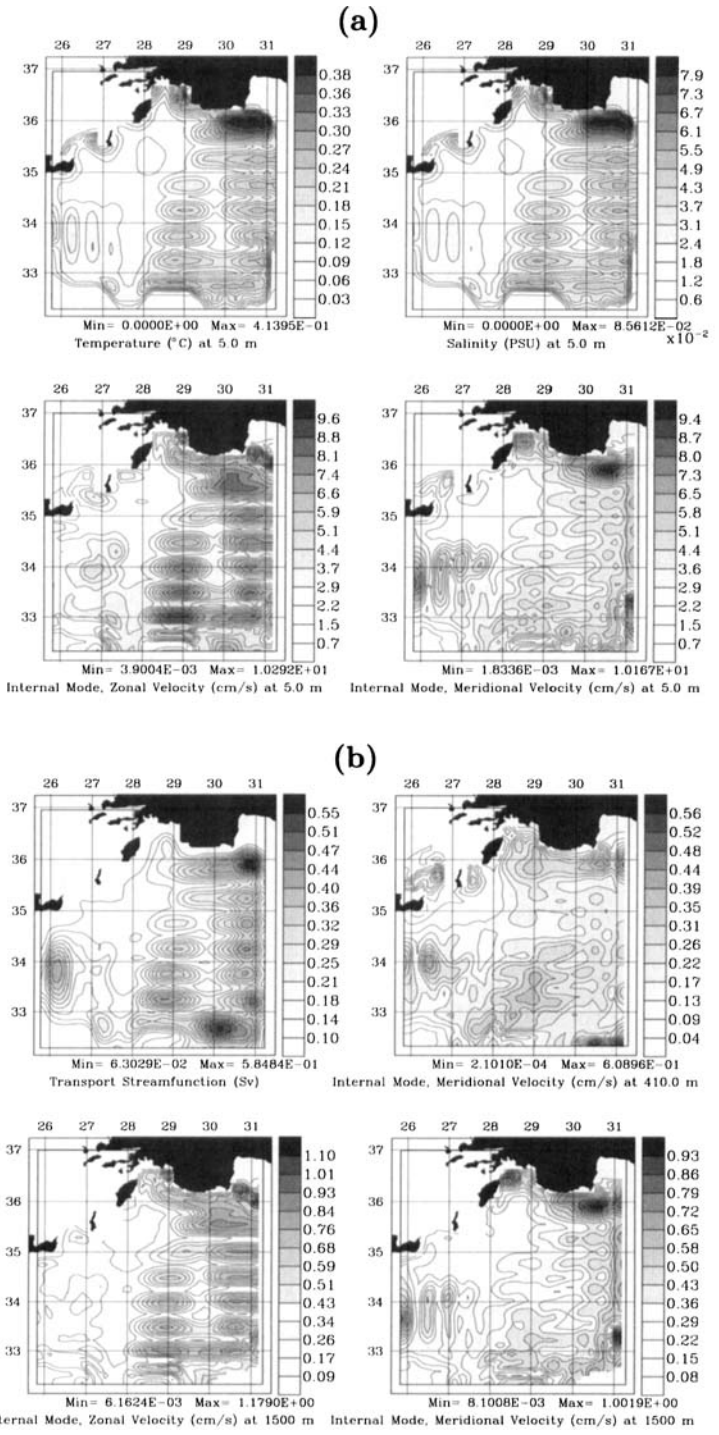

Figure 13. Multivariate standard deviation of the a posteriori mesoscale error covariance. The square root of the diagonal of $\mathbf{E}^a \mathbf{\Pi}^a \mathbf{E}^{a^T}$ is illustrated. (a) Surface (5 m) root-mean-square error (r.m.s.e.) values of four PE fields. (b) The ψ r.m.s.e., the $\widehat{v}$ r.m.s.e. at 410 m, and the $\widehat{u}$ and $\widehat{v}$ r.m.s.e. at 1500 m. Note that the a priori error was weighted as in Fig. 1. Further discussions are in the text.

some differences. These are due to the a priori correlations between T and S (e.g. Figs. 6 and 7); to the influence of sub-surface data which is non-uniform, in part because profiles have different lengths; and to the depth-dependent variance of the a priori T and S errors. The surface a posteriori errors of the internal velocities (panel (a)) are not similar to Fig. 8(c), for several reasons. First, the a priori covariance functions involving velocity fields are non-homogeneous. Second, the PE velocities are connected to T and S via gradients of density. This property enters $\mathbf{B}^p$ (23), the gain $\mathbf{K}^p$ and thus $\mathbf{E}^a\mathbf{\Pi}^a\mathbf{E}^{a^T}$ (appendix A). Presently, for the internal velocities orthogonal to ship tracks, along which the data resolution is usually a 1/4 of a degree (Fig. 8(b)), the uncertainty is low and almost uniform, mainly because: (i) the higher data-resolution along tracks leads to a good estimate of mesoscale density gradients, hence velocities, across tracks; (ii) the often longer correlations of velocity components along their direction compensate for the lower data-resolution across tracks. For $\widehat{v}$, these conditions clearly hold east of 28°E, where errors are almost uniform and the $\widehat{v}$ field is thus well constrained. For $\widehat{u}$, they hold west of 27.5°E, between 33°N and 34.5°N. At the Ierapetra, mixing leads to an almost uniform surface a posteriori $\widehat{u}$ error of about 6 cm sec^{-1} (total velocity error at 5 m is there about 9 cm sec^{-1}). For velocities along tracks, in several locations 1/2 of a degree apart (Fig. 8(b)), the opposite conditions occur (low data-resolution across tracks, longer velocity correlations along tracks): the velocity errors are thus low in between tracks, but high on the tracks, where they are still close to their a priori values. For the $\widehat{v}$ errors, this occurs west of 27.5°E, between 33°N and 34.5°N. For the $\widehat{u}$ errors, it holds east of 28°E.

Panel (b) of Fig. 13 shows the a posteriori error standard deviations of ψ, of $\widehat{v}$ at 410 m, and of $\widehat{u}$ and $\widehat{v}$ at 1500 m. The ψ field is connected to T and S via the vertical integration of buoyancy. Because of this proportionality (vs. gradients for $\widehat{u}$ and $\widehat{v}$), the pattern of the ψ error is somewhat similar to that of the tracer errors (panel (a)). A multivariate result is that 400 vectors here suffice to ensure this expected fact. The main reasons for differences include the non-uniform a priori ψ errors (see above), the varying lengths of the data profiles and, in a lesser role, the multivariate T and S correlations. The first two effects are observed at several locations. The uncertainty of the Mid-Mediterranean Jet inflow (west of 27°E) is high over a large area because it was already so a priori, and because all profiles south of 34.5°N and west of 27°E are only about 1000 m in length, even though the average depth there is around 2500 m. The deep flow is not constrained, hence the ψ uncertainty. For the Mid-Mediterranean Jet outflow, in the south-east corner, the data-resolution is too low (1/2 degree). In the north-east corner, data are lacking altogether and the ψ errors near the Asia Minor Current inflow (Fig. 11(b)) are as a priori.

