Available online at www.sciencedirect.com
ScienceDirect
Physica D 230 (2007) 172–196
PHYSICA D
www.elsevier.com/locate/physd

Adaptive modeling, adaptive data assimilation and adaptive sampling

Pierre F.J. Lermusiaux *

Massachusetts Institute of Technology, Department of Mechanical Engineering, 77 Massachusetts Avenue, Cambridge MA 02319, USA

Available online 27 March 2007
Communicated by C.K.R.T. Jones

Abstract

For efficient progress, model properties and measurement needs can adapt to oceanic events and interactions as they occur. The combination of models and data via data assimilation can also be adaptive. These adaptive concepts are discussed and exemplified within the context of comprehensive real-time ocean observing and prediction systems. Novel adaptive modeling approaches based on simplified maximum likelihood principles are developed and applied to physical and physical–biogeochemical dynamics. In the regional examples shown, they allow the joint calibration of parameter values and model structures. Adaptable components of the Error Subspace Statistical Estimation (ESSE) system are reviewed and illustrated. Results indicate that error estimates, ensemble sizes, error subspace ranks, covariance tapering parameters and stochastic error models can be calibrated by such quantitative adaptation. New adaptive sampling approaches and schemes are outlined. Illustrations suggest that these adaptive schemes can be used in real time with the potential for most efficient sampling.

Keywords: Physical and biogeochemical ocean modeling; Atmospheric and weather forecasting; Stochastic processes; Data assimilation; Observation targeting; System identification; Learning; Adaptive systems

1. Introduction

In the ocean, certain properties are more germane than others in governing specific dynamics. Similarly, certain observations are more relevant than others to explain specific features. These dynamics and features are a function of time and space, often non-homogeneous and non-stationary. For efficient representations of the ocean, the models of the dynamics and the sampling plans can exploit these characteristics. They can both be adapted to the ever-changing ocean events and interactions as they occur and so capture important properties. To reduce uncertainties in the representations, models and observations are combined via data assimilation. This combination can also be adaptive for increased efficiencies and accuracies. In the present manuscript, these concepts of adaptive models, adaptive sampling plans and adaptive data assimilation are developed. Novel quantitative adaptive approaches are outlined and exemplified, within the context of real-time estimations with comprehensive ocean models and real ocean data. The intent is to illustrate the feasibility and use of real-time adaptive approaches with today's ocean observing and prediction systems. Results are provided for physical and biogeochemical applications in regional ocean domains.

Differences between oceanic and atmospheric estimations strengthen the need for adaptive behaviors in ocean studies. This arises because of fundamental dynamical properties, practical reasons and societal needs. Densities are different, which affects time and space scales as well as the number of dynamical features that are modeled. If one divides the dynamically active spatial volume of a model domain, $L_h^2 L_z$, by the product of horizontal scales with a vertical scale, $\ell_h^2 \ell_z$, one obtains a rough order of magnitude for the number of features in the model domain. For example, consider the regional modeling of oceanic mesoscales (internal weather of the sea) versus that of atmospheric synoptic (weather) scales. These scales are linked to the internal Rossby radius of deformation. For today's regional models, one obtains $O(1)$ to $O(100)$ features in atmospheric domains but $O(10)$ to $O(1000)$ features

* Corresponding address: Massachusetts Institute of Technology, Department of Mechanical Engineering, Room 5-428, 77 Massachusetts Avenue, Cambridge, MA 02139-4301, USA. Tel.: +1 617 324 5172.
E-mail address: pierrel@mit.edu.
URL: http://web.mit.edu/pierrel/www/.

0167-2789/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.physd.2007.02.014

in oceanic domains.[1] Similar comments can be made for time scales. Practical differences involve the larger atmospheric data sets and the challenges to measure the ocean. For example, in most ocean mesoscale forecasting exercises, the state of the ocean at t_0 is unknown. Historical or climatological data can be used for model initialization but, due to seasonal and inter-annual variability, a dedicated ocean initialization survey is often needed to estimate the background state. Without such a survey, uncertainties can be as large as the whole variability. Due to our societal needs, atmospheric predictions and measurements have been emphasized. Ocean model parameters and parameterizations are thus less well known than their atmospheric counterpart. Modeling, assimilating and sampling with an adaptive component is most promising for ocean estimation.

A fundamental motivation for adaptive modeling is the ocean variability in time and space. For example, there are physical regime transitions or variations of biological assemblages. Practical reasons for model adaptation are our limited understanding and approximations. First, an ocean model usually focuses on a subset of spatial and temporal scales. Models are thus approximate because certain processes and scales are omitted and parameterizations are utilized. Significant inaccuracies also arise in model boundary and initial conditions. This is in part because oceanic observations are limited, with coarser coverage than atmospheric observations. Given all these uncertainties, several choices of parameterizations are often acceptable a priori. By adaptation to data, the best ones can be chosen a posteriori, in real time.

Adaptive modeling here refers to a *modeling approach that allows the definition, functionals and parameters of the model to quantitatively learn from observations and evolve with data as they are collected*. This process is an important feedback of data assimilation, leading to better scientific understanding. Based on model–data misfits, model properties that need to be improved are identified and the improvements estimated. Presently, a model property is said to be adaptive [15] if its formulation, classically assumed constant, is estimated or made variable as a function of data values. In general, three types of model properties can respond to data: (i) parameter values, (ii) model structures and (iii) computational properties (finite difference grid resolution, numerical scheme, etc). In this paper, model structures refer to everything that defines the model: its functionals and parameterizations but also its state variables and their linkages. The estimation of model errors and correction of model biases as well as the quantitative comparison of competing models and combination of multiple models are in some sense also forms of adaptive modeling.

A motivation for adaptive sampling is that ocean sensors and platforms will always be finite. At least some assets should be adaptively deployed as the ocean changes [5,62,40]. Resources are then best utilized for a given accuracy. Adaptive sampling is here defined as *predicting the types and locations of observations that are expected to be most useful, based on given estimation objectives and the constraints of the available assets*. Usually, it also implies carrying out the optimal sampling plans. General references on adaptive sampling include [72, 73]. Oceanic adaptive sampling has been done heuristically for a while: consider researchers following a coastal front, the Gulf Stream or a warm core ring. However, quantitative adaptive sampling is more recent [62,42]. It has been based more on predictions of oceanic variables than on predictions of the impacts of future data. In meteorology, a range of metrics and approaches have been used (e.g. [59]), from singular vectors techniques [11] to ensemble transforms [8, 9]. For recent applications, we refer for example to [69,47, 46]. Adaptive sampling in meteorology is also referred to as observation targeting, the result being targeted observations. Adaptive sampling differs from but is linked to optimal array design or antenna analysis [6]. Adaptive sampling usually progresses sequentially, as data are collected. Observations are then optimized for a future period, from the immediate future to the remaining duration of a sampling program. Array design computes optimal observations over the whole estimation period, e.g. [4,36]. It is usually carried out well in advance of the data collection based on some linearization.

Observing and prediction systems are now employed in various regions of the world's ocean for diverse purposes [54, 60,23]. Many such systems utilize data assimilation (DA). A number of DA methods are utilized in meteorology and oceanography [6,76,48,61,63,35]. Most schemes are derived from estimation theory, control theory or optimization theory. Almost all schemes are linked to a minimization of an error norm or DA criterion, such as least-squares norms and their approximations for real-time studies. The scheme employed here is Error Subspace Statistical Estimation (ESSE, [38,44]). Its error norms truncate the full error space to its dominant components. These components are forecast by a Monte Carlo ensemble approach, using stochastic forcing to represent model uncertainties. For a minimum error variance norm, ESSE is an approximate Kalman filter and smoother, where the dominant error covariance is estimated by an ensemble of nonlinear stochastic ocean simulations.

The focus here is on new adaptive components of modeling, sampling and DA that have been developed and used for real-time estimations within several regions. However, for a more unified presentation, illustrations are linked to the Autonomous Ocean Sampling Network-II (AOSN-II, [3]) exercise which occurred in August 2003 in the Monterey Bay and California Current System (CCS) region, e.g. [12,26]. The goal of this multi-institution project was to research an adaptive observing and prediction system, with the intent to assimilate various data types, adapt the deployment of platforms and allow the relocation of the system to other regions. A plethora of remote and in situ sensors and platforms were used, including gliders, drifters, moorings, Autonomous Underwater Vehicles (AUVs), research vessels and satellites for remote sensing. Nonlinear uncertainty forecasts of 2–3 days duration were issued using ESSE and the stochastic version of the Primitive Equation (PE)

[1] The ocean domain used here is 144 by 125 km. Horizontal mesoscales in the region are around 15–35 km, which gives 15–160 features in the thermocline volume.

model of the Harvard Ocean Prediction System (HOPS, [27]). Various data types were assimilated and suggestions for adaptive sampling and dynamical studies were provided. The synthesis of this real-time work and detailed investigation of dynamics in the region is not presented here.

In what follows, real-time adaptive modeling for physical and physical–biogeochemical dynamics and predictions is first investigated and exemplified (Section 2). In Section 3, adaptable features of the ESSE system are presented, discussed and illustrated. Adaptive sampling is investigated and exemplified in Section 4. Conclusions are in Section 5. The deterministic ocean models employed and their stochastic forcings are defined in Appendix A. The ESSE formalism and computations are outlined in Appendix B.

2. Adaptive modeling

In ocean modeling, model tuning and calibration is often based on experience and scientific knowledge. The comparison of models to data can be as much an art as a rigorous scientific method. There is a need to provide new quantitative and computational aids to the modeler. The methodological basis of this new adaptive modeling is the misfits among model estimates and data. When misfits are significant, models are adapted. Models can also be rejected when estimated to be too inadequate. The choice among approaches [16,15] that can be used for adaptive modeling depends on the application, robustness requirements, data availability and costs involved. The main cost constraints are the size of the ocean state vector, of $O(10^6 - 10^7)$, and large number of model parameters, of $O(10 - 10^2)$. Maximum likelihood schemes look for a single model estimate. Pure Bayesian methods aim to use data to sharpen a prior probability density on model estimates, leading to a posterior density. In both cases, an *a priori* set of model properties can be chosen. Adaptive modeling is an extension of classic parameter estimation (e.g. [71,66,17,52,68]) to the selection of model state variables, linkages, parameterizations and/or computational characteristics. The optimization is thus over both model structures and parameter values.

Possible issues in adaptive modeling include the robustness, scalability and convergence of the adaptation scheme. Some model structures and parameters can appear accurate, mainly because of insufficient data. This degeneracy has been referred to as equifinality (e.g., [7]). The results of [49] provide a biogeochemical example. When such degeneracy occurs, one can select the simplest of the models or weight and combine fields of equifinality models into multi-model estimates. Here, we develop new concepts in adaptive physical and adaptive physical–biogeochemical modeling. Results are illustrated within the context of AOSN-II [3], based on a simplified maximum likelihood approach. A small set of deterministic model configurations is chosen a priori and the configuration with the best averaged skill is assumed to be the most likely.

2.1. Adaptive physical modeling by Monte Carlo "Trial and Error" and optimization of predictive skill

For the ocean physics, fundamental equations are known, the Navier–Stokes equations. However, approximations occur because of equations are reduced to the scales of interest and because observations are limited. The resulting parameterizations and parameter values, including boundary and initial conditions, can thus learn from new data. There are essentially two categories of adaptation. The first category is hypothesis testing. A physical model with several candidate parameterizations is ran multiple times with these different parameterizations. Such test simulations can be carried out in parallel. The second category is the adaptive evolution that consists of a physical model with adaptive parameterizations. Of course, these categories can be combined, even though their numerical implementation differ. We present next an example of "hypothesis testing" adaptation, run in real time during AOSN-II.

Prior to AOSN-II, model parameters were calibrated to historical conditions expected to be similar to the unknown August 2003 conditions. To do so, about 260 different numerical simulations were carried out in the region. Four candidate initialization data and parameter sets were selected. Three were August climatologies, corresponding to El Nino, La Nina and regular year conditions, respectively. The fourth was based on an historical survey with good synoptic coverage over the region [55]. At the beginning of AOSN-II, ocean forecasts were started from these four initial conditions and parameters. Once the August 2003 initialization survey was completed, the physical ocean model was restarted from fresh initial conditions computed from this survey and other AOSN-II synoptic data sets. Model properties were then adapted to the new 2003 data. Based on comments from collaborators, on our own experience and on quick-look comparisons among model fields and measurements, several parameters and a parameterization were selected to be candidates for improvements. The parameters were: (i) the initial conditions and simulation restart time; (ii) values of empirical constants in the schemes for vertical mixing and dissipation of atmospheric fluxes; and (iii), horizontal viscosities. The parameterization that was allowed to vary was the formulation of the open boundary conditions. This adaptive modeling thus intends to optimize both over parameter values in the interior and over model functionals at boundaries. In particular, the use of parameter estimation in ocean modeling to estimate values at open boundaries (e.g. [77]) is extended to the estimation of functionals at these boundaries.

These selections are summarized by Eqs. (1)–(4), using notation defined in Appendix A.1. Let us review them one by one, starting with the initial state and time (Eq. (1)). The first choice for initial conditions was a two-scale objective analysis[2] of the ship-based initialization survey Fig. 1 and concurrent glider measurements during Aug. 2–7, 2003. The date of the start of the PE model, within Aug. 5–7, was chosen to be a parameter. The other initialization choices were snapshots in model simulations started from the climatological/historical data sets but that had assimilated the initialization survey and glider measurements. Vertical mixing parameters (Eqs. (2a)–(2c), see Appendix A.1) were also estimated because forecast

[2] A larger-scale map of the data is first estimated, then data residuals are computed and mapped into a mesoscale field which is added to the larger-scale map, see [39].

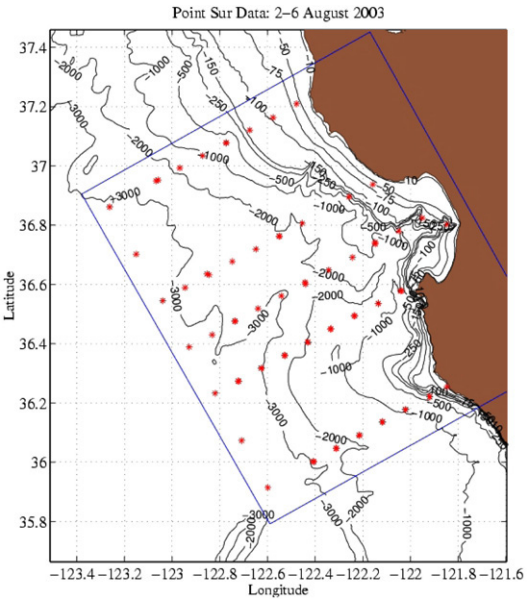

Fig. 1. Horizontal positions of the temperature and salinity profiles taken by the R/V Point Sur during 2–6 August 2003, overlaid on bottom bathymetry. The blue rectangle indicates the limits of the ocean model domain. More information on AOSN-II is available from [3]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

fields appeared too geostrophic at the surface (e.g., Leslie Rosenfeld, personal communication) and likely a bit too warm in the upper layers, especially along the north-western coastline of Monterey Bay. The parameters adapted are in the definition of the vertical eddy viscosities A_v and diffusivities K_v. They were: ν_0 (cm^2/s, shear viscosity at zero local gradient Richardson number), K_v^e (cm^2/s, eddy diffusion for tracers within the wind-mixing depth h^e) and E_k (Ekman depth factor). The horizontal viscosity on vorticity (Eq. (3)) which is modeled by a Shapiro filter [67] was also adapted.

The parameterizations adapted were the open boundary conditions (Eqs. (4a) and (4b)). Several new formulations were developed and evaluated in real time. The need for such an open boundary update arose because of (i) the limited synoptic data offshore and (ii) the tendency of the regional ocean conditions to relax [64], in response to weaker or reversed atmospheric forcing and/or larger-scale buoyancy-driven forcing by the California Current System [12,26]. Such natural relaxations have been observed, for example after upwelling events [64]. Denoting the dynamical tracers, temperature and salinity, by T and S, and the internal velocities by u and v, the formulations tested were: no relaxation; relaxation at some or all open boundaries; relaxation for T and S only; relaxation for all (T, S, u and v); relaxation dependent on depth or not; and, reduced relaxation at outflow points, either only for T and S, or for all four variables. Using T_i and u_i to represent the two dynamical tracers and internal velocities at boundary i, the final adapted formulation is given by Eqs. (4a) and (4b).

IC. State & Time $\quad \mathbf{T}_0, \mathbf{S}_0, \mathbf{u}_0, \mathbf{v}_0\ t_0$ (1)

Vertical Mixing Param. $\quad \dfrac{\partial A_v(\nu_0, A_v^e, E_k)\partial \mathbf{u}_h/\partial z}{\partial z}$ (2a)

$$\frac{\partial K_v(\nu_0, K_v^e, E_k)\partial T/\partial z}{\partial z} \tag{2b}$$

$$\frac{\partial K_v(\nu_0, K_v^e, E_k)\partial S/\partial z}{\partial z}. \tag{2c}$$

Horiz. Viscos. on Vort.

Shapiro filter param.: order, # of applic., freq. (3)

OBC functionals

(final formulation)

$$\frac{\partial T_i}{\partial t} \pm c_n \frac{\partial T_i}{\partial n} = -\beta_i \frac{(T_i - T_i^\ell(x, y, z, t))}{\tau_i(z)} \tag{4a}$$

$$\frac{\partial u_i}{\partial t} \pm c_n \frac{\partial u_i}{\partial n} = -\beta_i \frac{(u_i - u_i^\ell(x, y, z, t))}{\tau_i(z)}. \tag{4b}$$

Importantly, Eqs. (4a) and (4b) were not used in each model version. Different codes were compiled, one for each formulation of the open boundary conditions tested. It is only the adaptation and optimization of the boundary condition equations that led to the final formulation ((4a) and (4b)), which happened to encompass all formulations tested. The left-hand side contains an Orlanksi open boundary condition [57] based on the Sommerfeld radiation equation where c_n is the phase speed of the signal normal to and at the boundary. The right-hand side is the new boundary relaxation towards the most recent estimate of the larger-scale buoyancy state, denoted by $T_i^\ell(x, y, z, t)$ for the dynamical tracers and $u_i^\ell(x, y, z, t)$ for the internal velocities. These larger-scale buoyancy fields were updated at each assimilation step as a function of the error standard deviation of a larger-scale objective analysis of the new data. At each open boundary i, adaptive modeling was used to estimate the relaxation time scale τ_i in Eqs. (4a) and (4b). It was selected to be a function of depth, given by: $\tau_i(z) = \tau_i^b\ (1 - \mathrm{e}^{-\frac{z}{h_i}}) + \tau_i^s \mathrm{e}^{-\frac{z}{h_i}}$, where τ_i^b and τ_i^s are bottom and surface relaxation time scales, respectively, z is depth and h_i is a vertical depth scale beyond which $\tau_i(z) \to \tau_i^b$. This depth dependent formulation aims to allow the radiation in the upper layers of surface intensified dynamical events and waves. It also limits localized numerical uncertainties at depth. Finally, the parameter $\beta_i = \beta_i(u_i, v_i)$ is used to reduce the relaxation at outflow points. In the final adapted run, it was set at inflow points to $\beta_i = 1$ and at outflow points to $\beta_i = 0.25$.

