Ecosystem Modeling and Data Assimilation of Physical-Biogeochemical processes in Shelf and Regional Areas of the Mediterranean Sea

Triantafyllou George\textsuperscript{1}, Hoteit Ibrahim\textsuperscript{2}, Korres Gerassimos\textsuperscript{1}, and Petihakis George\textsuperscript{1}

\textsuperscript{1} Hellenic Center for Marine Research P.O. Box 712 - 19013 Greece
\textsuperscript{2} Scripps Institution of Oceanography, La Jolla, CA, USA

Received 5 November 2004, revised 10 May 2005, accepted 20 May 2005
Published online 10 July 2005

Key words Data assimilation, Singular Evolutive Extended Kalman (SEEK) filter, Model error
Subject classification 93A30, 81T80

In this work, a general overview of the recent progress and the existing difficulties in the implementation of physical and biogeochemical models accompanied with advanced Kalman filtering techniques in regional and shelf areas of the Mediterranean Sea is presented. The hydrodynamics of these areas were simulated with realistic implementations of the Princeton Ocean Model (POM) while the biogeochemistry of the ecosystems was tackled with the European Regional Seas Ecosystem Model (ERSEM). On the other hand, the Singular Evolutive Extended Kalman (SEEK) filter and its ensemble-based variant called the SEIK filter was used to assimilate all physical and biogeochemical data into the models. In these filters the error statistics were parameterized by means of a suitable basis of empirical orthogonal functions (EOFs). This contribution falls into the framework of the ‘Mediterranean ocean Forecasting System: Toward Environmental Predictions (MFSTEP)’ project and synthetically presents the modeling efforts in conjunction with data assimilation techniques for the development of operational systems in regional and shelf areas of the Mediterranean Sea.

1 Introduction

One of the major challenges for the coming years in oceanography is the design of operational data assimilation systems capable to effectively support global, regional and coastal management issues of the marine environment. The development of marine monitoring and forecasting facilities plays a key role in improving our understanding of physical, chemical and biological processes. For this purpose, several international marine monitoring and forecasting projects have been developed over the past 10 – 20 years. Recently, the Mediterranean basin has drawn increasing attention due to its comparatively small size and the relevance of biophysical processes occurring in its basin. Starting in 2003, the objective of the project (Mediterranean ocean Forecasting System: Toward Environmental Predictions) MFSTEP is to demonstrate the feasibility of a Mediterranean operational system for predictions of physical and biogeochemical parameters in the whole basin and in coastal/shelf areas, for time scales of weeks to months.

Ocean data assimilation systems consists of three major components: (i) a numerical model suitable for simulating the current state of the ocean and even predict, to a limited extent, the future state of the ocean given a suitable set of initial conditions, (ii) an observation network that provides reliable information on the state of the ocean, and (iii) an assimilation scheme that efficiently combines available data with the numerical model to obtain the best possible estimate of the ocean state.

By contrast to marine ecosystem modeling, physical ocean circulation models had extraordinary development in the last 10 years. This is mostly due to the fact that the ocean circulation is governed by the Navier-Stokes equations while our knowledge of the dynamics of the marine ecosystem is still rather limited. The hydrodynamic models vary according to the approximations, from the shallow water models, passing by the quasi-geostrophic...
models, to the Primitive equations (PE) models. The last type of models resolve the full Navier-Stokes equations and therefore they are the most demanding in computing resources. These highly sophisticated models are now widely used in the operational centers due to the continuous progress in computer power. Several modeling efforts [1, 2, 3] have shown that the various hydrodynamic models implemented into the Mediterranean basin are capable of reproducing the major circulation characteristics and the water mass variability of the basin. However, there are still uncertainties in the models’ setup and the atmospheric forcing which inevitably results in imperfect modeling simulations. Strong density gradients, important seasonal to interannual changes and sharp bathymetric changes occurring into the basin make the assimilation problem into a Mediterranean Sea model a challenging effort.

Despite the launching of different international ocean observations missions in the last two decades, ocean sciences still suffer of cruel lack of data, particularly in the deep ocean. These missions greatly increased the amount of available in-situ data, which are collected by buoys, ships, drifters, etc, in the last years. The most significant amount of ocean data is now provided by the continuous satellites missions (ERS1, ERS2, TOPEX, and recently JASON) which measure with great precision the sea surface height, sea color data, and sea surface temperature. However, satellite only observes the ocean surface and one of the challenges is to design data assimilation schemes capable of efficiently propagating the surface information into the deep ocean. In the Mediterranean basin, an extended observational network has been established within the framework of the MFSTEP project [4] aiming in building up the operational capacity of the Mediterranean area. These observations will be used to improve our understanding of the ocean circulation, to optimize and to guide the numerical models toward realistic trajectories using appropriate data assimilation schemes.

Data assimilation techniques have been originally developed for meteorological applications. Their theoretical framework is now well established and two separate directions are usually followed in oceanographic applications [5], one being a Kalman filtering approach, and the other being a variational approach. Filtering methods proceed by incrementally correcting the discrepancy between observations and a model prediction based on prior information about uncertainties in the model and data. By contrast, variational approaches seek to minimize the misfit between data and model trajectory over a given period of time. To this end, the model-data misfit is reduced through adjustments of a well chosen set of control parameters. Although assimilation methods have been found to be very efficient in improving the ocean model’s behavior, their computational demands remain one of the main obstacles. Attempts related to reducing the computational time of these algorithms have received considerable attention. For instance, different variants of the Kalman filter have been proposed to reduce the dimension of the system [6, 7, 8]. An alternative direction that makes use of Monte Carlo methods to represent the error statistics by an ensemble of state vectors has been also developed [9].

Assuming that a coupled model with a given set of parameters can reproduce the observations, variational methods can be strongly sensitive to the model unresolved processes. Although model error can be in principle considered by applying a weak constraint inverse formulation [10], model errors would make the dimension of the search so long that unpreconditioned searches for a best fit would be prohibitive. Sequential methods which allow to incorporate the model error seem therefore to be more appropriate for data assimilation into marine ecosystem models.

