A data assimilation tool for the Pagasitikos Gulf ecosystem dynamics: Methods and benefits

G. Korres \( ^{a,\ast} \), G. Triantafyllou \( ^{a} \), G. Petihakis \( ^{a} \), D.E. Raitsos \( ^{a} \), I. Hoteit \( ^{b} \), A. Pollani \( ^{a} \), S. Colella \( ^{c} \), K. Tsiaras \( ^{a} \)

\( ^{a} \) Hellenic Centre for Marine Research, P.O. Box 712, 19013, Anavyssos, Greece
\( ^{b} \) King Abdullah University of Science and Technology, Thuwai 23955-6900, Saudi Arabia
\( ^{c} \) Istituto di Scienze dell’Atmosfera e del Clima, CNR, Italy

1. Introduction

Pagasitikos is a semi-enclosed gulf situated on the western part of Aegean Sea north of the island of Evia connected at the south with the Aegean Sea through the 5.5 km wide, narrow channel of Trikeri. The predominant weak winds of the area result in small to moderate water currents while renewal occurs mainly through the deep-water layer of the Trikeri channel. During winter months the water mass of Pagasitikos is fairly mixed, forming a two-layer thermocline which remains for the rest of the year, with the exception of August that three layers are observed. Inflow of fresh waters in the areas of Volos and Almyros observed during winter and spring adds to the complexity of the system. The basin is highly influenced both by anthropogenic activities (inflow of nutrients at the north and west parts) as well as by water exchange through the Trikeri channel resulting in the development of functional sub-areas within the gulf. Thus the inner part is characterized by eutrophic conditions with sporadic formation of harmful algal blooms while the central part acts as a buffer with mesotrophic characteristics influenced by the oligotrophic outer area. Pagasitikos is a rather sensitive ecosystem due to its semi-enclosed nature and the shallow depths. The human activity in the coastal areas is not significant with agricultural farming being the major occupation. However during the last years there has been a shift towards intensive production of cereal and cotton with the major occupation. However during the last years there has been a shift towards intensive production of cereal and cotton with the major occupation. However during the last years there has been a shift towards intensive production of cereal and cotton with the major occupation. However during the last years there has been a shift towards intensive production of cereal and cotton with the major occupation. However during the last years there has been a shift towards intensive production of cereal and cotton with the major occupation.
Understanding and predicting the dynamics of the marine ecosystem is a fundamental issue for the exploration, management and protection of coastal areas. Although the study of the underlying processes is rather challenging for the scientific community this often proves to be a difficult task due to the significant complexity of the system. Especially coastal areas are characterized by a high degree of complexity requiring intensive long-term studies capable to reveal relations between the numerous variables. In the years to come, one of the major challenges 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.

The system of Pagasitikos has been studied since 1975 (e.g. Friligos, 1987; Friligos and Gotsis-Skretas, 1987; Friligos et al., 1990; Koliou-Mitsou, 2000). From the results of these projects it has been found that the water mass is homogeneous during winter in contrast with the rest of the year when two layers of different salinity and density are formed. This is caused mainly by the inflowing and warmer waters from the Aegean Sea. During August a third layer is observed, thus separating the water column into three layers. In winter, under certain environmental conditions, strong vertical mixing takes place resulting in the homogenization of the water column, although this is not a periodic phenomenon. The water renewal takes place along the Trikeri channel in the deep layers close to the bottom with waters entering the gulf from the east side and outflowing from the west side of the channel, a situation reversed in the surface layers. In Petihakis et al. (2005, 2012-this issue-b) the circulation fields and the development of water masses in the Pagasitikos Gulf were explored using in situ data and a high resolution hydrodynamic model based on the Princeton Ocean Model (POM). In these papers the capability of the high-resolution hydrodynamic model to describe the phenomenology of Pagasitikos was investigated. In a companion paper, Petihakis et al. (2012-this issue-a) investigated the interactions between the physical and biogeochemical systems in the Pagasitikos Gulf by coupling advanced hydrodynamic and ecological models. The simulation system comprises of two online coupled sub-models: a three-dimensional hydrodynamic model based on POM model and an ecological model adapted from the European Regional Seas Ecosystem Model (ERSEM) for the particular ecosystem. A cost function is used for the validation of model results with field data. Comparisons of simulation results with direct measurements show good agreement, suggesting that the model can reproduce fairly well the general circulation of the area. Furthermore, it helps to better understand the complicated hydrological structure and circulation patterns of the Pagasitikos Gulf and illustrates the role of the physical processes in determining the evolution and variability of the ecosystem.