The real-time adaptation proceeded as follows. Each parameter and parameterizations were first modified one at a time. Additional simulations with more than one modification were then also ran. In total, a set of 49 simulations was carried out in real time, in parallel batches. These simulations started either from August 5 or 7. All assimilated the same temperature and salinity data up to August 15 and issued ocean forecasts for August 16, 17 and 18. They were forced at the ocean surface by COAMPS atmospheric fluxes (J. Doyle, personal communication): these fluxes were analyses up to August 16 and forecasts beyond.

Table 1
Parameters and parameterization before and after adaptation

	IC/Re-start date	ν_0	K_v^e	E_k	Boundary relaxation				
Non-adapted	OA on Aug. 7 of Aug. 2–7 data	50.0	5.0	0.15	None				
Adapted	”	”	6.0	0.22	Bnd i	y/n	τ_i^b	τ_i^s	h_i
					West	1	1.5	4.5	50
					South	1	1.0	2.0	50
					East	0	–	–	–
					North	1	1.0	2.0	50

Units: ν_0 (cm^2/s), K_v^e (cm^2/s), E_k (non-dimensionalized), τ_i^b and τ_i^s (days), and h_i (m) (see also Appendix A.1).

The results of each simulation were compared to data qualitatively and quantitatively, daily up to August 18. The data used were T and S profiles from two sets of gliders (R. Davis and D. Fratantoni, personal communication). Model fields were compared both to measurements at data points and to objective analyses of the data at model points. The quantitative comparisons were done using horizontal mean estimates of the bias, Root Mean Square Error (RMSE) and Pattern Correlation Coefficient (PCC), for each model configuration. The bias and RMSE estimates are defined by $\langle T^f(x,y,z,t) - T^o(x,y,z,t)\rangle_h$ and $\sqrt{\langle (T^f(x,y,z,t) - T^o(x,y,z,t))^2\rangle_h}$, where $\langle\cdot\rangle_h$ denotes an horizontal average and, T^f and T^o are the forecast at data point and the observed value, respectively. Choosing a background mean estimate to be a large-scale field T^ℓ and defining $\Delta T^f = T^f(x,y,z,t) - T^\ell(x,y,z,t)$ and $\Delta T^o = T^o(x,y,z,t) - T^\ell(x,y,z,t)$, the mesoscale PCC is $\frac{\langle \Delta T^f \Delta T^o\rangle_h}{\sqrt{\langle \Delta T^{f\,2}\rangle_h}\sqrt{\langle \Delta T^{o\,2}\rangle_h}}$. The ultimate choice among the different model configurations was made by comparing the depth and time average of these three numbers. Depth averages were computed for each of the three forecast days (August 16, 17 and 18), for both T and S. These numbers were then non-dimensionalized and summed. In that sum, to reduce the effect of atmospheric forecast uncertainties, the August 16, 17 and 18 values were given a weight of 1, 0.8 and 0.5, respectively, based on a time decorrelation. Salinity values were also given two third of the weight of the corresponding temperature values because of the larger uncertainties of salinity sensors.

The adaptive modeling scheme used here is a simplified maximum likelihood. The model configuration with the best averaged skill is assumed to be the most likely model. The prior and posterior set of parameters and parameterization are given in Table 1. The initial conditions, parameter ν_0 and horizontal viscosity remained unchanged (changing them made things worse or had no significant impact). The vertical mixing in the wind-mixing surface layer and the depth of this layer were increased by 20% and 30%, respectively. The penetration of atmospheric effects is then increased and geostrophic tendencies reduced. A boundary relaxation scheme was combined with the radiation condition (Eqs. (4a) and (4b)), with a time dependent relaxation time scale. The relaxation is stronger in the alongshore direction (north and south boundary), in accord with the mean direction of the flow, with alongshore inflows from the north at the surface and from the south at depth. The relaxation is also weaker in the upper layers, so as to allow active surface dynamics (e.g. upwelling plume) to radiate out of, or exit, the domain. Of course, in the present adaptation, only 49 model configurations were run and the true optimum is likely not attained. To account for all uncertainties, a small ensemble of simulations should also be run out for each configuration. This ensemble statistics is here approximated by averaging over time. Our goal is simply to show benefits of sub-optimal but real-time and quantitative adaptation of a comprehensive ocean dynamical model.

Figs. 2–4 illustrate the results of the adaptation. The unadapted and adapted physical ocean model are compared to the T and S data collected on Aug 16, 17 and 18, 2003, as a function of depth and time. Quantitative comparisons at data points are illustrated by Fig. 2. The horizontally averaged bias estimate (Fig. 2(a)), Root Mean Square Error (RMSE) estimate (Fig. 2(b)) and mesoscale pattern correlation coefficient estimate (Fig. 2(c)) were computed at 20 depths, from 2 m to the largest data depth of 400 m. In these estimates, data values are assumed to represent truth. On Fig. 2(a–c), only values for the first 100 m are shown for visibility (the other depths 125, 150, 175, 200, 250, 300, 350 and 400 m are not plotted).

On average, the biases (Fig. 2(a)) for the unadapted and adapted forecasts are similar on Aug 16 and Aug 17. However, the adapted model has smaller biases on Aug 18, for both T and S. Note that bias profiles change shape and sign from day to day, in part due to limited data coverage (Fig. 3). However, bias amplitudes remain small and similar with time. Considering RMSE values (Fig. 2(a)), the adaptation has substantially reduced the RMSEs of the T forecasts, on each day and with averaged improvements of 20%–40%. For S, they are similar for the first two days but much smaller for the third. Note the shape of the RMSE profiles, similar to that of the variability. With adaptation, the mesoscale PCC values (Fig. 2(c)) of the T forecasts are much closer to one, every day and with an averaged PCC increase of 40%–60%. For S, adapted PCCs are better for Aug 16 and 18, by about 10%–20%, but slightly worse for Aug 17 (this may be due to sensor biofouling). In conclusion, the bias, RMSE and PCC estimates confirm that the adaptive scheme led to significant quantitative improvements of model forecast fields.

These improvements are illustrated on Fig. 3 by differences between the objectively analyzed (OAed) T data and the real-time T forecasts using the unadapted (left column) and

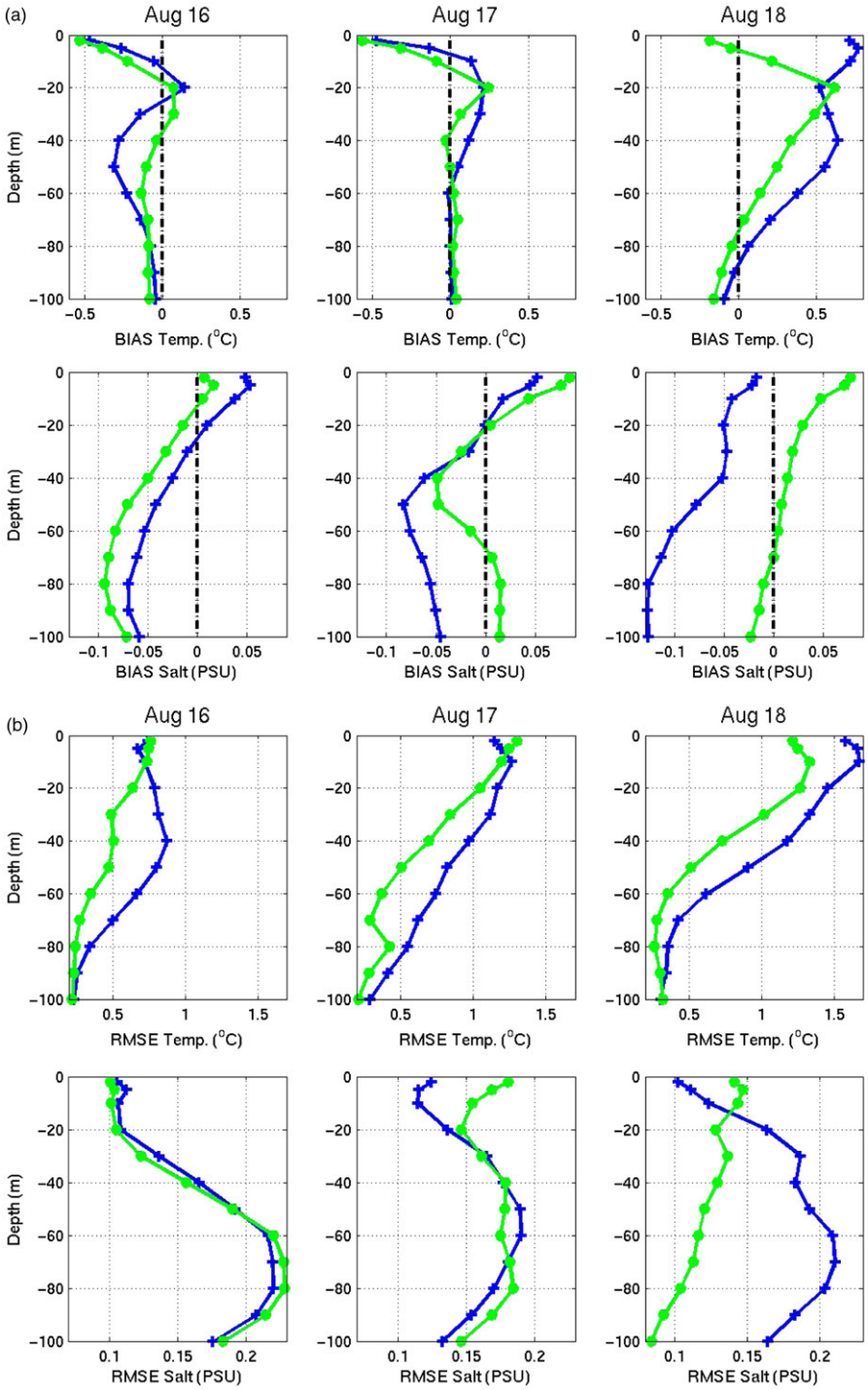

Fig. 2. Comparison between real-time unadapted (blue lines with a plus at each data point) and real-time adapted (green lines with filled circles) physical ocean model. (a) Bias estimate (Model - OAed data) for temperature and salinity, as a function of depth (m) and time (day). (b) As (a), but for the Root Mean Square Error (RMSE) estimate. (c) As (a), but for the mesoscale pattern correlation coefficient estimate. Note the overall decay of the PCC with the duration of the forecasts. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

adapted (right column) model. The depth of 30 m was selected because it is characteristic of the pycnocline. The forecasts and OAed data are estimates for noon GMT (4 a.m. local time). Differences are plotted only where the expected, mesoscale error standard deviation of the OAed data is less than 30% (in non-dimensional form). This also shows the limited data coverage, especially offshore. This offshore region is where the adapted model has enhanced the non-adapted model the most, for the 3 days. The effect of the adaptation is smaller near Monterey Bay. This is because data density was higher there

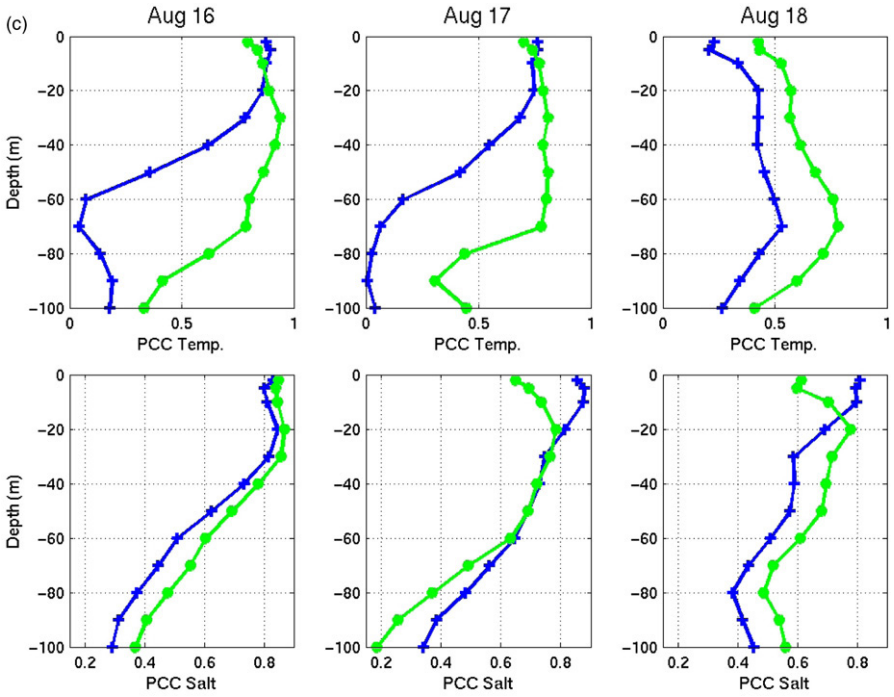

Fig. 2. (*continued*)

since the start of the experiment on Aug 2, which reduces the impacts of different model configurations.

The full temperature fields at 30 m are shown on Fig. 4. The effects of model updates are significant. The sustained warmer temperature at 30 m along the western open boundary is due to: (i) the increased Ekman depth factor and stronger eddy diffusion, which lead to a deeper mixing of warmer surface waters; and (ii), the boundary relaxation towards the data-driven larger-scale field. At 30 m, updates along the coastline and in Monterey Bay are smaller around data locations but still significant elsewhere (even though less visible on Fig. 4 due to the larger T gradient between offshore and shelf waters). Several structures in the flow and tracer fields have been modified by the adaptation. Our real-time example is based on a relatively small ensemble of simulations, but it indicates the potential of adaptive physical modeling.

2.2. Towards adaptive physical–biogeochemical modeling

Fundamental biological equations remain relatively unknown. Even though much progress has been made recently in biogeochemical understanding and modeling e.g. [37,28], model uncertainties remain relatively larger than in physical modeling. One challenge is the countless types of biological organisms and animals, which evolve and go through various life stages. Many biological processes, interactions and behaviors are possible and new ones continue to be discovered. Physical and chemical variables such as light, temperature, currents and dissolved nutrients also impact biological processes and parameters (e.g., growth rates, grazing and predation rates). For efficient modeling, biological quantities are first aggregated into a possibly large but limited number of model state variables. The interactions among these aggregated variables including biological behaviors and forcing functions are then parameterized. Currently, several formulations are based more on empirical relationships than on underlying ecological or physiological processes. A variety of parameterizations have been utilized and there are few standards. All of these fundamental model uncertainties compound with classic ocean simplifications such as model reductions to the scales of interest and limited data (Section 1). Adaptive modeling can then be used to select adequate biological model properties and so reduce these uncertainties. A corollary is to use adaptive modeling for automated biological aggregations or simplifications, by comparing complex and simple models with data.

Adaptive modeling is also useful for ecosystems simply because biological states evolve and go through transitions, and because interactions vary with time and space [50]. For efficiency, predictive models can have the same behavior and adapt to the ever-changing dynamics. For example, consider the variations of biological assemblages with time and space, with respect to both size classes and functional phytoplankton and zooplankton groupings. As the plankton community and nutrient composition shift over the course of a bloom, efficient models can adapt accordingly. Biogeochemical parameters (rates, ratios, etc) and model structures can adapt in such circumstances.

As for the physical case, there are essentially two basic categories of biological–physical adaptation (Fig. 5). The first (Fig. 5, left) is a physical model that drives multiple competing biological models (biological hypothesis testing). The second (Fig. 5, right) consists of a single adaptive

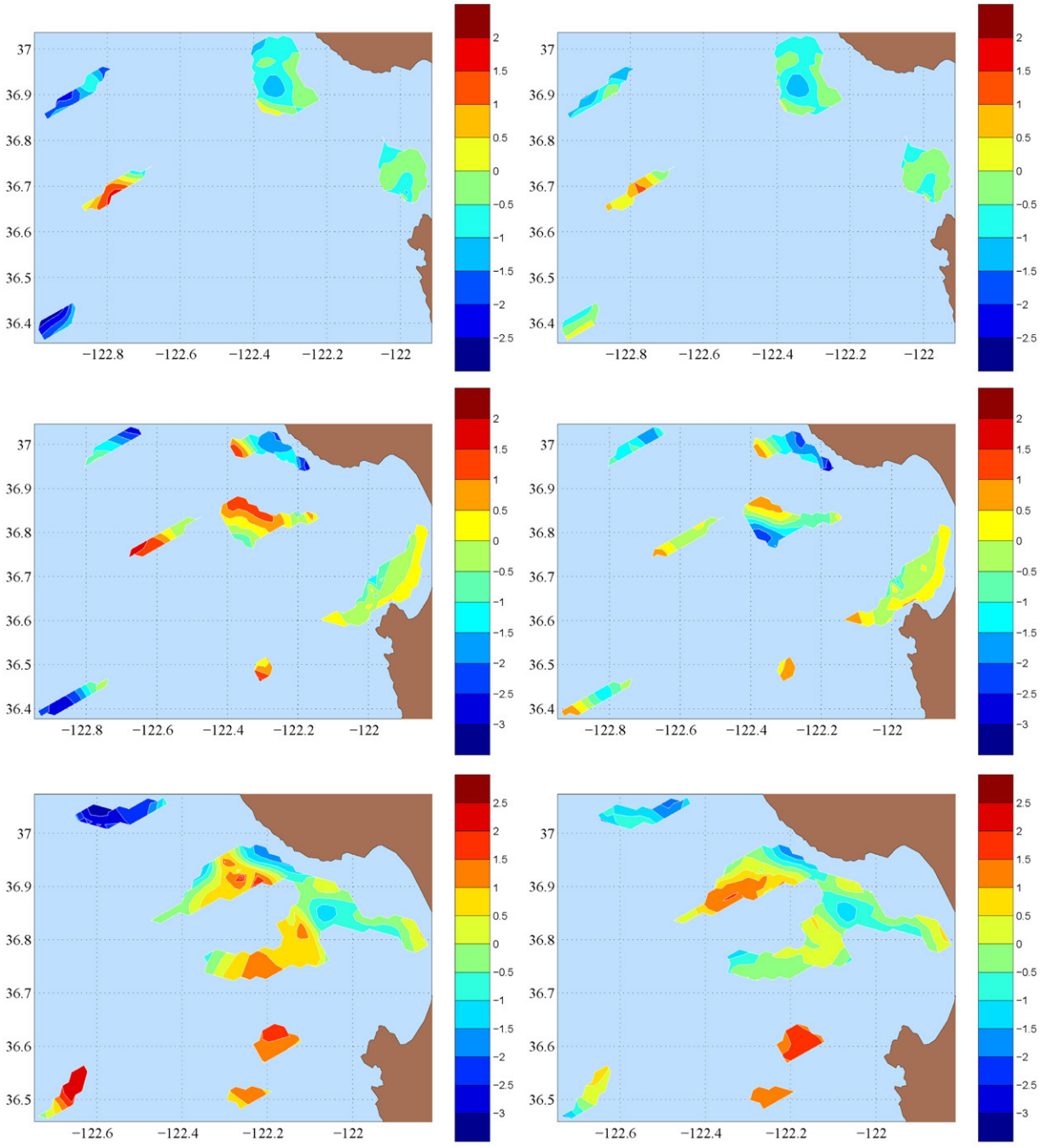

Fig. 3. Differences between the objectively analyzed (OAed) temperature data and the real-time unadapted (left column) and real-time adapted (right column) temperature forecasts at 30 m, for Aug 16 (top), 17 (middle) and 18 (bottom), 2003. Differences are shown only where the expected mesoscale error standard deviation of the OAed data is less then 30% (this standard deviation is the non-dimensional form of the square root of the a posteriori error variance of the mesoscale OA map). The region plotted varies from day to day, following glider motions.

interdisciplinary model: an adaptive physical and adaptive biological model proceed in parallel, with some independent adaptation. As shown on Fig. 5, physical and biological adaptations do not need to be computationally concurrent, but they must be dynamically compatible. Of course, there are several intermediate cases not sketched on Fig. 5. For the first category, the physical model could be adaptive while competing biological models are evaluated. A single physical model could also be coupled to a set of competing biological models whose unknown parameters are adapted. Finally, the combination of the two categories defines the general case of running competing adaptive interdisciplinary models. Note that many aspects of such new biological adaptive modeling are computationally challenging [45]. Due to the smaller physical uncertainties, physical model selection should likely precede biogeochemical model selection.