In this work we present our efforts to develop data assimilation systems for shelf and regional areas of the Mediterranean Sea (Figure 1) using the physical POM model and the ecological ERSEM and advanced Kalman filtering techniques. Such systems will be supplied in the future with several ocean observations collected in real-time by the MFSTEP observing system products (VOS-XBT, T-S profiles from ARGO floats and gliders) and satellite products (SLA and SST). After describing the ocean models in Section 2, the assimilation schemes are presented and discussed theoretically in Section 3. A general overview of the results obtained from the model simulations and assimilation systems implemented in different areas of the Mediterranean Sea is provided in Section 4. We further discuss the progress we made so far and the problems that still need to be addressed. A general discussion including a plan for our future research directions concludes the paper in Section 5.
2 Physical-Biogeochemical Modeling of Shelf and Regional Areas of the Mediterranean Sea

The coupled physical-biogeochemical model consists of two, on-line coupled, sub-models: the Princeton Ocean Model (POM) [11], which describes the hydrodynamics of the area, and provides the physical forcing to the second submodel, the European Regional Seas Ecosystem Model (ERSEM) [12]. ERSEM describes the biogeochemical cycles and has been successfully applied in a wide variety of regimes from coastal eutrophic to oligotrophic and on a variety of spatial scales [13, 14, 15, 16].

2.1 The physical model - POM

The hydrodynamic model (POM) has been successfully implemented in a number of coastal applications like the Mediterranean Sea [17, 18, 19], the Adriatic Sea [17] the Levantine Sea [20, 21] and the Aegean Sea [22]. POM is a three-dimensional primitive equation model. It consists of prognostic equations for the horizontal momentum, potential temperature, salinity and the free surface elevation. The hydrostatic equation, the equation of state and the vertical velocity that is derived from the mass continuity equation constitute the diagnostic equations. In addition the model contains an embedded second order turbulence closure sub-model, which solves two prognostic equations, one for the kinetic energy and one for the turbulence macroscale. POM is a sigma coordinate model in that the vertical coordinate is scaled on the water column depth, while in the horizontal grid it uses curvilinear orthogonal coordinates and an Arakawa-C differencing scheme. The sigma coordinate system is a necessary attribute in dealing with significant topographical variability, and together with the free surface - an important feature for high frequency forcing - makes the model suitable for applications in coastal seas and continental margins. Time integration is performed with a split (external/internal) time step in which the barotropic and baroclinic modes are integrated separately with a leapfrog scheme with different time steps.

2.2 The biogeochemical model - ERSEM

ERSEM is a differential equation model describing both the pelagic and benthic ecosystems, and the coupling between them in terms of the significant biogeochemical processes affecting the flow of carbon, nitrogen, phosphorus and silicon [23, 24, 25, 26, 27]. The dynamics of biological functional groups are described by population processes (growth, and mortality) and physiological processes (ingestion, respiration, excretion and egestion). The biotic system is subdivided into three functional types, producers (phytoplankton), decomposers (pelagic and benthic bacteria) and consumers (zooplankton and zoobenthos). These broad functional classifications are then further divided, according to their trophic level (derived according to size classes or feeding method) to create a foodweb. The phytoplankton pool is described by four functional groups based on size and ecological
properties. These are diatoms (silicate consumers, $20 - 200 \mu m$), nanophytoplankton ($2 - 20 \mu m$), picophytoplankton ($< 2 \mu m$) and dinoflagellates ($> 20 \mu m$). The nutrient uptake is controlled by the difference between the organism’s internal nutrient pool and the external nutrient concentration [28, 29]. The microbial loop contains bacteria, heterotrophic nano-flagellates and microzooplankton. All groups in the phytoplankton and the microbial loop have dynamically varying C:N:P ratios. Bacteria act in the decomposition of detritus and can compete for nutrients with phytoplankton. Heterotrophic nano-flagellates feed on bacteria and picophytoplankton, and are grazed by microzooplankton, which also consumes diatoms and nanophytoplankton, and is grazed by mesozooplankton. A schematic diagram of the food web is given in Figure 2. The chemical dynamics of nitrogen, phosphorus, silicate and oxygen are coupled with the biologically driven carbon dynamics. The use of a foodweb that carries both carbon and nutrient dynamics allows the model to adjust to spatial and temporal variations in carbon and nutrient availability and reproduce the different types of ecosystem behaviour. Considering the significant depth of the area of application, the weak coupling between pelagic and benthic systems and the heavy computational load, a simple benthic returns model was used instead of the standard dynamical benthic model [28]. Thus the benthic-pelagic coupling is described by a simple first order benthic returns module, which includes the settling of organic detritus into the benthos and diffusional nutrient fluxes into and out of the sediment. The parameter set used in this simulation and a detailed description of the model can be found in [30].

![Fig. 2 Schematic representation of the ecosystem foodweb.](image)

### 3 Description of the Assimilation Schemes

#### 3.1 The Kalman Filter

The Kalman filter is a data assimilation technique that recursively generates an optimal analysis, in the least-squares sense, of the state of a linear system given a set of measurements [31]. The filter is very powerful in several aspects: it supports estimations of past, present, and even future system states, and it can do so even when the model and the measurements are noisy. It however requires that the system and measurement noises are
additive, white and Gaussian. Indeed, linear models interact uniquely well with Gaussian noise making filter’s 
distributions Gaussian and the calculations are easy. Note that when the noise is not Gaussian, the Kalman filter 
still provides the best linear state estimator, given only the mean and covariance matrix of noise.

Starting from an initial estimate of the state \( X^a(t_0) \) and the corresponding error covariance matrix \( P^a(t_0) \), the 
Kalman filter works as a succession of two steps: (i) forecast step using the model, and (ii) analysis step to correct 
the forecast each time new observations are available. In general, very little is known concerning the filter’s initial 
conditions. The choice of these parameters is however not very important because of the stability property of the 
Kalman filter [32]. A common choice for \( X^f(t_0) \) and \( P^f(t_0) \) is the average and the sample covariance matrix of 
a set of observed, or possibly simulated, state sequences from the system. In the following, the notation proposed 
by [33] is adopted for the presentation of the filter’s algorithm. Consider a linear physical system described by

\[
X^f(t_k) = M(t_k, t_{k-1})X^f(t_{k-1}) + \eta(t_k),
\]

where \( X^f(t) \) is a vector representing the true state at time \( t \), \( M(t, s) \) is an operator describing the system transition 
from time \( s \) to time \( t \), and \( \eta(t) \) is the system noise vector. At each time \( t_k \), the state vector is observed according to

\[
Y^o_k = H_k X^f(t_k) + \varepsilon_k,
\]

where \( H_k \) is the observational operator and \( \varepsilon_k \) is the observational noise. The noises \( \eta(t_k) \) and \( \varepsilon_k \) are assumed to 
be independent random vectors with mean zero, and covariance matrices \( Q_k \) and \( R_k \), respectively. Assuming that 
an initial analysis state \( X^a(t_{k-1}) \) and its error covariance matrix \( P^a(t_{k-1}) \) are available at time \( t_{k-1} \), the Kalman 
filter allows the construction of the next analysis vector \( X^a(t_k) \) at the time of the available new observations \( t_k \), 
together with its error covariance matrix \( P^a(t_k) \) as follows.

