Data Assimilation via Error Subspace Statistical Estimation.
Part I: Theory and Schemes

P. F. J. LERMUSIAUX

Division of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts

A. R. ROBINSON

*Division of Engineering and Applied Sciences, Department of Earth and Planetary Sciences, Harvard University,
Cambridge, Massachusetts*

ABSTRACT

A rational approach is used to identify efficient schemes for data assimilation in nonlinear ocean–atmosphere models. The conditional mean, a minimum of several cost functionals, is chosen for an optimal estimate. After stating the present goals and describing some of the existing schemes, the constraints and issues particular to ocean–atmosphere data assimilation are emphasized. An approximation to the optimal criterion satisfying the goals and addressing the issues is obtained using heuristic characteristics of geophysical measurements and models. This leads to the notion of an evolving error subspace, of variable size, that spans and tracks the scales and processes where the dominant errors occur. The concept of error subspace statistical estimation (ESSE) is defined. In the present minimum error variance approach, the suboptimal criterion is based on a continued and energetically optimal reduction of the dimension of error covariance matrices. The evolving error subspace is characterized by error singular vectors and values, or in other words, the error principal components and coefficients.

Schemes for filtering and smoothing via ESSE are derived. The data–forecast melding minimizes variance in the error subspace. Nonlinear Monte Carlo forecasts integrate the error subspace in time. The smoothing is based on a statistical approximation approach. Comparisons with existing filtering and smoothing procedures are made. The theoretical and practical advantages of ESSE are discussed. The concepts introduced by the subspace approach are as useful as the practical benefits. The formalism forms a theoretical basis for the intercomparison of reduced dimension assimilation methods and for the validation of specific assumptions for tailored applications. The subspace approach is useful for a wide range of purposes, including nonlinear field and error forecasting, predictability and stability studies, objective analyses, data-driven simulations, model improvements, adaptive sampling, and parameter estimation.

1. Introduction

Data assimilation (DA) refers to the estimation of oceanic–atmospheric fields by melding sensor data with a model of the dynamics under study. Most DA schemes are rooted in statistical estimation theory: the state of a system is estimated by combining all knowledge of the system, like measurements and theoretical laws or empirical principles, in accord with their respective statistical uncertainty. The present challenge is that the state of the atmosphere and ocean system is complex, evolving on multiple time and space scales (Charnock 1981; Charney and Flierl 1981). Direct observations can be

difficult and costly to acquire on a sustained basis, especially in oceanography. The large breadth of scales and variables also leads to costly and challenging numerical simulations. Future advances in coupled ocean–atmosphere estimation will thus require efficient assimilation schemes. In Part I of this two-part paper, the main goal is to develop the basis of a comprehensive, portable, and versatile four-dimensional DA scheme for the estimation and simulation of realistic geophysical fields. The adjective *realistic* emphasizes that the scheme should capture the time and space scales of the real processes of interest. It implies the use of real ocean data, as well as appropriate theoretical models and numerical resources. The primary focus is on the physics; acoustical, biological, and chemical phenomena will be investigated later. The implementation presented is compatible with the Harvard Ocean Prediction System (HOPS; e.g., Lozano et al. 1996; Robinson 1996b) and the future of this work involves ocean–atmospheric

Corresponding author address: Dr. Pierre F. J. Lermusiaux, Division of Engineering and Applied Sciences, Harvard University, 29 Oxford Street, Cambridge, MA 02138.
E-mail: pierrel@pacific.harvard.edu

data-driven estimations. In Part II (Lermusiaux 1999a) of this paper, identical twin experiments based on Middle Atlantic Bight shelfbreak front simulations are employed to assess and exemplify the capabilities of the present DA scheme.

A description of the goals and uses of DA, with a review of most methods, is given in Robinson et al. (1998a, and references therein). The issue is that, with most existing approaches, the combination of our practical, accuracy, and realism goals is difficult to satisfy (sections 3, 4). In fact, several directions have been taken so as to determine feasible schemes for realistic studies. Examples of such attempts include simpler physics models to integrate errors (e.g., Dee et al. 1985; Dee 1990; Daley 1992b), variance-only error models (e.g., Daley 1991, 1992b), steady-state error models (e.g., Fukumori et al. 1993), and reduced dimension or coarse-grid Kalman filters (KF; e.g., Fukumori and Malanotte-Rizzoli 1995). Other reductions deal with the explicit computation of non-null elements of linearized transfer matrices (e.g., Parrish and Cohn 1985; Jiang and Ghil 1993), banded approximations (e.g., Parrish and Cohn 1985), extended filters (e.g., Evensen 1993), ensemble and Monte Carlo methods (e.g., Evensen 1994a,b; Miller et al. 1994), and possibly using the optimal and breed perturbations for assimilation (Ehrendorfer and Errico 1995; Toth and Kalnay 1993). It is important to realize that several of these attempts are based on incompatible hypotheses. Briefly, the coarse-grid KFs imply global-scale forecast errors while the variance-only and banded approximations assume local errors. Pure Monte Carlo methods acknowledge the importance of nonlinear terms, extended schemes neglect their effects locally in time, and linearized methods neglect them at all times. The forward integration of error fields using simpler physics models assumes that the dominant predictability error is never correlated to the complex physics. Steady-state error models are somewhat limited to fixed data arrays and statistically steady dynamics. For each attempt, the list of arguments for and against is long. Even if most a priori reduced methods have been successful data interpolators, they have logically led to controversies.

If one accepts that, in general, relatively little is known about dynamical and observational error fields, it is rational to limit the a priori assumptions. For the present comprehensive aims, the conditional mean, a minimum of several cost functionals or estimation criteria, is chosen for the optimal estimate. An approximation to the estimation criterion is obtained using heuristic characteristics of geophysical measurements and models. The resulting suboptimal approach is based on an objective, evolving truncation of the number and dimension of the parameters that characterize the conditional probability or error space. The ideal error (probability) subspace spans and tracks the scales and processes where the dominant errors (low probabilities) occur. The notion of dominant is naturally defined by the

error measure used in the chosen optimal criterion: to each estimation criterion corresponds an error subspace definition. For the present minimum error variance approach, the logical definition yields an evolving subspace, of variable size, characterized by error singular vectors and values, or, similarly, the error empirical orthogonal functions (EOFs) and coefficients. Data assimilation via error subspace statistical estimation (ESSE) combines data and dynamics in accord with their respective dominant uncertainties. Once successfully applied, ESSE can rationally validate specific a priori error truncations for future tailored applications. Organizing the error space as a function of relative importance in fact defines a theoretical basis for quantitative intercomparison of today's numerous reduced dimension methods. A first issue of course is the meaning and validity of the truncation of geophysical error spaces (section 5 and Part II). Another is that it is easy to define the concept of an evolving subspace, but it is harder to determine mathematical systems that describe and track its evolution. Most of the schemes for filtering and smoothing via ESSE (derived next) have variations. Focusing on the dominant errors fosters dynamical model testing and corrections. The error subspace also helps to identify areas and variables for which observations are most needed. ESSE provides a feasible quantitative approach to both dynamical model improvements and adequate sampling. Historically, these have been challenging issues. The accurate specification and tracking of the dominant errors hence appears of paramount importance from a fundamental point of view.

The text is organized as follows. Selected definitions and generalities are stated in section 2. Section 3 deals with the focus and specific objectives of the paper. Section 4 develops the main issues in realistic and comprehensive data assimilation today: the efficient reduction of error models and the powerful use of all information present in the few observations available. Section 5 addresses the meaning of the variations of variability and error subspaces. The correlations in geophysical systems are emphasized and the ESSE criteria introduced. Sections 6 and 7 derive schemes for filtering and smoothing via ESSE, respectively. These filtering and smoothing algorithms with nonlinear systems are succinctly compared with existing "suboptimal" procedures. Section 8 consists of the summary and conclusions. Appendix A describes most of the notation and assumptions. Appendix B addresses important specifics and variations of the ESSE schemes presented.

2. Definitions and generalities

The dynamics and sensor data are described by mathematical models. The *dynamical model* is an approximation of the basic laws to the phenomena and scales of interest. It defines the time–space evolution of the state variables and, in continuous form (appendix A),

$$\mathbf{d}\boldsymbol{\psi} = \mathbf{f}(\boldsymbol{\psi}, t) \, dt. \tag{1a}$$

Dynamical variability (true or model) refers to the statistics of the difference between the dynamical system evolution (true or model) and a reference mean state. The model (1a) usually considers both the variability and mean state. The *measurement model*[1] is a *directed* relation linking the state variables to the observations (appendix A):

$$\mathbf{d}_k = \mathbf{C}_k \boldsymbol{\psi}_k. \tag{1b}$$

The relation (1b) may comprise changes of variables (e.g., diagnostic relations or correlations), forward interpolations, or time–space filters. The data vectors $\mathbf{d}_k$ can be sensor, feature modeled, or structured data (Lozano et al. 1996). For simplicity, (1b) is assumed locally linear; integrating the nonlinear dynamics (1a) is often more critical. In ocean and atmosphere estimations, the determination of efficient measurement models is important. Depending on the sensor and state variables, they can be simple and straightforward, as a link between the heat equation and temperature data, or complex and indefinite, as a link between coupled physical–biological equations and remotely sensed data (e.g., ocean color, surface height, or temperature). By nature, (1b) is a well-posed mapping, from a large state space to a usually much smaller data space; but its inverse is not mathematically defined by (1b). For the inverse to be well posed, an additional data–dynamics melding criterion is required. The dynamical constraints enhance the observations and vice versa; this dual feedback is essential.

Since the dynamical (1a) and measurement models (1b) are approximations, statistical estimation theory formulates stochastic hypotheses for their respective imperfections or errors. This defines the statistics of the true models (2a)–(2b),

$$\mathbf{d}\boldsymbol{\psi}^t = \mathbf{f}(\boldsymbol{\psi}^t, t)dt + \mathbf{dw}, \tag{2a}$$

$$\mathbf{d}_k = \mathbf{C}_k \boldsymbol{\psi}_k^t + \mathbf{v}_k. \tag{2b}$$

The definitions and hypotheses employed here allow parameter estimations and time-correlated model errors (appendix A). All nonzero mean phenomena are assumed included into (1a)–(1b). Once the truth is stochastically described, the *estimation or melding criterion* determines the respective influence of the dynamics and observations on the state estimate. A *DA system* consists of three components: a dynamical model, a measurement model with sensor data, and an estimation criterion. The data assimilation problem is to determine the best possible field estimate that is in accord with the dynamical and measurement models, within their

respective uncertainties. By "best" it is meant "in closest statistical accord with the truth." The notion of close accord is defined mathematically by the estimation criterion. Within the criterion, all constraints are weak a priori, but to ease computations some may be assumed strong, depending on the relative estimated accuracy of each constraint. For instance, the nonlinear dynamical model can be considered either as a strong or weak constraint (e.g., Reid 1968; Sasaki 1970; Daley 1991; Bennett 1992; Wunsch 1996). Finally, it must be remembered that the central notion of an optimal estimate is a function of the melding criterion and associated statistical hypotheses. The ultimate arbiter consists of using the optimal estimate for ocean prediction.

3. Focus and specific objectives

A comprehensive assimilation system is suited to most nonlinear oceanic–atmospheric phenomena and most data types and coverage. Even though weather and ocean modeling differ (dynamics, data and models available, constraints), the approach chosen is general. The main restriction is a focus on the synoptic/mesoscale circulation and processes. The assimilation period is left arbitrary; it could be days to years. It is only assumed that the observations are time–space adequate with regard to the predictability limits of the phenomena of interest. This notion of adequate is subtle since even for known predictability limits and validated models (1a)–(1b), data sufficiency is still a function of the assimilation scheme. The better the scheme, the fewer the necessary data (e.g., Lorenc et al. 1991; Todling and Ghil 1994). The application considered in Part II mainly involves simulated temperature and salinity data for the control of shelfbreak front phenomena.

As far as specific objectives, the scheme should address model nonlinearities, estimate the uncertainty of the forecast with sufficient information on a posteriori errors, be suitable for assimilation in real time as well as for parallel computations, and allow adaptive filtering and parameter estimation. Most of the existing optimal schemes (section 4) cannot yet satisfy our practical objectives. The information, accuracy, and realistic goals may reduce the success of today's operational methods like the optimal interpolation (OI), commonly used in weather forecasting (Bengtsson et al. 1981; Ghil 1989; Lorenc et al. 1991) and in real-time, at-sea ocean prediction (Robinson et al. 1996a,b). One would like to improve the existing nowcast/forecast capabilities and hopefully increase general understanding; data assimilation should feed back to fundamental science.

4. Constraints and issues

The assimilation problem of section 2 is a nonlinear statistical estimation problem. Existing nonlinear schemes have first been analyzed with regard to our goals. In this text, it is simply our intention to focus on

---

[1] From among the terms used in the literature, "measurement model" was preferred. Observation model, measurement relation, or functional are also used (Gelb 1974; Catlin 1989; Daley 1991; Bennett 1992).

the essentials of nonlinear schemes and issues, and provide references for more complete discussions. Within estimation theory, *Bayesian estimation* and *maximum likelihood estimation* are the common approaches (e.g., Jazwinski 1970; Gelb 1974; Lorenc 1986; Boguslavskij 1988). In control theory, the DA problem is seen as a deterministic optimization issue, and for quadratic cost functions it amounts to *weighted least squares estimation* (Le Dimet and Talagrand 1986; Tarantola 1987; Sundqvist 1993). Statistical assumptions can be implicitly associated with all approaches (Robinson et al. 1998a). The discussion to follow involves Bayesian ideas.

The Bayesian conditional mean is the optimal estimator with respect to several cost functionals for arbitrary statistics. In this study, it is chosen as the optimal estimate. For nonlinear systems, it depends on all moments of the conditional density function, hence, on an infinite number of parameters. Solving for the conditional probability density governed by the Fokker–Planck or Kushner's equation (Jazwinski 1970) is today an informative guide with simple systems (Miller et al. 1994; Miller et al. 1998). In real ocean–atmosphere nonlinear estimation, the aim is to approximate the quantities of primary interest, the state, and its uncertainty. Taylor expansions and local linear extensions yield the common approximate schemes: the linearized, extended, higher-order and iterated Kalman filters/smoothers (KF/KS) and associated simplifications (e.g., Boguslavskij 1988). They provide an estimate of the uncertainty, but their truncated expansion may diverge (Evensen 1993, 1994a,b) and require frequent reinitialization. Most of the control and weighted least squares methods were derived for linear systems but can be iterated locally to solve nonlinear generalized inverse problems (e.g., Bennett et al. 1996; Bennett et al. 1997). The representer method considers model errors and minimizes the cost function in the data space but can be as costly as the estimation schemes if a posteriori state error covariance estimates are required. Direct global minimizations (e.g., Robinson et al. 1998a) are alternatives but a physically realizable solution is not assured. To limit expensive model integrations, a good first guess is required. In fact, the convergence of the iterated weighted least squares schemes is not yet proven (Evensen and van Leeuwen 1996). Smoothing prior to the minimization appears necessary (section 7).

The main advantages of all the above methods are the update and dynamical forecast of the error covariance. Even with linear models, forecast errors can differ considerably from the ones currently prescribed in OI (Parrish and Cohn 1985; Cohn and Parrish 1991; Miller and Cane 1989; Todling and Ghil 1990; Daley 1992a–c). Nonetheless, this error forecast and update is very expensive. For nonlinear systems, the required dimension is infinite. For discrete linear systems with n degrees of freedom, one needs $O(n^2)$ numbers to represent the error covariance, with n of $O(10^5$–$10^6)$ or more.

Even if the real-time aim is relaxed, today's parallel computers cannot manage such sizes. Since most schemes are only optimal for linear models, several sometimes conflicting hypotheses have been made to derive practical/operational reduced schemes (e.g., Parrish and Cohn 1985; Lorenc et al. 1991; Toth and Kalnay 1993; Todling and Cohn 1994), hence leading to controversies.

There are two constraints that a DA system needs to address. First, the dimension of the full error model associated with (2a)–(2b) is too large. Since less is known about errors than about dynamics, a careful reduction is necessary. The a priori hypotheses should be limited. With experience, for specific regions and processes, and particular data properties, some of today's hypotheses (e.g., Todling and Cohn 1994; appendix B) may be validated for use as a priori information. Even though the data coverage, type, and quality have increased in the past decades, the second constraint is the limited data sets. This concern is of special relevance in oceanography (Robinson et al. 1998a). To optimize the assimilation, it is very important to utilize at once all information contained in the few observations available. In summary, the issues are (a) how to reduce the size of the error model while explaining as accurately as possible the error structure and amplitude (many degrees of freedom, limited resources) and (b) how to optimize the extraction of reliable information from observations limited both in type and coverage.

5. Error subspace statistical estimation

Estimation criteria that address the objectives and questions raised in sections 3–4 are now identified. The approach is dynamic, reflecting basic properties of oceanic–atmospheric systems. Based on essential characteristics of geophysical measurements and models, the first property argues that efficient bases to describe, respectively, the dynamics, the variations of variability, and the errors, exist and are meaningful (section 5a). Section 5a is detailed since confusions between these evolving bases lead to controversies. The second property relates to the correlations between geophysical fields (section 5b). These facts are then used to determine the optimal representation of the error and a criterion for data–dynamics melding, leading to the ESSE concept (section 5c).

a. Efficient basis for describing variations of variability and error subspace

Even though it still needs to be rigorously proven that many observed geophysical phenomena can be associated with phase space attractors of low, finite dimension (West and Mackey 1991; Osborne and Pastorello 1993), statistical data analysis at least infers that most geophysical phenomena are colored random processes, for example, red in time and space. Nonetheless, some

field observations (e.g., satellite imagery) exemplify geophysical features (Monin 1974; Robinson 1989) at most energetic scales.[2] These structures occur intermittently, with strong similarities between occurrences. Several form anisotropic, nonhomogeneous, multiscale but coherent dynamical fields. Hence, most observed geophysical phenomena develop structures at many scales and have colored spectra. At the dynamical modeling end, many driven-dissipative systems as well as nonlinear conservative systems have been shown to possess "attractors" of finite dimension (e.g., Bergé et al. 1988; Osborne and Pastorello 1993). The existence of a global attractor for the Navier–Stokes equations has been proven in two dimensions (Foias and Teman 1977) and in three dimensions for remaining smooth fields (Foias and Teman 1987), with the dimension of the attractor being a function of the Reynolds number. As can also be deduced from Kolmogorov's physical principles (Kolmogorov 1941), the dynamical systems approach thus implies that the number of degrees of freedom necessary to describe most synoptic/mesoscale-to-large-scale turbulent geophysical flows is limited (Teman 1991).