To carry out biological adaptive modeling and allow automated dynamic switching of biogeochemical parameters and structures, new generalized biogeochemical computational models are needed. A version of such a flexible model

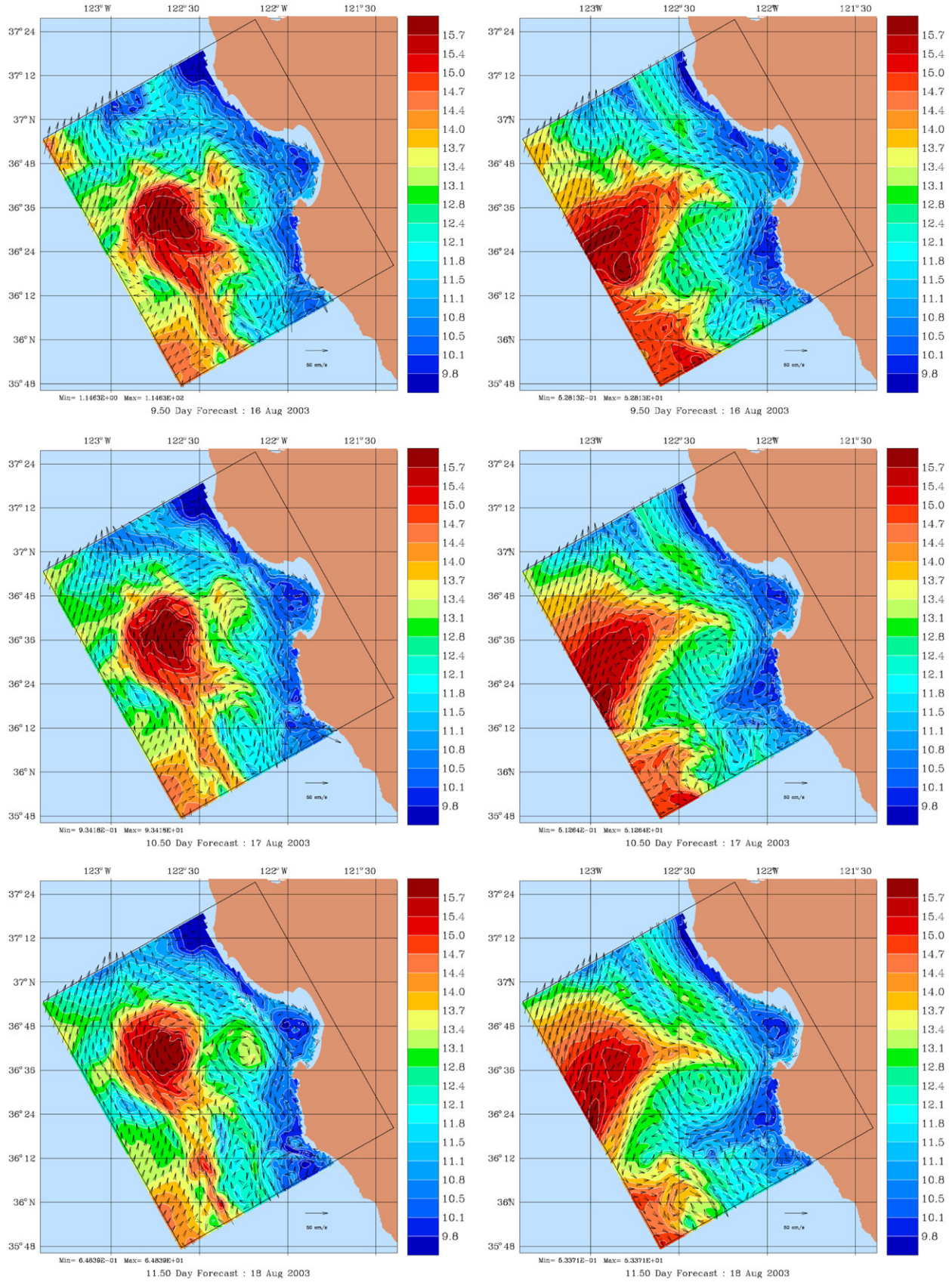

Fig. 4. Real-time unadapted (left column) and real-time adapted (right column) temperature forecasts at 30 m overlaid with horizontal velocity arrow vectors at that depth, for Aug 16 (top), 17 (middle) and 18 (bottom), 2003. Arrow vectors are only plotted every three grid points and the velocity scale (bottom right) is 50 cm/s.

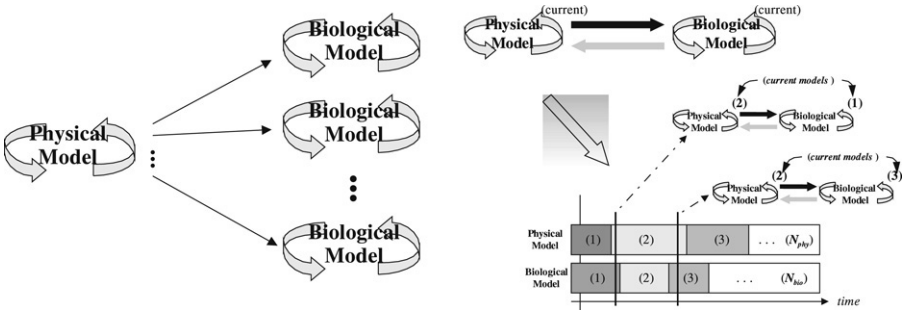

Fig. 5. Sketch illustrating adaptive coupled physical–biological modeling. Left: a set of biological models are coupled to the same physical model. Right: a physical and a biological model evolve during the simulation as a function of data.

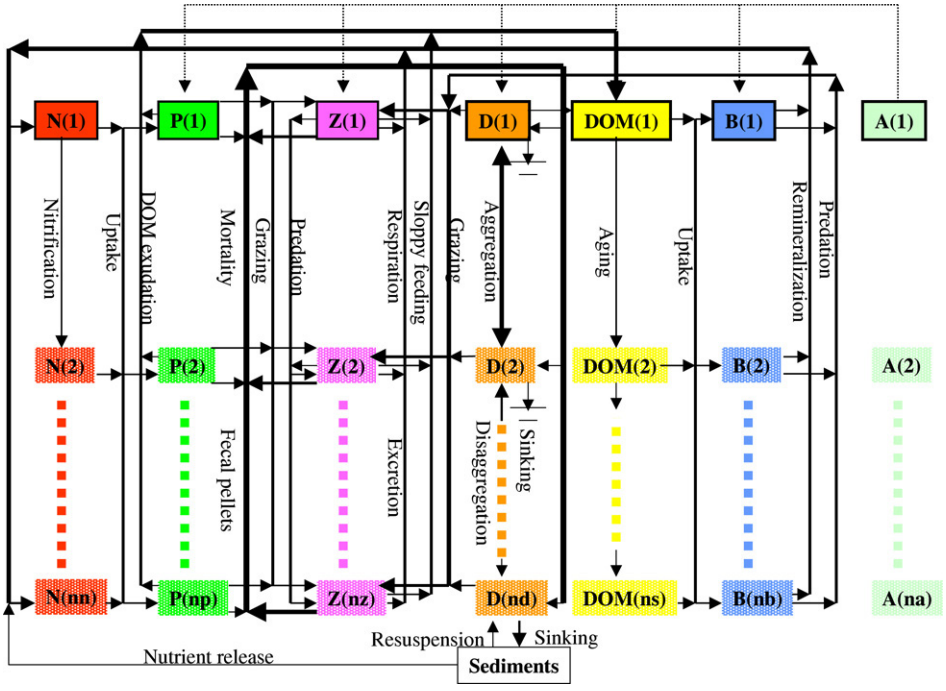

Fig. 6. Generalized biological model [70]. N: Nutrients; P: Phytoplankton; Z: Zooplankton; D: biogenic Detritus; DOM: Dissolved Organic Matter; B: Bacteria; A: Auxiliary variables. nn, np, nz, nd, ns and na: adaptable numbers of state variables for each functional group.

has been developed [70]. It was constructed based on a study of functional groups and parameterizations for coastal ecosystems. A general set of state variables and of mathematical structures representing their interactions was selected, based on importance, completeness and accuracy. As illustrated on Fig. 6, this led to a generalized model with the following generic functional groups of state variables: nutrients (N_i), phytoplankton (P_i), zooplankton (Z_i), detritus (D_i), dissolved organic matter (DOM_i), bacteria (B_i) and auxiliary variables (A_i). Within each functional group, the number of state variables can vary and the parameterizations of their interactions can be selected. With such flexibility, this new model can adapt to different ecosystems, scientific objectives and available measurements. Details of its properties and implementation are given in [70].

We now illustrate the use of this new model in two configurations for the Monterey Bay region. The first configuration is a simple NPZ model with two nutrients. It has four state variables: phytoplankton, zooplankton and two nutrients (NH_4^+ and NO_3^-). Chlorophyll is simply estimated from a scalar Chl : P ratio. The second configuration (Fig. 7) is what we refer to as an *a priori* model. As in classic Bayesian estimation, it is selected based on previous knowledge of the ocean region of interest and on the synoptic data available. In the present case, this *a priori* set-up has two nutrients (NH_4^+ and NO_3^-), two phytoplankton (small and large), two zooplankton (micro and meso), two detritus types (small suspended and large sinking detritus), one DOM (dissolved organic nitrogen), one bacterial pool and four auxiliary state variables (prokaryotic, eukaryotic and total chlorophyll and bioluminescence). Another *a priori* configuration uses two additional nutrients, PO_4 and $\mathrm{Si(OH)}_4$, but it is not used here. The units of biological variables are: m mol of nitrogen per meter cube (m M N m^{-3}). For more details on each configuration, we refer to [70].

These two biological set-ups were coupled to a physical re-analysis by the PE model of HOPS which assimilates physical AOSN-II data collected during August 2003 (Section 1).

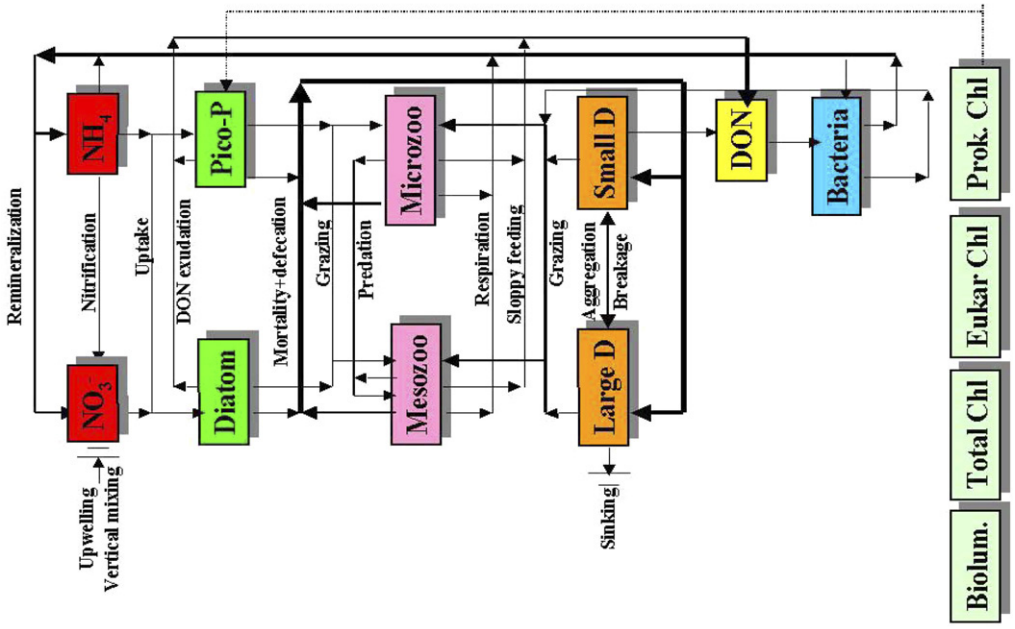

Fig. 7. *A priori* model configuration, for the Monterey Bay region: two nutrients (NH_4^+ and NO_3^-), two phytoplankton (small and large), two zooplankton (micro and meso), two detritus sizes (small and large), one DOM (DON), one bacterial pool and four auxiliary state variables (small, large and total chlorophyll and bioluminescence).

Biological initial conditions and parameters were selected based on historical data and the literature. A simulation for each set-up is illustrated on Fig. 8, by surface fields during an upwelling event. The surface Chl of the two set-ups are logically different, with the NPZ configuration having a stronger tendency to overestimate production offshore (when compared to sea surface color images, not shown). There are several reasons for this. One of them is the use of one instead of two classes of phytoplankton (illustrated by Chl fields on Fig. 8). The NPZ set-up (Fig. 8, bottom) simulates an aggregate of phytoplankton with unique growth and mortality rates. In the *a priori* set-up (Fig. 8, top), the larger phytoplankton requires more nutrients, light, etc, for optimal growth, hence the weaker total growth offshore. Comparisons of biological fields with the available fluorometer data further indicates the better adequacy of the *a priori* model. However, quantitative comparisons also show that other improvements are necessary, including an overall slower biological production, a higher or more rapid mortality offshore and possibly higher biological grid resolution. Some of this may be achieved by using four nutrients (see above). Ultimately, more biological data are needed. The new generalized model with adaptive behavior could then select model properties in an automated quantitative fashion.

3. Adaptive ESSE parameters and data assimilation

Most DA schemes used for realistic studies approximate fundamental principles for combining various sources of information. One of these principles is Bayes rule [34,16]. In part because of these approximations, the schemes involve several parameters, options and heuristic algorithms whose specifics impact the DA results. In the ESSE system, such specifics can vary with each application and with users inputs. They are also modified with time, as a function of the regional dynamics, available data or other considerations. In what follows, such adaptable features are reviewed and illustrated for the first time, for each computational components of the system. We refer to [10] for the related investigation of adaptive Kalman Filters and to [51] for covariance matching.

The main ESSE components are: (i) uncertainty initialization (see also [53]); (ii) deterministic–stochastic ensemble and error subspace forecasts; (iii) adaptive sampling; (iv) data processing and measurement model; (v) data assimilation; (vi) adaptive error learning; and, (vii) smoothing. These algorithms are outlined in Appendix B, excepted for adaptive sampling (Section 4) and stochastic forcing (Appendix A.2). For the *uncertainty initialization* [43], the properties that can vary include the number of initial error scales utilized, the type and parameters of the horizontal error covariances for each scale, and the factors setting the amplitude of the prior initial error conditions based on historical data or variability estimates. For the *ensemble and error subspace forecast*, they include the statistics of the random noise $\mathbf{n}_k^j$ (Appendix B) used to model the truncated errors and the adaptation of this statistics based on posterior data–model misfits. Properties of the stochastic forcings (Appendix A.2) also need to be selected and can be adapted. The choice of the ocean forecast from the ESSE ensemble is a parameter. One can select either the: central forecast, ensemble mean, most probable forecast (mode) or ensemble member closest to the mean/mode or to the new data. Finally, several convergence criteria and factors can be used to control the size of the error subspace. At the *data assimilation* step, the amplitudes of error forecasts can be scaled, e.g. inflated. This scaling can be adapted exploiting the prior data–ensemble misfits. When ensemble sizes are too small, error covariances

