Fifth Framework Programme of the European Commission 1998-2002
Energy, Environment and Sustainable Development (EESD) Programme

FerryBox

From On-line Oceanographic Observations to Environmental Information

Report on Suitability of Existing Data Assimilation Schemes for Use with Ferrybox Data

Contract number :	**EVK2-2002-00144**
Deliverable number :	**D-5-2**
Revision :	**2.0**

Co-ordinator:

Professor Dr. Franciscus Colijn

GKSS Research Centre
Institute for Coastal Research
Max-Planck-Strasse
D-21502 Geesthacht

http://www.ferrybox.org

Document Reference Sheet

This document has been elaborated and issued by the European FerryBox Consortium.

P 1	**GKSS**	**GKSS Research Centre** **Institute for Coastal Research**	**Coordinator**
P 2	**NERC.NOC**	**NERC.NOC – National Oceanography Centre** **Southampton University and National Environment Res. Council** formerly NERC.SOC – Southampton Oceanography Centre	
P 3	**NIOZ**	**Royal Netherlands Institute of Sea Research**	
P 4	**FIMR**	**Finnish Institute of Marine Research**	
P 5	**HCMR** (formerly NCMR)	**Hellenic Centre for Marine Research** (formerly National Centre for Marine Research)	
P 6	**NERC.POL**	**Proudman Oceanographic Laboratory**	
P 7	**NIVA**	**Norwegian Institute for Water Research**	
P 8	**HYDROMOD**	**HYDROMOD Scientific Consulting**	
P 9	**CTG** **(formerly CIL)**	**Chelsea Technology Group** **(formerly Chelsea Instruments Ltd.)**	
P 10	**IEO**	**Spanish Institute of Oceanography**	
P 11	**EMI**	**Estonian Marine Institute** (in cooperation with the Estonian Maritime Academy)	

The European FerryBox Project was co-funded by the European Commission under the Fifth Framework Programme of the European Commission 1998-2002 - Energy, Environment and Sustainable Development (EESD) Programme under contract no. EVK2-2002-00144

Document Control Table

Project acronym:	FerryBox		Contract no.:	EVK2-2002-00144
Deliverable No.:	D-5-2		Revision:	2.0
WP number and title:	FerryBox WP-5	Applications of FerryBox data		
Work Package Manager:	Roger Proctor – Proudman Oceanographic Laboratory			
Work Package Team:	FerryBox WP-5 Team			
Document title:	Report on Suitability of Existing Data Assimilation Schemes for Use with Ferrybox Data			
Document owner:	European FerryBox Project Consortium			
Document category:	Deliverable			
Document classification:	PU – Public			
Status:	Final			
Purpose of release:	Deliverable for the European Commission			
Contents of deliverable:	Report on Suitability of Existing Data Assimilation Schemes for Use with Ferrybox Data			
Pages (total):	22	Figures:	4	Tables: 0
Remarks:	Updated revision for publication on the FerryBox report CD and website			
Main author / editor:	Roger Proctor	FerryBox WP-5 Leader	NERC.POL	
Contributors:	FerryBox WP-5 Team			
Main contacts:	FerryBox project coordinator:		Contact for this report:	
	Professor Dr. Franciscus Colijn GKSS Research Centre Institute for Coastal Research Max-Planck-Strasse D-21502 Geesthacht, Germany Tel.: +49 4152 87 – 15 33 Fax.: +49 4152 87 – 20 20 E-mail: franciscus.colijn@gkss.de		Dr. Roger Proctor POL, Proudman Oceanographic Laboratory 6 Brownlow Street UK- Liverpool L35DA, United Kingdom Tel.: +44 151 795 4856 Fax.: +44 151 795 4801 E-mail: rp@pol.ac.uk	
Project website:	http://www.ferrybox.org			

Table of Contents

List of Figures

Report on Suitability of Existing Data Assimilation Schemes for Use with Ferrybox Data

Within the FerryBox Project (see http://www.ferrybox.org), ferries operating along 9 lines in European Seas[1] are equipped with boxes capable of measuring temperature, salinity, turbidity and fluorescence (chlorophyll-a) at the surface. In certain lines, additional equipment is available: oxygen/pH sensors, nutrient sensors/analysers, photosynthetic available radiation (PAR) for validation of satellite remote sensing data, Acoustic Doppler Current Profiler (ADCP) for water current measurement, Fast Repetition Rate Fluorometer (FRRF) and an algal classes detector.

In FerryBox work package WP-5, one of the tasks is to test and optimize existing data assimilation algorithms for their use with FerryBox data. Fields addressed are temperature, salinity, nutrients and suspended sediment concentrations. The modelled areas will be the Irish Sea, the Southern North Sea and the Mediterranean Sea, in all of which coastal processes play a key role.

This document introduces the particularities of the coastal ocean, where strong nonlinearities and anisotropies limit the range of suitable data assimilation methods. Some of the methods which have already proved successful in coastal systems are then presented with the focus on those which are coded into FerryBox models.

1 The Coastal Ocean

A thorough review on data assimilation and its specificities in coastal systems was written by Robinson *et al.* (1998).

In coastal areas, processes range from the mesoscale ocean in the inner shelf, to quicker and smaller scale processes near the shore. The flow can be strongly *nonlinear*, reducing the limit of *predictability*[2]. (Lorenz, 1963a; Lorenz, 1963b, Lorenz, 1965; Lorenz, 1968). As model resolution increases and smaller-scale phenomena are resolved, the validity of model linearisations is reduced, because error growth is quicker and the energy levels lower.

The particularities of coastal geometry and of topography and the influence of local forcing regimes make this an anisotropic system, with at least different scales in the long-shelf versus the cross-shelf directions. The spatial structure of the background (forecast) error statistics can not be stated *a priori* (Echevin, 2000; Echevin *et al.*, 2000).

A suitable assimilation scheme for high resolution monitoring/prediction systems has to account for the nonlinearities of the flow. It also has to have a fair representation of the spatial anisotropy of the background/forecast error and estimate its time evolution.

[1] Baltic Sea, Skagerrak, North Sea, Wadden Sea, Irish Sea, Solent Estuary, Bay of Biscay and Mediterranean Sea.