Equation (1a) is of high dimension more for numerical accuracy than for physical variability dimensionality. The above-mentioned observations's/models's dual properties imply that the dominant geophysical variability consists of dynamical patterns and structures, with, in general, a colored spectrum that can be efficiently described at each instant by a limited number of functions or modes. The time–space physical nature of these functions evolves with the system's position in the phase space and local structure of the attractor, if it exists (e.g., Anderson 1996). In practice, the ideal choice of functions is a concern; it defines the evolving dynamics subspace. One aim is to reduce the number of functions to a minimum while still describing most of the variability of interest. Common techniques are dynamical normal modes or singular vectors, empirical modes or EOFs (Lorenz 1965; Davis 1977b; Weare and Nasstrom 1982; Wallace et al. 1992; von Storch and Frankignoul 1998), principal oscillation and interaction patterns (POPs and PIPs) (Hasselmann 1988; von Storch et al. 1988; Penland 1989; Schnur et al. 1993), and radial functions and wavelets (e.g., Gamage and Blumen 1993).

The leap from the above reduction of variability to an evolving efficient reduction of the error model (section 4) is now made in three successive steps. The model errors and data available are first assumed null; that is, $\mathbf{dw} = 0$ and $\mathbf{d}_k = \mathbf{v}_k = 0$ for $k > 0$ (appendix A). The dynamics of the true and model systems then only differ because of the predictability error. The uncertainties are

a subset of the local variations of variability, which have structural and spectral properties analogous to those of the dynamics subspace. If model errors and data are nonexistent, the dominant uncertainties can be described by a finite set of state-space vectors and the additional error variance explained by each new vector decays rapidly (e.g., hyperbolic, exponential, or power decay).

Observations are now considered. The data type and coverage, and their evolutions, influence the estimate's uncertainty but the dominant errors are still variations of geophysical variability. Only the nature and physical location of the dominant uncertainty depends on the phenomena and scales that are (or are not) controlled by observations. For instance, if energetic variations of variability are controlled at a given time by high quality sensors, they are not dominant errors at that time. Less energetic variations not controlled by measurements will be. The data properties thus do not alter the conclusion of the former paragraph. It is now argued that model errors do not affect this conclusion either. Using the conditional mean $\hat{\boldsymbol{\psi}}^t$ for the prediction and subtracting it from (2a), a forecast error has two components, the predictability and model errors,

$$\mathbf{d}\boldsymbol{\psi}^t - \mathbf{d}\hat{\boldsymbol{\psi}}^t = [\mathbf{f}(\boldsymbol{\psi}^t, t) - \hat{\mathbf{f}}(\boldsymbol{\psi}^t, t)]\, dt + \mathbf{dw}, \quad (3)$$

where $\hat{\mathbf{f}}$ is the expected value of $\mathbf{f}$ (appendix A). For meaningful modeling, the ratio of model to predictability errors should not be much larger than one. If at any given time model errors have amplitudes similar to predictability errors, they represent energetic physical processes not captured by the deterministic model (1a). When model errors are important, they are thus local variations of variability and the previous structured, power decay property applies. Model error covariances $\mathbf{Q}(t)$ can thus be assumed to have a limited number of dominant modes.[3] Combining the three conclusions, a time-evolving, limited dimension subspace contains most of the error. It is influenced by dynamics, data and model errors.

b. Correlations in geophysical systems

The limited datasets issue is addressed by the multiscale, multivariate correlations between geophysical field variations. For instance, dynamical and statistical studies show that a tracer transect is related to the reference velocity in an ocean basin (Wunsch 1988); the type and strength of local precipitation informs us about remote weather; an El Niño event can lead to abnormal conditions at other times/places; the surface temperature in the Gulf Stream at a given time implies some of the deeper water properties; the upwelling/downwelling variations along a coast relate to the coastal distribution

---

[2] In the present study, the term energy can also refer to a pseudoenergy or field squared amplitude.

[3] Relations between the dominant eigenvectors of $\mathbf{Q}(t)$ and dominant stochastic optimals are discussed in Lermusiaux (1997).

of nutrients; each fish species has optimum water properties; the sea color correlates with phytoplankton concentration; etc. These examples appear simple but in reality correlation issues are subtle due to the multiscale, multivariate, inhomogeneous, anisotropic, or nonstationary properties (e.g., McWilliams et al. 1986; Lorenc 1986, 1992; Daley 1991). In summary, 3D multivariate DA, in accord with the phenomena and time–space scales considered, is necessary. The expression *3D multivariate DA* indicates here that each datum instantaneously influences all state variables and scales that matter. This impact must be in accord with the evolving dynamics. For instance, surface altimeter fields are not always at once linked to subsurface properties (e.g., Martel and Wunsch 1993; Fukumori et al. 1993).

c. ESSE

The conditional mean was chosen for optimal estimate. For all statistics, minimizing the expectation of a convex measure of the estimate's uncertainty leads to that optimum (appendix B, section d). Approximate definitions of uncertainty and measures thereof lead to approximate estimates. Here, both notions are determined using sections 5a and 5b.

The estimate's uncertainty can be represented by an error covariance, a common convex statistical measure of error fields.[4] The first conclusion (section 5a) supports its truncation to a most energetic error subspace while the second (section 5b) implies that the melding criterion should be 3D multivariate. A reduced-rank approximation of the error covariance $\mathbf{P}$ at time t_k is thus optimal for a given rank p_k if it explains the maximum variance and structure of the multivariate $\mathbf{P}_k$ that is possible. For the structured, power-decay property (section 5a) there is a relatively small number p_k for which the optimal reduction explains most of $\mathbf{P}_k$. Denoting for convenience this time-variant p_k by p, the associated reduction is called the principal error covariance $\mathbf{P}_k^p$. The difference between $\mathbf{P}_k$ and $\mathbf{P}_k^p$, or complementary error covariance $\mathbf{P}_k^c$,

$$\mathbf{P}_k^c \doteq \mathbf{P}_k - \mathbf{P}_k^p, \tag{4}$$

should have a minimal norm. An orthonormal decomposition of $\mathbf{P}_k^p$, $\mathbf{E}_k \mathbf{\Pi}_k \mathbf{E}_k^T$, with variance $\mathbf{\Pi}_k \in \mathbb{R}^{p \times p}$ and structure $\mathbf{E}_k \in \mathbb{R}^{n \times p}$, hence satisfies

$$\left\{ \mathbf{\Pi}_k \text{ and } \mathbf{E}_k, \text{ with rank}(\mathbf{E}_k) = p \; \middle| \; \min_{\mathbf{\Pi}_k, \mathbf{E}_k} \| \mathbf{P}_k^c = \mathbf{P}_k - \mathbf{E}_k \mathbf{\Pi}_k \mathbf{E}_k^T \| \right\}. \tag{5}$$

In the sense of any unitarily invariant norm (e.g., the

two-norm $\| \mathbf{P}_k^c \|_2$ and Frobenius norm $\| \mathbf{P}_k^c \|_F$), the optimum in (5) is the dominant rank-p singular value decomposition of $\mathbf{P}_k$ (Horn and Johnson 1985, 1991). The matrix $\mathbf{\Pi}_k$ is the ordered diagonal of dominant-p singular values and the columns of $\mathbf{E}_k$ are the associated singular vectors. Since $\mathbf{P}_k$ is positive semidefinite, $\mathbf{E}_k \mathbf{\Pi}_k \mathbf{E}_k^T$ is also its rank-p eigendecomposition. The columns of $\mathbf{E}_k$ form a basis for the 3D multivariate error subspace; $\mathbf{\Pi}_k$ is the error subspace covariance. At a given time, t_k, and for a given p, these matrices characterize the error subspace (ES). They answer the first issue raised in section 4.

With the ES theoretically defined by (5) at all t_k, only the error measure in this ES still needs to be chosen. Several exist (e.g., Horn and Johnson 1985), but the most logical is the Euclidean one. Combining the criterion (5) with the Euclidean minimum error variance approach leads to the present notion of ESSE. Data and dynamics are melded such that the a posteriori ES variance is minimized. The estimation criteria are

$$\left\{ \hat{\boldsymbol{\psi}}_k \; \middle| \; \min_{\hat{\boldsymbol{\psi}}_k} J_k = \text{tr}[\mathbf{P}_k^p(+)] \text{ using } [\mathbf{d}_k] \right\} \tag{6a}$$

for objective analysis (OA) via ESSE at t_k;

$$\left\{ \hat{\boldsymbol{\psi}}_k \; \middle| \; \min_{\hat{\boldsymbol{\psi}}_k} J_k = \text{tr}[\mathbf{P}_k^p(+)] \text{ using } [\mathbf{d}_0, \ldots, \mathbf{d}_k] \right\}, \tag{6b}$$

for filtering via ESSE at t_k; and

$$\left\{ \hat{\boldsymbol{\psi}}_{k/N} \; \middle| \; \min_{\hat{\boldsymbol{\psi}}_{k/N}} J_k = \text{tr}[\mathbf{P}_k^p(+)] \text{ using } [\mathbf{d}_0, \ldots, \mathbf{d}_N] \right\}, \tag{6c}$$

for smoothing via ESSE at t_k within the data interval $[t_0, t_N]$; where each of (6a)–(6c) are subject to the dynamical and measurement model constraints (2a)–(2b). To distinguish between a priori and a posteriori quantities, the symbols $(-)$ and $(+)$ are employed (see appendix A for more on notation). Except for the ignored small errors, criteria (6a)–(6c) follow the Bayesian minimum error variance nonlinear estimation. They lead to efficient, 3D multivariate analyses since the meldings occur within the ES. The second question raised in section 4 is answered. The criteria (6a)–(6c) also relate to the efficient concept of "minimax assimilation": the maximum errors, here in the Euclidean sense, are minimized. The general goal of ESSE is to determine the considered ocean–atmosphere state evolution by minimizing the dominant errors, in accord with the full dynamical and measurement models, and their respective uncertainties. Of course, the criteria (6a)–(6c) are only theoretical. In practice, efficient schemes for finding (6a) and tracking (6b)–(6c) the ES have to be determined (sections 6, 7).

Objective analysis via ESSE (6a) or "fixed-time ESSE" emphasizes the general applicability of the approach. In fact, Anderson (1996) has shown that for the Lorenz model (Lorenz 1963) the projection of classical OA error correlation onto the local attractor sheet is the

---

[4] In this text, the term covariance relates to a matrix quantity. The covariances are dimensional but all eigendecompositions or SVDs are made on nondimensional fields so that the ordering of eigen- or singular values is unit independent. To simplify notations, the normalization is presented in appendix B, section a.

TABLE 1. Filtering via ESSE at t_k: Continuous–discrete problem statement.

		Equation
Dynamical model:	$d\boldsymbol{\psi}^t = \mathbf{f}(\boldsymbol{\psi}^t, t)\, dt + d\mathbf{w}, \quad \text{with } \hat{\boldsymbol{\psi}}_0 = \boldsymbol{\Psi}_0,$	(2a)
Measurement model:	$\mathbf{d}_k = \mathbf{C}_k \boldsymbol{\psi}_k^t + \mathbf{v}_k.$	(2b)
Error subpace:	$\left\{ \boldsymbol{\Pi}_k,\, \mathbf{E}_k \text{ with } \mathbf{P}_k^p = \mathbf{E}_k \boldsymbol{\Pi}_k \mathbf{E}_k^\mathrm{T} \text{ and } \mathrm{rank}(\mathbf{E}_k) = p \,\middle|\, \min_{\boldsymbol{\Pi}_k, \mathbf{E}_k} \|\mathbf{P}_k - \mathbf{P}_k^p\| \right\}.$	(5)
ES melding criterion:	$\left\{ \hat{\boldsymbol{\psi}}_k \,\middle|\, \min_{\hat{\boldsymbol{\psi}}_k} J_k = \mathrm{tr}[\mathbf{P}_k^p(+)] \text{ using } [\mathbf{d}_0, \ldots, \mathbf{d}_k] \right\}.$	(6b)

most effective selection of initial conditions for ensemble forecasts. This conclusion has also been exemplified in primitive-equation (PE) modeling (Lermusiaux 1997). Since the ES is a subset of the local variations of variability, (6a) is the most efficient analysis for given resources and in the Euclidean framework. In Part II of this study, the focus is on filtering via ESSE (6b). The problem statement is given in Table 1.

Section 6 outlines filtering via ESSE schemes using a Monte Carlo approach; dynamical systems for adaptive ESSE are also derived in Lermusiaux (1997). The smoothing via ESSE problem statement is as in Table 1, but with criterion (6c) replacing (6b). For linear models (2a), this corresponds to the generalized inverse problem restricted to the dominant errors. Such smoothing schemes with a posteriori error estimates are obtained in section 7. The five main components of the present ESSE system (initialization, field and ES nonlinear forecasts, minimum variance within the ES, and smoothing) are illustrated in Fig. 1.

Some general properties are now discussed. First, in (6a)–(6c) the data only correct the most erroneous parts of the forecast. The accurate portion of the forecast is corrected by dynamical adjustments and interpolations. Second, the ES in (5) is time dependent. The reduction to the principal error covariance is dynamic. The ES tracks the scales and processes where the dominant er-

rors occur. The time rate of change of the ES is a function of the (i) initial uncertainty conditions; (ii) evolving model errors; and (iii) data type, quality, and coverage; and of the nonlinear, interactive evolution of these three components. All these factors influence the nature of the ES (e.g., multiscale, anisotropic, homogeneous, or not). In fact, the successes of the optimal perturbations (OP) to determine initial conditions for ensemble forecasting (e.g., Mureau et al. 1993; Toth and Kalnay 1993; Molteni and Palmer 1993; Molteni et al. 1996) can be improved by quantitatively taking into account data quality and model errors. The OP spectrum and structures, which only consider the predictability error, should be modified accordingly, especially in oceanography. ESSE provides a theoretical framework to do so. Third, the error and dynamics subspaces in general differ. Fourth, the statistical estimation of the ES (sections 6, 7) yields the notion of error EOFs. In fact, there are several quantitative definitions for the ES, each associated with slightly modified criteria (6a)–(6c): for example, singular or normal error modes; extended, complex, or frequency error EOFs; error POPs and PIPs; and synoptic and wavelet-based ES. If the maximum norm had been chosen in (5), a maximum measure would have been logical in (6a)–(6c). These particular formulations are not discussed further. Priority is given to the central concept, common to all representations. Fifth, the ideal ES dimension p (e.g., appendix B, section b) evolves in time, in accord with the dynamics, model errors, and available data. It is only for statistically stationary data, dynamics, and model errors that p should stay constant. In all schemes of sections 6 and 7, the size of the ES is thus time dependent.
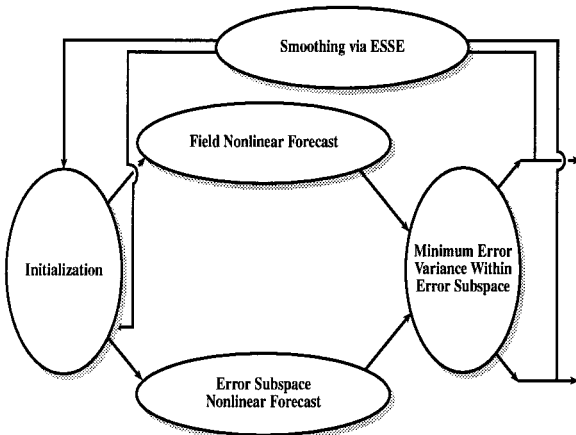
6. Filtering via ESSE schemes

A recursive scheme is now derived for filtering in the ES (5) corresponding to the models (2a)–(2b). One needs to track the ES evolution, which is not trivial. The two-step root of the algorithm consists of a forecast–data melding when data are available (section 6a), followed by the dynamical state and ES forecasts to the next assimilation time (section 6b). It is assumed that an estimate of the conditional mean state and associated ES have been integrated from t_0 to t_k using (1)–(2)

FIG. 1. The five main components of the present ESSE system.

Table 2. Eigendecomposition of the minimum error variance linear update (index k omitted).

		Equation
Dynamical state update:	$\hat{\boldsymbol{\psi}}(+) = \hat{\boldsymbol{\psi}}(-) + \mathbf{K}[\mathbf{d} - \mathbf{C}\hat{\boldsymbol{\psi}}(-)]$	(8)
Optimal gain:	$\mathbf{K} = \mathbf{U}_-\boldsymbol{\Lambda}(-)\tilde{\mathbf{C}}^{\mathrm{T}}[\tilde{\mathbf{C}}\boldsymbol{\Lambda}(-)\tilde{\mathbf{C}}^{\mathrm{T}} + \mathbf{R}]^{-1}, \quad \text{with } \tilde{\mathbf{C}} \doteq \mathbf{C}\mathbf{U}_-$	(12)
Error eigenvalue update:	$\mathbf{H}\boldsymbol{\Lambda}(+)\mathbf{H}^{\mathrm{T}} = \tilde{\boldsymbol{\Lambda}}(+) = \boldsymbol{\Lambda}(-) - \boldsymbol{\Lambda}(-)\tilde{\mathbf{C}}^{\mathrm{T}}[\tilde{\mathbf{C}}\boldsymbol{\Lambda}(-)\tilde{\mathbf{C}}^{\mathrm{T}} + \mathbf{R}]^{-1}\tilde{\mathbf{C}}\boldsymbol{\Lambda}(-)$	(14)
Error eigenvector update:	$\mathbf{U}_+ = \mathbf{U}_-\mathbf{H}$	(16)

and (6b) for $[\mathbf{d}_0, \ldots, \mathbf{d}_{k-1}]$. Hence, $\hat{\boldsymbol{\psi}}_k(-)$, $\mathbf{E}_k(-)$, and $\boldsymbol{\Pi}_k(-)$ are available. Specifics and variations of the algorithm are discussed in appendix B.

a. ES melding or ES analysis scheme

As in most existing schemes, the data–forecast melding is chosen linear a priori. This causes a departure from the strict Bayesian approach (e.g., Jazwinski 1970), which would solve (6a)–(6c) without making this simplification. Nonetheless, the melding weights are determined using criterion (6b) in the nonlinearly evolved ES (5). To simplify notations, the index k is omitted. The minimum error covariance melding is first decomposed exactly in terms of error eigenvalues and eigenvectors [section 6a(1)]. Its ES truncation is given in Lermusiaux (1997). In section 6a(2), the sample or empirical ES melding is derived.

