Lermusiaux, P.F.J., A.R. Robinson, P.J. Haley and W.G. Leslie, 2002. Advanced interdisciplinary data assimilation: Filtering and smoothing via Error Subspace Statistical Estimation (.pdf). Proceedings of "The OCEANS 2002 MTS/IEEE" conference, IEEE, Holland Publications, 795-802.

Advanced Interdisciplinary Data Assimilation: Filtering and Smoothing via Error Subspace Statistical Estimation

P.F.J. Lermusiaux, A.R. Robinson,
P.J.H. Haley and W.G. Leslie

Harvard University
Division of Engineering and Applied Sciences
Pierce Hall G2A
29 Oxford Street
Cambridge MA 02318
USA
web: www.deas.harvard.edu/~pierrel
e-mail: pierrel@pacific.deas.harvard.edu

Abstract. The efficient interdisciplinary 4D data assimilation with nonlinear models via Error Subspace Statistical Estimation (ESSE) is reviewed and exemplified. ESSE is based on evolving an error subspace, of variable size, that spans and tracks the scales and processes where the dominant errors occur. A specific focus here is the use of ESSE in interdisciplinary smoothing which allows the correction of past estimates based on future data, dynamics and model errors. ESSE is useful for a wide range of purposes which are illustrated by three investigations: (i) smoothing estimation of physical ocean fields in the Eastern Mediterranean, (ii) coupled physical-acoustical data assimilation in the Middle Atlantic Bight shelfbreak, and (iii) coupled physical-biological smoothing and dynamics in Massachusetts Bay.

I. INTRODUCTION

All dynamical models are to some extent approximate, and all data sets are finite and to some extent limited by error bounds. The purpose of data assimilation is to provide quantitative estimates of nature which are better estimates than can be obtained by using only observations or a dynamical model. State variables, parameters and their respective uncertainties can be estimated, processes inferred, dynamical hypothesis tested, necessary data identified, and fundamental models developed.

The data assimilation process is optimal because dynamical models and data sets are combined based on the quantitative minimization of an assimilation criterion or cost function. The links between observational data and dynamical model fields and parameters are provided by measurement models. Since dynamical models, data sets and measurement models are all approximate, they all involve an uncertainty component, i.e. the error models. In general, it is the probability density functions (pdf's) of the errors which are modeled. A specific realization such a pdf is here represented by a random noise, i.e. stochastic forcings.

The assimilation criterion usually weights each information according to its uncertainty. Accurate uncertainty estimates are thus essential. Since the dimension of error models is in principal infinite, an important challenge in data assimilation is the efficient attribution, representation and estimation of errors. This challenge involves scientific as well as computational issues. Two examples are: (i) what are the dominant uncertain processes in mesoscale or climate predictions?, and (ii) how can errors be accurately forecast at feasible costs? Error Subspace Statistical Estimation (ESSE; [7,6]) is a data assimilation methodology which can address such issues. Instead of characterizing and capturing all uncertainties, ESSE focuses on the uncertainties that matter. It is based on evolving an error subspace, of variable size, that spans and tracks the scales and processes where the dominant errors occur.

After briefly reviewing and illustrating a few data assimilation schemes and their applications to physical and interdisciplinary research, the oral presentation discussed a few current data assimilation research issues. We refer here to [18, 19, 20] for such reviews and discussions. Presently, the ESSE schemes are first outlined in Sect. 2. Their results are then illustrated and evaluated for three physical, acoustical and biological applications of scientific and operational relevance (Sect. 3). A focus here is the use of ESSE in interdisciplinary smoothing, specifically, the correction of past estimates based on the future interdisciplinary data, dynamics and model errors. Conclusions are provided in Sect. 4.

Lermusiaux, P.F.J., A.R. Robinson, P.J. Haley and W.G. Leslie, 2002. Advanced interdisciplinary data assimilation: Filtering and smoothing via Error Subspace Statistical Estimation (.pdf). Proceedings of "The OCEANS 2002 MTS/IEEE" conference, IEEE, Holland Publications, 795-802.

II. FILTERING AND SMOOTHING VIA ESSE

The spatially discretized interdisciplinary ocean state is represented by a coupled state vector, $\mathbf{x}$, which is evolved from $\mathbf{x}(t_0) = \hat{\mathbf{x}}_0$ based on, $d\mathbf{x} = M(\mathbf{x})dt + d\boldsymbol{\eta}$, where M is the coupled model operator and $d\boldsymbol{\eta}$ are the stochastic uncertainties. At time t_k, measurement models are of the form, $\mathbf{y}_k = H(\mathbf{x}_k) + \boldsymbol{\varepsilon}_k$, where $\mathbf{y}_k$ is the observed data, H the measurement model operator and $\boldsymbol{\varepsilon}_k$ the stochastic uncertainties. The goal of the present four-dimensional data assimilation is to minimize the trace of the *a posteriori* error covariance, $\mathbf{P}_k^p(+)$, of the coupled state, i.e. find $\mathbf{x}_k$ such that $J_k = tr[\mathbf{P}_k^p(+)]$ is minimized using $[\mathbf{y}_0,\ldots,\mathbf{y}_k/\mathbf{y}_N]$. If data are employed up to time t_N to estimate the state at time t_k, the assimilation is a smoothing problem. If only data up to t_k are employed, it is a filtering problem.

The ESSE methodology provides data assimilation schemes which solve an optimal reduction of the above problem. The goal of ESSE is to determine the ocean evolution by minimizing the dominant errors, in agreement with the full dynamical model and measurement model (data) constraints, and their respective uncertainties. The "dominant errors" are here defined by the dominant eigendecomposition of a normalized form of the error covariance matrix. The dominant error eigenvectors and eigenvalues determine the "error subspace."

A. Error Subspace Statistical Estimation Schemes

Presently (Fig. 1), the error subspace is initialized by decomposition on multiple scales [9, 13]. The dominant uncertainties at t_0 are assumed to be the variability that is uncertain in the initial state. For parts of this uncertain variability, some data are usually available. These "observed portions" of the dominant error covariances are directly specified from differences between the initial state and those data, and/or from a statistical model fit to these differences.

