Regional Ocean Data Assimilation

Christopher A. Edwards,[1] Andrew M. Moore,[1] Ibrahim Hoteit,[2] and Bruce D. Cornuelle[3]

[1]Department of Ocean Sciences, University of California, Santa Cruz, California 95064; email: cedwards@ucsc.edu

[2]King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia

[3]Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093-0230

Annu. Rev. Mar. Sci. 2015. 7:21–42

First published online as a Review in Advance on August 6, 2014

The *Annual Review of Marine Science* is online at marine.annualreviews.org

This article's doi: 10.1146/annurev-marine-010814-015821

Keywords

data assimilation, ocean modeling, ecosystem modeling

Abstract

This article reviews the past 15 years of developments in regional ocean data assimilation. A variety of scientific, management, and safety-related objectives motivate marine scientists to characterize many ocean environments, including coastal regions. As in weather prediction, the accurate representation of physical, chemical, and/or biological properties in the ocean is challenging. Models and observations alone provide imperfect representations of the ocean state, but together they can offer improved estimates. Variational and sequential methods are among the most widely used in regional ocean systems, and there have been exciting recent advances in ensemble and four-dimensional variational approaches. These techniques are increasingly being tested and adapted for biogeochemical applications.

1. INTRODUCTION

In marine science, data assimilation (DA) refers to any process by which a model representing the ocean state (i.e., the time evolution of physical, chemical, and/or biological fields) is constrained rigorously by observations to reduce discrepancies between model output and data. Realistically configured but unconstrained numerical models are often statistically realistic, with broad measures of field means and variances loosely consistent with observational estimates, but there is also substantial disagreement when modeled and observed variables are compared directly at specific times and locations. Such discrepancies are fundamental and unavoidable. By definition, ocean models are imperfect representations of the natural system, involving both simplifying mathematical parameterizations for complex processes and errors associated with a variety of factors, such as finite model grid resolution. Even with a perfect model, errors result from uncertainties in model initial conditions and forcing, and inherent nonlinearities in underlying dynamics cause initially small discrepancies to grow. Of course, observations also offer an imperfect realization of the true ocean state owing to measurement error and the use of point measurements to represent broader ocean conditions. DA methods mathematically estimate changes to a control vector consisting of adjustable variables or parameters that control model predictions, so that the resulting model-data misfit is more consistent with model and observation uncertainties than an unconstrained model solution.

Alternatively, DA can be viewed as a dynamical approach to interpolating and extrapolating sparse oceanic data. Various functions can be used to interpolate observations separated in time or space. Linear interpolation, for example, is straightforward but not generally sensible for the ocean, as it does not account for correlations in fields across space and time resulting from geophysically relevant phenomena such as mesoscale eddies. By embodying ocean dynamics, an ocean model accounts for such correlations and should be a better interpolant, producing a more reliable ocean state estimate in places for which no data exist.

The basic DA problem resembles the well-known problem of fitting a line to data. Although many lines can be drawn through a cluster of points, a reasonable choice for a best solution is the pair of values representing the slope and intercept that minimizes the overall discrepancy between model values and data. Minimization can be defined in various ways (e.g., reduction of the mean absolute deviation between model predictions and data) but is most often performed in a least-squares sense, minimizing the sum of the squares of model-data differences. The ocean DA problem is similar in that it also seeks to minimize residuals between model predictions and observations, but different challenges result from a considerably larger control vector dimension and limited sampling, yielding a woefully underdetermined system, and from correlations between elements of the control vector. For example, when using a control vector consisting only of initial conditions, a model that has 500×500 horizontal grid points with 50 vertical levels and five state variables (two horizontal velocities, temperature, salinity, and sea surface height) yields a dimension of more than 5×10^7. If surface forcing or the model state at other times is to be adjusted as well, the state-vector dimension increases further. Although remotely sensed estimates of some surface variables are plentiful, subsurface information is extremely scarce, resulting in a vector of observations that is typically orders of magnitude smaller. Different approaches to handling these challenges have led to a variety of techniques that differ in methodology but share a similar goal.

The first DA efforts occurred in meteorology, carried out by Lewis Fry Richardson during World War I (Shuman 1989). However, the operations proved challenging, not only because finite difference calculations were computed by hand but also because of the then-unknown effects of numerical instabilities. It was not until the mid-twentieth century, following World War II, that

numerical weather prediction was born as its own discipline and major advances began. Developments in oceanography emerged in earnest in the 1980s, with influential papers documenting early forecast and DA attempts (Robinson et al. 1984, 1986; Robinson & Leslie 1985; De Mey & Robinson 1987). Multiple books and book chapters have described important methodological approaches and major contributions (Ghil & Malanotte-Rizzoli 1991, Bennett 1992, Malanotte-Rizzoli 1996, Wunsch 1996, Robinson et al. 1998). The present article is intended not to review the early history again but rather to focus on major developments over the past 15 years, a period marked by substantial observational, methodological, and computational advances that have resulted in the now-routine practice of ocean DA on both global and regional scales. To limit the scope of the article, we address primarily regional models.

Why perform regional DA in the ocean? Similar methods are used at global and regional scales, but, as in unconstrained ocean modeling, global and regional efforts are directed at different scales of motion and scientific problems. Today, global ocean models span ocean basins at resolutions of roughly 1/12° to 1° (e.g., Carton & Giese 2008, Chassignet et al. 2009, Cummings et al. 2009, Wunsch et al. 2009, Enomoto et al. 2013), with similarly coarse atmospheric forcing; among other goals, global ocean state estimates identify large-scale transport pathways, constrain upper-ocean heat content, and quantify meridional oceanic heat flux. By contrast, regional ocean models resolve 1–10-km features, usually driven by similarly resolved atmospheric models. Such high resolution is required for coastal applications, in which continental slopes and shelves vary on distances of kilometers and internal scales of motion in both the atmosphere and the ocean decrease rapidly as the distance to the coastal boundary diminishes. Although regional models can contribute to global efforts, they are generally motivated by more local issues of along- and cross-shelf transport for pollutant dispersal or search and rescue operations and by interdisciplinary problems such as hypoxia, coastal fishery management, harmful algal bloom development and dispersal, and sediment transport. Present-day computational limitations prohibit calculating basin-scale dynamics at resolutions required for regional applications.

One major development of the past decade was the advancement of operational oceanography and, specifically, the implementation of ocean observing systems that encompass observations, models, and analysis to yield societally relevant oceanographic information in near real time. The Global Ocean Observing System is a collaborative, international effort that includes both global and coastal foci. Regional ocean observing systems have emerged throughout the world, including in Europe, many Asian countries, Australia, and the Americas. Complex collaborations between universities, government agencies, and private organizations contribute complementary functionality to their success. Ocean observing systems include a vast array of in situ and remotely sensed measurements of ocean properties; the present platforms that distinguish coastal from global operations are glider lines that collect subsurface hydrography (Rudnick et al. 2004) and high-frequency radar estimates of surface currents (Paduan & Washburn 2013). Regional DA is central to the goal of providing accurate marine information in near real time by synthesizing sparse observations and model estimates to yield best estimates of the ocean state.

Regional DA has one unavoidable challenge relative to its global counterpart: open boundary conditions. By definition, a regional ocean model includes open boundaries over at least a portion of its perimeter, which formally results in a mathematically ill-posed problem (Oliger & Sundström 1978). Various approaches have been developed over the decades to address this issue for both barotropic (Flather 1976, Orlanski 1976, Chapman 1985, Mason et al. 2010) and baroclinic (Marchesiello et al. 2001) dynamics by approximating boundary values through a combination of prescribed fields from a larger-domain model (or climatology) and specified or diagnosed wave motion within the model domain. Near boundaries, regions of enhanced viscosity and diffusivity (sponges) are widely used to damp energetic features that develop over model integrations, and

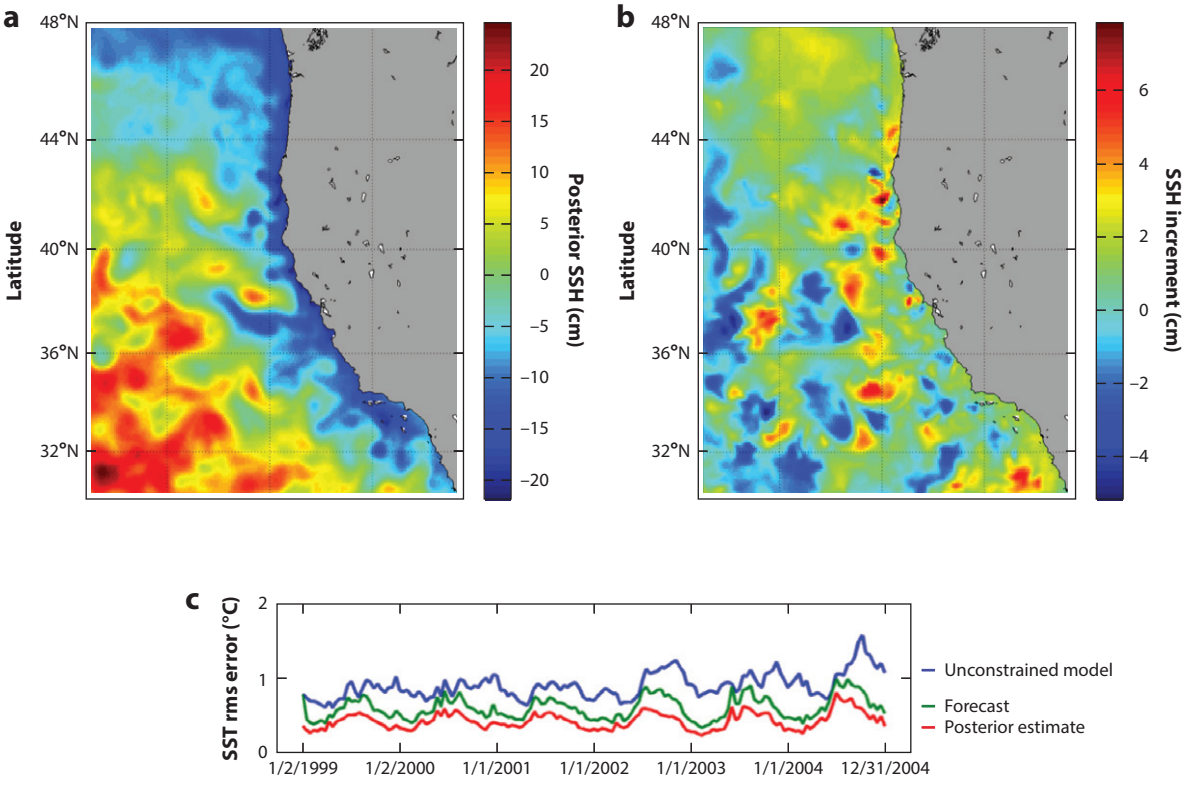

Figure 1

An example result from a near-real-time, four-dimensional variational system operating over the California Current System, assimilating data such as sea surface height (SSH), sea surface temperature (SST), tide gauge sea level, and in situ hydrography from a coastal glider (Moore et al. 2013). (*a*,*b*) Posterior SSH (panel *a*) and SSH increments (panel *b*) from May 20, 2013, resulting from assimilation. Corrections are predominantly mesoscale, with small amplitudes relative to the mean structure. (*c*) SST root-mean-square (rms) errors for an unconstrained model (*top*), posterior estimate (*bottom*), and forecast (*center*) from a series of 14-day assimilation cycles in a six-year reanalysis using a methodology similar to that used in panels *a* and *b*. Assimilation reduces the rms error to approximately half that of the free run, and skill is sustained for at least two weeks beyond the assimilation window. Panel *c* adapted from Broquet et al. (2009) with permission from Elsevier.

fields are sometimes relaxed to climatological conditions. Although these procedures are commonplace, offline nesting is imperfect. Experience has shown that boundary artifacts are reduced when two-way nesting is implemented and boundary gradients are determined dynamically on the model time step (Debreu et al. 2012). DA has the potential to alter the boundary information substantially if the control vector includes boundary values (Moore et al. 2011a). However, at present, such systems are designed to improve low-frequency boundary estimates, and their impact on wave reflection is not well established.