[2] At times longer than this limit, two different estimations of a same dynamical model with exceedingly small perturbations in the initial conditions will have become as different as two randomly chosen states of the model.

2 Some Suitable Methods for Assimilation in Coastal Systems

In this section, we comment on some methods that are potentially adequate for assimilation in coastal systems. These methods are grouped as sequential and variational methods.

2.1 Sequential Estimation

Generalisation of *Linear Estimation Theory* for *nonlinear* estimation led to the *Extended Kalman Filter (EKF)* (Jazwinski, 1970). In its formulation, it is assumed that the *nonlinear* observation (*H*), and model (*M*) operators can be approximated by their Jacobians if both the *background* an the *analyzed* state of the system lie close enough to the *true* state of the system.

Under this (*tangent linear*) hypothesis, the analysis $\mathbf{x}_k^a$ obtained at an observation time, t_k, is the sum of the background state $\mathbf{x}_k^b$ plus a correction term that weights the innovation **(**$\mathbf{y}_k^o$-$H\mathbf{x}_k^b$) brought up by the observation vector, $\mathbf{y}_k^o$, with respect to the background state at the observation locations, $H\mathbf{x}_k^b$.

$$x_k^a = x_k^b + P_k^b H_k^{'T}\left[H'_k P_k^b H_k^{'T} + R_k\right]^{-1}\left(y_k^o - H_k x_k^b\right),$$

The weights are given by the *Kalman gain matrix*:

$$K_k = P_k^b H_k^{'T}\left[H'_k P_k^b H_k^{'T} + R_k\right]^{-1},$$

where $\mathbf{P}^b_k$ is the background error covariance matrix, and $\mathbf{R}_k$ the observation error covariance matrix. *H* is the observation operator, that projects the background error onto the space generated by the observations (this can be interpolation of the model output from the model grid onto the observation points, but can also include transformations that express the model variables in terms of the observed fields, which can sometimes be different).

At each analysis step, the error statistics are also updated following:

$$P_k^a = P_k^b - P_k^b H_k^{'T}\left[H'_k P_k^b H_k^{'T} + R_k\right]^{-1} H'_k P_k^b,$$

Here, Pak is the new analyzed error variance, that gives an estimate of the accuracy of the analyzed state xak. H' is the Jacobian operator of the nonlinear operator H.

Between two analysis(/observation) times, the model calculates the evolution of the error covariance matrix as:

$$P_{k+1}^b = M' P_k^a M^{'T} + Q\ .$$

Approximations of the EKF have been successful in some cases (Fukumori and Malanotte-Rizzoli, 1995). Nevertheless, an article written by Evensen (1992) pointed out that this method could lead to unbounded instability of the error, due to linearization of the error covariance propagation equation. In that same article, Evensen concluded that the use of the full EKF would require either extensive data coverage able to control the instability, or some higher order closure equations that damp it. In a latter publication, Evensen (1993) showed the difficulty of handling open boundary conditions in the treatment of the approximate error evolution equation. He then proposed an alternative method, capable of coping with the problems associated to the EKF in *nonlinear* systems.

2.2 The Ensemble Kalman Filter (EnKF)

The alternative method proposed by Evensen(1994) was the Ensemble Kalman Filter (EnKF). In this scheme, the probability density function of the initial state is represented by an ensemble of initial states of mean equal to the best guess initial condition. Initial variance of the ensemble is set equal to that first guess estimated uncertainty. At each analysis step, the first and second order statistical moments of the system are estimated from the ensemble of states attained through integration of the initial ensemble (*a priori ensemble*).

Brusdal *et al.* (2003) discuss three ensemble based methods (Singular Evolutive Extended Kalman, Ensemble Kalman Filter and Ensemble Kalman Smoother) from the perspective of operational marine monitoring and forecasting systems.

2.3 The Singular Evolutive Extended Kalman Filter (SEEK)

Pham *et al.* (1998) present a Single Evolutive Extended Kalman Filter, that parameterizes the initial error covariance matrix as a low rank matrix in the phase subspace spanned by the fastest error growing directions:

$$P_k^a = L_k U_k L_k^T,$$

with Lk,the column vector containing the directions of greatest error amplitude (EOFs) and Uk a positive definite matrix containing the associated eigenvalues.

The error subspace spanned by these EOFs evolves during the assimilation according to model dynamics.

Brasseur *et al.* (1999) implemented an improved version of the filter in which the statistical structure of the error covariance is learned from the innovation vector at each analysis step.

The SEEK filter provides a correction in accordance with the fluid situation at each time-step. It is suitable for coastal applications, as it doesn't need assumptions on homogeneity or symmetry of **P**.

This method is widely used at present, with applications in ecosystem modelling (Hoteit *et al.*, 2003) amongst others.

2.4 The Error Subspace Statistical Estimation (ESSE) Approach

ESSE (Lermusiaux, 1997; Lermusiaux and Robinson, 1999) is an efficient statistical estimation scheme for data assimilation in realistic nonlinear ocean models. In this approach the error covariance matrix is truncated on it's dominant EOFs at each assimilation time. Nonlinear Monte Carlo forecasts that use the full nonlinear model integrate this error subspace (ES) in time. The members of the *a priori* ensemble are chosen so as to optimally sample the error subspace. The *a posteriori* members are used to sample the new ES. The melding criterion minimizes the variance in this ES. Therefore, the dynamical forecast is corrected by data there where the errors are more energetic. The assimilation is multivariate and 3D in physical space. It also takes into account model nonlinearities and errors in an explicit manner. This method can be implemented through parallel computing.

ESSE is much less costly than classical analyses involving full covariances. Furthermore, the evolving ES is allowed to vary in size, eigenbase and eigenvalues with time and on multiple scales (Lermusiaux, 2001).

The general concept, objectives and details of *Error Subspace Statistical Estimation* are discussed in Lermusiaux (1997). Sloan (1996) compares results obtained through this method and through OI in a twin experiment of an idealized shelf-break-front simulation: ESSE improves the quality of the retrieval of the true ocean, with better representation of the non-homogeneous and anisotropic physical characteristics (shape, location and scales of physical characteristics).

