Chapter 20. DATA ASSIMILATION

ALLAN R. ROBINSON, PIERRE F. J. LERMUSIAUX AND N. QUINCY SLOAN III

Harvard University

Contents

1. Introduction
2. Goals and applications
3. Overview of data assimilation methods
4. Coastal ocean data assimilation
5. Progress to date
6. Middle Atlantic Bight studies
7. Conclusions

Appendix: DA methods under common generic assumptions

Bibliography

1. Introduction

The coastal ocean is a complex system in which it can be difficult and costly to measure and predict quantities of interest for scientific and management purposes. The estimation of a quantity of interest via data assimilation involves the combination of observational data with the underlying dynamical principles governing the system under observation. The melding of data and dynamics is a powerful methodology that makes possible efficient, accurate and realistic estimations which might not otherwise be feasible. It is a methodology that can optimize the extraction of reliable information from observations. Data assimilation has recently entered oceanography from the related fields of meteorology and engineering. It is expected to provide rapid advances in important aspects of both basic ocean science and applied marine technology and operations. In this section we present the concepts of field estimation, give an overview of the general methodology and goals of data assimilation, and discuss implications for the coastal ocean in the context of marine science and technology in general.

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

1.1. Field Estimation

Ocean science, and marine technology and operations, require a knowledge of the distribution and evolution in space and time of the physical, biological and chemical characteristics of the sea. The functions of space and time, or *state variables*, which characterize the state of the sea under observation are classically designated as *fields*. The determination of the distribution or evolution of the state variables poses problems of *field estimation* in three or four dimensions. The fundamental problem of ocean science may be stated simply as: Given the state of the ocean at one time, what is the state of the ocean at a later time? It is the dynamics (i.e., the basic laws and principles of oceanic physics, biology and chemistry) that evolve the state variables forward in time. Thus, also from a practical viewpoint, predicting the present and future state of oceanic variables for applications is intimately linked to fundamental ocean science.

The physical state variables are usually the velocity components, the pressure, density, temperature and salinity. Examples of biological and chemical state variables are concentration distributions of nutrients, plankton, dissolved and particulate matter, and so on. Because of the complexity of the marine biogeochemical systems and ecosystems, the number of possible state variables is extremely large, and the limitation to a finite subset of *critical* state variables is an important contemporary research problem (Flierl and Davis, 1995; GLOBEC, 1995). For example, life-stage classes within individual species and size classes within particle types need to be considered. However, we shall not be concerned with these issues here and will focus primarily, but not entirely, on physical fields. A complexity with which we shall be concerned is associated with the vast range of phenomena, and the multitude of concurrent and interactive scales in space and time, which occur in the ocean. This complexity has two important consequences. First, state variable definitions *relevant to phenomena and scales of interest* need to be developed from the basic definitions. Second, *approximate dynamics* that govern the evolution of the scale restricted state variables, and their interaction with other scales, must be developed from the basic dynamical model equations. A familiar example is the derivation of a primitive equation model for a thin film of fluid on a rotating spherical earth from the basic Navier–Stokes equations of fluid dynamics (Batchelor, 1967; Le Blond and Mysak, 1978; Holton, 1992). By decomposing the fields into slow and fast time scales and averaging over the shorter scales, these equations can be adapted to govern synoptic/mesoscale-resolution state variables over a large-scale oceanic domain, with faster smaller-scale phenomena represented as parameterized fluctuation correlations (Reynolds stresses) (Reynolds, 1895; Goldstein, 1965; Tennekes and Lumley, 1972; Landahl and Mollo-Christensen, 1992). There is, of course, a great variety of other scale-restricted state variables and approximate dynamics of vital interest in ocean science. We shall refer to scale-restricted state variables and approximate dynamics simply as *state variables* and *dynamics*.

The use of dynamics is of fundamental importance for efficient and accurate field estimation in oceanography. A field estimate obtained by melding observational data with dynamics is referred to as a field estimation via *data assimilation*. It is perhaps obvious that dynamics is of paramount importance for a forecast, or the evolution of the state of the system forward in time. If the present state of the system used to initialize the forecast is estimated from observational data, we refer to the future esti-

mate as involving *assimilation via initialization*. It is less obvious that dynamics can be of paramount importance for efficient and accurate melded estimates of the present state of the system (*nowcasts*). Today and in the foreseeable future, data acquisition in the ocean is sufficiently difficult and costly so as to make field estimates by direct measurements, on a substantial and sustained basis, and for state variables, sampling rates, spatial domains, and measurement intervals of interest, essentially prohibitive. However, data acquisition for field estimate via data assimilation is feasible, but substantial resources must be applied to obtain adequate observations.

A schematic of the data assimilation process is shown in Fig. 20.1. The insertion of the reliable data into the dynamical model serves to adjust the data dynamically and therefore to improve the quality and accuracy of the estimate. Dynamics provides linkages among the state variables and thereby extends the influence or impact and value of the measurements. The model dynamically interpolates the data, which allows for efficient estimation by the reduction of space/time sampling requirements. Forward time extrapolation provides the forecast.

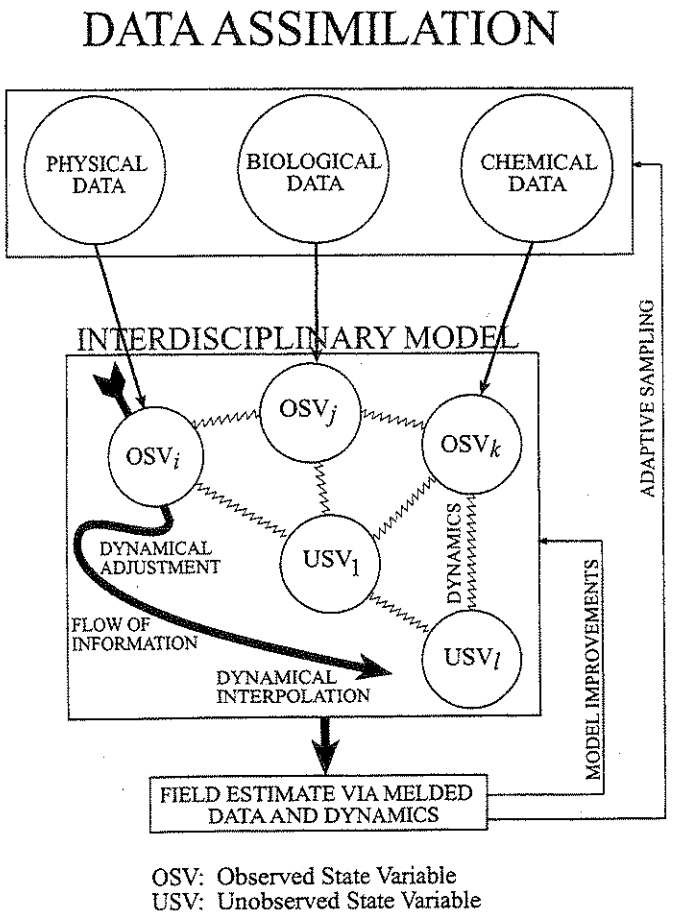


Fig. 20.1. Data assimilation process schematic.

1.2. Data Assimilation

Central to the concept of data assimilation is the concept of errors, error estimation and error modeling. A *data assimilation system* consists of three components: a set of observations, a dynamical model and a data assimilation scheme or *melding scheme*. The field of observations has errors arising from various sources, including instrumental noise, environmental noise, sampling and the interpretation of sensor measurements as state variables. All oceanic dynamical models are imperfect, with errors arising from various sources, including the approximate physics (or biology or chemistry), which govern the explicit evolution of the scale-restricted state variables; the approximate closures, which parameterizes the interaction of the state variables with smaller and faster scales, and the discretization of continuum physics into a computational model. Several melding schemes will be discussed, but an aspect common to all of them is that the quantitative basis of melding is the relative uncertainties of the dynamics and the observations. *A well-constructed melded estimate agrees with all observations within data error bounds and satisfies the dynamical model within model error bounds* (Ghil, 1989; Lermusiaux, 1997). Thus the melded estimate does not degrade the reliable information of the observational data but rather, enhances that information content.

A quantitative assessment of the accuracy of the melded field estimate is highly desirable but may be difficult to achieve because of the quantity and quality of the data required for verification of the data assimilation system. Such verification involves the quantitative assessment of the subcomponents, including the dynamical model, the observational network, the associated error models and the melding scheme. In our discussion of verification we introduce the concepts of validation and calibration. Validation is the establishment of the general adequacy of the system and its components to deal with the phenomena of interest. For example, avoid the use of a barotropic model for baroclinic phenomena, and avoid the use of an instrument whose threshold is higher than the accuracy of the required measurement. In reality, validation issues can be quite subtle. Calibration involves the tuning of system parameters to the phenomena and regional characteristics of interest. Throughout the verification process, sensitivity studies are extremely important.

At this point it is useful to classify types of estimates with respect to the time interval of the data input to the estimate for time t (Gelb, 1974). If only past and present data are utilized, the estimation is a *filtering* process. A nowcast via filtering provides the best possible initialization for a real-time forecast. A hindcast is an *a posteriori* forecast. However, for an estimate to be made based on a time series of data of duration $(0, T)$ after the entire time series is available, the estimate for any time $0 \leq t \leq T$ is best based on the entire data set. Thus future as well as past data are utilized, and the estimation is a *smoothing* process. In oceanography, an estimate for all t in the interval $(0, T)$ by either a filtering or a smoothing process is referred to as a *data-driven simulation*.

Data assimilation in oceanography is only a few years old (Bennett and McIntosh, 1982; McIntosh and Bennett, 1984; Robinson and Leslie, 1985; Mooers et al., 1986). Data assimilation methods being used or adapted today for ocean science have their roots in engineering (Gelb, 1974; Catlin, 1989) and meteorology (Lorenc, 1986; Ghil, 1989; Daley, 1991; National Research Council, 1991) and are generally

based on estimation theory (Jazwinski, 1970; Sage and Melsa, 1971) or control theory (Lions, 1971; Berkovitz, 1974; Boguslavskij, 1988). Estimation theory usually involves a sequential estimation process with the melding scheme based on a minimization of the expected error of the estimate in terms of the statistics of both model and data errors. The application of Kalman filtering (Section 3.2) to oceanography, with dynamics evolving both the state of the system and the model errors, has been explored extensively (e.g., Kalman, 1960; Miller, 1986; Bennett and Budgell, 1989; Cohn and Parrish, 1991; Todling and Ghil, 1990, 1994). This approach is hindered theoretically by its inherent linearity and practically by the enormous computational resources required. Computation can be reduced drastically by the relaxation of the exact error minimization constraint. A common practice is to replace the dynamical evolution of the forecast errors by an empirical, generally fixed error hypothesis, thereby reducing an *optimal* estimate via Kalman filtering to a *suboptimal estimate* via so-called *optimal interpolation*. Optimal interpolation has been commonly used for operational weather forecasting (Bengtsson et al., 1981; Phillips, 1982; Lorenc et al., 1991) and real-time at-sea ocean prediction (Robinson et al., 1996a,b).

Control theory methods are generally based on a variational principle and are mostly smoothing processes. A *cost* or *penalty* function is defined that measures the discrepancy (*misfit*) between the state of the system as estimated by dynamical evolution and the observed state. The time/space discrepancy is then minimized (e.g., in a least-squares sense) in terms of the statistics of model and data errors. It is generally an iterative process and computationally demanding; a good first-guess estimate is necessary. A very important aspect of the control theory approach is that it can easily be generalized to include *parameter estimation*. Uncertainties in internal parameters of the dynamical model, and parameters characterizing external forcings and initial and boundary conditions, can all be incorporated in the definition of the penalty function. If control weight assumptions can be adequately formulated, this can provide a powerful method for exploiting the information content of feasible observations. An important example is the generalized inverse method (Section 3.3; Bennett, 1992; Bennett et al., 1996, 1997), including the adjoint method (Section 3.3; e.g., Le Dimet and Talagrand, 1986; Thacker and Long, 1988; Daley, 1991). Finally, we mention that, in practice, hybrid suboptimal methods (a combination of approximate methods) could be utilized for practical estimates (Section 3.6). An example is the use of a simplified Kalman filter or optimal interpolation estimate to provide a first guess for an adjoint or generalized inverse estimate.

1.3. The Coastal Ocean

The coastal ocean, together with its exchanges and connectivities with the deep seas and estuaries, is a complex natural system with a wealth of phenomena and variabilities occurring over a wide range of space and time scales. The physical dynamics may be dominated singly or in combination by wind, buoyancy, tidal or deep-sea exchanges; open boundaries are always present. Complex and severe coastal geometries and steep topographies exist and statistics must generally be expected to be nonstationary and anisotropic. Thus, on the one hand, the coastal ocean is a very challenging regime for the application of data assimilation methods. On the other hand, data assimilation provides a powerful modern technology with which to work efficiently in this difficult but critically important region of the global ocean.

In the remainder of this chapter we develop and illustrate data assimilation concepts with an emphasis on their application to coastal ocean science and management. Section 2 deals with the goals and uses of data assimilations. Section 3 provides a general overview of methodology. Section 4 introduces the special considerations characterizing the coastal ocean problems. Reviews of some progress to date and aspects of the present status of coastal ocean data assimilation are presented in Section 5. Section 6 is an application to the Middle Atlantic Bight off the northeast coast of the United States. Section 7 consists of the summary and conclusions. In the presentation an attempt has been made both to introduce the subject to scientists who are not yet familiar with it, and to provide a useful summary overview at a more advanced level. In particular, many of the mathematical details of Section 3 are not necessary in order to understand most of the subsequent material.

2. Goals and Applications

The specific uses of data assimilation depend on the relative quality of data sets and models and the desired purposes of the field estimates. These uses include the control of errors (predictability and model), the estimation of parameters, the elucidation of real ocean dynamical processes, the design of experimental networks, and ocean monitoring and prediction.

First consider ocean prediction for scientific and practical purposes, which is the analog of meteorological numerical weather prediction (Bengtsson et al., 1981; Ghil, 1989). In the best-case scenario, the dynamical model correctly represents both the basic internal dynamical processes (e.g., synoptic/mesoscale) and the response mechanisms to external forcings (e.g., boundary conditions, surface forcings). Additionally, the observational network is well designed and adequate to efficiently provide initialization data of desired accuracy. The phenomenon of *loss of predictability* nonetheless inhibits accurate forecasts beyond the *predictability limit* for the region and system (Lorenz, 1963, 1975; Bergé et al., 1984; Tennekes et al., 1987; Palmer, 1993; Sheinin and Mellor, 1994). This limit for the global atmosphere is 1–2 weeks (e.g., Houghton, 1991) and for the midocean eddy field of the northwest Atlantic, on the order of weeks to months (Carton, 1987; Adamec, 1989; Walstad and Robinson, 1990). The phenomenon is associated with the nonlinear scale transfer and growth of initial errors (uncertainties in initial internal and initial external conditions). The early forecasts will accurately track the state of the real ocean, but longer forecasts, although representing plausible and realistic synoptical dynamical events, will not agree with contemporary nature. However, this *predictability error* can be controlled by the continual assimilation of data, with adequate spatial coverage and at intervals less than the predictability limit. This is a major use of data assimilation in meteorology and oceanography.

Next, consider the case of a field estimate with adequate data but a somewhat deficient dynamical model. Assimilated data can compensate for the imperfect physics so as to provide estimates in agreement with nature. This is possible if dynamical *model errors* are treated adequately. For instance, if a barotropic model is considered perfect (e.g., adjoint method), and baroclinic real ocean data are assimilated, the field estimate will remain barotropic. In more realistic situations, forecast errors are formed of both model deficiencies and predictability errors; consider the broadening or breakdown of an intense current in long simulations without assimilation.

Assimilation can keep the current system qualitatively correct. In a tropical Pacific study, data assimilation corrected the tendency of the model to thicken the vertical temperature stratification (Fisher and Latif, 1995). Melded estimates with deficient models can be useful, but it is of course important to analyze the discrepancy in the dynamics and to attempt to correct the model physics.

In all ocean inverse problems, one should thus analyze the physical validity of the estimated dynamical model characteristics. If the characteristics are very sensitive and/or tend to have values quite different from their a priori given ones, it is important to search for the sources of the discrepancy. Plausible causes are a wrong estimation of the a priori parameters, a dynamical model not adapted to the data set collected, or a parameter estimation method that has converged to wrong parameter values.

Parameter estimation via data assimilation should make an increasingly significant impact on ocean science in the coming years. The general concept is traditional in oceanography. Ekman (1905) used his original theory of the wind-driven surface boundary layer spiral to estimate the vertical turbulent eddy diffusion of momentum. For many years thereafter, ocean models could be philosophically regarded predominantly as mechanisms for the determination of both vertical and horizontal eddy viscosities and diffusivities (Sverdrup et al., 1942; Stommel, 1965; von Schwind, 1980). Within the last 10 years, parameter estimation has been used successfully in meteorology and oceanography for determining internal and external parameter values (Section 5.2). Inverse models are an established important component of modern oceanography (e.g., Wunsch, 1988; Bennett, 1992; Martel and Wunsch, 1993a,b). Regional field estimates can be improved substantially by boundary condition estimation. For interdisciplinary ocean science, parameter estimation is particularly promising. Biological modelers have been severely hampered by the inability to directly measure in situ rates (e.g., grazing and mortality). By this approach, feasible measurements of quantities, such as concentration distributions of planktons, together with a realistic coupled biological–chemical–physical model, can be used for in situ rate estimation.

Data-driven simulations can provide four-dimensional time series of fully dynamically adjusted fields that are realistic and governed by real ocean physical, biological and chemical dynamical processes. These fields, regarded as (numerical) experimental data, can thus serve as high-resolution, accurate and complete data sets for dynamical studies. Balance of terms studies can be carried out to determine fluxes and rates for energy, vorticity, productivity, grazing, carbon flux and so on. Case studies can be carried out and statistics and general processes can be inferred for simulations of sufficient duration. Of particular importance are observation system simulation experiments (OSSEs), which first entered meteorology almost 30 years ago (Charney et al., 1969). By subsampling the simulated "true" ocean, experimental networks and monitoring arrays can be designed to provide efficient field estimates of requisite accuracies. Data assimilation and OSSEs develop the concepts of data, theory and their relationship beyond those of the classical scientific methodology; Smith's (1993) schematic of these ideas is shown in Fig. 20.2.

For a period of almost 300 years, scientific methodology was powerfully established on the basis of two essential elements: (1) experiments/observations and (2) theory/models. Today, due to powerful computers and evolving concepts of information, science is based on three fundamental concepts: experiment, theory and simulation. Simulation with validated dynamics provide numerically generated databases

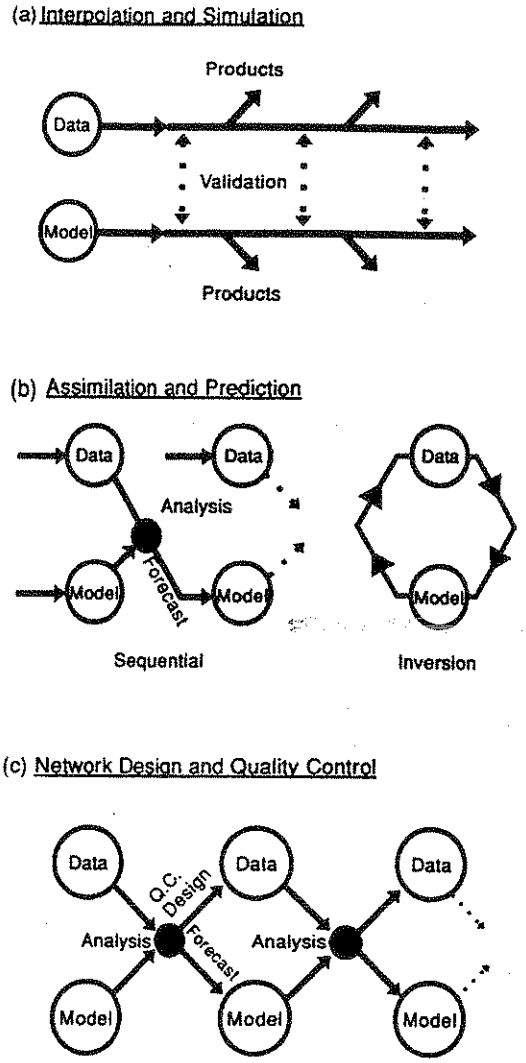


Fig. 20.2. Schematic of the different interactions between data and models: validation, assimilation, OSSEs and quality control. (From Smith, 1993.)

for serious dynamical studies. Furthermore, since our best field estimates today are based on data assimilation (i.e., the melding of observation and dynamics), our very perception and conceptions of nature and reality require philosophical development.

It is apparent from the discussion above that marine operations and ocean management must essentially depend on data assimilation methods for efficient and accurate practical field estimates. Regional data-driven simulations should be coupled to multipurpose management models for risk assessments and for the design of operational procedures. Regional multiscale ocean prediction and monitoring systems, designed by OSSEs, should be established to provide ongoing nowcasts and forecasts with predictability error controlled by updating. Both simple and sophisticated versions of such systems are possible and relevant.

3. Overview of Data Assimilation Methods

The purpose of this section is to give an overview of most methods available for data assimilation (DA) in ocean sciences. DA is an estimation problem by definition (Section 1), but the schemes for solving the assimilation problem have different backgrounds. They can generally be related to either estimation theory or control theory (Sections 3.2 and 3.3). For the methods that are not derived directly from either estimation or control theory, or that can be used without distinction in both frameworks, three other classes have been introduced: direct minimization, stochastic and hybrid methods (Sections 3.4–3.6). In this overview, within each class, the schemes have been ordered by increasing complexity, which generally implies increasing computer resources required but also, hopefully, increasing assimilation quality. We distinguish between theoretically "optimal" and "suboptimal" schemes. Although optimal schemes are preferred because they provide better estimates, in fact, suboptimal methods are generally the ones in operational use in oceanography and meteorology today (Todling and Cohn, 1994).

The three types of estimation problems are filtering, forecasting and smoothing (Gelb, 1974). In the three cases, the problem is to estimate the state of the system under study at a certain time t but with different use or availability of data (Section 1.2). The notions of *dynamical model* and *measurement model* are defined in Section 3.1. Common assumptions are made throughout to facilitate the theoretical understanding of the differences and similarities among DA schemes (Lermusiaux, 1997). Here the description of the equations is kept to a minimum and it is not expected that complete understanding of the methods can be achieved from this presentation alone. It is simply our intention to provide a guide to the methods available, with references to facilitate exploring methods of possible interest. The presentation advances from simple to complex methods and some readers may wish to look at complex schemes first (end of Sections 3.2 and 3.3). The notation is uniform so as to provide the basis for mathematical reference and intercomparisons. The assumptions made about the generic true ocean, dynamical and measurement models as well as tables summarizing the equations of the schemes addressed are also given for easy reference in the Appendix. For comprehensive reviews or reports on theories and applications of DA in atmosphere and ocean sciences as well as practical considerations on computer resources required, we refer to Bengtsson et al. (1981), Haidvogel and Robinson (1989), Anderson and Willebrand (1989), Ghil and Malanotte-Rizzoli (1991), Daley (1991), Bennett (1992), Tellus (1993), Brasseur and Nihoul (1994), Evensen (1994b), Brasseur (1995), Malanotte-Rizzoli (1996) and Wunsch (1996). We mention again that some readers may wish to omit some mathematical details.