Figure 1 presents an example result from a DA experiment using a four-dimensional variational (4D-Var) approach over the California Current System, showing an estimate of sea surface height following assimilation (**Figure 1*a***) along with adjustments to the initial guess that were calculated to obtain the estimate (**Figure 1*b***). Various measures of goodness of fit are possible. Formally, for Gaussian distributed data, the cost function is a chi-squared variable with known degrees of freedom (Wunsch 1996, Bennett 2002), and hypothesis testing can determine whether the estimate is consistent with observations and assumed errors. Although the normalized cost function

is presented in some cases (e.g., Broquet et al. 2009), more often a straightforward calculation of the root-mean-square (rms) difference between the model and observations is determined. Useful information can be drawn when this calculation uses observations that were assimilated, but evaluation against unassimilated observations provides a more independent performance evaluation. **Figure 1*c*** shows the rms error in sea surface temperature from a six-year reanalysis by Broquet et al. (2009); overall, assimilation reduces the rms error to approximately half that of the free run, and the forecast error (calculated with independent data) reveals that skill is sustained for at least two weeks beyond the assimilation window.

This review is organized as follows. Section 2 describes the development and implementation of common methods for assimilating physical oceanographic data. This section is divided into two subsections that focus respectively on variational and sequential methods; although other methods exist, we address only 3D- and 4D-Var approaches in Section 2.1 and optimum interpolation (OI) and Kalman filter (KF) methods in Section 2.2. The past decade has also seen a dramatic rise in biological DA efforts, which we review in Section 3.

2. METHODS AND REGIONAL APPLICATIONS ASSIMILATING PHYSICAL DATA

The foundations of DA are encapsulated by Bayes' theorem (Wikle & Berliner 2007), although other independent approaches lead to the same general result. If we denote the ocean state vector by $\mathbf{x}$ and the vector of observations by $\mathbf{y}$, then Bayes' theorem states that

$$p(\mathbf{x}\,|\,\mathbf{y}) = \frac{p(\mathbf{y}\,|\,\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}, \tag{1}$$

where $p(\mathbf{x}\,|\,\mathbf{y})$ denotes the posterior conditional probability distribution for $\mathbf{x}$ given the observations and represents an update of the prior distribution $p(\mathbf{x})$ for $\mathbf{x}$, $p(\mathbf{y}\,|\,\mathbf{x})$ is referred to as the data distribution (or likelihood) and is the distribution of the observation errors given the prior estimate of $\mathbf{x}$, and $p(\mathbf{y})$ is the so-called marginal distribution and represents a normalizing constant to ensure the necessary condition that the integral of $p(\mathbf{x}\,|\,\mathbf{y})$ is unity. Assuming a normal distribution for the prior and observation errors, $p(\mathbf{y}\,|\,\mathbf{x}) = \exp\{-[\mathbf{y} - H(\mathbf{x})]^T \mathbf{R}^{-1}[\mathbf{y} - H(\mathbf{x})]\}$, where H is the observation operator that maps the prior into the observation space and $\mathbf{R}$ is the observation error covariance matrix. Similarly, $p(\mathbf{x}) = \exp[-(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b)]$, where $\mathbf{x}_b$ represents the prior estimate for $\mathbf{x}$ (also called the background) and $\mathbf{B}$ is the prior error covariance matrix (also called the background error covariance matrix). From Equation 1, the posterior distribution $p(\mathbf{x}\,|\,\mathbf{y}) \propto \exp(-J)$, where $J = (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) + [\mathbf{y} - H(\mathbf{x})]^T \mathbf{R}^{-1}[\mathbf{y} - H(\mathbf{x})]$. In this context, DA becomes a problem of identifying the state vector $\mathbf{x}_a$ that maximizes $p(\mathbf{x}\,|\,\mathbf{y})$, which of course is equivalent to minimizing J. As outlined below, both variational and sequential methods share this foundation in approaching the DA problem, with important differences in their methods of solutions and their characterization of the prior error covariance.

2.1. Variational Methods

In optimal control theory, J is referred to as the cost function, and methods based on the calculus of variations naturally lend themselves to identifying the cost function minimum. Variational approaches in geophysical contexts are divided into 3D and 4D problems.

Generally, the role of the observation operator H is to interpolate $\mathbf{x}$ in space and time to the observation locations. If the observations are grouped closely in time, then as a first approximation, time interpolation can be ignored, and H is a map from the state space to the observation space.

If the observations take the form of direct measurements of the state variables (e.g., temperature, salinity, and velocity), then H takes the form of a space-interpolation operator. By contrast, if the observations are complicated nonlinear functions of the state variables, such as sound speed, then H may have a complicated functional form. In either case, the minimization of J involves the solution of a 3D-Var problem. In 3D-Var, the ocean is considered static, and only data spanning a sufficiently short time window are sensibly selected to contribute to the analysis.

As the observations become more widely separated in time, or if one considers a time window over which the ocean is not static, it becomes necessary to interpolate $\mathbf{x}$ in both space and time. This is typically best done using prior knowledge of the dynamics that govern the ocean circulation (e.g., advection and wave dynamics), which is provided by an ocean model. In this case, H takes the form of an ocean model appropriately sampled at the space-time locations of the observations. As before, H may include simple space interpolation as well as more complicated functional forms for mapping the state variable into the appropriate observation space at the observation time. The minimization of J in this case involves the solution of a 4D-Var problem. In 4D-Var, the analysis is influenced by data collected both in advance of and following the analysis time through explicitly modeled ocean dynamics.

Because H is generally nonlinear, the topology of J can be quite complicated and may, in fact, possess multiple minima. In practice, it is common to linearize the problem about the prior and identify the increments (Courtier et al. 1994). In this case, H is replaced by a linearized observation operator $\mathbf{H}$, and the resulting cost function $J = \delta\mathbf{x}^T\mathbf{B}^{-1}\delta\mathbf{x}+(\mathbf{H}\delta\mathbf{x}-\mathbf{d})^T\mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{x}-\mathbf{d})$—where $\delta\mathbf{x} = \mathbf{x} - \mathbf{x}_b$ is the state-vector increment and $\mathbf{d} = \mathbf{y} - H(\mathbf{x}_b)$ is a vector of innovations (differences between the background and observations)—is quadratic with a unique minimum. Direct minimization of the cost function for the linearized problem (identified by finding the stationary point $\partial J/\partial\delta\mathbf{x} = 0$) yields the well-known solution for the posterior solution, called the analysis (Daley 1991):

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}[\mathbf{y} - H(\mathbf{x}_b)]. \quad (2)$$

The posterior increment is given by $\delta\mathbf{x}_a = \mathbf{x}_a - \mathbf{x}_b = \mathbf{B}\mathbf{g}$, where $\mathbf{g} = \mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T+\mathbf{R})^{-1}[\mathbf{y}-H(\mathbf{x}_b)]$, which shows that the posterior increment lies completely within the space spanned by the prior error covariance matrix, $\mathbf{B}$ (Weaver 2013). Because of the critical role played by $\mathbf{B}$, one of the greatest challenges of both the 3D- and 4D-Var approaches is the specification of the prior error and observation error covariance matrices.

The posterior circulation in Equation 2 can also be written in terms of the Kalman gain matrix, $\mathbf{K}$:

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}[\mathbf{y} - H(\mathbf{x}_b)], \quad (3)$$

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} = (\mathbf{B}^{-1}+\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}. \quad (4)$$

Although $\mathbf{K}$ involves a sequence of matrix products, matrices are generally not used in practice (except for $\mathbf{R}$). Instead, $\mathbf{H}$, $\mathbf{H}^T$, and $\mathbf{B}$ represent models for the tangent linear dynamical operators, adjoint dynamical operators, and prior error covariance matrix, respectively. The cost function J is minimized iteratively, which amounts to an iterative solution of the inverse matrices in Equation 4. When the first equality in Equation 4 is used, the resulting variational problem is referred to as the dual formulation and involves an iterative search for $\mathbf{x}$ in the space spanned by the observations; when the second equality is used, the resulting problem is referred to as the primal formulation and involves an iterative search for $\mathbf{x}$ in the space spanned by the state vector. Given that the number of observations is generally orders of magnitude smaller than the ocean state vector, the dual formulation offers a potential computational advantage over the primal formulation.