1) Eigendecomposition of the minimum error variance linear update

Since linear melding is enforced, determining the optimum of

$$\min_{\hat{\boldsymbol{\psi}}^t} [\hat{\boldsymbol{\psi}}(-) - \boldsymbol{\psi}^t]^{\mathrm{T}}\mathbf{P}^{-1}(-)[\hat{\boldsymbol{\psi}}(-) - \boldsymbol{\psi}^t]$$
$$+ (\mathbf{d} - \mathbf{C}\boldsymbol{\psi}^t)^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{d} - \mathbf{C}\boldsymbol{\psi}^t) \quad (7)$$

is equivalent to minimizing the trace of $\mathbf{P}(+)$ (e.g., Gelb 1974). Taking the derivative of (7) with respect to $\boldsymbol{\psi}^t$ yields the melded estimate

$$\hat{\boldsymbol{\psi}}(+) = \hat{\boldsymbol{\psi}}(-) + \mathbf{K}[\mathbf{d} - \mathbf{C}\hat{\boldsymbol{\psi}}(-)] \quad (8)$$

and updated error covariance

$$\mathbf{P}(+) = \mathbf{P}(-) - \mathbf{K}\mathbf{C}\mathbf{P}(-), \quad (9)$$

where $\mathbf{K}$ is the Kalman gain

$$\mathbf{K} = \mathbf{P}(-)\mathbf{C}^{\mathrm{T}}[\mathbf{C}\mathbf{P}(-)\mathbf{C}^{\mathrm{T}} + \mathbf{R}]^{-1}. \quad (10)$$

Note that (8) contains the data update of all boundary variables and external forcings since they are assumed to be part of (1a)–(2a). Introducing the eigendecompositions of the a priori and a posteriori error covariances (see appendix A for notation),

$$\mathbf{P}(-) = \mathbf{U}_-\boldsymbol{\Lambda}(-)\mathbf{U}_-^{\mathrm{T}},$$
$$\mathbf{P}(+) = \mathbf{U}_+\boldsymbol{\Lambda}(+)\mathbf{U}_+^{\mathrm{T}}, \quad (11)$$

with $\mathbf{U}_-\mathbf{U}_-^{\mathrm{T}} = \mathbf{U}_-^{\mathrm{T}}\mathbf{U}_- = \mathbf{U}_+\mathbf{U}_+^{\mathrm{T}} = \mathbf{U}_+^{\mathrm{T}}\mathbf{U}_+ = \mathbf{I}$, into (8)–(10), the Kalman gain and error covariance update are exactly rewritten, respectively, as

$$\mathbf{K} = \mathbf{U}_-\boldsymbol{\Lambda}(-)\tilde{\mathbf{C}}^{\mathrm{T}}[\tilde{\mathbf{C}}\boldsymbol{\Lambda}(-)\tilde{\mathbf{C}}^{\mathrm{T}} + \mathbf{R}]^{-1}, \quad (12)$$

$$\mathbf{U}_+\boldsymbol{\Lambda}(+)\mathbf{U}_+^{\mathrm{T}} = \mathbf{U}_-\{\boldsymbol{\Lambda}(-) - \boldsymbol{\Lambda}(-)\tilde{\mathbf{C}}^{\mathrm{T}}$$
$$\times [\tilde{\mathbf{C}}\boldsymbol{\Lambda}(-)\tilde{\mathbf{C}}^{\mathrm{T}} + \mathbf{R}]^{-1}\tilde{\mathbf{C}}\boldsymbol{\Lambda}(-)\}\mathbf{U}_-^{\mathrm{T}}$$
$$\doteq \mathbf{U}_-\tilde{\boldsymbol{\Lambda}}(+)\mathbf{U}_-^{\mathrm{T}}, \quad (13)$$

where $\tilde{\mathbf{C}} \doteq \mathbf{C}\mathbf{U}_-$. The eigendecomposition of the nonnegative definite $\tilde{\boldsymbol{\Lambda}}(+)$ yields

$$\tilde{\boldsymbol{\Lambda}}(+) = \boldsymbol{\Lambda}(-) - \boldsymbol{\Lambda}(-)\tilde{\mathbf{C}}^{\mathrm{T}}[\tilde{\mathbf{C}}\boldsymbol{\Lambda}(-)\tilde{\mathbf{C}}^{\mathrm{T}} + \mathbf{R}]^{-1}\tilde{\mathbf{C}}\boldsymbol{\Lambda}(-)$$
$$= \mathbf{H}\boldsymbol{\Lambda}(+)\mathbf{H}^{\mathrm{T}}, \quad (14)$$

where $\boldsymbol{\Lambda}(+)$ is diagonal and the columns of $\mathbf{H} \in \mathbb{R}^{n \times n}$ are a set of orthonormal eigenvectors for $\tilde{\boldsymbol{\Lambda}}(+)$. Hence,

$$\mathbf{P}(+) = \mathbf{U}_+\boldsymbol{\Lambda}(+)\mathbf{U}_+^{\mathrm{T}} = \mathbf{U}_-\mathbf{H}\boldsymbol{\Lambda}(+)\mathbf{H}^{\mathrm{T}}\mathbf{U}_-^{\mathrm{T}}. \quad (15)$$

The a posteriori diagonal matrix of error eigenvalues $\boldsymbol{\Lambda}(+)$ is given by (14) and the a posteriori error eigenvectors by

$$\mathbf{U}_+ = \mathbf{U}_-\mathbf{H}. \quad (16)$$

The a posteriori error covariance derives from (15). For linear melding, the state (8), (12) and error eigenupdates (14), (16) are the minimum error variance estimates (Table 2).

2) Minimum sample ES variance linear update

In this section, a *sample* ES forecast described by $\mathbf{E}_-$ and $\boldsymbol{\Pi}(-)$ is assumed available (section 5b). The data-forecast melding within the sample ES is outlined, with updates of the fields and ES covariance. For the details, we refer to (Lermusiaux 1997).

(i) Dynamical state update

The field update can be derived either by truncation of (8) and (12) to the sample ES or by minimization analogous to (7), but within the ES. One gets

$$\hat{\boldsymbol{\psi}}(+) = \hat{\boldsymbol{\psi}}(-) + \mathbf{K}^p[\mathbf{d} - \mathbf{C}\hat{\boldsymbol{\psi}}(-)], \quad (17)$$

$$\mathbf{K}^p = \mathbf{P}^p(-)\mathbf{C}^{\mathrm{T}}[\mathbf{C}\mathbf{P}^p(-)\mathbf{C}^{\mathrm{T}} + \mathbf{R}]^{-1},$$
$$= \mathbf{E}_-\boldsymbol{\Pi}(-)\tilde{\mathbf{C}}^{p\mathrm{T}}[\tilde{\mathbf{C}}^p\boldsymbol{\Pi}(-)\tilde{\mathbf{C}}^{p\mathrm{T}} + \mathbf{R}]^{-1} \doteq \mathbf{E}_-\tilde{\mathbf{K}}^p, \quad (18)$$

where $\tilde{\mathbf{C}}^p \doteq \mathbf{CE}_-$. For adequate a priori sampling, that is, $\mathbf{P}^p(-) = \mathbf{E}_-\mathbf{\Pi}(-)\mathbf{E}_-^T$ converges to $\mathbf{P}(-)$, (17)–(18) converge to (8) and (12) at the infinite sample limit.

(ii) Sample ES update

The derivation of the ES covariance update requires care since the present ES forecast (section 5b) is obtained from an ensemble forecast. As discussed in (Evensen 1997b; Lermusiaux 1997; Burgers et al. 1998), the original ensemble update algorithm (e.g., Evensen 1994a) underestimates a posteriori errors. The two ES algorithms outlined next give a correct error estimate at the infinite ensemble limit. The scheme A directly estimates $\mathbf{\Pi}(-)$ and $\mathbf{E}_+$. The scheme B updates the SVD of the ensemble spread.

Scheme A: Update of the sample ES covariance. An ensemble of *q unbiased* dynamical states is denoted by $\hat{\boldsymbol{\psi}}^j(-)$, $j = 1, \ldots, q$. The associated a priori and a posteriori error sample matrices, $\mathbf{M}(-)$ and $\mathbf{M}(+) \in \mathbb{R}^{n \times q}$, are[5]

$$\mathbf{M}(-) \doteq [\hat{\boldsymbol{\psi}}^j(-) - \hat{\boldsymbol{\psi}}(-)], \qquad (19\text{a})$$

$$\mathbf{M}(+) \doteq [\hat{\boldsymbol{\psi}}^j(+) - \hat{\boldsymbol{\psi}}(+)]. \qquad (19\text{b})$$

These expressions are matrices whose column *j* consists of the ensemble member *j* minus the mean estimates $\hat{\boldsymbol{\psi}}(-)$ and $\hat{\boldsymbol{\psi}}(+)$, respectively. The update of the sample error covariance, $\mathbf{P}^s \doteq \mathbf{MM}^T/q$, is now obtained. Denoting by $\mathbf{d}^j$ a set of *q* data vectors perturbed with noise of zero mean and covariance $\mathbf{R}$, the updates of the ensemble (e.g., Burgers et al. 1998) and conditional mean estimate are, respectively,

$$\hat{\boldsymbol{\psi}}^j(+) = \hat{\boldsymbol{\psi}}^j(-) + \mathbf{K}^s[\mathbf{d}^j - \mathbf{C}\hat{\boldsymbol{\psi}}^j(-)], \quad (20\text{a})$$

$$\hat{\boldsymbol{\psi}}(+) = \hat{\boldsymbol{\psi}}(-) + \mathbf{K}^s[\mathbf{d} - \mathbf{C}\hat{\boldsymbol{\psi}}(-)], \qquad (20\text{b})$$

where the gain $\mathbf{K}^s$ has to be optimized. Subtracting (20b) from (20a) and using (19a)–(19b),

$$\mathbf{M}(+) = (\mathbf{I} - \mathbf{K}^s\mathbf{C})\mathbf{M}(-) + \mathbf{K}^s\overline{\mathbf{V}}. \qquad (21)$$

The columns of $\overline{\mathbf{V}} = [\mathbf{v}^j] = [\mathbf{d}^j - \mathbf{d}] \in \mathbb{R}^{m \times q}$ are realizations of the random processes $\mathbf{v}$. The update of the sample error covariance derives from (21)

$$\mathbf{P}^s(+) = (\mathbf{I} - \mathbf{K}^s\mathbf{C})\mathbf{P}^s(-)(\mathbf{I} - \mathbf{K}^s\mathbf{C})^T + \mathbf{K}^s\mathbf{R}^s\mathbf{K}^{sT}$$
$$+ (\mathbf{I} - \mathbf{K}^s\mathbf{C})\mathbf{\Omega}^s\mathbf{K}^{sT} + \mathbf{K}^s\mathbf{\Omega}^{sT}(\mathbf{I} - \mathbf{K}^s\mathbf{C})^T, \quad (22)$$

where $\mathbf{R}^s \doteq (1/q)\overline{\mathbf{V}}\overline{\mathbf{V}}^T = \mathcal{E}^q\{\mathbf{v}^j\mathbf{v}^{jT}\}$ and $\mathbf{\Omega}^s \doteq (1/q)\mathbf{M}(-)$
$\overline{\mathbf{V}}^T = \mathcal{E}^q\{[\hat{\boldsymbol{\psi}}^j(-) - \hat{\boldsymbol{\psi}}(-)]\mathbf{v}^{jT}\}$. For the gain $\mathbf{K}^s$ minimizing the trace of $\mathbf{P}^s(+)$, $\mathbf{K}^s = \mathbf{P}^s(-)\mathbf{C}^T(\mathbf{C}\mathbf{P}^s(-)\mathbf{C}^T + \mathbf{R}^s)^{-1}$, one obtains

---

[5] For coherence with other works (e.g., Burgers et al. 1998), (19a,b) differ from the derivation of Lermusiaux (1997) which used true error sample matrices. The present scheme (A) still leads to the same results since the variance of the error incurred for using an estimate of the truth in (19a)–(19b) instead of the truth itself is of $O(1/q)$ (e.g., Bendat and Piersol 1986), which is the order of the truncated terms.

$$\mathbf{P}^s(+) = (\mathbf{I} - \mathbf{K}^s\mathbf{C})\mathbf{P}^s(-) + (\mathbf{I} - \mathbf{K}^s\mathbf{C})\mathbf{\Omega}^s\mathbf{K}^{sT}$$
$$+ \mathbf{K}^s\mathbf{\Omega}^{sT}(\mathbf{I} - \mathbf{K}^s\mathbf{C})^T. \qquad (23)$$

By hypotheses (appendix A), $\mathbf{\Omega}^s \to 0$ for $q \to \infty$. Neglecting the associated symmetric sum in (23) yields an estimate of $\mathbf{P}(+)$ of standard deviation error decay of $O(1/\sqrt{q})$,

$$\mathbf{P}^s(+) = (\mathbf{I} - \mathbf{K}^s\mathbf{C})\mathbf{P}^s(-). \qquad (24)$$

The sample ES update derives from the dominant rank-*p* reduction of (19)–(24). It is efficiently estimated based on the SVD of $\mathbf{M}(-)$ and unknown $\mathbf{M}(+)$, respectively,

$$\text{SVD}_p[\mathbf{M}(-)] = \mathbf{E}_-\mathbf{\Sigma}(-)\mathbf{V}_-^T, \qquad (25\text{a})$$

$$\text{SVD}_p[\mathbf{M}(+)] = \mathbf{E}_+\mathbf{\Sigma}(+)\mathbf{V}_+^T, \qquad (25\text{b})$$

where the operator $\text{SVD}_p(\cdot)$ selects the dominant rank-*p* SVD. After melding, the *p* dominant left singular vectors, columns of $\mathbf{E}_+$, form the ordered basis for the ES of dimension $p \leq q$. The corresponding singular values yield the diagonals $\mathbf{\Pi}_k(-)$ and $\mathbf{\Pi}_k(+)$,

$$\mathbf{\Pi}(-) = \frac{1}{q}\mathbf{\Sigma}^2(-), \qquad (26\text{a})$$

$$\mathbf{\Pi}(+) = \frac{1}{q}\mathbf{\Sigma}^2(+). \qquad (26\text{b})$$

Performing computations similar to (21)–(24), but starting from (25)–(26), leads to the optimal gain (18) and to the equations for the ES update. Using the orthogonality of singular vectors, these are

$$\tilde{\mathbf{\Pi}}(+) = \mathbf{E}_-^T\mathbf{P}^p(+)\mathbf{E}_- = (\mathbf{I}^p - \tilde{\mathbf{K}}^p\tilde{\mathbf{C}}^p)\mathbf{\Pi}(-)$$
$$= \mathbf{H}\mathbf{\Pi}(+)\mathbf{H}^T, \qquad (27)$$

$$\mathbf{E}_+ = \mathbf{E}_-\mathbf{H}. \qquad (28)$$

The columns of $\mathbf{H}$ in (27) are ordered orthonormal eigenvectors of $\tilde{\mathbf{\Pi}}(+)$, projection of $\mathbf{P}^p(+)$ onto the columns of $\mathbf{E}_-$. The corresponding eigenvalues form the diagonal $\mathbf{\Pi}(+)$ and $\mathbf{I}^p \in \mathbb{R}^{p \times p}$ is the identity matrix. With (18) and (24)–(25), the columns of $\mathbf{E}_-$ and $\mathbf{E}_+$ in (28) span the same space. Table 3 summarizes the sample ES scheme. Algorithmic and computing issues are discussed in appendix B, section e.

In this scheme A, the ensemble update (20a) is not carried out. It was only utilized to derive the ES covariance update. Only one melding (17) is necessary. The number *p* of significant error EOFs or singular vectors is smaller than the ensemble size *q,* which reduces computations in Table 3 (appendix B, section c). Of course, for efficient ensemble forecasts, *q* should not be much larger than *p* (i.e., at most, an order of magnitude larger).

Scheme B: Update of the SVD of the ensemble spread. A disadvantage of (27)–(28) is that the information contained in the right singular vectors of (25a) is lost. Once $\mathbf{E}_+$, $\mathbf{\Pi}(+)$, and $\hat{\boldsymbol{\psi}}(+)$ are computed, an adequate ensemble still has to be constructed [section 6b(2)]. Hence,

Table 3. Minimum sample ES variance linear update (index k omitted).