3.1. Dynamical and Measurement Models

The notions of dynamics of the ocean system under study, of state variables associated with that dynamics and of dynamical model were introduced in Section 1.1. To perform an assimilation, one needs to meld the dynamics and the incoming data which are characterized by a *dynamical model* and a *measurement model*, respectively.

The *dynamical model* is a mathematical representation of the dynamics of the ocean system which defines the continuous time/space evolution of the *state variables* of the system (e.g., the primitive equation model). State variables, however, might not be measurable directly, and this must be taken into account.

The *measurement model* is a time/space relation (physical, statistical) that links dynamical model state variables to sensor data. It is, by definition, a directed mapping, from the state space to the space of observables or measurement space (Sage and Melsa, 1971; Gelb, 1974; Yavin, 1985; Catlin, 1989), and the available sensor data have to be filtered to the scales of interests. For a measurement model example, consider the case where one measures temperature and salinity but uses a quasi-geostrophic dynamical model. The measurement model is then a model that links the gridded values of the quasi-geostrophic stream function, Ψ, to the (T, S) data. The measurement model does not necessarily involve all dynamical state variables. A relationship that relates gridded salinity of a primitive equation model to temperature observations (e.g., by a T–S relationship) is also a measurement model. In oceanography, the number of discrete state variables is generally much larger than the amount of measurements (sparse, noisy data), and the measurement model strictly by itself is formally not directly invertible; an additional criterion is needed (e.g., weighted least squares). In DA methods, the measurement model should thus be perceived as a *weak constraint on the dynamics*. It is the data-dynamics melding criterion that determines how the data influence the state variables. In fact, even for the initialization of a dynamical model from historical data, one should combine the set of sparse measurements with dynamical considerations (e.g., basic state in geostrophic balance) to infer the complete fields of state variables and their associated errors (e.g., feature models, structured data models; Lozano et al., 1996). The sole direct least-squares inversion of the sparse data set would give in most cases a poor estimate of the ocean state. Finally, the measurement constraint can be nonlinear and time dependent since both the observation network and the measurement type may vary in time. For a given observation, there are many possible measurement models, even for the same dynamical model (e.g., altimetry data) and the determination of efficient measurement models is an important issue.

Once both the dynamical model and the measurement model have been defined analytically by their respective equations, a discretized version in time and space is used in computer simulations. The discrete models used, as well as the assumptions made for easy comparisons of methods, are discussed in the Appendix. We no longer use the adjective *discretized* unless necessary to prevent confusion. Equation 1 describes the statistics of the generic true ocean dynamical model and equation 2 the true measurement model:

$$\boldsymbol{\psi}_k^t = \mathbf{A}_{k-1}\boldsymbol{\psi}_{k-1}^t + \mathbf{w}_{k-1} \tag{1}$$

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

The classic deterministic versions of these models are obtained by canceling the random noise $\mathbf{w}_{k-1}$ and $\mathbf{v}_k$.

3.2. *Estimation Theory (Sequential Estimation)*

Estimation theory encompasses theories used to estimate the state of a system by combining, usually with a statistical approach, all available reliable knowledge of the system, including measurements and theoretical laws or empirical principles. In oceanography, estimation theory first formulates crucial *statistical hypotheses* about

the true ocean dynamical model, the measurement model and their respective uncertainties (see, e.g., the Appendix). An *estimation criterion* is then associated with those hypotheses so as to determine the optimal melding of all reliable information. The estimation or melding criterion is as crucial as the *a priori* hypotheses since it determines the respective influence of dynamics and data onto the ocean state estimate.

In practice, discrete estimation schemes are usually a succession of two steps. A dynamical model is first employed to issue a *forecast* (i.e., state vector and error forecasts) and a statistical criterion is then used to meld the incoming data with the dynamical forecast (referred to as the *analysis* step). For simplicity, the analysis step is usually chosen to be a linear combination of the data and the forecast (linear melding), and this will be the case in this generic discussion. The estimation schemes discussed hereafter mainly differ from each other by the respective coefficients of data and dynamics in that linear combination.

In estimation theory, an important part of the a priori hypotheses is the explicit definition of the statistics of model noise and data uncertainties (correlations, error covariance matrices or the full probability density function). Common sense and some theoretical studies are generally used to characterize the uncertainties in the initial conditions, dynamical model and measurement model. Error theory is in fact an area of active research in data assimilation today (e.g., Balgovind et al., 1983; Lönnberg and Hollingsworth, 1986; Phillips, 1986; Cohn and Parrish, 1991; Daley, 1992a–c; Jiang and Ghil, 1993; Cohn, 1993; Bouttier, 1994). The more precise the error specifications, the greater are the potentials of data assimilation (e.g., Section 6.2). Here, for generic comparison purposes, classical assumptions were made (see the Appendix). But assimilation results depend on the statistical hypotheses and one should study the sensitivity of the estimate to those a priori choices and definitions (Gelb, 1974). For instance, most schemes presented here are related in some fashion to least-squares criteria, which are very useful andhave had great success. However, least-squares or average solutions can sometimes be far from reality, especially in systems with multimodal probability density functions and very noisy, sparse data. Other melding criteria, such as the general maximum likelihood, the minimax criterion or associated variations, might thus be more appropriate in special situations (Boguslavskij, 1988).

Estimation of the model parameters is possible with sequential estimation methods. A common approach is to add to the initial model equations a simple stochastic evolution equation for each parameter to be estimated. For each spatially (time) varying parameter, discretized on a three-dimensional grid, the size of the state vector is increased by the number of grid points. To reduce this additional number of variables, predetermined functional parameter expansions are used. Since the model characteristics are model dependent and their estimation leads to a nonlinear estimation problem (products of parameters and state variables), parameter estimation techniques will not be discussed further in this generic section on estimation methods. Parameter estimation in coastal ocean models is reviewed separately in Section 4.2. In the following discussion, equation numbers refer to the tables in the Appendix, where all symbols are also defined.

Direct Insertion

At all data points, one replaces the forecast values by the observed ones; the direct insertion estimate (3) is data at observation points (4),

$$\hat{\boldsymbol{\psi}}_k(-) = \mathbf{A}_{k-1}\hat{\boldsymbol{\psi}}_{k-1}(+) \tag{3}$$

with

$$(\hat{\boldsymbol{\psi}}_k)_d(+) = \mathbf{d}_k \qquad [\mathbf{d}_k = \mathbf{C}_k\hat{\boldsymbol{\psi}}_k(+)] \tag{4}$$

at the m data points and $\hat{\boldsymbol{\psi}}_k(+) = \hat{\boldsymbol{\psi}}_k(-)$ elsewhere. The a priori statistical hypothesis is that data are always exact. For a comparison with the Kalman filter (KF) scheme (12–16), the direct insertion method assumes that $\mathbf{R}_k = 0$ and $\mathbf{C}_k\mathbf{Q}_k\mathbf{C}_k^T = \infty$ with $\mathbf{P}_k = \mathbf{Q}_k = 0$ elsewhere (model errors are infinite at the data points).

Blending

At all data points, the blending estimate (6) is a scalar linear combination of the forecast (5) and data values:

$$\hat{\boldsymbol{\psi}}_k(-) = \mathbf{A}_{k-1}\hat{\boldsymbol{\psi}}_{k-1}(+) \tag{5}$$

with

$$(\hat{\boldsymbol{\psi}}_k)_d(+) = \alpha\mathbf{d}_k + (1-\alpha)\mathbf{C}_k\hat{\boldsymbol{\psi}}_k(-) \tag{6}$$

at the m data points and $\hat{\boldsymbol{\psi}}_k(+) = \hat{\boldsymbol{\psi}}_k(-)$ elsewhere. Within the assumptions discussed in the Appendix, it is an unbiased estimate. The parameter α is a user-assigned parameter. For a comparison with the Kalman filter scheme (12–16), the blending method assumes that $\mathbf{R}_k = (1-\alpha)\mathbf{I}$ and $\mathbf{C}_k\mathbf{P}_k(-)\mathbf{C}_k^T = \alpha\mathbf{I}$ with $\mathbf{P}_k = 0$ elsewhere (model errors at data points = α and = 0 elsewhere).

Nudging or Newtonian Relaxation Scheme

The nudging scheme (Anthes, 1974) can be seen either as a simplification of the KF or as the dynamical model relaxed toward the observations. In discrete notation, both points of view lead to the same equations, but the relaxation scheme is clearer in the continuous notation. For the model forecast to relax to the data, one adds a term proportional to $(\mathbf{d} - \mathbf{C}\hat{\boldsymbol{\psi}})$ in the state continuous evolution equation, to obtain

$$\frac{\partial\hat{\boldsymbol{\psi}}}{\partial t} = \mathbf{L}\hat{\boldsymbol{\psi}} + \mathbf{N}\mathbf{C}^+(\mathbf{d} - \mathbf{C}\hat{\boldsymbol{\psi}})$$

where $\hat{\boldsymbol{\psi}}$ is the vector of state variables, $\mathbf{C}^+ \in \mathbb{R}^{n\times m}$ is a generalized inverse of $\mathbf{C}$ and $\mathbf{N} \in \mathbb{R}^{n\times n}$ the *diagonal* matrix of nudging coefficients. In most cases, one only relaxes the state variables that are, up to a scalar linear transformation, directly observed (direct relaxation). In our notation, $n-m$ diagonal elements of N are then zero and one can order the equations so that the only nonzero elements of N, C^+, and $\mathbf{C}$ are on their respective diagonal. But if all state variables are relaxed, C^+ is the inverse mapping of the measurement model. This should perform better than direct relaxation as long as C^+ is derived from appropriate dynamical hypotheses such as feature- or data-structured models (e.g., Lozano et al., 1996); the Moore–Penrose generalized inverse would be poor. The nudging coefficients contained in N are in s^{-1}. They can be chosen to be time dependent but

in our notation need to be positive for $\hat{\boldsymbol{\psi}}$ to relax to the observations $\mathbf{d}$. They cannot be too large, to avoid model disruption (Stauffer and Bao, 1993). They are related to the relaxation time of the state variable and should ideally depend on the evolution of the basic state dynamics. They should also be a function of some a priori estimates of dynamical model uncertainties and data errors. Obviously, this relaxation time must be less than the predictability threshold or the e-folding time by which the data lose much of their influence on the forecast state.

Using, for instance, the Euler scheme to discretize time, and evaluating the relaxation term at time t_k, the continuous nudging scheme becomes

$$\hat{\boldsymbol{\psi}}_k = \mathbf{A}_{k-1}\hat{\boldsymbol{\psi}}_{k-1} + \tilde{\mathbf{K}}_k\mathbf{C}_k^+(\mathbf{d}_k - \mathbf{C}_k\mathbf{A}_{k-1}\hat{\boldsymbol{\psi}}_{k-1}) \tag{7}$$

where the *diagonal* matrix $\tilde{\mathbf{K}}_k$ is the discretized analog of N. The elements of the diagonal $\tilde{\mathbf{K}}_k$ are assigned a positive value. From the filtering point of view, the nudging scheme is a simplification of the KF, where the Kalman gain is assigned and diagonal; see (12) and (14) with $\mathbf{R}_k$, $\mathbf{Q}_k$ and $\mathbf{P}_k(+)$ diagonal. Usually, the simplified diagonal gain $\tilde{\mathbf{K}}_k$ is chosen to be a constant, but in principle it could vary with time. Referring to continuous-time relaxation and using the KF as the optimal scheme, the nudging coefficients should be functions of simplified error hypotheses for the dynamics, the model and the data, [i.e., simple $\mathbf{P}_k(+)$, $\mathbf{Q}_k$ and $\mathbf{R}_k$]. For nudging examples, we refer to Zou et al. (1992) and Fukumori and Malanotte-Rizzoli (1995).

Optimal Interpolation or Statistical Interpolation

The term *optimal interpolation* (OI) is usually used in oceanography and *statistical interpolation* (SI) in meteorology (Daley, 1991) but both terms refer to the same melding method (9). The assimilation via OI is described by

$$\hat{\boldsymbol{\psi}}_k(-) = \mathbf{A}_{k-1}\hat{\boldsymbol{\psi}}_{k-1}(+) \tag{8}$$

and

$$\hat{\boldsymbol{\psi}}_k(+) = \hat{\boldsymbol{\psi}}_k(-) + \tilde{\tilde{\mathbf{K}}}_k[\mathbf{d}_k - \mathbf{C}_k\hat{\boldsymbol{\psi}}_k(-)] \tag{9}$$

with gains $\tilde{\tilde{\mathbf{K}}}_k$ empirically assigned. The sequence of $\tilde{\tilde{\mathbf{K}}}_k$ is user assigned and usually based on an empirical forecast error covariance matrix.

OI is a simplification of the KF scheme as shown by equations (12–16) and (8, 9). In both schemes, the analysis step (9 or 14) is a linear combination of the dynamical model forecast and of the difference between the actual observations $\mathbf{d}_k$ and the model predicted values for those observations, $\mathbf{C}_k\boldsymbol{\psi}_k(-)$. This difference is commonly called the model-data misfits, the observation residuals or the innovation vector (Daley, 1991). The matrix weighting these misfits, $\tilde{\tilde{\mathbf{K}}}_k$, is called the gain matrix. In the KF algorithm, the gain $\mathbf{K}_k$ is computed and updated internally (15), whereas in the OI scheme it is empirically assigned. If the assigned gain is diagonal and kept constant in time, the OI (8, 9) and the constant gain version of the nudging scheme (7) are equivalent. But the OI gain is not generally diagonal, as it is usually analytically expressed as a function of the nondiagonal *field-to-measurement-point* correlation matrix and

of the *measurement error* covariance matrix (Bretherton et al., 1976; Daley, 1991; Bennett, 1992). For instance, the operational OI scheme at Harvard (HOPS, Section 6) first locally objectively analyzes (two-step OA) the data to produce a three-dimensional gridded field and then blends this local OA three-dimensional gridded data with the forecast fields, using the objectively analyzed data error covariance to empirically determine the OI gain matrix (Robinson et al., 1996a; Lozano et al., 1996).

Method of Successive Corrections

Instead of correcting the forecast fields only once by linear combination of the data and forecast as in the previous methods, multiple iterations are performed. The forecast,

$$\hat{\boldsymbol{\psi}}_k(-) = \mathbf{A}_{k-1}\hat{\boldsymbol{\psi}}_{k-1}(+) \tag{10}$$

is iteratively melded with the data,

$$\hat{\boldsymbol{\psi}}_k^{j+1}(+) = \hat{\boldsymbol{\psi}}_k^{j}(+) + \mathbf{W}^j[\mathbf{d}_k - \mathbf{C}_k\hat{\boldsymbol{\psi}}_k^{j}(+)] \tag{11}$$

with $\hat{\boldsymbol{\psi}}_k^{0}(+) = \hat{\boldsymbol{\psi}}_k(-)$ and $\mathbf{W}^j$ assigned weight matrices, in general function of the iteration j. In practice, only two or three iterations (11) are performed (Daley, 1991) and the final estimate depends on the finite sequence of weights $\mathbf{W}^j$. For example, one may choose to correct the scales or processes of interest one after the other (e.g., geostrophic before ageostrophic, implicitly assuming that the data errors of each of these processes are uncorrelated). $\mathbf{W}^j$ should then be a function of the correlations between features (processes) of scales that are corrected at melding j. In the infinite sequence limit, if (11) converges, then $\hat{\boldsymbol{\psi}}_k^{j+1} = \hat{\boldsymbol{\psi}}_k^{j} = \hat{\boldsymbol{\psi}}_k^{\infty}$ and (11) converges to the data values at the observation locations, $\mathbf{C}_k\hat{\boldsymbol{\psi}}_k^{\infty} = \mathbf{d}_k$, regardless of the sequence $\mathbf{W}^j$ chosen. The sequence $\mathbf{W}^j$ then only determines how the data are extrapolated at nonobserved variable locations. To converge to the Kalman update (14, 15), simple modifications of (11) are thus necessary (Bratseth, 1986; Daley, 1991; Lorenc, 1992). Conditions for equivalence and convergence are addressed by those authors, but the exact update (14, 15) is still optimal in the least-squares sense only if the covariances used are correct. In some cases these covariances are not well known, and criteria such as computer time and/or sensible representation of observed features have favored (11) over the update (14, 15).

Kalman Filter

For a linear system such as that we considered in (1, 2), the Kalman filter (KF) is the sequential, unbiased, minimum error variance estimate based on a *linear* combination of all past measurements and past linear dynamics (Kalman, 1960):

$$\left\{ \hat{\boldsymbol{\psi}}_k \,|\, \min_{\hat{\boldsymbol{\psi}}_k} J_k = \text{trace}[\mathbf{P}_k(+)], \text{using } [\mathbf{d}_1, \ \ldots \ , \mathbf{d}_k] \right\}$$

The error variance of the KF estimate is less than or equal to that of any other unbiased, linear filtering estimate. Only the error mean (0) and covariance ($\mathbf{P}_k$) need to be known.

The KF can be viewed as a simplification of the general Bayesian estimation. Both Bayesian theory with the maximum entropy principle, and the direct minimum error variance criterion with a linear estimate, lead to the same classical KF algorithm (12–16). But the assumption of Gaussian probability density function is not necessary in either approach. The implicit assumption that the probability density function of the errors is well described locally by a mean and a covariance is sufficient. The derivation of the two steps of the KF algorithm, the forecast of the state vector and of the error,

$$\hat{\boldsymbol{\psi}}_k(-) = \mathbf{A}_{k-1}\hat{\boldsymbol{\psi}}_{k-1}(+) \tag{12}$$

$$\mathbf{P}_k(-) = \mathbf{A}_{k-1}\mathbf{P}_{k-1}(+)\mathbf{A}_{k-1}^T + \mathbf{Q}_{k-1} \tag{13}$$

and the melding of the data with the two forecasts,

$$\hat{\boldsymbol{\psi}}_k(+) = \hat{\boldsymbol{\psi}}_k(-) + \mathbf{K}_k[\mathbf{d}_k - \mathbf{C}_k\hat{\boldsymbol{\psi}}_k(-)] \tag{14}$$

$$\mathbf{K}_k = \mathbf{P}_k(-)\,\mathbf{C}_k^T\,[\mathbf{C}_k\mathbf{P}_k(-)\mathbf{C}_k^T + \mathbf{R}_k]^{-1} \tag{15}$$

$$\mathbf{P}_k(+) = \mathbf{P}_k(-) - \mathbf{K}_k\mathbf{C}_k\mathbf{P}_k(-) \tag{16}$$

are discussed in Gelb (1974), Daley (1991), Bennett (1992) and Lermusiaux (1997). In the light of (13) and (16), it is commonly said that the KF generates its own error analysis. Two examples are given by Todling and Ghil (1990) and Miller and Cane (1989).

Kalman Smoother

The Kalman smoother (KS) is also a linear, unbiased, minimum error variance estimate. However, it solves a smoothing problem and uses all measurements available within the complete assimilation period $[t_0, t_N]$ to estimate the state of the system at all times t_k, where $t_0 \le t_k \le t_N$ (fixed interval smoothing). If we denote $\boldsymbol{\psi}_{k/N}$ as a smoothed estimate at time t_k, using all available data in $[t_0, t_N]$, the KS estimate $\hat{\boldsymbol{\psi}}_{k/N}$ is defined as follows:

$$\left\{\hat{\boldsymbol{\psi}}_{k/N} \middle| \min_{\hat{\boldsymbol{\psi}}_{k/N}} J_k = \text{trace}[\mathbf{P}_k(+)] \text{ using } [\mathbf{d}_1, \ldots, \mathbf{d}_{N-1}]\right\}$$

There are different formulations of the fixed interval smoothing (e.g., Wunsch, 1996). A simple formulation is to use first the forward KF (12–16) and then backward filtering with the inverse of the dynamical model to carry back to time t_k the information contained in the observations made after time t_k. This requires running the forward dynamical model backward in time. Even though this is possible theoretically, it is equivalent and usually easier from a practical point of view to use the Rauch–Tung–Striebel (RTS) algorithm (Gelb, 1974), which does not require running the dynamical model backward, nor does it involve the backward processing of actual measurement data. Future data information is propagated backward by the error covariances and the adjoint dynamical transition matrix A_k^T. Thus both the KF state and error covariances need to be stored at all analysis times, which is in gen-

eral very demanding on memory resources. The filter part of the RTS is identical to (12–16). The smoothing steps are given by

$$\hat{\boldsymbol{\psi}}_{k/N} = \hat{\boldsymbol{\psi}}_k(+) + \mathbf{L}_k[\hat{\boldsymbol{\psi}}_{k+1/N} - \hat{\boldsymbol{\psi}}_{k+1}(-)] \tag{17}$$

$$\mathbf{P}_{k/N} = \mathbf{P}_k(+) + \mathbf{L}_k[\mathbf{P}_{k+1/N} - \mathbf{P}_{k+1}(-)]\mathbf{L}_k^T \tag{18}$$

with $\hat{\boldsymbol{\psi}}_{N/N} = \hat{\boldsymbol{\psi}}_N(+)$, $\mathbf{L}_k = \mathbf{P}_k(+)\,\mathbf{A}_k^T\,\mathbf{P}_{k+1}^{-1}(-)$ and $\mathbf{P}_{N/N} = \mathbf{P}_N(+)$. Another example of fixed interval smoothing is the Gelfand and Fomin (1963) sweep algorithm, which requires the storage of the gains K_k but not the error covariances (e.g., Bennett, 1992; Wunsch, 1996). To our knowledge, only simplified RTS algorithms have been used in practice (Budgell, 1986; Fukumori et al., 1993). In (17) and (18), $\hat{\boldsymbol{\psi}}_{k/N}$ and $\mathbf{P}_{k/N}$ represent, respectively, the smoothed state estimate and smoothed error covariance at time t_k, using all observations up to time t_N. Note that both are continuous; their filtered versions (14) and (16) were not. For examples, we refer to (Bennett and Budgell (1989) and Fukumori et al. (1993).

3.3. Control Theory (Calculus-of-Variation Approach, Variational Assimilation)

All variational assimilation approaches perform a global time/space adjustment of the model solution to the complete set of available observations and thus solve a smoothing problem. Except for the KS, all sequential estimation methods presented in Section 3.2 solve a filtering problem. Theoretically, this means that each individual observation is being used only once without feedback to earlier times. Real-time shipboard assimilation can be performed with techniques presented in this section only if for all new data acquired, a global time/space readjustment of the previous ocean state estimate is made. In most realistic cases, this real-time smoothing is beyond the computing power available today.