Despite the important development in the modeling of the Pagasitikos system, accurate monitoring and forecasting of the circulation in this area may still greatly benefit from the assimilation of data into the models. This approach, referred to as data assimilation, is now widely used in meteorology and oceanography. It basically provides a framework for adjusting the model behavior on the basis of available observations, taking into account the error characteristics of both the model and the observations. Recently, the use of data assimilation to complement ecosystem modeling efforts gained widespread attention (e.g. Triantafyllou et al., 2003b; Hoteit et al., 2003; Allen et al., 2003; Natvik and Evensen, 2003; Hoteit et al., 2004; Hoteit et al., 2005a; Triantafyllou et al., 2005; Triantafyllou et al., 2007). Assimilation methods have been historically classified into two categories: sequential approach and variational approach (Ghil and Malanotte-Rizzioli, 1991). Sequential methods proceed by incrementally correcting the discrepancy between observations and a model prediction based on prior information about uncertainties in the model and data. Variational methods seek to minimize the misfit between data and model trajectory over a given period of time through the adjustments of a well chosen set of control parameters.

The Singular Evolutive Extended Kalman (SEEK) filter is a sequential data assimilation scheme based on the extended Kalman (EK) filter (Pham et al., 1997). The filter operates with low-rank error covariance matrices to avoid the associated prohibitive computational cost in the EK filter (Cane et al., 1996; Fukumori and Malanotte-Rizzioli, 1995). This leads to adjusting the model forecast in the directions of error growth only. The SEEK filter further supports different degrees of simplifications in the evolution of its ‘correction directions’ at a minimal loss of performance (Hoteit et al., 2002; Hoteit et al., 2005b). The SEEK filter and its variants are particularly popular among the ecosystem modeling community and were successfully implemented in several marine ecosystem data assimilation studies. Carmillet et al. (2001) used a simplified variant of the SEEK filter with an invariant set of Empirical Orthogonal Functions (EOFs) correction directions, called Singular Fixed basis Extended Kalman filter — SFEK, to assimilate pseudo-ocean color data into a 3-D physical–biochemical model of the North Atlantic Ocean. Hoteit et al. (2003) used the same filter with a 1-D complex ecosystem model of the Cretan Sea assimilating real observations of oxygen and nitrate and validating the filter with independent chlorophyll data. Hoteit et al. (2005a) successfully tested a variant of the SEEK filter with partially local EOFs correction directions in a complex three dimensional ecosystem model of the Cretan Sea. The reader is referred to Triantafyllou et al. (2005) for a review on the implementation of the SEEK filter and its variants in shelf and regional areas of the Mediterranean Sea. The SEIK filter is a square root ensemble Kalman filter (Hoteit et al., 2005b) and has strong similarities with the ensemble Kalman (EnK) filter (Evensen and Van Leeuwen, 1994). The main differences between the two filters lie in the procedure of generating the analysis ensemble. The analysis ensemble of the EnK filter is determined by assimilating perturbed observations to each member of the background ensemble (Tippett et al., 2003). The SEIK filter does not require perturbing the observations and applies the analysis only once, to provide both the mean analysis and the corresponding error covariance matrix. The filter combines the low rank approximation with an ensemble representation of the error covariance matrix. It therefore avoids the linearization used in the SEEK filter and substitutes it by a linear interpolation resulting in less errors especially with highly non-linear dynamical systems (Pham, 2001). The SEEK and the SEIK filters have the same analysis formula and the only difference between them is the use of the tangent linear model or of a nonlinear ensemble forecasting scheme to perform the evolution of the statistics of the forecast state. Triantafyllou et al. (2003b) tested SEIK filter to assimilate synthetic chlorophyll data into a 3D complex ecosystem model. Recently, Korres et al. (2007) successfully implemented the SEIK filter to assimilate satellite and in situ data into a high resolution POM model of the Mediterranean Sea.