The $\widehat{v}$ a posteriori errors at 410 m (panel (b)) clearly show the influence of the non-uniform and 3D properties of the a priori missing PE variability. At the Ierapetra (Fig. 6), 410 m is near the depth of minimum internal velocity ($\leqslant$4 cm sec^{-1}, see LLA98), where thermal-wind balance does not dominate. Errors are thus small and of different shapes than at 5 m (panel (a)). In the region of the Mersa Matruh–Shikmona Gyre complex (south of 34.5°N, east of 28°E), 410 m is near the local bottom of the main thermocline. The dominant mesoscale variabilities, hence the error pattern, thus also differ from that in the surface. Along the Asia Minor Current such variability effects, as well as boundary layer processes, influence the error. In the south-east corner, the higher errors are due to lack of data and open boundary condition uncertainties. At 1500 m (panel (b)), most internal velocities are opposite to their surface values by constraint, but of much smaller amplitudes (e.g. Fig. 6). These properties are reflected in the $\widehat{u}$ and $\widehat{v}$ errors: e.g. even though several profiles are shallower than 1500 m, the zero vertical-average constraint

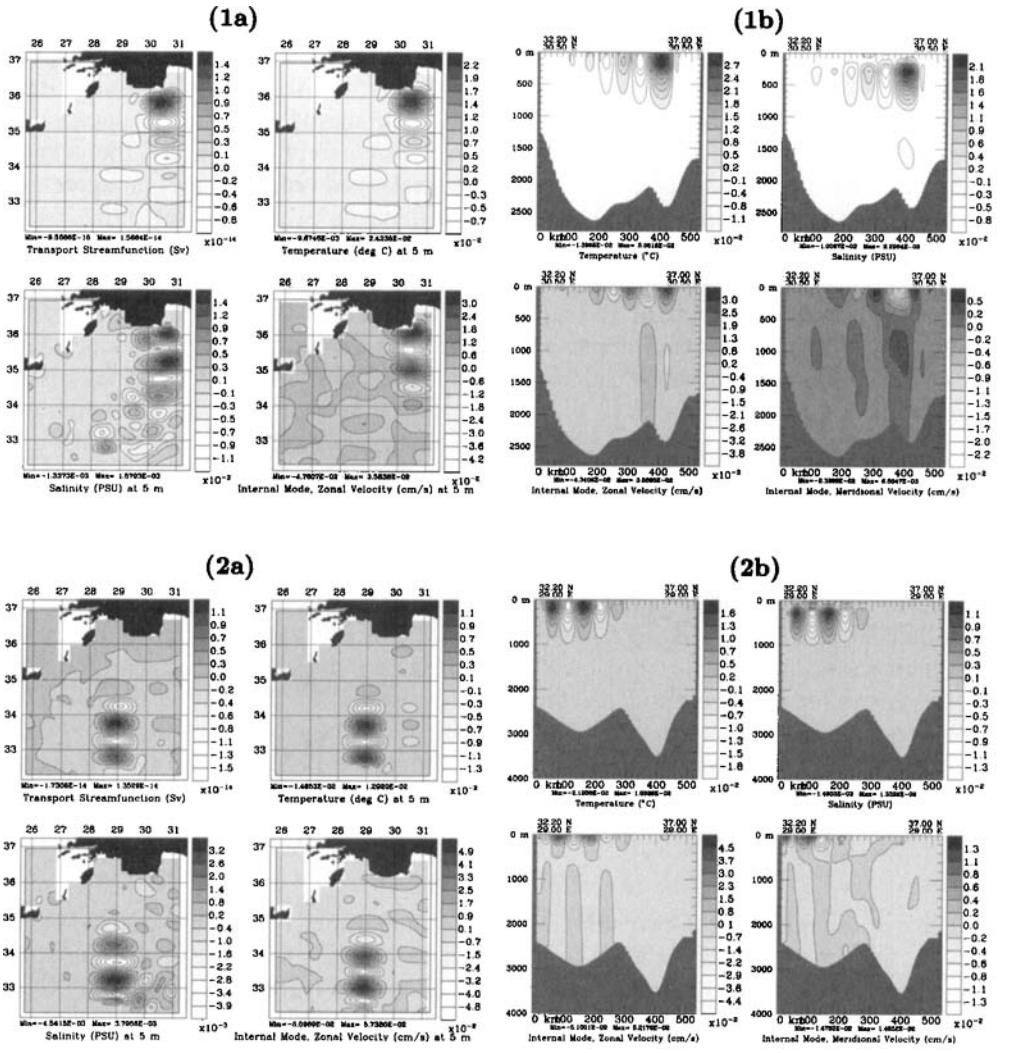

Figure 14. Multivariate eigenvectors of the a posteriori mesoscale error covariance. The panel number is the eigenvector number, with index (a) for the surface (5 m) level, and (b) for a vertical cross-section along an axis of large amplitudes (30.50°E for vector 1, 29°E for vector 2). The size of the covariance is [218943 × 218943]. Further discussions are in the text.

allows $\widehat{u}$ and $\widehat{v}$ to be relatively well estimated at depth. The T and S error maps near 1500 m are similar overall to those at 5 m (panel (a)), except for the influences of the varying length of the profiles and of the multivariate correlations, non-uniform in the vertical.

Figure 14 shows the first and second non-dimensional a posteriori error singular vectors, columns of $\mathbf{E}^{a^*}$. The first vector accounts for 4.25% of the a posteriori error variance; the second for 4.05%. By comparison with the a priori vectors (Fig. 6), the data $\mathbf{y}^o$ has reorganized the variance in the error subspace, leading to very different dominant a posteriori vectors. The first vector (panels 1(a) and (b)) is associated with the lack of data in the north-east corner (north of 35.5°N, east of 29.5°E) and low data-resolution along 30.75°E (Fig. 8(b)). In the no-data corner, errors are as a priori; they

correspond to unknown mesoscale variability. This is confirmed, for example, by the intensified surface patterns in T, subsurface maxima in S and first baroclinic $\widehat{u}$ and $\widehat{v}$ on panel (1(b)). South of that, in the vicinity of 30.75°E, the large mesoscales have been corrected and the dominant errors are at small mesoscales. In the open ocean (panels 1(a) and (b)), the ψ pattern is in phase with those of the tracers, the $\widehat{u}$ and $\widehat{v}$ patterns are often close to phase quadrature (thermal-wind balance), etc., in agreement with PE-adjusted variability. However, except for these facts, the first vector has no 3D dynamical meaning since a detailed study of its structures (LLA98) shows that they are due to data. The second vector (panels 2(a) and (b)) correspond to the low data-resolution patches centred on 29°E. Its properties are similar to these of the first vector. In summary, the locations and depths of the profiles dictate the overall shape of the patterns, and non-homogeneous, small mesoscale variability dictates the relationships between these patterns. Similar facts hold for the other vectors, some of which also account for smaller-scale patterns. The 50 dominant a posteriori vectors explain 71.5% of the error variance explained by the 400 vectors; the 100 dominant vectors 84.5%.

