

# A Reduced-Order Kalman Filter for Data Assimilation in Physical Oceanography\*

D. Rozier<sup>†</sup>  
F. Birol<sup>‡</sup>  
E. Cosme<sup>§</sup>  
P. Brasseur<sup>†</sup>  
J. M. Brankart<sup>†</sup>  
J. Verron<sup>†</sup>

**Abstract.** A central task of physical oceanography is the prediction of ocean circulation at various time scales. Mathematical techniques are used in this domain not only for the modeling of ocean circulation but also for the enhancement of simulation through data assimilation. The ocean circulation model of concern here, namely, HYCOM, is briefly presented through its variables, equations, and specific vertical coordinate system. The main part of this paper focuses on the Kalman filter as a data assimilation method, and especially on how this mathematical technique, usually associated with a prohibitively high computing cost for operational sciences, is simplified in order to make it applicable to the simulation of realistic ocean circulation models. Some practical issues are presented, such as a brief explanation about ocean observation systems, together with examples of data assimilation results.

**Key words.** data assimilation, Kalman filter, physical oceanography, numerical simulation

**AMS subject classifications.** 65P99, 37M05, 37N10

**DOI.** 10.1137/050635717

**1. Introduction.** Physical oceanography is a rapidly growing field. The increasing pressure for work on climate systems (global climate warming, El Niño oscillation, etc.) is one important reason. Also, the recent progress in oceanography has triggered more demands from various fields of application aside from climate studies, such as weather forecasting, marine transport, the offshore petrol industry, management of marine environment and fish resources, as well as military marine activities.

---

\*Received by the editors July 8, 2005; accepted for publication (in revised form) June 22, 2006; published electronically July 30, 2007. This work was partially supported by the French Service Hydrographique et Océanographique de la Marine under contract 183 EPSHOM/CMO/BRESM/NP, the U.S. Office of Naval Research under grant N00014-05-1-0731, and the EU-funded MERSEA project under contract SIP3-CT-2003-502885.

<http://www.siam.org/journals/sirev/49-3/63571.html>

<sup>†</sup>Laboratoire des Ecoulements Géophysiques et Industriels (LEGI), Centre National de la Recherche Scientifique/Université Joseph Fourier, BP53, 38041 Grenoble Cedex 9, France (David.Rozier@hmg.inpg.fr, Pierre.Brasseur@hmg.inpg.fr, brankart@hmg.inpg.fr, Jacques.Verron@hmg.inpg.fr).

<sup>‡</sup>Laboratoire d'Etudes en Géophysique et Océanographie Spatiale (LEGOS), 18 Avenue E. Belin, 31401 Toulouse Cedex 9, France (Birol@notos.cst.cnes.fr).

<sup>§</sup>Corresponding author. Laboratoire des Ecoulements Géophysiques et Industriels (LEGI), Centre National de la Recherche Scientifique/Université Joseph Fourier, BP53, 38041 Grenoble Cedex 9, France (Emmanuel.Cosme@hmg.inpg.fr).

The operational branch of physical oceanography became progressively viable thanks to the progress made in observing and measuring the oceans, in numerical modeling, and in data assimilation (see [2] for an overview of current operational oceanography). Indeed numerical models have reached a level of complexity that enables them to reproduce the spatiotemporal variability of ocean circulation with increased realism, for both global and coastal applications. Despite this progress, simulations still need to be enhanced in order to improve their coherence with the accurate measurements provided by satellite-borne observation systems.

A better fit between simulations and observations can be achieved through data assimilation, which is a mathematical technique combining the information provided by measured data with the information provided by a model's forecast to obtain the best possible estimate of the ocean state. This paper is concerned with data assimilation systems, and more particularly with a particular technique, namely, a reduced-order Kalman filter called the SEEK filter (singular evolutive extended Kalman filter), applied to an ocean model called HYCOM, standing for *hybrid coordinates ocean model*. The success of this assimilation system has led to the SEEK filter being utilized by the MERCATOR group (see [1], [9]) and the HYCOM consortium, two of the main international operational oceanography projects.

This paper does not describe new advances in the field. It aims to present the SEEK filter, the technique itself together with its context of use and its achievements, to the community of applied mathematicians. Indeed, although they are central to operational oceanography, oceanographic data assimilation and the SEEK filter are little known by the wider community of applied mathematicians.

After a presentation of the typical equations used in ocean circulation models, the model HYCOM is presented in section 2 focusing on its original vertical coordinates system. Central to this paper, section 3 focuses on the SEEK filter data assimilation system. The principles of sequential data assimilation are presented, first on a general level followed by the specifics of Kalman filters. The problems raised by the application of Kalman filtering to geophysical fluid models are exposed. The SEEK filter, specifically designed to tackle these problems, is then described in detail. The last section illustrates the SEEK filter's practical functioning, including general data acquisition issues, examples of assimilation experiments, and the description of an ongoing real-time experiment. In conclusion, we identify some current trends of operational oceanography where the use and development of mathematical techniques are key points to future success, thereby proving the long-term necessity of interactions between mathematicians and oceanographers.

**2. OGCMs and the HYCOM Model.** The development of data assimilation schemes makes use of a configuration built around a model. The work described in this paper was conducted with the model HYCOM [3], developed at RSMAS, University of Miami. It was built upon the model MICOM (Miami isopycnal coordinate ocean model) from the same university, by adopting an alternative vertical coordinate system, as developed in section 2.2. HYCOM belongs to the most widely used class of numerical ocean models, called ocean global circulation models or OGCMs (see [13] and [28] for an extensive presentation of OGCMs).

**2.1. The Primitive Equations.** OGCMs are built on primitive equations describing the fundamental laws of geophysical fluids, i.e., stratified fluids on the rotating earth. The general formalism aims to ensure conservation of mass, momentum, heat, and salt. The different existing ocean models are therefore distinguished by specific selection of the grid definition and parameterization of subgrid-scale motions, and

of the vertical coordinate system. The primitive equations are given below as they are written in HYCOM, using  $(x, y, s)$  coordinates where  $s$  is a generalized vertical coordinate. They consist of

- a momentum equation (2.1a) for the horizontal velocity vector  $\mathbf{v} = (u, v)$ ;
- a mass continuity equation (2.1b) for the pressure  $p$ ;
- one conservation equation (2.1c) for each thermodynamic variable represented by  $\theta$  (temperature and salinity):

$$(2.1a) \quad \frac{\partial \mathbf{v}}{\partial t_s} + \nabla_s \frac{\mathbf{v}^2}{2} + (\zeta + f)\mathbf{k} \times \mathbf{v} + \left( \frac{\partial s}{\partial t} \frac{\partial p}{\partial s} \right) \frac{\partial \mathbf{v}}{\partial p} + \nabla_s M - p \nabla_s \alpha \\ = -g \frac{\partial \tau}{\partial p} + \left( \frac{\partial p}{\partial s} \right)^{-1} \nabla_s \cdot \left( \nu \frac{\partial p}{\partial s} \nabla_s \mathbf{v} \right),$$

$$(2.1b) \quad \frac{\partial}{\partial t_s} \left( \frac{\partial p}{\partial s} \right) + \nabla_s \cdot \left( \mathbf{v} \frac{\partial p}{\partial s} \right) + \frac{\partial}{\partial s} \left( \frac{\partial s}{\partial t} \frac{\partial p}{\partial s} \right) = 0,$$

$$(2.1c) \quad \frac{\partial}{\partial t_s} \left( \frac{\partial p}{\partial s} \theta \right) + \nabla_s \cdot \left( \mathbf{v} \frac{\partial p}{\partial s} \theta \right) + \frac{\partial}{\partial s} \left( \frac{\partial s}{\partial t} \frac{\partial p}{\partial s} \theta \right) = \nabla_s \cdot \left( \nu \frac{\partial p}{\partial s} \nabla_s \theta \right) + \mathcal{H}_\theta,$$

where  $\alpha$  is the potential specific volume, inverse of the potential density,  $\zeta \equiv \partial v / \partial x_s - \partial u / \partial y_s$  is the relative vorticity,  $M \equiv gz + p\alpha$  is the Montgomery potential,  $f$  is the Coriolis parameter,  $\mathbf{k}$  is the vertical unit vector,  $\nu$  is a variable eddy viscosity/diffusivity coefficient,  $\tau$  is the wind-drag and/or bottom-drag induced shear stress vector, and  $\mathcal{H}_\theta$  represents the sum of diabatic source terms acting on  $\theta$ . Subscripts indicate which variable is held constant during partial differentiation.