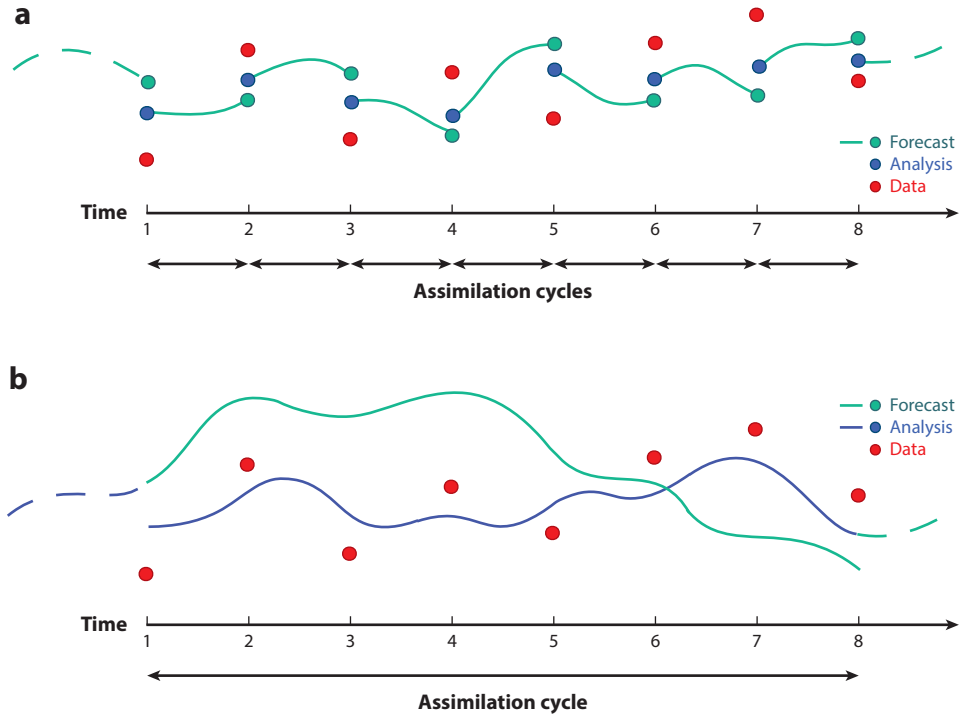

Figure 2

Schematic diagram of data assimilation procedures. (*a*) During the forecast step, three-dimensional variational (3D-Var), optimum interpolation, and Kalman filter methods integrate the nonlinear model forward until a datum or set of data becomes available, at which time an analysis is determined. Seven cycles are shown. (*b*) The four-dimensional variational (4D-Var) method uses all data available within an assimilation window to determine the analysis, which extends over the full cycle. One 4D-Var cycle is shown over a time window similar to that of the multiple 3D-Var cycles in panel *a*.

Regardless of the formulation, cost function minimization requires the repeated use of $\mathbf{H}$ and $\mathbf{H}^T$, which for 4D-Var is more computationally intensive than the nonlinear model itself. Thus, the computational cost of 4D-Var is considerably larger than that of 3D-Var (in practice between 10 and 100 times more costly, depending on the number of iterations). For the same computational capacity, 3D-Var applications can be practical at a higher resolution than 4D-Var applications.

It is common to run 3D-Var and 4D-Var sequentially, where $\mathbf{x}_b(t)$ is the solution of the forecast model initialized from the posterior from an earlier 3D- or 4D-Var calculation. Thus, $\mathbf{x}_b(t_{k+1}) = M\,[t_{k+1}, t_k, \mathbf{x}_a(t_k)]$, where t_k is the current time, t_{k+1} is a future time, and M represents the forecast ocean model. **Figure 2** schematically presents both 3D- and 4D-Var procedures. As shown, 3D-Var applications update the state vector as each datum (or data set collected in a narrow time window) becomes available. In practical regional applications, this update occurs on 12-h or daily timescales. 4D-Var, by contrast, exploits data collected over an extended time window [e.g., weekly or biweekly in Regional Ocean Modeling System (ROMS) applications and annually or on longer timescales in MIT General Circulation Model (MITgcm) applications].

If the model is assumed to be error free, then the posterior circulation estimate $\mathbf{x}_a$ is an exact solution of the model equations, and the resulting variational problem is said to be subject to the strong constraint (Sasaki 1970). Alternatively, if the inevitable errors in the model are accounted

for, then the resulting circulation estimate is said to be subject to the weak constraint, and $\mathbf{x}_a$ is no longer an exact solution of the model.

In the context of coastal oceanography, Li et al. (2008) have applied incremental, strong-constraint 3D-Var using the ROMS in the California Current System and in Prince William Sound in the Gulf of Alaska. The prior error covariance matrix in this system is approximated using the properties of the Kronecker product as described by Li et al. (2008). To address the differences in the scales resolved by satellite observations and in situ observations (such as those made by gliders and Argo profiling floats), Farrara et al. (2013) described a multiscale 3D-Var method involving a two-stage DA procedure that first minimizes the cost function to identify the large-scale circulation and then performs a second minimization to resolve smaller scales. Pan et al. (2011) have applied 3D-Var in the Navy Coastal Ocean Model to assimilate hydrographic data from the Monterey Bay. The authors used a hybrid approach to model the prior error covariance matrix **B**, where **B** is composed of a flow-dependent component derived from an ensemble and a static component based on a Laplacian smoother (Weaver & Courtier 2001). Dobricic & Pinardi (2008) described a 3D-Var scheme using the OPA (Océan Parallélisé) model (Madec et al. 1998) that is applied in the Mediterranean Sea as part of the Mediterranean Forecasting System. In this scheme, the vertical covariance component of **B** is modeled using empirical orthogonal functions (EOFs), whereas the horizontal covariance is modeled using a recursive filter (Lorenc 1992) and is allowed to vary in the presence of coastlines. More recently, Dobricic et al. (2014) have employed a Bayesian hierarchical modeling technique as a means of estimating flow-dependent vertical error covariance components of **B**.

4D-Var has also been applied in ROMS to estimate the circulation along the west coast of the United States. Kurapov et al. (2009, 2011) and Yu et al. (2012) have used the Advanced Variational Regional Ocean Representer Analyzer (AVRORA)—an incremental, strong-constraint, indirect representer formulation of dual 4D-Var (Egbert et al. 1994)—to recover the ocean circulation along the Oregon and Washington coasts. In a larger domain extending from near Vancouver Island in the north to midway down the Baja California peninsula in the south, Broquet et al. (2009, 2011) have applied an incremental, strong-constraint, primal 4D-Var formulation in the Rutgers University community ROMS code, and this approach is presently applied in near real time (Moore et al. 2013). More recent efforts include Moore et al.'s (2011b) application of the dual formulation to the California Current System and the computation of a sequence of historical analyses spanning the period 1980–2010 (Moore et al. 2013; E. Neveu, A.M. Moore, C.A. Edwards, J. Fiechter, P. Drake, et al., manuscript in preparation). Both AVRORA and the Rutgers ROMS model the prior error covariance using a pseudo-heat-diffusion equation, following the approach developed by Weaver & Courtier (2001). A dual 4D-Var system has also been developed for the Navy Coastal Ocean Model, and Ngodock & Carrier (2013) have used this system in the vicinity of Monterey to assimilate data from multiple platforms, including ocean gliders. This system also models **B** as the solution of a diffusion equation.

The community ROMS 4D-Var system described by Moore et al. (2011a) has been widely applied in other coastal and shelf-sea regions, including the Caribbean and the Gulf of Mexico (Powell et al. 2008, 2009), the New York Bight (Zhang et al. 2010a,b), the Mid-Atlantic Bight off the east coast of the United States (Wilkin & Hunter 2013, Chen et al. 2014, Zavala-Garay et al. 2014), the East Australia Current (Zavala-Garay et al. 2012), the Philippine Sea (Arango et al. 2010), the coastal Gulf of Alaska (Fiechter et al. 2011, Fiechter & Moore 2012), and the Hawaiian Islands (Matthews et al. 2012).

The MITgcm 4D-Var system employs a nonincremental, primal approach (i.e., in which the nonlinear cost function is minimized directly) (Stammer et al. 2003). It has been applied regionally in many locations, including the Azores Current region (Gebbie et al. 2006), the tropical Pacific

(Hoteit et al. 2010), and coastal environments such as the Southern California Bight (Hoteit et al. 2009), the California Current System (Todd et al. 2011), and the Gulf of Mexico (Gopalakrishnan et al. 2013).

2.2. Sequential Methods

Sequential DA methods assimilate data as they become available. As a result, state estimates are not constrained by future observations, as they are in 4D-Var. These methods originate from recursive Bayesian estimation theory but could also be viewed as time-recursive solutions of the variational DA methods, following linear estimation theory under the general conditions of linear dynamics and Gaussian-distributed errors (as discussed in Section 2.1).

2.2.1. Optimal interpolation. OI refers to 3D sequential DA methods that originate from linear statistical estimation theory and use a background covariance that does not change over time. The solution of this problem is equivalent to the 3D-Var problem, with the weights of the data and prior terms in the cost function being the inverse of the error covariance matrices for the observations and background state. OI was first described by Gandin (1965) and became the operational analysis scheme of choice during the 1980s and 1990s. It is still widely used in many research centers.

Although the solutions are identical to 3D-Var when the same assumptions are used and no approximations are made, OI methods involve direct matrix inversion for the solution of the problem (Daley 1991) and can also provide estimates of uncertainty. In the linear case, the OI solution is given by Equation 2, and the analysis error covariance matrix can be expressed as $\mathbf{P}_a = \mathbf{B}-\mathbf{KHB}$. When the observational operator is nonlinear, it is linearized about the prior solution $\mathbf{x}_b$.

OI methods are therefore dual formulations, inverting the stabilized representer matrix ($\mathbf{HBH}^T + \mathbf{R}$) in the observation space as in the first equality of Equation 4. When the number of observations to be assimilated is large, this inversion is computationally expensive, and approximations are employed. Two common approximations are the localization of the observation error covariance matrix and the use of low-rank forms of the background covariance.

Data selection is commonly used for efficient OI inversion. Limiting the number of observations that influence a given analysis point reduces the dimension of the stabilized representer matrix. Generally, background error correlations decrease with separation distance, reaching near-zero values within a finite distance. A cutoff radius beyond which observations are discarded is thus applied. However, this approach prevents the exact satisfaction of global constraints applied through the background **B**, because each analysis point is updated by different observations. In addition, the smoothness of the analysis is not guaranteed even when making geostrophic and hydrostatic assumptions, and the analysis must be filtered to prevent initial state imbalances from exciting spurious gravity wave activity that negatively impacts the forecast.

Several OI schemes have been introduced that differ only in the parameterization of **B** and its numerical representation. One of the most popular methods is the multivariate OI scheme (Daley 1991), which expresses **B** as the product of a correlation matrix and a diagonal-variance matrix. Multivariate OI forms the base of the DA component in the Navy Coupled Ocean Data Assimilation system (Smedstad et al. 2003). Other forms of multivariate OI have been also implemented, including in the Mediterranean Forecasting System (Pinardi et al. 2003), the Canada-Newfoundland Operational Ocean Forecasting System in the Northwestern Atlantic (**http://www.c-noofs.gc.ca**), and the global Forecasting Ocean Assimilation Model in the United Kingdom (Bell et al. 2000). The reduced-order information filter (Chin et al. 1999) uses a Markov random field to model the background **B** and has been applied for DA in the North Atlantic (Chin

et al. 2001). In the ensemble OI scheme (Oke et al. 2008a), **B** is calculated as the sample covariance matrix from a time-independent ensemble of anomalies. Lima et al. (2013) described an ensemble OI system off the Brazilian coast that used the Hybrid Coordinate Ocean Model and was based on OI experiments performed by Tanajura et al. (2013), and Halliwell et al. (2014) applied this method within an observing system simulation experiment in the Gulf of Mexico. Ensemble OI is very similar to the fixed-basis variant of the singular evolutive extended Kalman (SEEK) filter (Hoteit et al. 2002), which determines **B** through an EOF analysis of a stationary ensemble of anomalies. Ensemble OI and the SEEK filter have been applied in forecasting the North Atlantic and the Mediterranean as part of the MERCATOR operational assimilation system (Brasseur et al. 2005), in forecasting the Loop Current in the Gulf of Mexico (Counillon & Bertino 2009), and as part of the Bluelink Ocean Data Assimilation System in Australia (Oke et al. 2008b). Srinivasan et al. (2001) recently compared these particular OI schemes within the Gulf of Mexico and found that, given sufficient tuning, they performed similarly.