This study presents the first attempt to use advanced Kalman filtering techniques, for the assimilation of satellite ocean color data into the complex three-dimensional marine ecosystem model of Pagasitikos Gulf. The controllability of the ecosystem variability in the deep layers using satellite measurements is a major question in marine ecology and will be addressed here. The assimilation system is further validated with the assimilation of two different data sets, GlobColour and SeaWiFS data, during two different time periods. Section 2 presents the hydrodynamic and ecological components of the ecosystem model and the nesting and initialization techniques implemented for the hydrodynamic model. The assimilation scheme is described in Section 3. Assimilation results of GlobColour and SeaWiFS data into the coupled model are reported and the behavior of the assimilation system is discussed in Sections 4 and 5. A general conclusion, including a discussion on the progress we have made so far and the problems that still need to be addressed, is offered in Section 6.
2. Description of the ecosystem model

2.1. The numerical models

The 3D ecosystem model developed for the Pagasitikos Gulf consists of two on-line coupled sub-models: the three-dimensional Princeton Ocean Model (POM, Blumberg and Mellor, 1987) described in Petihakis et al., 2012-this issue-b, and the ecological model described in Petihakis et al., 2012-this issue-a. The physical model solves for the hydrodynamics of the area and provides the background physical information to the ecological model, which describes the biogeochemical cycles. The computational domain of Pagasitikos Gulf model covers the geographical area between 22.8125°E to 23.3025°E and 39.0°N to 39.43°N (Fig. 1). It uses a Cartesian coordinate system consisting of 49×45 grid points with a horizontal grid resolution of 0.01° deg both in latitude and longitude. In the vertical there are 25 layers of variable thickness with logarithmic distribution near the surface and the bottom for greater accuracy where velocity gradients are larger. The tidal signal is not considered in the model dynamics. According to Balopoulos et al. (1977) who performed current measurements during the summer and autumn of 1976, the most likely factors affecting the circulation patterns of the gulf are the meteorological forcing and the intrusion of the Aegean Sea waters into the gulf. Tidally induced currents were found to be very small. The same conclusion can be drawn from the work of Lascaratos and Theocharis, 1984.

The Pagasitikos high resolution coastal model is nested into the lower resolution (1/30°) model of the Aegean Sea being part of the Greek POSEIDON forecasting system. The Aegean Sea model is further nested to a 1/10° implementation of POM in the Mediterranean Sea (Fig. 1). 25 sigma levels were used in both models to solve the vertical column with logarithmic distribution near the sea surface and the bottom. All three models are forced with the 72 h forecasting fields of the POSEIDON meteorological model, which is an implementation of the ETA limited area model in the Mediterranean and the Aegean Seas with 20/10 km respectively. The coupling between the hydrodynamic and the atmospheric model allows one-way feedback ocean–atmosphere mechanisms to take place: although the atmospheric model heat, moisture and momentum budget is not influenced/corrected by the oceanic model, the latter can adjust the evaporative, upward long-wave radiation and sensible heat flux consistently with its own surface temperature. This is done by using properly tuned bulk formulae for the computation of the surface momentum, heat and freshwater fluxes at each time step of model integration using the atmospheric parameters and the SST as predicted by the hydrodynamic model itself. The air–sea interaction scheme is the same as the one described in Korres et al. (2002) for the Aegean Sea model.