**2.2. The Different Vertical Coordinate Systems and the Hybrid System in HYCOM.** In order to represent fully the ocean's complex vertical structure, a model must have an appropriate vertical coordinate system. Within the set of OGCMs, the most important distinction therefore lies in the choice of vertical coordinate system. The three common vertical coordinates are as follows:

- The  $z$ -level coordinate, which represents the vertical distance relative to a geopotential reference level. This is a simple and intuitive coordinate matching the natural concept of depth. The surface mixed layer and unstratified regions are well represented by this coordinate.  $z$ -models have been used for decades and are still nowadays at the forefront, represented, among others, by the modular ocean model (MOM) or the océan parallélisé (OPA) model [28].
- The  $\sigma$ -coordinate is a terrain-following coordinate, equal to zero at the surface and  $-1$  at the ocean bottom. This coordinate provides a smooth representation of the topography of the ocean's bed, called the bottom topography. This type of model, however, does not give a good account of the surface layer or of some physical phenomena in the ocean interior. The use of this  $\sigma$ -coordinate, such as in the Princeton ocean model (POM) [6], has therefore been restricted to regional studies or smaller basin-wide studies.
- The  $\rho$ -coordinate is equal to the potential density  $\rho$ . The density is appropriate as a vertical coordinate because most diffusion processes in the ocean happen along isopycnal surfaces, i.e., surfaces of same density. Since in most OGCMs the hydrostatic equilibrium is assumed, the density increases monotonically from the surface to the ocean bottom.  $\rho$ -coordinate models are

therefore most appropriate in the stably stratified regions of the ocean's interior. Bottom topography is also adequately represented. However, for areas where the ocean is not strongly stratified, such a coordinate system will not perform well. Models using this coordinate include MICOM [4], HIM (Hallberg isopycnal model) [29], and POSUM (parallel Oregon State University model).

Evolving from the  $\rho$ -coordinate model MICOM, the originality of HYCOM lies in its specific hybrid coordinates system, aimed at preserving the advantages of  $\rho$ -coordinate models while acquiring the advantages of the other coordinate systems in regions where the  $\rho$ -coordinate is not the most appropriate. Along the same lines, the model POSEIDON also uses a hybrid but different vertical coordinate system [31].

HYCOM switches between different vertical coordinate systems in different three-dimensional regions of the ocean [3], [14]. The reason for this choice is the inability of any vertical grid to perform well everywhere. For example, the  $\sigma$ -coordinate (terrain following) performs well on the shelf (the shallow part of the ocean, close to the coast) where increased vertical resolution is desired in shallow water, but it introduces errors in the calculation of horizontal gradients near steeply sloping bottom topography. The  $z$ -level (geopotential) coordinate eliminates this problem near steep slopes but is not appropriate for both shallow and deep water areas, and represents the topography by steps.

Due to its hybrid coordinate system, HYCOM uses an appropriate coordinate system within any three-dimensional region in the domain. It switches between

- the  $\sigma$ -coordinate in unstratified shallow coastal regions,
- the  $z$ -level coordinate in the mixed layer and in other unstratified regions,
- and the  $\rho$ -coordinate everywhere the ocean is stratified.

The choice of coordinate type varies in time and space, and the optimal choice is updated every time step. The layered continuity equation (2.1b) assures a smooth transition between the different coordinate types.

This hybrid coordinate system is a definite advantage as soon as the studied ocean region includes shallow or unstratified regions as well as open ocean areas, for example, the North Atlantic ocean including the Gulf of Mexico [14]. Further details concerning all characteristics of HYCOM can be found in [3] and at <http://hycom.rsmas.miami.edu>. Because of its original vertical coordinate system, HYCOM is a state-of-the-art ocean circulation model that is widely used by oceanographers worldwide, including our team, for developing configurations allowing the study of advanced data assimilation schemes.

**3. Sequential Data Assimilation and the Kalman Filter.** Ocean circulation simulations suffer mainly from three types of drawbacks:

- Model deficiencies: problems due to the resolution, the numerical scheme, the physical misrepresentation of some process, etc.
- Erroneous forcing: false information can be given to the model through incorrect forcing data such as wind, heat flux, or fresh water flux forcing.
- Incorrect initial conditions: this is an obvious and unavoidable difficulty in such a complex system.

The adequacy of a model's reproduction of reality with independent measurements is therefore limited, to a different extent depending on the model. Consequently, data assimilation techniques are widely used in order to enhance the quality of simulations.

In a generic way, data assimilation can be defined as the synthesis of various sources of information available about one system in order to get the best possible

estimates of this system. In oceanography, the two sources of information are the knowledge of the physics governing the system, transcribed in the model, and the observations (measurements) made of this system. Thus data assimilation can be summarized here as a dynamical interpolation of observations using a nonlinear model of the ocean's dynamics. Data assimilation problems belong to the category of inverse problems, since they are concerned with the reconstitution of a system from some data produced by this system.

**3.1. Sequential Data Assimilation.** Data assimilation consists practically of modifying the state vector of a numerical model (in primitive equation models such as HYCOM, this typically includes the temperature, salinity, velocities, and sea-surface height two-dimensional field) to account for information contained in observations of the real system.

Various stochastic and nonstochastic data assimilation methods exist (see [26], [25] for reviews). Nonstochastic methods in use for oceanography include in particular the four-dimensional variational (4D-Var) scheme. Here, we are concerned with a subset of stochastic methods called the sequential methods. Sequential data assimilation filters operate in three steps:

- An initial probability distribution of state variables at the initial time is provided.
- The statistics of the state variables evolve over time with the system dynamics.
- The observations modify the probability distribution of state variables. In Kalman filtering techniques, the mean of this distribution is updated by a linear combination of the innovations (innovations are the misfits between the observations and the system's variables), and it is assumed that the measurement and model errors are unbiased and Gaussian.

Sequential data assimilation techniques include the Kalman filters, on which we will now focus.

**3.2. The Kalman Filter.** The Kalman filter (see [32], [27], [15] for reviews) is a sequential method, composed of two steps: a forecast step followed by an analysis step.

- The forecast step uses the dynamical model to provide the system's current forecast, i.e., the state of the system obtained by integrating the model's equations from the previously estimated state of the system,
- The analysis step provides a new estimate of the system's state. This estimate is obtained by correcting the forecasted state of the system using the innovation, while taking into account the errors in both the model's state and the observations.