In the constrained optimization framework, the smoothing problem for oceanography can be defined, in general, as the minimization of a cost function penalizing the time/space misfits between the data and the ocean model state estimate which is constrained by the dynamical model equations. The control theory approach interprets the misfits as part of the unknown controls of the ocean system (Lions, 1971; Le Dimet and Talagrand, 1986), but still tries to solve a field estimation problem. Additional penalty functionals can be added to the cost function, like a time/space smooth field constraint or expressions restraining the variations of the a priori chosen dynamical model's parameters and forcings if these model characteristics are allowed to vary. Variational assimilation (e.g., Courtier, 1995) thus allows the estimation of model parameters, including initial and boundary conditions, subgrid scale processes and uncertainties in forcing functions.

Estimation theory requires a priori statistical assumptions for the model noise and data errors. Similarly, the control theory results are dependent on the a priori control weights and on the expression of any added penalties such as a smooth estimate constraint. In most cases, the cost function is naturally chosen to be quadratic. It then measures the Euclidean time/space distance between the dynamical model solution and the available observations (weighted least squares).

In the now-classical terminology (Sasaki, 1970), the dynamical model can be either considered as a *strong constraint* (leading to the adjoint method) or as a *weak con-*

straint (leading to generalized inverses) (Reid, 1968; Daley, 1991; Bennett, 1992). The strong constraint approach assumes the model to be exact and provides exact consistency between the dynamics of the model and of the final smoothed estimate. The evolution of all state variables is deterministic and fixed once the initial and boundary conditions are given. In contrast to Section 3.2, the boundary variables here are not explicitly part of the state vector, since the boundary conditions are explicitly treated as separate constraints. The auxiliary conditions are thus the only free variables of the problem; usually, the boundary values are fixed and the free variables are the initial conditions only. On the other hand, the estimation of model parameters is sometimes the only goal of the variational problem. The cost function then comprises data-model misfits and parameter penalties, but the initial conditions are fixed and the only free variables are the parameters to be estimated. We refer to Section 5.2 for coastal ocean parameter estimation.

If the model is considered as a weak constraint, the final smoothed estimate is only in approximate consistency with the dynamical model equation. At least implicitly, one thus assumes dynamical model errors and implies stochastic "true" model state equations as in sequential estimation (Section 3.2). The weak constraint approach solves a "stochastic control" problem (Boguslavskij, 1988). The particular form of the weak dynamical model constraint and its weight relative to the data constraint defines the bounds of the approximate dynamical consistency.

In engineering, the weights and form of the cost function are determined by the objectives of the control/optimization problem to be solved. Similarly, in oceanography, the data and model weights and the shape of the cost function can be chosen as one pleases. This freedom exposes the weak constraint formulation to the criticism that any desired field estimate could be made a solution of the variational problem by manipulation of the cost function. In fact, the strong constraint approach is just one of these particular manipulations, the limiting choice of infinite weights for the model equations. Due to the approximate dynamics and the restrictions to relevant phenomena and scales of interest (Section 1.1), the strong dynamical constraint formulation in ocean science certainly needs to be justified as much as any other weighted smoothed estimate. These considerations are also applicable to estimation theory (Section 3.2). A rational choice for the weights and form of the cost function is thus very important. Statistically, a natural selection is to pick the dynamical model (data) weight inversely proportional to an a priori specified model (data) error, especially when the cost function has a weighted least-squares form. The lower the specified model (data) errors, the higher the weight given to the model (data). For such a quadratic cost function, where the data and model weights are the inverses of the error covariances, the KS (Section 3.2) and the generalized inverse give exactly the same estimate for any n-dimensional linear dynamical system (Lermusiaux, 1997). Finally, the remarks made in Section 3.2 on DA sensitivities to the hypotheses and melding criterion also apply here.

Adjoint Method

In the adjoint method, the dynamical model is a strong constraint. In this section we omit forcing and internal parameter estimation. To simplify notations, the boundary variables and associated conditions are included in the state vector evolution (1). The only free variables are thus the initial conditions. The functional to be minimized is the sum of two penalties. One penalty weights the initial condition uncertainty with the a priori initial error covariance $\mathbf{P}_0$. The other is the sum over all observation times

t_k of the measurement model errors (data-model misfits at observation locations), balanced by measurement error covariances $\mathbf{R}_k$. To optimize the functional, one may take its derivative with respect to $\boldsymbol{\psi}_0^t$, but this approach requires knowledge of the dynamical model solution in closed form [i.e., the expression of all $\boldsymbol{\psi}_k^t$ as a function of $\boldsymbol{\psi}_0^t$ using (1)]. Since this is not normally feasible for complex models, a classical approach to solving the strong constraint minimization is to add to the cost function the product of undetermined Lagrange multipliers ($\boldsymbol{\lambda}_k \in \mathbb{R}^n$) with the dynamical constraints, leading to the expression

$$\min_{\hat{\boldsymbol{\psi}}_k} J_N = \boldsymbol{\epsilon}_0^T \mathbf{P}_0^{-1} \boldsymbol{\epsilon}_0 + \sum_{k=1}^{N-1} \mathbf{v}_k^T \mathbf{R}_k^{-1} \mathbf{v}_k + \sum_{k=1}^{N} 2\boldsymbol{\lambda}_{k-1}^T \mathbf{w}_{k-1} \tag{19}$$

The initial uncertainty is denoted by $\boldsymbol{\epsilon}_0 = \boldsymbol{\psi}_0^t - \boldsymbol{\Psi}_0$, where $\boldsymbol{\Psi}_0$ is the initial guess at the initial state conditions. All other notations use the same conventions as in previous sections. A necessary condition for an extremum of the quadratic form J_N is to set to zero the derivatives with respect to all free variables. This yields the following Euler–Lagrange equations,

$$\hat{\boldsymbol{\psi}}_k = \mathbf{A}_{k-1} \hat{\boldsymbol{\psi}}_{k-1}, \qquad k = 1, \ldots, N \tag{20}$$

$$\hat{\boldsymbol{\psi}}_0 = \boldsymbol{\Psi}_0 + \mathbf{P}_0 \mathbf{A}_0^T \boldsymbol{\lambda}_0 \tag{21}$$

$$\boldsymbol{\lambda}_{k-1} = \mathbf{A}_k^T \boldsymbol{\lambda}_k + \mathbf{C}_k^T \mathbf{R}_k^{-1} (\mathbf{d}_k - \mathbf{C}_k \hat{\boldsymbol{\psi}}_k), \qquad k = 1, \ldots, N-1 \tag{22}$$

$$\boldsymbol{\lambda}_{N-1} = 0 \tag{23}$$

defining a two-point boundary value problem in $[t_0, t_N]$ for the $2N+1$ estimates $\hat{\boldsymbol{\psi}}_k$ and $\boldsymbol{\lambda}_k$, both $\in \mathbb{R}^n$ (Lermusiaux, 1997). The Euler–Lagrange equation (22) is the so-called adjoint equation and (23) its boundary condition (Brockett, 1970). The Lagrange multipliers or adjoint variables measure the cost function sensitivity to the dynamical effects of the reliable data, $\frac{1}{2}(\partial J_N / \partial \mathbf{w}_{k-1} = \boldsymbol{\lambda}_{k-1}$. The physics of the adjoint model (22) between time t_k and t_{k-1} is described by the backward adjoint operator $\mathbf{A}_k^T$. Even in linear ocean science, the dynamical model is in general not self-adjoint, and the backward propagation of reliable data (22) can have quite different physics from the forward model (20). The study of the adjoint model has received much attention for these and other reasons: fastest-growing error modes at t_0 and t_N, sensitivities to auxiliary conditions and identification of instability regions (Le Dimet and Talagrand, 1986; Lorenc, 1986; Lacarra and Talagrand, 1988; Thacker and Long, 1988; Derber, 1989; Farrell and Moore, 1992; Tziperman et al., 1992a,b; Bergamasco et al., 1993; and others).

The following iterative descent algorithm for solving the system (20–23) has often been included in the term *adjoint method*. First, the dynamical model is integrated from t_0 to t_N using a first guess $\boldsymbol{\Psi}_0^1$ for initial condition. Next, the adjoint model is integrated backward using (23) as initial condition (no data at t_N) to find $\boldsymbol{\lambda}_{N-1}$, $\ldots$, $\boldsymbol{\lambda}_0$. By construction, the only component of ∇J_N left is

$$\frac{1}{2} \frac{\partial J_N}{\partial \boldsymbol{\psi}_0^t} = \mathbf{P}_0^{-1} (\boldsymbol{\psi}_0^t - \boldsymbol{\Psi}_0^1) - \mathbf{A}_0^T \boldsymbol{\lambda}_0$$

which is equal to $-A_0^T\boldsymbol{\lambda}_0$ if $\hat{\boldsymbol{\psi}}_0$, the estimate of $\boldsymbol{\psi}_0^f$, is chosen equal to $\boldsymbol{\Psi}_0^1$. If the amplitude of the gradient is too big, $\{\hat{\boldsymbol{\psi}}\}_0$ is too far from its optimum value and one needs to iterate. A line search is made along the direction given by ∇J_N and the value of $\hat{\boldsymbol{\psi}}_0$ canceling ∇J_N is given by (21). A new iteration can then be started from $\boldsymbol{\Psi}_0^2 = \hat{\boldsymbol{\psi}}_0^1 = \boldsymbol{\Psi}_0^1 + \mathbf{P}_0^1\mathbf{A}_0^T\boldsymbol{\lambda}_0^1$ and so on, until an acceptable value is reached for ∇J_N. Sometimes, other control algorithms (Boguslavskij, 1988; Daley, 1991; Wunsch, 1996) based on adjoint equations have been called adjoint methods by oceanographers, and readers should carefully discern what this term means in each case.

Generalized Inverse Problem

Any problem where one tries to estimate the unknown causes of known consequences is normally called an inverse problem. Classical finite-dimensional inverse methods in oceanography solve such problems (Bennett, 1992), the consequences usually being insufficient measurements, and the causes, hypothesized processes. Expanding this approach to the simultaneous best fit of a data set with a complete time-space continuous dynamical model leads to generalized inverses. The best fit is here again defined in a least-squares sense as by Bennett (1992). In discrete terms, the quadratic penalty to be minimized is given by

$$\min_{\hat{\boldsymbol{\psi}}_k} J_N = \boldsymbol{\epsilon}_0^T\mathbf{P}_0^{-1}\boldsymbol{\epsilon}_0 + \sum_{k=1}^{N-1}\mathbf{v}_k^T\mathbf{R}_k^{-1}\mathbf{v}_k + \sum_{k=1}^{N}\mathbf{w}_{k-1}^T\mathbf{Q}_{k-1}^{-1}\mathbf{w}_{k-1} \tag{24}$$

The last term in (24) now quadratically weights the white noise in time dynamical model uncertainties, $\mathbf{w}_{k-1}$, with a priori model error covariances ($\mathbf{Q}_{k-1}$). Using equations 1 and 2, the free variables are the $(N+1)$ state vectors values $\boldsymbol{\psi}_k^f$, $k = 0, \ldots, N, \in \mathbb{R}^n$. The Euler–Lagrange equations 25–28 are obtained by setting to zero the derivatives of J_N with respect to these free variables. Strictly for notational convenience, the variables $\boldsymbol{\lambda}_k$, $k = 0, \ldots, N-1$, are added and defined by the backward evolution equation 27. This equation is the same adjoint model as equation 22. It has similar properties and the $\boldsymbol{\lambda}$'s are called adjoint variables:

$$\hat{\boldsymbol{\psi}}_k = \mathbf{A}_{k-1}\hat{\boldsymbol{\psi}}_{k-1} + \mathbf{Q}_{k-1}\boldsymbol{\lambda}_{k-1}, \qquad k = 1, \ldots, N \tag{25}$$

$$\hat{\boldsymbol{\psi}}_0 = \boldsymbol{\Psi}_0 + \mathbf{P}_0\mathbf{A}_0^T\boldsymbol{\lambda}_0 \tag{26}$$

$$\boldsymbol{\lambda}_{k-1} = \mathbf{A}_k^T\boldsymbol{\lambda}_k + \mathbf{C}_k^T\mathbf{R}_k^{-1}[\mathbf{d}_k - \mathbf{C}_k\hat{\boldsymbol{\psi}}_k], \qquad k = 1, \ldots, N-1 \tag{27}$$

$$\boldsymbol{\lambda}_{N-1} = 0 \tag{28}$$

Comparison of this Euler–Lagrange system (25–28) with (20–23) shows that the generalized inverse estimates the dynamical model uncertainties by $\mathbf{Q}_{k-1}\boldsymbol{\lambda}_{k-1}$. This term couples the state vector evolution (25) with the adjoint evolution (27). Without further manipulations, this coupling renders the iterative solution of the two-point boundary problem (25–28) very difficult. For any linear system, this could be solved by using the KS scheme (Section 3.2), which also minimizes J_N given by (24). Two other approaches can also be used and are discussed next: the representer method and the direct minimization methods.

Representer Method

Our objective of uniform notation and common hypothesis (see the Appendix) have led to a presentation of the representer method differing from the classic continuous-discrete-time version reviewed by Bennett (Bennett, 1992; Malanotte-Rizzoli, 1996). The essence is the same, however, and the representer method is an algorithm for solving the system (25–28). First recall that for all linear systems (1), one can show that J_N in (24) has a unique global minimum. If a solution is found, it is the unique solution. Using this fact, it can be shown (Bennett, 1992; Lermusiaux, 1997) that the system (25–28) can be reduced to initial value problems by first decomposing $\hat{\boldsymbol{\psi}}_k$ into the expression

$$\hat{\boldsymbol{\psi}}_k = \boldsymbol{\psi}_k^f + \tilde{\mathbf{R}}_{kl}\mathbf{b}_l, \qquad k = 0, \ldots, N \tag{29}$$

This decomposition (29) and its developments could be simply stated as mathematical tools to transform the system (25–28) into initial value problems. However, to ease comprehension, descriptive allusions will be used. The $(N-1)$ time evolving matrices of representers, $\tilde{\mathbf{R}}_{kl} = [\mathbf{r}_{kl}^1, \ldots, \mathbf{r}_{kl}^m] \in \mathbb{R}^{n\times m}$ are matrices whose columns i contains the value of the representer (i, l) at time t_k at each grid point, $\mathbf{r}_{kl}^i \in \mathbb{R}^n$. The index $i = 1, \ldots, m$, numbers the m observations made at time t_l, with $l = 1, \ldots, N-1$. In (29) and hereafter the summation convention is used on the index l only. In total, there are $m(N-1)$ discrete representer vectors to be determined, one for each datum. The continuous analog of a discrete representer vector, $\mathbf{r}_{kl}^i$, is a representer field $\mathbf{r}_l^i(x,t)$ of Bennett (1992). The $(N-1)$ vectors $\mathbf{b}_l \in \mathbb{R}^m$ are vectors of coefficients that need to be determined from the system (25–28) in function of the data residuals, model errors and measurement uncertainties. The vector $\mathbf{b}_l$ balances the representer vectors of the m measurements made at time t_l. In words, it will be shown that the representer $\mathbf{r}_{kl}^i$ maps a weighted combination of all forecast misfits, as seen from the position of scalar datum i, available at time t_l, onto the field estimate at time t_k, $\hat{\boldsymbol{\psi}}_k$ (36).

Determining the $\tilde{\mathbf{R}}_{kl}$ and $\mathbf{b}_l$ in (29) can be done by further assuming that the following initial value problems for the forecast $\boldsymbol{\psi}_k^f$ (30), the matrices of representers $\tilde{\mathbf{R}}_{kl}$ (31) and the matrices of so-called representers adjoints $\boldsymbol{\Gamma}_{kl}$ (32) are satisfied:

$$\boldsymbol{\psi}_k^f = \mathbf{A}_{k-1}\,\boldsymbol{\psi}_{k-1}^f \quad \text{with} \quad \boldsymbol{\psi}_0^f = \boldsymbol{\Psi}_0 \tag{30}$$

$$\tilde{\mathbf{R}}_{kl} = \mathbf{A}_{k-1}\tilde{\mathbf{R}}_{k-1,l} + \mathbf{Q}_{k-1}\boldsymbol{\Gamma}_{k-1,l} \quad \text{with} \quad \tilde{\mathbf{R}}_{0,l} = \mathbf{P}_0\mathbf{A}_0^T\boldsymbol{\Gamma}_{0,l} \tag{31}$$

$$\boldsymbol{\Gamma}_{k-1,l} = \mathbf{A}_k^T\boldsymbol{\Gamma}_{kl} + \mathbf{C}_k^T\boldsymbol{\delta}_{kl} \quad \text{with} \quad \boldsymbol{\Gamma}_{N-1,l} = 0 \tag{32}$$

The $(N-1)$ backward time evolving $\boldsymbol{\Gamma}_{kl} \in \mathbb{R}^{n\times m}$ are the matrices of the adjoints of the representers. Their column i contains the adjoint to $\mathbf{r}_{kl}^i$. In the adjoint evolution (32), the operator $\boldsymbol{\delta}_{kl}$ is the classical Kronecker delta, equal to 1 if $k = l$ and to zero if not. Within the data time series, $l = 1, \ldots, N-1$, $\boldsymbol{\delta}_{kl}$ selects the measurements made at time t_k. The matrix $\boldsymbol{\Gamma}_{k-1,l}$ of representer adjoints, associated with the observations made at time t_l, feels its data array $\mathbf{C}_l$ at time $t_k = t_l$ only. The backward initial value problem (32) is uncoupled from the two others. It is what renders the representers method iteratively attractive when compared to (20–23) and (25–28).

Before solving (29) as a function of solutions to (30–32), it is useful to discuss an interesting property of the representers $\mathbf{r}_{kl}^i$ and their adjoints. For direct measurements

of state variables, at isolated points in time and space (in our notation, each row of $\mathbf{C}_k \in \mathbb{R}^{m\times n}$ contains one element equal to 1, all others are null), the columns of $\mathbf{\Gamma}_{kl}$ in (32) are precisely discrete Green's functions or influence functions for the adjoint operator $\mathbf{A}_k^T$ or fundamental solutions of the backward problem (27). Using the symmetry property of Green's functions of adjoint operators, they are also equal to Green's functions of the forward problem (30), in which associated paired indices are permuted (e.g., Roach, 1982). For the more general measurement model (2), one may deduce the following illustrative relations from (30–32). The matrices of representers adjoints are Green's functions projected in the measurement space or "generalized" Green's functions; in discrete terms, $\mathbf{\Gamma}_{0,k}^T = \mathbf{C}_k\mathbf{A}_{k-1}\cdots\mathbf{A}_1$, or in general, $\mathbf{\Gamma}_{k-k',k}^T = \mathbf{C}_k\mathbf{A}_{k-1}\cdots\mathbf{A}_{k-k'+1}$, for $k \geq 0$ and $k' \geq 1$. Thus, for all $\mathbf{C}_k$ in (2), one obtains

$$\mathbf{C}_k\boldsymbol{\psi}_k^f = \mathbf{\Gamma}_{0,k}^T\mathbf{A}_0\boldsymbol{\Psi}_0$$

and

$$\mathbf{C}_k\tilde{\mathbf{R}}_{kl} = \mathbf{\Gamma}_{0,k}^T\mathbf{A}_0\mathbf{P}_0\mathbf{A}_0^T\mathbf{\Gamma}_{0,l} + \sum_{k'=1}^{k}\mathbf{\Gamma}_{k-k',k}^T\mathbf{Q}_{k-k'}\mathbf{\Gamma}_{k-k',l}$$

The $(N-1)$ matrices $\mathbf{C}_k\tilde{\mathbf{R}}_{kl} \in \mathbb{R}^{m\times m}$ are the discrete equivalents of the representer matrix of Bennett (1992). From the last expression, each of these matrices is symmetric, positive definite since initial and model error covariances satisfy these properties.

It can be shown (Lermusiaux, 1997; and the Appendix) that the following expressions for the $\mathbf{b}_l$'s,

$$\mathbf{\Gamma}_{k-1,l}\mathbf{b}_l = \boldsymbol{\lambda}_{k-1} \tag{33}$$

$$\mathbf{b}_l = \mathbf{R}_l^{-1}\cdot[\mathbf{d}_l - \mathbf{C}_l\hat{\boldsymbol{\psi}}_l] \tag{34}$$

$$\underline{\mathbf{b}} = [\underline{\mathbf{R}} + \underline{\mathbf{C}\tilde{\mathbf{R}}}]^{-1}\cdot[\underline{\mathbf{d}} - \underline{\mathbf{C}\boldsymbol{\psi}^f}] \tag{35}$$

are such that equations (29–32) satisfy the system (25–28). Since the representer matrix $\underline{\mathbf{C}\tilde{\mathbf{R}}}$ was proven symmetric, positive definite and since data error covariances in $\underline{\mathbf{R}}$ are symmetric, positive definite, the matrix $[\underline{\mathbf{R}}+\underline{\mathbf{C}\tilde{\mathbf{R}}}]$ is invertible. Hence equation 35 gives the unique solution $\hat{\underline{\boldsymbol{\psi}}}_k$ for the Euler–Lagrange system (25–28). In summary, the weak constraint estimate $\hat{\boldsymbol{\psi}}_k$ minimum of J_N in (24) can be expressed as a function of the matrices of representers $\tilde{\mathbf{R}}_{kl}$ and of the forecast $\boldsymbol{\psi}_k^f$ by

$$\hat{\boldsymbol{\psi}}_k = \boldsymbol{\psi}_k^f + \tilde{\mathbf{R}}_{kl}[\underline{\mathbf{R}} + \underline{\mathbf{C}\tilde{\mathbf{R}}}]_l^{-1}\cdot[\underline{\mathbf{d}} - \underline{\mathbf{C}\boldsymbol{\psi}^f}] \tag{36}$$

Integrating (30–32) and inserting the solutions into (36), one solves the generalized inverse problem. Notice that for linear systems, the field estimates given by (36) and (17) are precisely the same. In practice, there is no need to use (36), which requires the storage of the matrices of representers for all t_k. In Bennett (1992) and Lermusiaux (1997) it is shown that the direct evaluation of (36), well suited for parallel computations, is dominated by $2m(N-1)+3$ dynamical model type of integrations.