**1- Forecast step:** Starting from \( X^a(t_{k-1}) \), the model (1) is used to forecast the state,

\[
X^f(t_k) = M(t_k, t_{k-1})X^a(t_{k-1}).
\]

The forecast error covariance matrix is then given by

\[
P^f(t_k) = M(t_k, t_{k-1})P^a(t_{k-1})M^T(t_k, t_{k-1}) + Q_k.
\]

**2- Analysis step:** The new observation \( Y^o_k \) at time \( t_k \) is then used to correct the forecast according to the 
formula

\[
X^a(t_k) = X^f(t_k) + G_k[Y^o_k - H_k X^f(t_k)],
\]

where \( G_k \) is the so-called Kalman gain matrix,

\[
G_k = P^f(t_k)H_k^T[H_k P^f(t_k)H_k^T + R_k]^{-1}
\]

The corresponding analysis error has the covariance matrix

\[
P^a(t_k) = P^f(t_k) - G_k H_k P^f(t_k).
\]

At a first glance, the implementation of the Kalman filters seems easy. However, its use for data assimilation 
with realistic ocean models is far from being straightforward. The most obvious difficulty is the nonlinear nature 
of the ocean models. In this case the system equations are generally linearized about the current state estimate 
leading to the popular, but no longer optimal, extended Kalman (EK) filter [32]. The algorithm of the EK filter 
is identical to the algorithm of the Kalman filter; only \( M(t_k, t_{k-1}) \) and \( H_k \) are respectively replaced in (4), (6) 
and (7) by the gradient \( M(t_k, t_{k-1}) \) of \( M(t_k, t_{k-1}) \) evaluated at \( X^a(t_{k-1}) \) and the gradient \( H_k \) of \( H_k \) evaluated 
at \( X^f(t_k) \). The use of the linearized system however amounts to neglect higher-order statistical moments and 
several studies have shown that this might produce instabilities, even divergence, when implemented with strongly
nonlinear systems [34, 35]. The other difficulty is related to the huge dimension \( (n) \) of the ocean state \( (\approx 10^8) \) making the manipulation of the error covariance matrices practically impossible. Approximations are therefore unavoidable. Several simplified versions of the EK filter have been proposed to reduce its excessive computational burden [7, 6]. These sub-optimal Kalman filters basically consist of projecting the state of the system onto a low dimensional subspace. The SEEK filter and its variants are alternatives to the Kalman filter, designed for data assimilation with realistic ocean models. These filters are summarized below. The reader is referred to [36, 37, 38] for a detailed description and theoretical discussion of the filters.

3.2 The SEEK, SFEK, and SEIK Filters

The SEEK filter is a simplified square-root EK filter [39] which aims at reducing the computational burden of the EK filter by using low rank \( (r << n) \) matrices approximations of the filter’s error covariance matrices \( (P) \). More precisely, this filter is based on the decomposition \( P = LUL^T \), where \( L \) and \( U \) are \( n \times r \) and \( r \times r \) matrices, allowing to replace the evolution equations of the error covariance matrices by two less costly (in terms of storage and computing time) equations of the evolution of \( L \) and \( U \), while keeping the EK filter’s algorithm mostly unchanged. Starting from an initial low-rank \( r \) error covariance matrix, [36] showed that the filter’s error covariance matrices will always be of rank \( r \), if the model is perfect. When the model is imperfect, the model error can be projected onto the subspace spanned by the columns of \( L \) to avoid continuous increase in the rank of the error covariance matrices. As for the EK filter, the implementation of the SEEK filter with strongly nonlinear models can be problematic. To avoid the linearization of the model equations, [37] introduced the ensemble variant of the SEEK filter, called SEIK filter. This filter basically consists of using a stochastic nonlinear ensemble forecasting scheme to perform the evolution of the statistics of the forecast. Another variant of the SEEK filter that has been widely used is the Singular Fixed Extended Kalman (SFEK) filter which simply neglects the evolution of the matrix \( L \) for further computing time reduction. Several other variants of the SEEK and SEIK filters, not considered here, with different degrees of simplification of the evolution of the matrix \( L \) have been also developed by [38] in case of lack of computing resources. The SEEK filter and its variants have the same initialization and analysis steps and they only differ in the way they perform the evolution of the forecast statistics. The algorithms of the SEEK, SFEK and SEIK filters are jointly presented below to highlight their similarities as well as their differences.

3.2.1 Initialization Step

An initial analysis state \( X^n(t_0) \) and a low-rank \((r)\) error covariance matrix \( P^n(t_0) = L_0U_0L_0^T \) are needed for the initialization of the filters algorithms. Since a long sequence of observed state vectors is rarely available in oceanography, \( X^n(t_0) \) is taken as the average of a simulated sequence of model state vectors and \( P^n(t_0) \) as the low-rank approximation of the sample covariance matrix \( P_0 \) of these vectors. To obtain such an approximation, we use an Empirical Orthogonal Functions (EOFs) technique\(^1\), which provides the best approximation (Eckart-Young theorem) in the sense of mean-squares error [40]. If \( l_1, l_2, \ldots, \) the eigenvectors of \( P_0 \), called EOFs, are ordered according to their eigenvalues \( \lambda_1, \lambda_2, \ldots, \), the EOF analysis approximates \( P_0 \) to \( L_0U_0L_0^T \) where \( L_0 = [l_1 \ldots l_r] \) and \( U_0 = Diag[\lambda_1, \ldots, \lambda_r] \). It can be further seen that the error of this ‘projection’ onto the linear subspace spanned by \( l_1, \ldots, l_r \) has squared norm \( \sum_{i>r} \lambda_i \), since the squared norm of \( P_0 \) is \( Trace(P_0) = \sum_i \lambda_i \), the ratio \( \frac{\sum_{i>r} \lambda_i}{Trace(P_0)} \) represents the relative error of the approximation and can therefore be used to assess the accuracy of the approximation for the appropriate choice of \( r \).