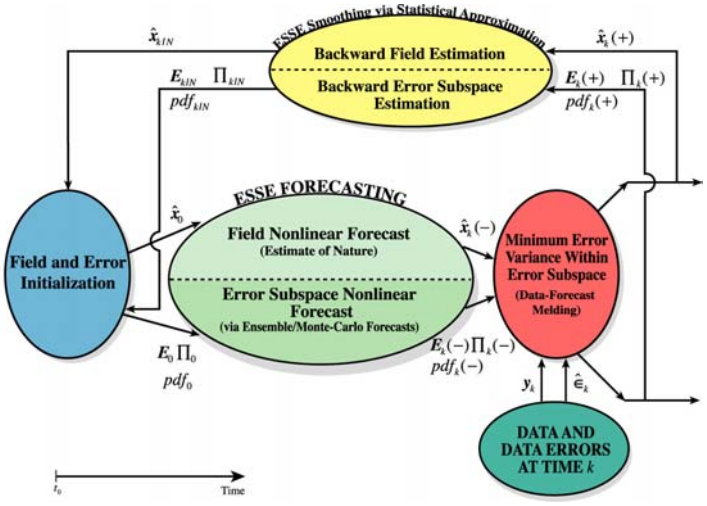

Fig. 1. Five main components of the present ESSE system.

The "non-observed" portions are constructed by an ensemble of stochastic dynamical model integrations: by dynamical cross-covariances, these unknown portions of the covariance are created and adjusted to the specified observed ones. The resulting estimate of the initial error eigendecomposition ($\mathbf{E}_0$, $\boldsymbol{\Pi}_0$) or error pdf (Fig. 1: blue, left oval) is used to perturb the initial state $\mathbf{x}_0$.

To evolve the ocean fields and uncertainties (Fig. 1: light green, central oval) up to time t_k, an ensemble of stochastic ocean model integrations are carried out in parallel. The error covariance forecast is computed from this ensemble. The ensemble size is increased and ultimately controlled by convergence criteria; when satisfied, the ensemble of states leads to the forecast of nature $\mathbf{x}_k(-)$ and to its error estimate, e.g. the error eigenvectors $\mathbf{E}_k(-)$ and eigenvalues $\boldsymbol{\Pi}_k(-)$ obtained by normalized SVD. It is important to note that in this process the biochemical, physical and acoustical ensembles are concatenated to provide the coupled predicted fields and uncertainties.

At this stage, the data and their error estimates (Fig. 1: dark green, bottom oval) are employed. Data-forecast misfits are computed and used to correct the predicted fields by minimum error variance estimation in the interdisciplinary error subspace (Fig. 1: red, right oval). During this melding, the data influence is across disciplines, e.g. biochemical data correct the physics. The outputs are the filtering estimates: the *a posteriori* coupled fields $\mathbf{x}_k(+)$ and *a posteriori* coupled errors, e.g. $\mathbf{E}_k(+)$, $\Pi_k(+)$. *A posteriori* data misfits are then calculated and used for adaptive learning of the dominant errors, e.g. [8]. This learning of errors from misfits can be necessary because error estimates are themselves uncertain.

Ultimately, the smoothing via ESSE [7] is carried out (Fig. 1: yellow, top oval) to correct, based on future data, the past coupled fields and uncertainties. Starting 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 first derived. The approximation is importantly written in a backward form. Presently, it is a backward statistical linearization based on the *a posteriori* error subspace at times t_{k-1} and nonlinear forecast error subspace at times t_k. It is then utilized to minimize the smoothing cost function and so compute the smoothing estimate and its errors. Carrying out the smoothing process recursively up to t_0 leads to the smoothed initial fields and errors, e.g. $\mathbf{E}_{0/N}$, $\Pi_{0/N}$.

III. INTERDISCIPLINARY INVESTIGATIONS

A. Physical Smoothing in the Levantine Sea

The first smoothing application illustrated is in the Levantine Sea. The focus is on the mesoscale to subbasin-scale physical fields and errors. To introduce this region and its contemporary upper-thermocline features, a multivariate 3D objective analysis (computed by ESSE, [13]) of the potential density at 105 m and velocity at 5 m is shown on Fig. 2. Because of thermal-wind effects, surface circulation

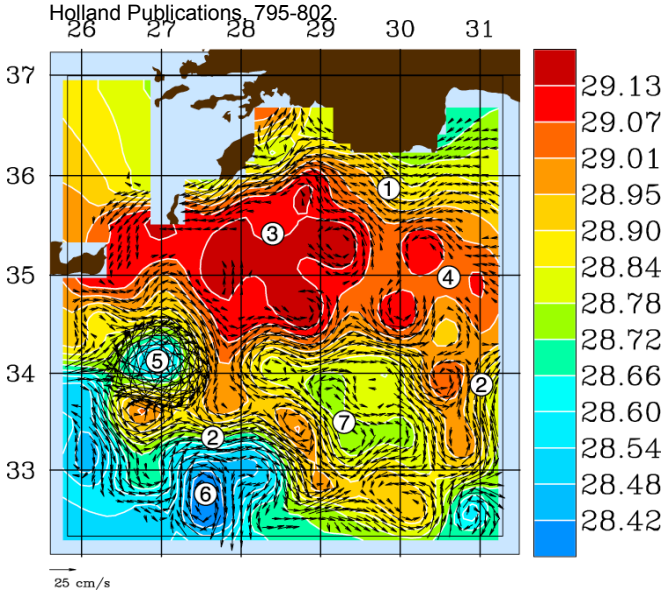

Fig. 2. March 27 analysis of potential density σ_θ at 105 m, overlaid with horizontal velocity vectors at 5 m (vectors are plotted only if the analyzed $\|\mathbf{u}\|$ is larger than 6 cm/s). The numbers identify the location of the main upper-thermocline features: Asia Minor Current (1), Mid-Mediterranean Jet (2), Rhodes Gyre (3), West Cyprus Gyre (4), Ierapetra Eddy (5), a lobe of the Mersa Matruh Gyre (6) and main anticyclone in the Mersa Matruh-Shikmona Gyre complex (7).