Fig. 1. The three nested hydrodynamic models and their bathymetry: Mediterranean (1/10°), Aegean Sea (1/30°) and Pagasitikos Gulf (1/100°) models.
The ecosystem model extends the 1-D European Regional Seas Ecosystem Model (ERSEM) (Baretta et al., 1995; Blackford et al., 2004). ERSEM is a generic model that can be applied into a wide range of ecosystems allowing for adequate level of complexity. The ecological model has already been tested at sub-basin (Aegean–Levantine), shelf and coastal areas of the Eastern Mediterranean (Triantafyllou et al., 2003a; Petihakis et al., 2002). The biological part of the model consists of interlinked modules which describe the biological and chemical interactions between the state variables. A general description of the model is offered by Baretta et al. (1995). In ERSEM organisms are clustered into functional groups following their role in the system. All important physiological (ingestion, respiration, excretion, egestion etc.) and population (growth, migration, mortality) processes are included, while carbon dynamics are coupled to chemical dynamics of nitrogen, phosphate, silicate and oxygen. Primary producers are subdivided into four size groups, picoplankton, nanoplankton flagellates and diatoms with the latter two being differentiated by the silicate uptake. Heterotrophs include heterotrophic flagellates, microzooplankton and mesozooplankton while decomposers are represented by heterotrophic free living pelagic bacteria. Although ERSEM has the option of a full benthic system in this particular application the simpler benthic returns module was used in order to avoid unnecessary expensive calculations. Thus the coupling between the water column and the benthos is done through the sedimentation of detritus and the diffusional nutrient fluxes into and out of the sediment due to advection and remineralization.

2.2. Nesting with the Aegean Sea hydrodynamic model

Nesting is a finite-difference technique to simulate a high-resolution domain embedded in a coarse resolution model. In our case, the coarse resolution model is the Aegean Sea model, which is one-way nested with the Pagasitikos coastal model (fine grid model) with a 1/100° × 1/100° resolution. By one-way nesting we mean that the boundary conditions of the fine grid model are prescribed by external data taken from the coarse resolution model while the solution of the latter is not modified by the solution of the fine grid model in their common overlapping area. Of major importance in nesting techniques is the conservation of properties between the coarse and fine grid model and the treatment of fine grid interior noise to avoid inconsistencies between the two models.

The coarse resolution model is part of the POSEIDON-II system (Nittis et al., 2006a) and is implemented within the Aegean Sea region with a 1/30° × 1/30° horizontal resolution and 25 sigma-levels in the vertical (Fig. 1). It was integrated for a 6-year period (2000–2006) nested with a basin scale model (Mediterranean Sea model 1/10° × 1/10°) using the high resolution atmospheric momentum, water and heat flux fields produced for the INSEAN project by running a 1/10° × 1/10° version of the ETA atmospheric model over the Mediterranean region. During the Aegean model integration, model prognostic variables over specific sections are instantaneously stored every 1-hour for further use in the Pagasitikos model.

Nesting with the Aegean Sea model is applied along the eastern boundary of the Pagasitikos model (located at 23.3025°E) following the methodology described in Korres and Lascaratos (2003) and successfully tested in other applications (Triantafyllou et al., 2003a; Zodiatis et al., 2003; Nittis et al., 2006b). The nesting procedure involves the variables \( U_c, V_c \) (zonal and meridional velocity components), \( T_c, S_c \) (temperature and salinity) and \( n_c \) (the sea surface elevation) of the coarse grid model (Aegean Sea) and the prognostic variables \( U_f, V_f \) (the external mode zonal and meridional velocity components), \( U_i, V_i \) (the internal mode zonal and meridional velocity components), \( T_i, S_i \) and \( n_i \) (free surface elevation) of the fine grid model (Pagasitikos model). During the Pagasitikos model run, the Aegean Sea model variables \( U_c, V_c, T_c, S_c \) and \( n_c \) are interpolated onto the y-z open boundary section of Pagasitikos model at each time step. The linearly interpolated variables are denoted here by \( U_{INT}^c, V_{INT}^c, T_{INT}^c, S_{INT}^c \) and \( n_{INT}^c \), respectively. The spatial interpolation is bilinear for all Aegean Sea model variables, with the additional constraint of volume conservation through the open boundary: the interpolated field \( U_{INT}^c \) is corrected in such a way that it guarantees volume conservation between the coarse and the fine grid model, i.e.

\[
\int_{y_1}^{y_2} \int_{z_1}^{z_f} U_{INT}^c \, dz \, dy = \int_{y_1}^{y_2} \int_{z_1}^{z_f} U_c \, dz \, dy
\]

where \( y_1, y_2 \) and \( z_1, z_2 \) are the extremes of the open boundary section for the fine and coarse grid models, respectively, while \( H_c \) and \( H_i \) are the corresponding bathymetries along the section.