2.2.2. Kalman filter. The celebrated KF (Kalman 1960) is a sequential, 4D development of OI in which both the background and its error covariance matrix are updated at each assimilation time. Assuming an additive, unbiased model and uncorrelated, Gaussian-distributed observation errors, the KF provides the optimal way (in the least-squares sense) to sequentially assimilate linear observations into a linear dynamical system. When implemented using the same observations and background, and assuming linearity and no model error, the KF solution is equivalent to the 4D-Var solution at the end of the 4D-Var assimilation window (i.e., when only observations collected prior to the analysis time exist) (Li & Navon 2001). Thus, the KF solves the least-squares problem through a sequence of sequential 3D update steps, whereas 4D-Var solves the 4D assimilation problem over a broader time window generally equivalent to multiple KF update steps (**Figure 2**).

Applying the KF is a two-stage process consisting of a forecast and analysis. As in 4D-Var, during the first stage a forecast is made by integrating the dynamical model M forward in time to compute a prior solution $\mathbf{x}_b(t)$. However, an advantage of the KF over 4D-Var is that, during the second stage, the prior error covariance matrix is explicitly evolved forward in time. Specifically, $\mathbf{B}(t_{k+1}) = \mathbf{M}(t_{k+1}, t_k)\mathbf{P}_a(t_k)\mathbf{M}^T(t_{k+1}, t_k) + \mathbf{Q}(t_k)$, where $\mathbf{Q}(t_k)$ and $\mathbf{P}_a(t_k) = \mathbf{B}(t_k) - \mathbf{K}(t_k)\mathbf{H}(t_k)\mathbf{B}(t_k)$ are the model error and expected analysis error covariance matrices at time t_k, respectively. **B** is therefore flow dependent in the KF. In 4D-Var, the prior error covariance matrix is implicitly evolved in time, with the result that $\mathbf{B}(t_{k+1})$ is not available for use in future assimilation cycles. The KF posterior circulation estimate $\mathbf{x}_a$ is again computed using Equation 3, with **B** and **H** evaluated at the appropriate times. In this case, the observation operator H represents a single observation time, unlike in 4D-Var, where H includes the model M.

The KF cannot be used directly for DA into realistic ocean models because ocean dynamics (and some observations) are nonlinear and the state dimension can be very large, often making the calculation required to compute the KF error covariance matrices computationally intractable. As with other methods, model dynamics and the nonlinear observation operators can be linearized, which is implied in the above development. This linearization leads to the so-called extended Kalman filter (EKF) (Jazwinski 1970). To avoid the prohibitive computational requirements resulting from large state vectors, different forms of reduced-order state and low-rank error covariance matrix approximations have been proposed (e.g., Fukumori & Malanotte-Rizzoli 1995, Cane et al. 1996, Cohn & Todling 1996, Verlaan & Heemink 1997, Pham et al. 1998, Lermusiaux & Robinson 1999, Farrell & Ioannou 2001, Hoteit et al. 2002). A common feature of these simplified EKFs is that they exploit information from a representative subspace of the full ocean state space and ignore information from the less influential complement subspace. This simplification is supported by the dissipative and driven nature of the ocean dynamics, which concentrates energy at large scales,

meaning a red spectrum of variability (Daley 1991, Pham et al. 1998, Lermusiaux & Robinson 1999). Consequently, computations using these filters are conducted on the retained subspace, dramatically reducing the EKF computational cost.

Simplified EKFs have been applied for DA in the global ocean as part of the Estimating the Circulation and Climate of the Ocean–Global Ocean Data Assimilation Experiment (ECCO-GODAE) system (Kim et al. 2006), in the Pacific Ocean (Cane et al. 1996, Verron et al. 1999, Hoteit et al. 2002), and regionally in a nested implementation of the Ligurian Sea (Barth et al. 2007). They are also used operationally in the Greek national POSEIDON-II system for the Mediterranean (Korres et al. 2010) and in Monterey Bay (Haley et al. 2009). These methods have lost popularity in the past decade owing to the advances in ensemble Kalman filters (EnKFs).

2.2.3. Ensemble Kalman filters. Because the EKF provides only a first-order approximation of the KF (often computed numerically using finite differences to avoid the need for the tangent linear models and their adjoints), this algorithm may produce instabilities or not converge to a solution when applied to strongly nonlinear systems (Evensen 1992, Gauthier et al. 1993). The EnKF is another form of nonlinear Kalman filtering that was specifically introduced to overcome the KF limitations with realistic ocean models. The EnKF uses a Monte Carlo approach. At the analysis step of an EnKF, an ensemble of system states called the analysis ensemble, $\mathbf{X}_a = [\mathbf{x}_a^1 \mathbf{x}_a^2 \ldots]$, is generated with a sample mean and covariance equal to the KF analysis state and error covariance matrix, respectively. For typical oceanic applications, the ensemble size n is much smaller than the state dimension m_x. Propagating the analysis ensemble forward through the dynamical model determines a forecast ensemble for the next DA cycle (**Figure 3**): $\mathbf{x}_f^i(t_{k+1}) = \mathbf{M}(t_{k+1}, t_k)\mathbf{x}_a^i(t_k)$. The

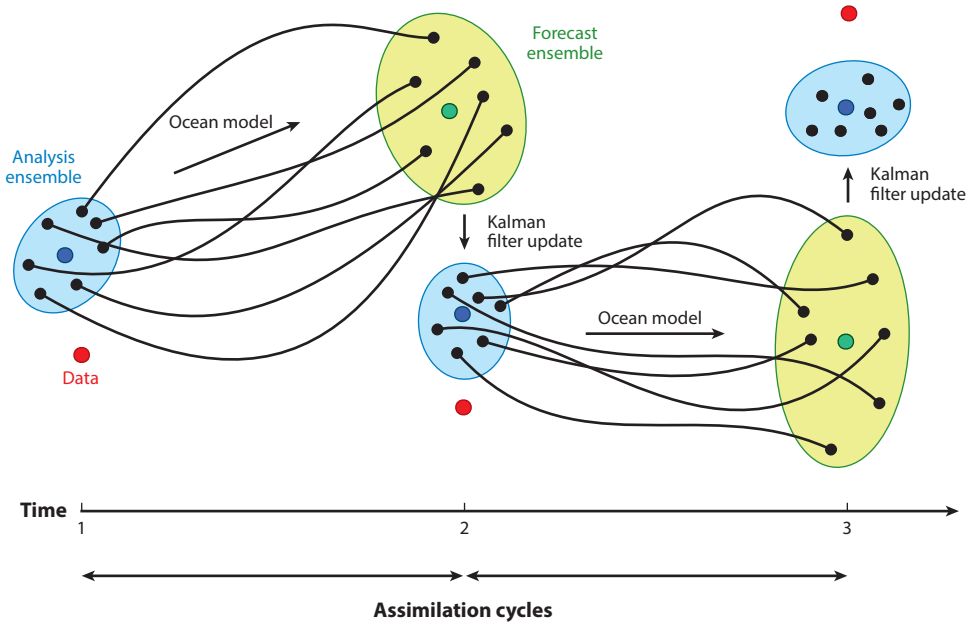

Figure 3

Schematic diagram of the ensemble Kalman filter procedure. During the forecast step, an ensemble of simulations (*black*) is integrated by the nonlinear ocean model until data are available. The forecast ensemble provides an estimate of the model error covariance, which is used in a Kalman filter update step to provide an analysis and a new analysis ensemble. Two assimilation cycles are shown.

forecast ensemble is used to estimate the background covariance matrix for this step: $\mathbf{B}(t_{k+1}) = [1/(n-1)]\mathbf{X}'_f(t_{k+1})\mathbf{X}'^T_f(t_{k+1})$, where $\mathbf{X}'_f(t_{k+1})$ consists of ensemble members $\mathbf{x}^i_f(t_{k+1})$ with the mean removed.

In this respect, the EnKF is similar to the reduced-rank nonlinear KF, but it is computationally more efficient in that it does not need to conduct (covariance) matrix factorization in the state space. For example, the computational cost of the prediction step is $\mathcal{O}(m_x^2 n)$ in a linear system, typically much lower than the order $\mathcal{O}(m_x^3)$ calculation of the original KF (Evensen 2006). Moreover, the EnKF does not require linearizing the models, making its implementation less intrusive. When a new observation becomes available, the KF update step is used to compute the analysis ensemble from its forecast counterpart based on the sample covariance matrix of the forecast ensemble. This forecast-analysis process is repeated as new observations are made available. EnKF methods may further benefit from the ensemble formulation to account for modeling uncertainties through perturbations of the parameters and inputs during the forecasting step of the ensemble members (Houtekamer et al. 2009). This allows the relaxation of the assumption of additive model error and the efficient treatment of some model uncertainties, which are otherwise difficult to include in the reduced-order EKFs (Pham et al. 1998).

Because of their ease of implementation, remarkable efficiency and robustness, and reasonable computational burden, sequential EnKF methods have become popular in many geophysical applications. Different EnKF variants have been developed in recent years. Based on whether the observations are perturbed before assimilation, these EnKF variants are customarily classified as one of two types (Tippett et al. 2003): stochastic EnKFs (Burgers et al. 1998, Houtekamer & Mitchell 2005) and deterministic EnKFs (Anderson 2001, Bishop et al. 2001, Whitaker & Hamill 2002, Hoteit et al. 2002). A stochastic EnKF essentially updates each forecast ensemble member with perturbed observations during the KF correction step. By contrast, a deterministic EnKF updates the ensemble mean and a specific square-root form of the sample error covariance matrix without perturbing the observations. An analysis ensemble is then produced as the sample mean plus the transformed background perturbations prior to the forecast step. The most popular deterministic EnKFs with publicly available codes are the singular evolutive interpolated KF (Pham 2001, Hoteit et al. 2002), ensemble transform KF (Bishop et al. 2001, Wang et al. 2004), and ensemble adjustment KF (Anderson 2001, Anderson et al. 2009), the last of which was developed under the umbrella of the Data Assimilation Research Testbed at the National Center for Atmospheric Research.