3.2.2 Analysis Step

Assume that at time \( t_k \) of the available new observation \( Y_k^n \), a forecast state \( X^f(t_k) \) and a low-rank approximation of the error covariance matrix \( P^f(t_k) = L_kU_{k-1}L_k^T \) have been already computed. Using this approximation in the EK filters, the analysis step can be rewritten as

\[
X^n(t_k) = X^f(t_k) + L_kU_{k+1}(H_kL_k)^T[Y_k^n - H_kX^f_k],
\]

\(^1\) also known under Principal Components (PCs)
where $U$ is updated according to

$$U_k = U_k^{-1} + P_k L_k^T Q_k P_k L_k^{-1} + (H_k L_k)^T R_k^{-1} H_k L_k,$$

(9)

with $P_k = [L_k^T L_k]^{-1} L_k$ is the projection operator onto the subspace spanned by the columns of $L_k$. The corresponding filter error covariance matrix is then

$$P^a(t_k) = L_k U_k L_k^T.$$  (10)

The analysis estimate in (8) is therefore expressed as the forecast plus a linear combination of the columns of $L_k$, which therefore constitute the correction basis of the filter.

3.2.3 Forecast Step

Once the analysis state $X^a(t_k)$ and the corresponding error covariance matrix $P^a(t_k) = L_k U_k L_k^T$ have been obtained as presented above, the model (1) is used to make the forecast and to update its statistics.

**SEEK Filter:** The forecast state is computed as in (3) and the corresponding error covariance matrix is obtained by simply replacing $P^a(t_k)$ by $L_k U_k L_k^T$ in (4), which implies

$$P^f(t_{k+1}) = L_{k+1} U_k L_{k+1}^T + Q_k,$$

(11)

where the new correction basis of the filter $L_k$ is updated according to the tangent linear model,

$$L_{k+1} = M(t_{k+1}, t_k) L_k.$$  (12)

[36] showed that the columns of $L_k$ converge toward the directions of model error growth. Since the correction of the filters is only applied in the directions of $L_k$, this implies that the SEEK filter corrects the forecast in the directions for which the error was not sufficiently attenuated by the dynamics of the model.

**SFEK Filter:** Motivated by the fact that most of the estimation errors in the numerical experiments of [36] was immediately reduced after the first correction, i.e. while the evolution of the EOFs was not been yet applied, [41] proposed to keep the initial correction basis of the SEEK filter fixed in time to further reduce the implementation cost of the SEEK filter. This can be supported by the fact that the state of the ocean evolves very slowly in time, allowing the crude approximation,

$$M(t_{k+1}, t_k) = I_d.$$  (13)

The error covariance matrix is therefore parametrized by a set of EOFs which describe the dominant modes of the system’s variability [40], and is decomposed as

$$P^f(t_{k+1}) = L_0 U_k L_0^T + Q_k.$$  (14)

**SEIK Filter:** The SEIK filter uses a nonlinear ensemble forecasting to avoid the linearization of the SEEK filter. The basic idea of this filter consists of representing the statistics of the SEEK filter’s analysis $X^a(t_k)$ by an ensemble of $N$ state vectors $X^a_1(t_k), \ldots, X^a_N(t_k)$, i.e.

$$X^a(t_k) = \frac{1}{N} \sum_{i=1}^N X^a_i(t_k),$$

(15)

$$P^a(t_k) = \frac{1}{N} \sum_{i=1}^N [X^a_i(t_k) - X^a(t_k)][X^a_i(t_k) - X^a(t_k)]^T,$$

(16)

using a Monte-Carlo approach, called the second-order exact sampling scheme [37]. Since $P^a(t_k)$ has a low-rank $r$, one can consider the lowest number of members, namely $N = r + 1$, to represent $P^a(t_k)$. Once such an ensemble is generated, the model (1) is used to integrate the members forward in time to obtain the so-called forecast ensemble members $X^f_1(t_{k+1}), \ldots, X^f_{r+1}(t_{k+1})$. The forecast state and the corresponding error covariance matrix are then taken as the average and the sample covariance matrix of $X^f(t_{k+1})$ (plus the model error.
covariance matrix), respectively. It can then be seen that $P^f(t_k)$ can be represented in the form $L_{k+1}^T \hat{U}_k L_{k+1}^T$, with

$$L_{k+1} = [X^f_1(t_{k+1}) \cdots X^f_r(t_{k+1})] \cdot T$$

and

$$\hat{U}_k = [(r+1)T^T]^{-1}$$

and $T$ is a $(r+1) \times r$ full rank matrix with zero column sums. The ensemble members can be also used to avoid the linearization of the observational operator $H_k$ by replacing it with a “linear interpolation” around the $X^f(t_k)$. This amounts to replace $H_k L_k$ in (8) and (9) by the matrix $H_k X^f_1(t_k) \cdots H_k X^f_r(t_k)]$.

Finally, the drawing of the ensemble members is performed as follows. Let $C_k$ be the Cholesky decomposition of $U_k^{-1}$, one can write

$$P^a(t_k-1) = L_{k-1}(C_{k-1}^{-1})^T \Omega_{k-1}^{-1} C_{k-1}^{-1} L_{k-1}^T$$

for any $(r+1) \times r$ matrix $\Omega_k$ with orthonormal columns and zero column sums. The latter matrix is drawn randomly using the procedure described in [37]. The ensemble members can be then taken as

$$X^a_i(t_k-1) = X^a(t_k-1) + \sqrt{r+1} L_{k-1}(\Omega_{k-1,i} C_{k-1}^{-1})^T,$$

where $\Omega_{k-1,i}$ denotes the $i^{th}$ row of $\Omega_{k-1}$.

The use of the ensemble members is very similar to the ensemble Kalman (EnK) filter proposed by [9]. However, the SEIK filter works under the assumption of low rank $(r)$ error covariance matrices allowing the use of the smallest number of ensemble members $(r+1)$, hence possesses an advantage in term of computing cost. The main difference between the two filters is in the generation of the analysis ensemble: by correcting each member of the background ensemble using perturbed observations in the EnK filter, or by sampling according to the analysis state and the corresponding error covariance matrix in the SEIK filter. [42] compared the performances of the two filters and found that the SEIK filter generally outperforms the EnK filter since the perturbed observations tends to increase the sampling error of the filter.