2.2.1. OBCs for the external mode

The barotropic velocity (external mode) normal to the open boundary of the Pagasitikos model is specified according to a modified Flather (1976) condition which efficiently allows interior disturbances – due to possible mismatches between coarse and nested values – to pass out through the lateral boundary. The Flather boundary condition, initially proposed for tidal models, combines a Sommerfeld-type radiation condition:

\[
\frac{\partial n}{\partial t} - \sqrt{gH_f} \frac{\partial n}{\partial x} = 0
\]

with an one-dimensional version of the continuity equation:

\[
\frac{\partial ((H + n)U_f)}{\partial x} + \frac{\partial n}{\partial t} = 0
\]

to yield a boundary condition for the normal barotropic velocity \( U_f \) of Pagasitikos model:

\[
U_f = U_{INT}^c + \frac{\sqrt{gH_f}}{H_f + n_f} \left( n_{INT}^c - n_f \right)
\]

where \( U_{INT}^c \) is

\[
\frac{1}{n_f + n_{INT}^c} \int_{z_1}^{z_f} U_{INT}^c \, dz.
\]

Sommerfeld boundary condition does not in general respect volume conservation. Thus, one is forced to apply volume conservation constraints in cases of significant imbalances between the net volume transport at the open boundary and the time variation of the total volume of the modeling domain. As a result, the interpolated field \( U_{INT}^c \) is corrected in such a way that it guarantees volume conservation between the coarse and the fine grid model.

The tangential barotropic velocity at the open boundary is directly prescribed from the Aegean Sea model:

\[
V_f = V_{INT}^c
\]

2.2.2. OBCs for the internal mode

The internal mode velocities \( U_f \) and \( V_f \) (normal and tangential) at the open boundary of Pagasitikos model are directly prescribed from the Aegean Sea model:

\[
U_f = U_{INT}^c, \quad V_f = V_{INT}^c
\]

2.2.3. OBCs for temperature and salinity

The temperature and salinity profiles \( T_f \) and \( S_f \) at the open boundary of Pagasitikos model are updated by using an upstream advection scheme whenever the normal velocity is directed outwards from the modeling area:

\[
\frac{\partial T_f}{\partial t} + U_f \frac{\partial T_f}{\partial x} = 0, \quad \frac{\partial S_f}{\partial t} + U_f \frac{\partial S_f}{\partial x} = 0, \quad U_f > 0
\]
In cases of inflow through the open boundary, temperature and salinity are prescribed directly from the interpolated OGCM temperature and salinity profiles \(T_{c}^{\text{INT}}, S_{c}^{\text{INT}}\):

\[
T_{f} = T_{c}^{\text{INT}}, \quad S_{f} = S_{c}^{\text{INT}}, \quad U_{f} > 0
\]  

(8)

2.2.4. OBC for the free surface elevation

For the specification of the free surface elevation \(n_{y}\) at the open boundaries of the Pagasitikos model, we have adopted a zero-gradient condition:

\[
\frac{\partial n_{y}}{\partial x} = 0
\]  

(9)

2.3. Initialization of the hydrodynamic model

The model initialization problem is of central importance for a short-range near real time forecast system. A proper initialization procedure should produce balanced initial conditions that do not excite inertia-gravity oscillations during the model integration, which contaminates the forecasting result. This is of great importance for coastal models which, due to the bathymetric steep slopes they usually encompass, are prone to produce high frequency oscillations when the coarser model solution is downscaled by simple interpolation techniques onto their model grid. Moreover, these models have usually a free-surface which in contrast to the rigid-lid assumption cannot filter out fast moving waves with often catastrophic consequences for the limited time forecast itself. Among different approaches to treat the initialization problem (damping time integration procedures, adjoint model) regional and shelf models within the POSEIDON system use the VIFOP technique (Variational Initialization and Forcing Optimization Platform), a 3D variational initialization method which minimizes a cost function based on data constraints and dynamical penalties which involve the tangent linear model. This technique, which has been designed to satisfy both statistical and dynamical constraints, leads to a drastic reduction of numerically generated external gravity waves and produces a dynamically consistent model initialization field (Aucclair et al., 2000).