Implementing an EnKF for DA in realistic models is feasible only with relatively small ensembles. This limitation leads to some undesired effects, such as rank deficiency and underestimation of the sample error variances of the system state as well as overestimation of the corresponding cross-covariances (Whitaker & Hamill 2002, Hamill et al. 2009). It is customary to introduce covariance inflation and localization to mitigate these effects.

Covariance inflation is used to address variance underestimation (Anderson 2001), on the basis that relatively small ensembles and neglected model errors tend to cause underestimates of the sample error variances of the system state. In many situations, proper covariance inflation not only improves the estimation accuracy of the filter but also enhances its robustness from the point of view of robust filtering theory (Luo & Hoteit 2011). Various inflation methods have been proposed and studied (Anderson 2003, Hamill & Whitaker 2011, Miyoshi 2011, Whitaker & Hamill 2012).

Localization is applied to address rank deficiency and spuriously large cross-covariances between different state variables. Two popular localization methods are covariance localization (Hamill et al. 2001) and local analysis (Cohn et al. 1998). In covariance localization, a modified covariance matrix is obtained from the Schur product of the original covariance matrix and a tapering matrix based on the distances between the grid points of a physical model. In local analysis,

the whole state space is divided into a set of disjoint local analysis domains, and the system state in each is updated using only observations within a preset distance. Greybush et al. (2011), Janjić et al. (2011), and Sakov & Bertino (2011) have examined the relationship between these two techniques.

Important research is still being conducted to enhance the EnKF methods, understand their behavior and sensitivity to the assumption of Gaussian statistics, and determine the practical impact of limited ensemble size. Most EnKF-based regional DA systems are still in the development phase, with several of them moving toward using deterministic EnKFs to avoid the issue of observation error undersampling in the (stochastic) EnKF (Nerger et al. 2005). Keppenne & Rienecker (2002) presented one of the first applications of an EnKF in a general circulation ocean model, implementing a multivariate massively parallel EnKF for DA of temperature, salinity, and velocity measurements into a Pacific Ocean model. A comparison of different KF-based applications by Brusdal et al. (2003) demonstrated the efficiency of the EnKF. EnKF methods have also been successful in the Gulf of Mexico and with various circulation models: Counillon et al. (2009) implemented the stochastic EnKF with the Hybrid Coordinate Ocean Model, Hoteit et al. (2012) applied the Data Assimilation Research Testbed ensemble assimilation system with the MITgcm based on the ensemble adjustment KF, and Xu et al. (2013) investigated the ensemble transform EnKF using the Princeton Ocean Model. The EnKF is used operationally in the North Atlantic and Arctic Oceans (including ice) as part of the TOPAZ ocean forecasting project (Bertino & Lister 2008).

3. BIOLOGICAL DATA ASSIMILATION

Regional oceanographic interests extend well beyond physical fields. Oceanographers also seek to better understand and quantify biogeochemical distributions and processes by estimating ecosystem variables (e.g., phytoplankton biomass, community composition, pH, and oxygen content) and the fluxes governing their changes (e.g., primary production, grazing, respiration, air-sea carbon exchange, and organic matter export from surface waters). A wide variety of models have been developed over several decades for various research and management purposes. Marine ecosystem models usually include inorganic nutrients, primary producers, consumers, and nonliving organic matter; differences between models generally reside in the number of overall components and the parameterization of processes (e.g., linear as opposed to quadratic zooplankton mortality). Biological state variables are coupled to circulation models through tracer advection and diffusion.

Relative to DA for physical systems, biogeochemical data assimilation (BDA) must address various additional challenges. The number of tracers ranges from three to $\mathcal{O}(100)$, making the computational cost substantial. In optimal conditions, biological processes such as phytoplankton growth are rapid [with doubling times of $\mathcal{O}(1$ day)] compared with adjustments in mesoscale physical fields. As a result, linearity approximations embedded in some DA approaches are rapidly violated. Also, biogeochemical state variables represent concentrations that must be nonnegative (unlike velocity, for example). Biological variables often range over orders of magnitude, with near-zero means and large positive skew; thus, the statistics of their errors are distinctly non-Gaussian. Despite these issues, BDA has advanced dramatically over the past 20 years with the development of multiple approaches. Gregg (2008) presented a detailed review of early BDA efforts, and the remainder of this section emphasizes advancements primarily since that review.

Because biogeochemical models are based on simple parameterizations of complex processes measured in the field and laboratory (often with considerable scatter) rather than on fundamental laws (as in physical systems), these models include many poorly known parameters; it has long been recognized that DA may help to constrain such parameters (Matear 1995). From one perspective, parameter estimation should be less challenging than state estimation because the number of

parameters is orders of magnitude smaller than the dimension of biological variables. However, DA efforts in 1D systems have shown that even the simplest ecosystem models are underdetermined by presently available data and that improved model performance results from constraining only a small number of carefully chosen, uncorrelated parameters (Matear 1995, Friedrichs et al. 2006, Ward et al. 2010).

Variational methods have been popular for parameter estimation, but emulators now offer a computationally efficient alternative. An ensemble of forward simulations is calculated for a range of each parameter of interest, and a distance function that quantifies model-data misfit is calculated for each ensemble member. The optimal parameters correspond to the distance function minimum, which may occur at parameter values that are not included in the original ensemble but rather are represented through an expansion in the parameter space of model evolution by a limited number of basis functions. Computational efficiency results from the finite number of model runs that effectively span continuous parameter space through the interpolating functions. Once the emulator is created, it can be rapidly executed and applied repeatedly to data sets that span different spatial regions or temporal periods. Emulators have been demonstrated to be efficient alternatives for parameter optimization studies in the northern Gulf of Alaska (Hooten et al. 2011, Leeds et al. 2013) and Mid-Atlantic Bight (Mattern et al. 2012).

In biological state estimation problems, the large-dimensionality of the background error covariance matrix compared with that in physical assimilation problems leads to high computational costs and generally prohibits direct application of the full KF. The simplified SEEK filter (Pham et al. 1998), which develops a reduced-rank and spatially compact approximation of the full error covariance, has mitigated this issue in some biological applications. This method or variants have found wide use in 3D circulation models, including in the Pagasitikos Gulf of the Aegean Sea (Korres et al. 2012), the Sea of Crete (Hoteit et al. 2005), the eastern Mediterranean Sea (Triantafyllou et al. 2007), the French Mediterranean coast (Fontana et al. 2009, 2010), and the broader North Atlantic (Berline et al. 2007, Ourmières et al. 2009). Shulman et al. (2013) recently demonstrated the successful assimilation of bio-optical properties using an alternate EOF-based representation of the error covariance matrix in a sequential OI application within Monterey Bay.

Whereas the SEEK filter includes linearity assumptions from the original KF, this restriction is relaxed in the EnKF (as discussed above). In biological applications, ensembles can be generated through either state-vector or parameter adjustments. Successful implementations of biological localized EnKFs have been developed for a 1D application using Bermuda Atlantic Time-Series Study data (Mattern et al. 2010) and for 3D state estimation within the western English Channel (Ciavatta et al. 2011) and Mid-Atlantic Bight (Hu et al. 2012).

One specific challenge for BDA using 4D-Var is the requirement for tangent linear and adjoint models ($\mathbf{H}$ and $\mathbf{H}^T$, respectively). Although automated methods for model construction exist (Giering & Kaminski 1998), their use is nontrivial, and regional models such as ROMS code these models by hand. This labor-intensive process has no doubt limited the development of 4D-Var BDA systems, although biogeochemical systems that apply 4D-Var only to physical variables have been investigated (Fiechter et al. 2011). One remedy, proposed by Vermeulen & Heemink (2006), is a reduced 4D-Var implementation in which cost function minimization occurs in a model subspace determined by the leading EOFs from an ensemble of nonlinear model calculations. Adjoint and tangent linear models are written once and do not require recoding for different nonlinear models. Although the overall computational burden of the approach is not necessarily small, its advantage lies in the ease with which alternative models can be considered. Pelc et al. (2012) and Pelc (2013) demonstrated the viability of this EOF-based 4D-Var technique for marine ecosystems.

In the past decade, considerable attention has focused on the non-Gaussian error statistics characterizing BDA. Observationally, Campbell (1995) demonstrated that several bio-optical oceanic

variables (e.g., chlorophyll *a*) are well represented by lognormal distributions. A logarithm transformation of biogeochemical variables is a common approach in BDA and has the added benefit of ensuring nonnegative concentrations. One example came from Song et al. (2012), who implemented an incremental form of variational DA in a 1D model twin configuration, including a logarithm transformation developed for meteorological models by Fletcher & Zupanski (2006). Incremental logarithmic 4D-Var has been incorporated into ROMS, and this approach has been tested more fully in a realistic California Current System configuration that simultaneously assimilates both physical and biological data (H. Song, C.A. Edwards, A.M. Moore & J. Fiechter, manuscripts in preparation).

Although the logarithm transformation is analytically convenient, the methods applying this step remain arguably suboptimal because biogeochemical variables are not truly lognormal. An alternative approach applies a transformation, empirically determined from nonlinear model results, that renders variables Gaussian. This so-called Gaussian anamorphosis is performed for each non-Gaussian variable at each point in the model domain prior to a KF update step, with a reverse transformation applied to the analysis. This approach has been tested in conjunction with EnKF methods in model twin and realistic North Atlantic experiments (Simon & Bertino 2009, Béal et al. 2010, Doron et al. 2013) and holds promise for future regional modeling.

Finally, so-called particle filter and related methods also eliminate the Gaussian assumption. As in the EnKF, an ensemble of simulations (each called a particle) is integrated through the nonlinear model, but here this is done for the purpose of representing the posterior probability density function without any assumption of its distribution. The analysis is determined from a new ensemble, generated by resampling the existing members using probabilities based on an estimated likelihood function. Simple 0D and 1D investigations have demonstrated the potential utility of these approaches for ecosystem assimilation (Losa et al. 2003; Dowd 2006, 2011; Mattern et al. 2010), and Mattern et al. (2013) presented an implementation in a 3D configuration of the Mid-Atlantic Bight. Improving the robustness and applicability of particle filters for DA into large-dimensional geophysical systems is currently an active area of research (Ades & van Leeuwen 2013).