A posteriori errors are not uniform in oceanography: they are not everywhere in geostrophic balance, they depend on the complexities of the variability and are strongly influenced by the non-homogeneous data types and coverages. The a posteriori error covariances are useful for sampling-design. For example, in mapping a buoyancy controlled jet (e.g. Mid-Mediterranean Jet, Asia Minor Current) using hydrographic data, high-resolution across and low-resolution along the jet is best. For the Mid-Mediterranean Jet, this was applied at the inflow (west of 27.5°E, from 33°N to 35.5°N), but less so near the outflow (east of 28°E and south of 35°N) where the tracks are mainly suited to meridional flows. In practice this is difficult to do, since the flow is locally variable and adaptive sampling is required (e.g. Lermusiaux 1999b).

(*c*) *Computational cost*

The elapsed times of the two mappings were comparable. For the mesoscale PE field analysis, the univariate 2D scheme took 4.02 h on a Sun Ultra. With the same single CPU, the multivariate 3D scheme would have taken 69.5 h for the computation of the error subspace (tracer eigendecomposition, 40 min; ensemble dynamical adjustments, 66.6 h; and SVDs, 1.4 h) but only 50 min for the 3D analysis. The actual computation which used a network of workstations in parallel, equivalent to 16 Sun Ultras, took 7.06 h (40 min + 66.6/16 h + 1.4 h + 50 min). After several simulations, it was noticed that further reductions were possible. For example, the time step in the adjustment runs could have been increased by a factor of 15 while still satisfying the Courant–Friedrichs–Lewy conditions. In general, the number of operations involved in (A.2a) to (A.2c) of appendix A grows quadratically with the size of the a priori error subspace and, for sequential data processing, linearly with the number of data points. In comparison with other practical schemes (e.g. the univariate 2D scheme which specifies $\mathbf{HBH}^{\mathrm{T}}$ and $\mathbf{BH}^{\mathrm{T}}$), the larger the size of the multivariate state space and the larger the number of data points, the more attractive the present mapping becomes. If interest lies only in the a posteriori state (no error estimate is desired) or if the number of data points is small (e.g. smaller than the size of the error subspace), the Method of Representers (Bennett 1992), possibly reduced to their subspace (LER97), is more efficient. In another limit, for small enough states ($\leqslant \mathcal{O}(10^4)$, e.g. horizontal grid-mapping) and number of data points ($\leqslant \mathcal{O}(10^3)$), the full error covariance scheme is feasible today. The error subspace concepts are then mainly useful to estimate, understand, and so perhaps to improve, the error structure.

6. SUMMARY AND CONCLUSIONS

Efficient methods were described and exemplified for the first-guess mapping of 3D multivariate geophysical fields and their dominant errors. The search for the Bayesian conditional mean estimate was approximated by algorithms adopting the minimum-error-variance criterion. With these choices, error subspaces are specified by the dominant eigendecomposition of error covariances.

For the first objective stated in section 2, algorithms for the estimation of a priori error subspaces were outlined. As for the fields themselves (appendix A, (A.1a) and (A.1b)) the data and models are combined, in accordance with their respective limitations. The 3D multivariate structures of the dominant a priori errors are hypothesized to be those of the covariance of the variability from the a priori state. The observed portions of the error subspace are formulated based on vertical expansions and non-homogeneous, non-separable, and anisotropic covariance functions (section 3). The non-observed portions of the error subspace are built in statistical accord with the observed ones via an ensemble of adjustment dynamical integrations (section 4). An ensemble of perturbed a priori states is first created based on the observed variability. The nonlinear and stochastic dynamical equations governing the lesser sampled variables, regions and regimes are then integrated numerically so as to construct the non-observed variability by cross-covariance with the observed one and by auto-covariance with itself.

For the second objective, the algorithms obtained were illustrated and evaluated in the Levantine Sea and in a Middle Atlantic Bight shelfbreak front simulation. Tracer data were utilized to specify the observed variability and a PE model to build the non-observed flow variability. The estimates of dominant error covariances were studied and compared to what is known about these regions and regime. For example, the description of vectors corresponding to the Ierapetra (section 4(c)) illustrated the use of variability subspaces for the understanding of geophysical features. For the main evaluation, a global, multiscale and multivariate, 3D analysis of PE fields and errors in the Levantine Sea was carried out and its results compared to these of a 'benchmark', the univariate 2D scheme of HOPS (section 5). Even though the multivariate 3D approach was simplified for fair comparisons, the outcomes are encouraging. For example, total velocity fields and their error covariances were computed accurately in real time from the tracer data.

For the third objective, an algorithm for the first-guess initialization of fields and errors was obtained. Such estimates can be used to start iterative initialization schemes based on adjoint models (e.g. Talagrand and Courtier 1987; Farrell and Moore 1992; Mureau *et al.* 1993; Molteni *et al.* 1996) or be directly employed in statistical estimation methods evolving the error subspace (LER97). After smoothing, the fields and associated error subspace are in accord with all available data and dynamical information. In LER97, Lermusiaux (1999b) and LLA98, field and error forecasts issued from the present analyses are compared with future *in situ* and sea surface temperature data.

The present approach applies to any geophysical system as long as multiscale, multivariate correlations exist and the number of degrees of freedom necessary to describe most of the variability is limited (Lermusiaux and Robinson 1999). If the correlations between the measured variables and some of the fields to be mapped is weak, the a posteriori values of these fields and their errors remain close to what they were a priori. The filtering of the environmental noise (i.e. errors of representativeness) is facilitated by the use of the dominant, 3D and multivariate error eigenvectors. In some cases, such efficient removal of scales is desirable (e.g. Pinardi and Navarra 1993). Another advantage is the estimation and decomposition of the a posteriori error covariance matrix in real time, which is not often available (e.g. Courtier *et al.* 1998;

Barkmeijer *et al.* 1998). The integration of the complete set of model equations during the dynamical adjustments can also address a few recent issues, including the proper coupling of the mass and wind fields in the Tropics (Andersson *et al.* 1998), the proper covariances between humidity and other fields (Courtier *et al.* 1998) and, in oceanography, tidal rectification and boundary layer effects. Of course, a challenge is the need for a calibrated model to build the non-observed variability. The parametrization of Reynolds stresses, boundary conditions and external forcings must be adequate. With the geostrophic model there are no such parameters to estimate and the code is also much less complex.