The VIFOP platform has been implemented for the two models of the POSEIDON system (Mediterranean & Aegean Sea models) and the Pagasitikos coastal model (INSEA project). The Mediterranean model is initialized from the MERSEA relevant products while the other two models (and in particular the Pagasitikos coastal model) are initialized directly from the POSEIDON system solutions (Mediterranean model → Aegean Sea model → Pagasitikos model). The initialization procedure optimizes only the external mode (barotropic) of the background field in such a way that spurious gravity waves are completely filtered out during the initialization process. The free surface elevation of the Pagasitikos model 12, 18 and 24 h after its initialization from the Aegean Sea model solution performed at 01/01/2003 18 UTC in the case where the external model is not optimized by the VIFOP package is plotted in the left column of Fig. 2. It is evident that the model solution even 18 h after the initialization process is contaminated by spurious gravity waves developed near the land boundaries. On the other hand the corresponding model solution is totally free of such gravity waves when the model is initialized with the VIFOP package optimized the external mode initial fields (right column of Fig. 2).

3. Implementation of a data assimilation system for the Pagasitikos ecosystem

The SEEK, SFEK, and SEIK filters were implemented to assimilate satellite ocean color data into the ERSEM model describing the Pagasitikos ecosystem. The physics simulated by POM and used to force the biology were assumed perfect. Because of the significant computational burden, the SEEK and SEIK filters were implemented in a parallel MPI architecture in which the multiple model runs needed to update the forecast error covariance matrices were performed simultaneously. This section describes the filters and discusses their characteristics.

3.1. The SEEK filter

The SEEK filter is a sequential data assimilation scheme derived from the extended Kalman filter (Pham et al., 1997). In a sequential scheme, the estimation of the state of the system uses only observations up to the estimation time. The solution of this problem has been solved by the well-known Kalman filter in the linear case. In the non-linear case, the system is often linearized around the most recent state estimate, yielding to the so-called the extended Kalman (EK) filter (Pham et al., 1997). Brute-force implementation of the EK filter is, however, not possible in practice because of the prohibitive computational requirements of the model and the huge dimension of the system \(n\).

The SEEK filter aims at simplifying the algorithm of the EK filter in order to make its implementation cost reasonable for meteorological and oceanographic applications. The main idea is to approximate the error covariance matrix by singular low rank \(r \ll n\) matrices. The covariance matrix is then represented by a limited number of vectors, describing the dominant modes of the system and defining in this way the structure of the directions along which the filter correction is applied. In its general form, the “correction directions” evolve in time to follow changes in the model dynamics (Pham et al., 1997). The Empirical Orthogonal Functions (EOFs) computed from model outputs is the most common approach to initialize the correction directions of the filter.

To present the algorithm of the SEEK filter, we adopt the notation proposed by Ide et al. (1997). Consider a physical system described by:

\[
X_{k}^{0} = M(t, t_{k-1})X_{k-1}^{0} + \eta_{k}
\]  

(10)

where \(X_{k}^{0}\) denotes the vector representing the true state at time \(t\), \(M(s, t)\) is an operator describing the system transition from time \(s\) to time \(t\) and \(\eta\) is the system noise vector. At each time \(t_{k}\), one observes

\[
Y_{k}^{0} = H_{k}X_{k}^{0} + \varepsilon_{k}
\]  

(11)

where \(H_{k}\) is the observational operator and \(\varepsilon_{k}\) is the observational noise. The noise \(\eta_{k}\) and \(\varepsilon_{k}\) are assumed to be independent random vectors with mean zero and covariance matrices \(Q_{k}\) and \(R_{k}\), respectively.