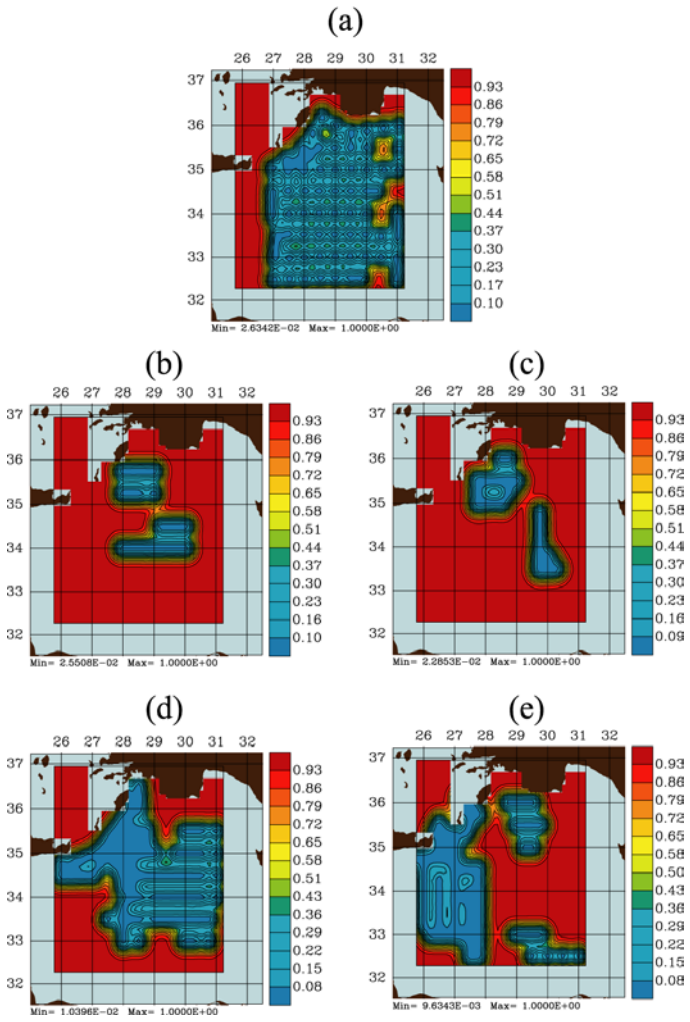

Fig. 3. Surface expected normalized mesoscale error variances (0–1) of the hydrographic data used in the smoothing, as computed by 2D objective analysis. Plots are ordered chronologically: (a) AXBT, March 15; (b) Week1 CTD, Mar 19–Mar 26; (c) Week2 CTD, Mar 27–Apr 2 (d) Week3 CTD, Apr 3–Apr 9; (e) Week4 CTD, Apr 10–Apr 16. In the 2D objective analysis, mesoscale zero-crossings are 150 km and decay-scales 30 km.

structures are easily distinguished along the steepest slopes of isopycnals. Several water masses of the Levantine Sea are also visible. The data, dynamical model and smoothing results are outlined next.

Ocean Physics Data. The measurements were collected during the Levantine Intermediate Water-95 campaign which involved five intensive cruises carried out by research vessels of Germany, Turkey, Italy, Greece and Israel, from January to April 1995 in the Eastern Mediterranean [16]. For the present smoothing application, the focus is on the spring period, from March 15 up to April 16, 1996. To rapidly re-sample the physical fields just before the spring cruises, temperature vs. depths probes were deployed by aircraft (AXBTs) of the US Navy on March 15, 1995. Surface error variance fields computed by a quick 2D objective analysis of the available data are shown on Fig. 3. From these data coverage plots, it is obvious that it is only by the end of the campaign that the model domain has been covered by temperature and salinity profiles. Determining initial conditions for the model simulations is thus an issue. In addition, velocities are not measured. The estimation of the nonlinear physical evolution is thus an interesting smoothing problem. It is a prime candidate for ESSE.

Ocean Physics Model. The physical state variables are temperature T, salinity S, velocity $\mathbf{u}$ and pressure p_w. For this study, their deterministic mesoscale evolution is computed by the primitive-equation model of the numerical Harvard Ocean Prediction System, e.g. [17]. Boundary condition formulations are employed at the surface and at the open-ocean boundaries. In the present examples (Sects. III.A–C), these conditions and the model parameters are calibrated using data and sensitivity studies. In the uncertainty forecasts, stochastic forcings are added to the primitive-equation model.

ESSE Filtering and Smoothing. To obtain the ESSE smoothed estimates, ESSE filtering is first carried out until the end of the campaign, assimilating the hydrographic profiles a time advances and here learning errors adaptively(see [5,8]). To carry this out, uncertainty forecasts are computed (Sect. II). Initial conditions are perturbed in accord with the initial error covariance and a parallel ensemble of perturbed forecasts is evolved. To account for model errors, each forecast is forced at each time-step by stochastic noise. This noise represents here the eventual model deficiencies in the surface external forcings and the sub-mesoscale to mesoscale processes. It consists of a stationary first-order Gauss-Markov process in time, with an autocorrelation time of half a day. In space, it has a correlation length of about one grid point (10

km). The amplitude of the noise is depth-dependent and was bracketed to be 10% of the dominant horizontally-averaged primitive-equation balance. We refer to [5] for more details.

The filtering and smoothing fields are illustrated by Fig. 4. Starting from the filtering estimate on April 16, the ESSE smoothing progresses backward in time, based on the evolution of the error subspace during the filtering step (the error covariances of the filtering need to be saved). Presently, the error subspace size evolved within 318 and 329. Comparing the smoothed (Fig. 4b) and filtered (Fig. 4a) temperature fields on April 6, the smoothing process has corrected the positions, shapes and strengths of the main features of the Levantine Sea (Fig. 2), in accord with the future data, dynamics and model errors. The positive effects of the continued stochastic error forcings are also visible on Fig. 4b. Since they increase the forecast error variance, they allow for better data corrections and, importantly, also lead to higher sub-meso- to mesoscale variability in the smoothed field, in accord with the data.