2.5 Variational Estimation

In variational estimation, the solution of the *Generalized Inverse problem,* is a set of states (*modal points*) that minimize the objective or cost function $\mathbf{J_N}$. This accounts for the misfit between the real initial conditions of the system and the initial conditions actually used ($\mathbf{J_0}$) , the misfit between real and available boundary conditions ($\mathbf{J_{ob}}$), that between model forecast and observations at observation times ($\mathbf{J_o}$) and the contribution of model error ($\mathbf{J_m}$). In each of the terms of $\mathbf{J_N}$, misfits are weighted by the corresponding accuracies, represented as the inverses of the error covariance matrices.

$$
\begin{aligned}
J_N &= \frac{1}{2}\left(x_b^0 - x^0\right)\left(P_b^0\right)^{-1}\left(x_b^0 - x^0\right)^T \\
&+ \frac{1}{2}\left(x_b^{ob} - x_k^{ob}\right)\left(P_b^{ob}\right)^{-1}\left(x_b^{ob} - x_k^{ob}\right)^T \\
&+ \frac{1}{2}\sum_{k=1}^{N}\left(Hx_k - y^o\right)\left(R_k\right)^{-1}\left(Hx_b - y^o\right)^T \\
&+ \frac{1}{2}\sum_{k=1}^{N}\left(x_k - Mx_{k-1}^t\right)\left(Q_{k-1}\right)^{-1}\left(x_k - Mx_{k-1}^t\right)^T \\
&= J_0 + J_{ob} + J_o + J_m
\end{aligned}
$$

In variational assimilation the problems arising from *strongly nonlinear* dynamics suggest the use of direct substitution algorithms, that solve the *generalized inverse problem* without the integration of forward and adjoint model equations. The method that we would retain is a *Gradient Descent Method*.

2.6 Gradient Descent Methods

These methods iteratively determine directions descending locally along the cost function surface. At each iteration, a line search minimization is performed along that local direction and a new descending direction is found. Examples of *descent methods* are the *conjugate gradient method* and the *Newton and quasi-Newton methods*.

Evensen and Fario (1997) used a gradient descent method that minimized the weak constraint problem by calculating the gradient of the cost function with respect to the full state in space and time. It consisted of an iterative method in which a new candidate for the solution was substituted after each computation. The conclusions of this work were that the method would work well for good data coverage. For poor data coverage *Simulated Annealing* was proposed as a better alternative.

Evensen (1997) compares the goodness of the EnKF, of an ensemble smoother method (van Leeuwen and Evensen, 1996) and of the gradient descent method to solve the highly non linear Lorenz model (Lorenz, 1963a).

The major drawback of descent methods is that they are all initialization sensitive. For nonlinear cost functions they must be restarted several times to avoid local minima.

3 Data Assimilation Schemes within FerryBox

3.1 In the HCMR POSEIDON Model System for the Mediterranean

An assimilation scheme based on the SEEK filter has been implemented into the Princeton Ocean Model which is the core of the POSEIDON pre-operational high resolution hydrodynamic model. The latter is a component of the POSEIDON operational monitoring and forecasting system that operates in the Aegean Sea since 1999 (Nittis et al., 2001).

SEEK filter being a multivariate sequential assimilation method is an efficient tool for optimally handling observation data sets such as those (sea surface temperature and salinity) that will be derived through the FerryBox activities at HCMR.

We describe briefly the theory behind the SEEK filter and its variant the SFEK filter as well as their implementation into the POM model.

3.1.1 The SEEK/SFEK Filters

The Singular Evolutive Extended Kalman filter (SEEK) introduced by Pham et al. (1998) is an extended Kalman filter based on a singular low rank error covariance matrix. SEEK approximates the full error covariance matrix with a singular low rank matrix reducing in this way the implementation cost and on the other hand improving the filter stability. Thus, instead of reducing the system (model) state vector the filter approximates the error propagation respecting in this way the system dynamics.

Assume an ensemble of random members that describe the probability distribution of a given system. An EOF (Empirical Orthogonal Functions) decomposition of the ensemble usually defines those few dominant directions that are necessary to describe the spreading of the ensemble. SEEK filter uses these dominant directions to construct its error subspace initially and evolves them in time according to the extended Kalman filter equations.

Let the error covariance matrix P_i^a be factorized as $L_i U_i L_i^T$.

The SEEK filter operates in two stages:

Forecast Stage

The model is used to compute the forecast state X^f at time t_i initialized from the analysis state X^a available at time t_{i-1}:

$$X^f(t_i) = M\left(t_{i-1}, t_i\right) X^a(t_{i-1}) \quad (1)$$

where $M(t_{i-1}, t_i)$ is the model transition operator.

Correction Stage

The analysis state X^a at time t_i is estimated as:

$$X^a(t_i) = X^f(t_i) + K_i\left[Y_i^o - H_i X^f(t_i)\right] \qquad (2)$$

where the Kalman gain matrix K_i is given by

$$K_i = L_i U_i L_i^T \mathrm{H}_i^T R_i^{-1} \quad (3)$$

and

$$U_i^{-1} = \rho U_{i-1}^{-1} + L_i^T \mathrm{H}_i^T R_i^{-1} \mathrm{H}_i L_i \qquad 0 < \rho < 1 \quad (4)$$

In Eq.2, Y_i^o is the observations vector (available at time t_i) and H_i is the observation operator which simply projects the model's state onto the locations of the observations. In Eq.3, H_i is the gradient of the observation operator and R_i is the observations error covariance matrix. The latter is usually taken as a diagonal matrix by assuming that the observations are spatially uncorrelated. Note that in cases of a linear observation operator $\mathrm{H}_i = H_i$.

The correction basis L_i evolves in time according to:

$$L_i = \mathrm{M}(t_{i-1}, t_i)\, L_{i-1} \qquad (5)$$

where $\mathrm{M}(t_{i-1}, t_i)$ is the gradient of the model transition operator evaluated at $X^a(t_{i-1})$. The analysis error covariance matrix $P^a(t_i)$ is given by:

$$P^a(t_i) = L_i U_i L_i^T \qquad (6)$$

The factor ρ appearing in Eq.4 may be interpreted as a forgetting factor and stands for the increment of uncertainty during the model integration from t_{i-1} to t_i due to model errors. Practically, with $\rho < 1$ recent observational data are exponentially more weighted than old data.