		Equation
Dynamical state update:	$\hat{\boldsymbol{\psi}}(+) = \hat{\boldsymbol{\psi}}(-) + \mathbf{K}^p[\mathbf{d} - \mathbf{C}\hat{\boldsymbol{\psi}}(-)]$	(17)
Sample ES optimal gain:	$\mathbf{K}^p = \mathbf{E}_-\mathbf{\Pi}(-)\tilde{\mathbf{C}}^{p^{\mathrm{T}}}[\tilde{\mathbf{C}}^p\mathbf{\Pi}(-)\tilde{\mathbf{C}}^{p^{\mathrm{T}}} + \mathbf{R}]^{-1} \doteq \mathbf{E}_-\tilde{\mathbf{K}}^p,$	(18)
	with $\tilde{\mathbf{C}}^p \doteq \mathbf{C}\mathbf{E}_-,$　　$\mathrm{SVD}_p[\mathbf{M}(-)] = \mathbf{E}_-\mathbf{\Sigma}(-)\mathbf{V}_-^{\mathrm{T}}$　and　$\mathbf{\Pi}(-) = \frac{1}{q}\mathbf{\Sigma}^2(-)$	
Sample ES covariance update:	$\mathbf{H}\mathbf{\Pi}(+)\mathbf{H}^{\mathrm{T}} = \tilde{\mathbf{\Pi}}(+) = (\mathbf{I}^p - \tilde{\mathbf{K}}^p\tilde{\mathbf{C}}^p)\mathbf{\Pi}(-)$	(27)
	$\mathbf{E}_+ = \mathbf{E}_-\mathbf{H}$	(28)

the complete rank-p SVD update is now sought. Using (25a)–(25b) to reduce (19a)–(19b), rewritten for clarity in a vector form [with $(\mathbf{A})^j$ denoting the column j of $\mathbf{A}$], one has

$$\hat{\boldsymbol{\psi}}^j(-) = \hat{\boldsymbol{\psi}}(-) + \mathbf{E}_-\mathbf{\Sigma}(-)(\mathbf{V}_-^{\mathrm{T}})^j, \qquad (29\text{a})$$

$$\hat{\boldsymbol{\psi}}^j(+) = \hat{\boldsymbol{\psi}}(+) + \mathbf{E}_+\mathbf{\Sigma}(+)(\mathbf{V}_+^{\mathrm{T}})^j. \qquad (29\text{b})$$

Deriving such an update (29b) based on the original ensemble algorithm (e.g., Evensen 1994a), Lermusiaux (1997) showed that a posteriori errors were underestimated; that is, that the terms $\mathbf{K}^s\overline{\mathbf{V}}$ in (21) and thus $\mathbf{K}^s\mathbf{R}^s\mathbf{K}^{s^{\mathrm{T}}}$ in (22) were missing in the original ensemble algorithm. The approach of Lermusiaux (1997) is now utilized but based on the modified ensemble update (20a)–(20b). This reduces the analysis of Burgers et al. (1998) to its significant subspace. The direct derivation evaluates the SVD of (21),

$$\mathbf{E}_+\mathbf{\Sigma}(+)\mathbf{V}_+^{\mathrm{T}} = \mathrm{SVD}_p[(\mathbf{I} - \mathbf{K}^s\mathbf{C})\mathbf{M}(-) + \mathbf{K}^s\overline{\mathbf{V}}]. \quad (30)$$

To reduce computations, the ensembles in the rhs of (30) can be truncated a priori to their significant rank-p SVDs. These are denoted by (25a) for the ensemble of states and by $\mathrm{SVD}_p(\overline{\mathbf{V}}) = \overline{\mathbf{V}}^p \in \mathbb{R}^{p \times p}$ for the perturbed data. With these rank-p approximations, using (25b) and the optimal gain (18), (21) reduces to, at $O(1/\sqrt{p})$ in standard deviation,

$$\mathbf{E}_+\mathbf{\Sigma}(+)\mathbf{V}_+^{\mathrm{T}} = (\mathbf{I} - \mathbf{K}^p\mathbf{C})\mathbf{E}_-\mathbf{\Sigma}(-)\mathbf{V}_-^{\mathrm{T}} + \mathbf{K}^p\overline{\mathbf{V}}^p. \quad (31)$$

Inserting $\mathbf{V}_-\mathbf{V}_-^{\mathrm{T}} = \mathbf{I}^p$ at the right of the second term in the rhs of (31) gives, with (18),

$$\mathbf{E}_+\mathbf{\Sigma}(+)\mathbf{V}_+^{\mathrm{T}} = \mathbf{E}_-[(\mathbf{I} - \tilde{\mathbf{K}}^p\tilde{\mathbf{C}}^p)\mathbf{\Sigma}(-)$$
$$+ \tilde{\mathbf{K}}^p\overline{\mathbf{V}}^p\mathbf{V}_-]\mathbf{V}_-^{\mathrm{T}}. \qquad (32)$$

In (32), computing the SVD of the term in bracket, already of rank p, leads to

$$\mathbf{F}\mathbf{\Sigma}(+)\mathbf{G}^{\mathrm{T}} \doteq \mathrm{SVD}_p[(\mathbf{I} - \tilde{\mathbf{K}}^p\tilde{\mathbf{C}}^p)\mathbf{\Sigma}(-) + \tilde{\mathbf{K}}^p\overline{\mathbf{V}}^p\mathbf{V}_-], \quad (33\text{a})$$

$$\mathbf{E}_+ = \mathbf{E}_-\mathbf{F}, \qquad (33\text{b})$$

$$\mathbf{V}_+ = \mathbf{V}_-\mathbf{G}. \qquad (33\text{c})$$

With (17)–(18), (33a)–(33c) update the complete rank-p SVD (29b), as summarized in Table 4. The advantage of (30) or (33a)–(33c) over (27)–(28) is the right singular vector update. The a posteriori states [(29b)] are physically balanced in the sense of the estimation criterion (section 2). If one uses Table 3, techniques to create an ensemble of a posteriori states from (27)–(28) are needed [section 6b(2)]. On the other hand, scheme A is an efficient procedure to compute the a posteriori ES covariance (27)–(28). It is in fact straightforward to show that the a posteriori rank-p sample error covariance obtained from (29b) and (33a)–(33c) is, at $O(1/\sqrt{p})$ in standard deviation, identical to that obtained in Table 3. In practice, the simulation requirements dictate the adequate scheme from Table 3 or 4. In both cases, one expects the size p of the ES to be time variant and algorithms for evolving p are discussed next in section 6b.

b. Dynamical state and error subspace forecast

The quantities $\hat{\boldsymbol{\psi}}_k(+)$, $\mathbf{E}_k(+)$, and $\mathbf{\Pi}_k(+)$ obtained in (section 6a) are now known. The goal is to issue their forecast to the next data time t_{k+1}. For large nonlinear models, we expect that for adequate sampling of the initial error conditions, an ensemble forecast is effi-

Table 4. SVD of the ensemble spread linear update (index k omitted).

		Equation
Dynamical state update:	(Same as Table 3)	(17)
Sample ES optimal gain:	(Same as Table 3)	(18)
Ensemble spread SVD update:	$\mathbf{F}\mathbf{\Sigma}(+)\mathbf{G}^{\mathrm{T}} = \mathrm{SVD}_p((\mathbf{I} - \tilde{\mathbf{K}}^p\tilde{\mathbf{C}}^p)\mathbf{\Sigma}(-) + \tilde{\mathbf{K}}^p\overline{\mathbf{V}}^p\mathbf{V}_-),$	(33a)
	$\mathbf{E}_+ = \mathbf{E}_-\mathbf{F},$	(33b)
	$\mathbf{V}_+ = \mathbf{V}_-\mathbf{G}$	(33c)
Ensemble update:	$\hat{\boldsymbol{\psi}}^j(+) = \hat{\boldsymbol{\psi}}(+) + \mathbf{E}_+\mathbf{\Sigma}(+)(\mathbf{V}_+^{\mathrm{T}})^j$	(29b)

cient to estimate the evolution of the state and its uncertainty. Monte Carlo forecasting is, thus, the approach followed. Several alternatives are discussed in appendix B, section c.

1) DYNAMICAL STATE FORECAST

The conditional mean of (2a), $\hat{\boldsymbol{\psi}}^t$ (appendix A), evolves according to

$$\mathbf{d}\hat{\boldsymbol{\psi}}^t = \hat{\mathbf{f}}(\boldsymbol{\psi}^t, t)\, dt. \qquad (34)$$

The nonlinear *central forecast* to t_{k+1}, $\hat{\boldsymbol{\psi}}_{k+1}^{\text{cf}}(-)$, is obtained from

$$\mathbf{d}\hat{\boldsymbol{\psi}} = \mathbf{f}(\hat{\boldsymbol{\psi}}, t)\, dt, \quad \text{with } \hat{\boldsymbol{\psi}}_k = \hat{\boldsymbol{\psi}}_k(+). \qquad (35)$$

Statistically, it is the first-order estimate of the conditional mean $\hat{\boldsymbol{\psi}}_{k+1}^t$ (Jazwinski 1970). It is also the classic deterministic forecast (1a). With the ensemble ES forecast approach [section 6b(2)], each member evolves during Δt_{k+1} as in (2a),

$$\mathbf{d}\hat{\boldsymbol{\psi}}^j = \mathbf{f}(\hat{\boldsymbol{\psi}}^j, t)\, dt + \mathbf{dw}, \quad \text{with } \hat{\boldsymbol{\psi}}_k^j = \hat{\boldsymbol{\psi}}_k^j(+). \qquad (36a)$$

The corresponding *ensemble mean* at t_{k+1} estimates $\hat{\boldsymbol{\psi}}_{k+1}^t$ with a standard deviation of $O(1/\sqrt{q})$:

$$\hat{\boldsymbol{\psi}}_{k+1}^{\text{em}}(-) \doteq \mathcal{E}^q\{\hat{\boldsymbol{\psi}}_{k+1}^j(-)\}. \qquad (36b)$$

Other estimates are the forecast of minimum data misfits (section 7c) or the most probable forecast (maximum likelihood). They are further discussed in appendix B, section d. In section 6b(2), the algorithm simply denotes the chosen conditional mean estimate as $\hat{\boldsymbol{\psi}}_{k+1}(-)$.

2) ERROR SUBSPACE FORECAST

Using (2)–(3), the error covariance of $\hat{\boldsymbol{\psi}}^t$, $\mathbf{P}$, evolves during Δt_{k+1} according to

$$\mathbf{dP}/dt = \mathcal{E}\{\boldsymbol{\psi}^t \mathbf{f}^{\mathrm{T}}(\boldsymbol{\psi}^t, t)\} - \hat{\boldsymbol{\psi}}^t \hat{\mathbf{f}}^{\mathrm{T}}(\boldsymbol{\psi}^t, t)$$
$$+ \mathcal{E}\{\mathbf{f}(\boldsymbol{\psi}^t, t)\boldsymbol{\psi}^{t\mathrm{T}}\} - \hat{\mathbf{f}}(\boldsymbol{\psi}^t, t)\hat{\boldsymbol{\psi}}^{t\mathrm{T}} + \mathbf{Q}. \qquad (37)$$

Here, the forecast of the principal covariance of (37), $\mathbf{P}_{k+1}^p(-)$, is estimated by an ensemble of stochastic evolutions to t_{k+1} of states initially sampling the a posteriori ES structure $\mathbf{E}_k(+)$ and amplitude $\boldsymbol{\Pi}_k(+)$. The three local steps involved are described next.

(i) Create an ensemble whose covariance from $\hat{\boldsymbol{\psi}}_k(+)$ tends to $\mathbf{P}_k^p(+)$

The a posteriori ensemble is defined by

$$\hat{\boldsymbol{\psi}}_k^j(+) = \hat{\boldsymbol{\psi}}_k(+) + \mathbf{E}_k(+)\boldsymbol{\pi}_k^j(+), \quad j = 1, \ldots, q, \quad (38)$$

where the coefficients $\boldsymbol{\pi}_k^j(+) \in \mathbb{R}^p$ have to be determined. The simplest choice is

$$\boldsymbol{\pi}_k^j(+) = \boldsymbol{\Pi}_k^{1/2}(+)\mathbf{u}^j. \qquad (39a)$$

The vectors $\mathbf{u}^j \in \mathbb{R}^p$ are q realizations of a random vector $\mathbf{u}$ of zero mean and covariance $\mathbf{I}^p$. For $q = q_k(+) \rightarrow$

∞ in (38)–(39a), the sample covariance with respect to $\hat{\boldsymbol{\psi}}_k(+)$ tends by construction toward $\mathbf{P}_k^p(+)$. Constraints can be added to (39a) so that all realizations $\hat{\boldsymbol{\psi}}_k^j(+)$ are physically acceptable,

$$\boldsymbol{\pi}_k^j(+) = \boldsymbol{\Pi}_k^{1/2}(+)\mathbf{u}^j,$$

with dynamical/data constraints. (39b)

One may wonder why (39b) should be used since the states $\hat{\boldsymbol{\psi}}_k^j(+)$ are in accord with the data and dynamics, and their dominant error covariances. One argument is that only dominant covariances are considered in (39a), but not higher moments. The signs of the $\mathbf{u}^j$ are free in (39a). Because of the orthogonality condition, some combinations of singular vectors can also lead to unrealistic variability, even if the true error subspace is spanned. Finally, some of the randomly generated $\mathbf{u}^j$ (e.g., Gaussian) can have values quite far from their statistical mean and variance. Hence, constraining the combinations (39a) is useful to reject the few states of possibly too unrealistic or unlikely physics. A simple constraint is to force the $\hat{\boldsymbol{\psi}}_k^j(+)$ vectors to have data residuals in strong accord with the measurement model errors. Rejecting members that have residuals of horizontal-averaged variance smaller than a factor of the local observation error variance has been implemented. Such residual constraints are limited to data regions. Other restrictions can be applied globally, for example, a water column in static equilibrium, weak geostrophic balance, or other considerations (e.g., Mureau et al. 1993). A third option for the $\boldsymbol{\pi}_k^j(+)$ in (38) is to directly use scheme B. From (29b), the combination coefficients are then

$$\boldsymbol{\pi}_k^j(+) = \boldsymbol{\Sigma}(+)(\mathbf{V}_+^{\mathrm{T}})^j, \qquad (39c)$$

where the rhs is obtained from (32) or (33a), (33c). The three linear combinations (39a)–(39c) have been utilized in Lermusiaux (1997) and in several regions of the world's oceans (section 8) for weeks of simulations. For the lack of sufficient time–space data coverage, real data cases for which an approach has always been found to be better than the others have not yet been determined. However, using (39b) instead of (39a) is advantageous since it involves little extra cost compared to that of running a very unlikely simulation in (36a).

An important advantage of (38)–(39) is that the size of the a posteriori ensemble $q_k(+)$ is easily made larger than that of the a priori one, $q_k(-)$. Since $q_k(+) = q_{k+1}(-)$, this is carried out as a function of the convergence of the ES forecast to t_{k+1} (appendix B, section b). Note that increasing the size of the ensemble using (38)–(39) does not extend the base spanning the ES at t_k [in our two-step recursive assumption, $\mathbf{E}_k(+)$ is assumed adequate]. However, for nonlinear models, each new integration of (36a) during Δt_{k+1} increases the size of the ES forecast to t_{k+1}, the significance of which grows with the duration of integration. The nonlinearities lead to an evolving ES (section 5c). This fact is

TABLE 5. Nonlinear dynamical state and ES ensemble forecast.

		Equation
Conditional mean estimates:		
Central forecast:	$\hat{\boldsymbol{\psi}}_{k+1}^{\mathrm{cf}}(-)\,\|\,\mathbf{d}\hat{\boldsymbol{\psi}} = \mathbf{f}(\hat{\boldsymbol{\psi}},\,t)\,dt,$ with $\hat{\boldsymbol{\psi}}_k = \hat{\boldsymbol{\psi}}_k(+)$	(35)
Ensemble mean:	$\hat{\boldsymbol{\psi}}_{k+1}^{\mathrm{em}}(-) \doteq \mathcal{E}^q\{\hat{\boldsymbol{\psi}}_{k+1}^{j}(-)\}\,\|\,(38\!-\!40)$	(36b)
ES initial conditions:	$\hat{\boldsymbol{\psi}}_k^{j}(+) = \hat{\boldsymbol{\psi}}_k(+) + \mathbf{E}_k(+)\boldsymbol{\pi}_k^{j}(+),$ $j = 1, \ldots, q,$	(38)
	with either	
	$\boldsymbol{\pi}_k^{j}(+) = \boldsymbol{\Pi}_k^{1/2}(+)\mathbf{u}^{j},$	(39a)
	$\boldsymbol{\pi}_k^{j}(+) = \boldsymbol{\Pi}_k^{1/2}(+)\mathbf{u}^{j},$ with dynamical/data constraints, or	(39b)
	$\boldsymbol{\pi}_k^{j}(+) = \boldsymbol{\Sigma}(+)(\mathbf{V}_+^{\mathrm{T}})^{j},$ with (32) or (33a)$-$(33c),	(39c)
	where $\mathbf{u}$ is of zero mean and covariance $\mathbf{I}^{p}$	
Ensemble forecast:	$\hat{\boldsymbol{\psi}}_{k+1}^{j}(-)\,\|\,\mathbf{d}\hat{\boldsymbol{\psi}}^{j} = \mathbf{f}(\hat{\boldsymbol{\psi}}^{j},\,t)\,dt + \mathbf{dw},$ with $\hat{\boldsymbol{\psi}}_k^{j} = \hat{\boldsymbol{\psi}}_k^{j}(+),$ $j = 1, \ldots, q,$	(40)
	where $\mathcal{E}\{\mathbf{dw}(t)\mathbf{dw}^{\mathrm{T}}(t)\} \doteq \mathbf{Q}(t)\,dt \doteq \mathbf{B}(t)\mathbf{B}^{\mathrm{T}}(t)\,dt$ and $\mathbf{B}(t) \in \Re^{n \times r}$	
ES forecast:	$\mathbf{M}_{k+1}(-) = \big[\hat{\boldsymbol{\psi}}_{k+1}^{j}(-) - \hat{\boldsymbol{\psi}}_{k+1}(-)\big],$ $j = 1, \ldots, q,$	
	decomposed into $\boldsymbol{\Pi}_{k+1}(-) \doteq \dfrac{1}{q}\boldsymbol{\Sigma}_{k+1}^{2}(-)$ and $\mathbf{E}_{k+1}(-)$ of rank $p \le q,$ where	
	$\big[\boldsymbol{\Sigma}_{k+1}(-),\,\mathbf{E}_{k+1}(-)\,\big|\,\mathrm{SVD}_p(\mathbf{M}_{k+1}(-)) = \mathbf{E}_{k+1}(-)\boldsymbol{\Sigma}_{k+1}(-)\mathbf{V}_{k+1}^{\mathrm{T}}(-)\big]$	(41)
	and the operator $\mathrm{SVD}_p(\cdot)$ selects the rank-p SVD	

illustrated in Part II. Stochastic model errors in (36a) also excite growing modes of variability and thus favor the ES evolution. With linear models, the size of the ES is only modified by this stochastic forcing. If model errors are null, to evolve the ES one must then add new columns to $\mathbf{E}_k$.