For quantitative evaluation, root-mean-square differences between ESSE field estimates and data were computed at data points. As shown on Fig. 5 (dots), the smoothing improves the filtering, especially the initial conditions. On two days, the filtering estimate is closer to data. This is statistically possible, but in this case, the two filtering estimates are in fact too close to data. This is in part supported by comparing the quality of forecasts starting from filtering estimates with these starting from smoothing estimates. As shown on Fig. 5 (squares), forecasts from smoothing estimates are better. The forecast error growth is much larger from filtering estimates than it is from smoothing estimates. In fact, it is mainly the non-linearities and the error subspace truncation which lead to differences between the smoothing estimate at t_{k+1} and the forecast to t_{k+1} starting from the smoothing estimate at t_k.

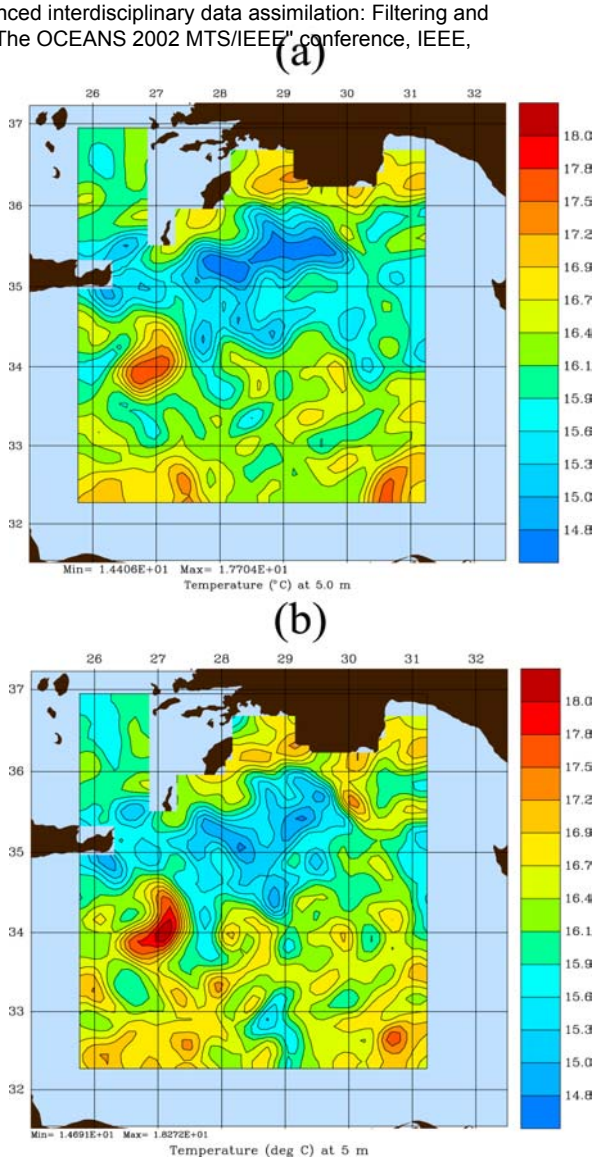

Fig. 4. (a) ESSE filtering estimate of the temperature at 5 m on April 6, 1995. (b) As (a), but for the ESSE smoothing estimate.

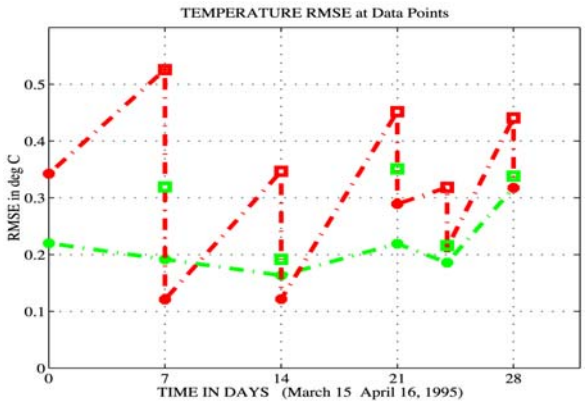

Fig. 5. Root-mean-square differences between temperature data at 5 m and ESSE field estimates at these data-points, on 6 different days. The "RMS-error" shown are these of the: filtering estimates (red dots), smoothing estimates (green dots), forecasts from the previous filtering estimate (red squares) and forecasts from the previous smoothing estimate (green squares). The dashed-dotted lines illustrate the forward filtering (red zig-zag) and backward smoothing (green continuous line) evolutions.

B. Physical-Acoustical Filtering in a Shelfbreak Environment

Coupled physical-acoustical data assimilation is now illustrated. The physics is the mesoscale dynamics of the Middle Atlantic Bight shelfbreak front, including remote influences from the shelf, slope and deep ocean. The acoustics is the transmission of low-frequency sound from the continental slope, through the shelfbreak front, onto the shelf. The period is the summer of 1996. In the present case, the coupled assimilation is an "identical twin experiment." Data are extracted from a physical-acoustical simulation that is independent from the ensemble of realizations carried out during ESSE. This independent realization is called the "true" ocean. Goals in such an experiment are to study the assimilation in a simulated situation and to find out if the *a posteriori* fields become close to the known "true" fields.

Physical and Acoustical Data. The physical data are profiles of temperature and salinity collected during the 1996 Shelfbreak-PRIMER [15] experiment. Several of these data are here assimilated in the physical simulation to create a relatively realistic "synthetic" true ocean. After five days of

Lermusiaux, P.F.J., A.R. Robinson, P.J. Haley and W.G. Leslie, 2002. Advanced interdisciplinary data assimilation: Filtering and smoothing via Error Subspace Statistical Estimation (.pdf). Proceedings of "The OCEANS 2002 MTS/IEEE" conference, IEEE, Holland Publications, 795-802.