For the general case, the equations for these two steps are given below:

- Forecast step: The current forecast state is given by the integration of the model, assumed linear here, from the previously estimated state:

$$(3.1) \quad \mathbf{x}_k^f = \mathbf{M}_{k-1,k} \mathbf{x}_{k-1}^a.$$

The current forecast of the error covariance matrix  $\mathbf{P}$  is given by the transformation of the previously estimated error covariance matrix using the model  $\mathbf{M}$  and the model error matrix  $\mathbf{Q}_k$ :

$$(3.2) \quad \mathbf{P}_k^f = \mathbf{M}_{k-1,k} \mathbf{P}_{k-1}^a \mathbf{M}_{k-1,k}^T + \mathbf{Q}_k.$$

- Analysis step: The current estimated state  $\mathbf{x}_k^a$  is given by the current forecast state  $\mathbf{x}_k^f$  corrected by the Kalman gain  $\mathbf{K}_k$  applied to the innovation vector ( $\mathbf{y}_k - \mathbf{H}_k \mathbf{x}_k^f$ ):

$$(3.3) \quad \mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}_k [ \mathbf{y}_k - \mathbf{H}_k \mathbf{x}_k^f ],$$

using the Kalman gain  $\mathbf{K}_k$  given by

$$(3.4) \quad \mathbf{K}_k = \mathbf{P}_k^f \mathbf{H}_k^T [ \mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T + \mathbf{R}_k ]^{-1},$$

where it can be demonstrated that  $\mathbf{K}_k$  corresponds to the minimization of the trace of the error covariance on  $\mathbf{x}_k^a$ , given at the minimum by

$$(3.5) \quad \mathbf{P}_k^a = [ \mathbf{I} - \mathbf{K}_k \mathbf{H}_k ] \mathbf{P}_k^f,$$

where  $\mathbf{x}_k^f$  is the forecasted model state vector at step  $k$ ,  $\mathbf{x}_k^a$  is the model state vector resulting from the analysis at step  $k$ ,  $\mathbf{M}_{k-1,k}$  is the model operator between step  $k-1$  and  $k$ ,  $\mathbf{P}_k^f$  is the background error covariance matrix after the forecast step  $k$ ,  $\mathbf{P}_k^a$  is the background error covariance matrix after the analysis step  $k$ ,  $\mathbf{Q}_k$  is the model error covariance matrix at step  $k$ ,  $\mathbf{K}_k$  is the Kalman gain matrix at step  $k$ ,  $\mathbf{H}_k$  is the observation operator at step  $k$ ,  $\mathbf{R}_k$  is the observation error covariance matrix at step  $k$ , and  $\mathbf{I}$  is the identity matrix.

At the end of each forecast step, the covariance matrix for the forecast error is calculated, and at the end of each analysis step, a new covariance matrix for the analysis error is specified.

**3.3. Kalman Filters for Geophysical Applications.** The canonical form of the Kalman filter is difficult to implement with realistic geophysical fluid models because of (i) nonlinearities in the models and observation operators; (ii) poorly known error statistics; and (iii) a prohibitive computational cost. Points (i) and (ii) affect the behavior or the quality of the data assimilation system. Point (iii) is numerically limiting, so that making the Kalman filter applicable to realistic geophysical problems requires simplifications and approximations. Such attempts have led to the development of suboptimal Kalman filter schemes [25], [17].

The problem of nonlinearities is partially solved by extending the Kalman filter algorithm to nonlinear systems, to what is called the extended Kalman filter (EKF). In the EKF,  $\mathbf{M}$  and  $\mathbf{H}$  are no longer linear applications, and have to be replaced by  $\mathbf{M}'$  and  $\mathbf{H}'$ , the tangent linear models, in the equations related to the model's error, i.e., equations (3.2)–(3.5). In most applications, however, the linearized models  $\mathbf{M}'$  and  $\mathbf{H}'$  are not used and calculations are performed with a finite difference approximation: If  $\mathbf{x}$  and  $\delta\mathbf{x}$  denote the state and a state perturbation or error,  $\mathbf{M}' \delta\mathbf{x} \approx \mathbf{M}(\mathbf{x} + \delta\mathbf{x}) - \mathbf{M}(\mathbf{x})$ .

The prescription of error covariance matrices  $\mathbf{R}$ ,  $\mathbf{Q}$ , and  $\mathbf{P}_0^f$  ( $\mathbf{P}^f$  at the first integration step) is required by the Kalman filter. However, little is known about these matrices for they cannot be inferred from their theoretical definition. The observation error statistics are generally deduced from consideration of instrumental errors, representativeness errors (due to the mismatch between the scales of spatial variability in the observations and the model grid), and errors in the observation operator. The model error statistics are often represented by simplified parameterizations, possibly adjusted online [21]. The background error statistics are often, though not always, determined from an ensemble of model simulations, supposedly representative of the

model’s uncertainty. Section 3.5 illustrates this by presenting such an approach to initialize the SEEK filter.

The prohibitive computational cost essentially comes from the propagation of the error covariance matrix  $\mathbf{P}$  (see (3.2)). Equation (3.2) implies a number of model integrations equal to the size of the state vector, i.e., larger than  $10^6$  in many oceanic applications. This is unfeasible with present-day ocean models and computation capabilities. Moreover, given the uncertainties in the prescribed error statistics, it appears inappropriate to spend a huge amount of computational time to propagate probably erroneous information. Several directions have been followed in the past to simplify the forecast step, often referred to as “suboptimal Kalman filter” schemes. They include the following ideas: considering error statistics constant in time (this corresponds to the optimal interpolation approach) [34], [18]; propagating only error variances and covariances for neighboring grid points [37], [42]; using a simplified model [19], [20] or a reduced-order model [17], [23] to propagate the error statistics; reducing the order of the state space [10], [24]; and reducing the order of the error space [22], [17], [44], [39], [33].

In the next section, we focus on how these three difficulties are tackled with the SEEK filter.

**3.4. The SEEK Filter: A Reduced-Order Kalman Filter.** The SEEK filter is a Kalman filter in which the dimension of the state error space is reduced. It was founded by Pham, Verron, and Roubaud [39], based on earlier ideas of Cohn and Todling [16], [17], and Verlaan and Heemink [43]. The integration of the matrix  $\mathbf{P}$  is made possible by the order reduction. This matrix is real and symmetric (thus Hermitian), and is therefore diagonalizable, with real eigenvalues and orthogonal eigenvectors. It can be written as

$$(3.6) \quad \mathbf{P} = \mathbf{N} \mathbf{B} \mathbf{N}^T,$$

where  $\mathbf{B}$  is a diagonal matrix of order  $n$  ( $n$  being the dimension of the dynamical system) containing the eigenvalues of  $\mathbf{P}$  and  $\mathbf{N}$  is a matrix containing its eigenvectors. The reduction of order consists of using only a small number  $r$  of eigenvectors for expressing  $\mathbf{P}$ , i.e., using a matrix  $\mathbf{N}$  of order  $n \times r$  rather than  $n \times n$ . How this subset of eigenvectors is determined in practice, i.e., what size it should have (what value does  $r$  take) and which of the eigenvectors should be included, is the subject of section 3.5. Traditionally, by introducing  $\mathbf{S} = \mathbf{N} \mathbf{B}^{1/2}$  in (3.6), the matrix  $\mathbf{P}$  is usually written as

$$(3.7) \quad \mathbf{P} = \mathbf{S} \mathbf{S}^T,$$

where  $\mathbf{S}$  is a matrix of order  $n \times r$ .

The equations of the EKF must now be rewritten using the matrix  $\mathbf{S}$  instead of  $\mathbf{P}$ . In order to reach an algorithm for the forecast/analysis cycle, we need to obtain equations for  $\mathbf{x}^f$ ,  $\mathbf{x}^a$ , the gain  $\mathbf{K}$  on which it depends, and the matrices  $\mathbf{S}^f$  and  $\mathbf{S}^a$ .