The a priori error decomposition algorithm is subject to several variations and/or improvements. The oceanic emphasis led to a vertical expansion in the specification of the observed variability, considering the Cartesian, terrain-following and hybrid coordinate systems (Lozano *et al.* 1994). For specific cases pressure or potential density coordinates, and spherical harmonics or Bessel function expansions, may be more efficient. In general, the observed/non-observed portions of the missing variability are problem-dependent. For example, variability observations could be sea surface winds, temperature and height for a general circulation model simulation of the Pacific ocean, or surface temperature, wind and pressure for a PE simulation of the European weather. The data to be mapped (e.g. current meter) can likewise be of a different type from the (historical) data used in the construction of the error subspace (e.g. hydrographic data). One should also remember that the present scheme only gives a first-guess error subspace. This guess can be computed in advance, e.g. before the start of a real-time forecast experiment. During the experiment dynamical forecasts of the error subspace then lead to 'educated guesses' (e.g. Lermusiaux 1999b). These forecasts can be employed to improve the statistical models fitted to past data or to provide sample error estimates. The method of Parrish and Derber (1992) is a sub-optimal version of the latter, using the differences between forecasts of varied durations as approximate error samples. Finally, the scheme can be iterated. Dynamics can be a weak constraint on the 'observed' portions of $\mathbf{B}$. Similarly, if a few data are available for the 'non-observed' portions, the adjustment integrations can be constrained by these data. Computing $\mathbf{B}^p$ is a data-assimilation problem in itself. Schemes for learning (Brockett 1990) the dominant errors have in fact been derived and utilized (Dee 1995; Blanchet *et al.* 1997; LER97).

The variability of most observed natural processes has a scale-dependent and structured organization. Techniques that take advantage of these properties should continue to prove fruitful in geophysical applications.

ACKNOWLEDGEMENTS

We appreciated the editing and constructive comments of Dr A. C. Lorenc. We are grateful to two anonymous referees for their helpful reviews. One referee pointed out that some of the relevant European literature was originally missing. This led to a useful positioning of our work with respect to several manuscripts published during the revision of the present text. We thank the individuals who helped in collecting the data. We thank Ms M. Armstrong, Ms S. Coons and Mr M. Landes for preparing some of the figures. We benefited greatly from several members of the Harvard oceanography group, past and present. In particular, PFJL thanks his dissertation advisor, Professor A. R. Robinson, and Dr J. Dusenberry, Dr D. V. Kroujiline, Dr P. J. Haley, Mr W. G. Leslie and Ms P. Zaldivar for their help in this work. PFJL and CJL were supported in part by the Office of Naval Research under grants N00014-95-1-0371, N00014-97-1-0239 and N00014-97-1-1018 to Harvard University.

APPENDIX A

Notation and multivariate 3D error subspace mapping

The main notation and definitions are first reviewed. Wherever possible, the notation of Ide *et al.* (1997) is used. The framework is that of continuous-discrete estimation (Jazwinski 1970). The true state vector $\mathbf{x}^t \in \mathrm{R}^n$ is assumed subject to the stochastic dynamical and measurement models, respectively,

$$\mathrm{d}\mathbf{x}^t = \mathcal{M}(\mathbf{x}^t, t)\,\mathrm{d}t + \mathrm{d}\boldsymbol{\eta}^t(t), \tag{A.1a}$$

$$\mathbf{y}^o = \mathcal{H}\{\mathbf{x}^t(t_0)\} + \boldsymbol{\epsilon}, \tag{A.1b}$$

where $\mathcal{M}$ is the dynamics operator, $\boldsymbol{\eta}^t$ is a Wiener process of zero mean and covariance matrix $\mathbf{Q}$, $\mathbf{y}^o \in \mathrm{R}^m$ is the data vector, $\mathcal{H}$ is the measurement operator and $\boldsymbol{\epsilon}$ is a Wiener process of zero mean and covariance matrix $\mathbf{R}$. The time t_0 for the estimation is fixed. In (A.1b), for direct comparisons with the univariate 2D scheme of HOPS, the explicitly time-dependent measurement model of Ide *et al.* (1997), $\mathbf{y}_i^o = \mathcal{H}_i\{\mathbf{x}^t(t_i)\} + \boldsymbol{\epsilon}_i$, where each t_i denotes the time of an observation, is concatenated. The vector $\mathbf{y}^o$ and operator $\mathcal{H}$ are the end-to-end concatenation of the $\mathbf{y}_i^o$ and $\mathcal{H}_i$ terms, respectively ($m = \sum_i m_i$). The vector $\boldsymbol{\epsilon}$ is the end-to-end concatenation of the $\boldsymbol{\epsilon}_i$'s multiplied by $\exp\{(\Delta t_i/\tau)^2/2\}$ to model, as in HOPS, the error increase with the interval Δt_i between the data time t_i and estimation time t_0, τ being a decorrelation time uniform in space. Details on (A.1a and b) are in LLA98.

An unbiased estimate of $\mathbf{x}^t(t_0)$ is denoted by $\mathbf{x} \in \mathrm{R}^n$. The state error covariance is defined by $\mathbf{P} \doteq \mathcal{E}[\{\mathbf{x} - \mathbf{x}^t(t_0)\}\{\mathbf{x} - \mathbf{x}^t(t_0)\}^{\mathrm{T}}] \in \mathrm{R}^{n\times n}$, where $\mathcal{E}\{\cdot\}$ is the statistical mean operator. An estimate of $\mathbf{P}$ is denoted by $\mathbf{B}$. The adjectives *a priori* (^{b}) and *a posteriori* (^{a}) refer to quantities at time t_0 before and after mapping, respectively. Quantities marked with asterisks are normalized. Sample error or residual matrices are denoted by the symbol $\mathbf{S}$. Their columns are field samples, the size of $\mathbf{S}$ depending on the context. The number of horizontal, vertical and total grid points are denoted by ℓ_h, ℓ_v and $\ell \doteq \ell_h\ell_v$, respectively.

To be consistent with the error measure of the present minimum-error-variance criterion (section 2), the covariances are optimally approximated, for an error subspace of dimension p, by their dominant rank-p eigendecomposition. A subspace is the portion of R^n of dimension p which is spanned by the eigenvectors corresponding to the dominant p eigenvalues of a covariance. The appropriate dimension p can be determined by a quantitative criterion, e.g. (22) in the main text. The classic, fixed-time, minimum-error-variance estimate (Gelb 1974) is thus reduced to:

$$\left\{\mathbf{x} \;\middle|\; \min_{\mathbf{x}} \mathrm{tr}[\mathbf{P}^{a^p}],\ \text{knowing } [\mathbf{y}^o, \mathbf{R}] \text{ and } [\mathbf{x}^b, \mathbf{P}^{b^p}]\right\}, \tag{A.1c}$$

in which $\mathbf{x}^t(t_0)$ is subject to (A.1a) and (A.1b). The superscript (^{p}) on a matrix indicates a rank-p error subspace approximation. The a posteriori principal covariance $\mathbf{P}^{a^p}$ is the rank-p eigendecomposition of $\mathbf{P}^a$, $\mathbf{P}^{a^p} \doteq \mathbf{E}^a\boldsymbol{\Pi}^a\mathbf{E}^{a^{\mathrm{T}}}$. The ordered diagonal matrix $\boldsymbol{\Pi}^a$ contains the largest p eigenvalues. The corresponding eigenvectors, columns of $\mathbf{E}^a$, span the a posteriori error subspace. Similar statements apply to $\mathbf{P}^{b^p} = \mathbf{E}\boldsymbol{\Pi}\mathbf{E}^{\mathrm{T}}$, but in general with different rank and eigendecomposition. In this paper, the superscript (^{b}) on the a priori error eigenvalues and eigenvectors is omitted.