The equations formulating the analysis and forecast error covariance matrices are not needed for the filters algorithms, but they have been presented to introduce the reader in the assimilation scheme. The SEEK and SEIK filters therefore drastically reduce the computational burden of the EnK filter. More precisely, the SEEK filter requires about $r+1$ times the computing cost of the numerical integration of the model to update its correction basis by the equation (12). The cost of the remaining operations is negligible. Since the same number of model integrations is also needed for the integration of the $(r+1)$ ensemble members of the SEIK filter, the computing cost of these two filters is mainly the same. The SFEK filter requires one model integration only to make the forecast and can be therefore as much as $r+1$ times faster than the SEEK and the SEIK filter, thus providing a very efficient tool to validate the assimilation system and to tune the filters parameters before using more expensive filters as the SEEK and SEIK filters.

### 3.2.4 Model and Observational Errors

The observational and model error covariance matrices $R$ and $Q$ need to be specified for the algorithms of the filters. These matrices are generally unknown. This is particularly true for $Q$ since its specification requires the estimation of a huge number of parameters, exactly $n \times n$ parameters where $n$ is the dimension of the system, and this is obviously not possible. In our assimilation experiments, a compensation technique is used to replace the term $P^T_{k|k} Q_k P_{k|k}$ in (9) by a forgetting factor $\rho$ which artificially amplifies the background error covariance matrix. This leads to the new update formula for the matrix $U$,

$$U_k^{-1} = 1/\rho U_{k-1}^{-1} + (H_k L_k)^T R_k^{-1} H_k L_k.$$  

The use of this factor can be also beneficial to avoid the common problem of simplified and ensemble Kalman filters to diverge because of the underestimation of the error covariance matrices (by low rank matrices) and/or to attenuate the propagation of the filters deficiencies in time, while giving more weight to recent observations.

Concerning the observational errors, it is natural to assume that the observations are independent Gaussian variables of zero mean and variance $\sigma^2$. The use of a fixed factor $\sigma^2$ is not recommended since this generally leads to under estimating the filter error, especially when the observation error is small [37]. Here, we adapt the
value of this factor to reflect the true filter errors. Specifically, we estimate \( \sigma^2 \) at time \( t_k \) by the quantity \( e_k^2 / n_k^o \) where \( e_k^2 \) and \( n_k^o \) are updated recursively according to

\[
e_k^2 = \rho e_{k-1}^2 + ||Y_k^o - H_k X^a_{t_k-1}||^2, \\
n_k^o = \rho n_{k-1}^o + \{ \text{dimension of } Y_k^o - r \}.
\]

(21)

(22)

This estimator actually estimates the spatial average of the squared prediction errors and its use would somewhat over-estimate the filter error covariance matrix, which is obviously better than to under-estimate it.

4 Modeling and Assimilation Results

4.1 Ocean Modelling

4.1.1 Mediterranean Sea (POM)

The Mediterranean Sea is a semi-enclosed concentration basin (evaporation exceeds precipitation and river runoff) rich of basin, sub-basin and mesoscale features interacting in order to compose the general circulation picture of the basin. The basin is subject to a strong seasonal cycle due to relevant changes of the atmospheric forcing [1]. Temperature gradients in the north-south direction can be as large as the seasonal changes while strong salinity contrasts is the standard due to the Atlantic and Black Sea fresh waters entering into the basin, rivers runoff and the high evaporation rates occurring over the basin at the same time. Wind, thermohaline and source forcing (Gibraltar and Dardanelles Strait inflow/outflow) are important driving mechanisms of the Mediterranean circulation. The atmospheric circulation over the basin is dominated by a general westerly flow which interacts with the rich and complicated orography of the Mediterranean region. A typical example of this interaction is the Mistral and the Bora wind patterns. During summer the tropical African circulation regimes expand over the Eastern Mediterranean giving rise to subsidence [43]. These climatic conditions make the marine system highly seasonal and prone to respond to large-scale interannual variability of the atmospheric forcing over the area.

The thermohaline circulation of the basin is of great importance as it can be considered as a direct analogue of the Atlantic Ocean conveyor belt. Intense air-sea interactions and mixing transform the inflowing Atlantic water into the T-S characteristics of the outflowing Levantine Intermediate water (LIW). The relatively warm and fresh Atlantic water that enters the basin through the upper layer of the Strait of Gibraltar flows eastwards (Algerian Current, Ionian-Atlantic Stream, Mid Mediterranean Jet) and progressively becomes saltier in its route to the Levantine at the easternmost part of the basin where during wintertime it sinks at intermediate depths to become the well known Levantine Intermediate water. This intermediate water mass (LIW) can be traced all the way to the Strait of Gibraltar but also branches into the Aegean, the Adriatic Sea and the Gulf of Lions where preconditions the deep water formation processes occurring into the basin. Previous modeling efforts [17, 1, 2, 3, 18] shown that the various models implemented into the basin are capable of reproducing the major circulation characteristics and water mass variability of the basin. However, there are still uncertainties in the models setup and the atmospheric forcing used to drive them which inevitably result in incorrect forecast.

The Mediterranean hydrodynamic model used in [44, 45] is based on POM code and has been implemented with a spatial resolution of \( 1/4^\circ \times 1/4^\circ \) and 25 sigma levels in the vertical in order to hindcast the Mediterranean general circulation for the period 1979 - 1993. In both works, the model was initialized after 19 years of climatological integration and was then forced with the 1979 - 1994 ECMWF \( 1^\circ \times 1^\circ \) re-analysis 6-hour atmospheric data (wind velocity, air temperature and relative humidity) and cloud cover and precipitation data taken from COADS monthly \( 1^\circ \times 1^\circ \) fields for the same period.

4.1.2 Cretan Sea (POM-ERSEM)

The 1D biogeochemical model was applied in three stations along a South to North transect in the Cretan Sea [46, 47]. The aim was to explore the underlying dynamics of this oligotrophic system with emphasis on the cycling of carbon. Additionally the limitations of this modeling effort were analyzed and described.

The results presented in these papers were in good agreement with the observed in situ data in all three stations following the increasing oligotrophy and lower carbon fluxes trend with increasing distance from the coast. Interesting conclusions arise from the simulation analysis. Low nutrient concentrations and in particular
phosphorus, are responsible for limiting primary production and hence low levels of phytoplankton biomass. Only during winter when mixing is significant the system changes from phosphorus to nitrogen limitation, on rather short time scales (week). At all stations the ecosystem was found to be under hydrodynamic control, with mixing events in winter and early-spring significantly influencing the biology by transporting nutrients from below the thermocline into the euphotic zone and passing organic matter either to the benthos or into deep waters. The model stratification which starts in mid-spring and is sustained until October (reaching 60 - 80m) acts as a trap for nutrients and dissolved organic compounds creating thus a deep zone of biologic activity.