An iterative method for evaluating (36) is also described in Bennett et al. (1996). Instead of solving (35) directly, which would require exact evaluation of the representers, it is solved iteratively. Preconditioning can be used by evaluating $[\underline{\mathbf{R}} + \underline{\mathbf{C}}\tilde{\mathbf{R}}]$ for a coarse resolution numerical grid, requiring $2m(N-1)$ coarse model integrations. Then (35) can be solved and a first estimate $\hat{\boldsymbol{\psi}}_k^1$ evaluated by insertion of (35) into (27) and integrations of (27) and (25), requiring three fine model integrations in total. For other iterations, new $\mathbf{b}_l$'s can be estimated from (34), then inserted into the adjoint model (27) to finally integrate the forward model, thus two additional fine model integrations per iteration. The total count is $2m(N-1)$ coarse and $3+2f$ fine model integrations, where f is the number of iterations. Since ocean data sets usually have much smaller dimensions than ocean model state vectors, the representer algorithm has at least a cost of order m/N smaller than a KS. However, the representer scheme as stated (29–36) does not estimate the a posteriori errors, which is a major disadvantage compared to the KS. For ocean applications, we refer to (Bennett (1992), Bennett and Thorburn (1992) and Bennett and Chua (1994).

3.4. Direct Minimization Methods

Direct minimization methods are all methods that directly minimize cost functions similar to (24), generally without using the Euler–Lagrange system (25–28). Nonlinear penalty functionals, with derivative smoothing terms, and nonlinear dynamical models are considered here. Direct methods can be modified for constrained minimization (e.g., Luenberger, 1984) and thus can limit the range of variables (e.g., nonnegative constraints for biochemical concentrations). There is a large literature available on large-scale minimization in engineering and oceanography (e.g., Gill et al., 1982; Schröter and Wunsch, 1986). Only the concepts are addressed here.

Descent Methods

Descent methods iteratively determine directions locally descending along the cost function surface. At each iteration, a line search minimization (exact or approximate) is performed along that local direction and a new descending direction is found. In the *steepest descent* method, the local direction is the opposite of the local cost function gradient. This is the simplest, but slowest method (converges linearly). The search directions of the *conjugate-gradient* method are orthogonal with respect to the local Hessian matrix, the first direction usually chosen equal to the local steepest descent. Only gradients of the cost function need to be evaluated and all conjugate-gradient methods have low storage requirements (a few gradients) for good convergence rates. For linear ocean systems and for any quadratic penalty functionals of a total of n free scalar variables, the conjugate-gradient converges exactly in at most n steps. For nonlinear models or other convex functionals, the conjugate-gradient is an iterative procedure and after n local conjugate-gradient searches, it is usually restarted. Conjugate-gradient methods differ in the algorithm used to generate the initial guess, the definition of the new search direction and the restart criteria. Well-established schemes include the Fletcher–Reeves (1964) and the Polak–Ribiere (1969) algorithms, Beale's (1972) restart and Powell's (1977) restart. *Newton and quasi-Newton* methods iteratively approximate the general penalty functional by a local quadratic form and explicitly try to minimize that local approximation. They can have quadratic rates of convergence. The Newton method requires the evaluation and storage of the local $(n \times n)$ Hessian of the cost function, which makes it very expensive to use in

ocean DA. The quasi-Newton methods only evaluate an approximation of the Hessian, generally by building up curvature information from local cost function and gradient evaluations. The storage requirements are similar to the Newton method, but the computational efficiency is better (only first derivatives evaluated). The most popular scheme are the BFGS (Broyden–Fletcher–Goldfarb–Shanno) quasi-Newton and its modifications, such as the Shanno (1978) limited-memory quasi-Newton method. To speed up the convergence, a preconditioning matrix can be applied to the cost function or its local quadratic approximation, leading to preconditioned conjugate gradient or preconditioned quasi-Newton methods (Zupanski, 1993).

For new developments of large-scale minimization problems, we refer to Navon and Legler (1987) and Li et al. (1994). The major drawback of all descent methods is that they are initialization sensitive. For nonlinear cost functions (e.g., a nonlinear coastal model), they must be restarted many times to avoid local minima. Nonlocal methods such as simulated annealing and genetic algorithms then become attractive.

Simulated Annealing

Simulated annealing is suitable for large-scale optimization problems (Press et al., 1989, Aart and Korst, 1991). The scheme is based on an analogy to the way that slowly cooling solids arrange themselves into a state of perfect crystal, with a minimum global energy. To achieve a perfect crystal, the solid is heated until it becomes an amorphous liquid, in a nearly random state. The liquid is then cooled very slowly using a specific scheme of decreasing temperatures such that thermal equilibrium is reached at each temperature. From a statistical point of view, the Boltzmann distribution (proportional to $e^{-E/kT}$) gives the probability of the solid having a certain energy E at thermal equilibrium at temperature T. During the cooling there is thus the possibility for a configuration with higher energy to be accepted. For simulating this evolution, Metropolis et al. (1953) introduced the *importance sampling* algorithm, which generates a sequence of states such that new states with higher energy ($\Delta E > 0$) are accepted with probability $e^{-\Delta E/kT}$. If the new state has lower energy ($\Delta E < 0$), this probability is greater than 1 and the system always accepts a new, lower-energy configuration. However, if $\Delta E > 0$, the Metropolis algorithm allows the system to get out of a local energy minimum. As indicated by the Boltzmann distribution, the lower the temperature, the less likely are "uphill" choices. A random number generator is needed to define the possible options available to the system. Importance sampling is a means to generate, from successive options, a sequence with a desired asymptotic distribution (Bennett, 1992).

In optimization using simulated annealing, the cost function plays the role of energy and temperature is replaced by a control parameter. An *annealing schedule* defining how and at what speed the control parameter (T) is decreased needs to be determined. For sufficiently slow schedules (too slow in practice), convergence to the global minimum is guaranteed for an infinite number of iterations. A *jump routine* is also necessary to reach through the space of possible free variables. Given a set of cost function arguments, the jump routine gives another set. Both the annealing schedule and jump routine are problem specific and need to be tuned to the problem. Kirkpatrick et al. (1983) introduced the method as a combinatorial optimization scheme. For ocean applications, we refer to the array design of Barth and Wunsch (1990) and to the fit to data of a steady-state solution of a nonlinear Q-G model (Krüger, 1993). The advantages of simulated annealing are its origin in verified physical the-

ory, which leads to convergence criteria, and its independence of the specific structure of the cost function and initial guess, allowing for a nonlocal minimum search. Its main disadvantages are its large computer requirements and, for affordable annealing schedules, an uncertain convergence to the global minimum.

Genetic Algorithms

Genetic algorithms are direct methods based on searches generated in analogy to the genetic evolution of natural organisms. Barth (1992) describes the philosophy as similar to nature's survival of the fittest. At each iteration or generation of the search for the optimum, the genetic scheme keeps a population of approximate solutions. The population is evolved by manipulations of past populations that mimic genetic transformations (i.e., breeding, offspring, etc.) such that the likelihood of producing better data-fitted generation increases for new populations. The more complex the penalty functional, the larger the population expected to be necessary. Genetic algorithms allow nonlocal minimum searches, but convergence to the global minimum is not assured. The major disadvantage is the lack of a theoretical base. In fact, almost any minimization method is a genetic one and the problem of determining a genetic algorithm physically adapted to ocean problems is still unresolved. Like simulated annealing, it has generally high computer requirements, but it has been found to be faster in some cases (Barth, 1992).

3.5. Stochastic Methods

In this section we consider methods directed toward a nonlinear stochastic dynamical model and stochastic optimal control. The ideas are new in ocean DA and further investigations are required. Once a stochastic cost function is defined, brute-force methods such as simulated annealing or genetic algorithms could be used for optimization. That can be very expensive and one could instead try to solve the conditional probability density equation (Fokker–Planck equation) associated with nonlinear versions of (1, 2). Minimum error variance, maximum likelihood or minimax estimate can, for instance, be determined from the probability density function (Boguslavskij, 1988; Kloeden and Platen, 1992; Carter, 1993). No assumptions are required, but for large-dimension problems, the cost can also be prohibitive, since the Fokker–Planck integration is generally implemented by large Monte Carlo ensemble calculations (Evensen, 1994a,b; Lermusiaux, 1997; van Leeuwen and Evensen, in press). Even for low-dimensional realistic models, parallel machines with large storage capabilities are necessary.

3.6. Hybrid Methods

Hybrid methods are different combinations of previously discussed data assimilation schemes. This is a new concept that requires further research. An example is as follows. A method based on the Kalman filter idea, but better adapted to ocean dynamical models, can be used initially to obtain a first estimate of the evolution of the state variables within the observation interval. This suboptimal first estimate can then be used as a first "good" guess in a smoothing cost function, suboptimal filtering being prerequisite to smoothing with nonlinear models. The technique called error subspace statistical estimation (ESSE), presented briefly in Section 6.2, is such an example (Lermusiaux and Robinson, in preparation). The cost function can also

be minimized by one of the methods described earlier, such as a variational inverse method or a direct nonlinear minimization method. For parameter estimation, one can always add to the smoothing cost function an additional term based on the difference between the parameters used for the first guess and the smoothed parameter values sought. Hybrid methods appear promising for oceanographic applications.

4. Coastal Ocean Data Assimilation

Ocean prediction systems, composed of an observational network, a dynamical model and a data assimilation (DA) scheme, are useful for the coastal oceans, for both scientific and practical purposes. Such systems can contribute to important societal requirements for management of multiuse coastal zones. The DA methods were reviewed in Section 3. Coastal processes, observations and models are now discussed and specific concerns for coastal ocean prediction systems addressed. A comprehensive system must account for a variety of disciplines (physics, biology, chemical and geological, and their mutual interactions), a multitude of time and space scales, and the complex geometries (land/sea boundaries, topography) of coastal and ocean regions. In this section we discuss the processes and scales of coastal oceans (4.1) and the issues and complexities particular to the coastal oceans that need to be addressed in the observational network and dynamical model components for a coastal ocean monitoring and prediction system (4.2). The final subsection (4.3) on the third component, the data assimilation schemes, relates the issues presented here to the overview of methodologies presented in Section 3.

4.1. Processes and Scales

To predict, monitor and simulate the coastal oceans, it is necessary to consider a particularly large range of phenomena and scales in both time and space. The coastal ocean, in its broadest definition, includes estuaries and the region between the shoreline and the beginning of the deep ocean or abyssal plain. Therefore, most processes that occur in the ocean in general must be studied to understand coastal oceans. Many of the physical processes are reviewed in other chapters of this volume. A comprehensive summary is presented in Table I. These include the effect of tides, both internal and external; temperature and salinity fronts caused by a variety of processes; waves with scales ranging from seconds to years and from meters to thousands of kilometers; currents and eddies arising from many processes, both internally and externally driven; boundary layers, at the surface, the bottom and near shore; surf zone phenomena, including storm surges and rip tides; sedimentary processes; and estuarine processes.

The scales on which these processes occur ranges from very small (millimeters) to very large (thousands of kilometers). Each process has a different range of scales that is important. A particular coastal ocean region will be affected by various combinations of processes (Table I). Which ones are most important can depend on, among other factors, latitude, geological particulars (of both the land and ocean), weather and climate, and the orientation and location relative to the deep-ocean general circulation. Often, a given region can be dominated by one type of forcing for a short period of time (e.g., a hurricane), and later dominated by a completely different type of phenomena (e.g., internal baroclinic instability of a tidally mixed front).

TABLE I
Physical Processes Occurring in the Coastal Oceans

Tides	
	Barotropic, internal, overtides (harmonic generation)
Fronts	
	Upwelling, shelfbreak, buoyancy, topographic, mixing, tidal, deep-ocean
Waves	
	Surface, internal-inertial, edge and infra-gravity, planetary, Frontal-trapped and shear, coastal-trapped, topographically trapped
Currents and eddies	
	Wind driven, buoyancy driven, squirts, filaments, eddies, external pressure gradients, boundary currents, and undercurrents, tidal rectification, topographic or geometrical eddies, gravity-wave related
Boundary layers	
	Mixing layer, atmospheric and river fluxes, Ekman transport, stratification, entrainment, ice interactions, surface films and microlayer, bottom shear stress, shear dispersion, topographic and roughness interactions, sediment transport
Surf zone effects	
	Sediment transport, rip currents, mean flow generation, storm surges, surf beat
Estuarine processes	
	Entrainment, layered circulations, salt wedges, tidal fronts

As an example, consider the Middle Atlantic Bight (MAB), off the east coast of the United States. The shelf is about 100 km wide, extending from Hatteras northward toward the Canadian shelf. Variability on the inner and middle shelf is dominated by winds and tides, while closer to the shelf break, deep-ocean interaction becomes more important. Tidal processes can be large scale, O(1000 km), or they can be smaller scale, O(1 km), such as internal wave breaking on the shelf forced by tidal interactions with the shelf break (Brickman and Loder, 1993). Wind events can also be large scale, generating forced and free shelf waves of O(1000 km), or smaller scale, generating strong currents locally. Due to the geometry of shelves such as the MAB, many processes have different scales in the cross-shelf versus long-shelf direction; upwelling events on the inner shelf can be long in the long-shelf direction, O(100 km), but much shorter in the cross-shelf direction, O(10 km). At the shelf break, interactions with the Gulf Stream and its warm-core rings, as well as the instability of the shelf-break front itself, generate meso- and submesoscale variability, O(10 km) (e.g., Churchill et al., 1986; Garvine et al., 1988; Gawarkiewicz, 1991; Sloan, 1996). Other scales are imposed by atmospheric fluxes and river runoff. Different time scales result from these processes. Table II shows processes that have been studied in the MAB, ordered from longest to shortest time scale. Nonlinearity also transfers energy from one scale to another.

Such complexities present a challenge for data assimilation. Different methods may be appropriate for some processes but not for others. Desirable observations need to be defined, but the best use of available observations should also be identified. A schematic of this twofold problem is shown in Fig. 20.3. One way to consider the problem is to ask: Given a set of processes that one wishes to monitor, what are the observational tools and models needed to capture those processes? Another view is

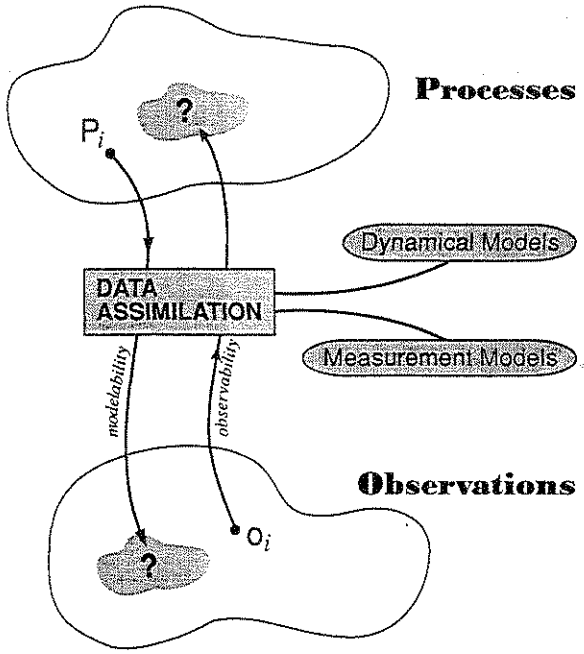
TABLE II
Some of the Operative Physical Processes on the Middle Atlantic Bight Shelf

Process	Reference	Mechanism/Notes
Interannual shelf water variability	Mountain (1991), Manning (1991)	Volume of shelf water (<34 psu) correlated to upstream inflow; salinity correlated to river discharge and precipitation.
Mean shelf circulation	Beardsley and Winant (1979)	Mean southwestward flow driven by along-shelf pressure gradient.
Seasonal stratification	Beardsley et al. (1981)	Driven by wind and buoyancy fluxes.
Cold pool	Houghton et al. (1982)	Resides in deep shelf waters between 50- and 80-m isobath; "remnant" of winter-cooled water.
Shelf-break front	Houghton et al. (1988)	Water mass front between cold fresh shelf waters and warm saline slope waters.
Estuarine forcing of inner shelf (e.g., Delaware coastal current)	Munchow and Garvine (1993)	Delaware Coastal Current 20 cm s^{-1} forced by Delaware River discharge.
Hudson shelf valley steered flow	Mayer et al. (1982)	Topography steers cross-shelf currents of up to 25 cm s^{-1}.
Atlantic coast northeaster storms	Davis et al. (1993)	Phenomenology of atmospheric synoptic patterns.
Subtidal coastal sea-level fluctuations	Noble and Butman (1979)	Strong correlation between along-shelf winds and coastal sea level in synoptic band (2- to 10-day period).
Shelf waves, Kelvin and edge waves	Ou et al. (1981), Huthnance (1986)	Shelf waves have $\sigma < f$; Kelvin and edge waves have $\sigma > f$.
Coastal shelf upwelling front	Crowley and Glenn (1994)	Along-shelf winds drive upwelling, forming surface thermal fronts.
Shelf-break exchanges	Garvine et al. (1988)	Shelf-break phenomenology includes interleaving, calving, ring entrainment and small-scale shelf-break eddies.
Gulf Stream water penetration	Gawarkiewicz et al. (1992)	Intrusion to 60 km north of Hatteras, to the 25-m isobath during summer.
Tides	Moody et al. (1983)	Dominant constituents are M_2 and K_1.
Internal waves, tides, solitons	Zheng et al. (1993), Flagg (1988), Brickman and Loder (1993)	Barotropic tidal motion of stratified fluid over sloping topography generates internal waves.
Suspended particulate matter	Glenn (1994), Palanques and Biscaye (1992)	Strong offshore sediment transport observed during winter storms due to wind, tides, waves.

to ask: Given a set of observations, what are the processes that can appropriately be monitored? The latter recognizes that only when a particular observational network is in place is it possible to determine fully which processes can be modeled successfully. In addition, historical observations might be useful for new purposes.

4.2. Coastal Ocean Monitoring and Prediction Systems: Measurements and Models

A complete coastal ocean monitoring and prediction system is made up of three components: an *observational network*, consisting of a wide variety of platforms and sen-



1) Given a vector of processes P_i, what observations are necessary and what models?

2) Given a vector of observations and model, what processes can be observed?

Fig. 20.3. Twofold adequacy of processes and observations for data assimilation. © PFJL, NQS.

sors measuring all aspects of the ocean state (observable variables); a set of *dynamical models*, which predict the ocean state (dynamical state variables) into the future; and a *data assimilation scheme*, which combines the observations with the dynamical models to accomplish the variety of goals mentioned in Section 2. It is challenging to combine all three elements and to include all the processes and scales necessary for desired purposes. Much progress has been made in recent years in the development of coastal ocean monitoring and prediction systems. Several preliminary systems have been implemented (e.g., Johannessen et al., 1993; Gerritsen et al., 1995). An operational forecast system is under development for the U.S. east coast that will include data assimilation (Aikman et al., 1996). A schematic of the Harvard Ocean Prediction System (HOPS) (Lozano et al., 1996) is shown in Fig. 20.4. HOPS is an interdisciplinary system, including physical, acoustical, biological and chemical dynamics (Fig. 20.4*a*), which is modular and flexible. HOPS assimilates a variety of data types (Fig. 20.4*b*), either directly or via structured data models including feature models and EOFs (Lozano et al., 1996). It can be set up in any region of the world ocean (deep sea, coastal and across the shelf break) and has been utilized and verified in several regions, in very simple or quite complicated configurations (Robinson, 1996; Robinson et al., 1996a,b).

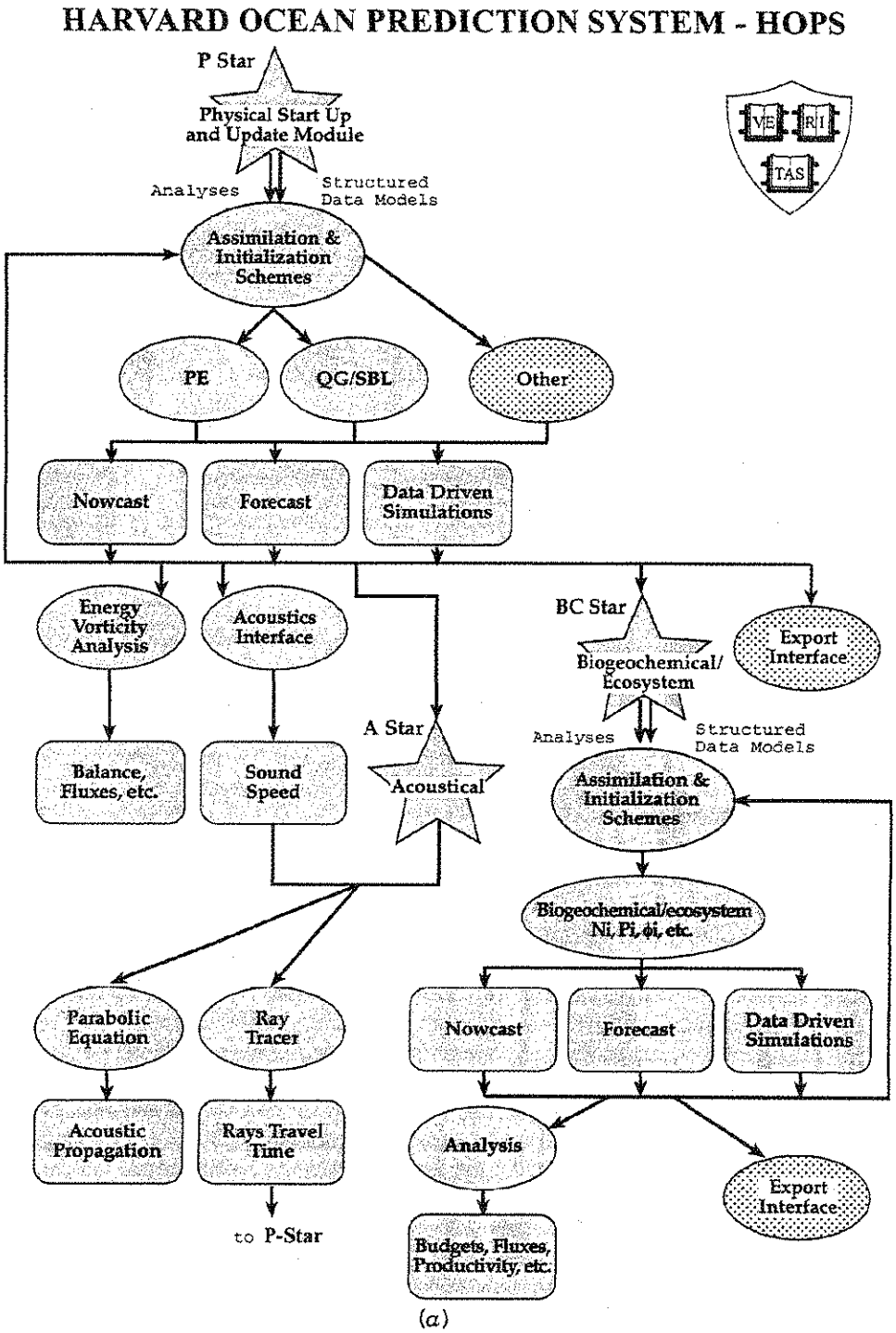
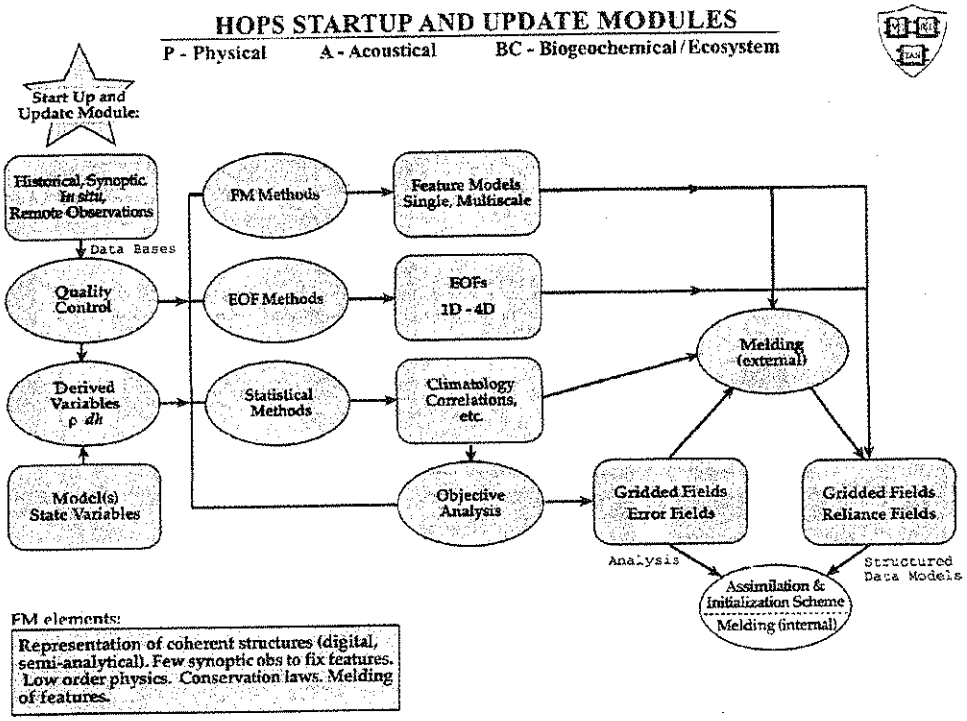


Fig. 20.4. (*a*) HOPS.