The extremum of (A.1c) yields $\mathbf{x}^a$, $\mathbf{E}^a$ and $\boldsymbol{\Pi}^a$. This could be solved for either by a descent algorithm (e.g. quasi-Newton) in the error subspace or by direct inversion. The latter approach is used here. To do so, (A.1b) is linearized in the vicinity of $\mathbf{x}^b$, as in the

incremental 3D-Var (Courtier *et al.* 1998). For $\mathbf{x}^a$ hypothesized to be a linear function of $\mathbf{x}^b$ and $\mathbf{y}^o - \mathcal{H}(\mathbf{x}^b)$, each of which is assumed to be unbiased, the extremum is (LER97):

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{E}\mathbf{\Pi}\mathbf{H}^{p^T}(\mathbf{H}^p\mathbf{\Pi}\mathbf{H}^{p^T} + \mathbf{R})^{-1}\{\mathbf{y}^o - \mathcal{H}(\mathbf{x}^b)\}, \tag{A.2a}$$

$$\mathbf{U}\mathbf{\Pi}^a\mathbf{U}^T = \tilde{\mathbf{\Pi}}^a \doteq \mathbf{\Pi} - \mathbf{\Pi}\mathbf{H}^{p^T}(\mathbf{H}^p\mathbf{\Pi}\mathbf{H}^{p^T} + \mathbf{R})^{-1}\mathbf{H}^p\mathbf{\Pi}, \tag{A.2b}$$

$$\mathbf{E}^a = \mathbf{E}\mathbf{U}, \tag{A.2c}$$

where $\mathbf{H}$ is the linearization of $\mathcal{H}$ in the vicinity of $\mathbf{x}^b$, $\mathbf{H}^p \doteq \mathbf{H}\mathbf{E}$, the columns of $\mathbf{U}$ are ordered orthonormal eigenvectors of $\tilde{\mathbf{\Pi}}^a$ and $\mathbf{\Pi}^a$ is diagonal. In (A.2a) to (A.2c), the gain $\mathbf{K}^p$ is $\mathbf{P}^{b^p}\mathbf{H}^T[\mathbf{H}\mathbf{P}^{b^p}\mathbf{H}^T + \mathbf{R}]^{-1}$ and the rank of $\mathbf{P}^{a^p}$ given by (A.2b) and (A.2c) is assumed equal to that of $\mathbf{P}^{b^p}$. The data residuals corresponding to a state in the complement of the error subspace are neglected. The differences between the Kalman update (e.g. Gelb 1974) and estimate (A.2a) as a function of the properties of (A.1b) are further discussed in LLA98.

APPENDIX B

EOF decomposition of tracer vertical covariances $\mathbf{C}^z_{trc}$

The algorithm for estimating the dominant decomposition of $\mathbf{C}^z_{trc}$ from the vertical EOFs of the a priori data residuals (section 3(b)) is outlined. Usually, the vertical resolution of hydrographic profiles is high and, if the number of profiles is significant, such estimates of $\mathbf{C}^z_{trc}$ are good. The assumptions are as in section 4(a). For each scale w, the regional historical (synoptic) profiles $i = 1, \ldots, q$ are first vertically interpolated onto the ℓ_v dynamical-model surfaces, here by vertical box-averaging. The intersections between the profiles and model surfaces are the data points in this context. The vertically interpolated profiles are the $\mathbf{y}^o_i$ terms (in this appendix, $\mathbf{y}^o_i$ is in fact a scale-restricted $\mathbf{y}^o_{i_w}$, but w is omitted for ease of notation). The differences between the $\mathbf{y}^o_i$ terms and the tracer fields $\mathbf{x}^b_{trc}$ of the previous scale-correction are then evaluated at data points, leading to the a priori data residuals,

$$\mathbf{d}_i = \mathbf{y}^o_i - \mathcal{H}_i(\mathbf{x}^b). \tag{B.1}$$

The operators $\mathcal{H}_i$ in (B.1) interpolate $\mathbf{x}^b_{trc}$ onto the data points. They consist here of bilinear interpolators along horizontal (level) PE surfaces. The horizontal averages $\overline{\mathbf{d}_i}$ of the residuals $\mathbf{d}_i$ (B.1) are then removed and the resulting zero-mean residuals normalized by their sample and volume average,

$$\mathbf{d}^*_i = \mathbf{N}_d^{-1}(\mathbf{d}_i - \overline{\mathbf{d}_i}). \tag{B.2}$$

The normalization matrix $\mathbf{N}_d$ in (B.2) is block-diagonal. Forming the matrix $\mathbf{S}^*_d$ of normalized sample tracer residuals,

$$\mathbf{S}^*_d = [\mathbf{d}^*_1; \ldots; \mathbf{d}^*_q], \tag{B.3}$$

the sample covariance $\mathbf{S}^*_d\mathbf{S}^{*T}_d/q$ is obtained. Its dominant rank-p_d eigendecomposition is sought and efficiently computed from, $\mathrm{SVD}_{p_d}(\mathbf{S}^*_d) = \mathbf{E}^*_d\mathbf{\Sigma}^*_d\mathbf{V}^{*^T}_d$, where $p_d \leqslant 2\ell_v$ and $\mathrm{SVD}_{p_d}(\cdot)$ denotes the operator selecting the rank-p_d SVD. For each scale, this yields the vertical EOFs and coefficients of the $\mathbf{d}_i$ (B.1), $\mathbf{E}^z_{trc} = \mathbf{N}_d\mathbf{E}^*_d$ and $\mathbf{\Pi}^z_{trc} = \mathbf{\Sigma}^{*^2}_d/q$, which

specify the dominant decomposition of $\mathbf{C}^z_{\text{trc}}$,

$$\mathbf{C}^z_{\text{trc}} = \mathbf{E}^z_{\text{trc}} \mathbf{\Pi}^z_{\text{trc}} \mathbf{E}^{z^{\mathrm{T}}}_{\text{trc}}. \tag{B.4}$$

The generalization of (B.1) to (B.4) to a weighted SVD approach (e.g. Thacker 1996; Wunsch 1996), accounting for varied error sources (e.g. in our case the error increase due to the time decorrelation, $\exp\{(\Delta t_i/\tau)^2/2\}$, see appendix A), is left as an exercise for the interested reader.