A significant outcome of this paper was that the primary production rate was coupled to maximum biomass in contrast to the bacterial production rate which was coupled to the DOC concentration exhibiting a time lag of approximately one month due to the slow growth of nutrient limited bacteria. Primary production rate decreases moving away from the coast and at higher levels compared to bacterial production. Increased primary production rates are exhibited during winter - spring because intense mixing enhances the supply of nutrients to the euphotic zone triggering phytoplankton growth. Production rates are sustained during summer due to nutrient recycling. However the rates decrease with time as sedimenting organic material is lost to the deeper layers. Phytoplankton biomass was maximum during winter - spring associated with mixing and new production while bacterial biomass was lower at all three stations with only exception the outer station during August.

Discussing the limitations of the particular approach, the authors argue that due to the 1D nature of the model vertical mixing processes below the thermocline are not sufficiently resolved, hence remineralised nutrients from the seabed cannot reach the surface over appropriate timescales. Thus the model exhibits a very tight fit to the measured mean concentrations at the shallow station, in contrast with the outer stations where simulations fall short. Additionally the Cretan Sea circulation is very complex with a number of transient or recurrent eddies interconnected by jets without defined time scales. Such cyclonic and anticyclonic eddies and the sporadic presence of TMW (Transient Mediterranean Waters) may act as nutrient source to the euphotic zone or as sinks for organic material to the deeper layers.

Having analysed the key features of the Cretan Sea system through the application of a 1D biogeochemical model, the next logical step was the integration of the modeling effort to a fully coupled 3D biogeochemical model capable of describing the observed spatial and temporal variability [30].

A common problem with 3D applications is the validation of the model results as the significant amount of data produced is difficult to be compared with the usually scarce in space and in time field measurements. To overcome this problem the authors separated the water column into two layers, the upper (0 - 150 m) where all biological activity takes place and the lower (150 - bottom) characterised by more or less stable conditions and hence reduced variability. The simulations were validated with a cost function which proved to be an important mathematical tool for the comparison of model results with \textit{in situ} data.

At the coastal areas, the model was very close to the reality for most of the variables in contrast with the outer areas where the cost function score was rather poor at the bottom layer. This was due to the aperiodic presence of water masses with distinct characteristics (TMW, LIW) penetrating the area and significantly affecting the concentrations of nutrients. Additionally throughout this simulation study, a double gyre system consisting of an anticyclone to the west and a cyclone to the east interconnected by meandering currents was persistent in the region. An analogous picture was obtained in an East - West transect at the outer area of the simulation grid, where the majority of model results was very good at the upper layer, but falling significantly at the deeper layer. Although from the biological point of view it is the top layer which is important, in this particular system the two layers are not completely uncoupled as deep mixing events in winter can penetrate deeper than 150m.

As the authors suggest, the biology of the Cretan sea is largely governed by the complex hydrodynamical patterns and in particular by the gyral dipole, with chlorophyll concentrations closely following the circulation patterns (Figure 3). The cyclonic circulation to the north of the central and eastern part of Crete is prominent both in March and August while the anticyclonic circulation at the north central part decreases significantly during summer. This pattern results in areas of increased production around the cyclone and very low production at the centre of the anticyclone.

Measurements have shown that the Cretan Sea is oligotrophic with low annual productivity and maximum rates between late winter and early spring due to intense mixing and subsequent supply of nutrients to the photic zone triggering production [48]. During this period highest model values of primary productivity were found between the two main gyral systems while lower concentrations were as expected at the centre of the anticyclone.
Primary and bacterial production decreased with increasing distance from the coast, while increased rates were retained during summer due to activity below the thermocline.

An interesting outcome of this work was the tight coupling between primary and bacterial production as indicated by the temporal evolution of mean values. The annual integrated values of primary and bacterial production exhibit three areas of high production, the west part of Crete, the eastern part at latitude 25.8E and the north part at latitude 25.4E. The increasing oligotrophy toward the east is exhibited during all four seasons with the west part having twice as high rates. This pattern is the outcome of the prevailed flow regime where waters are pushed to the west with a distinct cyclone at the eastern part.

It is worth noting that the existence of a numerical model that efficiently describes the spatial and temporal variability of the Cretan Sea ecosystem establishes the numerical basis for the development of a forecasting system. Such a system will be capable of supporting coastal zone management issues through the incorporation of numerical models, observational data and assimilation techniques (see Section 4.2.2).

4.1.3 Pagasitikos Gulf (POM-ERSEM)

Pagasitikos Gulf is a semi enclosed rather shallow (mean depth 69m) system, communicating with the Aegean Sea through the narrow (5.5 km) and relatively deep (80m) Trikeri channel. An important characteristic is that it receives considerable quantities of rural, industrial and agricultural effluents, which significantly affect the ecosystem structure and functioning often leading in the appearance of extensive algal blooms. In an attempt to reveal the governing dynamics in the gulf and the mechanism behind algal blooms, a complex ecosystem model was used [49, 50, 15, 16]. Model results were very close to reality at all areas and at all depths with the only exception the highly variable channel area, which is affected by the complex hydrological conditions created by the contribution of the Aegean Sea, the North Evoikos and Pagasitikos Gulfs.

A number of interesting outcomes were discussed in these modelling efforts. An important feature, the significant increase of nutrient concentrations with increasing depth was explored. The authors argue that this characteristic is partly attributed to the formation of a seasonal thermocline dividing the water column into two layers for an extended period of time, separating thus the coastal shallow areas from the central deep areas, and partly to the release of nutrients through the biodegradation of organic matter from the sediment. Additionally the light winds blowing in the area prohibit mixing and transport of nutrients in the euphotic zone, while the presence of an almost dominant anticyclone at the central part (deeper part of gulf) creates a deposition zone were organic material is concentrated. As a result higher nutrient concentrations were simulated at the deeper areas of the gulf. According to the 3D applications the coastal areas, and particularly the north part exhibit characteristics of a mesotrophic system with elevated chlorophyll concentrations and increased production rates. Their shallow depth in conjunction with the dominant big phytoplanktonic cells result in the fast sedimentation
and subsequent biodegradation of organic matter by benthic organisms, while the released nutrients are rapidly consumed by primary producers and bacteria in the carbon fixation process. On the opposite the central gulf area is characterised by small phytoplankton cells with bacteria playing a significant role in the recycling of nutrients. During stratification periods, bacteria strongly compete with phytoplankton for nutrients while when the water column is mixed during winter, nutrients are transported into the euphotic zone, triggering primary production. It is during this period that organic matter is passed to higher trophic layers and not recycled exclusively inside the microbial food web.