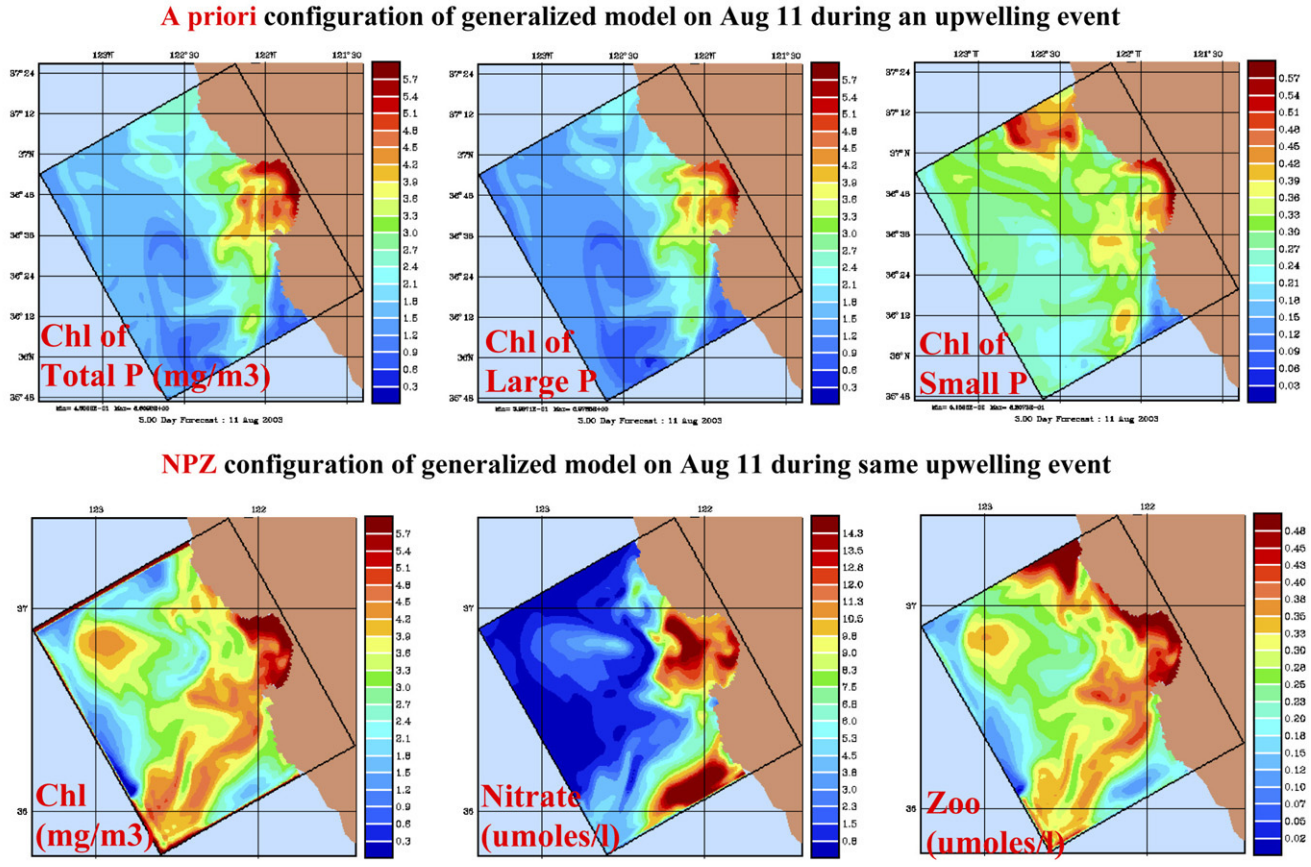

Fig. 8. Surface horizontal maps for the *a priori* and NPZ model configurations on Aug 11, 2003, during an upwelling event in the Monterey Bay region. Top: Chlorophyll associated with the total, large and small phytoplankton for the *a priori* configuration. Bottom: chlorophyll, nitrate and zooplankton for the NPZ configuration.

can also be reduced at long distances before DA, based on a Schur product of the background covariance matrix $\mathbf{B}_k^p$ with another covariance that decays at long distances [32,2,24,25]. After DA, *adaptive error learning* updates uncertainty estimates from the misfits among ESSE analyses and data (posterior data–ensemble misfits). For the *smoothing* step, options are limited because the backward smoothing equations depend on prior and posterior filtering estimates which are set. Next, we discuss and illustrate adaptive ESSE filtering.

(i) *Adaptive error scaling*. The error variances of an ESSE analysis or ensemble prediction can be scaled, e.g. tuned, by multiplication with a diagonal matrix of factors [40,41]. The scalar version of this multiplication is often referred to as error covariance inflation in weather forecasting [1,75]. In regional ocean modeling exercises such as predictive skill experiments, such scaling is most needed at the start of the exercise. This is because the internal state of the ocean at t_0 is often poorly known. For example, consider the case where the only synoptic data to start with are a few remotely sensed surface observations. The initial state is then estimated from historical in situ data sets, remotely sensed data and feature models [18]. Ocean state estimation at t_0 can thus be very different from an initialization in weather forecasting where the atmospheric state at t_0 is observed with a much larger accuracy. Even simple tuning of the initial ocean error estimate is useful.

In ESSE, the estimate of the dominant eigendecomposition $\mathbf{E}\boldsymbol{\Pi}\mathbf{E}^{\mathrm{T}}$, of the error covariance, $\mathbf{P}^p$, at t_0 (or at any t_k) can be scaled using a non-dimensional block diagonal matrix $\boldsymbol{\Gamma}$, leading to $\mathbf{B}^p = \boldsymbol{\Gamma}\mathbf{E}\boldsymbol{\Pi}\mathbf{E}^{\mathrm{T}}\boldsymbol{\Gamma}^{\mathrm{T}}$ [41]. Each block of $\boldsymbol{\Gamma}$ corresponds to a state variable (Appendix A.1) and is presently defined by a single scale factor. The procedure to determine the values of these factors is as follows. A preliminary estimate is first chosen. At t_0, values in $\boldsymbol{\Gamma}$ are usually set within 0.3–0.7 since $\mathbf{E}\boldsymbol{\Pi}\mathbf{E}^{\mathrm{T}}$ corresponds to historical variability estimates [43] which are often larger than initial uncertainties. At all times $t_k > t_0$, the preliminary estimate is 1, i.e. $\boldsymbol{\Gamma} = \mathbf{I}$, because it is first assumed that error amplitudes are adequate. Starting from this preliminary estimate at t_0 (t_k), a batch of usually 10–25 ocean forecast simulations is ran to t_1 (t_{k+1}) using Eqs. (B.11) and (B.12) of Appendix B and distributed computing. The standard deviation of this first batch of forecasts is a preliminary error prediction. If it is found to be too small (sometimes too large), either from experience or from the misfits between this preliminary ensemble and the most recent data, the factors in $\boldsymbol{\Gamma}$ are increased (sometimes decreased). The process is repeated until values in $\boldsymbol{\Gamma}$ are judged acceptable. For coastal mesoscale ocean forecasting exercises, it usually takes a few assimilation cycles and several days after t_0 for the values of $\boldsymbol{\Gamma}$ to stabilize and remain set to one. Obviously, for fully operational systems, such an adaptive procedure can be automated based on data–ensemble misfits.

(ii) *Adaptive parameterization of the truncated errors*. Uncertainties not represented by the posterior error subspace are modeled by the random noise $\mathbf{n}_k^j$ whose statistics needs to be specified. These errors are in theory orthogonal to the error subspace, i.e. to the eigenvectors of a normalized form of $\mathbf{P}^p$. In practice, the ESSE scheme utilizes an additive white/red noise and a multiplicative white noise. For each portion of $\mathbf{n}_k^j$ corresponding to a variable v, this amounts to $(\mathbf{n}_k^j = \alpha_k^j + \epsilon_k^j \mathbf{x}_k)_v$ where $(\alpha_k^j)_v$ is a random noise white or red in space and $(\epsilon_k^j)_v$ a non-dimensional white noise factor of amplitude $\ll 1$ (usually 1%–5%). Of course, amplitudes of the noise are limited by numerics since the noise is added to the initial conditions. Based on experience, the difference between using a red or white noise is small because the deterministic ocean model reddens the white noise (by a Shapiro filter in our model). The additive and multiplicative white noise scheme is thus utilized.

In the ESSE system, the amplitudes of $(\alpha_k^j)_v$ and $(\epsilon_k^j)_v$ above can be updated, similarly to the error subspace scaling factors. A batch of ensemble forecasts to t_{k+1} is carried out. If the uncertainty forecast is estimated to be too small (too large) or inadequate, the standard deviations of $(\alpha_k^j)_v$ and $(\epsilon_k^j)_v$ are increased (decreased). Currently, this adaptation is carried out manually, by quick-look evaluation of the standard deviation of the ensemble forecast. This ESSE parameterization of the truncated errors is complementary to the scaling of the error subspace variances (see (i) above) and to the adaptive error learning (Appendix B) which updates the posterior error subspace decomposition.

(iii) *Modification of error covariances based on a Schur product*. Ideally, the size of the ensemble should be as large or larger than the number of degrees of freedom of the dominant uncertain variability. When ensembles are too small, an issue is covariance estimates with too large amplitudes at long distances. This leads to erroneous corrections and artificial reductions of the error variance. With time, the ensemble variance tends to become null. An engineered remedy to this issue is to taper the amplitudes of covariance estimates at long distances by multiplication. This can be achieved by Schur product of the error covariance matrix with a positive definite matrix whose amplitudes decay at long distances (e.g. [21]). This Schur product approach has been successful in weather prediction studies (e.g. [25]). Other localization schemes include the use of a cut-off radius [31] and local DA [58].

Within ESSE, the size of the ensemble is controlled by a criterion (Appendix B, Eq. (B.9)) which can prevent the need to taper covariances. The criterion is based on a measure of the similarity between subspaces, e.g. [29,30]. Such criteria are of course only necessary conditions for convergence. Ensemble sizes are also limited by the speed and storage capacity of the computer system available. If the convergence criterion is not satisfied, the prior error covariance $\mathbf{P}_-^p$ is tapered by Schur product with a matrix $\mathbf{S}$. The ESSE assimilation (Eqs. (B.10)–(B.21) in Appendix B) is then modified by replacing $\mathbf{P}_-^p = \mathbf{E}_- \boldsymbol{\Pi}_- \mathbf{E}_-^{\mathrm{T}}$ with the Schur product $\mathbf{P}_-^p \circ \mathbf{S}$. In our ocean cases, elements of $\mathbf{S}$ are set to vary in the horizontal only. Each observation has then a different tapering covariance function which varies with the data position in the horizontal and is 1 at the data position.

To reduce the increased computational costs due to the Schur product, two approximations are made based on the facts that most measurement operators $\mathbf{H}$ are mainly local interpolations (e.g. bilinear stencil) and that sequential processing of observations is used. First, in the sum $\mathbf{H}\mathbf{P}_-^p\mathbf{H}^{\mathrm{T}} + \mathbf{R}$, the Schur product is ignored. The forecast error covariance at data points $\mathbf{H}\mathbf{P}_-^p\mathbf{H}^{\mathrm{T}}$ is directly added to the data error covariance, because sequential processing of observations is used (data scalars or vertical profiles are assimilated one at a time) and the non-zero elements in each row of $\mathbf{H}$ thus correspond to local interpolation stencils on the model grid which are very close to data points. Second, in computing $(\mathbf{P}_-^p \circ \mathbf{S})\mathbf{H}^{\mathrm{T}}$ which are covariances among model and data points, the sums and multiplications with the non-zero elements of $\mathbf{H}^{\mathrm{T}}$ (interpolation at data points) are simply approximated by the covariance with the stencil point the nearest to the data point. For a bilinear interpolation, only one covariance field is then computed for each data point instead of four.

For predictions of ocean mesoscale physics over $O(100)$ by $O(100)$ km regions, ESSE criteria are usually satisfied for ensemble sizes of 200–600. This is only a necessary condition for convergence. During AOSN-II, ensemble sizes were at times limited by computer speeds and the tapering of $\mathbf{P}_-^p$ was carried out. Elements of tapering matrices $\mathbf{S}$ were defined using an exponential decorrelation function, $\exp(-|\Delta r|/L_s - |\Delta t|/\tau_s)$, where L_s and τ_s are spatial and temporal exponential decay scales, Δr the horizontal distance between a data and stencil (model grid) points, and Δt the difference between the data and forecast times. Its effects are exemplified on Fig. 9 for an ESSE assimilation on Aug 28, 2003, using an ensemble 500 members. The position of the quality-controlled in situ T and S profiles that were assimilated for that day are shown on Fig. 9(a). The ESSE prediction of the error standard deviation for T at the ocean surface is on Fig. 9(b). It is smaller where the data density had been higher (e.g. near the coastline) and larger where dynamics has led to uncertainty growth, e.g. near the coastal front and its meanders, filaments and eddies. The mean surface T uncertainty is about .4 °C. Estimates of the reduction in the error standard deviation of the surface T due to ESSE assimilation are shown on Fig. 9(c) and 9(d), for the case without and with a Schur product correction, respectively. The decay scale used was $L_s = 25$ km, just a bit larger than the estimated mesoscale decorrelation length. The time decay was 100 days (which amounts to no time effects in the Schur product). Near data points (Fig. 9(a)), the two error reductions are almost indistinguishable. Away from data points, for the case without Schur tapering, error reductions are small but not negligible (e.g. ∼.05–.2 °C offshore). However, with tapering, they are smaller (e.g. almost null offshore). The corresponding results of the multivariate 3D ESSE assimilation are shown on Fig. 9(e) and (f), respectively. As for the error reduction estimates, the DA increments for the surface T are very similar

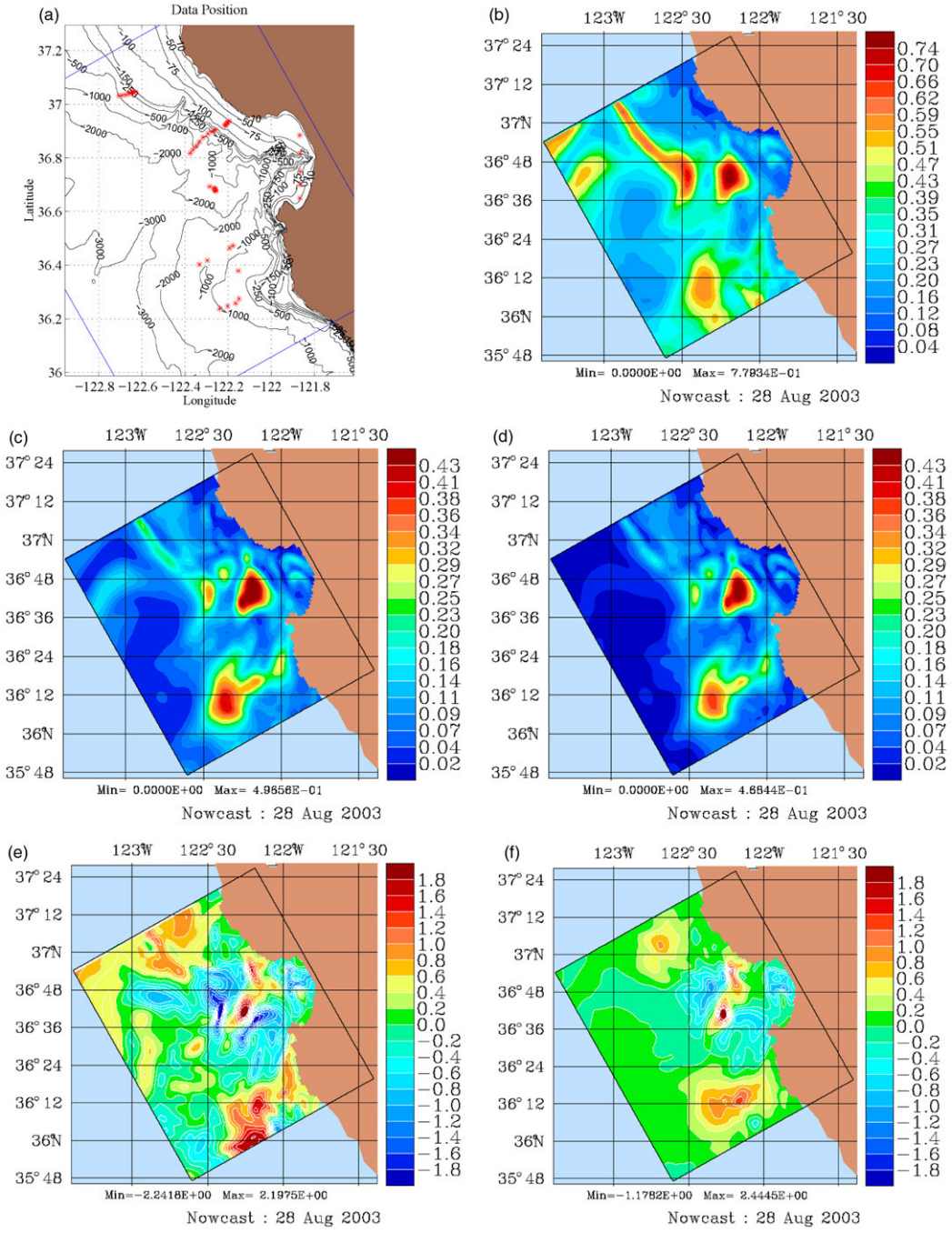

Fig. 9. Illustration of the effect of the Schur product of ESSE error covariances with a matrix whose values decay with distances. (a) Horizontal position of the in situ T and S profiles that are assimilated on Aug 28, 2003. (b) ESSE error standard deviation prediction for the surface T. (c) Reduction of the error standard deviation (prior–posterior error) due to the 3D multivariate ESSE DA of the T and S profiles, without tapering by Schur product. (d) as (c), but with tapering by Schur product (note the smaller reduction of error standard deviations offshore). (e) DA increments for the surface T, without Schur product. (f) as (e), but with Schur product.

near data points. Away from data points, the Schur tapering reduces the amplitudes of the field correction.

After study of other assimilation days (not shown), we find that ESSE increments and correlation estimates suggest that the region is characterized by mesoscale ($\lesssim$20 km) and larger-scale ($\gtrsim$60–100 km) spatial correlations. This is in accord with the mesoscale features (e.g. coastal current and upwelling centers) and the larger-scale California Current System forcing, respectively.

(iv) *ESSE convergence and adaptable ensemble size, error subspace rank and stochastic forcing*. The ensemble size is here controlled by the criterion $\rho \geq \alpha$ (see Eq. (B.9) and above). For the ESSE forecast of Aug. 28, the criterion was evaluated for every new 100 forecast realizations. It was almost satisfied when ensembles of size 500 and 400 were compared: the coefficient ρ was then found equal to 0.96, just under the chosen .97 limit. This suggests that covariance estimates may not need any tapering by Schur product (see Fig. 9 above).