4. SUMMARY

Driven by the twenty-first-century emergence of operational oceanography and a desire to maximally synthesize historical data sets, regional oceanographic DA has exhibited extraordinary growth over the past 15 years. Early methods based on OI, 3D-Var, and suboptimal KF-based methods have been in routine use in various coastal environments for years. More complex ensemble and 4D-Var approaches are increasingly mature, and advances in computational speed and capacity are enabling their application in near real time. Historical, regional reanalyses have been created and should offer much-needed insights into multidecadal changes in coastal environments. Although initially developed in oceanography for physical circulation models, DA methods are being successfully adapted to characterize biogeochemical properties or improve biogeochemical models through parameter estimation. New methods are also being developed that relax the fundamental linearity and statistical assumptions of existing DA procedures. The developments in biological DA have important implications for the future because additional coupled systems, such as synchronously nested ocean and ocean-atmosphere models, are on the horizon.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The authors gratefully acknowledge research support from the National Science Foundation (grant OCE-1061434), the Office of Naval Research (grant N00014-10-1-0476), and the National Oceanographic and Atmospheric Administration (grant NA10OAR4320156).

LITERATURE CITED

Ades M, van Leeuwen PJ. 2013. An exploration of the equivalent weights particle filter. *Q. J. R. Meteorol. Soc.* 139:820–40

Anderson JL. 2001. An ensemble adjustment Kalman filter for data assimilation. *Mon. Weath. Rev.* 129:2884–903

Anderson JL. 2003. A local least squares framework for ensemble filtering. *Mon. Weath. Rev.* 131:634–42

Anderson JL, Hoar T, Raeder K, Liu H, Collins N, et al. 2009. The Data Assimilation Research Testbed: a community facility. *Bull. Am. Meteorol. Soc.* 90:1283–96

Arango HG, Levin JC, Curchitser EN, Zhang B, Moore AM, et al. 2010. Development of a hindcast/forecast model for the Philippine Archipelago. *Oceanography* 24(1):58–69

Barth A, Alvera-Azcrate A, Beckers JM, Rixen M, Vandenbulcke L. 2007. Multigrid state vector for data assimilation in a two-way nested model of the Ligurian Sea. *J. Mar. Syst.* 65:41–59

Béal D, Brasseur P, Brankart JM, Ourmières Y, Verron J. 2010. Characterization of mixing errors in a coupled physical biogeochemical model of the North Atlantic: implications for nonlinear estimation using Gaussian anamorphosis. *Ocean Sci.* 6:247–62

Bell MJ, Forbes RM, Hines A. 2000. Assessment of the FOAM global data assimilation system for real-time operational ocean forecasting. *J. Mar. Syst.* 25:1–22

Bennett AF. 1992. *Inverse Methods in Physical Oceanography*. Cambridge, UK: Cambridge Univ. Press

Bennett AF. 2002. *Inverse Modeling of the Ocean and Atmosphere*. Cambridge, UK: Cambridge Univ. Press

Berline L, Brankart JM, Brasseur P, Ourmières Y, Verron J. 2007. Improving the physics of a coupled physical–biogeochemical model of the North Atlantic through data assimilation: impact on the ecosystem. *J. Mar. Syst.* 64:153–72

Bertino L, Lister K. 2008. The TOPAZ monitoring and prediction system for the Atlantic and Arctic Oceans. *J. Oper. Oceanogr.* 1:15–18

Bishop CH, Etherton BJ, Majumdar SJ. 2001. Adaptive sampling with ensemble transform Kalman filter. Part I: theoretical aspects. *Mon. Weath. Rev.* 129:420–36

Brasseur P, Bahurel P, Bertino L, Birol F, Brankart JM, et al. 2005. Data assimilation for marine monitoring and prediction: the MERCATOR operational assimilation systems and the MERSEA developments. *Q. J. R. Meteorol. Soc.* 131:3561–82

Broquet G, Edwards CA, Moore AM, Powell BS, Veneziani M, Doyle JD. 2009. Application of 4D-variational data assimilation to the California Current System. *Dyn. Atmos. Oceans* 48:69–92

Broquet G, Moore AM, Arango HG, Edwards CA. 2011. Corrections to ocean surface forcing in the California Current System using 4D-variational data assimilation. *Ocean Model.* 36:116–32

Brusdal K, Brankart JM, Halberstadt G, Evensen G, Brasseur P, et al. 2003. A demonstration of ensemble-based assimilation methods with a layered OGCM from the perspective of operational ocean forecasting systems. *J. Mar. Syst.* 40–41:253–89

Burgers G, van Leeuwen PJ, Evensen G. 1998. On the analysis scheme in the ensemble Kalman filter. *Mon. Weath. Rev.* 126:1719–24

Campbell JW. 1995. The lognormal distribution as a model for bio-optical variability in the sea. *J. Geophys. Res.* 100:13237–54

Cane MA, Kaplan A, Miller RN, Tang B, Hackert EC, Busalacchi AJ. 1996. Mapping tropical Pacific sea level: data assimilation via a reduced state Kalman filter. *J. Geophys. Res.* 101:22599–617

Carton JA, Giese BS. 2008. A reanalysis of ocean climate using simple ocean data assimilation (SODA). *Mon. Weather Rev.* 136:2999–3017

Chapman DC. 1985. Numerical treatment of cross-shelf open boundaries in a barotropic coastal ocean model. *J. Phys. Oceanogr.* 15:1060–75

Chassignet EP, Hurlburt HE, Metzger EJ, Smedstad OM, Cummings JA, et al. 2009. US GODAE: Global Ocean Prediction with the HYbrid Coordinate Ocean Model (HYCOM). *Oceanography* 22(2):64–75

Chen K, He R, Powell BS, Gawarkiewicz G, Moore AM, Arango HG. 2014. Data assimilative modeling investigation of Gulf Stream Warm Core Ring interaction with continental shelf and slope circulation. *J. Geophys. Res.* 119:5968–91

Chin TM, Haza AC, Mariano AJ. 2001. A reduced-order information filter for multi-layer shallow water models: profiling and assimilation of sea surface height. *J. Atmos. Ocean. Technol.* 19:517–33

Chin TM, Mariano AJ, Chassignet EP. 1999. Spatial regression and multi-scale approximations for sequential data assimilation in ocean models. *J. Geophys. Res.* 104:7991–8014

Ciavatta S, Torres R, Saux-Picart S, Allen JI. 2011. Can ocean color assimilation improve biogeochemical hindcasts in shelf seas? *J. Geophys. Res.* 116:C12043

Cohn SE, da Silva A, Guo J, Sienkiewicz M, Lamich D. 1998. Assessing the effects of data selection with the DAO physical-space statistical analysis system. *Mon. Weath. Rev.* 126:2913–26

Cohn SE, Todling R. 1996. Approximate data assimilation schemes for stable and unstable dynamics. *J. Meteorol. Soc. Jpn.* 74:63–75

Counillon F, Bertino L. 2009. Ensemble Optimal Interpolation: multivariate properties in the Gulf of Mexico. *Tellus A* 61:296–308

Counillon F, Sakov P, Bertino L. 2009. Application of a hybrid EnKF-OI to ocean forecasting. *Ocean Sci.* 5:389–401

Courtier P, Thépaut JN, Hollingsworth A. 1994. A strategy for operational implemenation of 4D-Var using an incremental approach. *Q. J. R. Meteorol. Soc.* 120:1367–88

Cummings JA, Bertino L, Brasseur P, Fukumori I, Kamachi M, et al. 2009. Ocean data assimilation systems for GODAE. *Oceanography* 22(3):96–109

Daley R. 1991. *Atmospheric Data Analysis*. Cambridge, UK: Cambridge Univ. Press

De Mey P, Robinson AR. 1987. Assimilation of altimeter eddy fields in a limited-area quasi-geostrophic model. *J. Phys. Oceanogr.* 17:2280–93

Debreu L, Marchesiello P, Penven P, Cambon G. 2012. Two-way nesting in split-explicit ocean models: algorithms, implementation and validation. *Ocean Model.* 49–50:1–21

Dobricic S, Pinardi N. 2008. An oceanographic three-dimensional variational data assimilation scheme. *Ocean Model.* 22:89–105

Dobricic S, Wikle CK, Milliff RF, Pinardi N, Berliner LB. 2014. Assimilation of oceanographic observations with estimates of vertical background-error covariances by a Bayesian hierarchical model. *Q. J. R. Meteorol. Soc.* In press. doi: 10.1002/qj.2348

Doron M, Brasseur P, Brankart JM, Losa SN, Melet A. 2013. Stochastic estimation of biogeochemical parameters from Globcolour ocean colour satellite data in a North Atlantic 3D ocean coupled physical-biogeochemical model. *J. Mar. Syst.* 117–118:81–95

Dowd M. 2006. A sequential Monte Carlo approach for marine ecological prediction. *Environmetrics* 17:435–55

Dowd M. 2011. Estimating parameters for a stochastic dynamic marine ecological system. *Environmetrics* 22:501–15

Egbert GD, Bennett AF, Foreman MGG. 1994. TOPEX/POSEIDON tides estimated using a global inverse method. *J. Geophys. Res.* 99:24,821–52

Enomoto T, Miyoshi T, Moteki Q, Inoue J, Hattori M, et al. 2013. Observing-system research and ensemble data assimilation at JAMSTEC. See Park & Xu 2013, pp. 509–26

Evensen G. 1992. Using the extended Kalman filter with a multilayer quasi-geostrophic ocean model. *J. Geophys. Res.* 97:17905–24

Evensen G. 2006. *Data Assimilation: The Ensemble Kalman Filter*. Berlin: Springer-Verlag

Farrara JD, Chao Y, Li Z, Wang X, Jin X, et al. 2013. A data-assimilative ocean forecasting system for the Prince William sound and an evaluation of its performance during sound Predictions 2009. *Cont. Shelf Res.* 63(Suppl.):S193–208

Farrell BF, Ioannou PJ. 2001. State estimation using a reduced-order Kalman filter. *J. Geophys. Res.* 58:3666–80

Fiechter J, Broquet G, Moore AM, Arango HG. 2011. A data assimilative, coupled physical-biological model for the Coastal Gulf of Alaska. *Dyn. Atmos. Oceans* 52:95–118