The gain  $\mathbf{K}$ , defined by (3.4), can be expressed in the following manner by taking (3.7) into account:

$$(3.8) \quad \mathbf{K} = \mathbf{S}^f (\mathbf{H} \mathbf{S}^f)^T [ (\mathbf{H} \mathbf{S}^f) (\mathbf{H} \mathbf{S}^f)^T + \mathbf{R} ]^{-1}.$$

The equality of the matrices,

$$\begin{aligned} & [ \mathbf{X}_1 + \mathbf{X}_{12} \mathbf{X}_2^{-1} \mathbf{X}_{21} ]^{-1} \\ &= \mathbf{X}_1^{-1} - \mathbf{X}_1^{-1} \mathbf{X}_{12} [ \mathbf{X}_2 + \mathbf{X}_{21} \mathbf{X}_1^{-1} \mathbf{X}_{12} ]^{-1} \mathbf{X}_{21} \mathbf{X}_1^{-1} \end{aligned}$$

allows us to develop (3.8) into the following expression of the Kalman gain  $\mathbf{K}$ , using  $\mathbf{X}_1 = \mathbf{R}$ ,  $\mathbf{X}_{12} = \mathbf{HS}$ ,  $\mathbf{X}_2^{-1} = \mathbf{I}$ , and  $\mathbf{X}_{21} = (\mathbf{HS})^T$ :

$$(3.9) \quad \mathbf{K} = \mathbf{S}^f [ \mathbf{I} + (\mathbf{HS}^f)^T \mathbf{R}^{-1} (\mathbf{HS}^f) ]^{-1} (\mathbf{HS}^f)^T \mathbf{R}^{-1}.$$

We also need to represent the analysis error covariance matrix  $\mathbf{S}^a$ . By introducing (3.9) (and (3.7)) into (3.5), we can develop  $\mathbf{P}^a$  as

$$\begin{aligned} \mathbf{P}^a &= [ \mathbf{I} - \mathbf{S}^f [ \mathbf{I} + (\mathbf{HS}^f)^T \mathbf{R}^{-1} (\mathbf{HS}^f) ]^{-1} (\mathbf{HS}^f)^T \mathbf{R}^{-1} \mathbf{H} ] \mathbf{S}^f \mathbf{S}^{fT} \\ &= \mathbf{S}^f [ \mathbf{I} + (\mathbf{HS}^f)^T \mathbf{R}^{-1} (\mathbf{HS}^f) ]^{-1} \mathbf{S}^{fT}. \end{aligned}$$

Since  $\mathbf{P}^a = \mathbf{S}^a \mathbf{S}^{aT}$ , we obtain

$$(3.10) \quad \mathbf{S}_k^a = \mathbf{S}_k^f [ \mathbf{I} + (\mathbf{H}_k \mathbf{S}_k^f)^T \mathbf{R}_k^{-1} (\mathbf{H}_k \mathbf{S}_k^f) ]^{-1/2}.$$

Then we perform the dynamical propagation of  $\mathbf{S}_k^a$ , i.e., calculate the vectors  $\{\mathbf{S}_k^f\}_l$  for each column  $l$  of  $\mathbf{S}^f$ , with the finite difference approximation

$$(3.11) \quad \{\mathbf{S}_{k+1}^f\}_l = \mathbf{M} [ \mathbf{x}_k^a + \{\mathbf{S}_k^a\}_l ] - \mathbf{M}[\mathbf{x}_k^a], \quad l = 1, \dots, r,$$

and add the model error covariances (see (3.2))

$$(3.12) \quad \mathbf{P}_{k+1}^f = \mathbf{S}_k^f \mathbf{S}_k^{fT} + \mathbf{Q}_k.$$

As mentioned in section 3.3,  $\mathbf{Q}_k$  is often represented by simple parameterizations. A natural constraint here is to conserve the rank of the  $\mathbf{P}^f$  from step  $k$  to  $k+1$ , then to define  $\mathbf{Q}_k$  in the same subspace as  $\mathbf{S}_k^f \mathbf{S}_k^{fT}$ . A common recipe is to consider  $\mathbf{Q}_k$  to be proportional to  $\mathbf{S}_k^f \mathbf{S}_k^{fT}$  [7], [9]. We have now obtained all of the equations forming the general SEEK filter algorithm.

Initialization (instant  $t_0$ ):

$$\begin{aligned} \mathbf{x}_{k=0} &= \mathbf{x}_0, \\ \mathbf{P}_{k=0} &= \mathbf{S}^0 \mathbf{S}^{0T} \quad (\text{cf. (3.7)}). \end{aligned}$$

Analysis step:

$$\begin{aligned} \mathbf{K}_k &= \mathbf{S}_k^f [ \mathbf{I} + (\mathbf{H}_k \mathbf{S}_k^f)^T \mathbf{R}_k^{-1} (\mathbf{H}_k \mathbf{S}_k^f) ]^{-1} (\mathbf{H}_k \mathbf{S}_k^f)^T \mathbf{R}_k^{-1} \quad (\text{cf. (3.9)}), \\ \mathbf{x}_k^a &= \mathbf{x}_k^f + \mathbf{K}_k [ \mathbf{y}_k - \mathbf{H}_k \mathbf{x}_k^f ] \quad (\text{cf. (3.3)}), \\ \mathbf{S}_k^a &= \mathbf{S}_k^f [ \mathbf{I} + (\mathbf{H}_k \mathbf{S}_k^f)^T \mathbf{R}_k^{-1} (\mathbf{H}_k \mathbf{S}_k^f) ]^{-1/2} \quad (\text{cf. (3.10)}). \end{aligned}$$

Forecast step:

$$\begin{aligned} \mathbf{x}_{k+1}^f &= \mathbf{M}_{k,k+1} \mathbf{x}_k^a \quad (\text{cf. (3.1)}), \\ \{\mathbf{S}_{k+1}^f\}_r &= \mathbf{M} [ \mathbf{x}_k^a + \{\mathbf{S}_k^a\}_l ] - \mathbf{M}[\mathbf{x}_k^a], \quad l = 1, \dots, r \quad (\text{cf. (3.11)}), \\ \mathbf{P}_{k+1}^f &= \mathbf{S}_k^f \mathbf{S}_k^{fT} + \mathbf{Q}_k \quad (\text{cf. (3.12)}). \end{aligned}$$

In oceanographic applications of reduced-order Kalman filters, the major amount of computing time is spent in the model evolution. Hence, the computational complexity of the SEEK algorithm is mainly due to the forecast step: it is equal to  $(r+1)N$ ,



where  $N$  is the number of operations required to integrate the model equations between times  $t_k$  and  $t_{k+1}$ . This explains why the reduction of rank (i.e., using a rank  $r$  much smaller than  $n$ ) is the decisive factor that enables a Kalman filter to be used for data assimilation in physical oceanography.

However, in some applications [38], [41] it appears to be not essential to dynamically propagate the error matrix, either because the effect is weak or because the model errors are large and make it useless. In these cases, the error matrix is kept fixed and equation (3.2) of the EKF is replaced by

$$\mathbf{P}_k^f = \mathbf{P}_{k-1}^a + \mathbf{Q}_{k-1}.$$

In such cases, the computational complexity of the forecast step is reduced to  $N$  (one single model integration) so that the computational cost of the analysis step may become nonnegligible. The leading terms of the computational complexity of the SEEK analysis described in this paper (see (3.3), (3.9), and (3.10)) are  $m(r^2 + r + 2) + r^3 + n(r^2 + r)$ , where  $n$  is the size of the state vector and  $m$  is the size of the observation vector. It is worth noting that for a large observation vector, the computation of the Kalman gain using the SEEK formula (3.9) is much cheaper than the original formula (3.8), for which the computational complexity is proportional to  $m^3$ . This is a key originality of the SEEK algorithm with respect to other reduced-rank Kalman filters, which most often use formulas like (3.8) to evaluate the Kalman gain (see [35] for more details).

**3.5. Reduction of Order in Practice through the Use of Empirical Orthogonal Functions (EOFs).** As explained in the previous section, (3.6) results in a reduction of order if we use a number  $r$  of eigenvectors that is small compared with  $n$ . Practically, we have to identify an appropriate number  $r$  and choose which  $r$  eigenvectors should be used.