As the EK filter, the SEEK filter proceeds in two steps, apart from an initialization step. To initialize the filter, an objective analysis is used to improve the initial a priori estimate of the system state with the first observation \(Y_{0}^{0}\)

\[
X_{0}^{a} = X_{0} + L_{0}U_{0}^{T}H_{0}R_{0}^{-1}(Y_{0}^{0} - H_{0}X_{0})
\]  

(12)

where \(L_{0}\) is a \(n \times r\) matrix containing the \(r\) retained EOFs on its columns,

\[
U_{0} = [L_{0}^{T}H_{0}R_{0}^{-1}H_{0}^{T}]^{-1},
\]  

(13)

\(X_{0}^{a}\) is an estimation of the model state at the time of the first available observation and \(H_{0}\) is the gradient of \(H_{0}\) evaluated at \(X_{0}\). The initial analysis error covariance matrix is then

\[
P_{0}^{a} = L_{0}U_{0}^{T}L_{0}^{T},
\]  

(14)

3.1.1. Forecast step

At time \(t_{k-1}\), the estimate of the system state is \(X_{k-1}^{a}\) and the associated error covariance matrix \(P_{k-1}^{a}\), in the factorized form is expressed as:

\[
P_{k-1}^{a} = P_{k-1}^{a}U_{k-1}^{T}L_{k-1}^{T},
\]  

(15)
where the matrix $U_{k-1}$ is of dimension $r \times r$. The model (10) is then used to forecast the state as:

$$X_f^k = M(t_k, t_{k-1})X_a^{k-1}. \quad (16)$$

The matrix $L_k$ is updated with the tangent linear model

$$L_k = \frac{\partial M}{\partial X} |_{X_a^{k-1}, L_{k-1}}. \quad (17)$$

The corresponding forecast error covariance matrix is approximated by:

$$P_f^k = L_k U_{k-1} L_k^T + Q_k. \quad (18)$$

### 3.1.2. Correction step

The new observation $Y^k_o$ at time $t_k$ is used to correct the forecast according to:

$$X_a^k = X_f^k + G_k \left[ Y^k_o - H_k X_a^f \right]. \quad (19)$$

where $G_k$ is the gain matrix given by:

$$G_k = L_k U_k L_k^T H_k^T R_k^{-1} \quad (20)$$

with $H_k$ the gradient of $H_k$ evaluated at $X_a^f$ and $U_k$ computed from:

$$U_k^{-1} = \left[ U_{k-1} + \left( L_k^T L_k \right)^{-1} L_k Q_k L_k \left( L_k^T L_k \right)^{-1} \right]^{-1} + L_k^T H_k^T R_k^{-1} H_k L_k \quad (21)$$

The corresponding filter error covariance matrix is then equal to:

$$P_a^k = L_k U_k L_k^T. \quad (22)$$

Note that the filter update is applied along the directions of $L_k$ only, hence it is called correction basis of the filter. Since the computation of $P_f^k$ and $P_a^k$ in Eqs. (9) and (13) are not needed for the algorithm, the SEEK filter drastically reduces the computational cost with respect to the EK filter. It mainly requires $r+1$ model integrations of the numerical model to update the forecast state and the correction basis.

To take into account model deficiencies, we follow the same approach as Pham et al. (1997). A so-called compensation technique that amplifies the pre-existing error modes during the forecast is used to parameterize the model error. The term $(L_k^T L_k)^{-1} L_k Q_k L_k (L_k^T L_k)^{-1}$ expressing this error in the Eq. (21) is then taken into account by means of a forgetting factor $\rho$. This equation is then rewritten as:

$$U_k^{-1} = \rho U_{k-1}^{-1} + L_k^T H_k^T R_k^{-1} H_k L_k \quad (23)$$

Such a factor replaces the contribution of the imperfect model by amplifying the already existing modes of the background error. The use of a forgetting factor further improves the stability of the filter and alleviates the problems due to the underestimation of the filter's error covariance matrices (by low rank matrices).

To avoid the linearization of the model equations, Pham (2001) introduced the ensemble variant of the SEEK filter, called the Singular Evolutive Interpolated Kalman (SEIK) filter. This filter uses 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 because of its low implementation cost, is the Singular Fixed Extended Kalman (SFEK) filter which simply neglects the evolution of the correction directions. Several other variants of the SEEK and SEIK filters, not considered here, with different degrees of simplification for the evolution of matrix \( L \) have been also developed by Hoteit et al. (2002) 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.