Fiechter J, Moore AM. 2012. Iron limitation impact on eddy-induced ecosystem variability in the coastal Gulf of Alaska. *J. Mar. Syst.* 92:1–15
Flather RA. 1976. A tidal model of the northwest European continental shelf. *Mem. Soc. R. Sci. Liege* 10:141–64
Fletcher SJ, Zupanski M. 2006. A data assimilation method for log-normally distributed observational errors. *Q. J. R. Meteorol. Soc.* 132:2505–19
Fontana C, Grenz C, Pinazo C. 2010. Sequential assimilation of a year-long time-series of SeaWiFS chlorophyll data into a 3D biogeochemical model on the French Mediterranean coast. *Cont. Shelf Res.* 30:1761–71
Fontana C, Grenz C, Pinazo C, Marsaleix P, Diaz F. 2009. Assimilation of SeaWiFS chlorophyll data into a 3D-coupled physical-biogeochemical model applied to a freshwater-influenced coastal zone. *Cont. Shelf Res.* 29:1397–409
Friedrichs MA, Hood RR, Wiggert JD. 2006. Ecosystem model complexity versus physical forcing: quantification of their relative impact with assimilated Arabian sea data. *Deep-Sea Res. II* 53:576–600
Fukumori I, Malanotte-Rizzoli P. 1995. An approximate Kalman filter for ocean data assimilation: an example with an idealized Gulf Stream model. *J. Geophys. Res.* 100:6777–93
Gandin LS. 1965. *Objective Analysis of Meteorological Fields*. Jerusalem: Israel Program Sci. Transl.
Gauthier P, Courtier P, Moll P. 1993. Assimilation of simulated wind lidar data with a Kalman filter. *Mon. Weath. Rev.* 121:1803–20
Gebbie G, Heimbach P, Wunsch C. 2006. Strategies for nested and eddy-permitting state estimation. *J. Geophys. Res.* 111:C10073
Ghil M, Malanotte-Rizzoli P. 1991. Data assimilation in meteorology and oceanography. In *Advances in Geophysics*, Vol. 33, ed. R Dmowska, B Saltzman, pp. 141–266. San Diego, CA: Academic
Giering R, Kaminski T. 1998. Recipes for adjoint code construction. *ACM Trans. Math. Softw.* 24:437–74
Gopalakrishnan G, Cornuelle BD, Hoteit I, Rudnick DL, Owens WB. 2013. State estimates and forecasts of the loop current in the Gulf of Mexico using the MITgcm and its adjoint. *J. Geophys. Res.* 118:3292–314
Gregg WW. 2008. Assimilation of SeaWiFS ocean chlorophyll data into a three-dimensional global ocean model. *J. Mar. Syst.* 69:205–25
Greybush SJ, Kalnay E, Miyoshi T, Ide K, Hunt BR. 2011. Balance and ensemble Kalman filter localization techniques. *Mon. Weath. Rev.* 139:511–22
Haley PJ Jr, Lermusiaux PFJ, Robinson AR, Leslie WG, Logutov O, et al. 2009. Forecasting and reanalysis in the Monterey Bay/California Current region for the Autonomous Ocean Sampling Network-II Experiment. *Deep-Sea Res. II* 56:127–48
Halliwell GR, Srinivasan A, Kourafalou V, Yang H, Willey D, et al. 2014. Rigorous evaluation of a fraternal twin ocean OSSE system for the open Gulf of Mexico. *J. Atmos. Ocean. Technol.* 31:105–30
Hamill TM, Whitaker JS. 2011. What constrains spread growth in forecasts initialized from ensemble Kalman filters? *Mon. Weath. Rev.* 139:117–31
Hamill TM, Whitaker JS, Anderson JL, Snyder C. 2009. Comments on "Sigma-point Kalman filter data assimilation methods for strongly nonlinear systems." *J. Atmos. Sci.* 66:3498–500
Hamill TM, Whitaker JS, Snyder C. 2001. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Weath. Rev.* 129:2776–90
Hooten MB, Leeds WB, Fiechter J, Wikle CK. 2011. Assessing first-order emulator inference for physical parameters in nonlinear mechanistic models. *J. Agric. Biol. Environ. Stat.* 16:475–94
Hoteit I, Cornuelle B, Heimbach P. 2010. An eddy-permitting, dynamically consistent adjoint-based assimilation system for the tropical Pacific: hindcast experiments in 2000. *J. Geophys. Res.* 115:C03001
Hoteit I, Cornuelle B, Kim SY, Forget G, Kohl A, Terrill E. 2009. Assessing 4D-VAR for dynamical mapping of coastal high-frequency radar in San Diego. *Dyn. Atmos. Oceans* 48:175–97
Hoteit I, Hoar T, Gopalakrishnan G, Anderson J, Collins N, et al. 2012. A MITgcm/DART ensemble analysis and prediction system with application to the Gulf of Mexico. *Dyn. Atmos. Oceans* 63:1–23
Hoteit I, Pham DT, Blum J. 2002. A simplified reduced order Kalman filtering and application to altimetric data assimilation in tropical Pacific. *J. Mar. Syst.* 36:101–27
Hoteit I, Triantafyllou G, Petihakis G. 2005. Efficient data assimilation into a complex, 3-D physical-biogeochemical model using partially-local Kalman filters. *Ann. Geophys.* 23:3171–85
Houtekamer PL, Mitchell HL. 2005. Ensemble Kalman filtering. *Q. J. R. Meteorol. Soc.* 131:3269–89

Houtekamer PL, Mitchell HL, Deng X. 2009. Model error representation in an operational ensemble Kalman filter. *Mon. Weath. Rev.* 137:2126–43

Hu J, Fennel K, Mattern JP, Wilkin JL. 2012. Data assimilation with a local ensemble Kalman filter applied to a three-dimensional biological model of the Middle Atlantic Bight. *J. Mar. Syst.* 94:145–56

Janjić T, Nerger L, Albertella A, Schröter J, Skachko S. 2011. On domain localization in ensemble-based Kalman filter algorithms. *Mon. Weath. Rev.* 139:2046–60

Jazwinski AH. 1970. *Stochastic Processes and Filtering Theory*. San Diego, CA: Academic

Kalman R. 1960. A new approach to linear filtering and prediction problems. *Trans. ASME D* 82:35–45

Keppenne C, Rienecker M. 2002. Initial testing of a massively parallel ensemble Kalman filter with the Poseidon Isopycnal Ocean General Circulation Model. *Mon. Weath. Rev.* 130:2951–65

Kim SB, Fukumori I, Lee T. 2006. The closure of the ocean mixed layer temperature budget using level-coordinate model fields. *J. Atmos. Ocean. Technol.* 23:840–53

Korres G, Nittis K, Perivoliotis L, Tsiaras K, Papadopoulos A, et al. 2010. Forecasting the Aegean Sea hydrodynamics within the POSEIDON-II operational system. *J. Oper. Oceanogr.* 38:37–49

Korres G, Triantafyllou G, Petihakis G, Raitsos D, Hoteit I, et al. 2012. A data assimilation tool for the Pagasitikos Gulf ecosystem dynamics: methods and benefits. *J. Mar. Syst.* 94(Suppl.):S102–17

Kurapov AL, Egbert GD, Allen JS, Miller RN. 2009. Representer-based analyses in the coastal upwelling system. *Dyn. Atmos. Oceans* 48:198–218

Kurapov AL, Foley D, Strub PT, Egbert GD, Allen JS. 2011. Variational assimilation of satellite observations in a coastal ocean model off Oregon. *J. Geophys. Res.* 116:C05006

Leeds WB, Wikle CK, Fiechter J, Brown J, Milliff RF. 2013. Modeling 3-D spatio–temporal biogeochemical processes with a forest of 1-D statistical emulators. *Environmetrics* 24:1–12

Lermusiaux P, Robinson A. 1999. Data assimilation via error subspace statistical estimation. Part I: theory and schemes. *Mon. Weath. Rev.* 127:1385–407

Li Z, Chao Y, McWilliams JC, Ide K. 2008. A three-dimensional variational data assimilation scheme for the Regional Ocean Modeling System. *J. Atmos. Ocean. Technol.* 25:2074–90

Li Z, Navon IM. 2001. Optimality of variational data assimilation and its relationship with the Kalman filter and smoother. *Q. J. R. Meteorol. Soc.* 127:661–83

Lima JAM, Martins RP, Tanajura CAS, de Moraes Paiva A, Cirano M, et al. 2013. Design and implementation of the Oceanographic Modeling and Observation Network (REMO) for operational oceanography and ocean forecasting. *Rev. Bras. Geofis.* 31:209–28

Lorenc A. 1992. Iterative analysis using covariance functions and filters. *Q. J. R. Meteorol. Soc.* 118:569–91

Losa SN, Kivman GA, Schröter J, Wenzel M. 2003. Sequential weak constraint parameter estimation in an ecosystem model. *J. Mar. Syst.* 43:31–49

Luo X, Hoteit I. 2011. Robust ensemble filtering and its relation to covariance inflation in the ensemble Kalman filter. *Mon. Weath. Rev.* 139:3938–53

Madec G, Delecluse P, Imbard M, Lévy C. 1998. *OPA 8.1 Ocean General Circulation Model Reference Manual*. Notes Pôle Modél. 11. Paris: Inst. Pierre Simon Laplace Sci. Environ. Glob.

Malanotte-Rizzoli P, ed. 1996. *Modern Approaches to Data Assimilation in Ocean Modeling*. Elsevier Oceanogr. Ser. 61. Amsterdam: Elsevier

Marchesiello P, McWilliams JC, Shchepetkin A. 2001. Open boundary conditions for long-term integration of regional oceanic models. *Ocean Model.* 3:1–20

Mason E, Molemaker J, Shchepetkin AF, Colas F, McWilliams JC, Sangrà P. 2010. Procedures for offline grid nesting in regional ocean models. *Ocean Model.* 35:1–15

Matear RJ. 1995. Parameter optimization and analysis of ecosystem models using simulated annealing: a case study at Station P. *J. Mar. Res.* 53:571–607

Mattern JP, Dowd M, Fennel K. 2010. Sequential data assimilation applied to a physical-biological model for the Bermuda Atlantic time series station. *J. Mar. Syst.* 79:144–56

Mattern JP, Dowd M, Fennel K. 2013. Particle filter-based data assimilation for a three-dimensional biological ocean model and satellite observations. *J. Geophys. Res.* 118:2746–60

Mattern JP, Fennel K, Dowd M. 2012. Estimating time-dependent parameters for a biological ocean model using an emulator approach. *J. Mar. Syst.* 96–97:32–47

Matthews D, Powell BS, Janeković I. 2012. Analysis of four-dimensional variational state estimation of the Hawaiian waters. *J. Geophys. Res.* 117:C03013

Miyoshi T. 2011. The Gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform Kalman filter. *Mon. Weath. Rev.* 139:1519–35

Moore AM, Arango HG, Broquet G, Powell BS, Weaver AT, Zavala-Garay J. 2011a. The Regional Ocean Modeling System (ROMS) 4-dimensional variational data assimilation systems. Part I: system overview and formulation. *Prog. Oceanogr.* 91:34–49

Moore AM, Arango HG, Broquet G, Edwards CA, Veneziani M, et al. 2011b. The Regional Ocean Modeling System (ROMS) 4-dimensional variational data assimilation systems. Part II: performance and application to the California Current System. *Prog. Oceanogr.* 91:50–73

Moore AM, Edwards CA, Fiechter J, Drake P, Arango HG, et al. 2013. A prototype for an operational regional ocean data assimilation system. See Park & Xu 2013, pp. 346–66