In order to initialize the  $\mathbf{P}$  matrix, the first step is usually to find a suitable random or deterministic process to simulate the background error statistics. According to the kind of error present in the system, this process can be, for instance, a simulation of the time variability of the system or the generation of an ensemble of model simulations from random perturbations in the model parameters, the model forcing function, or the model dynamics. If the error in the system is due to an incorrect time phasing of a turbulent flow, the system's variability is certainly appropriate to describe the covariance of the errors. On the contrary, if the error is due to inaccurate model parameters or forcing functions, it is certainly more suitable to parameterize the error covariance using the covariance of an ensemble of model simulations obtained from an ensemble of perturbations in the parameters or in the forcing function, since they can be assumed to be the real source of error. Usually, that kind of proxy for the error statistics (system's variability, ensemble of perturbed simulations) lead to a low-rank parameterization of the background error covariance matrix.

However, in the SEEK filter parameterization, a second step is needed, consisting of keeping only the first  $r$  EOFs from the ensemble of model state vectors that have been produced by one of the techniques described above. This truncation leads to a covariance matrix taking the form (3.6), where the columns of  $\mathbf{N}$  are the normalized EOFs and where  $\mathbf{B}$  is diagonal, with the corresponding eigenvalues [39]. This is justified by the following result [40]: "If  $x$  is a random vector of zero mean and covariance  $\mathbf{P}$ , then among all its projections onto a linear subspace of dimension  $r$ , the one for which the error vector has smallest expected squared norm is the projection onto the linear space spanned by the first  $r$  eigenvectors of  $\mathbf{P}$ ."

The accuracy of the above approximation for the background error covariance matrix mainly depends on three factors: (i) the suitability of the process that is used to simulate the error statistics, (ii) the size of the sample of independent states that have been produced using that process, and (iii) the number  $r$  of EOFs that are retained in the parameterization. It is difficult to provide general statements on the first factor, since this closely depends on the knowledge of the sources of errors in the system. Conversely, the second factor leads to the well-known confidence interval on the error statistics, so that it is easy to determine the size of the ensemble to obtain a desired accuracy on the background error covariance structure. The effect of the third factor is also easy to evaluate because, if  $\lambda_j$  are the eigenvalues of  $\mathbf{P}$  sorted in decreasing order, the relative error in the representation of a random vector  $x$  of covariance  $\mathbf{P}$  in the reduced space spanned by the first  $r$  eigenvectors of  $\mathbf{P}$  is given by the ratio  $\sum_{j>r} \lambda_j / \text{tr} \mathbf{P}$  [39].

However, in many practical applications the rank  $r$  of the matrix (factor 3) or even the size of the initial ensemble (factor 2) that would be needed to represent the error correlation structure with a sufficient accuracy is often too large to be numerically tractable. In order to overcome this difficulty, several authors have proposed a subsequent approximation to the reduced-rank Kalman filters (the local approximation [30], [41], [7]), consisting of an artificial increase of the rank of the  $\mathbf{P}$  matrix by nullifying (in a consistent manner) the long-range correlation coefficients. Such approximations complexify in several ways the structure of the analysis algorithm, which is beyond the scope of this paper to describe.

EOF decomposition techniques can also be used to reduce the dimension of the work space of the model itself, rather than the error state space. Such an approach has been recently introduced for oceanographic problems [11], [12] and belongs to the family of reduced-order models mentioned in section 3.3.

**4. Applying the SEEK Filter: Examples of Assimilation Experiments.** The reduced-order Kalman filter described in the previous section and implemented in the SEEK filter has been applied to various simulation tasks; see, for example, [8], [7]. In order to illustrate the use of the SEEK filter, we first describe how data are obtained, since without any data available, there would be no possible application of data assimilation. We then look at an assimilation experiment undertaken on a typical North Atlantic simulation and the enhancements obtained through data assimilation. Finally, we will consider one weakness of sequential assimilation schemes, the creation of dynamical shocks in the model, together with the track currently being explored by our team in order to address this weakness.

**4.1. Obtaining Data: Ocean Observation Systems.** Operational oceanography requires good quality systematic data for assimilation into numerical models to provide analyses of a range of ocean phenomena.

Satellite remote sensing has revolutionized the observation of the oceans: despite the recent deployment of a large number of floats (the ARGO floats campaign), in-situ data are limited in coverage and predominantly confined to shipping lanes. Satellite measurements offer the potential of surveying a range of key parameters over the entire ocean surface in just a few days. The most important of these parameters is the sea surface height (SSH), since altimetric data (i.e., measurements of SSH) form the common backbone of all ocean assimilation systems. It is measured through a satellite radar technique where short pulses are reflected at the ocean's surface. The analysis of the received echo allows, with the known satellite position, a precise, fast, and global monitoring of the SSH and its variability. SSH variability is due to and

so can be related to ocean dynamics. Other useful data sets offered by remote ocean sensing include sea surface temperature (SST). IR sensors on environmental satellites can be used to measure the temperature across large expanses of the ocean surface. The assimilation of SST allows a better monitoring of changes in ocean temperatures or circulation, the detection of newly formed sea ice, or the location of specific thermal features related to ocean dynamics such as oceanic fronts, rings, or eddies.

Critical in-situ observation systems are also used: the international ARGO program, which consists of a global network of profiling floats delivering vertical temperature and salinity profiles, and also the global tropical mooring network, the ship-of-opportunity programs, and the ocean sites time series station network.

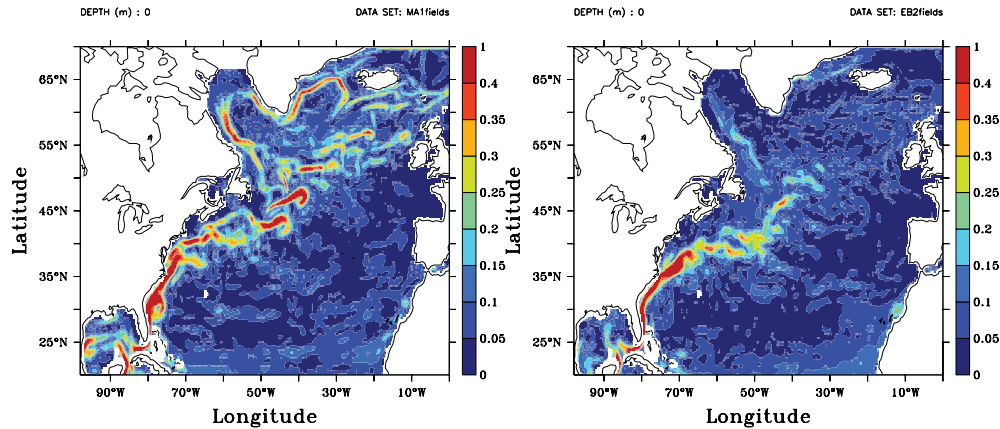
Moreover, a number of new observation systems are in the process of being deployed from space or at sea, some already able to deliver data while others are still at the design stage. Forthcoming ocean prediction systems will have to be able to assimilate new data sets such as large-scale sea surface salinity or ocean color measurements, ice concentration, or coastal in-situ data.