REFERENCES

Andersson, E., Haseler, J., Under, P., Courtier, P., Kelly, G., Vasiljevic, D., Brankovic, C., Cardinali, C., Gaffard, C., Hollingsworth, A., Jakob, C., Janssen, P., Klinker, E., Lanzinger, A., Miller, M., Rabier, F., Simmons, A., Strauss, B., Thépaut, J.-N. and Viterbo, P. 1998 The ECMWF implementation of three-dimensional variational assimilation (3D-Var). III: Experimental results. *Q. J. R. Meteorol. Soc.*, **124,** 1831–1860

Balgovind, R., Dalcher, A., Ghil, M. and Kalnay, E. 1983 A stochastic–dynamic model for the spatial structure of forecast error statistics. *Mon. Weather Rev.*, **111,** 701–722

Barkmeijer, J., Van Gijzen, M. and Bouttier, F. 1998 Singular vectors and estimates of the analysis-error covariance metric. *Q. J. R. Meteorol. Soc.*, **124,** 1695–1713

Barnes, S. L. 1964 A technique for maximizing details in numerical weather map analysis. *J. Appl. Meteorol.*, **3,** 369–409

1994 Applications of the Barnes objective analysis scheme. Part I: Effects of undersampling, wave position, and station randomness. *J. Atmos. Oceanic Technol.*, **11,** 1433–1448

Bartello, P. and Mitchell, H. L. 1992 A continuous three-dimensional model of short-range forecast error covariance. *Tellus*, **44A,** 217–235

Bennett, A. F. 1992 Inverse methods in physical oceanography. *Cambridge Monographs on mechanics and applied mathematics.* Cambridge University Press, UK

Bennett, A. F., Chua, B. S. and Leslie, L. M. 1995 Generalized inversion of a global numerical weather prediction model. *Meteorol. Atmos. Phys.*, **60**(1–3), 165–178

Blanchet, I., Frankignoul, C. and Cane, M. A. 1997 A comparison of adaptive Kalman filters for a tropical Pacific Ocean model. *Mon. Weather Rev.*, **125**(1), 40–58

Bouttier, F. 1994 A dynamical estimation of forecast error covariances in an assimilation system. *Mon. Weather Rev.*, **122,** 2376–2390

Brankart, J. M. and Brasseur, P. 1996 Optimal analysis of *in situ* data in the western Mediterranean using statistics and cross-validation. *J. Atmos. Oceanic Technol.*, **16,** (2), 477–491

Brenner, S. 1993 Long-term evolution and dynamics of a persistent warm core eddy in the Eastern Mediterranean Sea. *Deep-Sea Res. II*, **40**(6), 1193–1206

Bretherton, F. P., Davis, R. E. and Fandry, C. B. 1976 A technique for objective analysis and design of oceanographic experiments applied to MODE-73. *Deep-Sea Res.*, **23**(7), 559–582

Brockett, R. W. 1990 Dynamical systems that learn subspaces. Pp. 579–592 in *Mathematical system theory: the influence of R. E. Kalman.* Ed. A. C. Antoulis. Springer-Verlag, Berlin, Germany

Cane, M. A., Kamenkovich, V. M. and Krupitsky, A. 1998 On the utility and disutility of JEBAR. *J. Phys. Oceanogr.*, **28**(3), 519–526

Carter, E. F. and Robinson, A. R. 1987 Analysis models for the estimation of oceanic fields. *J. Atmos. Oceanic Technol.*, **4,** No. 1, 49–74

Carton, J. A. and Hackert, E. C. 1989 Application of multi-variate statistical objective analysis to the circulation in the tropical Atlantic Ocean. *Dyn. Atmos. Ocean*, **13,** 491–515

Courtier, P., Andersson, E., Heckley, W., Pailleux, J., Vasiljevic, D., Hamrud, M., Hollingsworth, A., Rabier, F. and Fisher, M. 1998 The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation. *Q. J. R. Meteorol. Soc.*, **124,** 1783–1807

Cressman, G. P. 1959 An operational objective analysis system. *Mon. Weather Rev.*, **87,** 367–374

Cushman-Roisin, B. 1994 *Introduction to physical oceanography.* Prentice-Hall Inc., New York, USA

Daley, R. 1991 *Atmospheric data analysis.* Cambridge University Press, New York, USA

1992a The lagged innovation covariance: A performance diagnostic for atmospheric data assimilation. *Mon. Weather Rev.*, **120,** 178–196

1992b Forecast-error statistics for homogeneous and inhomogeneous observation networks. *Mon. Weather Rev.*, **120,** 627–643

1992c Estimating model-error covariances for application to atmospheric data assimilation. *Mon. Weather Rev.*, **120,** 1735–1746

1996 Generation of global multivariate error covariances by singular-value decomposition of the linear balance equation. *Mon. Weather Rev.*, **124,** 2574–2587

Dee, D. P. 1995 On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Weather Rev.*, **123,** 1128–1145

Evans, G. T. and Fasham, M. J. R. (Eds.) 1993 *Towards a model of ocean biogeochemical processes.* NATO ASI series. Series I, Global environmental change, vol. 10. Springer-Verlag, New York, USA

Farrell, B. F. and Moore, A. M. 1992 An adjoint method for obtaining the most rapidly growing perturbation to the oceanic flows. *J. Phys. Oceanogr.*, **22,** 338–349

Fieguth, P. W., Willsky, A. S., Menemenlis, D. and Wunsch, C. I. 1996 'A general multiresolution approach to the estimation of dense fields in remote sensing'. Pp. 16–19 in Proceedings of the 3rd IEEE international conference on image processing, Lausanne, Switzerland

Gamage, N. and Blumen, W. 1993 Comparative analysis of low-level cold fronts: Wavelet, Fourier and empirical orthogonal function decompositions. *Mon. Weather Rev.*, **121,** 2867–2878

Gandin, L. S. 1965 Objective analysis of meteorological fields (Ob"ektivnyi analiz meteorologicheskikh polei, Leningrad 1963). Translated from Russian. Israel Program for Scientific Translations. Jerusalem

Gangopadhyay, A., Robinson, A. R. and Arango, H. G. 1997 Circulation and dynamics of the Western North Atlantic. Part I: Multiscale feature models. *J. Atmos. Oceanic Technol.*, **14,** 1314–1332

Gaspari, S. and Cohn, S. E. 1999 Construction of correlation functions in two and three dimensions. *Q. J. R. Meteorol. Soc.*, **125,** 723–757

Gauthier, P., Fillion, L., Koclas, P. and Charette, C. 1996 'Implementation of a 3D-variational analysis at the Canadian meteorological centre'. Pp. 19–23 in Proceedings of the 11th AMS conference on numerical weather prediction, Vancouver. American Meteorological Society, Boston, USA

Gavart, M. and DeMey, P. 1997 Isopycnal EOFs in the Azores Current region: A statistical tool for dynamical analysis and data assimilation. *J. Phys. Oceanogr.*, **27**(10), 2146–2157

Gelb, A. (ed.) 1974 *Applied optimal estimation.* MIT Press, Cambridge, USA

Gill, A. E. 1982 *Atmosphere–ocean dynamics.* International Geophysics Series, Vol. 30, Academic Press, New York, USA

Glimm, J. and Sharp, D. H. 1997 Stochastic partial differential equations: selected applications in continuum physics. In *Stochastic partial differential equations: Six perspectives.* Eds. R. A. Carmona and B. L. Rozoskii. Mathematical surveys and monographs. Birkhauser, New York, USA