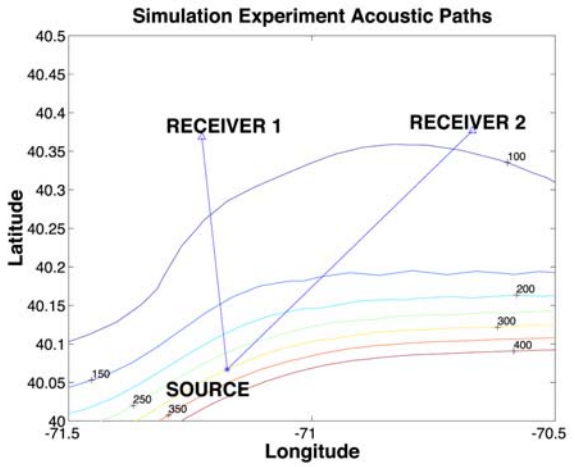

Fig. 6. Acoustic paths considered, overlaid on bathymetry.

physical evolution, the acoustic model is run using the sound-speed field of the physical model. The coupled assimilation of acoustic data is then carried out via ESSE, correcting the 3D physical fields and 2D transmission loss (TL) along an actual PRIMER acoustic path (Fig. 6). The 224 Hz source is at 300 m depth. The acoustical data assimilated here are simulated towed-receiver TL data along path 1, i.e. TL1. The TL observations are made at constant 70 m depth, every 50 m from a range of 150 m to almost receiver 1. These are very sub-sampled data since the (r, z) acoustic grid resolution is 5 m by 5 m (see [14] for more detail).

Coupled Ocean Physics and Acoustic Models. The ocean physics model is the primitive-equation model used in Sect. III.A. Here, it is forced by atmospheric fluxes imposed at the ocean surface. The physical-acoustical couplings occur through the sound-speed field which is the main water-column parameter in the acoustics. The acoustic model is a coupled normal mode model [2, 3] which solves a linearized wave equation governing sound pressure p_s. This pressure is decomposed in the frequency domain into slowly-varying complex envelopes that modulate (mode by mode) analytic, rapidly-varying, adiabatic-mode solutions. Given the sound speed, density, attenuation rate and bathymetry vertical cross-sections, the acoustic state is obtained by integrating differential equations governing the complex modal envelopes. The model output contains sound pressure, transmission loss, and travel time, phase and amplitude of the individual modes.

ESSE Filtering. The results of the coupled assimilation are illustrated on Fig. 7. The simulated true TL, *a priori* TL (i.e. the mean TL), *a posteriori* TL and the TL realization closest to the *a posteriori* TL are shown on Fig. 7a. Even though the true TL is challenging to retrieve (TL of high-order modal interactions) and the sub-sampled data are limited, the *a posteriori* TL is substantially closer to the true TL than the mean TL. From the ESSE ensemble of predicted TLs, one can select for best estimate the TL the closest (in some metric sense, here the RMS measure over the *r, z* grid) to the *a posteriori* TL. This realization is at some locations closer to the true TL than the *a posteriori* TL.

The differences between the *a priori* and true TLs, and between the *a posteriori* and true TLs, are shown on Fig. 7b. The *a posteriori* residuals are much smaller than the *a priori* ones at most locations, except above the thermocline near the surface on the shelf. This is due to the refractive effects of the thermocline (data are below at 70 m) and to the limited error subspace size (here 79). With ESSE, error covariances are also estimated: the diagonals of the *a priori* and *a posteriori* error covariances are plotted on Fig. 7b. Overall, these standard deviations agree with the averages of the residuals (note that their accuracy increases with the subspace size). In particular, the expected error along the simulated towed-receiver at 70 m has been reduced.

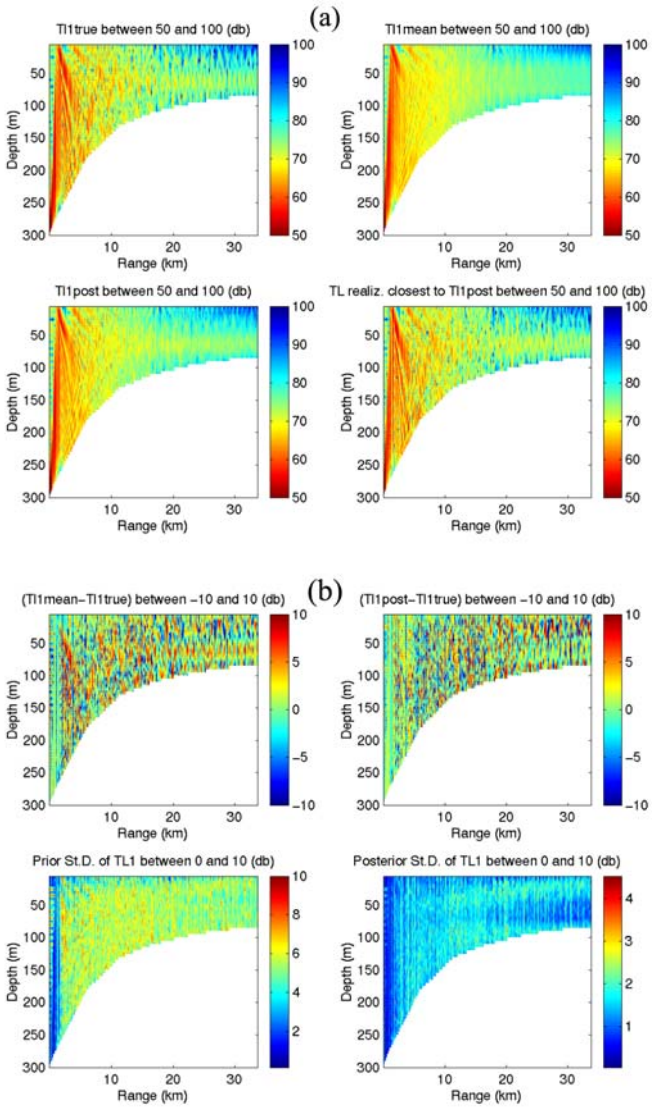

Fig.7. (a) "True" TL, *a priori* TL, *a posteriori* TL and TL realization closest to *a posteriori* TL, all along path 1. (b) *A posteriori* residuals and *a posteriori* error St.Dv., all for TL along path 1.