(ii) Integrate each ensemble member from t_{k+1} using the sample path (2a)

Renumbering (36a), an ensemble of stochastic forecasts are evolved during Δt_{k+1},

$$\mathbf{d}\hat{\boldsymbol{\psi}}^{j} = \mathbf{f}(\hat{\boldsymbol{\psi}}^{j},\,t)dt + \mathbf{dw}, \quad \text{with } \hat{\boldsymbol{\psi}}_k^{j} = \hat{\boldsymbol{\psi}}_k^{j}(+),$$
$$j = 1, \ldots, q, \qquad (40)$$

where $\mathbf{dw}(t)$ is a vector Brownian motion process, representing a priori model errors. Its covariance over dt is $\mathcal{E}\{\mathbf{dw}(t)\mathbf{dw}^{\mathrm{T}}(t)\} \doteq \mathbf{Q}(t)dt \doteq \mathbf{B}(t)\mathbf{B}^{\mathrm{T}}(t)dt,$ where $\mathbf{B}(t) \in \mathbb{R}^{n \times r}$. The ES concept argues that restricting $\mathbf{B}(t)$ to a few $r \ll n$ columns corresponding to the rank-r eigendecomposition of $\mathbf{Q}(t)$ is efficient. Several authors have implicitly used this fact (e.g., Dee at al. 1985; Phillips 1986; Cohn and Parrish 1991; Daley 1991; Cohn 1993; Jiang and Ghil 1993). Stochastic modeling of the dominant model errors $\mathbf{B}(t)$ is becoming an area of active research. Their specification is discussed in Lermusiaux (1997).

(iii) Compute the forecast of the ES structure and amplitude at t_{k+1}

Once $\hat{\boldsymbol{\psi}}_{k+1}(-)$ is chosen [section 6b(1)], the matrix of sample error forecasts, $\mathbf{M}_{k+1}(-) = [\hat{\boldsymbol{\psi}}_{k+1}^{j}(-) - \hat{\boldsymbol{\psi}}_{k+1}(-)] \in \mathbb{R}^{n \times q},$ is evaluated. The sample estimates $\boldsymbol{\Pi}_{k+1}(-)$ and $\mathbf{E}_{k+1}(-)$ of rank $p \le q$ are then most efficiently obtained from the rank-p SVD of $\mathbf{M}_{k+1}(-),$

$$\big[\boldsymbol{\Sigma}_{k+1}(-),\,\mathbf{E}_{k+1}(-)\,\big|\,\mathrm{SVD}_p[\mathbf{M}_{k+1}(-)]$$
$$= \mathbf{E}_{k+1}(-)\boldsymbol{\Sigma}_{k+1}(-)\mathbf{V}_{k+1}^{\mathrm{T}}(-)\big], \qquad (41)$$

with $\boldsymbol{\Pi}_{k+1}(-) \doteq (1/q)\boldsymbol{\Sigma}_{k+1}^{2}(-)$. Table 5 summarizes the steps (i–iii). Among the conditional mean estimates (appendix B, section d), only the central and ensemble mean forecasts are stated.

The basis of the present ESSE filtering scheme consists of Tables 3 and 5. A flow schematic of the algorithm is shown in Fig. 2. As in Fig. 1, each operation consists of several subcomputations and options, with corresponding equations (e.g., appendix B). For instance, the ES initialization and the adoptive error subspace learning are challenging (Lermusiaux 1997; Lermusiaux et al. 1998, manuscript submitted to *Quart. J. Roy. Meteor. Soc.*).

7. Smoothing via ESSE schemes

The improvement of the filtering solution $\hat{\boldsymbol{\psi}}_k(+)$ (section 6) based on future data leads to a smoothing so-
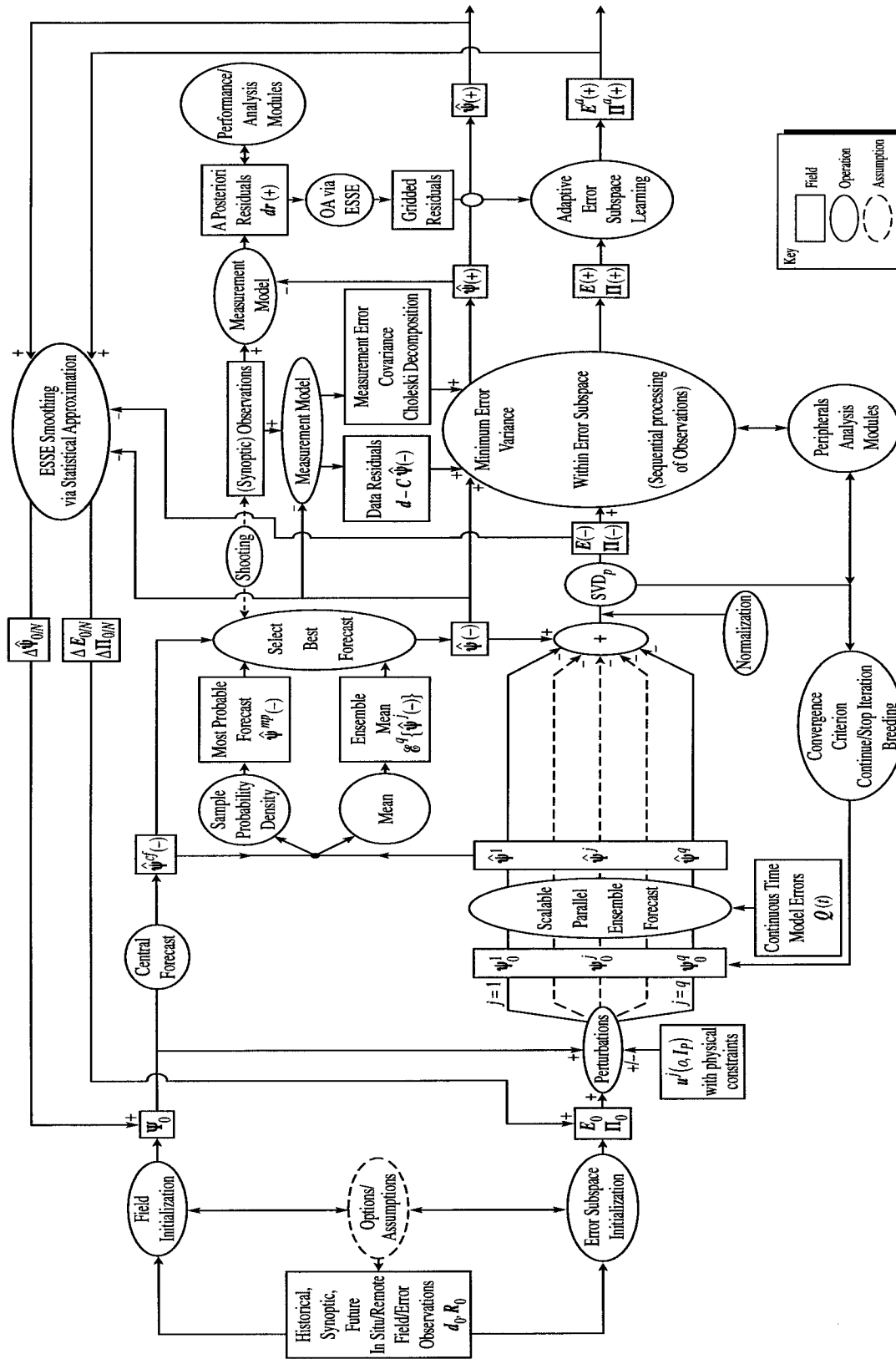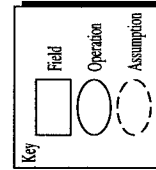
FIG. 2. ESSE flow diagram.

lution $\hat{\psi}_{k/N}$. The present smoothing criterion was defined by (6c). The data interval $[t_0, t_N]$ is fixed: fixed-interval smoothing is considered. The complete problem statement is as in Table 1, but with (6b) replaced by (6c). The issues in nonlinear smoothing are that a forward integration of (2a) is rarely invertible and that running nonlinear geophysical models backward in time can be difficult. In fact, most nonlinear smoothers are based on some of sort of localized (and iterated) approximation (Jazwinski 1970). The present approach falls into that category. However, the philosophy somewhat differs from that of some schemes previously utilized in geophysical studies: it is argued that i) for the lack of data, an accurate filtering solution (section 6) is an essential prerequisite to the smoothing; ii) the linearization, if any, should be local enough; and iii) in real-time smoothing, a few iterations of approximate schemes can be very valuable (section 7c). With this in mind, smoothing via statistical approximation is developed in section 7a. Its ESSE version is outlined in section 7b. A discussion, with additional approximate ESSE algorithms, is presented in section 7c. Specifics are provided in appendix B.

a. Smoothing via statistical approximation

The approach consists of nonlinear filtering until t_N (sections 4–6), followed by the update of the conditional mean and error covariance, backward in time, based on future data. This later component is now outlined. It is recursive as was the filtering. The derivation presumes that the smoothing estimates $\hat{\psi}_{k+1/N}$ and $\mathbf{P}_{k+1/N}$ have been obtained. The unknowns are $\hat{\psi}_{k/N}$ and $\mathbf{P}_{k/N}$. A statistical approximation (e.g., Gelb 1974; Austin and Leondes 1981) to the forward integration of (2a) between data times t_k and t_{k+1} is first derived, assuming that ψ_{k+1}^t is perfectly known. The approximation is written in a backward form. Based on the smoothing conditions at t_{k+1}, it is used to compute $\hat{\psi}_{k/N}$ and $\mathbf{P}_{k/N}$. The resulting smoothing is shown to include a few classic schemes as particular cases. Its truncation to a significant subspace is outlined in section 7b.

The present statistical approximation is chosen to be locally linear. This implies that Δt_{k+1} is small enough and that the statistical linearization is made around the data-corrected filtering solution, characterized by the initial conditions $\hat{\psi}_k(+)$ and $\mathbf{P}_k(+)$, and forecasts, $\hat{\psi}_{k+1}(-)$ and $\mathbf{P}_{k+1}(-)$. We seek a linear relation that estimates how to correct $\hat{\psi}_k(+)$ based on its forecast error. Assuming for now that ψ_{k+1}^t is known, this relation is of the form

$$\hat{\psi}_k = \hat{\psi}_k(+) + \mathbf{L}_k[\psi_{k+1}^t - \hat{\psi}_{k+1}(-)]. \quad (42)$$

Within the present minimum error variance approach, the unknown matrix $\mathbf{L}_k$ should be such that (42) minimizes the error variance of $\hat{\psi}_k$, hence,

$$\left\{ \mathbf{L}_k \middle| \min_{\mathbf{L}_k} \mathrm{tr}[\mathbf{P}_k] \doteq \mathrm{tr}[\mathcal{E}\{(\hat{\psi}_k - \psi_k^t)(\hat{\psi}_k - \psi_k^t)^{\mathrm{T}}\}], \right.$$

$$\left. \text{using (42)} \right\}. \quad (43)$$

Subtracting ψ_k^t from both sides of (42) and forming $\mathbf{P}_k$ in (43) leads to

$$\mathbf{P}_k = \mathbf{P}_k(+) + \mathcal{E}\{[\hat{\psi}_k(+) - \psi_k^t][\psi_{k+1}^t - \hat{\psi}_{k+1}(-)]^{\mathrm{T}}\}\mathbf{L}_k^{\mathrm{T}}$$
$$+ \mathbf{L}_k\mathcal{E}\{[\psi_{k+1}^t - \hat{\psi}_{k+1}(-)][\hat{\psi}_k(+) - \psi_k^t]^{\mathrm{T}}\}$$
$$+ \mathbf{L}_k\mathbf{P}_{k+1}(-)\mathbf{L}_k^{\mathrm{T}}, \quad (44)$$

where $\mathbf{P}_{k+1}(-)$ is the error covariance forecast carried out from $\mathbf{P}_k(+)$ based on (1a)–(2a). Inserting (44) into (43) and taking the derivative with respect to $\mathbf{L}_k$ yields

$$\mathbf{L}_k = \mathcal{E}\{[\psi_k^t - \hat{\psi}_k(+)][\psi_{k+1}^t - \hat{\psi}_{k+1}(-)]^{\mathrm{T}}\}\mathbf{P}_{k+1}^{-1}(-), \quad (45)$$

which is a minimum since $\mathbf{P}_{k+1}(-)$ is positive semidefinite. The optimum (45) and relation (42) define the statistical backward linearization sought. It is now employed to compute the smoothing conditions at t_k. The best available unbiased estimate of ψ_{k+1}^t is $\hat{\psi}_{k+1/N}$, of error covariance $\mathbf{P}_{k+1/N}$. Using it in (42) gives the smoothing estimate

$$\hat{\psi}_{k/N} = \hat{\psi}_k(+) + \mathbf{L}_k[\hat{\psi}_{k+1/N} - \hat{\psi}_{k+1}(-)], \quad (46)$$

with $\hat{\psi}_{N/N} = \hat{\psi}_N(+)$. The error covariance associated with (46) is derived as follows. Subtracting ψ_k^t from (46) and rearranging the terms gives

$$(\hat{\psi}_{k/N} - \psi_k^t) - \mathbf{L}_k\hat{\psi}_{k+1/N}$$
$$= [\hat{\psi}_k(+) - \psi_k^t] - \mathbf{L}_k\hat{\psi}_{k+1}(-). \quad (47)$$

Multiplying (47) by its transpose and taking expectations leads to

$$\mathbf{P}_{k/N} + \mathbf{L}_k\mathcal{E}\{\hat{\psi}_{k+1/N}\hat{\psi}_{k+1/N}^{\mathrm{T}}\}\mathbf{L}_k^{\mathrm{T}}$$
$$= \mathbf{P}_k(+) + \mathbf{L}_k\mathcal{E}\{\hat{\psi}_{k+1}(-)\hat{\psi}_{k+1}^{\mathrm{T}}(-)\}, \quad (48)$$

since the cross-terms, $\mathcal{E}\{(\hat{\psi}_k(+) - \psi_k^t)\hat{\psi}_{k+1}^{\mathrm{T}}(-)\}$ and $\mathcal{E}\{(\hat{\psi}_{k/N} - \psi_k^t)\hat{\psi}_{k+1/N}^{\mathrm{T}}\}$, are null. The first of these is null because it involves unbiased error fields multiplied by the expected state $\hat{\psi}_{k+1}^{\mathrm{T}}(-)$. To see that the second is also null, one can replace $\hat{\psi}_{k+1/N}$ using the linear relation (46), use the previous argument for $\hat{\psi}_{k+1}^{\mathrm{T}}(-)$, and invoke the orthogonality principle (Davis 1977a), which states that the error $\hat{\psi}_{k/N} - \psi_k^t$ is orthogonal to functions of already used measurements, hence to $\hat{\psi}_{k/N}$ and $\hat{\psi}_k(+)$. For similar reasons, the relation $\mathcal{E}\{\psi_{k+1}^t\psi_{k+1}^{t\mathrm{T}}\} = \mathbf{P}_{k+1} + \mathcal{E}\{\hat{\psi}_{k+1}\hat{\psi}_{k+1}\}$ holds for all estimates $\hat{\psi}_{k+1}$. Using it for $\hat{\psi}_{k+1/N}$ and $\hat{\psi}_{k+1}(-)$ into (48) yields the error covariance of $\hat{\psi}_{k/N}$,

$$\mathbf{P}_{k/N} = \mathbf{P}_k(+) + \mathbf{L}_k(\mathbf{P}_{k+1/N} - \mathbf{P}_{k+1}(-))\mathbf{L}_k^{\mathrm{T}}, \quad (49)$$

with $\mathbf{P}_{N/N} = \mathbf{P}_N(+)$. The complete scheme is stated in Table 6. In passing, if $\hat{\psi}_{k+1/N}$ in (46) had been perfect,

TABLE 6. Smoothing via statistical linearization.

		Equation
Filtering estimate (forward in time):	Forecast: Eqs. (34) and (37), with $\hat{\boldsymbol{\psi}}_0 = \boldsymbol{\Psi}_0$ and $\mathbf{P}_0 = \hat{\mathbf{P}}_0$ Update: Table 2 during $[t_0, t_N]$	
Smoothing estimate (backward in time):	$\hat{\boldsymbol{\psi}}_{k/N} = \hat{\boldsymbol{\psi}}_k(+) + \mathbf{L}_k[\hat{\boldsymbol{\psi}}_{k+1/N} - \hat{\boldsymbol{\psi}}_{k+1}(-)], \quad$ with $\hat{\boldsymbol{\psi}}_{N/N} = \hat{\boldsymbol{\psi}}_N(+)$	(46)
Smoothing gain:	$\mathbf{L}_k = \mathcal{E}\big[[\hat{\boldsymbol{\psi}}_k^t - \hat{\boldsymbol{\psi}}_k(+)][\hat{\boldsymbol{\psi}}_{k+1}^t - \hat{\boldsymbol{\psi}}_{k+1}(-)]^T\big]\mathbf{P}_{k+1}^{-1}(-)$	(45)
Smoothing error covariance (backward in time):	$\mathbf{P}_{k/N} = \mathbf{P}_k(+) + \mathbf{L}_k[\mathbf{P}_{k+1/N} - \mathbf{P}_{k+1}(-)]\mathbf{L}_k^T$	(49)

(49) would have been equal to (44). One can verify that (44) does not contain the positive-definite term involving $\mathbf{P}_{k+1/N}$. In simple words, $\mathbf{L}_k\mathbf{P}_{k+1/N}\mathbf{L}_k^T$ in (49) is the error cost incurred for using (46) instead of (42).