(*b*)

Fig. 20.4. (*Continued*) (*b*) startup module of HOPS.

Observational Networks

An observational network is a collection of platforms and sensors for field measurements (see Chapters 12–15). The measured quantities include physical variables (e.g., temperature, salinity, velocity), biological variables (e.g., phytoplankton, zooplankton, fish stocks), chemical constituents (e.g., oxygen, nutrients) and geological variables (bathymetry, sediments). Both internal budgets and boundary fluxes are important, so some of the measured quantities are internal variables of the system, while others are measured as fluxes from outside the coastal system: atmospheric fluxes, river fluxes, deep-ocean exchange, bottom interactions. The wind field is a particularly crucial variable for the coastal oceans. Bathymetry is a particularly important variable to measure; recent coastal ocean programs have discovered that traditional bathymetric data contains error levels that are too high for the purpose at hand. In some cases, models allow bathymetry to be adjusted as a parameter, which can improve the model results.

The sensors monitoring these variables include CTD/XBT/XCTD, which measure temperature and salinity (conductivity); acoustic devices, which can measure velocity (ADCP), integrated temperature and velocity (tomography), particles and biomass, and bathymetry; passive and active electromagnetic devices [e.g., color (SeaWiFS), infrared (AVHRR), LIDOR]. The number of oceanographic sensors is large and rapidly growing.

A variety of platforms deliver the sensors to their desired location. The plat-

forms can be manned or unmanned, remote or in situ, stationary or moving. Present-day platforms include satellites, ships, aircraft, moorings, free-floating instruments, remotely operated vehicles, autonomous underwater vehicles and fixed stations on land. Various platforms offer advantages and disadvantages. Moored instruments have the most nearly continuous measurements below the surface but cover only one data point in the horizontal. Satellite remote sensing offers the advantage of fine resolution and large coverage in space and time, but only for near-surface fields or surface height. Ships and aircraft can provide high resolution in all three spatial dimensions, but often only at discrete stations which are not simultaneous. Most platforms are affected by other considerations, such as weather, sea state, and interaction with local fishing activity. A cost–benefit analysis of the optimal sampling system is very difficult. In general, the mix of platforms and sensors chosen should attempt to minimize the resources required for a given purpose. In GLOBEC (1994), methods of sampling are reviewed in the context of physical and biological interactions. Calibration and interpretation, especially for novel systems, remains an important issue, and management and archiving of data are important concerns presently being addressed. Observing systems are generally divided into operational systems, for management and monitoring purposes, and research systems, which are used to further the study of the basic ocean science, which have somewhat different requirements. The optimal, efficient distribution of sensors is a research topic, which can be addressed through the use of observational system simulation experiments. OSSEs can help design the components of a prediction/observation system, optimize the use of resources, and improve and validate the system performance (GLOBEC, 1994).

An example of a multicomponent research network is the one maintained by Rutgers University (Fig. 20.5). It includes a remote sensing receiver, a meteorological tower, a research vessel and a long-term ecosystem observatory (LEO) (Grassle et al., 1996). The remote sensing receiver has access to AVHRR and will access SeaWiFS ocean color when it becomes available. The meteorological tower records air temperature, pressure, humidity, wind and shortwave solar radiation. The ship maintains a CTD, transmissometer, fluorometer, surface and underwater radiometer. The surface platform (the ship) can be used to calibrate and validate the satellite platform. The LEO is located 10 km offshore at 15 m depth. It presently has current meters and optical backscatter sensors, and benthic acoustic stress sensors. Plans include remotely operated vehicles or autonomous underwater vehicles to measure pressure, temperature conductivity, optical backscatter, fluorometer, radiometer, hydrophonics and ADCP current measurements.

Dynamical Models

A wide variety of dynamical models have been employed to tackle the problem of estimating fields in the coastal ocean. Due to the variety of processes and scales mentioned above, different models have been developed for different purposes. The models vary according to the different physical assumptions (for resolved scales), subgrid-scale parameterization for unresolved scales, computational methods, boundary conditions and geometry. One can divide the models into general models such as primitive equation or shallow water, which attempt to resolve a wide range of processes, and special-purpose models, such as tidal and storm surge models, and surface and bottom boundary layer models.

Complex, fully three-dimensional models have been developed and refined

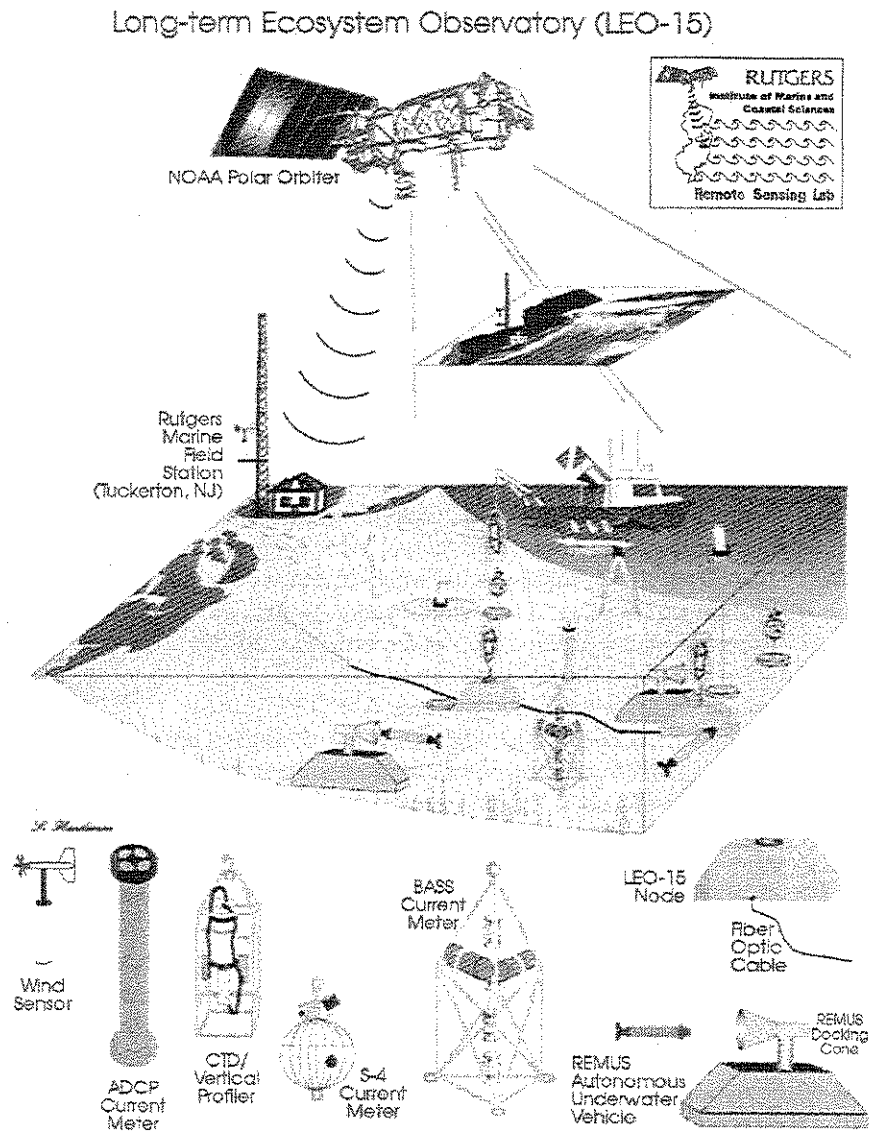


Fig. 20.5. Rutgers multicomponent research network. (Drawing by L. Henderson; from Grassle et al., 1996.)

recently which can handle many processes in a single model. Some, like primitive equation models, resemble deep-ocean models but in addition must handle the complexities of the coastal ocean: coastal geometry, different resolution requirements for different regions, steep and variable topography. Haidvogel and Beckmann (see Chapter 17) have reviewed the present status of many of these models.

A recent review of the problem of validation of coastal ocean models (Lynch and Davies, 1995) demonstrates the large number of models that are presently in use for

operational and scientific use and elucidates verification issues. Test problems have been developed to validate and compare models to models and models to data (for tidal problems, Davies and Xing, 1995; for convection and diffusion problems, Baptista et al., 1995; for steep topography, Chapter 17, this volume). These benchmarks provide an important starting point, but more work is required. Validation, calibration and verification are essential for data assimilation models and systems.

Geometrical complexities manifest themselves primarily in two ways. The complex geometry of coastlines has often motivated the use of curvilinear coordinates. A research topic of interest today is the generation of grids and the accuracy of different meshes (Carey, 1995). Variable terrain (both steep and high topography) has led to the development of different techniques for vertical coordinates, such as terrain-following coordinates (sigma) (e.g., Gerdes, 1993). One problem that has received much attention with these types of grids are truncation errors that different vertical coordinate systems introduce (e.g., Haney, 1991; McCalpin, 1994).

Open boundary conditions continue to present vexing problems for coastal ocean modelers. The deep ocean influences shallow seas at many time and space scales (waves, tides, boundary currents, etc.). Equally important, signals generated in the coastal domain must be able to propagate outward. For example, Johnsen and Lynch (1995) apply a second-order radiative open boundary condition for a shallow-water model forced by both winds and tides, with useful results. In general, the efficacy of different open boundary conditions depends on the particular problem (Chapman, 1985).

Sea-surface height has been one of the principal concerns for many coastal populations. Among the most advanced coastal ocean dynamical models are those that predict tides and storm surges. The North Sea–Baltic warning system provides operational forecasts based on the shallow-water equations. The model includes triply nested domains with a 3 : 1 ratio. Calibration, verification and validation techniques have been applied, such as tuning the bed friction, horizontal eddy viscosity and bathymetry. Model results compare well with data and provide useful forecasts (Lynch and Davies, 1995).

The shallow shelves and seas are often dominated by atmospheric fluxes, wind, heat and fresh water. Therefore, many attempts have focused on the vertical mixing component, and special bottom boundary layer and surface boundary layers have been employed. Turbulence closure schemes have been studied extensively and form an important part of any modeling system (Davies et al., 1995; Davies and Xing, 1995).

As mentioned above, processes occur at many scales, especially in the coastal oceans, but computer power limits resolution of the explicitly resolved dynamical model. Therefore, it becomes necessary to choose which models will be used and at what resolution. Variable resolution grids and nested grids are employed to achieve higher resolution in particular regions. For example, by using nested grids, a large domain can provide boundary conditions for a domain with higher grid resolution. Different dynamical assumptions can even be used in the nested models (Sloan, 1996). Drawing from the meteorological experience, nested ocean models have been applied successfully (Spall and Holland, 1991; Oey and Chen, 1992). Recent progress has been made in applying similar nested models in the coastal oceans. Two-way communications between the grids is often essential, as it allows information to propagate both into and out of the domain. A telescope of relocatable nested grids can

allow efficient, small coastal and estuarine models to be employed while still benefiting from an accurate representation of the larger-scale processes. The complete set of models (e.g., wave, tidal, primitive equation, biological chemical, geological) forms a hierarchy of models that can be employed. For the future, a more general system of coupled, nested models with physics, biology, chemistry and geological processes is foreseen.

4.3. Coastal Ocean Monitoring and Prediction Systems: Assimilation Schemes

The monitoring of coastal sea level has been the dominant component of the special-purpose assimilation studies. Within the last decade, such simulations have been successfully achieved, generally with parameter estimation. Relatively complex two- or three-dimensional models but simplified DA methods (adjoint method with diagonal weights, constant-gain KF, OI), or, simplified, usually one-dimensional dynamical models, but advanced methods (generalized inverse, EKF) have lead to significant improvements (Sections 5.1 and 5.2). For complete coastal circulation studies, it is only very recently that DA has been used.

The lack of sufficient coastal data coverage and the complex coastal dynamics advises against the simplest methods, like the classical versions of the direct insertion, blending and nudging methods (Section 3.2), which assume a one-to-one measurement model and a diagonal gain matrix. Those methods rely on the internal dynamics to propagate the local data impact, which can thus be transferred too slowly to the rest of the state vector. However, for an increased time/space data coverage and reliable observations, these techniques have a good cost–benefit ratio and are simple to implement.

Successive corrections and OI assimilation (Section 3.2) are better as long as their OI gain is based on a careful design of the forecast and data error covariances. Most of today's OI schemes have isotropic and homogeneous covariances, but other schemes (e.g., KF) have shown that this assumption is not generally correct (Cohn and Parrish, 1991; Daley, 1992a–c; Todling and Ghil, 1994). Since coastal variability is usually nonhomogeneous and nonisotropic, and since the forecast errors should be mainly in that variability, one expects nonhomogeneous and anisotropic forecast error covariances. The application of KF ideas with three-dimensional realistic coastal circulation models is still very expensive today. Parallel computers with gigabyte memories are the only hope for the strict KF in realistic modeling. Less costly, suboptimal KF versions have been investigated (e.g., Todling and Cohn, 1994; Evensen, 1994a; Fukumori and Malanotte-Rizzoli, 1995), some of which might be appropriate for coastal flows. In principle, the KF is valid only for linear models, but its extended or linearized version could be used. However, both extensions are based on a second-order truncation in the series of error moments, which precludes nonlinear saturation of the errors and can produce unbounded forecast errors (Evensen, 1993). Some nonlinear filters also exist, but they appear too expensive for most ocean models (Daum, 1994). Finally, in nonlinear dynamics, the ocean state probability density can have multiple maxima, and least-squares criteria for model-data melding are only appropriate locally; even the assumption of unbiased estimate might have to be reconsidered.

The variational smoothing methods based on control theory (Section 3.3) have the cost advantage that the *a posteriori* errors do not need to be computed to get the final estimate, even though some error information should be computed for con-

fidence assessment. Coastal flow nonlinearities are also a problem in all of today's smoothing methods, which were originally derived for linear systems (e.g., representer method). All concerns mentioned for today's filtering methods apply for control theory–based methods. The cost functions have multiple minima and convergence is not guaranteed (e.g., Tziperman, 1992a,b). In low data coverage, the global minimum might not even be physical, and it can be hard to determine the proper minimum. Global search algorithm such as simulated annealing and genetic algorithms could be used (Section 3.4), but they are very expensive and require parallel computers for realistic studies. Hybrid methods (Section 3.6), taking the advantages of both estimation and control theory without the disadvantages, might thus be the only choice for high-dimensional, nonlinear coastal flows, with a low data coverage. For parameter estimation, simple smoothing techniques based on the adjoint method have been used more than schemes based on estimation theory. First, stochastic modeling is still in its infancy, and second, with filtering techniques, past data do not improve previous parameter estimates. Finally, specific issues relevant to coastal DA are the open boundaries and geometries (Bennett and McIntosh, 1982; Miller, 1986; Shulman and Lewis, 1994; Bennett and Chua, 1994; Zou et al., 1995; Evensen, 1993, 1994a,b; Evensen and van Leeuwen, 1996); the convection and internal processes (e.g., Miller et al., 1994b), the nonlinearities (e.g., Miller et al., 1994a; Evensen, 1994b; Lermusiaux, 1997) and the best use of observations (e.g., Jiang and Ghil, 1993; Malanotte-Rizzoli and Young, 1992, 1995).

For all oceanographic studies, a good DA system should report a minimal information on the error evolution of the ocean state estimate. An error estimate should thus be combined with the description of the direct insertion, blending, nudging and adjoint method given in Section 3.2. Second, one should study the sensitivity of the estimate to the a priori melding assumptions and parameters (weights); the less sensitive the estimate, the better the scheme. Finally, the numerical complexity and cost, as well as the cost dependence with the state vector's size, are issues to consider in choosing the ideal DA scheme.

5. Progress to Date

It is only during the last decade that, for scientific as well as management and operational purposes, progress has been made in data-driven modeling and prediction of all energetically important coastal processes at once, with realistic coastal geometries, shelf-break topographies, shelf–deep sea interactions and adequate parametrization of subgrid-scale coastal physics. Most earlier realistic estuarian and coastal models were centered on specific processes such as the prediction of the sea-surface height, tides and storm surges. Thus many of the early coastal assimilation studies were naturally targeted at those specific processes and scales (Sections 5.1 and 5.2). Today, for coastal regions that have many energetically important processes, the development of a comprehensive assimilation approach is an important research subject. This includes, importantly, the determination of data impacts, and matched data and model sets which are adequately accurate for their specific purposes, but also efficient enough to be run routinely.

In the remainder of this section we present some selected examples of progress to data for tides, circulation, parameter estimation and real-time forecasting.

5.1. *Early Examples*

The earliest important example of coastal assimilation is the modeling of tides in open coastal straits or bay (Bennett and McIntosh, 1982; McIntosh and Bennett, 1984). The authors argued (1982) and confirmed (1984) that for the Bass Strait, the classical integration of a linear barotropic shallow water wave model of the M_2 tide could not be in agreement with tide gauge and current meter data available at interior stations if the model were driven only at the open boundaries by local sea level data. A weighted least-squares variational formalism was used to obtain a solution satisfying, within the respective error bounds, the dynamics, the coastal and open boundary conditions, and *all* the data. The results thus verified the consistency of the linearized shallow-water theory with all available measurements. Budgell (1986) combined simulated tide gauge data with a one-dimensional, *nonlinear*, shallow-water model for branched channels. The filtering techniques used were an EKF, or for weakly nonlinear cases, an incremental covariance simplification of the EKF. Nonlinear modeling improved the representation of coastal processes, including momentum dissipation, interaction of storm surges with tides, and shoaling of long waves in shallow water. Heemink and Kloosterhuis (1990) also developed simplified KF techniques to assimilate sea level data in a nonlinear two-dimensional shallow-water tidal model of the North Sea shelf. The model noise was expected to be large scale and was approximated on a coarse grid. To reduce the computational burden further, model nonlinearities were assumed to be weak and a constant-gain EKF was used. In an example of assimilation via initialization, Walstad et al. (1991) analyzed dynamical processes occurring in late spring of 1987 in the coastal transition zone (CTZ) inshore edge of the Northern California Current system. The authors used a regional baroclinic quasi-geostrophic model, initialized by objectively analyzed hydrographic and ADCP data and forced at the boundaries by a linear interpolation between the initial and final data sets. Barotropic instabilities and eddy-jet interactions in the CTZ were identified and supported the characterization of the CTZ flow as a meandering jet that gradually propagates offshore. This example confirms that when combined with a trial-and-error parameter estimation, even a very simple melding of data with realistic dynamics both improves the description of the field and allows for a realistic dynamical analysis.

5.2. *Parameter Estimation*

The approximate dynamical equations and forcing functions for governing the scale-restricted state variables of the coastal ocean (Sections 1.1 and 4.2) contain many parameters, the values of which are not given directly by fundamental dynamical considerations or are uncertain (eddy and drag coefficients, deformation radii, initial and boundary conditions, biological and chemical rates, etc.). Coastal parameter tuning by trial-and-error comparisons with data is both cumbersome and suboptimal for complex coastal models. Parameter estimation via data assimilation is thus important for calibration and verification of coastal models and for accurate coastal simulations and predictions.

Parameter estimation has always been of great interest to engineers. The related literature (Ljung and Söderström, 1987) and meteorological examples (e.g., Navon and Legler, 1987) provide valuable guidance. Some applications in general oceanography are (Thacker and Long, 1988; Yu and O'Brien, 1991; Smedstad and O'Brien, 1991; Tziperman, et al., 1992a,b; Egbert et al., 1994; Bogden et al., 1995; Lawson, et al.,

1995). Coastal parameter estimation has received much attention and some studies are discussed here. In most cases, the computer requirements preclude the estimation of parameters as field functions of (x, y, z, t), since it leads to the estimation of a value per grid point and per parameter. To reduce the number of degrees of freedom, prechosen profiles (linear, quadratic) in the vertical or finite-element representations have been examined. Unless important to the application considered, those technical issues are not discussed here.

To test the calibration of a one-dimensional shallow-water model of the eastern Scheldt Estuary, ten Brummelhuis (1990) estimated the friction coefficients from simulated data. A simple stochastic evolution equation for the parameters was added to the equations and an EKF was used to estimate the water levels, velocities and friction coefficients. The technique was very useful for one-dimensional tidal prediction, but in the two-dimensional study, for computer limitations, the estimated state variables were restricted to the boundary values and friction coefficients.