Lermusiaux, P.F.J., A.R. Robinson, P.J. Haley and W.G. Leslie, 2002. Advanced interdisciplinary data assimilation: Filtering and smoothing via Error Subspace Statistical Estimation (.pdf). Proceedings of "The OCEANS 2002 MTS/IEEE" conference, IEEE, Holland Publications, 795-802.

C. Biochemical-Physical Smoothing in Massachusetts Bay

A joint estimation of biochemical-physical variabilities and uncertainties in Massachusetts Bay is carried out via ESSE for the late summer of 1998. It is based on multiscale interdisciplinary data sets collected during the Littoral Ocean Observing and Predicting System (LOOPS-98) experiment. The LOOPS-98 scientific focus was phytoplankton and zooplankton patchiness, in particular, the spatial variability of zooplankton and its relationship to physical and phytoplankton variabilities. Synoptic physical and biochemical data sets were obtained simultaneously over a range of spatial and temporal scales. Real-time interdisciplinary nowcasts and forecasts were issued [1] and quantitative adaptive sampling was carried out [10]. The focus here is on the ESSE smoothing during the Aug. 25–Sep. 2, 1998 period.

Biochemical and Physical data. During LOOPS-98, observations were gathered on multiple scales using ships and Autonomous Underwater Vehicles. The gathering occurred in three phases: the initialization surveys (17–21 Aug.), update surveys (2–4 Sep.) and two weeks of intensive operations (17 Sep.–5 Oct.). The main physical data consisted of temperature and salinity profiles at Bay-scale, mesoscale and sub-mesoscale resolutions. For the ecosystem, profiles of chlorophyll-a fluorescence and light levels, and bottles of nitrate, ammonium, chlorophyll-a and pheaopigment, were collected at mesoscales and sub-mesoscales resolution in Cape Cod Bay and at mesoscale resolution north of Cape Cod Bay. The Massachusetts Water Resources Authority (MWRA) also collected biochemical samples which are employed here.

Coupled Ocean Physics and Biochemical Models. The ocean physics model is the primitive-equation model used in Sect. III.A. Here, it is coupled to a biochemical model which governs the interactive evolution and spatial distribution of phytoplankton, zooplankton, detritus, nitrate, ammonium and chlorophyll-a (Chl-a). The model is nitrogen-limited: fluxes and state variables are expressed in terms of nitrogen (μ *mol N/l*) except the Chl-a compartment which is in ($mgCh/m^3$). For all biochemical state variables, a four-dimensional advective-diffusive-reactive equation is employed. The model parameters were estimated from a combination of *in situ* data, literature surveys and approximate dynamical constraints. After the real-time cruises, a wide range of parameter values were investigated and the optimal values found are used here. Further details on model parameters and structures are given in [1].

ESSE smoothing. The biochemical-physical fields and their dominant errors are first initialized. This initialization is carried out following the 3D methodology of [9] (as in Sect. III.A), using the above interdisciplinary data and dynamical models. Presently, only the period Aug. 25–Sep. 2 of the ESSE simulation is discussed. The results of the coupled dynamical forecasts and of the assimilation via ESSE filtering and ESSE smoothing backward in time are illustrated on Fig. 8.

The nowcast on Aug.25 (Fig.8a) clearly indicates patchiness in the Chl-a field. At the end of the 1998 summer, the Chl-a maxima are around 4–7 $mgChl/m^3$ and located near 20 m depth. As September nears, storms of increasing strength and frequency pass over the region, sub-mesoscale and