Graham, A. 1981 *Kronecker products and matrix calculus: with applications.* Halsted Press, New York, USA

Haidvogel, D. B. 1983 Periodic and regional models. In *Eddies in marine sciences.* Ed. A. R. Robinson. Springer-Verlag, New York, USA

Haney, R. L., Hale, R. A. and Collins, C. A. 1995 Estimating subpycnocline density-fluctuations in the California current region from upper ocean observations. *J. Atmos. Oceanic Technol.*, **12**(3), 550–566

Hollingsworth, A. and Lönnberg, P. 1986 The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: the wind field. *Tellus*, **38A,** 111–136

Houghton, J. 1991 The Bahesian Lecture, 1991: The predictability of weather and climate. *Philos. Trans. R. Soc. London A*, **337,** 521–572

Hua, B. L., McWilliams, J. C. and Owens, W. B. 1986 An objective analysis of the POLYMODE local dynamics experiment. Part II: Streamfunction and potential vorticity fields during the intensive period. *J. Phys. Oceanogr.*, **16,** 506–522

Ide, K., Courtier, P., Ghil, M. and Lorenc, A. C. 1997 Unified notation for data assimilation: operational, sequential and variational. *J. Meteorol. Soc. Jpn.*, **75**(1B), 181–189

Jazwinski, A. H. 1970 *Stochastic processes and filtering theory.* Academic Press, New York, USA

Jiang, S. and Ghil, M. 1993 Dynamical properties of error statistics in a shallow-water model. *J. Phys. Oceanogr.*, **23**(12), 2541–2566

Kaplan, A., Kushnir, Y., Cane, M. A. and Blumenthal, M. B. 1997 Reduced space optimal analysis for historical data sets: 136 years of Atlantic sea surface temperatures. *J. Geophys. Res.*, **102**(C13), 27835–27860

Killworth, P. D. 1986 A Bernoulli inverse method for determining the ocean circulation. *J. Phys. Oceanogr.*, **16,** 2031–2051

Killworth, P. D. and Bigg, G. R. 1988 An intercomparison of inverse methods using an eddy-resolving general circulation model. *J. Phys. Oceanogr.*, **18,** 987–1008

Lermusiaux, P. F. J. 1997 Error subspace data assimilation methods for ocean field estimation: theory, validation and applications. PhD Thesis, Harvard University, Cambridge MA

1999a Data assimilation via error subspace statistical estimation, Part II: Middle Atlantic Bight shelfbreak front simulations and ESSE validation. *Mon. Weather Rev.*, **127**(7), 1408–1432

1999b Estimation and study of mesoscale variability in the strait of Sicily. *Dyn. Atmos. Ocean*, Special issue in honor of Professor A. R. Robinson. **29,** 255–303

Lermusiaux, P. F. J. and Robinson, A. R. 1999 Data assimilation via error subspace statistical estimation, Part I: theory and schemes. *Mon. Weather Rev.*, **127**(7), 1385–1407

Lermusiaux, P. F. J., Lozano, C. J. and Anderson, D. G. 1998 'On the mapping of multivariate geophysical fields: studies of the sensitivity to error subspace parameters'. Harvard Open Ocean Model report No. 58. Harvard University, Cambridge MA, USA

Lönnberg, P. and Hollingsworth, A. 1986 The statistical structure of short-range forecast errors as determined from radiosonde data. Part II: The covariance of height and wind errors. *Tellus*, **38A,** 137–161

Lorenc, A. C. 1981 A global three-dimensional multivariate statistical analysis scheme. *Mon. Weather Rev.*, **109,** 701–721

1986 Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.*, **112,** 1177–1194

1992 Iterative analysis using covariance functions and filters. *Q. J. R. Meteorol. Soc.*, **118,** 569–591

Lorenz, E. N. 1965 A study of the predictability of a 28-variable atmospheric model. *Tellus*, **17,** 321–333

Lozano, C. J. and Candela, J. 1995 The M_2 tide in the Mediterranean Sea: dynamic analysis and data assimilation. *Oceanol. Acta*, **18**(4), 419–441

Lozano, C. J., Haley, P. J., Arango, H. G., Sloan, Q. and Robinson, A. R. 1994 'Harvard coastal/deep water primitive equation model'. Harvard open ocean model report No. 52. Harvard University, Cambridge, MA, USA

Lozano, C. J., Robinson, A. R., Arango, H. G., Gangopadhyay, A., Sloan, N. Q., Haley, P. J. and Leslie, W. G. 1996 An interdisciplinary ocean prediction system: assimilation strategies and structured data models. *Modern approaches to data assimilation in ocean modelling.* Ed. P. Malanotte-Rizzoli. Elsevier, New York, USA

Malanotte-Rizzoli, P., Robinson, A. R., Roether, W., Manca, B., Bergamasco, A., Brenner, S., Civitarese, G., Georgopoulos, D., Haley, P. J., Kioroglou, S., Kontoyannis, H., Kress, N., Latif, M. A., Leslie, W. G., Ozsoy, E., Ribera d'Alcala, M., Salihoglu, I., Sansone, E. and Theocharis, A. 1996 Experiment in Eastern Mediterranean probes origin of deep water masses. EOS, **77**(32)

Mariano, A. J. and Brown, O. B. 1992 Efficient objective analysis of dynamically heterogeneous and nonstationary fields via the parameter matrix. *Deep Sea Res., A,* **39**(7–8A), 1255–1271

McIntosh, P. C. 1990 Oceanographic data interpolation: objective analysis and splines. *J. Geophys. Res.*, **95**, (C8), 13529–12541

McWilliams, J. C., Owens, W. B. and Hua, B. L. 1986 An objective analysis of the POLYMODE local dynamics experiment. Part I: General formalism and statistical model selection. *J. Phys. Oceanogr.*, **16**, 483–504

Miller, A. J., Lermusiaux, P. F. J. and Poulain, P. M. 1996 A topographic-Rossby mode resonance over the Iceland–Faeroe ridge. *J. Phys. Oceanogr.*, **26**(12), 2735–2747

Molteni, F. and Palmer, T. N. 1993 Predictability and finite-time instability of the northern winter circulation. *Q. J. R. Meteorol. Soc.*, **119**, 269–298

Molteni, F., Buizza, R., Palmer, T. N. and Petroliagis, T. 1996 The ECMWF ensemble prediction system: Methodology and validation. *Q. J. R. Meteorol. Soc.*, **122**, 73–119

Monin, A. S. 1990 *Theoretical geophysical fluid dynamics*. Kluwer Academic Publishers, Dordrecht, the Netherlands

Munk, W., Worcester, P. and Wunsch, C. 1995 *Ocean Acoustic Tomography*. Cambridge University Press, Cambridge, UK

Mureau, R., Molteni, F. and Palmer, T. N. 1993 Ensemble prediction using dynamically conditioned perturbations. *Q. J. R. Meteorol. Soc.*, **119**, 299–323