The assumption of exact dynamics also reduces the number of free variables (Section 3.3). With this approach, Das and Lardner (1991) estimated bottom friction coefficients and water depths in a one-dimensional shallow-water channel model. The adjoint equation was used to construct the gradient of the functional to be minimized. The authors compared different minimization algorithms and all estimated values were in good agreement with the exact ones. For piecewise linear parameter profiles, the authors empirically concluded that for a reliable smooth estimation, the number of data stations had to be at least half of the number of parameter nodal values. Ten Brummelhuis et al. (1993) also applied the adjoint method with parameter estimation to a shallow-sea model of the entire European continental shelf. Along with the initial and open boundary conditions, estimations of the space-varying bottom friction coefficient and of the parameterization of the wind stress coefficient as a function of the surface wind speed were made. Considering the open boundary conditions as uncertain variables greatly improved the robustness of the estimation and produced realistic estimates. Lardner (1993) also addressed the estimation of open boundary conditions for such a tidal model. Panchang and Richardson (1993) and Lardner and Das (1994) used an adjoint method to estimate the vertical eddy viscosity profiles of a two-and-a-half-dimensional linear coastal circulation model (three-dimensional PE equations split into the standard two-dimensional shallow-water model and Ekman equations for the vertically varying part of the horizontal velocities) from simulated current data. Since the vertical eddy viscosities appear only in the vertically varying equations, the adjoint of the barotropic mode is not needed and the computer requirements are greatly reduced. To reduce the number of data necessary for a reliable, smooth parameter estimation, a smoothing term for the parameters was added to the cost function. This restricted the possible parameter profiles, and the resulting eddy coefficients were good for both a wind-driven and a tidally driven test problem. Lardner and Song (1995, in preparation) used the same model, but they jointly estimated all parameters to which currents are sensitive: the eddy viscosities, water depth, and wind-drag and bottom friction coefficients. The barotropic mode then enters the minimization, and computer requirements are increased. For a 16×16 horizontal domain, one or two current meter stations, with two or more vertical observations, and one or two tide gauges were sufficient for a good estimation. They found no evidence that forecasts with the estimated parameters were better near data points than at other points. The adjoint method was also used by Song and Lardner (submitted) to estimate the

eddy viscosity, wind drag and bottom friction coefficients of two-dimensional alongshore averaged and linearized PE equations from model-simulated current meter and tide gauge data. The method was tested on a typical across-shelf coastal topography, using generalized terrain-following coordinates (S-coordinate). To reduce the effective number of parameters, uniform, linear and quadratic across-shelf finite-element expansions were tested. For their idealized problem, the authors concluded that the number of data stations needed to be at least equal to the number of parameter nodal values, which is a stronger criterion than the one obtained by Das and Lardner (1991).

It appears evident that parameter estimation is very important for the understanding and modeling of coastal processes, as well as for model validation, calibration and verification. However, the development of efficient techniques for real-time coastal ocean parameter estimation is an ongoing research topic.

5.3. Real-Time Prediction

Real-time forecasting with data assimilation for the coastal and shelf oceans, for a variety of scientific and practical purposes, is feasible and has been initiated. The coastal ocean prediction systems (OPSs) required for many applications are quite complex to construct and can require large computational resources for real-time usefulness. Thus at the present early stage empirical and heuristic assumptions and some modeling compromises are common. Robinson et al. (1996b) overview the general problem of the development and verification of regional systems. We cite four examples as an introduction to this topic: the Scotian shelf off Canada, the Iceland Faeroe Islands Front (IFF), the Seto Inland Sea of Japan and the coast of Norway.

Some operational applications are most effectively carried out at sea on board a data-gathering vessel. Bowen et al. (1995) used a relatively sophisticated adjoint scheme aboard the MV *Petrel V* on the outer Scotian shelf in November 1992 for nowcasts and short forecasts of the circulation and related dispersion. The purpose was to track a cohort of cod larvae for a 3-week period in order to avoid advective bias in the sampling of the population. The observational network consisted of surface drifters, telemetering moorings, hydrographic and ADCP measurements, and towed nets and plankton counters. Computational efficiency was achieved by simplifying the dynamics via linearization and simplified decomposition into wind-drift, geostrophic and tidal components. The success of this real-time exercise in predicting the location of the larvae, including after a storm event, provides an important demonstration of the power of this approach. Robinson et al. (1996) carried out real-time operational forecasts of the IFF system in August 1993 aboard the NATO RV *Alliance* using the primitive equation dynamics and optimal interpolation scheme of HOPS (Section 4.2). The purpose was to verify quantitatively the skill of a regional forecast system for accurate and efficient prediction of fronts and eddies in general and with application for acoustic propagation (Carman and Robinson, 1994). There was significant success in predicting the development over a few days of a highly nonlinear deep sock meander from a straight current.

Two attempts at practical management modeling are provided by Yanagi et al. (1995) for the Seto Sea and by Johannessen et al. (1993) for the Norwegian coast. These are simulation and process studies directed toward the development of coastal

OPSs. The purpose of the former is to predict red tide formations in order to protect agriculture fish stocks, and of the latter is to develop a general monitoring and prediction system for algal blooms, water quality and oil spills. The Seto Sea study constructed and utilized an end-to-end system with coupled models and schemes for hydrodynamics, thermodynamics, biological dynamics and tracers. An engineering approach was adopted with many compromises and a successive correction method for data assimilation. Care must be exercised in the interpretation of such an ad hoc and complex system, and some sensitivity studies were carried out for this purpose. The Norwegian study presents a useful conceptual analysis of some detailed considerations necessary for the construction of a general OPS for coastal management, with examples of hindcasts and assimilations.

6. Middle Atlantic Bight Studies

The Middle Atlantic Bight (MAB) shelf break marks a dramatic change, not only in water depth but also in the dynamics of the waters that lie on either side. Shoreward of the shelf break, the shelf-water variability is dominated by wind forcing. Seaward of the shelf break, wind effects are much less and mesoscale variability is often dominated by Gulf Stream rings. Eddies are frequently observed in the shelf break, and their dynamics are not well understood.

In this section we summarize some recent results with the Harvard Ocean Prediction System for the MAB to illustrate the capability of data-driven simulations to provide realistic field estimates in the presence of steep topography, stratification, multiscales and nonlinearities. Two-way nesting is illustrated and optimal interpolation is intercompared with a new quasi-optimal assimilation scheme.

6.1. Dynamics at the Shelfbreak

Data-driven experiments have been performed to study shelf-break dynamics, with a focus on meso- and submeso-scale variability driven by internal instability as well as Gulf Stream ring forcing. The numerical experiments are performed in a 500×400 km domain south of New England. An ocean prediction system is employed, which includes a primitive equation model, data assimilation, grid nesting, and initialization and update methods (feature models, data fusion, objective analysis; Lozano et al., 1996). The experiment assimilates three types of data that were gathered in 1984: satellite IR to identify the size, location and strength of a Gulf Stream ring; hydrographic coastal survey data (MARMAP); and CTD data collected as part of the shelf-break eddies experiment (Garvine et al., 1988). The purpose of the experiment is to use the models, climatology and synoptic data to generate four-dimensional data sets to study the physical processes at the shelf break. Intercomparative studies are used to examine the influence of various factors, such as proximity of rings and shelf-break jets strength and structure.

The HOPS PE model is initialized from a combination of climatology and feature models. The feature models are tuned to reproduce dynamical behavior (wave growth, eddy generation) that is consistent with historical synoptic observations. This allows the model to be initialized without spin-up from rest. Data are then assimilated into the model when and where it is available. The model uses topography following sigma coordinates in the vertical. To compute with high accuracy over steep

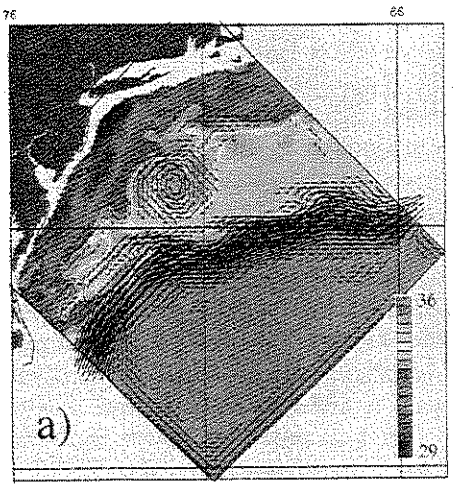
topography, the pressure gradient is computed on geopotential surfaces. To remove computational noise, without inordinate amount of mixing on the sigma surfaces, a filter is applied so as to remove only high-wavenumber signals (Lozano et al., 1994). To increase efficiency of computations, two-way grid nesting is employed. The larger domain, with coarse 5-km resolution is 500 × 400 km. The smaller 1.66-km grid is 167 km by 133 km and is contained within the coarse grid. The coarse grid provides the boundary conditions for the fine grid, while the fine grid averaged quantities are passed back to the coarse grid. Results after 15 days' integration from the MAB shelf to beyond the Gulf Stream are shown on Fig. 20.6*a,b*. Surface MAB temperature has evolved due to both ring interactions and instability of the front itself. Figure 20.6*c,d* shows another data-driven simulation in the shelf-break front region, without the Gulf Stream but with a nested grid. Future experiments will be performed with data sets that have synoptic high-resolution survey at different times, which will allow validation and verification of the prediction system, similar to the studies of the Iceland Faeroe Islands Front (Robinson et al., 1996a).

6.2. *ESSE: A Hybrid Approach*

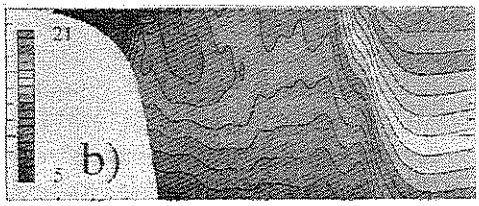
The most important defect of data assimilation via OI is its empirical error model (Section 3.2). Even for an advanced OI accounting for error growth, the error covariances are generally constructed from isotropic and homogeneous covariance functions, with parameters constant in time and independent of the evolution of the variability. On the other hand, for realistic, complex estimations with sufficient data, the KF, KS and generalized inverse (Sections 3.2 and 3.3) are still very expensive.

A rational approach was used to identify a new efficient statistical estimation (SE) scheme for DA in realistic nonlinear ocean models. The criterion used is based on an optimal reduction of the dimension of multivariate error covariances and has lead to the notion of evolving error subspaces (ESs), characterized by singular error vectors and values or, in other words, by error EOFs and coefficients. The ideal ES spans and tracks the three-dimensional scales and processes where the dominant, most energetic errors occur. The resulting general concept of *error subspace statistical estimation* (ESSE), along with its specific objectives and its mathematical details, are discussed by Lermusiaux (1997). Applications of ESSE are efficient three-dimensional objective analyses, real-time filtering assimilations and data-driven smoothing simulations. Importantly, the three-dimensional multivariate ESSE melding occurs in the ES and is thus much less costly than a full error covariance update. The time propagation of the ES is based on an ensemble forecast that uses the full nonlinear model. The members of the ensemble are chosen to optimally sample the current dominant three-dimensional multivariate ES. The implementation of ESSE, with scalable parallel computing for the ensemble forecast, was made compatible with HOPS and peripherals, aiming for a realistic, portable coastal and nonlinear ocean estimation system.

Identical twin experiments were designed for comparing the retrieval of the true ocean via OI and via ESSE, in the ideal exact dynamical model conditions. The experiment discussed here is an idealized MAB shelf-break-front simulation in summer conditions (Sloan, 1996). The domain extension is 100 km meridionally by 112.5 km zonally, with the depth varying from approximately 120 m on the shelf to 230 m in the deepest southern part, leading to 107,502 state variables. The main direction of the flow is east–west. A PE run of 39 days is defined as the true ocean evolution. Subsampled tracer



Panel (a): Salinity at 1 m from day 15 of a data-driven simulation of the Northwest Atlantic from the Middle Atlantic Bight continental shelf to beyond the Gulf Stream. A streamer of shelf water is extracted by a warm core ring; shelfbreak eddies exchange shelf and slope water; shelf water is entrained into the Gulf Stream near Cape Hatteras. Data-streams for initialization and assimilation include MARMAP/NMFS shelf hydrography, satellite sea-surface temperature, and feature models for the Gulf Stream, Warm Core Ring, and Shelfbreak Front.



Panel (b): Cross section of temperature at day 15 of the same simulation as (a). From left to right, the cross-section shows the shelfbreak front, the leading edge of a Warm Core Ring, and the Gulf Stream.

Panels (c) and (d): Near surface temperature from a nested-grid, data-driven simulation of shelfbreak eddies and ring interactions. The coarse grid (c) is 7.5 km, while the fine grid (d) is 2.5 km. Data-streams for initialization and assimilation into the model include MARMAP/NMFS hydrography, a feature model Warm Core Ring and a feature model for the shelfbreak front. Events include shelfbreak eddies, in which shelf water is drawn off the shelfbreak and slope water is pushed up onto the shelf; and a shelf streamer interacting with a warm core ring.

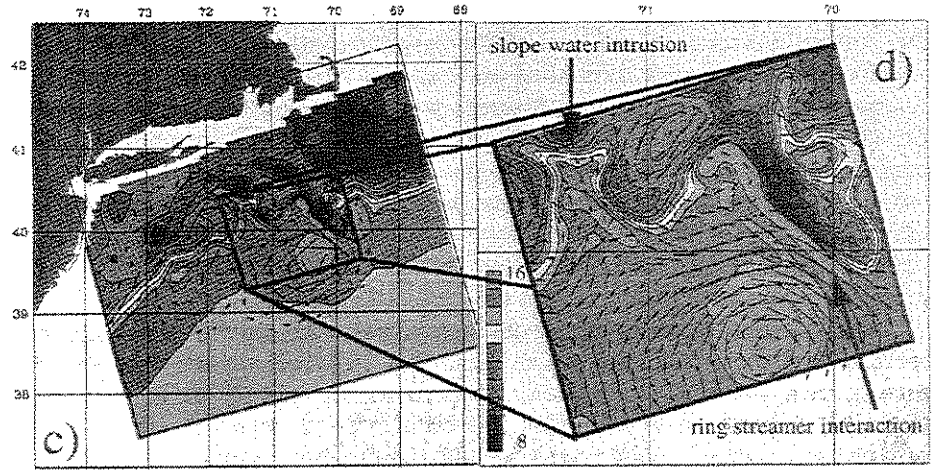


Fig. 20.6. MAB data-driven simulations.

data are extracted from that run, every 3 days, starting at day 18 up to day 36, with a low 10-km horizontal resolution. The initial conditions (ICs) of the false ocean are created by adding to the true ocean ICs both geostrophically balanced and white-noise perturbations. The OI and ESSE runs start from these same ICs; they only differ starting at

day 18, when the first batch of hydrographic data are assimilated. In ESSE, a total of 16 Sparc 10 Sun stations were used in parallel to propagate the ES.

As shown on Fig. 20.7, the ESSE improves the quality of the retrieval of the true ocean. Root-mean-square error and pattern correlation coefficient (Robinson et al., 1996a) analysis confirm qualitative arguments. The general gross features of the field have also been improved by OI. However, some of the physical characteristics and parameter ranges (frontal width, front position, eddy orientations, frontal wave packet variability) have been modified by the OI scheme. The obvious nonhomogeneous, nonisotropic and nonuniform properties (shape, locations and scales) of the frontal wave packets and cold eddies in the surface stratification are not considered. Everywhere in space and time, the same ellipsoidal error length scales are applied. The OI has difficulties coping with fields that have many scales that are not uniformly distributed. On the other hand, the forecast variability on panel (*a*) has statistically correct physical properties, but its sample path is different from the true one (*b*). Finally, the ESSE scheme (*d*) has kept the statistical properties of the true tracer and flow fields variability while correcting most of the erroneous components of the forecast. Another obvious advantage of ESSE is the singular decomposition of the multivariate error covariance, which facilitates the physical understanding and study of the dominant variabilities and uncertainties. We refer to Lermusiaux (1997) for a discussion on the ES evolution.

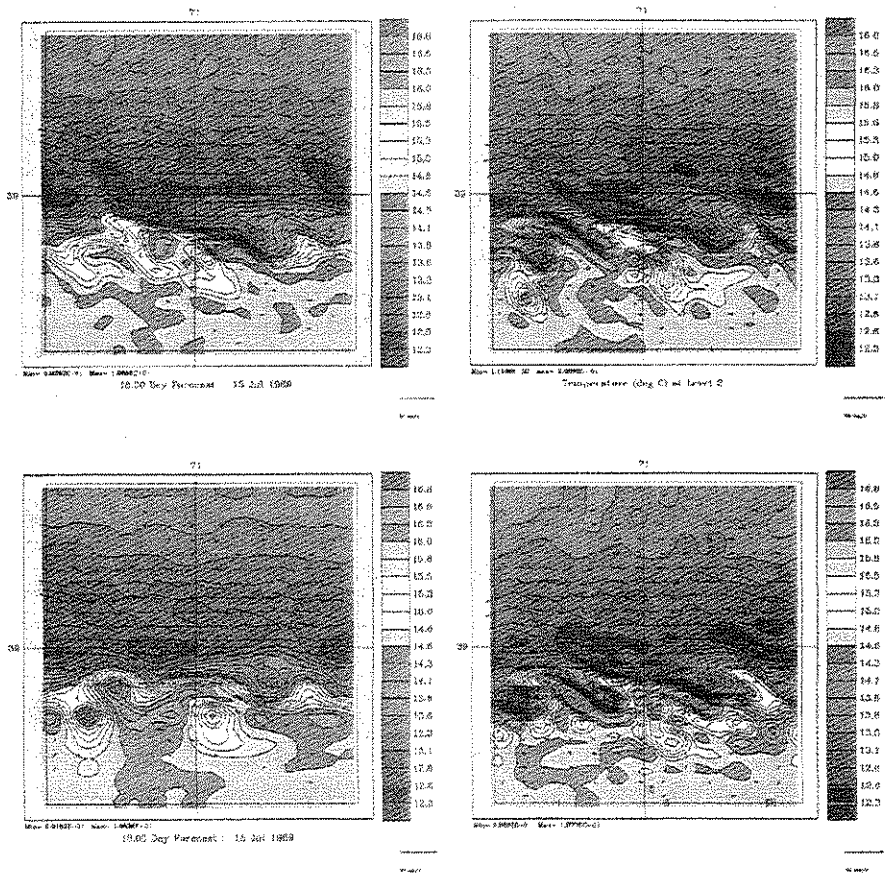
7. Conclusions

In this chapter we have presented the conceptual basis of field estimation via the melding of data and dynamics (i.e., data assimilation) and its several goals and applications. The mathematical basis, drawn primarily from estimation and control theories and rooted in error analyses for both the dynamics and the measurements, has been sketched in a uniform mathematical notation and with follow-up references. A field estimate via data assimilation both agrees with the observations within observational error bounds and satisfies the dynamical model within model error bounds. The concept of an ocean prediction system composed of an observational network, dynamical model set and data assimilation scheme has been introduced in general, and the special considerations for coastal ocean systems discussed. The progress and status of data assimilation for the coastal ocean, as well as opportunities and directions for research, have been illustrated by a few selected examples.

Data assimilation and ocean prediction systems today provide novel and feasible methodologies for fundamental ocean science and applied marine operations and management. A range of schemes and systems, from simple to complex, are now available. Application opportunities abound over a rich spectrum of needs. The use of data assimilation methodologies should significantly accelerate the advancement of coastal ocean science and technology over the next decade.

Acknowledgments

We are grateful to Dr. Carlos J. Lozano for critical comments and thank Ms. Marsha A. Glass and Mrs. Renate D'Arcangelo for preparation of the manuscript. This review was supported in part by the Office of Naval Research under Grants N00014-95-1-0371, N00014-93-1-0784 (Assert) and N00014-90-J-1612 to Harvard University.



Panel (a) shows the false ocean forecast at day 18, panel (b) the true ocean from which the day 18 hydrographic data are subsampled, panel (c) the result of the OI assimilation and panel (d) the result of the ESSE.

The ESSE (d) unstable intrusions of warm slope water into the cold shelf water have the proper (b) length scales, orientations, locations and magnitudes. The same remark applies to the shelf intrusions into the slope waters. The center field, cold surface baroclinic submesoscale eddy has been reinforced, relocated and reshaped by ESSE as it ought to be. The absolute position and tightness of the ESSE front (d) are appropriate: it is very narrow in the center of the domain and on the upstream side of warm slope water intrusions, while it has broader scales downstream of all warm intrusions into the shelf water.

The OI scheme (c) tends to create spatially uniform, ellipsoidal-like eddy structures, as it was designed to do. For instance, the westernmost cold intrusion into the slope water (b) has been almost estimated as a cold eddy by OI (c). Note that the southwest cold eddy present in the true ocean (b) is outside of the subsampled data domain and is thus not in the OI (c) nor in the ESSE estimate (d). The length scales of the OI wave packets are in agreement with the true ocean, but the OI front is wider than the true front and its position is not as accurate as the ESSE one.

Fig. 20.7. ESSE/OI MAB assimilation comparisons.

Appendix: DA Methods under Common Generic Assumptions

In ocean modeling, the discrete dynamical model's state variables are usually the *nodes* or *grid point values* of the continuous state variables. Here the dynamical state vector at time t_k is denoted $\boldsymbol{\psi}_k \in \mathbb{R}^n$, $k = 0, \ldots, N$. The number m of observations at time t_k could depend on k, but for simplicity we assume m time-invariant. All data available at time t_k are contained in the vector $\mathbf{d}_k \in \mathbb{R}^m$. For ease of notation, it is assumed that $\mathbf{d}_0 = \mathbf{d}_N = \mathbf{0}$ (no data collected at initial and final times). The time t_k is the time at which measurements k are collected. The interval $\Delta t_k = t_k - t_{k-1}$ is the time lag between the batch of synoptic measurements k and $(k-1)$ and should not be mistaken with the dynamical model time step δt. To simplify notations and ease comparisons of DA schemes, the model evolution during Δt_k is assumed linear and the known (deterministic) forcing term has been omitted since it only adds some complexity. In fact, external forcings could be assumed part of the dynamical state vector $\boldsymbol{\psi}_k$ since they may evolve with time and feedbacks between external and internal forcings exist.

A.1. Generic Assumptions: Statistical True Ocean ($\boldsymbol{\psi}_k^t$) and Measurement Models

In assimilation methods one usually defines, explicitly (estimation theory) or implicitly (control theory), the statistics of a "true" ocean dynamical system and a "true" measurement model. These models are here restrained by the following limitations. The sample path evolution of the stochastic true ocean system between two assimilation time steps is assumed to be described by the classical linear dynamical model evolution equation, $\boldsymbol{\psi}_k = \mathbf{A}_{k-1}\boldsymbol{\psi}_{k-1}$, forced by a model random noise $\mathbf{w}_{k-1}$:

$$\boldsymbol{\psi}_k^t = \mathbf{A}_{k-1}\boldsymbol{\psi}_{k-1}^t + \mathbf{w}_{k-1} \tag{1}$$

The discrete boundary conditions can be seen as diagnostic relations between boundary state variables at time t_k. Unless otherwise mentioned, the state vector contains the boundary variables and the boundary conditions are thus part of equation 1. The associated linear true measurement model is

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

where $\mathbf{d}_k = \mathbf{C}_k\boldsymbol{\psi}_k$ is the classical deterministic measurement model and $\mathbf{v}_k$ its random noise.