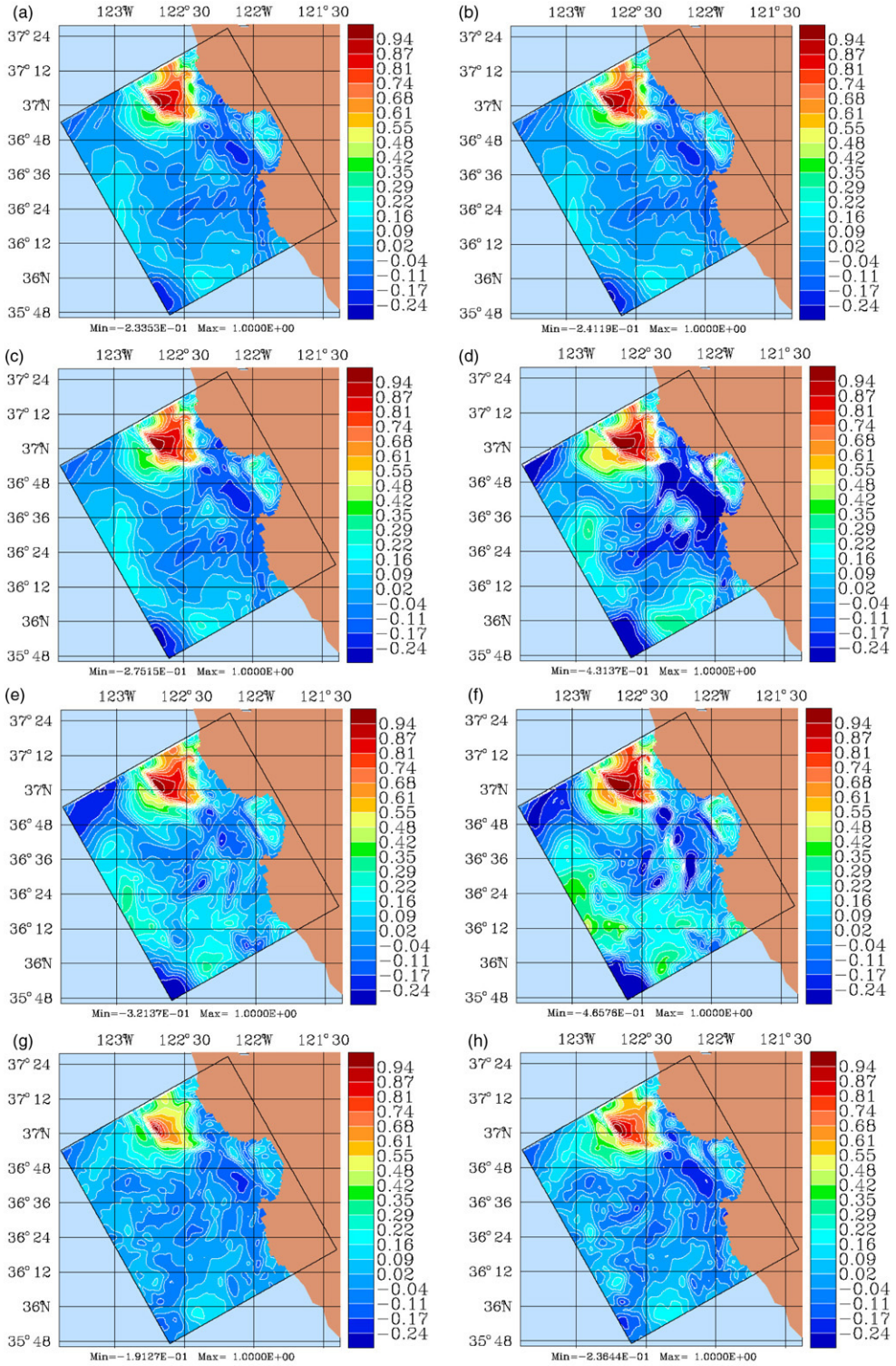

Fig. 10. Sensitivity of error correlation estimates to ensemble size, error subspace rank and stochastic forcing. Shown are two-day ESSE forecasts for Aug 28, 2003 of the error correlation of T at (37.031 N, −122.706 W, 0 m) with T at the surface. (a) Error correlation for an ensemble of size 500. (b) As (a), but using only the dominant rank-300 subspace of the ensemble. (c) As (a), but using the subspace of rank 100. (d) As (a), but using the subspace of rank 20. (e) Error correlation for an ensemble of size 100. (f) As (e), but using only its subspace of rank 20. (g) Error correlation for an ensemble of size 500 with stochastic forcings. (h) As (g), but using only its subspace of size 100.

To illustrate this and impacts of error subspace properties, we gauge next the sensitivity of error correlation estimates to the ensemble size, error subspace rank and stochastic forcing (Fig. 10).

All panels on Fig. 10 are two-day error auto-correlation forecasts for Aug 28, 2003, of T at the surface with T at (37.031 N, −122.706 W, 0 m). This location was chosen because it is on the edge of the equatorward coastal current,

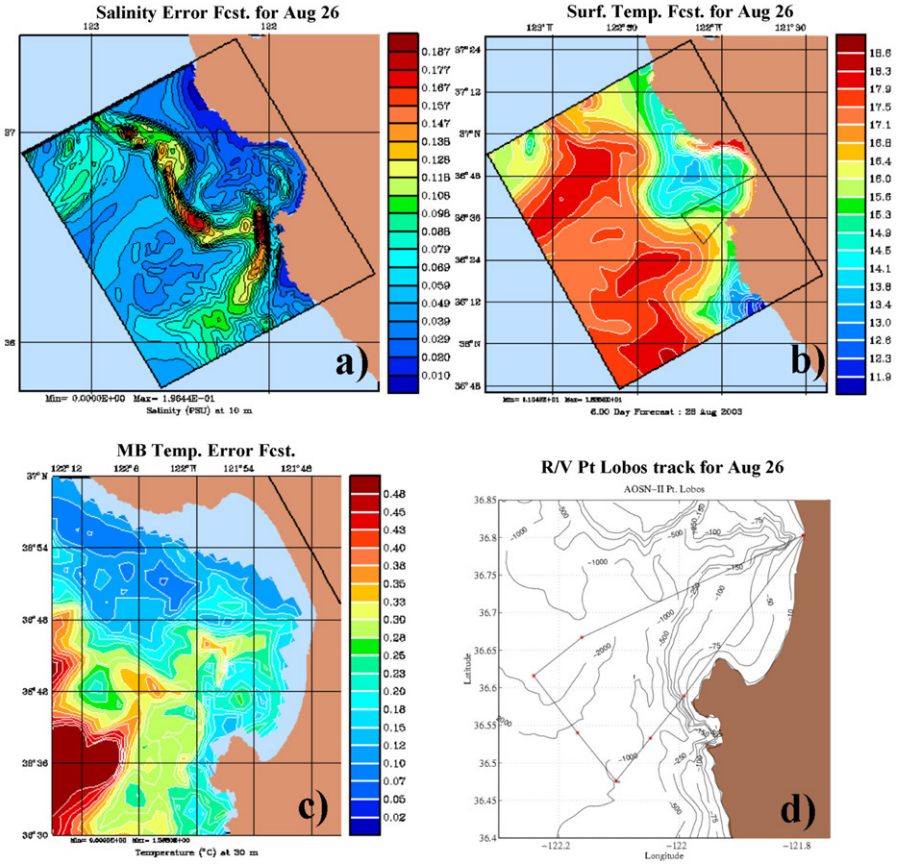

Fig. 11. Real-time adaptive sampling with ESSE and the R/V Pt. Lobos. (a) S error standard deviation forecast at 10 m for Aug 26, indicating the uncertain frontal position. (b) Central forecast of the surface T for Aug 26, indicating the meandering coastal current. (c) T error standard deviation forecast for Aug 26, at 30 m, zoomed over the Monterey Bay region. (d) Sampling plan designed and carried out (Francisco Chavez, personal communication) to investigate the position and strength of the meander and regions of high uncertainty.

the main inflow at the northern boundary of the modeling domain. Fig. 10(a–d) correspond to an ensemble of size 500. Fig. 10(a) is the auto-correlation estimated from the whole ensemble. Fig. 10(b)–(d) are auto-correlations estimated from its dominant singular value decomposition (SVD) of rank 300, 100 and 20, respectively. These dominant SVDs, which define error subspaces, capture 95.8%, 83.1% and 60.4% of the variance explained by the whole ensemble of 500 members, respectively. In accord with ESSE, using a subspace that captures more than 95% of the total variance leads to a correlation estimate (Fig. 10(b)) that is a good approximation of the full estimate (Fig. 10(a)) but that can substantially accelerate computations. Even 83.1% of the variance (Fig. 10(c)) could still be acceptable for fast ESSE analyses. However, using 60% of the variance (20 vectors, Fig. 10(d)) is not sufficient: the correlation estimate still remains accurate at short distances but then overshoots substantially at long distances, in accord with a Gibbs-like phenomenon. Fig. 10(e) shows the auto-correlation estimated from an ensemble size of only 100 (the first 100 members of the whole ensemble). In comparison with Fig. 10(a), an ensemble size of 100 is not large enough to estimate the correlation field: correlations are relatively similar at shorter distances but overshoot at longer distances. In fact, the convergence criterion (Appendix B) gives a similarity coefficient of only 76% between the ensembles of sizes 100 and 500, indicating non-convergence at 100. Note also that the estimate on Fig. 10(c) is more accurate than this Fig. 10(e) because Fig. 10(c) contains information from the other 400 members. Fig. 10(f) uses only the dominant 20 singular vectors of the small 100 member ensemble: its correlation estimate overshoots the most.

Comparing Fig. 10(a)–(f), one notices that amplitudes of correlation estimates decay at longer distances as the size of the ensemble or rank of its subspace increases. For example, consider the larger-scale positive correlation field offshore (from about 36 N, 122 W to 36.7 N, 123 W), the negative correlation patch in front of Monterey Bay (near 36.7 N, 122.25 W) and the positive correlations centers in the Bay itself. These long-distance correlations are still present in Fig. 10(a), (b) but with much smaller amplitudes. This confirms the needs for tapering at long distances when ensemble sizes are too small. Here, based on the similarity criterion (Appendix B), a size of 500 is almost large enough for 97% similarity.

Stochastic forcing (Fig. 10(g) and (h) also affect error covariances and correlations. Fig. 10(g) is the auto-correlation estimated from an ensemble of size 500, with stochastic forcing (Appendix A.2) during the two-day integrations of the PE model. Fig. 10(h) shows the same, but using only the error subspace of rank 100 (dominant SVD that captures 72.1% of the total variance explained). The stochastic forcing intend

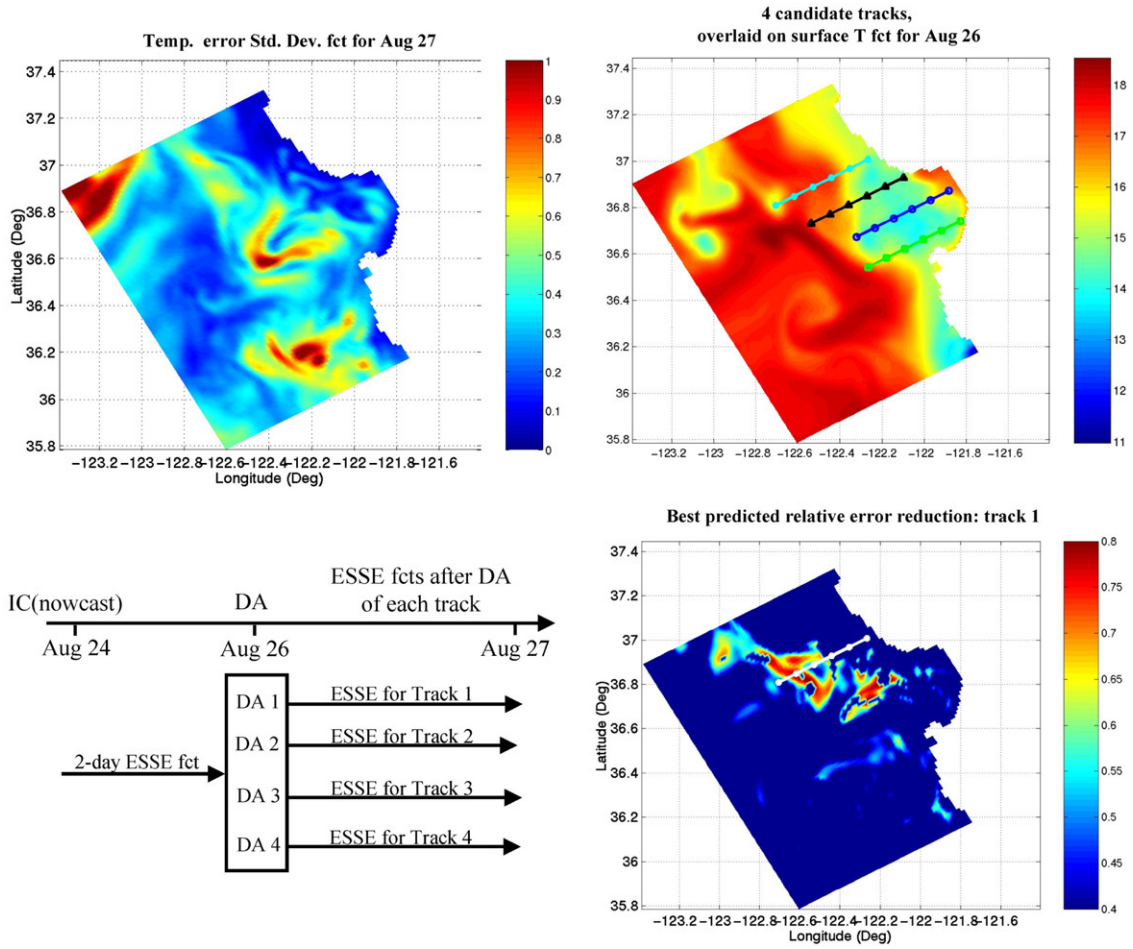

Fig. 12. Adaptive sampling via ESSE, for an Aug. 24 – Aug. 27 hindcast. (a) T error standard deviation forecast for Aug 27, at 10 m. (b) Central forecast of the surface T for Aug 26, overlaid with the four candidate tracks. (c) Time line for the ESSE scheme and computations. (d) Predicted optimal track for Aug. 26 and its relative error reduction for Aug. 27.

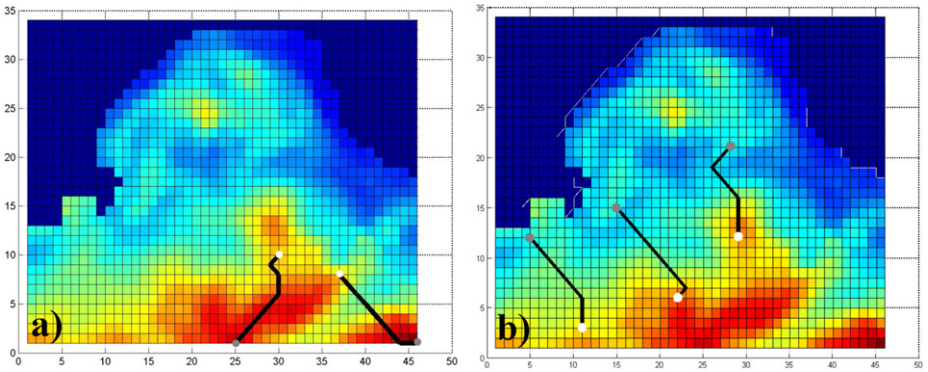

Fig. 13. Generation of sampling paths for fixed objective fields [74] using Mixed-Integer Programming (MIP). Examples are for two and three AUVs and a 2D objective field (the ESSE T error standard deviation averaged over the 0–50 m depths). Grey dots are starting points for the AUVs and white dots are the MIP optimal termination points. The ranges of the AUVs are set be around 20 km. The paths are constrained by a vicinity constraint such that vehicles remain away from each other by at least 10 km. (a) Optimal path of two vehicles. (b) Optimal path of three vehicles.

to represent the statistics of sub-grid-scale processes which have shorter correlations scales. In accord with measurements, they lead to noisier correlation fields (compare Fig. 10(a) and (g)) and reduce the main correlation length around (37.031 N, −122.706 W). However, note that the larger-scale correlation patches at longer distance remain on average similar. Comparing Fig. 10(g) and (h), the reduction to an error subspace of rank 100 modifies the correlation more than in the case without stochastic forcing (Fig. 10(a) and (c)). This is because the stochastic forcings flatten the error spectrum,

as confirmed by the percentages of variance explained, 83.1% and 72.1%, in the cases without and with stochastic forcings, respectively.

(v) *Adaptive stochastic model*. The amplitudes of the stochastic forcing representing uncertainties in the dynamical model can vary with time and be tuned. Details of the stochastic model are outlined in Appendix A.2. Parameters that can be modified in the current model are: the decorrelation time β and the amplitudes of model errors relative to the dominant dynamical model balance. These relative amplitudes are set by the factors $\epsilon_U, \epsilon_T, \epsilon_S$ and ϵ_ψ (see Eqs. (A.13a)–(A.13d)). In theory, the vertically varying characteristic scales in these Eqs. (A.13a)–(A.13d) could also be tuned or evolved as function of the background ocean regime. However, this has not yet been carried out: the noise factors are presently set at t_0 based on the initial state estimate.

4. Adaptive sampling

There are many facets to the adaptive sampling problem. Ideally, it should account for ocean dynamics, including nonlinear and coupled processes. Optimal sensors and sampling patterns should focus on regions of greatest dynamic variability and uncertainties, compounding model, data, initial condition and forcing error sources. In Bayesian estimation, the probability density of an ocean state, conditioned on data and governed by dynamics, is evolved by a prognostic equation which contains all these effects [34,16]. The objective of adaptive sampling can then be to optimize certain properties of this conditional probability density, for example, minimize specific future error variances, maximize the knowledge of dynamical hot-spots or optimize coverage of an area. Adaptive sampling can occur on various time and space scales, because of the dynamics itself but also because of the diverse capabilities of platforms and sensors. Operational constraints have to be represented. In setting up the problem, diverse assumptions can be made, e.g. assume a fixed or a variable environment, utilize an objective independent or dependent of unknown future data values. Finally, there are many optimization schemes to choose from.

The new adaptive sampling examples presented next illustrate these various facets. They are based on simulations for the AOSN-II exercise in the Monterey Bay and CCS region. The first one (Section 4.1) was carried out in real time. It aims to select tomorrow's sampling path that reduces tomorrow's uncertainties the most. This problem is radically simplified by assuming that the error reduction is instantaneous and by letting the user select the sampling path heuristically using uncertainty predictions. The other two examples are hindcast studies, but are directly applicable to real-time computations. They show the results of two new adaptive sampling schemes. The first scheme (Section 4.2) is a novel nonlinear ESSE adaptive sampling. It properly evolves fields and errors nonlinearly during the optimization and so predicts the impacts of data. Its only assumption is to pre-determine a number of candidate sampling regions or paths that is small enough so that the optimum can be computed in real time. The second new scheme (Section 4.3) rapidly generates the optimal path exactly, under the approximation that ESSE predictions remain constant and unaffected by the sampling. When this approximation is applicable, it simplifies the adaptive sampling problem such that its exact solution can be found.

4.1. Heuristic real-time adaptive sampling based on stochastic ensemble predictions

During AOSN-II, field and error forecasts were used with a priori experience to intuitively choose the future sampling. It is illustrated on Fig. 11. The objective is to find the sampling path to be carried out tomorrow that will most reduce errors in tomorrow's analysis. It ignores that sampling the ocean is not instantaneous. It assumes that uncertainties and ocean fields do not vary during the 8 h of sampling by the research vessel. It is a final time optimization problem.