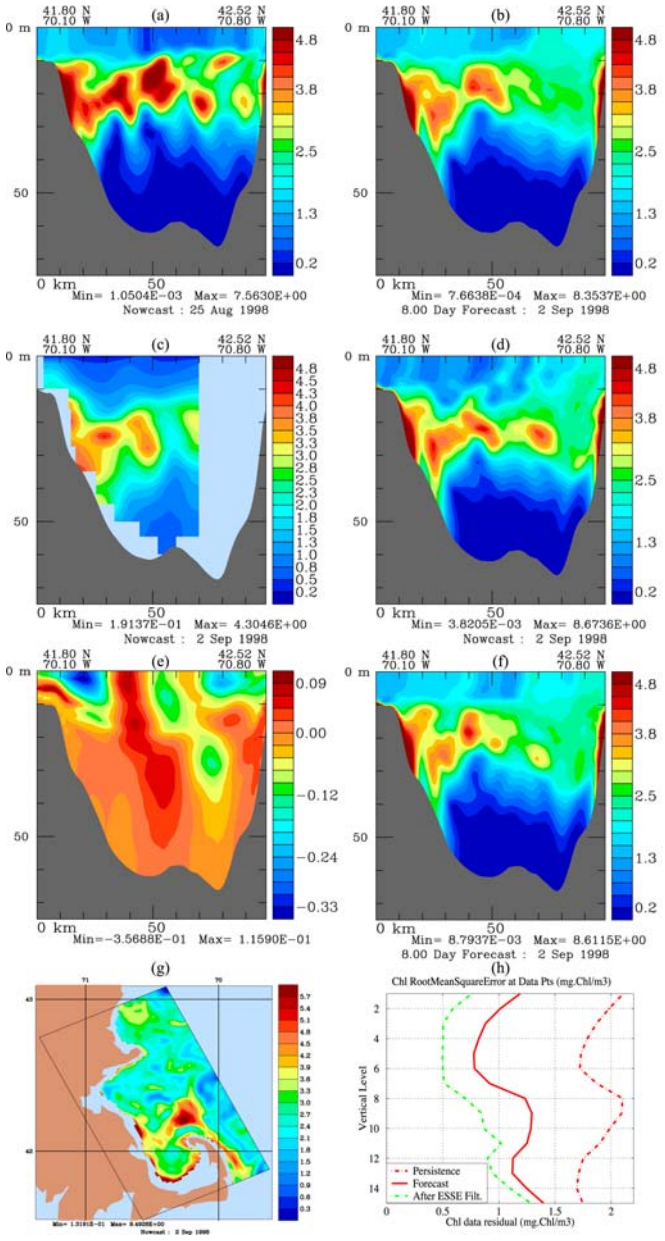

Fig. 8. Panels (a–f): Cross-sections in Chl-a fields, from south to north along the main axis of Massachusetts Bay, with, a) Nowcast conditions on Aug. 25; b) Forecast for Sep. 2; c) 2D objective analysis for Sep. 2 of the Chl-a data collected on Sep. 2–3; d) ESSE filtering estimate on Sep. 2; e) Difference between ESSE smoothing estimate on Aug. 25 and nowcast on Aug. 25; f) Forecast for Sep. 2, starting from ESSE smoothing estimate on Aug. 25. Panel (g): as d), but for Chl-a at 20 m depth. Panel (h): RMS differences between the Chl-a data on Sep. 2 and the field estimates at these data-points as a function of depth (specifically, "RMS-error" for persistence, dynamical forecast and ESSE

mesoscale variability builds up, and light levels start to decrease. The combination of these effects lead to a horizontal homogeneization and vertical mixing of the Chl-a field, an increase of Chl-a at the surface in response to the mixing and in some locations to decreased light inhibition, and a deepening of the Chl-a maxima. The results of these phenomena are visible on the forecast for Sep. 2 (Fig. 8b).

On Sep. 2–3, the available data are assimilated via ESSE. A cross-section in a 2D objective analysis for Sep. 2 of the Chl-a data collected during Sep. 2–3 is shown on Fig. 8c (note that values are masked for a non-dimensional data error larger than 70%). The result of the ESSE assimilation with an error subspace of size 600 is shown on Fig. 8d. Even though the interdisciplinary model forecast (Fig. 8b) was quite good, the data corrections are clearly visible on Fig. 8d. The differences between Fig. 8d and Fig. 8c provide a good example of the benefit of data assimilation and dynamical interpolation: on average, the 2D objective analysis (Fig. 8c) and the ESSE filtering field (Fig. 8d) have a similar departure from the actual data values (their RMS-differences with respect to data are similar). However, the ESSE filtering field is more realistic away from data positions, with multiple non-homogeneous scales, than the 2D objectively analyzed field.

Once the ESSE filtering is completed, the ESSE smoothing is carried out, leading to an update of the nowcast on Aug. 25. The difference between this ESSE smoothing field (not shown) and the nowcast on Aug. 25 (Fig. 8a) is shown on Fig. 8e. Starting from the ESSE smoothing estimate on Aug. 25, the forward dynamical model is integrated up to Sep. 2, leading to the forecast from the ESSE smoothing estimate. This forecast (Fig. 8f) is clearly closer to the Sep. 2 data than the first forecast (Fig. 8b) started from the Aug. 25 nowcast (Fig. 8a). However, the forecast from the ESSE smoothing estimate is not as good as the ESSE filtering field (Fig. 8d). This is expected and is due to the: (i) nonlinearities, (ii) perfect dynamical model assumption (model errors were used in Sect. III.A, but are not used here), and (iii), error subspace truncation.

To indicate the 4D aspects of the simulation, a horizontal map is provided on Fig. 8g, showing the Sep. 2 values of the Chl-a field at 20 m as estimated by ESSE filtering. The increased Chl-a values along the shallow coastline in Cape Cod Bay indicate an initial phase of the Fall bloom in the region. The locations of the high Chl-a just north of Cape Cod Bay, at the southern tip of Stellwagen Bank (north of the tip of Cape Cod) and to the south-east of Cape Cod, correspond to the regional physical circulations and to the path of the Gulf of Maine coastal current in Massachusetts Bay around that time. Higher sub-mesoscale mixing occur along the fronts and eddies associated to these circulation features, which leads to higher biological biomass.

Finally, the skill of the ESSE simulation is quantitatively evaluated as a function of depth. Fig. 8h shows the horizontally-averaged RMS differences between Chl-a data on Sep. 2 and three field estimates: persistence (initial conditions on Aug. 25), dynamical forecast for Sep. 2 and ESSE filtering estimate for Sep. 2. Based on these curves, the dynamical forecast is about 70 to 100% better than persistence, while the ESSE filtering estimate is about 15 to 50% better than the dynamical forecast. The forecast fields are relatively good because the dynamical model parameters have been tuned to the data and because the model is forced in surface by atmospheric analyses.

IV. CONCLUSIONS

ESSE has been shown to be useful for a wide range of purposes as illustrated by three particular investigations. The smoothing estimation of physical ocean fields in the Eastern Mediterranean provided dynamically consistent fields superior to those obtained by statistical methods alone. The coupled physical-acoustical data assimilation in the Middle Atlantic Bight shelfbreak demonstrated the multiscale nature of ESSE in recovering the fine scale transmission loss signal from coarser resolution data and physical ocean model. The impacts of interdisciplinary ESSE filtering and smoothing were quantified in coupled physical-biological dynamical simulations for Massachusetts Bay. The ESSE estimates provided evidence of patchiness in the Chl-a field and highlighted the effects of storms, of sub-mesoscale to mesoscale variability and of decreasing light levels on the Chl-a field at the end of the 1998 summer.

With ESSE, the sub-optimal reduction of errors is itself optimal. Without such reductions, real-time and accurate predictions of uncertainties are not directly feasible with the computers available today. On average, the computational cost of ESSE is about 10^3 to 10^4 times smaller than that of classic estimation schemes (e.g. Kalman smoother).