Some of the classic smoothing algorithms are simplifcations of Table 6. The extended Kalman smoother (EKS) is first considered. In the EKS, the nonlinear integrations (34) and (37) are reduced using Taylor series expansions. Successive relinearizations about the last available estimate of the conditional mean are carried out. The resulting state estimate is the central forecast (35). To obtain a covariance estimate, the evolution of the perturbation $(\boldsymbol{\psi}^t - \hat{\boldsymbol{\psi}})$ is linearized about (35), leading to

$$\mathbf{d}(\boldsymbol{\psi}^t - \hat{\boldsymbol{\psi}}) = \frac{\delta\mathbf{f}(\boldsymbol{\psi}, t)}{\delta\boldsymbol{\psi}}\bigg|_{\boldsymbol{\psi}=\hat{\boldsymbol{\psi}}(t)} (\boldsymbol{\psi}^t - \hat{\boldsymbol{\psi}})dt + \mathbf{dw}, \quad (50)$$

with initial conditions $[\boldsymbol{\psi}_k^t - \hat{\boldsymbol{\psi}}_k(+)]$. Integrating (50) over Δt_{k+1} and denoting by $\boldsymbol{\Phi}(t_{k+1}, t_k)$ the corresponding state transition matrix, one has

$$\boldsymbol{\psi}_{k+1}^t - \hat{\boldsymbol{\psi}}_{k+1}(-) = \boldsymbol{\Phi}(t_{k+1}, t_k)[\boldsymbol{\psi}_k^t - \hat{\boldsymbol{\psi}}_k(+)] + \mathbf{w}_{k+1},$$
$$(51)$$

where $\mathbf{w}_{k+1}$ accounts for the integrated effects of $\mathbf{dw}$ over Δt_{k+1} in (50). Inserting this simplification (51) into (45) yields

$$\mathbf{L}_k \simeq \mathbf{P}_k(+)\boldsymbol{\Phi}^T(t_{k+1}, t_k)\mathbf{P}_{k+1}^{-1}(-), \quad (52)$$

which, with (46) and (49), defines the Rauch–Tung–Striebel EKS (e.g., Jazwinski 1970). The linearized KS is also based on (52), but without the relinearizations. If (1a)–(2a) were linear models, $\boldsymbol{\Phi}$ in (52) would simply be the transition matrix of (1a). The statistical approximation (45) hence encompasses several truncations of lower order. In (45), all nonlinearities are kept in computing the expectations; derivatives do not need to exist. In fact, truncating the models (1a)–(2a) prior to computing these expectations, as was done in (50)–(52) using Taylor series, has been shown to be less accurate in several filtering cases (e.g., Austin and Leondes 1981; Uchino et al. 1993). In passing, expansions of higher order than (50)–(52) would still replace the global statistical properties of (37) and (45) by local derivatives.

b. ESSE and smoothing via statistical approximation

The smoothing scheme in Table 6 is now reduced to its significant components. The recursive derivation presumes that the smoothing estimates $\hat{\boldsymbol{\psi}}_{k+1/N}$, $\mathbf{E}_{k+1/N}$ and $\boldsymbol{\Pi}_{k+1/N}$ defining $\mathbf{P}_{k+1/N}^p$, have been obtained based on (6c). The unknowns are $\hat{\boldsymbol{\psi}}_{k/N}$, $\mathbf{E}_{k/N}$, and $\boldsymbol{\Pi}_{k/N}$. Starting from the nonlinear filtering ESSE scheme run forward until t_N (section 6), a Monte Carlo sample estimate of $\mathbf{L}_k$ in (45) is, at $O(1/\sqrt{q})$ in standard deviation,

$$\mathbf{L}_k = \frac{1}{q}\mathbf{M}_k(+)\mathbf{M}_{k+1}^T(-)\left(\frac{1}{q}\mathbf{M}_{k+1}(-)\mathbf{M}_{k+1}^T(-)\right)^\dagger, \quad (53)$$

where $\mathbf{M}_k(+)$ and $\mathbf{M}_{k+1}(-)$ are defined as in section 6. The $\dagger$ logically denotes the Moore–Penrose generalized inverse. Note that in (53), $q = q_k(+) = q_{k+1}(-)$ so that all multiplications are feasible. Truncating the sample matrices in (53) to their dominant SVD,

$$\text{SVD}_p[\mathbf{M}_{k+1}(-)] = \mathbf{E}_{k+1}(-)\boldsymbol{\Sigma}_{k+1}(-)\mathbf{V}_{k+1}^T(-), \quad (54a)$$

$$\text{SVD}_p[\mathbf{M}_k(+)] = \mathbf{E}_k(+)\boldsymbol{\Sigma}_k(+)\mathbf{V}_k^T(+), \quad (54b)$$

and using the orthogonality properties of the singular vectors yields the estimate

$$\mathbf{L}_k^p = \mathbf{E}_k(+)\boldsymbol{\Sigma}_k(+)\mathbf{V}_k^T(+)\mathbf{V}_{k+1}(-)\boldsymbol{\Sigma}_{k+1}^{-1}(-)\mathbf{E}_{k+1}^T(-). \quad (55)$$

The smoothing ESSE gain $\mathbf{L}_k^p$ is thus defined by $\text{SVD}_p[\mathbf{M}_k(+)]\text{SVD}_p[\mathbf{M}_{k+1}(-)]\dagger$. The covariance associated with (46) and (55) can be derived similarly to (47)–(49); one obtains

$$\mathbf{P}_{k/N}^p = \mathbf{P}_k^p(+) + \mathbf{L}_k^p[\mathbf{P}_{k+1/N}^p - \mathbf{P}_{k+1}^p(-)]\mathbf{L}_k^{p^T}. \quad (56)$$

For $\boldsymbol{\Gamma}_k \doteq \boldsymbol{\Sigma}_k(+)\mathbf{V}_k^T(+)\mathbf{V}_{k+1}(-)\boldsymbol{\Sigma}_{k+1}^{-1}(-)$ and $\boldsymbol{\theta}_{k+1} \doteq \mathbf{E}_{k+1}^T(-)\mathbf{E}_{k+1/N}$, (56) becomes

$$\mathbf{E}_{k/N}\boldsymbol{\Pi}_{k/N}\mathbf{E}_{k/N}^T$$
$$= \mathbf{E}_k(+)\{(\boldsymbol{\Pi}_k(+)$$
$$+ \boldsymbol{\Gamma}_k[\boldsymbol{\theta}_{k+1}\boldsymbol{\Pi}_{k+1/N}\boldsymbol{\theta}_{k+1}^T - \boldsymbol{\Pi}_{k+1}(-)]\boldsymbol{\Gamma}_k^T\}\mathbf{E}_k^T(+),$$
$$(57)$$

where $\boldsymbol{\Pi}_k \doteq (1/q_k)\boldsymbol{\Sigma}_k^2$, $\forall k$. Hence, the smoothing update of the ES characteristics consist of

TABLE 7. ESSE smoothing via statistical linearization.

		Equation
Filtering estimate:	Tables 3 and 5 during $[t_0, t_N]$	
Smoothing estimate (backward in time):	$\hat{\boldsymbol{\psi}}_{k/N} = \hat{\boldsymbol{\psi}}_k(+) + \mathbf{L}_k^p(\hat{\boldsymbol{\psi}}_{k+1/N} - \hat{\boldsymbol{\psi}}_{k+1}(-))$, with $\hat{\boldsymbol{\psi}}_{N/N} = \hat{\boldsymbol{\psi}}_N(+)$	(46)
Smoothing gain:	$\mathbf{L}_k^p = \mathbf{E}_k(+)\boldsymbol{\Gamma}_k\mathbf{E}_{k+1}^{\mathrm{T}}(-)$ and $\boldsymbol{\Gamma}_k = \boldsymbol{\Sigma}_k(+)\mathbf{V}_k^{\mathrm{T}}(+)\mathbf{V}_{k+1}(-)\boldsymbol{\Sigma}_{k+1}^{-1}(-)$	(55)
Smoothing error covariance (backward in time):	$\mathbf{H}_k\boldsymbol{\Pi}_{k/N}\mathbf{H}_k^{\mathrm{T}} = \boldsymbol{\Pi}_k(+) + \boldsymbol{\Gamma}_k[\boldsymbol{\theta}_{k+1}\boldsymbol{\Pi}_{k+1/N}\boldsymbol{\theta}_{k+1}^{\mathrm{T}} - \boldsymbol{\Pi}_{k+1}(-)]\boldsymbol{\Gamma}_k^{\mathrm{T}}$,	(58a)
	$\mathbf{E}_{k/N} = \mathbf{E}_k(+)\mathbf{H}_k$, with $\boldsymbol{\theta}_{k+1} \doteq \mathbf{E}_{k+1}^{\mathrm{T}}(-)\mathbf{E}_{k+1/N}$.	(58b)

$$\mathbf{H}_k\boldsymbol{\Pi}_{k/N}\mathbf{H}_k^{\mathrm{T}} = \boldsymbol{\Pi}_k(+)$$
$$+ \boldsymbol{\Gamma}_k[\boldsymbol{\theta}_{k+1}\boldsymbol{\Pi}_{k+1/N}\boldsymbol{\theta}_{k+1}^{\mathrm{T}} - \boldsymbol{\Pi}_{k+1}(-)]\boldsymbol{\Gamma}_k^{\mathrm{T}}, \quad (58a)$$
$$\mathbf{E}_{k/N} = \mathbf{E}_k(+)\mathbf{H}_k. \quad (58b)$$

The a posteriori covariance can then be obtained from $\mathbf{P}_{k/N}^p = \mathbf{E}_{k/N}\boldsymbol{\Pi}_{k/N}\mathbf{E}_{k/N}^{\mathrm{T}}$. The sequential scheme is summarized by Table 7. In cases with strong nonlinearities or long intervals without data, Table 7 can be iterated. Once $\hat{\boldsymbol{\psi}}_{0/N}^1$ is known, a new nonlinear ESSE filtering up to t_N can be carried out, followed by the statistical smoothing up to $\hat{\boldsymbol{\psi}}_{0/N}^2$ and so on. Such iterations are discussed for the EKS in Jazwinski (1970); similar equations apply here.

c. Discussion

An essential property of Tables 6–7 is that the four-dimensional statistics of the dynamical variability, model errors, and data available enter the cost function (6c). Another is that the statistical smoothing is made on the nonlinear filtering ESSE estimate, which is already a state evolution corrected by data. In fact, this contrasts with the ensemble smoother (Evensen 1997a) or iterated representer methods (e.g., Bennett et al. 1996, 1997). Evensen and van Leeuwen (1996) showed that the ensemble Kalman filter was superior to the ensemble representer smoother. An explanation for this unexpected issue is that in the linear representer approach, the first guess is the pure (without data assimilation) deterministic forecast up to t_N. Predictability errors, as well as numerical and analytical model deficiencies can then dominate before reaching t_N. Driving the pure forecast back to the true ocean evolution by quadratic minimization is therefore numerically difficult. This issue could be solved by modifying the original representer equations, $\hat{\boldsymbol{\psi}}_k = \hat{\boldsymbol{\psi}}_k^f + \tilde{\mathbf{R}}_{kl}\mathbf{b}_l$ (Robinson et al. 1998a), into

$$\hat{\boldsymbol{\psi}}_{k/N} = \hat{\boldsymbol{\psi}}_k(+) + \tilde{\mathbf{R}}_{kl}(+)\mathbf{b}_l(+), \quad k = 0, \ldots, N, \quad (59)$$

where $\tilde{\mathbf{R}}_{kl}(+) \doteq [\mathbf{r}_{kl}^1(+), \ldots, \mathbf{r}_{kl}^m(+)] \in \mathbb{R}^{n \times m}$ are a posteriori representers and $\mathbf{b}_l(+) \in \mathbb{R}^m$ their coefficients. Equations for $\tilde{\mathbf{R}}_{kl}(+)$ and $\mathbf{b}_l(+)$ can be derived from the linearization of Tables 6–7. Other advantages of Table 7 include the efficient reduction to the dominant errors; the orthogonality properties of the SVD leading to sim-

plified, efficient computations; and the evaluation of the a posteriori error covariances (58a)–(58b). Table 7 has been successfully utilized for PE process studies in the Levantine Basin, whose first results are partially described in Lermusiaux (1997).

In real-time smoothing with very large multivariate states, approximate approaches can be very useful benchmarks, as the OI scheme continues to be in filtering studies (e.g., Lermusiaux 1999b). For instance, one may impose a priori the form of the smoothing gains or simply best-fit the dynamics to the data. In the latter case, for an ensemble (40) of nonlinear forecasts during Δt_{k+1}, a first approach is to set $\hat{\boldsymbol{\psi}}_{k/N}$ to the initial conditions of the forecast that minimizes the forecast–data misfits:

$$\left\{ \hat{\boldsymbol{\psi}}_{k/N} \,\bigg|\, \min_j \|\mathbf{d}_{k+1} - \mathbf{C}_{k+1}\hat{\boldsymbol{\psi}}_{k+1}^j(-)\| \right\}. \quad (60a)$$

This smoothing is a *shooting* method toward future data. It is local, $\hat{\boldsymbol{\psi}}_{k/N} = \hat{\boldsymbol{\psi}}_{k/k+1}$, but it can be iterated. Taking the data errors into account, one obtains the cost function

$$\left\{ \hat{\boldsymbol{\psi}}_{k/N} \,\bigg|\, \min_j (\mathbf{d}_{k+1} - \mathbf{C}_{k+1}\hat{\boldsymbol{\psi}}_{k+1}^j(-))^{\mathrm{T}}\mathbf{R}_{k+1} \right.$$
$$\left. \times (\mathbf{d}_{k+1} - \mathbf{C}_{k+1}\hat{\boldsymbol{\psi}}_{k+1}^j(-)) \right\}. \quad (60b)$$

Examples of such smoothing are given in Lermusiaux (1997). A few gradient descents in an adjoint approach (Sundqvist 1993; Rabier et al. 1996) extend (60b) to the whole interval $[t_0, t_N]$, with an additional term for the model parameters (e.g., initial and boundary conditions). Deriving simple nonlinear subspace methods for iterative adjoint strategies is valuable. Miller and Cornuelle (1999) have successfully utilized such a reduced-state inverse method for initialization. Assuming perfect dynamics and a diagonal data error covariance matrix, the authors employed an ensemble of large-scale horizontal structure functions to adjust the initial conditions to the future data.

8. Summary and conclusions

Utilizing basic properties and heuristic characteristics of geophysical measurements and models, a nonlinear

assimilation approach has been outlined. The concept of an evolving error subspace (ES), of variable size, that spans and tracks the scales and processes where the dominant errors occur was defined. The word ''dominant'' is naturally specified by the error measure used in the chosen optimal estimation criterion. Truncating this criterion to its evolving ES defines the error subspace statistical estimation approach (ESSE). The general goal of ESSE is to determine the considered ocean–atmosphere state evolution by minimizing the dominant errors, in accord with the full dynamical and measurement model constraints, and their respective uncertainties. This rational approach, satisfying realistic and practical goals while addressing geophysical issues, leads to efficient estimation schemes. For the minimum error variance criterion, the ES is characterized by time-variant principal error vectors and values. In general, the size and span of the ES evolve in time. The meaning and validity of the dominant truncation of the error space was addressed, with a focus on synoptic/mesoscale to large-scale geophysical applications. The ES was intercompared to variations of variability subspaces and its properties were discussed. In Part II of this study, primitive-equation simulations of Middle Atlantic Bight shelfbreak front evolutions are employed to assess and exemplify the capabilities of an ESSE system. The approach has also been used in real-time forecasting for North Atlantic Treaty Organization (NATO) operations in the Strait of Sicily and in the simulation and study of the spreading of the Levantine intermediate water (Lermusiaux 1997). Other real-time simulations were carried out in the Ionian Sea and Gulf of Cadiz (Robinson et al. 1998b).

In this first part, filtering and smoothing schemes for nonlinear assimilation via ESSE are derived (Figs. 1, 2). The time integration of the ES is based on Monte Carlo ensemble forecasts. The a posteriori members sample the current ES and are forecast using the full nonlinear stochastic model. The melding criterion minimizes variance in the ES and is much less costly than classical analyses involving full covariances. The statistical smoothing keeps all nonlinearities in computing expectations and is carried out from the nonlinear filtering solution, which is already corrected by data. For given computer resources, the representation of errors is energetically optimal. The dynamical forecast is corrected by data where the forecast errors are most energetic. The assimilation is multivariate and three-dimensional in physical space. The model nonlinearities and model errors are accounted for, and their effects on forecast errors explicitly considered. The SVD facilitates the analysis of the evolving error covariance. The scheme is well suited to parallel computers. The formalism, while conceptually simple, can be complex in its details. Determining mathematical systems that describe and track the dominant errors is challenging. Several specific components and variations of the present schemes are provided in appendix B. Dynamical systems for tracking and learning the ES (Brockett 1991) are derived in Lermusiaux (1997).

The concepts introduced by the subspace approach are as useful as the practical benefits. The ESSE formalism defines a theoretical basis for rational intercomparison of other reduced dimension methods. As discussed in the text, these methods have led to numerous controversies and, in fact, several of their respective assumptions were simply shown to be incompatible. By definition, the present approach can rationally validate specific a priori error hypotheses for tailored applications. In accord with the evolution of the deterministic dynamics, model errors, and data available, ESSE may, for instance, lead to different simplified assimilation schemes for weather or ocean forecasting. In dynamical modeling, specific interests lead to verified a priori dynamical assumptions; similarly, in assimilation there are a priori error assumptions to verify. The focus on the dominant errors also fosters the testing and correction of existing dynamical models (e.g., coding mistakes are often associated with a large error growth). It equally implies future observations that span the dominant forecast errors or a search for data optimals (Lermusiaux 1997). ESSE in fact provides a feasible quantitative approach to both dynamical model improvements and adequate field sampling. Finally, aside from the assimilation framework, the present scheme has other applications. Turning off the assimilation in Fig. 2, one can study the impact of the dominant stochastic model errors. This is a new research area and ESSE can validate specific stochastic models for use in simulations. Without model errors and assimilation, the statistical estimation of the variations of variability and stability subspaces is considered. Predictability and stability properties (e.g., Farrell and Ioannou 1996a,b) can thus also be decomposed and analyzed. Fixing the estimation time yields an objective analysis scheme. In general, the range of applications includes nonlinear field and error forecasting, data-driven simulations, model improvements, adaptive sampling, and parameter estimation. The accurate tracking and specification of the dominant errors hence appears of paramount importance, even from a fundamental point of view.