## 4.2. An Assimilation Experiment with the SEEK Filter.

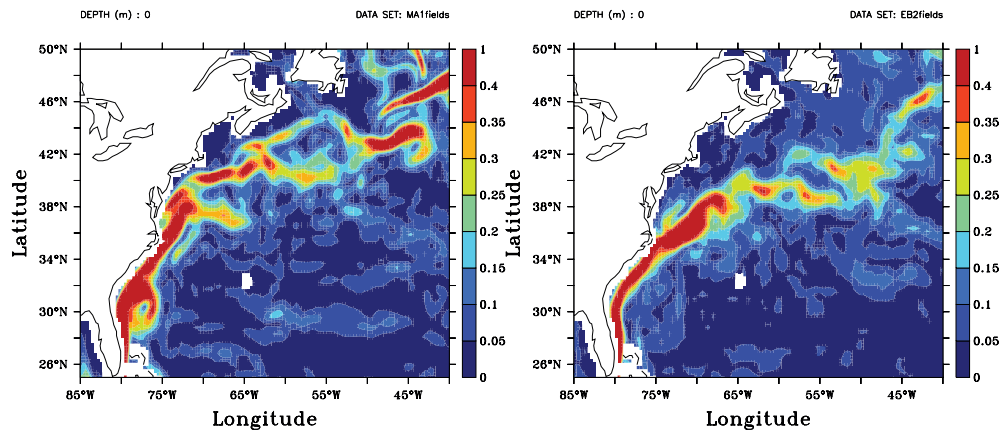
**4.2.1. Example of the Correction of a Surface Field.** Figures 4.1 and 4.2 show snapshots of the surface current velocity, from a simulation without assimilation and from the same simulation with assimilation. One major structure in the North Atlantic region shown in the figures is the Gulf Stream: a strong jet of hot water originating in the Gulf of Mexico and shooting up the east coast of the United States until it leaves the coast in the area of Cape Hatteras. The left-hand plots of both figures show snapshots of a simulation with no data assimilation. An important weakness of these two plots is the location where the Gulf Stream leaves the coast. Although it should be at Cape Hatteras, it takes place further north; we say that the stream is “over-shooting.” The right-hand plots were then taken from an equivalent simulation with data assimilation. The area where the Gulf Stream leaves the coast is appropriately located at Cape Hatteras. As a consequence of this correction, the Gulf Stream and its extension, the North Atlantic Current, follow a path that is much closer to reality. The structure of the stream is also affected and corrected by the data assimilation. It has a narrower and well-defined body, an important characteristic of such a jet, and the area of high turbulence is correctly located. The use of data assimilation allows for a more accurate simulation of the ocean’s surface currents.

**4.2.2. Example of the Correction of a Vertical Section.** In most cases, the data assimilated are acquired through surface measurements, such as SST or SSH. Indeed, the main sources of data with satisfactory coverage in time and space are satellite measurements, and satellites cannot yet produce measurements of the ocean’s interior. Obviously, the world’s ocean circulation cannot be narrowed down to the surface currents. Even if we were only interested in the prediction of surface currents, these are heavily dependent on what happens in the depth of the ocean.

The SEEK filter uses three-dimensional error statistics, allowing the inference of subsurface characteristics from surface measurements. The assimilation of surface data can therefore improve the simulation of circulation at any depth, from the surface to the bottom of the ocean. Figure 4.3 shows a section of the ocean’s temperature at a latitude of 46.5 North, between the longitudes of 15 West and 3 East, on August 3, 1993. This geographical area is located in the Bay of Biscay, where, in the summer, there is a warm pool in the southeastern part of the bay, separated from the warm waters of the Portugal polarward current by a colder current branching from



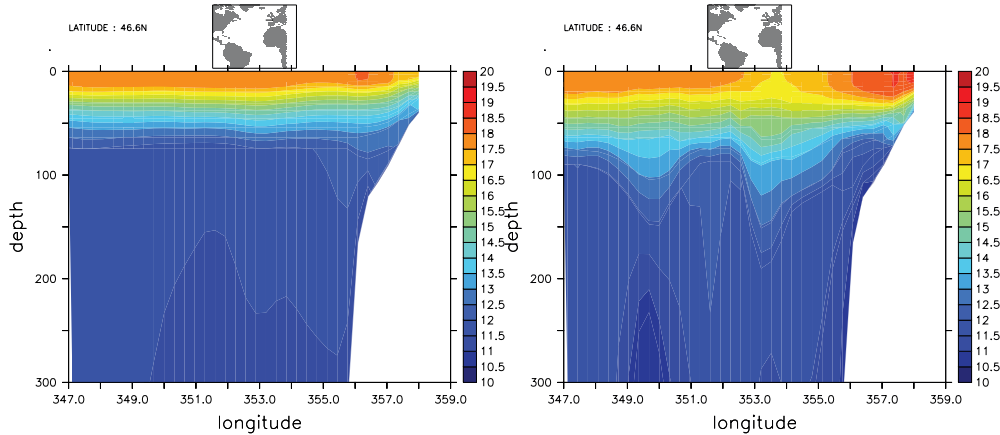
**Fig. 4.1** Surface current velocity in  $m.s^{-1}$  for the North Atlantic (18th August 1993) from HYCOM  $1/3^\circ$  simulation. Left: without data assimilation. Right: with data assimilation.



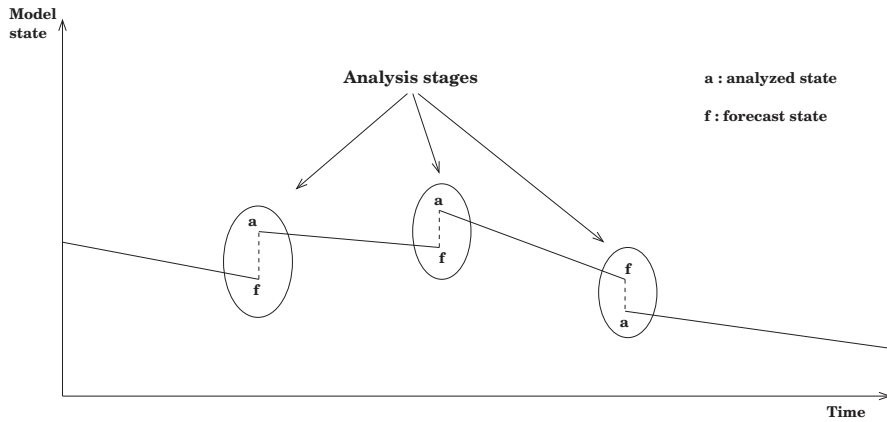
**Fig. 4.2** A zoom from Figure 4.1: the surface current velocity  $m.s^{-1}$  for the Gulf Stream region (18th August 1993) from HYCOM  $1/3^\circ$  simulation. Left: without data assimilation. Right: with data assimilation.

the North Atlantic Current. On the left-hand plot in Figure 4.3, resulting from a simulation without any assimilation, the temperature at the surface and subsurface does not reflect these structures and currents: the temperature is homogeneous along the section for a given depth. On the right-hand plot with assimilation, however, it is possible to identify the warm pool structure (colored red), on the east by the coast, separated from the other warm structure by a colder water mass (in yellow). The correction achieved by the data assimilation not only resulted in the formation of this cold water mass at the surface, but also strongly influenced the whole profile of temperature. Indeed this cold current is visible down to a depth of 150 meters and seems to be doubled by a second branch of colder waters between the depths of 50 to 150 meters that is not visible on the surface.

This example shows that despite the typical use of surface measurements as the main source for data assimilation, the correction achieved by this mathematical technique reaches down to water layers deep below the ocean's surface. This vertical



**Fig. 4.3** Section of temperature at 46.5°N from a simulation without data assimilation on the left, and with assimilation of SST, SSH, and salinity on the right.