Nihoul, J. C. J. and Djenidi, S. 1998 Coupled physical, chemical and biological models. In *The sea: The global coastal ocean: I Processes and methods*, Volume 10. Eds. K. H. Brink and A. R. Robinson. John Wiley and Sons, New York, USA

Parrish, D. F. and Cohn, S. E. 1985 'A Kalman filter for a two-dimensional shallow-water model: formulation and preliminary experiments'. Office Note 304, US Dept. of Commerce, NOAA, NWS, NMC, Washington DC, USA

Parrish, D. F. and Derber, J. C. 1992 The National Meteorological Center's spectral statistical interpolation analysis system. *Mon. Weather Rev.*, **120**, 1747–1763

Pedlosky, J. 1987 *Geophysical Fluid Dynamics*. Second edition, Springer-Verlag, New York, USA

Phillips, N. A. 1986 The spatial statistics of random geostrophic modes and first-guess errors. *Tellus*, **38A**, 314–322

Pinardi, N. and Navarra, A. 1993 Baroclinic wind adjustment processes in the Mediterranean Sea. *Deep-Sea Res. II.* **40**(6), 1299–1326

Preisendorfer, R. W. 1988 *Principal Component Analysis in Meteorology and Oceanogrphy*. Elsevier, New York, USA

Rabier, F., McNally, A., Andersson, E., Courtier, P., Unden, P., Eyre, J., Hollingsworth, A. and Bouttier, F. 1998 The ECMWF implementation of three-dimensional variational assimilation (3D-Var). II: Structure functions. *Q. J. R. Meteorol. Soc.*, **124**, 1809–1829

Rhines, P. B. 1977 The dynamics of unsteady currents. Pp. 189–318 in *The sea: Volume 6*. Eds. E. D. Goldberg, I. N. McCave, J. J. O'Brien and J. H. Steele. Wiley and Sons, New York, USA

Rienecker, M. M. and Adamec, D. 1995 Assimilation of altimeter data into a quasi-geostrophic ocean model using optimal interpolation and EOFs. *J. Mar. Syst.*, **6**(1–2), 125–143

Robinson, A. R. (Ed.) 1983 *Eddies in marine science*. Springer-Verlag, New York, USA

Robinson, A. R. 1996 Physical processes, field estimation and an approach to interdisciplinary ocean modeling. *Earth Sci. Rev.*, **40**, 3–54

Robinson, A. R. and Golnaraghi, M. 1993 Circulation and dynamics of the Eastern Mediterranean Sea; quasi-synoptic data-driven simulations. *Deep-Sea Res. II.* **40**(6), 1207–1246

Robinson, A. R. and Malanotte-Rizzoli, P. (Eds.) 1993 Topical studies in oceanography. *Deep-Sea Res., II*, **40**(6), 1073–1329

Robinson, A. R., Lermusiaux, P. F. J. and Sloan, N. Q. III 1998a Data assimilation. In *The sea, Volume 10: The global coastal ocean I: Processes and methods*. Eds. K. H. Brink and A. R. Robinson. John Wiley and Sons, New York, USA

Robinson, A. R., Haley Jr., P. J., Lermusiaux, P. F. J., Leslie, W. G., Lozano, C. J., Anderson, L. A. and Dusenberry, J. A. 1998b 'The rapid response 96, 97 and 98 exercises: the Strait of Sicily, Ionian Sea and Gulf of Cadiz'. Harvard open ocean model report No. 57. Harvard University, Cambridge MA, USA

Robinson, A. R., Sellschopp, J., Warn-Varnas, A., Leslie, W. G., Lozano, C. J., Haley P. J., Jr., Anderson, L. A. and Lermusiaux, P. F. J. 1999 The Atlantic Ionian Stream. *J. Mar. Sys.*, **20,** 129–156

Roether, W., Manca, B. B., Klein, B., Bregant, D., Georgopoulos, D., Beitzel, V., Kovačević, V. and Luchetta, A. 1996 Recent changes in Eastern Mediterranean deep water. *Science*, **271,** 333–335

Salmon, R. 1980 Baroclinic instability and geostrophic turbulence. *Geophys. Astrophys. Fluid Dyn.*, **15,** 167–211

Schott, F. and Stommel, H. 1978 Beta spirals and absolute velocities in different oceans. *Deep-Sea Res.*, **25,** 961–1010

Sloan, N. Q. 1996 Dynamics of a shelf/slope front: process studies and data-driven simulations. PhD Thesis, Harvard University, Cambridge MA

Smith, P. C. 1983 Eddies and coastal interactions. Pp. 446–480 in *Eddies in marine science*. Ed. A. R. Robinson. Springer-Verlag, New York, USA

Stammer, D. 1997 Geosat data assimilation with application to the eastern North Atlantic. *J. Phys. Oceanogr.*, **27**(1), 40–61

Talagrand, O. and Courtier, P. 1987 Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory. *Q. J. R. Meteorol. Soc.*, **113,** 1311–1328

Thacker, W. C. 1996 Metric-based principal components: data uncertainties. *Tellus*, **48A**(4), 584–592

Thacker, W. C. and Lewandowicz, R. 1997 A comparison of low-dimensional representations of sea-surface temperature anomalies in the North Atlantic. *Int. J. Climatol.*, **17**(9), 953–67

Thépaut, J. N., Courtier, P., Belaud, G. and Lemaitre, G. 1996 Dynamical structure functions in a four-dimensional variational assimilation: A case-study. *Q. J. R. Meteorol. Soc.*, **122,** 535–561

Thiebaux, H. J. 1976 Anisotropic correlation functions for objective analysis. *Mon. Weather Rev.*, **104,** 994–1002

Tsaoussi, L. S. and Koblinsky, C. J. 1994 An error covariance model for sea-surface topography and velocity derived from TOPEX/POSEIDON altimetry. *J. Geophys. Res.*, **99**(C12), 24669–24683

von Storch, H. and Frankignoul, C. 1998 Empirical modal decomposition in coastal oceanography. In *The sea: The global coastal ocean I: Processes and methods, Volume 10*. Eds. K. H. Brink and A. R. Robinson. John Wiley and Sons, New York, USA

Wahba, G. and Wendelberger, J. 1980 Some new mathematical methods for variational objective analysis using splines and cross validation. *Mon. Weather Rev.*, **108,** 1122–1143

Watts, D. R., Tracey, K. L. and Friedlander, A. I. 1989 Producing accurate maps of the Gulf Stream thermal front using objective analysis. *J. Geophys. Res.*, **94**(C6), 8040–8052

Woodgate, R. A. and Killworth, P. D. 1996 The problem of the barotropic mode in deriving pressure from density by using vertical normal modes. *J. Geophys. Res.*, **101**(C2), 3765–3768

Wunsch, C. 1978 The general circulation of the North Atlantic west of 50°W determined from inverse methods. *Rev. Geophys.*, **16,** 583–620

1996 *The ocean circulation inverse problem*. Cambridge University Press, New York, USA

1997 The vertical partition of oceanic horizontal kinetic energy. *J. Phys. Oceanogr.*, **27**(8), 1770–94