APPENDIX A

Generic Assumptions: Stochastic Dynamical and Measurement Models

In the present notation, the convention is that of continuous/discrete estimation. The dynamical state vector is denoted by $\boldsymbol{\psi} \in \mathbb{R}^n$ and its dynamics is continuous. The observations are made at discrete times t_k, with $k = 0, \ldots, N$, and contained in $\mathbf{d}_k \in \mathbb{R}^m$. The time lag in between observations is $\Delta t_k = t_k - t_{k-1}$. The values of the state vector at data times are denoted by $\boldsymbol{\psi}_k$. The sample path of the true ocean state, $\boldsymbol{\psi}^t$, is described by the deterministic evolution, $\mathbf{d}\boldsymbol{\psi} = \mathbf{f}(\boldsymbol{\psi}, t)dt$, forced by random processes $\mathbf{dw} \in \mathbb{R}^n$:

$$\mathbf{d}\boldsymbol{\psi}^t = \mathbf{f}(\boldsymbol{\psi}^t, t) \, dt + \mathbf{dw}. \tag{A1}$$

The Itô stochastic differential Eq. (A1) represents the evolution of the true phenomena and scales of interest. The measurement model is also assumed to be a stochastic extension of its deterministic version, here chosen to be linear, $\mathbf{d}_k = \mathbf{C}_k \boldsymbol{\psi}_k$, which leads to

$$\mathbf{d}_k = \mathbf{C}_k \boldsymbol{\psi}_k^t + \mathbf{v}_k, \tag{A2}$$

where $\mathbf{v}_k \in \mathbb{R}^m$ are the random processes. The Wiener processes $\mathbf{dw}(t)$ statistically represent the effect of model errors over dt and $\mathbf{v}_k$ represents the measurement noise and measurement model errors at t_k. Classic assumptions are made (e.g., Jazwinski 1970; Daley 1991). The forcings $\mathbf{dw}(t)$ and $\mathbf{v}_k$ have zero mean and respective covariance $\mathbf{Q}(t)dt$ and $\mathbf{R}_k$, with $\mathcal{E}\{\mathbf{dw}(t + \delta)\mathbf{dw}^T(t)\} = 0 \; \forall \delta \neq 0; \; \mathcal{E}\{\mathbf{v}_k\mathbf{v}_j^T\} = 0$ for $k \neq j$; and $\mathcal{E}\{\mathbf{dw}(t)\mathbf{v}_k^T\} = 0 \; \forall k$. In passing, the probability densities of the random forcings are not formally required to be Gaussian. Following Jazwinski (1970), for a given functional g, the notations $\mathcal{E}\{g\}$ and $\hat{g}$ refer to the statistical mean. For an ensemble of size q, the sample mean operator is written $\mathcal{E}^q\{\cdot\}$. The conditional mean state of $\boldsymbol{\psi}_k^t$ is denoted by $\hat{\boldsymbol{\psi}}_k^t$; an estimate of this mean state is denoted by $\hat{\boldsymbol{\psi}}_k$. At a given time t_k, the white random processes $\mathbf{dw}(t_k)$ are uncorrelated to the error, $(\hat{\boldsymbol{\psi}}_k - \boldsymbol{\psi}_k^t)$. The state error covariance at t_k is defined by $\mathbf{P}_k \doteq \mathcal{E}\{(\hat{\boldsymbol{\psi}}_k - \boldsymbol{\psi}_k^t)(\hat{\boldsymbol{\psi}}_k - \boldsymbol{\psi}_k^t)^T\} \in \mathbb{R}^{n \times n}$. To refer to quantities before and after assimilation, the adjectives a priori and a posteriori are used, respectively. In mathematical terms, a $(-)$ and a $(+)$ distinguish the two. For the singular vectors in section 6, the $(-)$ and $(+)$ are simplified to subscripts. A smoothing estimate at t_k is denoted by $\hat{\boldsymbol{\psi}}_{k/N}$ (e.g., Gelb 1974; Catlin 1989): the index k/N indicates that all observations made up to t_N are used.

State vector augmentation is used in (A1) to describe time-correlated random forcings of the deterministic model. For example, random walks, ramps, or exponentially time-correlated random processes are considered as dynamics such that the enlarged system (A1) is only excited by unbiased white noise (Gelb 1974). External forcings are assumed part of $\boldsymbol{\psi}^t$ in (A1) since they may evolve with time and feedback between external forcings and internal dynamics exist. Similarly, the boundary conditions, which are nonlinear relations between internal and boundary state variables, have an evolution equation. They are here part of (A1). Finally, parameter estimation is included in (A1) by adding a stochastic evolution equation for each parameter to be estimated. The products of parameters and original state variables then introduce new nonlinearities. To limit the size of the augmented $\boldsymbol{\psi}^t$ the parameters can be expanded into (local) unknown coefficients functionals (parameter EOFs) instead of gridpoint discretized fields.

APPENDIX B

Specifics and Variations of the ESSE Schemes

a. Normalization in multivariate ES

In most cases, ocean and atmosphere models are multivariate. For the ES estimation not to be sensitive to field units, a normalization is needed (Preisendorfer 1988). Field nondimensionalization is not adequate. For instance, salinity in psu and temperature in Celsius have similar orders of magnitude but, relative to temperature variations, small errors in salinity can lead to large errors in velocities. It is not the fields but their errors that need to be comparable. Each sample error field is thus divided by its volume and sample-averaged error variance. Details are in Lermusiaux (1997). The normalization is necessary in the SVDs and multivariate ES convergence criterion (appendix B, section b). For the minimum ES variance update (Table 3), the error singular vectors are redimensionalized prior to computations.

b. Quantitative ES divergence/convergence criteria

When the dominant error eigendecomposition for any given time is quasi-insensitive to new information, the principal covariance (5) has converged and the melding (6b) can be performed. Within the ensemble scheme chosen (Table 5), the new information is the value added by new forecasts. Monte Carlo integrations can thus be stopped when the dominant SVD of error samples (41) stablizes. Let us assume that $r \geq 1$ new forecasts have been carried out in a parallel batch (Fig. 2), and that the rank-p SVD of the "previous" error sample matrix,

$$\text{SVD}_p(\mathbf{M}) = \mathbf{E}\boldsymbol{\Sigma}\mathbf{V}^T \in \mathbb{R}^{n \times p}, \tag{B1}$$

and rank-$\tilde{p}$ SVD of the matrix formed of the previous and r new $\mathbf{M}^r$ error samples,

$$\text{SVD}_{\tilde{p}}([\mathbf{M}|\mathbf{M}^r]) = \tilde{\mathbf{E}}\tilde{\boldsymbol{\Sigma}}\tilde{\mathbf{V}}^T \in \mathbb{R}^{n \times \tilde{p}}, \tag{B2}$$

are available, where $\tilde{\mathbf{E}} \in \mathbb{R}^{n \times \tilde{p}}$, $\tilde{\boldsymbol{\Sigma}} \in \mathbb{R}^{\tilde{p} \times \tilde{p}}$, and $\tilde{\mathbf{V}}^T \in \mathbb{R}^{\tilde{p} \times \tilde{p}}$, with $\tilde{p} \geq p$. The associated principal error covariances are, respectively, $\mathbf{P}^p = \mathbf{E}\boldsymbol{\Pi}\mathbf{E}^T$, where $\boldsymbol{\Pi} = (1/q)\boldsymbol{\Sigma}^2$; and $\mathbf{P}^{\tilde{p}} = \tilde{\mathbf{E}}\tilde{\boldsymbol{\Pi}}\tilde{\mathbf{E}}^T$, where $\tilde{\boldsymbol{\Pi}} = (1/\tilde{q})\tilde{\boldsymbol{\Sigma}}^2$, with $\tilde{q}$

$= q + r > q$. In accord with sections 4 and 5c, the goal is to compute the similarity between the amplitude and structure of these two covariances. One would like to find out how close $\mathbf{E}\mathbf{\Pi}^{1/2} \in \mathbb{R}^{n \times p}$ is to $\tilde{\mathbf{E}}\tilde{\mathbf{\Pi}}^{1/2} \in \mathbb{R}^{n \times \tilde{p}}$. For coherence with the variance measures (6a)–(6c), a logical similarity coefficient ρ is

$$\rho = \frac{\sum_{i=1}^{k} \sigma_i(\mathbf{\Pi}^{1/2}\mathbf{E}^{\mathrm{T}}\tilde{\mathbf{E}}\tilde{\mathbf{\Pi}}^{1/2})}{\sum_{i=1}^{\tilde{p}} \sigma_i(\mathbf{P}^{\tilde{p}})} = \frac{\sum_{i=1}^{k} \sigma_i^{1/2}(\mathbf{P}^p\mathbf{P}^{\tilde{p}})}{\sum_{i=1}^{\tilde{p}} \sigma_i(\mathbf{P}^{\tilde{p}})}, \quad \text{(B3)}$$

where $k = \min(\tilde{p}, p)$ and $\sigma_i(\,\cdot\,)$ selects the singular value number i. If $\mathrm{tr}(\tilde{\mathbf{\Pi}}) \geq \mathrm{tr}(\mathbf{\Pi})$, the coefficient $\rho \leq 1$. The equality holds when $\mathbf{P}^{\tilde{p}} = \mathbf{P}^p$ and one stops the integrations when $\rho \simeq 1$. There are variations of (B3). One can also ensure that the variance and structure of each of the new r forecast can be sufficiently explained by $\mathbf{E}\mathbf{\Pi}^{1/2}$. There are then r coefficients ρ of form analogous to (B3) to evaluate. Other criteria consist of increasing the ensemble size until the new members yield insignificant reductions in the a posteriori data residuals or insignificant changes in $\hat{\boldsymbol{\psi}}(+)$ (17).

c. Error subspace forecast variations

For an efficient account of nonlinear effects, the ensemble method (Table 5) was preferred to integrate the dominant errors. Several alternate error forecasts are discussed next.

Iterative error breeding in between DA times generalizes the breeding of perturbations of Toth and Kalnay (1993). This new breeding (Fig. 2) uses the error ensemble forecast to time t_{k+1} (Table 5) for iterative improvements of the error initial condition at time t_k. Once the error ensemble forecast ℓ is made, the simplest approach is to rescale the ES forecast coefficients, $\mathbf{\Pi}_{k+1}^{\ell}(-)$ in (41), to their initial norm, hence imposing $\mathbf{\Pi}_k^{\ell+1}(+) = [\mathrm{tr}\mathbf{\Pi}_k^{\ell}(+)/\mathrm{tr}\mathbf{\Pi}_{k+1}^{\ell}(-)]\mathbf{\Pi}_{k+1}^{\ell}(-)$ and set $\mathbf{E}_k^{\ell+1}(+) = \mathbf{E}_{k+1}^{\ell}(-)$. This was used to test simulations in Part II. If the dynamics and model error statistics are locally steady, convergence to the local, steady-state principal component approximation of (37) is possible. For more realistic cases, the iterated $\mathbf{\Pi}_k^{\ell+1}(+)$ and $\mathbf{E}_k^{\ell+1}(+)$ can be obtained from an approximate inversion of (40)–(41). Breeding can be also combined with *shooting*. One may shoot for an initial ES that leads to $\hat{\boldsymbol{\psi}}_{k+1}(+)$ in best accord with the measurements at t_{k+1} (Fig. 2). This smoothing technique for determining optimal initial error conditions does not require an adjoint model.

A *tangent linear model* (TLM) and its adjoint can be used to search for the dominant right and left singular error vectors, embracing the classic search for optimal perturbations (e.g., Farrell and Moore 1992; Sundqvist 1993; Errico et al. 1993; Moore and Farrell 1994; Ehrendorfer and Errico 1995; Molteni et al. 1996). Since TLM forecasts are commonly shown to be similar to nonlinear forecasts for a limited duration, TLMs should perhaps only be used to derive local adjoint models. In a search for singular vectors, the nonlinear model would then be run forward and the linear adjoint backward for approximate back integrations. In fact, by nonlinear interactions, the fastest growing singular vectors of the TLMs interact/modify the basic state the most and the fastest. The duration for which a linearly estimated singular vector is reliable decreases proportionally with the vector's growth rate. Utilizing TLMs in forward computations for DA thus requires care.

The filtering algorithms described by Tables 3–5 are only based on *Monte Carlo nonlinear error forecasts* so as to satisfy our goals (section 3). It is related to but differs from the strict ensemble scheme (Evensen 1994a,b, 1997b; van Leeuwen and Evensen 1996). Its theoretical and practical advantages are now discussed. First, the ES approach brings a framework for validating the ensemble scheme. It permits the quantitative assessment of added value by new forecasts (appendix B, section b). For a given criterion, for example, $\rho = 98\%$ in (B3), the size of the ES is allowed to evolve with the dynamics and data available (2a)–(2b). In light of the intermittent and burst ocean processes, and of the often eclectic and variable data coverage, this property is important. Second, the central processing unit (CPU) requirements are reduced. For $p = q$, the ensemble and ESSE melding lead to the same a posteriori estimates, but in ESSE only one melding is necessary. For $p < q$, the present melding occurs in the significant subspace of the sample errors, further reducing computations at least by a factor q/p (Part II). Third, organizing errors according to their variance allows physical analyses of their dominant components. Such analyses can lead to adequate simplifications of DA schemes. If the errors are of numerical nature, they are usually distinguishable in the dominant ES structures. Algorithms and codes can be fixed; ESSE can be used for model verification. The main advantage of the present statistical smoothing (Tables 6, 7) is the use of the nonlinear filtering evolution as the starting point in the smoothing. The classic ensemble smoother (e.g., Evensen and van Leeuwen 1996) starts from a pure forecast, which increases the potential of divergence. Finally, the present schemes open the doors toward DA based on subspace trackers (Lermusiaux 1997).

In *real-time and management operations,* it might not be necessary to forecast the ES continuously. A stationary, historical, or climatological ES could suffice in specific conditions. If the dynamics and data statistics (2a)–(2b) are stationary, all possible dominant error vectors for the region studied can be evaluated in advance and stored, just as one can store the classic vertical EOFs of a region. Only the principal error values are then forecast. From experience, principal error value models could be derived. For instance, one may assume exponential growth in between each assimilation. Another practical method is to forecast the ES for a central as-

simulation time and to use it at other times for the time ramping of observations. *Analytical* ES can also be defined and projected onto the most dominant variability subspace of a given ocean state by a priori ensemble runs.

d. Criterion for "best" forecast selection

Even when the probability density forecast associated with (2a)–(2b) is available, the question of what should be the estimate of the geophysical state is still primordial, especially in practice (e.g., Robinson et al. 1989). Each sensible choice corresponds to a criterion defining the "best" or optimal forecast. Several such criteria are discussed next. The simple theoretical concepts presented are linked to the current schemes and illustrated with Gulf Stream scenarios.

A good estimate should obviously have small expected forecast errors. The state that is closest to the truth (2a)–(2b), in the sense of any convex loss function or measure[6] of the expected error, is the conditional mean (34). Here its logical statistical estimate is the *ensemble mean* $\hat{\boldsymbol{\psi}}_{k+1}^{\mathrm{em}}(-)$ given by (36b). It approximates the conditional mean with a standard deviation error decay of $O(1/\sqrt{q})$. However, $\hat{\boldsymbol{\psi}}_{k+1}^{\mathrm{em}}(-)$ may be "too smooth" or very unlikely, even though it tends to be the closest to the truth (2a) in the convex-measure sense. Considering the meandering Gulf Stream with several rings, (34) would smooth out certain scales and reduce physical gradients as a function of the predictability limits of the considered processes. A practical solution is to set the best forecast to the ensemble member $\hat{\boldsymbol{\psi}}_{k+1}^{j}(-)$ that is the closest to the ensemble mean. This combines the properties of likehood, that is, the available physical properties (2a) of the dynamics, with that of the "best guess" in the expected convex measure sense. For example, instead of smoothing out the meandering front and averaging uncertain rings, this criterion would select the forecast whose frontal axis is the closest to the mean one and that is in the process of shedding rings. All realistic gradients and other physical properties are then maintained.

Other options considered here are (Fig. 2) the central forecast (35), most probable forecast, and forecast of minimum data misfits. The nonlinear *central forecast* $\hat{\boldsymbol{\psi}}_{k+1}^{\mathrm{cf}}(-)$ is the classic deterministic forecast. It is the first-order estimate of the conditional mean. In Part II, which has no model errors, $\hat{\boldsymbol{\psi}}_{k+1}^{\mathrm{cf}}(-)$ is found slightly superior to $\hat{\boldsymbol{\psi}}_{k+1}^{\mathrm{em}}(-)$ in the rms data misfit sense. The central and ensemble mean forecasts are found close in the convex measure sense, but the central forecast contains more "realistic" features and, hence, is more like-

ly. These properties are as those of the member closest to the ensemble mean promoted above. In several other primitive equation simulations (e.g., Lermusiaux 1997), the central and ensemble mean forecasts had comparable data misfits but the central forecast was qualitatively better. Of course, this depends on the ensemble size and may vary with the simulation considered. For example, Evensen and van Leeuwen (1996) have reached other conclusions in quasigeostrophic simulations. The *most probable forecast* can also be estimated from the ensemble. A probability density (histogram) is first computed for each of the n state variables (i.e., elements of $\boldsymbol{\psi}$). The range of values taken by the q members is divided into a given number of intervals and the members assigned to their respective interval, forming the histogram. For each variable, the most probable value is the center of the tallest segment or bar. The most probable forecast $\hat{\boldsymbol{\psi}}_{k+1}^{\mathrm{mp}}(-)$ is defined by these values or, to ensure that (2a) is satisfied, by the ensemble member $\hat{\boldsymbol{\psi}}_{k+1}^{j}(-)$ of minimum averaged rms difference with these values. Selecting $\hat{\boldsymbol{\psi}}_{k+1}^{\mathrm{mp}}(-)$ as the best forecast is motivated by the non-Gaussian nature of the probabilities of the nonlinear model (2a). In fact, if $\hat{\boldsymbol{\psi}}_{k+1}^{\mathrm{em}}(-)$ is found to be quite different than $\hat{\boldsymbol{\psi}}_{k+1}^{\mathrm{mp}}(-)$, an option is to choose $\hat{\boldsymbol{\psi}}_{k+1}^{\mathrm{mp}}(-)$ as the best forecast and to reduce the ensemble to the subset of members of probability density locally convex around $\hat{\boldsymbol{\psi}}_{k+1}^{\mathrm{mp}}(-)$. With the outliers removed, minimum error variance can then be applied for *local* conditional mean estimation. If the subset is not selected, $\hat{\boldsymbol{\psi}}_{k+1}^{\mathrm{mp}}(-)$ should be updated from the conditional probability given the data, using Bayes theorem (e.g., Jazwinski 1970). A practical issue that we found is that the differences between the most probable and central forecasts were sometimes dominated by numerics (e.g., Shapiro filter). Finally, at analysis time, the *forecast of minimum data misfits* can be chosen among the stochastic ensemble. This is, in a sense, an adjoint approach in the data domain: instead of neglecting model errors, the more frequent small data errors are neglected. However, as for $\hat{\boldsymbol{\psi}}_{k+1}^{\mathrm{mp}}(-)$, one may need to subsample the ensemble. In addition, it is only the best forecast within the data domain, not over the whole modeling domain.

e. Algorithmic issues

The matrix $\tilde{\mathbf{C}}^p \doteq \mathbf{CE}_-$ is evaluated so that $\mathbf{E}_-\boldsymbol{\Pi}(-)\mathbf{E}^{\mathrm{T}}$ is never formed. The ES covariance mapped onto the data space, $\tilde{\mathbf{C}}^p\boldsymbol{\Pi}(-)\tilde{\mathbf{C}}^{p^{\mathrm{T}}}$, is not always invertible but the positive-definite $\mathbf{R}$ should guarantee the feasibility of the inversion in (18). For stability, a truncated SVD inverse can be used (e.g., Bennett 1992). Sequential processing of observations (Parrish and Cohn 1985; Cho et al. 1996) is adequate in Table 3 since (17)–(18) yield minimum variance in the ES. It is employed: data are assimilated one after the other, with, if necessary, a Choleski predecomposition of $\mathbf{R}$. In Table 3, the sparse matrices $\tilde{\mathbf{C}}^p = \mathbf{CE}_-$ are evaluated via pointers. Only non-

---

[6] For simplicity, the class of loss functions defined in Jazwinski (1970, 146–150) is here referred to as convex measures. This terminology is, in fact, a subclass, since the measure does not need to be convex.

null elements are retained. The size of the ES evolves in function of (B3). When ρ is larger than a threshold, iterations are stopped. Parallel computing is used, depending on availability. The cost of the ESSE system shown in Fig. 2 is often driven by the ensemble size. In terms of the number of forecasts made, this cost is of order q. The OI is then of order 1 and the full covariance scheme of order n. The representer method rescaled to filtering is of order $2m/2 = m,$ where m is the total number of scalar observations made during the assimilation period. Typical numbers for a 10-day period and three assimilations of 50 CTDs on 20 levels are $n = 3.10^5$, $q = 300$, and $m = 6000$. Assuming that a 1-day forecast is issued in 30 min and that 30 CPUs are available, the 10-day period ESSE takes 50 h. The full covariance scheme takes almost 6 yr, an iteration of the representer method of almost 42 days. Additional computational aspects are discussed in Part II and in Lermusiaux (1997).