On Aug. 25, the two-day central forecast for Aug. 26 predicted a meander of the coastal current (Fig. 11(b)) that advected warm and fresh waters towards the southern portions of the Bay (facing the Monterey Bay Peninsula). However, the salinity and temperature error fields (Fig. 11(a) and (c)) from a 450-member ESSE ensemble indicated a high degree of uncertainty in both the frontal position (Fig. 11(a)) and strength of the meander (Fig. 11(c), at the depth of the main thermocline). In fact, some ensemble members had either essentially no meander present or a meander shifted to the north. Data were needed west of the Peninsula. Based on this information, and constrained by operational limitations, a sampling pattern (Fig. 11(d)) was devised for the R/V Pt. Lobos. Once the 7 temperature and salinity profiles were sampled, it was found that the meander was most likely very weak. The assimilation of these data substantially reduced the meander. In fact, a hindcast (not shown) without the R/V Pt. Lobos but with re-calibrated glider data for Aug. 20–23 also weakened the meander.

4.2. Quantitative adaptive sampling via ESSE

The second example is a hindcast using a new fully fledged ESSE adaptive sampling scheme. The objective is to minimize the trace of the predicted error covariance matrix. Importantly, (i) the objective field is affected by the properties of the data to be collected and (ii) ocean fields and uncertainties change with time, as a function of the synthetic data assimilated and predicted ocean dynamics. The model of model errors (Appendix A.2), here homogeneous in the horizontal, ensures that the optimization for coverage is also accounted for. The time scales of the sampling considered are a day to a week. To reduce computational costs, a number of candidate sampling tracks or regions are pre-selected, based on available error forecasts and known operational constraints.

The generic evolution equations for the dynamical model and error covariances are given in Appendix B (Eqs. (B.1) and ((B.2)), respectively). The ESSE adaptive sampling objective is to find the future observation patterns $\mathbf{H}_i$, given the measurement uncertainties $\mathbf{R}_i$, that minimize the trace of the

error covariance matrix at final time or on average during the integration time:

$$\mathrm{Min}_{\mathbf{H}_i,\mathbf{R}_i}\ \mathrm{tr}(\mathbf{P}(t_f)) \quad \text{or} \tag{5a}$$

$$\mathrm{Min}_{\mathbf{H}_i,\mathbf{R}_i} \int_{t_0}^{t_f} \mathrm{tr}(\mathbf{P}(t))\mathrm{d}t. \tag{5b}$$

To address this problem, ESSE accounts for the nonlinear error covariance evolution, including impacts of synoptic DA at time t and their nonlinear evolution beyond time t. It is based on tree-structured multi-ensemble prediction.

Let us first consider the case of choosing the sampling path to be carried out at or near time $t < t_f$, with the objective of minimizing the error covariance at final time t_f (Eq. (2a)). One first selects a set of candidate sampling paths and data properties for each of these paths. Using the ESSE ensemble forecast for first observation time t, synthetic data (taken from the most probable forecast or from the ensemble mean) are assimilated via ESSE for each of these paths. The result is an analysis and its errors (the posterior ensemble) for each of the path. New ensemble forecasts are then started from each of these new analyses, to provide the updated nonlinear error covariance prediction for final time t_f. The sampling path that provides the largest error reduction at t_f is optimal among the pre-selected paths. This novel approach can be generalized easily to more complex cases, including integrated costs Eq. (2b). In general, a tree of ensembles with DA of synthetic data is evolved forward in time. Each branch of the tree corresponds to a candidate sampling plan. The information corresponding to each branch is the error reduction corresponding to that plan. The best candidate is the one that reduces an uncertainty metric the most.

In the example shown (Fig. 12), the problem is to compute today, the tracks to sample tomorrow, that will most reduce uncertainties the day after tomorrow. Precisely, the question asked is: considering an ESSE nowcast for Aug. 24 and prediction up to Aug. 27, which sampling track should be carried on Aug. 26 so as to optimally reduces uncertainties on Aug. 27? The new ESSE computations are schematized on Fig. 12(c). The error standard deviation for T at 10 m on Aug 27 is shown on Fig. 12(a) and the surface T on Aug 26 on Fig. 12(b). Four candidate tracks for Aug 26 are selected (Fig. 12(b)). After DA on Aug 26, four ESSE ensemble forecasts are computed for Aug 27, one for each track. These forecast ensembles can be relatively small since error variances usually converge faster than error covariances. The track on Aug 26 that reduces errors the most on Aug. 27 is the most northern one. The prediction of its relative error reduction on Aug. 27 is shown on Fig. 12(d). Note that this path crossed the coastal current the most and led to surface corrections that are elongated in the directions of main advections, both upstream (due to DA on Aug 26) and downstream (due to Aug 26–27 advections).

4.3. Optimal paths generation for a fixed objective field

The new second scheme computes the optimal paths exactly and rapidly by approximating the full objective field by a fixed, univariate and two-dimensional field [74]. The example provided is a hindcast but the scheme can be used in real time.

The objective is here to minimize the error standard deviation of a 2D ocean field. The hypotheses are that: (i) the speed of platforms (e.g. AUVs) are much larger than the time rate of change of environment, and (ii), the objective field is fixed during the computation of the path and is not affected by new data. The time scales of the adaptive sampling considered are hours. The practical problem solved is: assuming the error field is like that now and will remain so for the next few hours, where do we send the AUVs?

An exact solution is obtained by Mixed-Integer Programming (MIP, [56,65]) if the objective field is discretized as a piecewise-linear cost function (MIP is optimization approach for linear cost functions and constraints where some of the unknowns are integers). The approach is then a modified "traveling salesman" problem. The towns to be visited are hot-spots in discretized fields and the salesmen are the AUVs. The optimization is here carried out using the tool Xpress-MP from Dash [14]. An illustration is shown on Fig. 13 for the case of two and three AUVs. The objective field is the vertically averaged (0–50 m depths) error standard deviation of temperature around Monterey Bay. The ranges of the AUVs are set to be near 20 km and the distance between 2 AUVs is constrained to be larger than 10 km. The MIP computations take 2–30 min on today's computers. Results indicate that this scheme could be very useful for rapid adaptive sampling in marine operations.

5. Conclusions

Concepts of adaptive modeling, adaptive data assimilation and adaptive sampling were introduced, discussed and exemplified within the context of comprehensive real-time ocean observing and prediction systems. For a more unified presentation, illustrations were provided for the Monterey Bay and CCS region but the adaptive schemes outlined are being utilized in other regions. Novel but simple adaptive modeling approaches for hypothesis testing were first developed and applied to physical and physical–biogeochemical dynamics. The schemes are based on an ensemble of simulations using different model structures and parameter values, and on the automated selection of the model version with the best predictive skill. For physical adaptive modeling, predictive skill was measured by a weighted average of the bias, root mean square error and pattern correlation coefficient of model estimates. Adaptable components of the ESSE system were then reviewed, discussed and illustrated. This new synthesis of adaptable ESSE components includes error estimates (initial, prior and posterior covariance re-scaling, models of truncated errors, etc), ensemble sizes, error subspace ranks, covariance tapering by Schur product and parameters of stochastic error models. Such components are calibrated by quantitative adaptation. Illustrations were provided for influences of ensemble size, error subspace rank, covariance tapering and stochastic forcing. Adaptive sampling approaches and schemes were then outlined and their capabilities illustrated. The first scheme was heuristic, selecting the sampling manually based

on predictions and experience. The other two new schemes are quantitative. Adaptive sampling via ESSE selects the optimal sampling plan out of a set of candidates while the second scheme computes the optimum sampling path assuming a fixed objective.

Adaptive modeling is an extension of classic parameter estimations to the selection of model state variables, linkages, parameterizations and/or computational characteristics. The optimization is over both model structures and parameter values. The model changes and learns from data as they are collected. Adaptive modeling appears especially valuable for the parameterizations of sub-grid-scale effects and for biological investigations, two research areas for which fundamental models are being debated. The present Monte Carlo "trial and error" approach used for adaptive modeling is simple but already useful. It is quantitative, accounts for nonlinear effects and is representative of what is done manually by a human tuning a model. However, many other adaptive modeling methods need to be evaluated and used in the future. If a tangent linear model is a good approximation, a possibility is to use Green functions or adjoint equations for everything that is different in each of the model structures and parameter values, and optimize over the corresponding Euler–Lagrange equations, with weak or strong constraints.

Several aspects of adaptive modeling are computationally challenging [45] and require more research. Adaptation triggers can be provided to the software through external files which are checked during runtime and updated when model properties need to be modified. For example, for hypothesis testing (Fig. 5, left), the forecasts of competing models are compared based on their respective data–forecast misfits and the best functional forms selected for each data assimilation period. In the second category (Fig. 5, right), the code may modify parameters, increase the number of state variables (e.g. de-aggregation) or alter the form of a source term. Such structural adaptations are not trivial. They can be implemented using C function pointers to choose among Fortran modules. To account for forecast uncertainties, ESSE can be used and provide expected bounds on misfits. For each candidate model, these uncertainty bounds can be computed based on a small-size ensemble. This is because error variances often converge faster than covariances.

The first adaptive sampling scheme illustrated is the simple heuristic estimation of the ideal future sampling based on quantitative predictions of ocean fields and uncertainties. This is an approach often followed today in oceanic studies. However, it has the disadvantages of being manual and of usually ignoring the impact of the sampling plans on ocean estimates subsequent to that sampling. Adaptive sampling via ESSE is based on the nonlinear prediction of the impact of future observations on the predicted ocean state and uncertainty. A set of candidate sampling plans (data locations and types) is chosen a priori. For each plan, a nonlinear ESSE prediction and data assimilation is carried out. This estimates the impact of each plan on predicted fields and so the candidate plan with the best impact. Adaptive sampling via ESSE is theoretically a better algorithm but it requires prior selection of a finite set of candidate sampling plans. The last adaptive sampling scheme neglects the evolution of the objective field for the duration of the sampling, but computes the optimal sampling plan exactly. This is feasible by Mixed-Integer Programming. It is the quantitative version of the above heuristic adaptive sampling approach.

The present schemes for adaptive sampling and modeling were based on ensemble approaches and the ESSE system. An extension of these schemes is based on genetic algorithms, where new versions of the model or the sampling plans are generated based on the skill of previous generations, e.g. a previous ensemble of model versions or of sampling plans. There are also many ways to carry out adaptive DA and the adaptive ESSE components presented here were representative. Because of insufficient or limited data, the use of adaptive schemes in ocean studies can be challenging. Quantitative adaptive filters that work in engineering applications where the number of independent observations is typically large when compared to the number of control parameters may not work directly in oceanography. This does not mean however that adaptive schemes should not be employed. In fact, in oceanography, prior estimates can often be quite far from data. The adaptation of models, error models and assimilation schemes is then required.

The use of adaptive schemes in oceanography is recent and multiple research questions remain, from theoretical to applied subjects. Developing adaptive schemes and systems that quantify and automate the learning process of the ocean researcher should become more and more fruitful in the years to come.

Acknowledgments

I am grateful to K. Ide and C. Jones for inviting me to prepare a manuscript for this special issue of Physica D on Data Assimilation. I thank the Institute for Pure and Applied Mathematics (IPAM) at UCLA for their support for attending the workshop on "Mathematical Issues and Challenges in Data Assimilation for Geophysical Systems: Interdisciplinary Perspectives". I am grateful to P.J. Haley, W.G. Leslie, A.R. Robinson, J.J. McCarthy, R. Tian, N. Yilmaz, N. Patrikalakis, H. Schmidt and B. Miller for discussions and their collaboration. I thank the whole AOSN-II team including N. Leonard, S. Ramp, R. Davis, D. Fratantoni, J. Marsden, J. Bellingham, I. Shulman, Y. Chao, S. Majumdar and F. Lekien. I appreciated the useful comments and suggestions from the three anonymous reviewers. I thank T. Lehman and K. Ide for their skillful and patient editing. This work was funded by the Office of Naval Research under grant N00014-05-1-0335, N00014-05-1-0370, N00014-04-1-0534 and No. S05-06 (PLUSNet), and by Dr. F. Darema, under grant NSF/ITR EIA-0121263.

Appendix A. Stochastic–deterministic interdisciplinary ocean models

A.1. Deterministic coupled models

Physical Model. The PE model (Eqs. (A.1)–(A.7), of the Harvard Ocean Prediction System (HOPS, [27]) is used. The

state variables are temperature T, salinity S and horizontal total velocities $\mathbf{u}_h$. In Eqs. (A.1)–(A.7), $(x, y, z, t = \mathbf{r}, z, t)$ is the position vector and time, $\mathbf{u} = (\mathbf{u}_h, w)$ the velocity vector, ρ the density, ρ_0 the density of a state of reference, g gravity, f the Coriolis frequency, p_w the water pressure, A_v and K_v vertical eddy viscosities and diffusivities, and A_h and K_h their horizontal counterpart (modeled by a scale dependent Shapiro filter). Atmospheric fluxes from external atmospheric models are imposed at the surface (sometimes with a flux correction). Possible outputs consist of a wide range of prognostic and diagnostic variables and parameters.

Horiz. Mom.

$$\frac{D\mathbf{u}_h}{Dt} + f\mathbf{e}_3 \wedge \mathbf{u}_h = -\frac{1}{\rho_0}\nabla_h p_w + \nabla_h \cdot (A_h \nabla_h \mathbf{u}_h) + \frac{\partial A_v \partial \mathbf{u}_h/\partial z}{\partial z}. \quad \text{(A.1–2)}$$

Vert. Mom. $$\rho g + \frac{\partial p_w}{\partial z} = 0. \quad \text{(A.3)}$$

Thermal energy $$\frac{DT}{Dt} = \nabla_h \cdot (K_h \nabla_h T) + \frac{\partial K_v \partial T/\partial z}{\partial z}. \quad \text{(A.4)}$$

Cons. of salt $$\frac{DS}{Dt} = \nabla_h \cdot (K_h \nabla_h S) + \frac{\partial K_v \partial S/\partial z}{\partial z}. \quad \text{(A.5)}$$

Cons. of mass $$\nabla \cdot \mathbf{u} = 0. \quad \text{(A.6)}$$

Eqn. of state $$\rho(\mathbf{r}, z, t) = \rho(T, S, p_w). \quad \text{(A.7)}$$

The parameters of the PE model which were adapted in real time during AOSN-II (Section 2) included the profiles of vertical eddy viscosities A_v and diffusivities K_v, which are a function of space, time and local physical fields, see [42] and references therein. In the surface layers, a mixing-layer model is used to transfer and dissipate the atmospheric forcings (here COAMPS fluxes). This model first evaluates the local depth of turbulent wind-mixing or "Ekman depth" $h^e(x, y, t) = E_k \left(\sqrt{\|\boldsymbol{\tau}\|/\rho_0}\right)/f$ where $\boldsymbol{\tau}(x, y, t)$ is the wind stress vector, E_k an empirical factor, and ρ_0 and f as above. This depth h^e is constrained within adjustable bounds. Vertical eddy coefficients within h^e are empirically set to the scalars A_v^e and K_v^e. Below h^e, A_v and K_v are estimated based on the local gradient Richardson number Ri. Where $Ri(x, y, z, t)$ is ≥ 0, they are set to $A_v = A_v^b + \frac{\nu_0}{(1+5Ri)^2}$ and $K_v = K_v^b + \frac{\nu_0}{(1+5Ri)^3}$. In this shear vertical mixing scheme, the adjustable parameters are the background coefficients, A_v^b and K_v^b, and shear eddy viscosity at $Ri = 0$, denoted by ν_0. For negative Ri's, convective mixing values A_v^{cvct} and K_v^{cvct} are utilized.

Biogeochemical Model. The model stems from advection–diffusion–reaction (ADR) equations [70]. Its state variables $\phi_i(x, y, z, t)$ are governed by

$$\frac{\partial \phi_i}{\partial t} + \mathbf{u} \cdot \nabla \phi_i - \nabla_h (A_i \nabla_h \phi_i) - \frac{\partial K_i \partial \phi_i/\partial z}{\partial z} = \mathcal{B}_i(\phi_1, \ldots, \phi_i, \ldots, \phi_7). \quad \text{(A.8)}$$

In Eq. (A.8), the first term is the local time change at a point, the second is advection, and the third and fourth are eddy diffusions. The $\mathbf{u}$ are velocities, the K_i's eddy diffusivities and the $\mathcal{B}_i$'s biological dynamics or reactions. The latter model the sources and sinks of ϕ_i's due to, e.g., reproduction, life-stage transitions, natural mortality, predation, chemical reactions and behavior. Details on the different options that are coded in the numerical generalized model and can represent such processes are given in [70].

A.2. Stochastic models

In the present applications, a stochastic model is utilized to represent model errors in physical equations. These stochastic forcings aim to account for statistical effects of processes (e.g. sub-mesoscales) not resolved in the deterministic PEs (Eqs. (A1-2–A.7)). They are correlated in time and space, and added to the prognostic equations based on the following construction.

(a) *Time correlations*. In time, model errors are assumed to be stationary, zero-mean, random processes, exponentially decorrelated. To generate such processes, white noise is passed through a simple feedback equation. For a scalar noise $\tilde{w}(t)$, the sample path and variance equations are,

$$\mathrm{d}\tilde{w} + \beta \tilde{w} \mathrm{d}t = \mathrm{d}w \quad \text{(A.9)}$$

$$\dot{p}_{\tilde{w}} = -2\beta p_{\tilde{w}} + q \quad \text{(A.10)}$$

where $\frac{1}{\beta}$ is the autocorrelation time and $w \sim (0, q)$ a Gaussian noise white in time. Setting $\dot{p}_{\tilde{w}}$ to zero at all times yields $p_{\tilde{w}}(0) = \sigma^2 = \frac{q}{2\beta}$. For a stationary process (fixed fluctuation amplitude σ and autocorrelation time $\frac{1}{\beta}$), the variance of w is set to be a constant $q = 2\beta\sigma^2$.

(b) *3D Spatial covariances*. In space, model errors are set to be of vertically varying amplitude and of approximately two-grid-point decorrelation in all spatial directions. This is obtained from a white noise on a coarser grid, linearly interpolated onto the actual finer model grid. The result is an approximate two-grid-point correlated noise. For each prognostic equation, the noise variance at a given level is chosen equal to a small ϵ fraction of the amplitude of the terms involved in the dominant PE balance at that level (ϵ is usually set within .1 –.3).