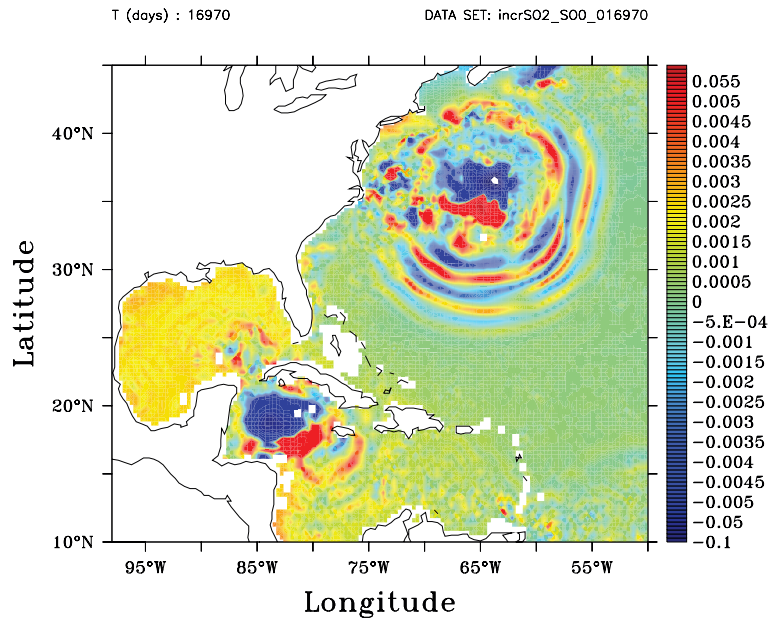


**Fig. 4.4** The temporal scheme of the classic Kalman filter.

extrapolation of surface correction enables data assimilation to be used as a tool for the enhancement of simulations for the entire ocean circulation.

**4.2.3. Intermittent and Incremental Assimilation Schemes.** The classic assimilation temporal scheme, used in the SEEK filter and all sequential assimilation methods, follows the time discretization described in section 3.2. The graphical representation of this scheme in Figure 4.4 demonstrates its intermittent nature, as shown by the discontinuous line. The trajectory is not continuous for every assimilated variable, since the analysis changes the value of these variables from the forecast value to the analyzed value at the same instant in time. Unfortunately this discontinuity is not compatible with physical processes, where changes can happen sharply but always continuously.

Let us take, for example, the variable SSH, which is typically assimilated with the SEEK filter. Its sudden and discontinuous variation creates waves on the ocean surface, created dynamically by the model through a normal adjustment process. These



**Fig. 4.5** *SSH increment between simulations with and without assimilation, 5 days after the latest assimilation step.*

waves can be seen in Figure 4.5, which represents the SSH increment between simulations with and without assimilation, 5 simulation days after the latest assimilation step. Although these waves have a small amplitude, their existence is spurious, and avoiding them would be important as part of a technique allowing us to avoid the discontinuities of the model's trajectory caused by the intermittent scheme. A potential solution consists of replacing the intermittent scheme by an incremental scheme, called incremental analysis updating (IAU) [5], [36]. The integration of this scheme into the SEEK filter is currently being evaluated. Following this method, every forecast/analysis cycle is integrated twice: the first instance allows the innovation to be computed by the analysis, and this innovation is then incremented into the forecast step all along the timespan of the second instance of this forecast step. This technique smooths the assimilated trajectory and should help to avoid the potentially dramatic consequences of discontinuities.

**5. Conclusion.** The amount of data currently available to oceanographers has drastically increased with the use of satellites. In addition, computing and storage possibilities are constantly enhanced by the fast progress of modern computers. These conditions motivate the use of increasingly high spatial resolution and increasingly complex models, integrating phenomena such as tide or the melting/freezing of sea ice. As a consequence, the mathematical techniques used by the community must cope with this increased complexity.

This paper has explained the particular case of a data assimilation technique called the SEEK filter using a Kalman filter. A review of the main characteristics of this type of filter showed that the computing and storage cost of such a method, if it is not modified, is prohibitive for the usual oceanographic configurations. The solution implemented in the SEEK filter consists of the reduction of the rank of the background error covariance matrix  $\mathbf{P}$ . This reduction of order allows the SEEK

filter to be applicable for typical physical and operational oceanography. Illustrating the practical use of the SEEK filter, we looked at some examples of assimilated simulations, highlighted the discontinuity drawback of this method, and described a potential solution to this drawback that is currently being explored.

Further theoretical work is still in progress concerning the assimilation scheme itself, such as the study of the IAU scheme mentioned in section 4.2.3, involving important mathematical skills. In operational oceanography, a highly active area of work is concerned with regional or even coastal predictions, as a large majority of human marine activities happen within only a few kilometers of the coast. Regional and coastal high-resolution configurations open up completely new sets of oceanographical challenges (such as the accurate account of tides), technical challenges (such as satellite measurements in shallow waters), and also mathematical challenges. The first mathematical challenge arises from the fact that regional or coastal domains include not only land borders, but also borders in the ocean, requiring the specification of open boundary conditions. These boundary conditions are provided by a coarser-resolution basin-scale model into which the regional model is said to be *nested*. The specification of these boundary conditions utilizes specific mathematical techniques (e.g., Flather), motivating the collaboration between oceanographers and mathematicians. Also, the control of these configurations covering a smaller area, close to a coast of interest, but with a higher resolution, is a complex problem that certainly cannot only be solved through the specification of boundary conditions. We believe that control can be enhanced through data assimilation techniques, and we are therefore working on applying the SEEK filter described in this paper to this particular area of nested configurations. and J. Verron

**Acknowledgments.** We would like to thank Eric Blayo from the Laboratory of Modelling and Calculus in Grenoble (LMC, IMAG) for his useful feedback on this paper, and Julie Dugdale for her thorough English proofreading.

#### REFERENCES

- [1] P. BAHUREL, *MERCATOR OCEAN global to regional ocean monitoring and forecasting*, in GODAE, an Integrated View of Oceanography: Ocean Weather Forecasting in the 21st Century, J. Verron and E. Chassignet, eds., Kluwer Academic, Dordrecht, The Netherlands, 2006, pp. 381–395.
- [2] J. M. BECKERS, Y. DESAUBIES, AND M. RIXEN, *Marine Environmental Monitoring and Predictions*, J. Marine Systems, 65 (2007). Special Issue dedicated to the 36th International Liège Colloquium on Ocean Dynamics.
- [3] R. BLECK, *An oceanic general circulation model framed in hybrid isopycnic-Cartesian coordinates*, Ocean Modelling, 37 (2002), pp. 55–88.
- [4] R. BLECK AND E. P. CHASSIGNET, *Simulating the oceanic circulation with isopycnic-coordinate models*, in The Oceans: Physical-Chemical Dynamics and Human Impact, S. K. Majumdar, E. W. Miller, G. S. Forbes, R. F. Schmalz, and A. A. Panah, eds., The Pennsylvania Academy of Science, 1994, pp. 17–39.
- [5] S. C. BLOOM, L. L. TAKACS, A. M. DASILVA, AND D. LEVINA, *Data assimilation using incremental analysis updates*, Monthly Weather Rev., 124 (1996), pp. 1256–1271.
- [6] A. F. BLUMBERG AND G. L. MELLOR, *A description of a three-dimensional coastal ocean circulation model*, in Three-Dimensional Coastal Ocean Models, N. S. Heaps, ed., American Geophysical Union, Washington, D.C., 1987, pp. 1–16.
- [7] J. M. BRANKART, C. E. TESTUT, P. BRASSEUR, AND J. VERRON, *Implementation of a multivariate data assimilation scheme for isopycnic ocean models: Application to a 1993–1996 hindcast of the North Atlantic Ocean circulation*, J. Geophys. Res., 108 (C3) (2003), pp. 1–20.
- [8] P. BRASSEUR, J. BALLABRERA-POY, AND J. VERRON, *Assimilation of altimetric data in the mid-latitude oceans using the Singular Evolutive Extended Kalman filter with an eddy-resolving, primitive equation model*, J. Marine Syst., 22 (1999), pp. 269–294.