The sample path (1) of the true ocean is thus hypothesized to be a linear vector Markov process stochastically forced by Wiener processes (Kloeden and Platen, 1992). Classical notations and assumptions are used; for extended discussions we refer to Gelb (1974), Daley (1991) and Lermusiaux (1997). In particular, we have:

- $\boldsymbol{\psi}_k^t$, $\mathbf{w}_k \in \mathbb{R}^n$ and $\mathbf{d}_k$, $\mathbf{v}_k \in \mathbb{R}^m$ (n state variables, m meas. with $m \ll n$)
- $\mathbf{A}_{k-1} \in \mathbb{R}^{n\times n}$ and $\mathbf{C}_k \in \mathbb{R}^{m\times n}$
- $\mathbf{w}_k \sim (0, \mathbf{Q}_k)$ and $\mathbf{v}_k \sim (0, \mathbf{R}_k)$ with

$$\mathcal{E}\{\mathbf{w}_k\mathbf{w}_j^T\} = \mathcal{E}\{\mathbf{v}_k\mathbf{v}_j^T\} = 0 \text{ for } k \neq j \quad \text{and} \quad \mathcal{E}\{\mathbf{w}_k\mathbf{v}_j^T\} = 0 \qquad \forall k, j$$

If we denote by $\hat{\boldsymbol{\psi}}_k$ the estimate of $\boldsymbol{\psi}_k^t$, the white random process $\mathbf{w}_{k-1}$ is also uncorrelated to the initial estimate's error at time t_{k-1}, $(\hat{\boldsymbol{\psi}}_{k-1} - \boldsymbol{\psi}_{k-1}^t)$. The probability density functions of the random forcings are not required to be Gaussian even though only their respective mean (0) and covariances $(\mathbf{Q}_k, \mathbf{R}_k)$ need to be specified. The estimate's error covariance matrix at time t_k is denoted by $\mathbf{P}_k \in \mathbb{R}^{n\times n}$ and is defined by $\mathbf{P}_k \doteq \mathcal{E}\{(\hat{\boldsymbol{\psi}}_k - \boldsymbol{\psi}_k^t)(\hat{\boldsymbol{\psi}}_k - \boldsymbol{\psi}_k^t)^T\}$. Its value at time t_0, $\mathbf{P}_0$, represents the initial condition error covariance.

A.2. *Estimation Theory (Sequential Estimation)*

Direct Insertion

Forecast $\hat{\boldsymbol{\psi}}_k(-) = \mathbf{A}_{k-1}\hat{\boldsymbol{\psi}}_{k-1}(+)$ (3)

Melding $(\hat{\boldsymbol{\psi}}_k)_d(+) = \mathbf{d}_k \qquad [\mathbf{d}_k = \mathbf{C}_k\hat{\boldsymbol{\psi}}_k(+)]$ (4)

at the m data points and $\hat{\boldsymbol{\psi}}_k(+) = \hat{\boldsymbol{\psi}}_k(-)$ elsewhere.

Blending

Forecast $\hat{\boldsymbol{\psi}}_k(-) = \mathbf{A}_{k-1}\hat{\boldsymbol{\psi}}_{k-1}(+)$ (5)

Melding $(\hat{\boldsymbol{\psi}}_k)_d(+) = \alpha\mathbf{d}_k + (1-\alpha)\mathbf{C}_k\hat{\boldsymbol{\psi}}_k(-)$ (6)

at the m data points and $\hat{\boldsymbol{\psi}}_k(+) = \hat{\boldsymbol{\psi}}_k(-)$ elsewhere.

Nudging or Newtonian Relaxation Scheme (Anthes, 1974)

$$\hat{\boldsymbol{\psi}}_k = \mathbf{A}_{k-1}\hat{\boldsymbol{\psi}}_{k-1} + \tilde{\mathbf{K}}_k\mathbf{C}_k^{+}(\mathbf{d}_k - \mathbf{C}_k\mathbf{A}_{k-1}\hat{\boldsymbol{\psi}}_{k-1}) \quad (7)$$

where $\tilde{\mathbf{K}}_k$ is user-assigned, diagonal and $\mathbf{C}_k^{+}$, e.g., is a feature model inverse of $\mathbf{C}_k$

Optimal Interpolation (OI) or Statistical Interpolation (SI)

Forecast $\hat{\boldsymbol{\psi}}_k(-) = \mathbf{A}_{k-1}\hat{\boldsymbol{\psi}}_{k-1}(+)$ (8)

Melding $\hat{\boldsymbol{\psi}}_k(+) = \hat{\boldsymbol{\psi}}_k(-) + \tilde{\tilde{\mathbf{K}}}_k[\mathbf{d}_k - \mathbf{C}_k\hat{\boldsymbol{\psi}}_k(-)]$ (9)

with gains $\tilde{\tilde{\mathbf{K}}}_k$ assigned empirically.

Method of Successive Corrections

Forecast $\hat{\boldsymbol{\psi}}_k(-) = \mathbf{A}_{k-1}\hat{\boldsymbol{\psi}}_{k-1}(+)$ (10)

Iterative Meldings $\hat{\boldsymbol{\psi}}_k^{j+1}(+) = \hat{\boldsymbol{\psi}}_k^{j}(+) + \mathbf{W}^j[\mathbf{d}_k - \mathbf{C}_k\hat{\boldsymbol{\psi}}_k^{j}(+)]$ (11)

with $\hat{\boldsymbol{\psi}}_k^{0}(+) = \hat{\boldsymbol{\psi}}_k(-)$ and $\mathbf{W}^j$ assigned function of j in general.

Kalman Filter

$$\{\hat{\boldsymbol{\psi}}_k \mid \min_{\hat{\boldsymbol{\psi}}_k} J_k = \text{trace}[\mathbf{P}_k(+)] \text{ using } [\mathbf{d}_1, \ldots, \mathbf{d}_k]\}$$

Forecast $\hat{\boldsymbol{\psi}}_k(-) = \mathbf{A}_{k-1}\hat{\boldsymbol{\psi}}_{k-1}(+)$ (12)

$\mathbf{P}_k(-) = \mathbf{A}_{k-1}\mathbf{P}_{k-1}(+)\,\mathbf{A}_{k-1}^T + \mathbf{Q}_{k-1}$ (13)

Analysis: $\hat{\boldsymbol{\psi}}_k(+) = \hat{\boldsymbol{\psi}}_k(-) + \mathbf{K}_k[\mathbf{d}_k - \mathbf{C}_k\hat{\boldsymbol{\psi}}_k(-)]$ (14)

Kalman Gain $\mathbf{K}_k = \mathbf{P}_k(-)\mathbf{C}_k^T[\mathbf{C}_k\mathbf{P}_k(-)\mathbf{C}_k^T + \mathbf{R}_k]^{-1}$ (15)

Error Covariance Update $\mathbf{P}_k(+) = \mathbf{P}_k(-) - \mathbf{K}_k\mathbf{C}_k\mathbf{P}_k(-)$ (16)

Kalman Smoother

$$\{\hat{\boldsymbol{\psi}}_{k/N} \mid \min_{\hat{\boldsymbol{\psi}}_{k/N}} J_k = \text{trace}[\mathbf{P}_k(+)] \text{ using } [\mathbf{d}_1, \ldots, \mathbf{d}_{N-1}]\}$$

Filter Estimate *See equations 12–16.*

Smoothed Estimate (Backward in Time) $\hat{\boldsymbol{\psi}}_{k/N} = \hat{\boldsymbol{\psi}}_k(+) + \mathbf{L}_k[\hat{\boldsymbol{\psi}}_{k+1/N} - \hat{\boldsymbol{\psi}}_{k+1}(-)]$ (17)

with $\hat{\boldsymbol{\psi}}_{N/N} = \hat{\boldsymbol{\psi}}_N(+)$ and $\mathbf{L}_k = \mathbf{P}_k(+)\mathbf{A}_k^T\mathbf{P}_{k+1}^{-1}(-)$.

Smoothed Error Covariance $\mathbf{P}_{k/N} = \mathbf{P}_k(+) + \mathbf{L}_k\,[\mathbf{P}_{k+1/N} - \mathbf{P}_{k+1}(-)]\,\mathbf{L}_k^T$ (18)

with $\mathbf{P}_{N/N} = \mathbf{P}_N(+)$.

A.3. Control Theory (Calculus of Variation Approach, Variational Assimilation)

Adjoint Method

$$\min_{\hat{\boldsymbol{\psi}}_k} J_N = \boldsymbol{\epsilon}_0^T \mathbf{P}_0^{-1} \boldsymbol{\epsilon}_0 + \sum_{k=1}^{N-1} \mathbf{v}_k^T \mathbf{R}_k^{-1} \mathbf{v}_k + \sum_{k=1}^{N} 2\boldsymbol{\lambda}_{k-1}^T \mathbf{w}_{k-1} \tag{19}$$

Dynamical Model $$\hat{\boldsymbol{\psi}}_k = \mathbf{A}_{k-1} \hat{\boldsymbol{\psi}}_{k-1}, \qquad k = 1, \ldots, N \tag{20}$$

Initial Condition $$\hat{\boldsymbol{\psi}}_0 = \boldsymbol{\Psi}_0 + \mathbf{P}_0 \mathbf{A}_0^T \boldsymbol{\lambda}_0 \tag{21}$$

Adjoint Model $$\boldsymbol{\lambda}_{k-1} = \mathbf{A}_k^T \boldsymbol{\lambda}_k + \mathbf{C}_k^T \mathbf{R}_k^{-1} (\mathbf{d}_k - \mathbf{C}_k \hat{\boldsymbol{\psi}}_k), \qquad k = 1, \ldots, N-1 \tag{22}$$

Initial Condition $$\boldsymbol{\lambda}_{N-1} = 0 \tag{23}$$

Generalized Inverse Problem

$$\min_{\hat{\boldsymbol{\psi}}_k} J_N = \boldsymbol{\epsilon}_0^T \mathbf{P}_0^{-1} \boldsymbol{\epsilon}_0 + \sum_{k=1}^{N-1} \mathbf{v}_k^T \mathbf{R}_k^{-1} \mathbf{v}_k + \sum_{k=1}^{N} \mathbf{w}_{k-1}^T \mathbf{Q}_{k-1}^{-1} \mathbf{w}_{k-1} \tag{24}$$

Weak Dynamical Model $$\hat{\boldsymbol{\psi}}_k = \mathbf{A}_{k-1} \hat{\boldsymbol{\psi}}_{k-1} + \mathbf{Q}_{k-1} \boldsymbol{\lambda}_{k-1}, \qquad k = 1, \ldots, N \tag{25}$$

Initial Condition $$\hat{\boldsymbol{\psi}}_0 = \boldsymbol{\Psi}_0 + \mathbf{P}_0 \mathbf{A}_0^T \boldsymbol{\lambda}_0 \tag{26}$$

Adjoint Model $$\boldsymbol{\lambda}_{k-1} = \mathbf{A}_k^T \boldsymbol{\lambda}_k + \mathbf{C}_k^T \mathbf{R}_k^{-1} [\mathbf{d}_k - \mathbf{C}_k \hat{\boldsymbol{\psi}}_k], \qquad k = 1, \ldots, N-1 \tag{27}$$

Initial Condition $$\boldsymbol{\lambda}_{N-1} = 0 \tag{28}$$

Representer Method for Solving the Generalized Inverse Problem

$$\hat{\boldsymbol{\psi}}_k = \boldsymbol{\psi}_k^f + \tilde{\mathbf{R}}_{kl} \mathbf{b}_l \tag{29}$$

Forecast $$\boldsymbol{\psi}_k^f = \mathbf{A}_{k-1} \boldsymbol{\psi}_{k-1}^f \quad \text{with } \boldsymbol{\psi}_0^f = \boldsymbol{\Psi}_0 \tag{30}$$

Representers Evolution $$\tilde{\mathbf{R}}_{kl} = \mathbf{A}_{k-1} \tilde{\mathbf{R}}_{k-1,l} + \mathbf{Q}_{k-1} \boldsymbol{\Gamma}_{k-1,l} \quad \text{with } \tilde{\mathbf{R}}_{0l} = \mathbf{P}_0 \mathbf{A}_0^T \boldsymbol{\Gamma}_{0l} \tag{31}$$

Adjoint Evolution $$\boldsymbol{\Gamma}_{k-1,l} = \mathbf{A}_k^T \boldsymbol{\Gamma}_{kl} + \mathbf{C}_k^T \boldsymbol{\delta}_{kl} \quad \text{with } \boldsymbol{\Gamma}_{N-1,l} = 0 \tag{32}$$

Representers Coefficient $$\boldsymbol{\Gamma}_{k-1,l}\, \mathbf{b}_l = \boldsymbol{\lambda}_{k-1} \tag{33}$$

$$\mathbf{b}_l = \mathbf{R}_l^{-1} \cdot [\mathbf{d}_l - \mathbf{C}_l \hat{\boldsymbol{\psi}}_l] \tag{34}$$

$$\underline{\mathbf{b}} = [\underline{\mathbf{R}} + \underline{\mathbf{C}\tilde{\mathbf{R}}}]^{-1} \cdot [\underline{\mathbf{d}} - \underline{\mathbf{C}\boldsymbol{\psi}^f}] \tag{35}$$

where $\underline{\mathbf{b}}$ is the $m(N-1)$ block vector of the $\mathbf{b}_l$'s; $\underline{\mathbf{R}}$ is an $m(N-1)$ block diagonal square matrix, with the data error covariances $\mathbf{R}_k$, $k = 1, \ldots, N-1$, on its diagonal; $\underline{\mathbf{C}\tilde{\mathbf{R}}}$ is the equivalent of the representer matrix of Bennett (1992) and is an $m(N-1)$ block square matrix, which square block kl is the size m matrix $\mathbf{C}_k \tilde{\mathbf{R}}_{kl}$; and $[\underline{\mathbf{d}} - \underline{\mathbf{C}\boldsymbol{\psi}^f}]$ is the $m(N-1)$ block vector of the forecast misfits $[\mathbf{d}_k - \mathbf{C}_k \boldsymbol{\psi}_k^f]$.

$$\hat{\boldsymbol{\psi}}_k = \boldsymbol{\psi}_k^f + \tilde{\mathbf{R}}_{kl} [\underline{\mathbf{R}} + \underline{\mathbf{C}\tilde{\mathbf{R}}}]_l^{-1} \cdot [\underline{\mathbf{d}} - \underline{\mathbf{C}\boldsymbol{\psi}^f}] \tag{36}$$

Bibliography

Aart, E. H. L. and J. H. M. Korst, 1991. Boltzmann machines as a model for parallel annealing. *Algorithmica*, **6**(3), 437–465.

Adamec, D., 1989. Predictability of quasi-geostrophic ocean flow: sensitivity to varying model vertical resolution. *J. Phys. Oceanogr.*, **19**(11), 1753–1764.

Aikman, F., III, G. L. Mellor, T. Eger, D. Sheinin, P. Chen, L. Breaker and D. B. Rao, 1996. Towards an operational nowcast/forecast system for the U.S. east coast. In *Modern Approaches to Data Assimilation in Ocean Modeling*, P. Malanotte-Rizzoli, ed. Elsevier Oceanography Series. Elsevier, New York.

Anderson, D. and J. Willebrand, eds., 1989. *Oceanic Circulation Models: Combining Data and Dynamics*. Kluwer, Dordrecht, The Netherlands.

Anthes, R. A., 1974. Data assimilation and initialization of hurricane prediction models. *J. Atmos. Sci.*, **31**, 702–719.

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

Baptista, A. M., E. E. Adams and P. Gresho, 1995. Benchmarks for the transport equation: the convection–diffusion forum and beyond. In *Coastal and Estuarine Studies*, Vol. 47, D. R. Lynch and A. M. Davies, eds. American Geophysical Union, Washington, D.C., pp. 241–269.

Barth, N. H., 1992. Oceanographic experiment design. II. Genetic algorithms. *J. Atmos. Ocean. Technol.*, **9**, 434–443.

Barth, N. H. and C. Wunsch, 1990. Oceanographic experiment design by simulated annealing. *J. Phys. Oceanogr.*, **20**, 1249–1263.

Batchelor, G. K., 1967. *An Introduction to Fluid Dynamics*. Cambridge University Press, Cambridge, 615pp.

Beardsley, R. C. and C. D. Winant, 1979. On the mean circulation in the Mid-Atlantic Bight. *J. Phys. Oceanogr.*, **9**(3), 612–619.

Beardsley, R. C., W. C. Boicourt, L. C. Huff, J. R. McCullough and J. Scott, 1981. CMICE: A near-surface current meter intercomparison experiment. *Deep-Sea Res.*, **28**(12A), 1577–1603.

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

Bennett, A. F., 1992. *Inverse Methods in Physical Oceanography*. Cambridge Monographs on Mechanics and Applied Mathematics. Cambridge University Press, Cambridge.

Bennett, A. F. and W. P. Budgell, 1989. The Kalman smoother for a linear quasigeostrophic model of ocean circulation. *Dyn. Atmos. Oceans*, **13**(3–4), 219–268.

Bennett, A. F. and B. S. Chua, 1994. Open-ocean modeling as an inverse problem: the primitive equations. *Mon. Weather Rev.*, **122**(6), 1326–1336.

Bennett, A. F. and P. G. McIntosh, 1982. Open ocean modeling as an inverse problem: tidal theory. *J. Phys. Oceanogr.*, **12**, 1004–1018.

Bennett, A. F. and M. A. Thorburn, 1992. The generalized inverse of a nonlinear quasigeostrophic ocean circulation model. *J. Phys. Oceanogr.*, **22**, 213–230.

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

Bennett, A. F., B. S. Chua and L. M. Leslie, 1997. Generalized inversion of a global numerical weather prediction model. II. Analysis and implementation. *Meteorol. Atmos. Phys.*, **62**(3–4), 129–140.

Bergamasco, A. et al., 1993. The seasonal steady circulation of the eastern Mediterranean determined with the adjoint method. *Deep-Sea Res.*, **40**(6), 1269–1299.

Bergé, P., Y. Pomeau and C. Vidal, 1984. *Order Within Chaos: Towards a Deterministic Approach to Turbulence*. Wiley-Interscience, New York, 329pp.

Berkovitz, L. D., 1974. *Optimal Control Theory*. Springer-Verlag, New York.

Bogden, P. S., P. Malanoote-Rizzoli and R. Signell, 1955. Open ocean boundary conditions from interior data: local and remote forcing of Massachusetts Bay. *J. Geophys. Res.*, **101**(C3), 6487–6500.

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

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

Bowen, A. J., D. A. Griffin, D. G. Hazen, S. A. Matheson and K. R. Thompson, 1995. Shipboard nowcasting of shelf circulation. *Cont. Shelf Res.*, **15**(1), 115–128.

Brasseur, P., ed., 1995. Data assimilation in marine science. *J. Mar. Syst.*, **6**(1–2), 175pp.

Brasseur, P. and J. C. J. Nihoul, eds. 1994. Data assimilation: tools for modelling the ocean in a global change perspective. In NATO ASI Series, Series I. *Global Environmental Change*, Vol. 19. Springer-Verlag, New York, 239pp.

Bratseth, A., 1986. Statistical interpolation by means of successive corrections. *Tellus*, **38A,** 439–447.

Bretherton, F. P., R. E. Davis and C. B. Fandry, 1976. A technique for objective analysis and design of oceanographic experiments applied to MODE-73. *Deep-Sea Res.*, **23,** 539–582.

Brickman, D. and J. W. Loder, 1993. Energetics of the internal tide on northern Georges Bank. *J. Phys. Oceanogr.*, **24,** 1464–1479.

Brockett, R. W., 1970. *Finite Dimensional Linear Systems*. Wiley, New York, 239pp.

Brummelhuis, P. G. S. T., 1990. The calibration of 2-D shallow water flow models. In *Realization and Modelling in System Theory*, Proc. Int. Symp. MTNS-89, Vol. 1, M. A. Kaashoek, J. H. van Schuppen and A. C. M. Ran, eds., Birkhäuser, Basel, Switzerland.

Brummelhuis, P. G. J. T., A. W. Heemink and H. F. P. van den Boogaard, 1993. Identification of shallow sea models. *Int. J. Numer. Methods Fluids*, **17,** 637–665.

Budgell, N. P., 1986. Nonlinear data assimilation for shallow water equation in branched channels. *J. Geophys. Res.*, **91,** 10633–10644.

Carey, G. F., 1995. Mesh generation, *a posteriori* error estimation and mesh refinement. In *Coastal and Estuarine Studies*, Vol. 47, D. R. Lynch and A. M. Davies, eds. American Geophysical Union, Washington, D.C., pp. 15–30.

Carman, J. C. and A. R. Robinson, 1994. Oceanographic/topographic interaction in acoustic propagation in the Iceland Faeroe Front region. *J. Acoust. Soc. Am.*, **95**(4), 1882–1894.

Carter, E. F., 1993. Data assimilation into nonlinear stochastic ocean models. Seminar, Harvard University, Cambridge, Mass.

Carton, J. A., 1987. How predictable are the geostrophic currents in the recirculation zone? *J. Phys. Oceanogr.*, **17**(6), 751–762.

Catlin, D. E., 1989. Estimation, control and the discrete Kalman filter. *Applied Mathematical Sciences*, Vol. 71. Springer-Verlag, New York, 274pp.

Chapman, D. C., 1985. Numerical treatment of cross-shelf open boundaries in barotropic coastal ocean model. *J. Phys. Oceanogr.*, **15,** 1060–1075.

Charney, J., M. Halem and R. Jastrow, 1969. Use of incomplete historical data to infer the present state of the atmosphere. *J. Atmos. Sci.*, **26,** 1162.

Churchill, J. H., P. C. Cornillon and G. W. Milkowski, 1986. A cyclonic eddy and shelf–slope water exchange associated with a Gulf Stream warm-core ring. *J. Geophys. Res.*, **91,** 9615–9623.

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

Cohn, S. E. and D. F. Parrish, 1991. The behavior of forecast error covariances for a Kalman filter in two dimensions. *Mon. Weather Rev.*, **119,** 1757–1785.

Courtier, P., 1995. Variational methods. *Geophys. Mag.*, Ser. 2, Vol. 1, Special Issue.

Crowley, T. J. and S. Glenn, 1994. Personal communication.

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

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

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

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

Das, S. and R. W. Lardner, 1991. On the estimation of parameters of hydraulic models by assimilation of periodic tidal data. *J. Geophys. Res.*, **96**(8), 15187–15196.

Daum, F., 1994. Nonlinear prediction and filtering. Seminar. Boston.