3.1.2 The SFEK filter

Brasseur et al. (1999) motivated by the fact that most of the error estimation in SEEK filter is reduced immediately after the first correction, proposed to keep the initial correction basis of the filter fixed in time. Thus in this implementation called the singular fixed extended Kalman filter (SFEK) the correction basis of the filter remains constant in time and equal to the initial EOF basis L_0. This approximation can be further justified by the fact that the state of the ocean evolves very slowly.

SFEK operates in two stages just like the SEEK filter but Eq.5 is not used. This drastically reduces the computational cost of the filter. However, the performance of the SFEK filter depends strongly on how well the model's variability is described by the initial EOF correction basis.

Forecast Stage

The model is used to compute the forecast state X^f at time t_i initialized from the analysis state X^a available at time t_{i-1}:

$$X^f(t_i) = M\left(t_{i-1}, t_i\right) X^a(t_{i-1}) \quad (7)$$

Correction Stage

The analysis state X^a at time t_i is estimated as:

$$X^a(t_i) = X^f(t_i) + K_i\left[Y_i^o - H_i X^f(t_i)\right] \quad (8)$$

where the Kalman gain matrix K_i is given by

$$K_i = L_0 U_i L_0^T \mathrm{H}_i^T R_i^{-1} \quad (9)$$

and

$$U_i^{-1} = \rho U_{i-1}^{-1} + L_0^T \mathrm{H}_i^T R_i^{-1} \mathrm{H}_i L_0 \qquad 0 < \rho < 1 \quad (10)$$

As can been seen by Eq.10, the projection of the error modes onto the error sub-space (U_i) evolves with the internal statistics of the filter. Thus although the error sub-space remains fixed in time the analysis error covariance matrix $P_i^a = L_0 U_i L_0^T$ is not steady.

Additionally, a third method, the interpolated ensemble-based variant of the SEEK filter, called SEIK filter (Hoteit et al., 2002) was implemented into the model and tested.

3.1.3 Initialisation of the Filters

The SEEK/SFEK filters are initialized from an analysis state vector $X^a(t_0)$ and an initial error covariance matrix P_0^a. The latter is estimated using an ensemble of sampled model states $X_1, X_2, \ldots\ldots, X_n$ extracted from a reference simulation. If $\overline{X}$ is the ensemble mean, the ensemble covariance matrix P_0 is given by

$$P_0 = \frac{1}{n}\mathbf{X}\mathbf{X}^T \quad (11)$$

where $\mathbf{X} = \left[X_1 - \overline{X}, \ldots\ldots, X_n - \overline{X}\right]$

The EOFs of the ensemble are the eigenvectors of the covariance matrix P_0 denoted as $S_1, S_2, \ldots S_{n-1}$ and ordered according to their eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots$.

Then the initial error covariance matrix can be approximated by selecting only the first r dominant EOFs which then determine the structure of the error sub-space at the initial time:

$$P_0^a = L_0 U_0 L_0^T \quad (12)$$

where $L_0 = [S_1, ..., S_r]$ and $U_0 = diag(\lambda_1, ..., \lambda_r)$

Such a procedure for the approximation of the initial error covariance matrix assumes that the covariance of the oceanic variability can be used as a proxy for the initial error covariance. Moreover it is assumed that the model variability is identical to the real ocean variability.

3.1.4 Filter Implementation into the Princeton Ocean Model

The SEEK/SFEK filter has already been applied with success to the Miami Isopycnic Coordinate Ocean Model (Brasseur et al., 1999; Brusdal et al., 2003) to the OPA model (Parent et al., 2003) and to the ERSEM ecosystem model (Triantafyllou et al., 2003). This is the first time that an attempt is made to apply the filter to the Princeton Ocean Model (POM) which is the core of the Aegean Sea hydrodynamic model (Korres et al., 2002)

POM is a free surface primitive equations general circulation model using curvilinear orthogonal horizontal coordinates and a sigma layers system in the vertical. It uses two sub-models for the computation of the eddy mixing (second order turbulence closure scheme) and horizontal non-linear viscosity. It also uses a time splitting technique for the external (barotropic) and internal (baroclinic) modes. POM has been extensively described in the literature (see for example Blumberg and Mellor, 1985, 1987). Potential temperature T, salinity S, velocity U,V, free surface elevation ζ, turbulent kinetic energy q^2 and turbulent kinetic energy times the turbulence length scale $q^2 l$ consist the prognostic variables of the model.

In order to validate the performance of the filter twin experiments were carried out using a ¼ × ¼ horizontal resolution – 25 sigma levels implementation of POM model into the Mediterranean basin forced with monthly climatological momentum, heat and freshwater fluxes.

The model state vector entering the filter algorithm consists of all prognostic variables of the model at each sea grid point, i.e. $X = [\zeta, T, S, U, V, q^2, q^2 l]$. For this particular application the dimension of the state vector is ~6×10^5.

As the SEEK/SFEK filter is a sequential algorithm, the assimilation proceeds through successive re-initializations of the model at each analysis step. Taking into account that POM is using a leap-frog temporal scheme, we have implemented an Euler time step in order to restart the model at each assimilation step. Moreover after the model state vector has been updated by the filter, the density field and the horizontal/vertical viscosity/diffusivity coefficients are recalculated for consistency.

The model initialized from middle-of-spring MODB-MED4 hydrological characteristics and zero velocities, has first been spun up for 15 years in order to reach a statistically steady seasonal cycle. An integration of 4 years was then performed in order to generate an ensemble of model states sampled every 2 days of model integration. Finally, restarting the model at the end of year 19 and doing a further integration of 2 years, pseudo-observations of sea surface elevation were extracted every 5 days.

This is considered as the reference or “true” run to be compared with the assimilation runs.

15 YEARS (SPIN-UP) → **4 YEARS (SAMPLES)** → **2 YEARS (TRUE RUN**

The ensemble of 720 model states sampled during the 4-year period was EOF decomposed in order to estimate the approximate error covariance matrix needed for the initialization of the assimilation scheme. Since the variables of the state vector are of different nature, each state variable was normalized by its standard deviation before the EOF analysis was conducted. In Figure 3-1 we show the percentage of variance explained by the number of retained EOFs for the first eighty of them. By truncating the EOF series to the first 20 modes almost 94% of the total ensemble variance is explained.