Nerger L, Hiller A, Schröter J. 2005. A comparison of error subspace Kalman filters. *Tellus A* 57:715–35

Ngodock H, Carrier M. 2013. A weak constraint 4D-Var assimilation system for the Navy Coastal Ocean Model using the representer method. See Park & Xu 2013, pp. 367–90

Oke PR, Allen JS, Miller RN, Egbert GD, Kosro PM. 2008a. Assimilation of surface velocity data into a primitive equation coastal ocean model. *J. Geophys. Res.* 107:3122

Oke PR, Brassington GB, Griffin DA, Schiller A. 2008b. The Bluelink Ocean Data Assimilation System (BODAS). *Ocean Model.* 21:46–70

Oliger J, Sundström A. 1978. Theoretical and practical aspects of some initial boundary value problems in fluid dynamics. *SIAM J. Appl. Math.* 35:419–46

Orlanski I. 1976. A simple boundary condition for unbounded hyperbolic flows. *J. Comput. Phys.* 21:251–69

Ourmières Y, Brasseur P, Lévy M, Brankart JM, Verron J. 2009. On the key role of nutrient data to constrain a coupled physical-biogeochemical assimilative model of the North Atlantic Ocean. *J. Mar. Syst.* 75:100–15

Paduan JD, Washburn L. 2013. High-frequency radar observations of ocean surface currents. *Annu. Rev. Mar. Sci.* 5:115–36

Pan C, Yaremchuk M, Nechaev D, Ngodock H. 2011. Variational assimilation of glider data in Monterey Bay. *J. Mar. Res.* 69:331–46

Park SK, Xu L, eds. 2013. *Data Assimilation for Atmospheric, Oceanic and Hydrological Applications*, Vol. 2. Berlin: Springer-Verlag

Pelc JS. 2013. *Data assimilation for marine ecosystem models*. PhD Thesis, Delft Univ. Technol., Delft, Neth.

Pelc JS, Ehouarn S, Bertino L, El Serafy G, Heemink AW. 2012. Application of model reduced 4D-Var to a 1D ecosystem model. *Ocean Model.* 57–58:16

Pham DT. 2001. Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Mon. Weath. Rev.* 129:1194–207

Pham DT, Verron J, Roubaud MC. 1998. Singular evolutive Kalman filter with EOF initialization for data assimilation in oceanography. *J. Mar. Syst.* 16:323–40

Pinardi N, Allen I, Demirov E, De Mey P, Korres G, et al. 2003. The Mediterranean ocean forecasting system: first phase of implementation (1998–2001). *Ann. Geophys.* 21:3–20

Powell BS, Arango HG, Moore AM, Di Lorenzo E, Milliff RF, Foley D. 2008. 4DVAR data assimilation in the Intra-Americas Sea with the Regional Ocean Modeling System (ROMS). *Ocean Model.* 23:130–45

Powell BS, Moore AM, Arango HG, Lorenzo ED, Milliff RF, Leben R. 2009. Near real-time assimilation and prediction in the Intra-Americas Sea with the Regional Ocean Modeling System (ROMS). *Dyn. Atmos. Oceans* 48:46–68

Robinson AR, Carton JA, Mooers CNK, Walstad LJ, Carter EF, et al. 1984. A real-time dynamical forecast of ocean synoptic/mesoscale eddies. *Nature* 309:781–83

Robinson AR, Carton JA, Pinardi N, Mooers CNK. 1986. Dynamical forecasting and dynamical interpolation: an experiment in the California Current. *J. Phys. Oceanogr.* 16:1561–79

Robinson AR, Lermusiaux PFJ, Sloan NQ III. 1998. Data assimilation. In *The Sea*, Vol. 10: *The Global Coastal Ocean: Processes and Method*, ed. KH Brink, AR Robinson, pp. 541–94. New York: Wiley & Sons

Robinson AR, Leslie WG. 1985. Estimation and prediction of ocean eddy fields. *Prog. Oceanogr.* 14:485–510

Rudnick DL, Davis RE, Eriksen CC, Fratantoni DM, Perry MJ. 2004. Underwater gliders for ocean research. *Mar. Technol. Soc. J.* 38:73–84

Sakov P, Bertino L. 2011. Relation between two common localisation methods for the EnKF. *Comput. Geosci.* 15:225–37
Sasaki Y. 1970. Some basic formulations in numerical variational analysis. *Mon. Weather Rev.* 98:875–83
Shulman I, Frolov S, Anderson S, Penta B, Gould R, et al. 2013. Impact of bio-optical data assimilation on short-term coupled physical, bio-optical model predictions. *J. Geophys. Res.* 118:2215–30
Shuman FG. 1989. History of numerical weather prediction at the National Meteorological Center. *Weath. Forecast.* 4:286–96
Simon E, Bertino L. 2009. Application of the Gaussian anamorphosis to assimilation in a 3-D coupled physical-ecosystem model of the North Atlantic with the EnKF: a twin experiment. *Ocean Sci.* 5:495–510
Smedstad OM, Hurlburt HE, Metzger EJ, Rhodes RC, Shriver JF, et al. 2003. An operational eddy-resolving 1/16° global ocean nowcast/forecast system. *J. Mar. Syst.* 40–41:341–61
Song H, Edwards CA, Moore AM, Fiechter J. 2012. Incremental four-dimensional variational data assimilation of positive-definite oceanic variables using a logarithm transformation. *Ocean Model.* 54–55:1–17
Srinivasan A, Chassignet EP, Bertino L, Brankart JM, Brasseur P, et al. 2001. A comparison of sequential assimilation schemes for ocean prediction with the HYbrid Coordinate Ocean Model (HYCOM): twin experiments with static forecast error covariances. *Ocean Model.* 37:85–111
Stammer D, Wunsch C, Giering R, Eckert C, Heimbach P, et al. 2003. Volume, heat, and freshwater transports of the global ocean circulation 1993–2000, estimated from a general circulation model constrained by World Ocean Circulation Experiment (WOCE) data. *J. Geophys. Res.* 108:3007–29
Tanajura CAS, Costa FB, da Silva RR, Ruggiero GA, Daher VB. 2013. Assimilation of sea surface height anomalies into HYCOM with an optimal interpolation scheme over the Atlantic Ocean METAREA V. *Rev. Bras. Geofis.* 31:257–70
Tippett M, Anderson J, Bishop C, Hamill T, Whitacker J. 2003. Ensemble square root filters. *Mon. Weath. Rev.* 131:1490–85
Todd RE, Rudnick DL, Mazloff MR, Davis RE, Cornuelle BD. 2011. Poleward flows in the southern California Current System: glider observations and numerical simulation. *J. Geophys. Res.* 116:C02026
Triantafyllou G, Korres G, Hoteit I, Petihakis G, Banks AC. 2007. Assimilation of ocean colour data into a biogeochemical flux model of the eastern Mediterranean Sea. *Ocean Sci.* 3:397–410
Verlaan M, Heemink AW. 1997. Tidal flow forecasting using reduced rank square root filters. *Stoch. Hydrol. Hydraul.* 11:349–68
Vermeulen PTM, Heemink AW. 2006. Model-reduced variational data assimilation. *Mon. Weath. Rev.* 134:2888–99
Verron J, Gourdeau L, Pham DT, Murtugudde R, Busalacchi AJ. 1999. An extended Kalman filter to assimilate satellite altimeter data into a nonlinear numerical model of the tropical Pacific: method and validation. *J. Geophys. Res.* 104:5441–58
Wang X, Bishop CH, Julier SJ. 2004. Which is better, an ensemble of positive–negative pairs or a centered spherical simplex ensemble? *Mon. Weath. Rev.* 132:1590–605
Ward BA, Friedrichs MA, Anderson TR, Oschlies A. 2010. Parameter optimisation techniques and the problem of underdetermination in marine biogeochemical models. *J. Mar. Syst.* 81:34–43
Weaver A. 2013. *Aspects of covariance modelling in variational ocean data assimilation*. HDR Thesis, Inst. Natl. Polytech. Toulouse, Toulouse, France
Weaver A, Courtier P. 2001. Correlation modelling on the sphere using a generalized diffusion equation. *Q. J. R. Meteorol. Soc.* 127:1815–46
Whitaker JS, Hamill TM. 2002. Ensemble data assimilation without perturbed observations. *Mon. Weath. Rev.* 130:1913–24
Whitaker JS, Hamill TM. 2012. Evaluating methods to account for system errors in ensemble data assimilation. *Mon. Weath. Rev.* 140:3078–89
Wikle CK, Berliner LB. 2007. A Bayesian tutorial for data assimilation. *Physica D* 230:1–16
Wilkin JL, Hunter EJ. 2013. An assessment of the skill of real-time models of Mid-Atlantic Bight continental shelf circulation. *J. Geophys. Res.* 118:2919–33
Wunsch C. 1996. *The Ocean Circulation Inverse Problem*. Cambridge, UK: Cambridge Univ. Press
Wunsch C, Heimbach P, Ponte R, Fukumori I, ECCO-GODAE Consort. Memb. 2009. The global general circulation of the ocean estimated by the ECCO-Consortium. *Oceanography* 22(2):88–103

Xu FH, Oey LY, Miyazawa Y, Hamilton P. 2013. Hindcasts and forecasts of Loop Current and eddies in the Gulf of Mexico using local ensemble transform Kalman filter and optimum-interpolation assimilation schemes. *Ocean Model.* 69:22–38

Yu P, Kurapov AL, Egbert GD, Allen JS, Kosro PM. 2012. Variational assimilation of HnF radar surface currents in a coastal ocean model off Oregon. *Ocean Model.* 49–50:86–104

Zavala-Garay J, Wilkin JL, Arango HG. 2012. Predictability of mesoscale variability in the East Australia Current given strong-constraint data assimilation. *J. Phys. Oceanogr.* 42:1402–20

Zavala-Garay J, Wilkin JL, Levin J. 2014. Data assimilation on coastal oceanography: IS4DVAR in the Regional Ocean Modelling System (ROMS). In *Advanced Data Assimilation for Geosciences*, ed. É Blayo, M Bocquet, E Cosme, LF Cugliandolo. Oxford, UK: Oxford Univ. Press. In press

Zhang WG, Wilkin JL, Arango HG. 2010a. Towards building an integrated observation and modeling system in the New York Bight using variational methods. Part I: 4DVAR data assimilation. *Ocean Model.* 35:119–33

Zhang WG, Wilkin JL, Levin JC. 2010b. Towards building an integrated observation and modeling system in the New York Bight using variational methods. Part II: representer-based observing system evaluation. *Ocean Model.* 35:134–45

Annual Review of
Marine Science

Volume 7, 2015

Contents

Errata

An online log of corrections to *Annual Review of Marine Science* articles may be found at http://www.annualreviews.org/errata/marine