(c) *Stochastic PE Model*. In continuous time and discrete space, equations are

$$\mathrm{d}\mathbf{x} = \mathcal{M}(\mathbf{x}, t)\mathrm{d}t + \mathbf{B}^{fc}(t)\mathrm{d}\tilde{\mathbf{w}}^c \quad \text{(A.11)}$$

$$\mathrm{d}\tilde{\mathbf{w}}^c = -\boldsymbol{\beta}^c \tilde{\mathbf{w}}^c \mathrm{d}t + \mathrm{d}\mathbf{w}^c \quad \text{(A.12)}$$

where: $\mathbf{x} \in R^n$ is the discrete-space PE state vector; $\mathcal{M}(\cdot, t)$ the discrete PE operator; $\mathbf{w}_k^c$ the multivariate coarse 3D white noise; $\tilde{\mathbf{w}}_k^c$ the coarse 3D Gauss–Markov process; and $\mathbf{B}^{fc}(t)$ the coarse-to-fine linear interpolation operator. The discrete PE model [27,13] state is $\mathbf{x} = (\hat{\mathbf{u}}, \hat{\mathbf{v}}, \mathbf{T}, \mathbf{S}, \mathbf{p})^{\mathrm{T}}$, where the vectors $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$ are internal horizontal zonal and meridional velocities, $\mathbf{T}$ temperatures, $\mathbf{S}$ salinities and $\mathbf{p}$ barotropic streamfunctions (discretized ψ). The vector components of $\mathbf{w}^c$ are $(\mathbf{w}_{\hat{u}}^c, \mathbf{w}_{\hat{v}}^c, \mathbf{w}_T^c, \mathbf{w}_S^c, \mathbf{w}_\psi^c)^{\mathrm{T}}$. Similar definitions hold for the correlated $\tilde{\mathbf{w}}^c$. Its diagonal sub-matrices of time decorrelations, $(\boldsymbol{\beta}_u, \boldsymbol{\beta}_v, \boldsymbol{\beta}_T, \boldsymbol{\beta}_S, \boldsymbol{\beta}_\psi) = \boldsymbol{\beta}^c$, are here set simply to $\boldsymbol{\beta}_X = \beta\mathbf{I}$. The noise decorrelation time is thus assumed homogeneous in space and across state variables. The diagonal sub-matrices

of noise variances, $(\boldsymbol{\Sigma}_u, \boldsymbol{\Sigma}_v, \boldsymbol{\Sigma}_T, \boldsymbol{\Sigma}_S, \boldsymbol{\Sigma}_\psi) = \boldsymbol{\Sigma}_{\tilde{\mathbf{w}}^c}$, are here functions of z only. Their amplitudes are set to "$\epsilon * \|\text{geostrophy}(z)\|$", which leads to:

$$\boldsymbol{\Sigma}_u = \boldsymbol{\Sigma}_v = \sigma^2(z)\mathbf{I}, \quad \text{with } \sigma_U(z) = \epsilon_U f_c U(z) \tag{A.13a}$$

$$\boldsymbol{\Sigma}_T = \sigma_T^2(z)\mathbf{I}, \quad \text{with } \sigma_T(z) = \epsilon_T U(z)\frac{\Delta T(z)}{L(z)} \tag{A.13b}$$

$$\boldsymbol{\Sigma}_S = \sigma_S^2(z)\mathbf{I}, \quad \text{with } \sigma_S(z) = \epsilon_S U(z)\frac{\Delta S(z)}{L(z)} \tag{A.13c}$$

$$\boldsymbol{\Sigma}_\psi = \sigma_\psi^2(z)\mathbf{I}, \quad \text{with } \sigma_\psi(z) = \epsilon_\psi \frac{\overline{\omega} L(z)}{U(z)} \tag{A.13d}$$

where f_c is the central Coriolis frequency. The values of the vertically varying characteristic scales, $U(z)$, $\Delta T(z)$, $\Delta S(z)$, $L(z)$ and $\overline{\omega}$, and corresponding ϵ_X's are estimated from a combination of observations and model runs.

Appendix B. Error subspace statistical estimation

B.1. Formalism

Using continuous–discrete Bayesian estimation [34] and the notation of [33], the spatially discretized version of the deterministic–stochastic ocean model and parameter equations are combined into a single equation for the augmented state vector $\mathbf{x}$, of large but finite dimensions. Observations are taken at discrete instants $t_k \geq t_0$ and are concatenated into a data vector $\mathbf{y}_k^o$. The dynamics, observations and DA criterion are then,

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

$$\mathbf{y}_k^o = \mathcal{H}(\mathbf{x}_k, t_k) + \boldsymbol{\epsilon}_k \tag{B.1b}$$

$$\min_{\mathbf{x}} J(\mathbf{x}, \mathbf{y}_k^o; \mathrm{d}\eta, \epsilon_k, \mathbf{Q}(t), \mathbf{R}_k) \tag{B.1c}$$

where $\mathcal{M}$ and $\mathcal{H}$ are the model and measurement model operator, respectively, J the objective function, and $\mathrm{d}\eta$ Wiener processes (Brownian motion), i.e. $\beta \sim \mathcal{N}(0, \mathbf{Q}(t))$ with $\mathcal{E}\{\mathrm{d}\eta(t)\mathrm{d}\eta^{\mathrm{T}}(t)\} \doteq \mathbf{Q}(t)\mathrm{d}t$. Note that the deterministic ocean dynamics and parameter equations are actually forced by noise processes correlated in time and space. State augmentation [34, 22,20] is used to re-write equations in the form of Eq. (B.1a) which are forced by intermediary processes $\mathrm{d}\eta$ white in time and space. Measurement model uncertainties $\boldsymbol{\epsilon}_k$ are assumed white Gaussian sequences, $\boldsymbol{\epsilon}_k \sim \mathcal{N}(0, \mathbf{R}_k)$. The initial conditions have a prior PDF, $p(\mathbf{x}(t_0))$, i.e. $\mathbf{x}(t_0) = \widehat{\mathbf{x}}_0 + \mathbf{n}(0)$ with $\mathbf{n}(0)$ random.

Defining the data collected prior to time t by $\mathbf{y}_{t-}$, the conditional PDF of $\mathbf{x}(t)$, $p(\mathbf{x}, t|\mathbf{y}_{t-})$, is the Bayesian prediction to time t. The Fokker–Planck equation [34,19] governs $p(\mathbf{x}, t|\mathbf{y}_{t-})$ which is itself random since it depends on data values prior to t. For realistic ocean applications, the focus has been on sub-optimal estimates of the conditional mean and error covariance matrix $\mathbf{P} = \mathcal{E}\{(\mathbf{x} - \widehat{\mathbf{x}})(\mathbf{x} - \widehat{\mathbf{x}})^{\mathrm{T}}\}$. One approximation is to aim for minimum error variance. The objective function J in Eq. (B.1c) is then the trace of the *a posteriori* error covariance $\mathbf{P}_k(+)$: the goal is to find $\mathbf{x}_k$ such that $J_k = \mathrm{tr}[\mathbf{P}_k(+)]$ is minimized using $[\mathbf{y}_0, \ldots, \mathbf{y}_k/\mathbf{y}_N]$ for filtering/smoothing. For accurate approximations, accurate estimates of $\mathbf{P}$ are required. In between observations, $\mathbf{P}$ is governed by Eq. (B.2a) where $\langle\cdot\rangle$ denotes expectations. It is obtained from Eq. (B.1a) and the Itô rule [34,22]. The update of $\mathbf{P}$ at data times t_k is given by Eq. (B.2b) where $\langle\cdot\rangle_-$ are expectations over $p(\mathbf{x}, t|\mathbf{y}_{t-})$.

$$\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}t} = \langle(\mathbf{x} - \widehat{\mathbf{x}})(\mathcal{M}(\mathbf{x}) - \mathcal{M}(\widehat{\mathbf{x}}))^{\mathrm{T}}\rangle + \langle(\mathcal{M}(\mathbf{x}) - \mathcal{M}(\widehat{\mathbf{x}}))(\mathbf{x} - \widehat{\mathbf{x}})^{\mathrm{T}}\rangle + \mathbf{Q} \tag{B.2a}$$

$$\mathbf{P}_k(+) = \frac{\langle \mathbf{x}_k \mathbf{x}_k^{\mathrm{T}} p(\mathbf{y}_k^o|\mathbf{x}_k)\rangle_-}{\langle p(\mathbf{y}_k^o|\mathbf{x}_k)\rangle_-} - \widehat{\mathbf{x}}_k(+)\widehat{\mathbf{x}}_k(+)^{\mathrm{T}}. \tag{B.2b}$$

The evolution of $\mathbf{P}$ depends on four factors: (i) its initial condition $\mathbf{P}(0)$; (ii) the deterministic dynamics, first two terms in the RHS of Eq. (B.2a); (iii) the model uncertainties which increase variance, last term in Eq. (B.2a); and (iv) the data impacts which reduce variance, Eq. (B.2b). Importantly for adaptive sampling, $\mathbf{P}$ depends on data and state values prior to t. It is only for linear systems that the covariance evolution does not depend on these values.

Error Subspace Statistical Estimation (ESSE, [38,44]) intends to estimate and predict the largest uncertainties, and combine models and data accordingly. When the DA criterion (Eq. (B.1c)) guides the definition of the largest uncertainties or "error subspace", the suboptimal truncation of errors in the full space is optimal. For minimum error variance estimation, the subspace is defined by a truncation of the ordered eigendecomposition of a normalized form of the multivariate $\mathbf{P}$ governed by Eq. (B.2). For a truncation of rank p, it is denoted here by $\mathbf{P}_k^p$ (Eq. (B.3)). The corresponding DA criterion is Eq. (B.4). ESSE intends to approximate Eqs. (B.2a) and (B.2b) using these Eqs. (B.3) and (B.4).

Er. Subspace:

$$\{\mathbf{P}_k^p = \mathbf{E}_k \boldsymbol{\Pi}_k \mathbf{E}_k^{\mathrm{T}} \text{ with rank } (\mathbf{E}_k) = p | \min_{\boldsymbol{\Pi}_k, \mathbf{E}_k} \|\mathbf{P}_k - \mathbf{P}_k^p\|\} \tag{B.3}$$

Min. Er. Var.:

$$\{\widehat{\mathbf{x}}_k | \min_{\widehat{\mathbf{x}}_k} J_k = \mathrm{tr}[\mathbf{P}_k^p(+)] \text{ using } [\mathbf{y}_0^o, \ldots, \mathbf{y}_k^o/\mathbf{y}_N^o]\}. \tag{B.4}$$

B.2. Computational schemes

Six of the main computational components of the present ESSE system are summarized next. Adaptive sampling components were described in Section 4 and details on our model of model errors in Appendix A.2.

(i) *Initialization*. The scheme utilized to initialize the multivariate error subspace (ES) is given in [43]. In summary, some data are often available to estimate parts of the dominant uncertainties in the initial mean or background state $\widehat{\mathbf{x}}_0$. These "observed portions" of $\mathbf{P}_0^p$ are directly specified from differences between $\widehat{\mathbf{x}}_0$ and these data. The remaining "non-observed" portions are then computed by dynamical inference, using an ensemble of model integrations (Eq. (B.1a)): the background state is perturbed with the observed portions, model

equations are integrated and the unknown portions arise by dynamical adjustments. The result is an estimate of the initial decomposition ($\mathbf{E}_0$, $\boldsymbol{\Pi}_0$).

(ii) *Ensemble and error subspace forecasts*. At t_k, $\widehat{\mathbf{x}}_k(+)$ is perturbed (Eq. (B.6)) using a combination of error modes $\mathbf{E}_k(+)$ with random coefficients $\boldsymbol{\pi}_k^j(+)$. These coefficients are weighted by $\boldsymbol{\Pi}_k(+)$ and constrained by dynamics [38]. The truncated tail of the error spectrum is modeled by random white noise $\mathbf{n}_k^j$. For the evolution to t_{k+1}, a central forecast (Eq. (B.5a)) and an ensemble of $j = 1,\ldots,q$ stochastic ocean model integrations is run (Eq. (B.7)), starting from the perturbed states $\mathbf{x}_k^j(+)$. The forcings $\mathrm{d}\eta(t)$are defined in Appendix A.2. The ES forecast (Eq. (B.8)) is computed from the ensemble. The matrix $\mathbf{M}_{k+1}(-) = [\widehat{\mathbf{x}}_{k+1}^j(-) - \widehat{\mathbf{x}}_{k+1}(-)]$ of differences between q realizations and an estimate of the conditional mean, e.g. $\widehat{\mathbf{x}}_{k+1}^{\mathrm{em}}(-)$ in Eq. (B.5b), is then computed. It is normalized and decomposed (Eq. (B.8)) into $\boldsymbol{\Pi}_{k+1}(-) \doteq \frac{1}{q}\boldsymbol{\Sigma}_{k+1}^2(-)$ and $\mathbf{E}_{k+1}(-)$ of rank $p \leq q$ by singular value decomposition (the operator $\mathrm{SVD}_p(\cdot)$ selects the rank-p SVD). The ensemble size is limited by a convergence criterion (Eq. (B.9)). The coefficient ρ used here measures the similarity between two subspaces of different sizes [29,30]. A "previous" estimate ($\mathbf{E}$, $\boldsymbol{\Pi}$) of rank p and "new" estimate ($\widetilde{\mathbf{E}}$, $\widetilde{\boldsymbol{\Pi}}$) of rank $\tilde{p} \geq p$ are compared, using singular values to weight singular vectors. The scalar limit α is chosen by the user ($1-\epsilon \leq \alpha \leq 1$). $\sigma_i(\cdot)$ selects the singular value number i and $k = \min(\tilde{p}, p)$. When ρ is close to one, ($\widetilde{\mathbf{E}}$, $\widetilde{\boldsymbol{\Pi}}$) is the error forecast for t_{k+1}: $\boldsymbol{\Pi}_{k+1}(-)$, $\mathbf{E}_{k+1}(-)$. The dimensions of the ensemble (q) and ES (p) hence vary with time, in accord with data and dynamics.

Central fcst:

$$\widehat{\mathbf{x}}_{k+1}^{\mathrm{cf}}(-) \mid \mathrm{d}\widehat{\mathbf{x}} = \mathcal{M}(\widehat{\mathbf{x}}, t)\mathrm{d}t, \quad \text{with } \widehat{\mathbf{x}}_k = \widehat{\mathbf{x}}_k(+). \tag{B.5a}$$

Ens. mean:

$$\widehat{\mathbf{x}}_{k+1}^{\mathrm{em}}(-) \doteq \mathcal{E}^q\left\{\widehat{\mathbf{x}}_{k+1}^j(-)\right\} \tag{B.5b}$$

ES In. Cond.:

$$\widehat{\mathbf{x}}_k^j(+) = \widehat{\mathbf{x}}_k(+) + \mathbf{E}_k(+)\boldsymbol{\pi}_k^j(+) + \mathbf{n}_k^j, \quad j = 1,\ldots,q. \tag{B.6}$$

Ens. Fcst:

$$\widehat{\mathbf{x}}_{k+1}^j(-) \mid \mathrm{d}\widehat{\mathbf{x}}^j = \mathcal{M}(\widehat{\mathbf{x}}^j, t)\mathrm{d}t + \mathrm{d}\eta, \quad \text{with } \widehat{\mathbf{x}}_k^j = \widehat{\mathbf{x}}_k^j(+). \tag{B.7}$$

ES Fcst:

$$\begin{aligned}\mathbf{M}_{k+1}(-) &= [\widehat{\mathbf{x}}_{k+1}^j(-) - \widehat{\mathbf{x}}_{k+1}(-)]\\ &\{\boldsymbol{\Sigma}_{k+1}(-), \mathbf{E}_{k+1}(-) | \mathrm{SVD}_p(\mathbf{M}_{k+1}(-))\\ &= \mathbf{E}_{k+1}(-)\boldsymbol{\Sigma}_{k+1}(-)\mathbf{V}_{k+1}^{\mathrm{T}}(-)\}.\end{aligned} \tag{B.8}$$

Conv. Crit.:

$$\rho = \frac{\sum_{i=1}^{k}\sigma_i(\boldsymbol{\Pi}^{\frac{1}{2}}\mathbf{E}^{\mathrm{T}}\widetilde{\mathbf{E}}\widetilde{\boldsymbol{\Pi}}^{\frac{1}{2}})}{\sum_{i=1}^{\tilde{p}}\sigma_i(\widetilde{\boldsymbol{\Pi}})} \geq \alpha. \tag{B.9}$$

(iii) *Measurement Models and Data*. In the present applications, the raw data from the gliders are processed into vertical profiles. Glider and ship temperature and salinity profiles are then interpolated onto model levels. The measurement operator ($\mathbf{H}$), linking data and model variables, is then straightforward. It consists of simple bilinear interpolations. Measurement error variances are assumed to vary in the vertical. They are set to be a fraction of the horizontally averaged variability, as estimated from historical data. Measurement error covariances are set to zero (assumed uncorrelated).

(iv) *ESSE data assimilation*. Once new data and their error estimates are available, forecasts are updated (Eqs. (B.10)–(B.13), $k+1$ omitted). Data–forecast misfits are computed and combined with the predicted state by multivariate minimum error variance estimation in the ES. Outputs are the filtering estimates: the *a posteriori* state $\mathbf{x}(+)$ and errors, e.g. $\mathbf{E}_+$, $\boldsymbol{\Pi}(+)$. The Eqs. (B.10)–(B.13) are a linear approximation of Eq. (B.2b). Filtering covariance estimates can also be obtained by update of the SVD of the ensemble spread (see [38]).