Other recent examples of the applications of ESSE include: dynamical studies of the dominant mesoscale features and water pathways in the Strait of Sicily [8,11]; operational adaptive sampling in the Gulf of Cadiz [20]; coupled physical-acoustical uncertainty predictions in the Middle Atlantic Bight shelfbreak [12]; physical variability forecast and dynamics in Massachusetts Bay [10]; and, visualization of uncertainties [4].

Acknowledgments

PFJL thanks Prof. C.-S. Chiu for his collaboration in the acoustic computations,the MWRA for their biochemical data, and Ms. M. Armstrong for help in preparing the document. This was work supported in part by the Office of Naval Research under grants N00014–95–1–0371, 97–1–0239, 97–1–1018 and 00–1–0771 and by the National Science Foundation under grant NSF-5710001319.

Lermusiaux, P.F.J., A.R. Robinson, P.J. Haley and W.G. Leslie, 2002. Advanced interdisciplinary data assimilation: Filtering and smoothing via Error Subspace Statistical Estimation (.pdf). Proceedings of "The OCEANS 2002 MTS/IEEE" conference, IEEE, Holland Publications, 795-802.

References

[1] Besiktepe S.T., P.F.J. Lermusiaux and A.R. Robinson. Coupled physical and biochemical data driven simulations of Massachusetts Bay in late summer: real-time and post-cruise data assimilation. Special issue on "The use of data assimilation in coupled hydrodynamic, ecological and bio-geo-chemical models of the oceans," M. Gregoire, P.F.J. Lermusiaux and P. Brasseur (Eds.), Submitted to the *J. of Marine Systems*, (2002).

[2] Chiu, C.-S., Downslope modal energy conversion, *J. Acoust. Soc. Am.*, **95**(3), 1654–657, (1994).

[3] Chiu, C.-S., J.H. Miller and J.F. Lynch, Forward coupled-mode propagation modeling for coastal acoustic tomography, *J. Acoust. Soc. Am.*, **99**(2), 793–802, (1996).

[4] Djurcilov S., K. Kim, P.F.J. Lermusiaux and A. Pang. Visualizing scalar volumetric data with uncertainty. *Computers and Graphics*, **26**(2): 239–248, (2002)

[5] Lermusiaux, P.F.J. Error subspace data assimilation methods for ocean field estimation: theory, validation and applications. PhD Thesis, May 1997, Harvard Univ., Cambridge, MA, (1997).

[6] Lermusiaux, P.F.J. Data assimilation via Error Subspace Statistical Estimation. Part II: Middle Atlantic Bight shelfbreak front simulations and ESSE validation. *Month. Weather Rev.*, **127**(7), 1408–1432. (1999a).

[7] Lermusiaux P.F.J. and A.R. Robinson. Data assimilation via Error Subspace Statistical Estimation. Part I: Theory and schemes. *Month. Weather Rev.*, **127**(7), 1385–1407. (1999)

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

[9] Lermusiaux P.F.J., D.G.M. Anderson and C.J. Lozano. On the mapping of multivariate geophysical fields: error and variability subspace estimates. *Q.J.R. Meteorol. Soc.*, April B, 1387–1430. (2000).

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

[11] Lermusiaux, P.F.J. and A.R. Robinson. Features of dominant mesoscale variability, circulation patterns and dynamics in the Strait of Sicily. *Deep Sea Research*, **48**(9), 1953–1997, (2001).

[12] Lermusiaux P.F.J., C.-S. Chiu and A.R. Robinson. Modeling Uncertainties in the Prediction of the Acoustic Wavefield in a Shelfbreak Environment. *Proceedings of the 5th International conference on theoretical and computational acoustics,* May 21–25, 2001. Beijing, China. In press, (2001).

[13] Lermusiaux P.F.J. On the mapping of multivariate geophysical fields: sensitivity to size, scales and dynamics. *J. of Atmos. Oceanic Tech.*, In press. (2002).

[14] Lermusiaux, P.F.J. and C.-S. Chiu. Four-dimensional data assimilation for coupled physical-acoustical fields. In "*Acoustic Variability, 2002*". Saclantcen. Kluwer Academic Press, In press, (2002).

[15] Lynch, J.F., G.G. Gawarkiewicz, C.-S. Chiu, R. Pickart, J.H. Miller, K.B. Smith, A. Robinson, K. Brink, R. Beardsley, B. Sperry, and G. Potty. Shelfbreak PRIMER - An integrated acoustic and oceanographic field study in the mid-Atlantic Bight, in *Shallow-Water Acoustics* (R. Zhang and J. Zhou Editors), China Ocean Press, 205–212, (1997).

[16] Malanotte-Rizzoli, P., A.R. Robinson, W. Roether, B. Manca, A. Bergamasco, S. Brenner, G. Civitarese, D. Georgopoulos, P.J. Haley, S. Kioroglou, H. Kontoyannis, N. Kress, M.A. Latif, W.G. Leslie, E. Ozsoy, M. Ribera d'Alcala, I. Salihoglu, E. Sansone, and A. Theocharis. Experiment in Eastern Mediterranean probes origin of deep water masses. *EOS*, **77**(32), (1996).

[17] Robinson, A.R. Physical processes, field estimation and an approach to interdisciplinary ocean modeling. *Earth-Science Rev.*, 40, 3–54. (1996)

[18] Robinson A.R., P.F.J. Lermusiaux and N.Q. Sloan, III. Data assimilation. In "The Sea: The Global Coastal Ocean I", Processes and Methods (K.H. Brink and A.R. Robinson, editors), Volume 10, John Wiley and Sons, New York, NY. (1998).

[19] Robinson A.R. and P.F.J. Lermusiaux. Data Assimilation in Models. *Encyclopedia of Ocean Sciences*, 623–634, Academic Press Ltd., London. (2001).

[20] Robinson A.R. and P.F.J. Lermusiaux. Data assimilation for modeling and predicting coupled physical-biological interactions in the sea. In "The Sea, Vol. 12: Biological-Physical Interactions in the Ocean.", Robinson A.R., J.R. McCarthy and B.J. Rothschild (Eds.). pp 475–536, (2002).