REFERENCES

Anderson, J. L., 1996: Selection of initial conditions for ensemble forecasts in a simple perfect model framework. *J. Atmos. Sci.,* **53,** 22–36.

Austin, J. W., and C. T. Leondes, 1981: Statistically linearized estimation of reentry trajectories. *IEEE Trans. Aerosp. Electron. Syst.,* **17** (1), 54–61.

Bendat, J. S., and A. G. Piersol, 1986: *Random Data Analysis and Measurement Procedures.* John Wiley and Sons, 566 pp.

Bengtsson, L., M. Ghil, and E. Kallen, Eds., 1981: *Dynamic Meteorology: Data Assimilation Methods.* Springer-Verlag, 330 pp.

Bennett, A. F., 1992: Inverse methods in physical oceanography. *Cambridge Monographs on Mechanics and Applied Mathematics,* Cambridge University Press, 346 pp.

——, B. S. Chua, and L. M. Leslie, 1996: Generalized inversion of a global numerical weather prediction model. *Meteor. Atmos. Phys.,* **60** (1–3), 165–178.

——, ——, and ——, 1997: Generalized inversion of a global numerical weather prediction model, II: Analysis and implementation. *Meteor. Atmos. Phys.,* **62** (3–4), 129–140.

Bergé, P., Y. Pomeau, and C. Vidal, 1988: *L'Ordre dans le Chaos. Vers une Approche Déterministe de la Turbulence.* Wiley Interscience, 329 pp.

Boguslavskij, I. A., 1988: *Filtering and Control.* Optimization Software, 380 pp.

Brockett, R. W., 1991: Dynamical systems that learn subspaces. *Mathematical Systems Theory: The Influence of R. E. Kalman,* A. Antoulas, Ed., Springer-Verlag, 579–592.

Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.,* **126,** 1719–1724.

Catlin, D. E., 1989: *Estimation, Control, and the Discrete Kalman Filter.* Vol. 71, *Applied Mathematical Sciences,* Springer-Verlag, 274 pp.

Charney, J. G., and G. R. Flierl, 1981: Oceanic analogues of large-scale atmospheric motions. *Evolution of Physical Oceanography: Scientific Surveys in Honor of Henry Stommel,* B. Warren and G. Wunsch, Eds., The MIT Press, 504–548.

Charnock, H., 1981: Air–sea interaction. *Evolution of Physical Oceanography: Scientific Surveys in Honor of Henry Stommel,* B. Warren and G. Wunsch, Eds., The MIT Press, 482–503.

Cho, Y., V. Shin, M. Oh, and Y. Lee, 1996: Suboptimal continuous filtering based on the decomposition of the observation vector. *Comput. Math. Appl.,* **32** (4), 23–31.

Cohn, S. E., 1993: Dynamics of short-term univariate forecast error covariances. *Mon. Wea. Rev.,* **121,** 3123–3149.

——, and D. F. Parrish, 1991: The behavior of forecast error covariances for a Kalman filter in two dimensions. *Mon. Wea. Rev.,* **119,** 1757–1785.

Daley, R., 1991: *Atmospheric Data Analysis.* Cambridge University Press, 457 pp.

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

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

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

Davis, M. H. A., 1977a: *Linear Estimation and Stochastic Control.* Chapman-Hall, 224 pp.

Davis, R. E., 1977b: Techniques for statistical analysis and prediction of geophysical fluid systems. *Geophys. Astrophys. Fluid Dyn.,* **8,** 245–277.

Dee, D. P., 1990: Simplified adaptive Kalman filtering for large-scale geophysical models. *Realization and Modelling in System Theory,* M. A. Kaashoek, J. H. van Schuppen, and A. C. M. Ran, Eds., *Proceedings of the International Symposium MTNS-89,* Vol. 1, Birkhäuser, 567–574.

——, S. E. Cohn, A. Dalcher, and M. Ghil, 1985: An efficient algorithm for estimating noise covariances in distributed systems. *IEEE Trans. Control.,* **AC-30,** 1057–1065.

Ehrendorfer, M., and R. M. Errico, 1995: Mesoscale predictability and the spectrum of optimal perturbations. *J. Atmos. Sci.,* **52,** 3475–3500.

Errico, R. M., T. E. Rosmond, and J. S. Goerss, 1993: A comparison of analysis and initialization increments in an operational data-assimilation system. *Mon. Wea. Rev.,* **121,** 579–588.

Evensen, G., 1993: Open boundary conditions for the extended Kalman filter with a quasi-geostrophic ocean model. *J. Geophys. Res.,* **98,** 16 529–16 546.

——, 1994a: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.,* **99** (C5), 10 143–10 162.

——, 1994b: Inverse methods and data assimilation in nonlinear ocean models. *Physica D,* **77,** 108–129.

——, 1997a: Advanced data assimilation for strongly nonlinear dynamics. *Mon. Wea. Rev.,* **125,** 1342–1354.

——, 1997b: Application of ensemble integrations for predictability studies and data assimilation. *Monte Carlo Simulations in Oceanography: Proc. Hawaiian Winter Workshop,* Honolulu, HI, Office of Naval Research and School of Ocean and Earth Science and Technology, University of Hawaii at Manoa, 11–22.

——, and P. J. van Leeuwen, 1996: Assimilation of Geosat altimeter data for the Agulhas Current using the ensemble Kalman filter with a quasigeostrophic model. *Mon. Wea. Rev.,* **124,** 85–96.

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

——, and P. J. Ioannou, 1996a: Generalized stability theory. Part I: Autonomous operators. *J. Atmos. Sci.,* **53,** 2025–2040.

——, and ——, 1996b: Generalized stability theory. Part II: Nonautonomous operators. *J. Atmos. Sci.,* **53,** 2041–2053.

Foias, C., and R. Teman, 1977: Structure of the set of stationary solutions of the Navier–Stokes equations. *Commun. Pure Appl. Math.,* **30,** 149–164.

——, and ——, 1987: The connection between the Navier–Stokes equations, dynamical systems and turbulence. *Directions in Partial Differential Equations,* M. G. Grandall, P. H. Rabinowitz, and E. E. L. Turner, Eds., Academic Press, 55–73.

Fukumori, I., and P. Malanotte-Rizzoli, 1995: An approximate Kalman filter for ocean data assimilation: An example with one idealized Gulf Stream model. *J. Geophys. Res.,* **100,** 6777–6793.

——, J. Benveniste, C. Wunsch, and D. B. Haidvogel, 1993: Assimilation of sea surface topography into an ocean circulation model using a steady-state smoother. *J. Phys. Oceanogr.,* **23,** 1831–1855.

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

Gelb, A., Ed., 1974: *Applied Optimal Estimation.* The MIT Press, 374 pp.

Ghil, M., 1989: Meteorological data assimilation for oceanographers. Part I: Description and theoretical framework. *Dyn. Atmos. Oceans,* **13** (3–4), 171–218.

Hasselmann, K., 1988: PIPs and POPs. A general formalism for the reduction of dynamical systems in terms of principal interaction patterns and principal oscillation patterns. *J. Geophys. Res.,* **93,** 11 015–11 021.

Horn, R. A., and C. R. Johnson, 1985: *Matrix Analysis.* Cambridge University Press, 561 pp.

——, and ——, 1991: *Topics in Matrix Analysis.* Cambridge University Press, 607 pp.

Jazwinski, A. H., 1970: *Stochastic Processes and Filtering Theory.* Academic Press, 376 pp.

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

Kolmogorov, A. N., 1941: *Dokl. Akad. Nauk SSSR,* **30,** 301; **32,** 16.

Le Dimet, F. X., and O. Talagrand, 1986: Variational algorithms for analysis and assimilation of meteorological observations. *Tellus,* **38A,** 97–110.

Lermusiaux, P. F. J., 1997: Error subspace data assimilation methods for ocean field estimation: Theory, validation and applications. Ph.D. thesis, Harvard University, Cambridge, MA, 402 pp.

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

——, 1999b: Estimation and study of mesoscale variability in the Strait of Sicily. *Dyn. Atmos. Oceans,* in press.

Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.,* **112,** 1177–1194.

——, 1992: Iterative analysis using covariance functions and filters. *Quart. J. Roy. Meteor. Soc.,* **118,** 569–591.

——, R. S. Bell, and B. Macpherson, 1991: The Meteorological Office analysis correction data assimilation scheme. *Quart. J. Roy. Meteor. Soc.,* **117,** 59–89.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.,* **20,** 130–141.

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

Lozano, C. J., A. R. Robinson, H. G. Arango, A. Gangopadhyay, N. Q. Sloan, P. J. Haley, and W. G. Leslie, 1996: An interdisciplinary ocean prediction system: Assimilation strategies and structured data models. *Modern Approaches to Data Assimilation in Ocean Modelling,* P. Malanotte-Rizzoli, Ed., Elsevier Oceanography Series, Elsevier Science, 413–432.

Martel, F., and C. Wunsch, 1993: Combined inversion of hydrography, current meter data and altimetric elevations for the North Atlantic circulation. *Manuscripta Geodaetica,* **18** (4), 219–226.

McWilliams, J. C., W. B. Owens, and B. L. Hua, 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., and B. D. Cornuelle, 1999: Forecasts from fits of frontal fluctuations. *Dyn. Atmos. Oceans,* in press.

Miller, R. N., and M. A. Cane, 1989: A Kalman filter analysis of sea level height in the tropical Pacific. *J. Phys. Oceanogr.,* **19,** 773–790.

——, M. Ghil, and F. Gauthier, 1994: Data assimilation in strongly nonlinear dynamical systems. *J. Atmos. Sci.,* **51,** 1037–1056.

——, E. F. Carter, and S. T. Blue, cited 1998: Data assimilation into nonlinear stochastic models. [Available online at http://tangaroa.oce.orst.edu/stochast.html.]

Molteni, F., and T. N. Palmer, 1993: Predictability and finite-time instability of the northern winter circulation. *Quart. J. Roy. Meteor. Soc.,* **119,** 269–298.

——, R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The ECMWF ensemble prediction system: Methodology and validation. *Quart. J. Roy. Meteor. Soc.,* **122,** 73–119.

Monin, A. S., 1974: *Variability of the Oceans.* Wiley, 241 pp.

Moore, A. M., and B. F. Farrell, 1994: Using adjoint models for stability and predictability analysis. *NATO ASI Ser.,* Vol. 119, 217–239.

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

Osborne, A. R., and A. Pastorello, 1993: Simultaneous occurrence of low-dimensional chaos and colored random noise in nonlinear physical systems. *Phys. Lett. A,* **181,** 159–171.

Parrish, D. F., and S. E. Cohn, 1985: A Kalman filter for a two-dimensional shallow-water model: Formulation and preliminary experiments. Office Note 304, NOAA/NWS/NMC, 64 pp.

Penland, C., 1989: Random forcing and forecasting using principal oscillation pattern analysis. *Mon. Wea. Rev.,* **117,** 2165–2185.

Phillips, N. A., 1986: The spatial statistics of random geostrophic modes and first-guess errors. *Tellus,* **38A,** 314–332.

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

Rabier, F., E. Klinker, P. Courtier, and A. Hollingsworth, 1996: Sensitivity of forecast errors to initial conditions. *Quart. J. Roy. Meteor. Soc.,* **122,** 121–150.

Reid, W. T., 1968: Generalized inverses of differential and integral operators. *Theory and Applications of Generalized Inverse of Matrices,* T. L. Bouillon and P. L. Odell, Eds., Lublock, 1–25.

Robinson, A. R., 1989: *Progress in Geophysical Fluid Dynamics.* Vol. 26, *Earth-Science Reviews,* Elsevier Science.

——, M. A. Spall, and W. G. Leslie, 1989: Data assimilation and dynamical interpolation in gulfcast experiments. *Dyn. Atmos. Oceans,* **13** (3–4), 301–316.

——, H. G. Arango, A. J. Miller, A. Warn-Varnas, P.-M. Poulain, and W. G. Leslie, 1996a: Real-time operational forecasting on shipboard of the Iceland–Faeroe frontal variability. *Bull. Amer. Meteor. Soc.,* **77,** 243–259.

——, A. Warn-Varnas, W. G. Leslie, A. J. Miller, P. J. Haley, and C. J. Lozano, 1996b: Real-time regional forecasting. *Modern Approaches to Data Assimilation in Ocean Modeling,* P. Malanotte-Rizzoli, Ed., Elsevier Science, 455 pp.

——, J. Sellschopp, A. Warn-Varnas, W. G. Leslie, C. J. Lozano, P. J. Haley Jr., L. A. Anderson, and P. F. J. Lermusiaux, 1997: The Atlantic Ionian Stream, *J. Mar. Syst.,* in press.

——, P. F. J. Lermusiaux, and N. Q. Sloan III, 1998a: Data assimilation. *Processes and Methods,* K. H. Brink and A. R. Robinson, Eds., *The Sea: The Global Coastal Ocean I,* Vol. 10, John Wiley and Sons.

——, and Coauthors, 1998b: The Rapid Response 96, 97 and 98 exercises: The Strait of Sicily, Ionian Sea and Gulf of Cadiz. Harvard Open Ocean Model Rep., Rep. in Meteorology and Oceanography 57, 45 pp. [Available from Harvard Oceanography Group, DEAS, 29 Oxford St., Cambridge, MA 02138.]

Sasaki, Y., 1970: Some basic formalism in numerical variational analysis. *Mon. Wea. Rev.,* **98,** 875–883.

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

Sundqvist, H., Ed., 1993: Special issue on adjoint applications in dynamic meteorology. *Tellus,* **45A** (5), 341–569.

Tarantola, A., 1987: *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation.* Elsevier, 613 pp.

Teman, R., 1991: Approximation of attractors, large eddy simulations and multiscale methods. *Proc. Roy. Soc. London,* **434A,** 23–29.

Todling, R., and M. Ghil, 1990: Kalman filtering for a two-layer two-dimensional shallow-water model. *Proc. WMO Int. Symp. on*

Assimilation of Observations in Meteorology and Oceanography, Clermont-Ferrand, France, WMO, 454–459.

——, and S. E. Cohn, 1994: Suboptimal schemes for atmospheric data assimilation based on the Kalman filter. *Mon. Wea. Rev.,* **122,** 2530–2557.

——, and M. Ghil, 1994: Tracking atmospheric instabilities with the Kalman filter. Part I: Methodology and one-layer results. *Mon. Wea. Rev.,* **122,** 183–204.

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.,* **74,** 2317–2330.

Uchino, E., M. Ohta, and H. Takata, 1993: A new state estimation method for a quantized stochastic sound system based on a generalized statistical linearization. *J. Sound Vibration,* **160** (2), 193–203.

van Leeuwen, P. J., and G. Evensen, 1996: Data assimilation and inverse methods in terms of a probabilistic formulation. *Mon. Wea. Rev.,* **124,** 2898–2913.

von Storch, H., and C. Frankignoul, 1998: Empirical modal decomposition in coastal oceanography. *Processes and Methods,* K. H. Brink and A. R. Robinson, Eds., *The Sea: The Global Coastal Ocean I,* Vol. 10, John Wiley and Sons, 419–455.

——, I. Bruns, I. Fishcler-Bruns, and K. Hasselmann, 1988: Principal oscillation pattern analysis of the 30- to 60-day oscillation in general circulation model equatorial troposphere. *J. Geophys. Res.,* **93** (D9), 11 022–11 036.

Wallace, J. M., C. Smith, and C. S. Bretherton, 1992: Singular value decomposition of wintertime sea surface temperature and 500-mb height anomalies. *J. Climate,* **5,** 561–576.

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

West, B. J., and H. J. Mackey, 1991: Geophysical attractors may be only colored noise. *J. Appl. Phys.,* **69** (9), 6747–6749.

Wunsch, C., 1988: Transient tracers as a problem in control theory. *J. Geophys. Res.,* **93,** 8099–8110.

——, 1996: *The Ocean Circulation Inverse Problem.* Cambridge University Press, 456 pp.