The significant variability of primary productivity as produced by the model, indicates that the ecosystem of Pagasitikos gulf although it exhibits oligotrophic characteristics possesses clear tendencies towards eutrophic conditions. In summer, elevated temperatures, adequate light conditions and shallow depths at the north part of the gulf in conjunction with high production rates can cause significant phytoplanktonic blooms. The coastal system and especially the north part was characterized as particularly sensitive to nutrient enrichment as it receives significant quantities of enriched waters, initiating harmful algal blooms.

4.2 Assimilation results

4.2.1 Assimilation of physical variables into POM hydrodynamic model

In [44, 45] the SEEK, SFEK and SEIK filters were implemented in the hydrodynamic models of the Mediterranean basin and the Pagasitikos Gulf. These studies aimed at investigating the performance and the capability of the filters in low (Mediterranean) and high (Pagasitikos) spatial resolution primitive equations models under a realistic setup. This was mainly achieved by performing identical twin experiments with both models and by assimilating real altimetric and surface temperature data into Mediterranean model. Twin experiments is a common approach in oceanography in which the observations are extracted from the model outputs (perturbed with some random noise). They will provide the best way of validating and inter-comparing assimilation schemes since all inputs are known by design. Large part of this work was dedicated to sensitivity analysis dealing with the rank approximation of the error covariance matrix, the state vector composition and the value of the forgetting factor in order to compare the filters’ performances as well as to determine the best possible parameters for the implementation of the filters.

In the case of the low resolution Mediterranean model, the twin experiments showed a similar skill for the SEEK and SEIK filters. 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 4. 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. 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 sea surface height (SSH). In these experiments the SFEK filter was found to provide an acceptable level of performance considering its time-invariance assumption of the correction basis.

An interesting issue that was also investigated with the low resolution Mediterranean model, concerns the saturation limit of the three filters. The saturation limit of a filter is practically the maximum number of directions of correction beyond which no further improvement in the filter’s performance is achieved. Moreover, [7, 51] found that increasing the number of modes in the error covariance matrix approximation can sometimes deteriorate the performance of the filter. The saturation limit of the SFEK filter was found to be approximately equal to 50. Covariance matrix approximations using a larger number of modes, were found to degrade the filter’s performance. However this finding was not relevant for the cases of the SEEK and SEIK filters and the use of higher ranks (than 50) presented a quasi significant improvement for both filters performances. For instance, the SEEK filters results were overall improved by 9% when 70 directions were used and this number was almost double (≈ 17%) for the SEIK filter. The limitation of these filters was not seen until the use of more than 100 correction directions. More precisely, the SEEK filter definitely stagnated after this number. By contrast, although the SEIK filter didn’t show any stagnation, the improvement in its performance was not significant when more than 100 directions were used. These results are not conflicting and they can be supported as follows. The fixed EOFs correction directions of the SFEK filter are only optimal in a time-mean sense and over the period on which they
have been computed. The last EOFs which generally represent fine-scale information can be therefore irrelevant during the assimilation period. In the case of the SEEK filter, the evolution of the last EOFs was not properly performed with the tangent linear model which is assumed to be more sensitive to fine-scale variabilities. Finally, the use of a Monte-Carlo technique in the SEIK was shown to efficiently attenuate the bad influence of the fine-scales by evolving statistically averaged combinations of correction directions weighted by their relevance (i.e. estimated variance).

![Graph showing the evolution of the global relative error for the SFEK, SEEK and SEIK filters for the Mediterranean Sea model (twin experiments)](image)

**Fig. 4** Evolution of the global relative error for the SFEK, SEEK and SEIK filters for the Mediterranean Sea model (twin experiments)

The intercompared performances of the SEEK and SEIK were also tested against real observations, where weekly altimetric sea level anomaly (SLA) and sea surface (SST) data were assimilated into the Mediterranean model for one year period (1993). The merged Topex/Poseidon & ERS SLA data used for that study, were processed by the Aviso altimetry group and mapped on a 1/3° × 1/3° Mercator grid using a space-time objective analysis method which accounts for long wavelength errors [52]. The SST data were analyzed by [53] from different observation missions. In these experiments, the rank of the error covariance matrices was set to 50 and the observational error was considered to be equal to 3cm. RMS differences between the SLA observations and the model SLA predictions are shown in Figure 5 for the model free-run (without assimilation) and the filters analysis. It can be seen that the model free run presents an annual average RMS difference of 5.9cm with peaks reaching approximately 9.5cm at the beginning of February 1993 and at the end of September 1993. The SEEK filter was able to reduce the average RMS difference by 2.4cm while the SEIK filter reduced the difference by almost 3cm reaching a value even lower than the specified observational error. When SST data were assimilated together with SLA data, the model SST data was greatly improved. By contrast, the filters SLA analysis was slightly degraded comparing with the assimilation of SLA data only, as the filter try to balance between the assimilated observations according to their weights (errors).

Pagasitikos Gulf, presents a totally different situation with respect to the Mediterranean case due to its strongly non-linear circulation characteristics. Pagasitikos Gulf is a very shallow coastal area (maximum depth equals to 90m) where the circulation picture is mainly set by the non-linear interactions and instabilities of a rich mesoscale
eddy field developing in the central part of the basin. In fact, the turbulent kinetic energy (as estimated by the model) of this coastal area undergoes a seasonal cycle to which quasi-chaotic fluctuations are superimposed. In this respect, Pagasitikos presents a highly non-linear system. The same conclusion can be drawn by estimating the norm of the second derivative with respect to the initial conditions (which is a measure of model non-linearity) for both cases (the second derivative can be approximated by taking the positively, the negatively perturbed and the unperturbed runs).

SEEK and SEIK filters were both tested by [45] in this non-linear regime using the twin experiments concept. In these experiments the error covariance matrix was approximated with 35 directions of correction while a forgetting factor of 0.65 was used. The global relative error for the two filters is shown in Figure 6 for one year integration period (1986) during which temperature and salinity profiles were assimilated into the model every 5 days. SEEK filter shows an unstable behaviour (sudden bursts of large error) with a tendency of the analysis error to increase over time. Such behaviour suggests that the linearized model approximation to evolve the filter statistics is inadequate for highly non-linear systems even if the linearization of the model is carried out over shorter time periods than 5 days. On the other hand, the ensemble representation of the error covariance matrix used by the SEIK filter, results in a more stable behaviour with respect to the SEEK filter and an improved overall performance. However, one can clearly see that the ensemble-based Kalman filtering definitely needs improvements in order to provide better estimates of the ocean state for highly non-linear systems.