Assimilation experiments are initialized from an early-winter state extracted during the 16th year of model integration. Free surface elevation data at a spatial resolution of 1 deg extracted from the “true” run of the model are assimilated every 5 days using either the SFEK of the SEEK filter with an error sub-space of rank 20. The observation error matrix is diagonal with a 2.6 cm RMS error and the observation operator is a bilinear interpolation of the model grid to the observation grid. Based upon sensitivity studies an optimal value for the forgetting factor was found to be equal to 0.8.

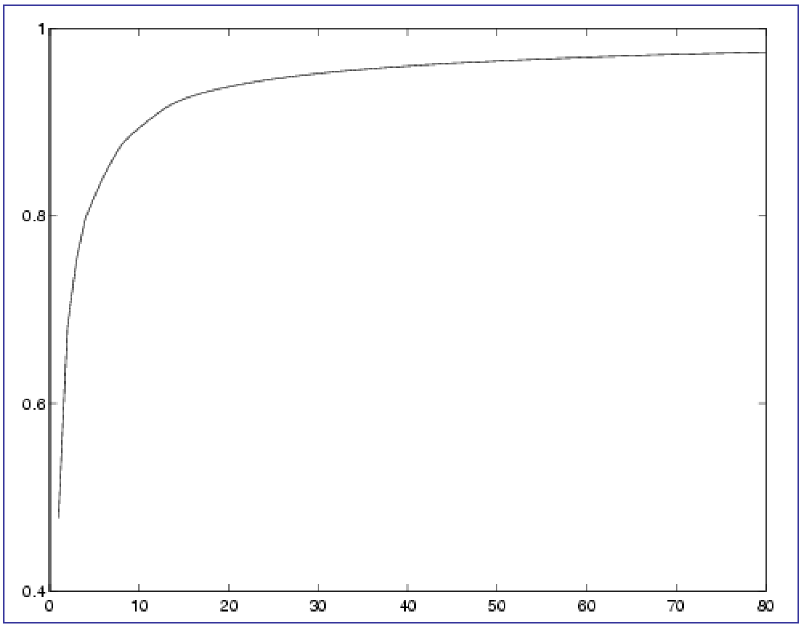

Figure 3-1: Percentage of variance explained versus number of retained EOFs.

In order to assess the performance of the filters, we compare the misfit between the true run, the assimilation and the pure forecast (the pure forecast is initialized from the same state as the assimilation experiment). The relative error of each state variable is defined as the ratio between assimilation error and forecast error:

$$e = \left[\frac{\int \left(x_{ASSIMILATION} - x_{TRUE} \right)^2 dV}{\int \left(x_{FORECAST} - x_{TRUE} \right)^2 dV} \right]^{1/2} \quad (13)$$

The relative error is evaluated according to (13) for temperature, salinity, velocity and free surface elevation. Relative errors smaller than unity indicate that the assimilation estimates are in average closer to the true run than pure predictions.

The global relative error (i.e. the mean relative error over all model variables) for the three filters (SFEK/SEEK and SEIK) is shown in Figure 3-2. After 5 assimilation cycles (25 days) the relative error for SFEK filter has been decreased to 34%, it is relatively higher for the SEEK filter (39.5%) and is almost 30% for SEIK filter. SEEK and SEIK filters converge at approximately the same rate during the first year of model integration while SFEK filter shows a relatively faster convergence.

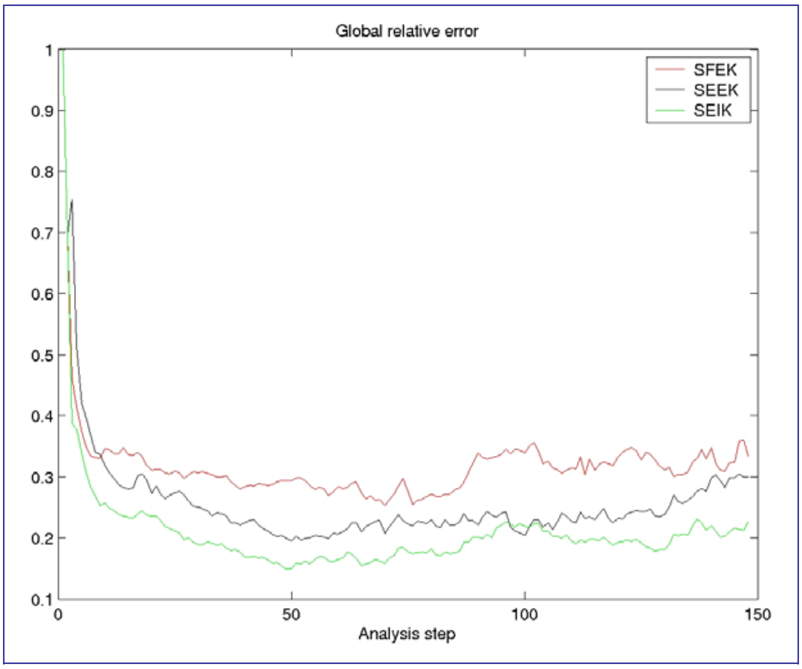

Figure 3-2: Evolution of global relative error as a function of time for the SFEK, SEEK and SEIK filters during the assimilation period 1986-1987.

After 50 assimilation cycles the global relative error for SFEK filter converges to a value of 30% and stays there with small undulations until the winter of the next year. The respective values for SEEK and SEIK filters are 19.5% and 15%. Overall, SEEK and SEIK filters were found to provide a reasonably good analysis for all model variables of the state vector and even for the fast evolving ones like the SSH. On the other hand, in these experiments the SFEK filter was found to provide an acceptable level of performance considering its time-invariance assumption of the correction basis.

3.1.5 References

Blumberg, A.F., Mellor, G.L.,1985. A simulation of the circulation in the Gulf of Mexico. Israel Journal of Earth Sciences, 34, 122-144.

Blumberg, A.F., Mellor, G.L., 1987. A description of a three-dimensional coastal ocean circulation model. In Three-Dimensional Coastal Ocean Circulation Models, Coastal Estuarine Sci., vol. 4, edited by N.S.Heaps, pp. 1-16, AGU, Washington, D.C.