State Upd.:

$$\widehat{\mathbf{x}}(+) = \widehat{\mathbf{x}}(-) + \mathbf{K}^p\left(\mathbf{y}^{\mathrm{o}} - \mathcal{H}\left(\widehat{\mathbf{x}}(-)\right)\right). \tag{B.10}$$

ES Optimal Gain:

$$\mathbf{K}^p = \mathbf{E}_-\boldsymbol{\Pi}(-)\mathbf{H}^{p^{\mathrm{T}}}\left(\mathbf{H}^p\boldsymbol{\Pi}(-)\mathbf{H}^{p^{\mathrm{T}}} + \mathbf{R}\right)^{-1}, \quad \text{where } \mathbf{H}^p \doteq \mathbf{H}\mathbf{E}_-. \tag{B.11}$$

ES Cov. Upd.:

$$\mathbf{L}\boldsymbol{\Pi}(+)\mathbf{L}^{\mathrm{T}} = \boldsymbol{\Pi}(-) - \boldsymbol{\Pi}(-)\mathbf{H}^{p^{\mathrm{T}}}(\mathbf{H}^p\boldsymbol{\Pi}(-)\mathbf{H}^{p^{\mathrm{T}}} + \mathbf{R})^{-1}\mathbf{H}^p\boldsymbol{\Pi}(-) \tag{B.12}$$

$$\mathbf{E}_+ = \mathbf{E}_-\mathbf{L} \tag{B.13}$$

(v) *ESSE adaptive error estimation*. Dominant errors are learned from *a posteriori* data–model misfits (e.g. [40]). This adaptive learning from misfits Eqs. (B.14)–(B.17) can be necessary because covariance estimates are truncated and themselves uncertain. Eqs. (B.14)–(B.17) correspond to the present case where measurements are tracer data. The posterior misfits are first gridded. This is done using Eqs. (B.16) and (B.17), where $\mathbf{E}_{\mathrm{trc}}(-)$ and $\boldsymbol{\Pi}_{\mathrm{trc}}(-)$ are a decomposition of a tracer misfit covariance and $\mathbf{H}_{\mathrm{trc}} \doteq \mathbf{H}\mathbf{E}_{\mathrm{trc}}(-)$. The result $\widehat{\mathbf{n}}(+)$ are added by SVD to the ES, increasing its rank by one. This lead to $\mathbf{E}_+^{\mathrm{a}}$ and $\boldsymbol{\Sigma}^{\mathrm{a}}(+)$.

$$\widehat{\mathbf{n}}(+) = \mathbf{K}_{\mathrm{trc}}(\mathbf{y}^{\mathrm{o}} - \mathcal{H}(\widehat{\mathbf{x}}(+))) \tag{B.14}$$

$$\mathbf{K}_{\mathrm{trc}} = \mathbf{E}_{\mathrm{trc}}(-)\boldsymbol{\Pi}_{\mathrm{trc}}(-)\mathbf{H}_{\mathrm{trc}}^{\mathrm{T}}\left(\mathbf{H}_{\mathrm{trc}}\boldsymbol{\Pi}_{\mathrm{trc}}(-)\mathbf{H}_{\mathrm{trc}}^{\mathrm{T}} + \mathbf{R}\right)^{-1} \tag{B.15}$$

$$\mathbf{E}_+^{\mathrm{a}}\boldsymbol{\Sigma}^{\mathrm{a}}(+)\mathbf{V}_+^{\mathrm{a}\,\mathrm{T}} = \mathrm{SVD}_{p+1}([\mathbf{E}_+\boldsymbol{\Sigma}(+), \widehat{\mathbf{n}}(+)]) \tag{B.16}$$

$$\boldsymbol{\Pi}^{\mathrm{a}}(+) = \frac{1}{q+1}\boldsymbol{\Sigma}^{\mathrm{a}2}(+). \tag{B.17}$$

(vi) *ESSE smoothing*. To correct, based on future data, the past states and uncertainties, smoothing via ESSE [38] is carried out backward in time Eqs. (B.18)–(B.21). The scheme starts from the filtering estimate. A statistical approximation to the forward integration of the dynamical model between two data times t_{k-1} and t_k is derived. The approximation is a backward statistical linearization Eq. (B.19) based on the *a posteriori* ES at t_{k-1}

and nonlinear ES forecast at t_k. This backward rule is then used to minimize the smoothing DA criterion. The results are the smoothing estimate Eq. (B.18) and its errors (Eqs. (B.20) and (B.21)). Carrying out this process recursively up to t_0 leads to the smoothed initial state $\mathbf{x}_{0/N}$ and its errors $\mathbf{E}_{0/N}$, $\boldsymbol{\Pi}_{0/N}$.

Filtering Est.: ESSE filtering during $[t_0, t_N]$.

Sm. Est.:

$$\widehat{\mathbf{x}}_{k/N} = \widehat{\mathbf{x}}_k(+) + \mathbf{L}_k^p(\widehat{\mathbf{x}}_{k+1/N} - \widehat{\mathbf{x}}_{k+1}(-)),$$
$$\text{with } \widehat{\mathbf{x}}_{N/N} = \widehat{\mathbf{x}}_N(+). \tag{B.18}$$

Sm. Gain:

$$\mathbf{L}_k^p = \mathbf{E}_k(+)\boldsymbol{\Gamma}_k\mathbf{E}_{k+1}^{\mathrm{T}}(-) \quad \text{and}$$
$$\boldsymbol{\Gamma}_k = \boldsymbol{\Sigma}_k(+)\mathbf{V}_k^{\mathrm{T}}(+)\mathbf{V}_{k+1}(-)\boldsymbol{\Sigma}_{k+1}^{-1}(-). \tag{B.19}$$

Sm. Err. Cov.:

$$\mathbf{H}_k\boldsymbol{\Pi}_{k/N}\mathbf{H}_k^{\mathrm{T}} = \boldsymbol{\Pi}_k(+) + \boldsymbol{\Gamma}_k(\boldsymbol{\theta}_{k+1}\boldsymbol{\Pi}_{k+1/N}\boldsymbol{\theta}_{k+1}^{\mathrm{T}} - \boldsymbol{\Pi}_{k+1}(-))\boldsymbol{\Gamma}_k^{\mathrm{T}}. \tag{B.20}$$

$$\mathbf{E}_{k/N} = \mathbf{E}_k(+)\mathbf{H}_k, \quad \text{with } \boldsymbol{\theta}_{k+1} \doteq \mathbf{E}_{k+1}^{\mathrm{T}}(-)\mathbf{E}_{k+1/N}. \tag{B.21}$$

References

[1] J.L. Anderson, S.L. Anderson, A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts, Mon. Weather Rev. 127 (1999) 2741–2758.

[2] J.L. Anderson, An ensemble adjustment Kalman filter for data assimilation, Mon. Weather Rev. 129 (2001) 2884–2903.

[3] HOPS-ESSE contributions to the Autonomous Ocean Sampling Network-II (AOSN-II) field exercise. http://www.deas.harvard.edu/leslie/AOSNII/index.html, August 2003.

[4] N.H. Barth, C. Wunsch, Oceanographic experiment design by simulated annealing, J. Phys. Oceanogr. 20 (1990) 1249–1263.

[5] J.G. Bellingham, J.S. Willcox, Optimizing AUV oceanographic surveys, in: IEEE Conference on Autonomous Underwater Vehicles Technology, Monterey, CA, 1996.

[6] A.F. Bennett, Inverse methods in physical oceanography, in: Cambridge Monographs on Mechanics and Applied Mathematics, Cambridge University Press, 1992.

[7] K.J. Beven, J. Freer, Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the glue methodology, J. Hydrology 249 (2001) 11–29.

[8] C.H. Bishop, Z. Toth, Ensemble transformation and adaptive observations, J. Atmos. Sci. 56 (1999) 1748–1765.

[9] C.H. Bishop, B.J. Etherton, S.J. Majumdar, Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects, Mon. Weather Rev. 129 (2001) 420–436.

[10] I. Blanchet, C. Frankignoul, M.A. Cane, A comparison of adaptive Kalman filters for a tropical pacific ocean model, Mon. Weather Rev. 125 (1997) 40–58.

[11] R. Buizza, A. Montani, Targeted observations using singular vectors, J. Atmospheric Sci. 56 (1999) 2965–2985.

[12] K.H. Brink, R.C. Beardsley, P.P. Niiler, M. Abbott, A. Huyer, S. Ramp, T. Stanton, D. Stuart, Statistical properties of near-surface flow in the California coastal transition zone, J. Geophys. Res. 96 (1991) 14,693–14,706.

[13] M.D. Cox, A primitive equation, 3-dimensional model of the ocean, Technical Report, Geophysical Fluid Dynamical Laboratory/NOAA, Princeton, 1984.

[14] Dash Optimization. http://www.dashoptimization.com/index.html.

[15] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, 2nd ed., Wiley-Interscience, New York, 2000.

[16] G.J. Erickson, C.R. Smith, Maximum-Entropy and Bayesian Methods in Science and Engineering, Kluwer Academic Publishers, Dordrecht, Boston, 1988.

[17] G. Evensen, D. Dee, J. Schroter, Parameter estimation in dynamical models, in: E.P. Chassignet, J. Verron (Eds.), Ocean Modeling and Parameterizations, NATO ASI, 1998.

[18] A. Gangopadhyay, A.R. Robinson, Feature oriented regional modeling of oceanic fronts, Dyn. Atmos. Ocean. 36 (2002) 201–232.

[19] T.C. Gard, Introduction to Stochastic Differential Equations, Marcel Dekker, New York, 1988, 234 pages.

[20] C.W. Gardiner, Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences, Springer-Verlag, 1983, 442 pages.

[21] G. Gaspari, S.E. Cohn, Construction of correlation functions in two and three dimensions, Quart. J. Roy. Meteor. Soc. 125 (1999) 723–757.

[22] A. Gelb (Ed.), Applied Optimal Estimation, MIT Press, Cambridge, MA, 1974.

[23] S. Griffies, Fundamentals of Ocean Climate Models, Princeton University Press, 2004, 496 pages.

[24] T.M Hamill, J.S. Whitaker, C. Snyder, Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter, Mon. Weather Rev. 129 (2001) 2776–2790.

[25] T.M. Hamill, Ensemble-based atmospheric data assimilation: A tutorial. University of Colorado and NOAA-CIRES Climate Diagnostics Center, Boulder, Colorado, 2003, 46 pages.

[26] B.M. Hickey, Coastal Oceanography of Western North America from the Tip of Baja California to Vancouver, in: K.H. Brink, A.R. Robinson (Eds.), The Sea, vol. 11, Wiley and Sons, Inc., 1998, pp. 345–393 (Chapter 12).

[27] P.J. Haley, P.F.J. Lermusiaux, W.G. Leslie, A.R. Robinson, Harvard Ocean Prediction System (HOPS). http://oceans.deas.harvard.edu/HOPS/HOPS.html.

[28] E.E. Hofmann, M.A.M. Friedrichs, Predictive modeling for marine ecosystems, in: A.R. Robinson, J.J. McCarthy, B.J. Rothschild (Eds.), in: THE SEA: Biological–Physical Interactions in the Sea, vol. 12, John Wiley and Sons, NY, 2002, pp. 537–565.

[29] R.A. Horn, C.R. Johnson, Matrix Analysis, Cambridge University Press, 1985, 561 pages.

[30] R.A. Horn, C.R. Johnson, Topics in Matrix Analysis, Cambridge University Press, 1991, 607 pages.

[31] P.L. Houtekamer, H.L. Mitchell, Data assimilation using an ensemble Kalman filter technique, Mon. Weather Rev. 126 (1998) 796–811.

[32] P.L. Houtekamer, H.L. Mitchell, A sequential ensemble Kalman filter for atmospheric data assimilation, Mon. Weather Rev. 129 (2001) 123–137.

[33] K. Ide, P. Courtier, M. Ghil, A.C. Lorenc, Unified notation for data assimilation: Operational, sequential and variational, J. Meteor. Soc. Japan 75 (1B) (1997) 181–189.

[34] A.H. Jazwinski, Stochastic Processes and Filtering Theory, Academic Press, 1970.

[35] E. Kalnay, Atmospheric Modeling, Data Assimilation and Predictability, Cambridge University Press, 2003, 341 pages.

[36] A. Kohl, D. Stammer, Optimal observations for variational data assimilation, J. Phys. Ocean 34 (2004) 529–542.

[37] C.M. Lalli, T.R. Parsons, Biological Oceanography: An Introduction, Butterworth-Heinemann, Oxford, 1997, 314 pages.

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

[39] P.F.J. Lermusiaux, Data assimilation via error subspace statistical estimation, Part II: Middle Atlantic Bight shelfbreak front simulations and ESSE validation, Mon. Weather Rev. 127 (7) (1999) 1408–1432.

[40] P.F.J. Lermusiaux, Estimation and study of mesoscale variability in the Strait of Sicily, Dyn. Atmos. Ocean. 29 (1999) 255–303.

[41] P.F.J. Lermusiaux, D.G. Anderson, C.J. Lozano, On the mapping of multivariate geophysical fields: Error and variability subspace estimates, Quart. J. Roy. Meteor. Soc. (April B) (2000) 1387–1430.

[42] P.F.J. Lermusiaux, Evolving the subspace of the three-dimensional multiscale ocean variability: Massachusetts Bay ("Three-dimensional ocean circulation: Lagrangian measurements and diagnostic analyses"), J. Marine Syst. 29/1–4 (2001) 385–422 (special issue).

[43] P.F.J. Lermusiaux, On the mapping of multivariate geophysical fields: Sensitivity to size, scales and dynamics, J. Atmos. Ocean. Tech. 19 (2002) 1602–1637.
[44] P.F.J. Lermusiaux, A.R. Robinson, P.J. Haley, W.G. Leslie, Advanced interdisciplinary data assimilation: Filtering and smoothing via error subspace statistical estimation, in: Proceedings of "The OCEANS 2002 MTS/IEEE" conference, Holland Publications, 2002, pp. 795–802.
[45] P.F.J. Lermusiaux, C. Evangelinos, R. Tian, P.J. Haley, J.J. McCarthy, N.M. Patrikalakis, A.R. Robinson, H. Schmidt, Adaptive coupled physical and biogeochemical ocean predictions: A conceptual basis, in: F. Darema (Ed.), Computational Science — ICCS 2004, in: Lecture Notes in Computer Science, vol. 3038, 2004, pp. 685–692.
[46] S.J. Majumdar, C.H. Bishop, B.J. Etherton, Adaptive sampling with the ensemble transform Kalman filter. Part II: Field program implementation, Mon. Weather Rev. 130 (2002) 1356–1369.
[47] S.J. Majumdar, C.H. Bishop, B.J. Etherton, I. Szunyogh, Z. Toth, Can an ensemble transform Kalman filter predict the reduction in forecast error variance produced by targeted observations Quart, J. Roy. Meteor. Soc. 127 (2001) 2803–2820.
[48] P. Malanotte-Rizzoli (Ed.), Modern Approaches to Data Assimilation in Ocean Modeling, Amsterdam, Elsevier, 1996.
[49] R.J. Matear, Parameter optimization and analysis of ecosystem models using simulated annealing: A case study at Station P, J. Marine Res. 53 (1995) 571–607.
[50] J.J. McCarthy, A.R. Robinson, B.J. Rothschild, in: A.R. Robinson, J.J. McCarthy, B.J. Rothschild (Eds.), Biological–Physical Interactions in the Sea: Emergent Findings and New Directions, in: The Sea, vol. 12, John Wiley & Sons, Inc., New York, 2002 (Chapter 1).
[51] D. Menemenlis, M. Chechelnitsky, Error Estimates for an Ocean General Circulation from Altimeter and Acoustic Tomography Data, Mon. Weather Rev. 128 (2000) 763–778.
[52] D. Menemenlis, I. Fukumori, T. Lee, Using Green's functions to calibrate an ocean general circulation model, Mon. Weather Rev. 133 (2005) 1224–1240.
[53] R.N. Miller, L.L. Ehret, Ensemble generation for models of multimodal systems, Mon. Weather Rev. 130 (9) (2002) 2313–2333.
[54] C.N.K. Mooers (Ed.), Coastal ocean prediction, AGU coastal and estuarine studies series, American Geophysical Union, Washington, DC, 1999, 523 pages.
[55] MOOS Upper-water-column Science Experiment (MUSE). http://www.mbari.org/muse/intro.htm.
[56] G.L. Nemhauser, L.A. Wolsey, Integer and Combinatorial Optimization, John Wiley & Sons, New York, 1988.
[57] I. Orlanski, A simple boundary condition for unbounded hyperbolic flows, J. Comput. Phys. 41 (1976) 251–269.
[58] E. Ott, B.R. Hunt, I. Szunyogh, A.V. Zimin, E.J. Kostelich, M. Corazza, E. Kalnay, D.J. Patil, J.A. Yorke, A local ensemble Kalman filter for atmospheric data assimilation, Tellus A 56 (5) (2004) 415.
[59] T.N. Palmer, R. Gelaro, J. Barkmeijer, R. Buizza, Singular vectors, metrics and adaptive observations, J. Atmospheric Sci. 55 (1998) 633–653.
[60] N. Pinardi, J. Woods (Eds.), Ocean Forecasting: Conceptual Basis and Applications, Springer-Verlag, Berlin, 2002, 472 pages.
[61] A.R. Robinson, P.F.J. Lermusiaux, N.Q. Sloan, in: K.H. Brink, A.R. Robinson (Eds.), THE SEA: The Global Coastal Ocean, in: Processes and Methods, vol. 10, John Wiley and Sons, NY, 1998, pp. 541–594.
[62] A.R. Robinson, S.M. Glenn, Adaptive sampling for ocean forecasting, Naval Res. Rev. 51 (2) (1999) 28–38.
[63] A.R. Robinson, P.F.J. Lermusiaux, Data assimilation in models, in: Encyclopedia of Ocean Sciences, Academic Press Ltd., London, 2001, pp. 623–634.
[64] L.K. Rosenfeld, F.B. Schwing, N. Garfield, D.E. Tracy, Bifurcated flow from an upwelling center — A cold-water source for Monterey Bay, Cont. Shelf Res. 14 (9) (1994) 931–964.
[65] A. Schrijver, Theory of Linear and Integer Programming, John Wiley and Sons, 1998.
[66] O.M. Smedstad, J.J. O'Brien, Variational data assimilation and parameter estimation in an equatorial Pacific Ocean model, Prog. Oceanogr. 26 (1991) 179–241. Pergamon.
[67] R. Shapiro, Smoothing, filtering, and boundary effects, Rev. Geophys. Space Phys. 8 (2) (1970) 359–387.
[68] D. Stammer, Adjusting internal model errors through ocean state estimation, J. Phys. Oceanogr. 35 (2005) 1143–1153.
[69] I. Szunyogh, Z. Toth, K.A. Emanuel, C.H. Bishop, C. Snyder, R.E. Morss, J.S. Woolen, T.P. Marchok, Ensemble-based targeting experiments during FASTEX: The effect of dropsonde data from the Lear jet, Quart. J. Roy. Meteor. Soc. 125 (1999) 3189–3217.
[70] R.C. Tian, P.F.J. Lermusiaux, J.J. McCarthy, A.R. Robinson, A generalized prognostic model of marine biogeochemical-ecosystem dynamics: Structure, parameterization and adaptive modeling, Harvard Reports in Physical/Interdisciplinary Ocean Science, No. 67, 2004, pp. 1– 65.
[71] W.G. Thacker, R.B. Long, Fitting dynamics to data, J. Geophys. Res. 93 (1988) 1227–1240.
[72] S.K. Thompson, G.A.F. Seber, Adaptive Sampling, in: Wiley Series in Probability and Statistics, 1996, 288 pages.
[73] S.K. Thompson, Sampling, 2nd ed., J. Wiley & Sons, New York, 2002.
[74] N.K. Yilmaz, Path planning of autonomous underwater vehicles for adaptive sampling, Ph.D. Thesis, Department of Mechanical Engineering MIT, September 2005.
[75] J.S. Whitaker, T.M. Hamill, Ensemble data assimilation without perturbed observations, Mon. Weather Rev. 130 (2002) 1913–1924.
[76] C. Wunsch, The Ocean Circulation Inverse Problem, Cambridge Univ. Press, 1996.
[77] K.Q. Zhang, J. Marotzke, The importance of open-boundary estimation for an Indian Ocean GCM-data synthesis, J. Marine Res. 57 (2) (1999) 305–334.