4.2.2 Assimilation of biogeochemical observations into the Cretan Sea ecosystem model

Based on the assumption of perfect physical model, the singular evolutive extended Kalman (SEEK) filter was applied to the 1D marine ecosystem model of the Cretan Sea, presented in 4.1.2, for the assimilation of real *in-situ* data [54, 55]. Two major numerical experiments were demonstrated, firstly by performing twin experiments with simulated data and secondly by assimilating real *in-situ* data.
In the simpler context of twin experiments, the filter clearly reduced the inconsistency between the model and the data and efficiently propagated the information on non-observed variables. The importance of these papers was however mostly related to the assimilation of high frequency in-situ data collected during the MFSPPP project at the north Crete during 30th January 2000 to 20th April 2001. The parameters measured on a 3-hourly basis were oxygen, nitrate and chlorophyll-a concentrations at 40, 65, 90 and 115m as described in [56]. For the experiments, in-situ measurements of oxygen and nitrate were assimilated, while chlorophyll data were used as independent measurements to validate the assimilation system. The filter successfully improved the model behaviour in the upper part of the water column with respect to reference simulation run of the derived chlorophyll variable (Figure 7). The improvement of the chlorophyll in the 115m was less evident and several problems arise comparing with the simplistic twin experiment approach: the authors report that part of this problem comes from the inconsistency between the real data and EOFs, since the latter was computed from the model which was not very successful to simulate chlorophyll concentrations at the lower part of the euphotic zone. In these experiments, the role of the forgetting factor \( \rho \) in attenuating the model errors was made quite clear; showing that a small \( \rho \) is needed when the model behaves badly (i.e. when the model error is large). The use of a more realistic (adaptive) model error covariance matrix is however expected to improve performance of the assimilation system. It is also believed that the performance of the model, and therefore of the assimilation system, will be improved if the C:CHL ratio is allowed to vary with depth. A convenient way to specify the values of these parameters which are badly known is to estimate them by including them in the filters state vector.

In the above papers, the ‘directions of correction’ of the filter were kept fixed in order to save computational cost. As a next step the evolution of these directions was carried out to follow changes in the model dynamics with the implementation of the SEIK filter in the 3D ecosystem model of the Cretan Sea (see Section 4.1.2) [57]. Following a twin experiments approach, the SEIK filter assimilated nitrate, phosphate, silicate, and ammonia pseudo-data sampled according to available 23 stations into a complex 3D coupled physical-biogeochemical model. This is consistent with the multivariate character of the error sub-space, which contains the dominant
modes of the model’s variability and therefore the cross correlations between different model variables. These experiments confirmed the usefulness of the evolution of the filter’s correction directions while showing a clearly better performance of the SEIK filter compared with the SFEK filter results. In order to improve the representativeness of the local variability in the filters correction directions and also to limit the impact of observations on distant variables, a so-called local EOFs [58] were recently computed by decomposing the Cretan Sea domain onto vertical as well as horizontal domains. The use of such local directions was shown to improve the filters performances. This was particularly true for the vertical local EOFs revealing strong sensitivities of the ecosystem models to ‘bad correlations’ in the vertical direction. The assimilation of real ocean color data into this model will be considered in the near future.

5 Discussion

The different assimilation systems developed in shelf and regional areas of the Mediterranean Sea clearly demonstrate the effectiveness of the Kalman filters in improving the consistency between the physical-biogeochemical ocean models and available observations. The numerical twin experiments were also very useful for assessing the impact of the filters on non-observed variables, showing that the filters provide good estimations of these variables too. This is in agreement with the multivariate character of the filters error sub-spaces which resume the cross correlations among different model variables. It has been also found that the performances of the SEEK and SEIK filters are quite similar when applied to weakly nonlinear systems. However, the SEIK filter was definitely superior when high resolution ocean model was considered. Indeed, the behavior of the SEEK filter was shown to significantly degrade because of the appearance of strong model nonlinearities while the SEIK filter remained remarkably more stable due to the use of a nonlinear ensemble technique to represent its error covariance matrices. The analysis step of the SEIK filter was however showing some weaknesses since it is based on the approximated assumption of prior Gaussian distribution. The use of the analysis step of the fully optimal nonlinear filter is expected to improve the analysis step. Since the latter filter is very demanding in computer resources, approximations will be inevitable. A work in this sense will be reported in the near future.

Assimilation of real data revealed a clear improvement in the model’s behavior with respect to the model free-run without assimilation. Particularly, the filters were remarkably efficient in bringing the physical model into agreement with the available observations. It was however noticed that the improvement of the ecosystem models was not as significant as for the physics, revealing the importance of incorporating a more realistic model error covariance matrix in the filters algorithms. This problem is very difficult because of the scarcity of real observations making a full comparison between the models and the data rather limited.

So far the filters were implemented independently with the physical and the ecological model. This is an important step to better understand the impact of the filters on both systems and to help understanding the problems related to the special characteristics of each model. The implementation of the filters to jointly assimilate physical and ecological observations is currently under development. Several difficulties were encountered related to the
significant computational cost involved with the estimation of the state variables for both systems. Such a full assimilation system is however expected to improve the assimilation results for the ecosystem variables, since the estimation of the physical variables would decrease the model error of the ecosystem model.

Kalman filtering techniques provide a very promising tool for efficient assimilation of observations into realistic physical–biogeochemical ocean models as well as to increase the predictive capability of these models. It is clear that special care should be taken in the implementation of these assimilation schemes, and sensitivity experiments are suggested for the proper configuration of the filters algorithms.

**Acknowledgements**

This work has been supported by Mediterranean ocean Forecasting System: Toward Environmental Predictions (MFSTEP) project.

**References**


G. Triantafyllou, G. Petihakis, J.I. Allen, and A. Tselepides. Assessing the performance of the cretan sea ecosystem model with the use of the extended kalman filter with a multilayer quasi-geostrophic ocean model.

G. Korres, I. Hoteit, and G. Triantafyllou. Data assimilation into a princeton ocean model of the mediterranean sea using ersem.

R.A. Varela, A. Cruzardo, and J.E. Gabaldon. Modeling the primary production in the north sea using ersem.


I. Hoteit, G. Korres, and G. Triantafyllou. Comparison of extended and ensemble based kalman filters with low and high resolution primitive ocean models.