Brasseur, P., Ballabrera-Poy, J., Verron, J., 1999. Assimilation of altimetric data in the mid-latitude oceans using the Singular Evolutive Kalman filter with an eddy-resolving, primitive equation model. J. Mar. Syst., 22, 269-294.

Brusdal, K., Brankart, J.M., Halberstadt, G., Evensen, G., Brasseur, P., Leeuwen, P.J., Dombrowsky, E., Verron, J., 2003. A demonstration of ensemble-based assimilation methods with a layered OGCM from the perspective of operational ocean forecasting systems. J. Mar. Syst., 40-41, 253-289.

Korres, G., Lascaratos, A., Hatziapostolou, E., Katsafados, P., 2002. Towards an ocean forecast system for the Aegean Sea. The Global Atmosphere and Ocean System, Vol. 8, No. 2-3, 191-218.

Nittis, K., Zervakis, V., Perivoliotis, L., Papadopoulos, A., Chronis, G., 2001. Operational Monitoring and Forecasting in the Aegean Sea: Limitations and Forecasting Skill Evaluation. Marine Pollution Bulletin, 43/7-12, 154-163.

Parent, L., Testut, C.E., Brankart, J.M., Verron, J., Brasseur, P., Gourdeau, L., 2003. Comparative assimilation of Topex/Poseidon and ERS altimeter data and of TAO temperature data in the tropical Pacific Ocean during 1994-1998, and the mean sea-surface height issue. J. Mar. Syst., 40-41, 381-401.

Pham, D.T., Verron, J., Roubaud, M.C., 1998. A singular evolutive extended Kalman filter for data assimilation in oceanography. J. Mar. Syst., 16, 323-340.

Triantafyllou, G., Hoteit, I., Petihakis, G., 2003. A singular evolutive interpolated Kalman filter for efficient data assimilation in a 3-D complex physical-biogeochemical model of the Cretan Sea. J. Mar. Syst., 40-41, 213-231.

3.2 In the Proudman Oceanographic Laboratory Coastal Ocean Modelling System (POLCOMS) for the Irish Sea

Two variants of the Kalman Filter are coded into POLCOMS: 1) a simplified Kalman Filter (Annan and Hargreaves, 1999) to assimilate sea surface properties, and, 2) the Ensemble Kalman Filter (Evensen, 1994, 2003, 2004) to assimilate sea surface properties combined with available water column profiles. Both these schemes are capable of assimilating FerryBox-type data, i.e. irregular in space and time.

3.2.1 A Simplified Kalman Filter (Annan and Hargreaves, 1999)

This algorithm was used to apply a correction of the forecast temperature within the mixed layer by weighting the difference between the observed Sea Surface Temperature (SST) and the model forecast SST. The weighting factor is a balance between the confidence attributed to both the data and the model forecast. This is expressed in equation (a), where the analysed temperature, $t_l^a(i,j,k)$, at the grid points *i, j, k* at time-step *l,* is expressed as the sum of the corresponding forecast temperature, $t_l^f(i,j,k)$, plus a correction given by the product of the model-to-observation difference and the Kalman gain $\mathrm{K}_l(i,j,k)$:

$$t_l^a(i,j,k) = t_l^f(i,j,k) + \mathrm{K}(i,j,k)\left[sst_l^o(i,j) - sst_l^f(i,j)\right] \quad \text{(a)}$$

In the present version of the algorithm, the correction operates vertically within the diagnosed Mixed Layer (ML). This is justified by the following hypotheses:

i) It is supposed that errors in the model upper thermal structure are mainly due to inaccuracies in the heat fluxes. Details on the estimation of model error can be found in Annan and Hargreaves (1999).

ii) The time-scale of error propagation across the horizontal grid due to residual horizontal currents is supposed to be longer than that associated to vertical mixing.

iii) The turbulent kinetic energy is supposed close to a quasi-equilibrium state and negligible mixing is assumed across the thermocline. Temperature within the mixed layer can therefore be adjusted without modifying the position of the thermocline.

At observation times, observations are read in and interpolated onto the model grid. The difference between observations and forecast is computed for every sea point, as well as the weighting factor for the correction. The analyzed temperature is then produced down the diagnosed mixed layer. [The mixed layer depth (MLD) is diagnosed as the depth (starting from the surface) at which the turbulent kinetic energy becomes weaker than a fixed threshold. This threshold has been set to 0.001 m^2s^{-1}]

As this algorithm corrects only the forecast for those grid-points for which observations are available, error must be bound to avoid unrealistic values for points where no assimilation takes place. A modification was therefore introduced to this algorithm to set a saturation value for the forecast error variance.

The code can run on parallel as well as vector machines and be fitted into different codes, using different SST data sets. It was first tested on a Linux Box in a 12 km setup of the Irish Sea, to assimilate 9 km resolution AVHRR SST observations. Results showed an improvement of the thermal representation in the upper layers of the water column. A second set of experiments involving a 1.8 km Irish Sea setup were then run through MPI to assimilate 9 km and 2 km resolution satellite SST products respectively. Results are presented in Andreu-Burillo *et al.* (2005).

3.2.2 The Ensemble Kalman Filter (Evensen, 1994 and 2004)

The EnKF is a Monte Carlo method that aims at sampling the evolution of the probability density function (*pdf*) of the model state (Evensen, 1994, *JGR*). First, a set of initial conditions is built to obtain an *a priori* ensemble. The mean of this ensemble must be equal to a known first guess of the corresponding state of the system and its variance represents the uncertainty associated to that first guess. The ensemble is then run forward in time under the effect of forcing and model noises. These can be parameterized as stochastic perturbations with or without spatio-temporal coherence, added to the first-guess forcing fields or model parameters that are supposed to mostly account for forecast errors. At each model time-step, the best estimate of the true state of the system is provided by the mean of the ensemble:

$$\overline{x} = \frac{1}{m}\sum_{i=1}^{m} x_i .$$

The error covariances can then be obtained through:

$$P = \left\langle x - \overline{x}, x - \overline{x} \right\rangle .$$

The EnKF does not need to calculate the evolution of the error covariance matrix, as the necessary statistical information for the analysis is carried by the ensemble itself.