- [9] P. BRASSEUR, P. BAHUREL, L. BERTINO, F. BIROL, J.-M. BRANKART, N. FERRY, S. LOSA, E. RÉMY, J. SCHRÖTER, S. SKACHKO, C.-E. TESTUT, B. TRANCHANT, P.-J. VAN LEUWEN, AND J. VERRON, *Data assimilation for marine monitoring and prediction: The MERCATOR operational assimilation systems and the MERSEA developments*, Quart. J. Roy. Meteor. Soc., 131 (2005), pp. 3561–3582.
- [10] M. A. CANE, A. KAPLAN, R. N. MILLER, B. TANG, E. C. HACKERT, AND A. J. BUSALACCHI, *Mapping tropical Pacific sea level: Data assimilation via a reduced state space Kalman filter*, J. Geophys. Res., 101 (C10) (1996), pp. 22599–22617.
- [11] Y. CAO, J. ZHU, Z. LUO, AND I. M. NAVON, *Reduced order modeling of the upper tropical Pacific Ocean model using proper orthogonal decomposition*, Comput. Math. Appl., 52 (2006), pp. 1373–1386.
- [12] Y. CAO, J. ZHU, I. M. NAVON, AND Z. LUO, *A reduced order approach to four-dimensional variational data assimilation using proper orthogonal decomposition*, Internat. J. Numer. Methods Fluids, 53 (2007), pp. 1571–1583.
- [13] E. P. CHASSIGNET AND J. VERRON, *Ocean Modeling and Parameterization*, NATO Sci. Ser. C Math. Phys. Sci. 516, Kluwer Academic, Dordrecht, The Netherlands, 1998.
- [14] E. P. CHASSIGNET, L. T. SMITH, G. R. HALLIWELL, AND R. BLECK, *North Atlantic simulations with the Hybrid Coordinate Ocean Model (HYCOM): Impact of the vertical coordinate choice, reference pressure, and thermobaricity*, J. Phys. Oceanogr., 33 (2003), pp. 2504–2526.
- [15] S. E. COHN, *An introduction to estimation theory*, J. Meteor. Soc. Japan, 75 (1B) (1997), pp. 257–288.
- [16] S. E. COHN AND R. TODLING, *Approximate Kalman filters for unstable dynamics*, in Proceedings of the Second International Symposium on Assimilation of Observations in Meteorology and Oceanography, WMO, 1995, pp. 241–246.
- [17] S. E. COHN AND R. TODLING, *Approximate Kalman filters for stable and unstable dynamics*, J. Meteor. Soc. Japan, 74 (1996), pp. 63–75.
- [18] R. DALEY, *Atmospheric Data Analysis*, Cambridge University Press, Cambridge, UK, 1991.
- [19] D. P. DEE, *An efficient algorithm for estimating noise covariances in distributed systems*, IEEE Trans. Automat. Control, 30 (1985), pp. 1057–1065.
- [20] D. P. DEE, *Simplification of the Kalman filter for meteorological data assimilation*, Quart. J. Roy. Meteor. Soc., 117 (1991), pp. 365–384.
- [21] D. P. DEE, *On-line estimation error covariance parameters for atmospheric data assimilation*, Monthly Weather Rev., 123 (1995), pp. 1128–1145.
- [22] G. EVENSEN, *Sequential data assimilation with a non linear quasigeostrophic model using Monte Carlo methods to forecast error statistics*, J. Geophys. Res., 99 (C5) (1994), pp. 10143–10162.
- [23] B. F. FARRELL AND P. J. IOANNOU, *State estimation using a reduced-order Kalman filter*, J. Atmos. Sci., 58 (2001), pp. 3666–3680.
- [24] I. FUKUMORI, *Assimilation of TOPEX sea level measurements with a reduced-gravity, shallow water model of the tropical Pacific Ocean*, J. Geophys. Res., 100 (C12) (1995), pp. 25027–25039.
- [25] M. GHIL AND P. MALANOTTE-RIZZOLI, *Data assimilation in meteorology and oceanography*, Adv. Geophys., 33 (1991), pp. 141–266.
- [26] M. GHIL, K. IDE, A. BENNETT, P. COURTIER, M. KIMOTO, M. NAGATA, M. SAIKI, AND N. SATO, EDs., *Data Assimilation in Meteorology and Oceanography: Theory and Practice*, J. Meteor. Soc. Japan, special issue, 1997.
- [27] M. S. GREWAL AND A. P. ANDREWS, *Kalman Filtering: Theory and Practice Using MATLAB*, Wiley-Interscience, New York, 2001.
- [28] S. M. GRIFFIES, C. BÖNING, F. O. BRYAN, E. P. CHASSIGNET, R. GERDES, H. HASUMI, A. HIRST, A.-M. TREGUIER, AND D. WEBB, *Developments in ocean climate modelling*, Ocean Modelling, 2 (2000), pp. 123–192.
- [29] R. HALLBERG, *HIM: The Hallberg Isopycnal Coordinate Primitive Equation Model*, NOAA Geophysical Fluid Dynamics Laboratory, Tech. Report, Princeton University, Princeton, NJ, 1997.
- [30] P. L. HOUTEKAMER AND H. L. MITCHELL, *Data assimilation using an ensemble Kalman filter technique*, Monthly Weather Rev., 126 (1998), pp. 796–811.
- [31] R. X. HUANG, *Real freshwater flux as a natural boundary condition for salinity balance and thermohaline circulation forced by evaporation and precipitation*, J. Phys. Oceanogr., 23 (1993), pp. 2428–2446.
- [32] A. H. JAZWINSKI, *Stochastic Processes and Filtering Theory*, Math. Sci. Engrg. 64, Academic Press, New York, 1970.



- [33] P. F. J. LERMUSIAUX AND A. R. ROBINSON, *Data assimilation via error subspace statistical estimation. Part I: Theory and schemes*, Monthly Weather Rev., 127 (1999), pp. 1385–1407.
- [34] A. LORENC, *A global three-dimensional multivariate statistical interpolation scheme*, Monthly Weather Rev., 109 (1981), pp. 701–721.
- [35] L. NERGER, W. HILLER, AND J. SCHRÖTER, *A comparison of subspace Kalman filters*, Tellus, 57A (2005), pp. 715–735.
- [36] Y. OURMIÈRES, J. M. BRANKART, L. BERLINE, J. VERRON, AND P. BRASSEUR, *Incremental analysis update implementation into a sequential ocean data assimilation system*, J. Atmospheric and Oceanic Technologies, 23 (2006), pp. 1729–1744.
- [37] D. F. PARRISH AND S. E. COHN, *A Kalman Filter for a Two-Dimensional Shallow-Water Model: Formulation and Preliminary Experiments*, Office Note 304, National Meteorological Center, Washington D.C., 1985.
- [38] T. PENDUFF, P. BRASSEUR, C.-E. TESTUT, B. BARNIER, AND J. VERRON, *A four-year eddy-permitting assimilation of sea-surface temperature and altimetric data in the South Atlantic Ocean*, J. Marine Res., 60 (2002), pp. 805–833.
- [39] D. PHAM, J. VERRON, AND M. ROUBAUD, *A Singular Evolutive Extended Kalman filter for data assimilation in oceanography*, J. Marine Syst., 16(3-4)(1998), pp. 323–340.
- [40] G. A. F. SEBER, *Multivariate Observations*, Wiley, New York, 1984.
- [41] C.-E. TESTUT, P. BRASSEUR, J.-M. BRANKART, AND J. VERRON, *Assimilation of sea-surface temperature and altimetric observations during 1992–1993 into an eddy permitting primitive equation model of the North Atlantic Ocean*, J. Marine Syst., 40-41 (2003), pp. 291-316.
- [42] R. TODLING AND S. E. COHN, *Suboptimal schemes for atmospheric data assimilation based on the Kalman filter*, Monthly Weather Rev., 122 (1994), pp. 2530–2557.
- [43] M. VERLAAN AND A. W. HEEMINK, *Reduced rank square root filters for large scale data assimilation problem*, in Proceedings of the Second International Symposium on Assimilation of Observations in Meteorology and Oceanography, WMO, 1995, pp. 247–252.
- [44] M. VERLAAN AND A. W. HEEMINK, *Tidal flow forecasting using reduced-rank square root filter*, Stochastic Hydro. Hydraul., 11 (1997), pp. 349–368.