Davies, A. M. and J. Xing, 1995. An intercomparison and validation of a range of turbulence closure schemes used in three-dimensional tidal models. In *Coastal and Estuarine Studies*, Vol. 47, D. R. Lynch and A. M. Davies, eds. American Geophysical Union, Washington, D.C., pp. 71–96.

Davies, A. M., P. J. Luyten and E. Deleersnijder, 1995. Turbulence energy models in shallow sea oceanography. In *Coastal and Estuarine Studies*, Vol. 47, D. R. Lynch and A. M. Davies, eds. American Geophysical Union, Washington, D.C., pp. 97–125.

Davis, R. E., R. Dolan and G. Demme, 1993. Synoptic climatology of Atlantic coast north-easters. *Int. J. Climatol.*, **13**(2), 171–189.

Derber, J. C., 1989. A variational continuous assimilation scheme. *Mon. Weather Rev.*, **117**, 2347–2446.

Egbert, G. D., A. F. Bennett and M. G. G. Foreman, 1994. TOPEX/POSEIDON tides estimated using a global inverse model. *J. Geophys. Res.*, **99**(C12), 24821–24852.

Ekman, V. W., 1905. On the influence of the earth's rotation on ocean currents. *Arch. Math. Astron. Phys.*, **2**(11).

Evensen, G., 1993. Open boundary conditions for the extended Kalman filter with a quasi-geostrophic ocean model. *J. Geophys. Res.*, **98**, 16529–16546.

Evensen, G., 1994a. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**(C5), 10143–10162.

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

Evensen, G. 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. Weather Rev.*, **124**(1), 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.

Fisher, M. and M. Latif, 1995. Assimilation of temperature and sea level observations into a primitive equation model of the tropical Pacific. *J. Mar. Syst.*, **6**, 31–46.

Flagg, C. N., 1988. Internal waves and mixing along the New England shelf-slope front. *Cont. Shelf Res.*, **8**(5–7), 737–756.

Flierl, G. R. and C. S. Davis, 1995. Reduction of complexity in biological/physical models. Draft.

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.

Fukumori, I., 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.

Garvine, R. W., K.-C. Wong, G. G. Gawarkiewicz, R. K. McCarthy, R. W. Houghton and F. Aikman III, 1988. The morphology of shelfbreak eddies. *J. Geophys. Res.*, **93**, 15593–15607.

Gawarkiewicz, G., 1991. Linear stability models of shelfbreak fronts. *J. Phys. Oceanogr.*, **21**, 471–488.

Gawarkiewicz, G., T. M. Church, G. W. Luther III, T. G. Ferdelman and M. Caruso, 1992. Large-scale penetration of Gulf Stream water onto the continental shelf north of Cape Hatteras. *Geophys. Res. Let.*, **19**(4), 373–376.

Gelb, A., ed., 1974. *Applied Optimal Estimation*. MIT Press, Cambridge, Mass.

Gelfand, I. M. and S. V. Fomin, 1963. *Calculus of Variations*. Prentice Hall, Englewood Cliffs, N.J.

Gerdes, R., 1993. A primitive equation ocean circulation model using a general vertical coordinate transformation. 1. Description and testing of the model. *J. Geophys. Res.*, **98**, 14683–14701.

Gerritsen, H., H. de Vries and M. Philippart, 1995. The Dutch continental shelf model. In *Coastal and Estuarine Studies*, Vol. 47, D. R. Lynch and A. M. Davies, eds., American Geophysical Union, Washington, D.C., pp. 425–469.

Ghil, M., 1989. Meteorological data assimilation for oceanographers. I. Description and theoretical framework. *Dyn. Atmos. Oceans*. Special Issue: Data assimilation, Vol. 13, Nos. 3/4, 171–218.

Ghil, M. and P. Malanotte-Rizzoli, 1991. Data assimilation in meteorology and oceanography. *Advances in Geophysics*, Vol. 33. Academic Press, San Diego, Calif., pp. 141–266.

Gill, P. E., W. Murray and M. H. Wright, 1982. *Practical Optimization*. Academic Press, London.

Glenn, S. M., 1994. Personnal communication on sediment transport.

GLOBEC, 1994. An advanced modeling/observation system (AMOS) for physical–biological–chemical ecosystem research and monitoring (concepts and methodology). *GLOBEC Special Contributions 2*. Technical Report. Harvard University, Cambridge, Mass., 149pp.

GLOBEC, 1995. Interdisciplinary model formulation and parameterization, Numerical Modeling Working Group 1995, Nantes, France, Meeting Report. *GLOBEC Report 7*.

Goldstein, S., ed., 1965. *Modern Developments in Fluid Dynamics*. Dover, New York, Vols. 1–2, 702pp.

Grassle, J. F., C. von Alt and S. M. Glenn, in preparation. An electro-optical-table linking a shore laboratory with a long-term ecosystem observatory on the continental shelf of New Jersey.

Haidvogel, D. B. and A. R. Robinson, eds., 1989. Special issue on data assimilation. *Dyn. Atmos. Oceans*, **13**, 171–517.

Haney, R. L., 1991. On the pressure gradient force over steep topography in sigma coordinate ocean models. *J. Phys. Oceanogr.*, **21**, 610–619.

Heemink, A. W. and H. Kloosterhuis, 1990. Data assimilation for non-linear tidal models. *Int. J. Numer. Methods Fluids*, **11**, 1097–1112.

Holton, J., 1992. *An Introduction to Dynamic Meteorology*, 3rd ed. International Geophysics Series, Vol. 48. Academic Press, San Diego, Calif., 511pp.

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

Houghton, R. W., R. Schlitz, R. C. Beardsley, B. Butman and J. Lockwood Chamberlin, 1982. The Middle Atlantic Bight cold pool: Evolution of the temperature structure during summer 1979. *J. Phys. Oceanogr.*, **12**(10), 1019–1029.

Houghton, R. W., F. Aikman III and H. W. Ou, 1988. Shelf-slope frontal structure and cross-shelf exchange at the New England shelf-break. *Cont. Shelf Res.*, **8**(5–7), 687–710.

Huthnance, J. M., 1986. The subtidal behaviour of the Celtic Sea, III: A model of shelf waves and surges on a wide shelf. *Cont. Shelf Res.*, **5**(3), 347–377.

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

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

Johannessen, J. A., L. P. Roed, O. M. Johannessen, G. Evensen, B. Hackett, L. H. Pettersson, P. M. Haugan, S. Sandven and R. Shuchman, 1993. Monitoring and modeling of the marine coastal environment. *Photogramm. Eng. Remote Sensing*, **59**(3), 351–361.

Johnsen, M. and D. R. Lynch, 1995. Assessment of a second-order radiation boundary condition for tidal and wind driven flows. In *Coastal and Estuarine Studies*, Vol. 47, D. R. Lynch and A. M. Davies, eds. American Geophysical Union, Washington, D.C., pp. 49–70.

Kalman, R. E., 1960. A new approach to linear filtering and prediction problems. *J. Basic Eng.*, **82D**, 35–45.

Kirkpatrick, S. et al., 1983. Optimization by simulated annealing. *Science*, **220**, 671–680.

Kloeden, P. E. and E. Platen, 1992. Numerical solution of stochastic differential equations. In *Applications of Mathematics*, Vol. 23. Springer-Verlag, New York, 632pp.

Krüger, J., 1993. Simulated annealing: a tool for data assimilation into an almost steady model state. *J. Phys. Oceanogr.*, **23**, 679–688.

Lacarra, J. F. and O. Talagrand, 1988. Short-range evolution of small perturbations in a barotropic model. *Tellus*, **40A**, 81–95.

Landahl, M. T. and E. Mollo-Christensen, 1992. *Turbulence and Random Processes in Fluid Mechanics*, 2nd ed. Cambridge University Press, London.

Lardner, R. W., 1993. Optimal control of open boundary conditions for a numerical tidal model. *Comput. Methods Appl. Mech. Eng.*, **102**, 367–387.

Lardner, R. W. and S. K. Das, 1994. Optimal estimation of eddy viscosity for a quasi-three-dimensional numerical tidal and storm surge model. *Int. J. Numer. Methods Fluids*, **18**(3), 295–312.

Lardner, R. W. and Y. Song, 1995. Variational data assimilation for a three-dimensional coastal-ocean model. *WMO Symposium*, Tokyo.

Lardner, R. W. and Y. Song, in preparation. Optimal estimation of eddy viscosity and friction coefficients for a quasi-three-dimensional numerical tidal model. Draft.

Lawson, L. M., Y. H. Spitz, E. E. Hofmann and R. B. Long, 1995. A data assimilation technique applied to a predator–prey model. *Bull. Math. Biol.*, **57**(4), 593–617.

Le Blond, P. and L. Mysak, 1978. *Waves in the Ocean*. Elsevier Oceanographic Series, Vol. 20. Elsevier, Amsterdam, 602pp.

Le Dimet, F. X. and O. Talagrand, 1986. Variational algorithms for analysis and assimilation of meteorological meterological 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, Mass.

Lermusiaux, P. F. J. and A. R. Robinson, in preparation. Data assimilation via error subspace statistical estimation.

Li, Y., I. M. Navon, W. Yang, X. Zou, J. R. Bates, S. Moorthi and R. W. Higgins, 1994. Four-dimensional variational data assimilation experiments with a multilevel semi-Lagrangian semi-implicit general circulation model. *Mon. Weather Rev.*, **122**, 966–983.

Lions, J. L., 1971. *Optimal Control of Systems Governed by Partial Differential Equations*. Springer-Verlag, Berlin.

Ljung, L. and T. Söderström, 1987. *Theory and Practice of Recursive Identification*. MIT Press, Cambridge, Mass.

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

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

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

Lorenc, A. C., R. S. Bell and B. Macpherson, 1991. The Meteorological Office analysis correction data assimilation scheme. *Q. J. R. Meteorol. Soc.*, **117**, 59–89.

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

Lorenz, E. N., 1975. Climate predictability: the physical basis of climate modelling. *WMO, GARP Publ. Ser.*, **16**, 132–136.

Lozano, C. J., P. J. Haley, H. G. Arango, Q. Sloan and A. R. Robinson, 1994. Harvard coastal/deep water primitive equation model. *Harvard Open Ocean Model Report 52*. Harvard University, Cambridge, Mass., 15pp.

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. In *Modern Approaches to Data Assimilation in Ocean Modelling*, P. Malanotte-Rizzoli, ed. Elsevier Oceanography Series. Elsevier, Amsterdam.

Luenberger, D. C., 1984. *Linear and Nonlinear Programming*. Addison-Wesley, Reading, Mass.

Lynch, D. R. and A, M. Davies, eds., 1995. Quantitative skill assessment for coastal ocean models. In *Coastal and Estuarine Studies*. American Geophysical Union, Washington, D.C., 510pp.

Malanotte-Rizzoli, P., ed., 1996. *Modern Approaches to Data Assimilation in Ocean Modeling*. Elsevier Oceanography Series. Elsevier, Amsterdam, 455pp.

Malanotte-Rizzoli, P. and R. E. Young, 1992. How useful are localized clusters of traditional oceanographic measurements for data assimilation? *Dyn. Atmos. Oceans*, **17**, 23–61.

Malanotte-Rizzoli, P. and R. E. Young, 1995. Assimilation of global versus local datasets into a regional model of the Gulf Stream system. I. Data effectiveness. *J. Geophys. Res.*, **100**(C12), 24773–24793.

Manning, J., 1991. Middle Atlantic Bight salinity: Interannual variability. *Cont. Shelf Res.*, **11**(2), 123–137.

Mayer, D. A., G. C. Han and D. V. Hansen, 1982. Circulation in the Hudson Shelf Valley: MESA physical oceanographic studies in New York Bight, I. *J. Geophys. Res.*, **87**(C12), 9563–9578.

Martel, F. and C. Wunsch, 1993a. The North Atlantic circulation in the early 1980s: an estimate from inversion of a finite-difference model. *J. Phys. Oceanogr.*, **23**, 898–924.

Martel, F. and C. Wunsch, 1993b. Combined inversion of a finite difference model and altimetric sea surface topography. *Man. Geod.*, **18**(4), 219–226.

McCalpin, J. D., 1994. A comparison of second-order and fourth-order pressure gradient algorithms in a sigma coordinate ocean model. *Int. J. Numer. Methods*, **18**, 361–383.

McIntosh, P. G. and A. F. Bennett, 1984. Open ocean modeling as an inverse problem: M_2 tides in Bass Strait. *J. Phys. Ocean.*, **14**, 601–614.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, 1953. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, **21**, 1987–1092.

Miller, R. N., 1986. Toward the application of the Kalman filter to regional open ocean modeling. *J. Phys. Oceanogr.*, **16**, 72–86.

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.

Miller, R. N., M. Ghil and F. Gauthier, 1994a. Data assimilation in strongly nonlinear dynamical systems. *J. Atmos. Sci.*, **51**, 1037–1056.

Miller, R. N., E. O. Zaron and A. F. Bennett, 1994b. Data assimilation in models with convective adjustment. *Mon. Weather Rev.*, **122**, 2607–2613.

Moody et al., 1983. Personal communication articles in the MAB.

Mooers, C. N. K., A. R. Robinson and J. D. Thompson, 1986. Ocean prediction workshop 1986: a status and prospectus report on the scientific basis and the Navy's needs. In *Proc. Ocean Prediction Workshop*. Institute of Naval Oceanography. NSTL, MS.

Mountain, D. G., 1991. The volume of Shelf Water in the Middle Atlantic Bight: Seasonal and interannual variability 1977–87. *Cont. Shelf Res.*, **11**(3), 251–267.

Munchow, A. and R. W. Garvine, 1993. Dynamical properties of a buoyancy-driven coastal current. *J. Geophys. Res.*, **98**(C11), 20063–20077.

National Research Council Panel on Model-Assimilated Data Sets for Atmospheric and Oceanic Research, 1991. *Four Dimensional Model Assimilation of Data: A Strategy for the Earth System Sciences*. National Academy Press, Washington, D.C.

Navon, I. M. and D. M. Legler, 1987. Conjugate-gradient methods for large-scale minimization in meteorology. *Mon. Weather Rev.*, **115**, 1479–1502.

Noble, M. and B. Butman, 1979. Low-frequency wind-induced sea level oscillations along the east coast of North America. *J. Geophys. Res.*, **84**(C6), 3227–3236.

Oey, L.-Y. and P. Chen, 1992. A nested-grid ocean model with application to the simulation of meanders and eddies in the Norwegian Coastal Current. *J. Geophys. Res.*, **97**, 20063–20086.

Ou, H. W., R. C. Beardsley, W. C. Boicourt and B. Butman, 1981. An analysis of subtidal current fluctuations in the Middle Atlantic Bight. *J. Phys. Oceanogr.*, **11**(10), 1383–1392.

Palanques, A. and P. E. Biscaye, 1992. Suspended matter distribution over the shelf and upper slope south of New England. *Cont. Shelf Res.*, **12**(5–6), 577–600.

Palmer, T. N., 1993. Extended-range atmospheric prediction and the Lorenz mode. *Bull. Am. Meteorol. Soc.*, **74**(1), 49–65.

Panchang, V. G. and J. E. Richardson, 1993. Inverse adjoint estimation of eddy viscosity for coastal flow models. *J. Hydraul. Eng.*, **119**(4), 506–525.

Phillips, N. A., 1982. On the completeness of multi-variate optimum interpolation for large-scale meteorological analysis. *Mon. Weather Rev.*, **110**, 1329–1334.

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

Press, W. H., B. P. Flannery, S. A. Teukolsky and W. T. Vetterling, 1989. *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, Cambridge.

Reid, W. T., 1968. Generalized inverses of differential and integral operators. In *Theory and Applications*

of Generalized Inverse of Matrices, T. L. Boullion and P. L. Odell, eds. Texas Technical College, Lubbock, Texas, pp. 1–25.

Reynolds, C. A., 1895. On the dynamical theory of incompressible viscous fluids and the determination of the criterion. *Philos. Trans. A*, **186,** 123–164.

Roach, G. F., 1982. *Green's Functions*, 2nd ed. Cambridge University Press, Cambridge, 325pp.

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

Robinson, A. R. and W. G. Leslie, 1985. Estimation and prediction of oceanic fields. In *Progress in Oceanography*, Vol. **14,** J. Crease, W. J. Gould and P. M. Saunders, eds. Essays on Oceanography. Pergamon Press, Oxford, pp. 485–510.

Robinson, A. R., 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. Am. Meterol. Soc.*, 77(2), 243–259.

Robinson, A. R., H. G. Arango, A. Warn-Varnas, W. G. Leslie, A. J. Miller, P. J. Haley and C. J. Lozano, 1996b. Real-time regional forecasting. In *Modern Approaches to Data Assimilation in Ocean Modeling*, P. Malanotte-Rizzoli, ed. Elsevier, Amsterdam, 455pp.

Sage, A. P. and J. L. Melsa, 1971. *Estimation Theory with Applications to Communications and Control.* McGraw-Hill, New York, 529pp.

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

Schröter, J. and C. Wunsch, 1986. Solution of nonlinear finite difference ocean models by optimization methods with sensitivity and observational strategy analysis. *J. Phys. Oceanogr.*, **16,** 1855–1874.

Sheinin, D. A. and G. L. Mellor, 1994. Predictability studies with a coastal forecast system for the U.S. east coast. *EOS, Trans. Am. Geophys. Union, Spring Meet.*, **75**(16), April 19, Suppl., p. 197.

Shulman, I. and K. J. Lewis, 1994. Modeling open boundary conditions by using the optimization approach. *COAM Report Tr-1/95*, pp. 1–19.

Sloan, N. Q., 1996. Dynamics of a shelf/slope front: process studies and data-driven simulations. Ph.D. thesis, Division of Applied Sciences, Harvard University, Cambridge, Mass.

Smedstad, O. M. and J. J. O'Brien, 1991. Variational data assimilation and parameter estimation in an equatorial Pacific Ocean model. *Prog. Oceanogr.*, **26,** 179–241. Pergamon Press, Elmsford, N.Y.

Smith, N. R., 1993. Ocean modeling in a global ocean observing system. *Rev. Geophys.*, **31**(3), 281–317.

Song, Y. and R. W. Lardner, submitted. Optimal estimation of eddy viscosity and friction coefficients for a shallow sea coastal model.

Spall, M. A. and W. R. Holland, 1991. A nested primitive equation studies of the Gulf Stream meander and ring formation region. *J. Phys. Oceanogr.*, **20,** 985–1016.

Stauffer, D. R. and J. W. Bao, 1993. Optimal determination of nudging coefficients using the adjoint equations, *Tellus*, Ser. A, Special Issue on Adjoint Application to Dynamic Meteorology, pp. 358–370.

Stommel, H., 1965. *The Gulf Stream*, 2nd ed. University of California Press, Berkeley, 248pp.

Sverdrup, H. V., M. W. Johnson and R. H. Fleming, 1942. *The Oceans: Their Physics, Chemistry, and General Biology*. Prentice Hall, Upper Saddle River, N.J., 1087pp.

Tellus, 1993. Series A: Dynamic meteorology and oceanography. Special Issue on Adjoint Applications in Dynamic Meteorology. *Tellus*, **45A**(5), 341–569.

Tennekes, H. and J. L. Lumley, 1972. *A First Course in Turbulence*. MIT Press, Cambridge, Mass.

Tennekes, H., A. P. M. Baede and J. D. Opsteegh, 1987. *ECMWF Workshop on Predictability in the Medium and Extended Range*. ECMWF, Reading, Berkshire, England, 415pp.

Thacker, W. G. and R. B. Long, 1988. Fitting dynamics to data. *J. Geophys. Res.*, **93,** 1227–1240.

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

Todling, R. and M. Ghil, 1990. Kalman filtering for a two-layer two-dimensional shallow-water model. In *Proc. International Symposium on Assimilation and Observations in Meteorology and Oceanography*, pp. 454–459.

Todling, R. and M. Ghil, 1994. Tracking atmospheric instabilities with the Kalman filter. I. Methodology and one-layer results. *Mon. Weather Rev.*, **122,** 183–204.

Tziperman, E., W. C. Thacker, R. B. Long and S.-M. Hwang, 1992a. Oceanic data analysis using a general circulation model. I. Simulations. *J. Phys. Oceanogr.*, **22,** 1434–1457.

Tziperman, E., W. C. Thacker, R. B. Long, S.-M. Hwang and S. R. Rintoul, 1992b. Oceanic data analysis using a general circulation model. II. A North Atlantic model. *J. Phys. Oceanogr.*, **22,** 1458–1485.

van Leeuwen, P. J. and G. Evensen, in press. Data assimilation and inverse methods in terms of a probabilistic formulation. *Month. Weath. Rev.*

von Schwind, J. J., 1980. *Geophysical Fluid Dynamics for Oceanographers*. Prentice Hall, Upper Saddle River, N.J., 307pp.

Walstad, L. J. and A. R. Robinson, 1990. Hindcasting and forecasting of the POLY-MODE dataset with the Harvard Open Ocean Model. *J. Phys. Oceanogr.*, **20**(11), 1682–1702.

Walstad, L. J., J. S. Allen, P. M. Kosro and A. Huyer, 1991. Dynamics of the coastal transition zone through data assimilation studies. *J. Geophys. Res.*, **96**(C8), 14959–14977.

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

Wunsch, C., 1996. *The Ocean Circulation Inverse Problem*. Cambridge University Press, Cambridge, 456pp.

Yanagi, T., T. Yamamoto, Y. Koizumi, T. Ikeda, M. Kamizono and H. Tamori, 1995. A numerical simulation of red tide formation. *J. Mar. Syst.*, **6,** 269–285.

Yavin, Y., 1985. Lecture notes in control and information sciences.In *Numerical Studies in Nonlinear Filtering*, Vol. 65. Springer-Verlag, New York, 273pp.

Yu, L. and J. J. O'Brien, 1991. Variational estimation of the wind stress drag coefficient and the oceanic eddy viscosity profile. *J. Phys. Oceanogr.*, **21,** 709–719.

Zheng, Quanan, Xiao-Hai and Yan V. Klemas, 1993. Statistical and dynamical analysis of internal waves on the continental shelf of the Middle Atlantic Bight from Space Shuttle photographs. *J. Geophys. Res.*, **98**(C5), 8495–8504.

Zou, X., I. M. Navon and F. X. LeDimet, 1992. An optimal nudging data assimilation scheme using parameter-estimation. *Q. J. R. Meteorol. Soc.*, **118,** 1163–1186.

Zou, J., W. W. Hsieh and I. M. Navon, 1995. Sequential open-boundary control by data assimilation in a limited-area model. *Mon. Weather Rev.*, **123,** 2899–2909.

Zupanski, M., 1993. Regional four-dimensional variational data assimilation in a quasi-operational forecasting environment. *Mon. Weather Rev.*, **121,** 2396–2408.