Our implementation of the EnKF is based on that proposed by Evensen (2003) and corrected in Evensen (2004). The code was generated by adapting the available routines on the website http://www.nersc.no/~geir/EnKF/Code/ for use with POLCOMS. At present, the forecast/analysis sequence is run through a shell script. The ensemble forecast is run sequentially, each member running with model domain decomposition.

Simulations for a high-resolution model setup were conducted in an Irish Sea model domain with a 1.8 km horizontal grid (301 x 173 gridpoints) and 32 vertical σ–levels.

Initial Conditions

An ensemble of 10 different initial conditions was obtained by adding a set of perturbations to the first guess. The perturbations were obtained by sampling (Evensen, 2004) daily model output, from the reference simulation, corresponding to the season during which the ensemble was to be started (e.g. for an ensemble simulation starting in March, the model output used to obtain the initial perturbations was that from January to April of the reference simulation).

Parameterization of Model Error

Model error was parameterized by imposing Gaussian noise onto those components of the system that were supposed to be main sources of background error.

Wind was perturbed by adding geostrophic perturbations to the corresponding reference field. The perturbations had a decorrelation time scale of 6 hours and a horizontal decorrelation scale of approximately 500 km. This length scale was estimated from that of the geopotential height fields in ECMWF analyses.

In order to account for possible errors in the turbulent closure scheme, two empirical constants were perturbed: von Karman's constant (κ) and the mixing length limit (ℓ). The actual perturbed values were different for each member but held constant throughout its integration in time.

Finally, cloud cover was perturbed with a horizontal decorrelation scale of approximately 200 km.

Model Setup and Experiments Protocol

A first experiment was performed running an ensemble of 10 members. The ensemble was started on 01/03/2001 and was run in forecast mode under the effect of wind and turbulent parameter errors. The ensemble variance of the SST field after two months of integration was concentrated on the deeper areas of the domain, presenting very fine scale structure. The ensemble was then run from that set of states for two additional months (starting 01/05/2001) with the additional contribution to model error parameterization of cloud cover perturbations. This extended the ensemble SST variance across the domain, its spatial structure distinguishing areas of the basin with different physical mechanisms (Figure 3-3(b)).

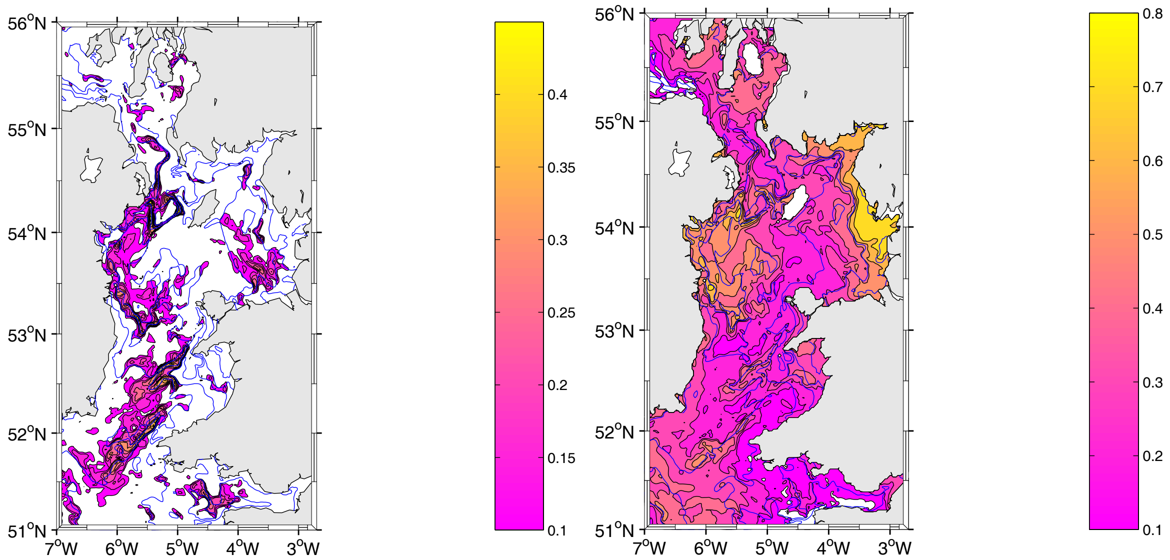

Figure 3-3: Ensemble forecast standard deviation of SST (° C) for a 10 member ensemble on 01/06/2001 (a) after three months of integration submitted to noise in the forcing wind field and the turbulent parameters, (b) with the contribution of errors in the cloud cover from the beginning of the third month. Note the difference in scales.

Assimilation of 2 km satellite SST products started on 01/07/2001, taking place at 0 hrs. every night. It was found that the ensemble spread decreased very quickly after assimilation was activated. In the analysis step, the correction introduced by the observations produced a decrease in the ensemble spread which is consistent with the EnKF formulation. However, model error parameterization did not succeed to build up enough forecast error to enable efficient assimilation of observations in subsequent analyses.

After a certain number of analysis-forecast cycles, the ensemble tended to coalesce. The time scales of forecast error build up revealed longer than the time scales on which we assimilated (Figure 3-4). This could be due to an insufficient model error parameterization, but it also brings up the question of whether we can actually account for most of model error using model sensitivity. This question remains open and will be subject to further investigation. In the mean time, an Ensemble Optimal Interpolation scheme (EnOI) scheme has been implemented[3], in order to assure efficient assimilation of observations. This scheme updates a single forecast state using an analysis scheme fed by a stationary ensemble, used to estimate the error covariances. This ensemble can change depending on the period or season.

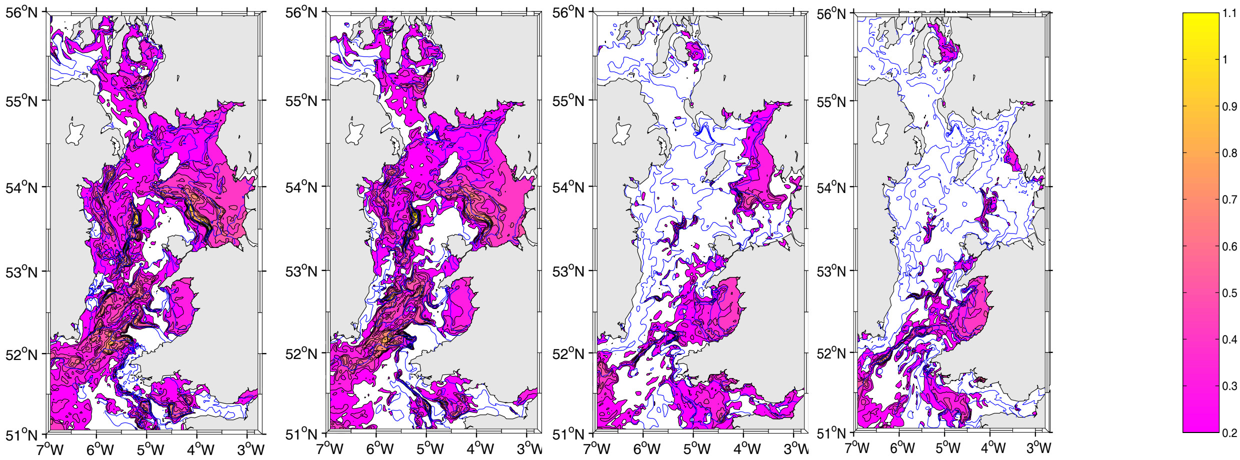

Figure 3-4: Ensemble standard deviation of forecast SST from 01/07/2001, first day of assimilation, and subsequent days. Assimilation takes place on a daily basis. Figures show the spread of the background ensemble before assimilation for each day and how the ensemble tends to coalesce. Scale is 0.2-1.1° C for all figures.

[3] This was done with the help of G. Evensen (Norsk Hydro), L. Bertino (NERSC) and François Counillon (NERSC).

3.2.3 Bibliography

Andreu-Burillo *et al*, Assimilation of Sea Surface Temperature in the POL Coastal Ocean Modelling system, *JMS accepted*, 2005.

Annan, J. D., and J. C. Hargreaves, Sea Surface Temperature assimilation for a three-dimensional baroclinic model of shelf areas, *Continental Shelf Research*, 19, 1507-15020, 1999.

Brusdal *et al*., A demonstration of ensemble-based assimilation methods with a layered OGCM from the perspective of operational ocean forecasting systems, *J. Marine. Syst*., 40-41, 253-289, 2003.

Brasseur *et al*., Assimilation of altimetric data in the mid-latitude oceans using the Singular Evolutive Extended Kalman filter with an eddy-resolving, primitive-equation model, *J. Marine Syst*., 22, 269-294, 1999.

Echevin, V., *Variabilite du niveau de la mer mesure par altimetrie dans le Golfe du Lion. Etude de l'influence de diverses observations sur la prevision de la circulation cotiere*, Ph.D. thesis, Universite Paul Sabatier, 2000.

Echevin, V., P. De Mey, and G. Evensen, Horizontal and vertical structure of the representer functions for sea surface measurements in a coastal circulation model, *J. Phys. Oceanogr*., 30, 2627-2634, 2000.

Evensen, G., Using the Extended Kalman Filter with a multilayer Quasi-Geostrophic ocean model, *J. Geophys. Res*., 97(C11), 17905-17924, 1992.

Evensen, G., Open boundary conditions for the Extended Kalman Filter with a Quasi-Geostrophic model, *J. Geophys. Res*., 98(C9), 16529-16546, 1993.

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

Evensen, G., Advanced data assimilation for strongly nonlinear dynamics, *Mon. Wea. Rev*., 125, 1342-1354, 1997

Evensen, G., and N. Fario, Solving for the generalized inverse of the Lorenz model, *J. Meteor. Soc. Japan*, 75(1B), 229-243, 1997.

Evensen, G., The Ensemble Kalman Filter: Theoretical Formulation and practical implementation, *Ocean Dynamics*, 53, 343-367, 2003.

Evensen, G., Sampling strategies and square root analysis schemes for the EnKF, *Ocean Dynamics*, 54, 539-560, 2004.

Fukumori, I., and P. Malanotte-Rizzoli, An approximate Kalman Filter for ocean data assimilation: An example with an idealized Gulf Stream model, *J. Geophys. Res*., 100(C4), 6777-6793, 1995.

Hoteit *et al*., A singular Evolutive Extended Kalman filter to assimilate real in situ data in a 1-D marine ecosystem model, *Advances in Geophysics*, 21, 389-397, 2003.

Jazwinsky, A. H., *Stochastic processes and filtering theory*, Academic Press, New York, 1970

Lermusiaux, P. J. F. *Error Subspace data assimilation methods for ocean field estimation: theory, validation and applications*, Ph.D. thesis, Harvard University, Cambridge, Mass., 1997.

Lermusiaux, P. J. F., and A. R. Robinson, Data assimilation via Error Subspace Statistical Estimation. Part I: Theory and schemes, *Mon. Wea. Rev*., 127, 1385-1407, 1999.

Lermusiaux, P. J. F. Evolving the subspace of the three-dimensional multiscale ocean variability: Massachusetts Bay, *J. Marine Syst*., 29, 385-422, 2001.

Lorenz, E. N, Deterministic non-periodic flow, *J. Atmos. Sci*., 20, 130-141, 1963a.

Lorenz, E. N, The predictability of hydrodynamic flow, *Trans, NY Acad. Sci*., Series II, 25, 409-432, 1963b.

Lorenz, E. N, A study of the predictability of a 28-variable atmospheric model, *Tellus*, 17, 321-333,1965.

Lorenz, E. N, The predictability of a flow which possesses many scales of motion, *Tellus*, 21, 289-307, 1968.

Pham *et al*., 1998, Singular Evolutive Kalman Filters for data assimilation in oceanography, *C. R. Acad. Sci. Paris, Sciences de la Terre et des planetes/Earth & Planetary Sciences*, 326, 255-260, 1998.

Robinson, A. R., P. F. J. Lermusiaux, and N. Q. Sloan, *The Sea*, vol. 10, chapter 20. DATA ASSIMILATION, pp. 541-594, 1998.

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

van Leeuwen, P. J., and G. Evensen, Data assimilation and inverse methods in terms of a probabilistic formulation, *Mon. Wea. Rev*., 124, 2898-2913, 1996.