3.2. The SEIK filter

The SEIK filter is a low rank square-root ensemble filter designed for applications with strongly nonlinear models. As the ensemble Kalman (EnK) filter, the SEIK filter applies an ensemble of model states to represent the error statistics of the model estimate. Additionally, it operates under the assumption of low rank \( r \) error covariance matrices allowing for the use of the smallest number of ensemble members \( r + 1 \). The initial state of the filter is assumed to be Gaussian with the mean and covariance set as the mean and the sample covariance of a long model run. An EOF analysis is then applied to compute a low-rank covariance matrix from the sample model state vectors. The initial ensemble can be generated using the second order exact sampling scheme which preserves the first two moments of the distribution as described by Pham (2001) and Hoteit et al. (2002).

The EnK and SEIK filters differ in the generation of the analysis ensemble: the EnK filter corrects each member of the background ensemble using perturbed observations while the SEIK filter samples the analysis ensemble from the distribution of the analysis state, using a so-called second-order exact sampling scheme (Pham, 2001). The three steps of the SEIK filter are summarized below. The reader is referred to Pham (2001) and Hoteit et al. (2002) for a detailed description of the filter.

3.2.1. Sampling step

Sampling of \( r + 1 \) ensemble members such as their mean and sample covariance matrix exactly matches \( X^f_{k-1} \), and \( P^f_{k-1} \).

3.2.2. Forecast stage

The model is used to integrate the analysis ensemble forward in time with the model to compute the forecast members \( X^f_k \):

\[
X^f_k = M(f_k, f_{k-1})X^a_{k-1}.
\]

(24)

The forecast state is then the mean of the ensemble and the corresponding forecast error covariance matrix is approximated by the sample covariance matrix of the \( X^f_k \). The later can be decomposed as

\[
P^f_k = L_k[(r + 1)T^T]^{-1}L_k^T + Q_k
\]

(25)

where \( T \) is a \((r + 1) \times r \) orthogonal matrix with zero columns sum and

\[
L_k = [X^f_k - X^f_{k-1}]T.
\]

(26)

3.2.3. Correction step

The correction of the forecast state is done according to the formulas

\[
H_k L_k = \left[H_kX^f_k - H_kX^a_{k-1}\right]T
\]

(27)

\[
U_{k-1} = \left[(r + 1)T^T\right]^{-1}
\]

(28)

3.3. The correction directions and the observations/model error specification

The initialization of the filters requires an initial state and a low rank approximation of the error covariance matrix. The strategy for the estimation of the error covariance matrix is to use model statistics as an approximation of the true system statistics. Then by sampling the model state vectors, one can obtain an error covariance matrix and approximate it through the dominant EOFs.

In this work, we followed several previous studies, e.g. Pham et al. (1998), Brasseur et al. (1999), Verron et al. (1999), Pham (2001), Nerger et al. (2005), Hoteit et al. (2002, 2003), and we set the filter initial error covariance matrix as the sample covariance matrix of the outputs of an adequately long model run as explained below. EOFs were then computed to reduce the rank of the error covariance matrix. This is a natural estimate in the absence of information at the initial times and has been already proven as a very reasonable choice in all the references listed above to the point that some of these studies considered using these statistics as an invariant error covariance matrix. However the risk in this approximation is that the filter may not account – especially at the beginning – for dynamical features that are not present in the error covariance matrix. Experience however suggests that the missing variability is not very important and the first filter correction step is in most cases very efficient in significantly decreasing the initial estimation error. More importantly, the initial EOFs are only used for the filters initialization and Kalman filters are known to be weakly sensitive to the specification of the initial conditions (and not sensitive when the model is linear). Kalman filters covariance matrices are thus expected to be very close to the true covariances after very few filtering steps regardless of the initial error covariance matrix.

The Pagasitikos ecosystem model has been integrated for 3 years (2000–2002) forced with the Poseidon 1/10° atmospheric model analysis. The initialization of the physics was done using the Aegean Sea model results and the VIFOP technique while ecology was initialized from a climatological run of the model. After 1 year spin up period (year 2000) of the ecosystem dynamics, model states were sampled every two days in order to produce a historical sequence of 365 ecosystem states to be used for the initialization of the filter for the first assimilation experiment performed within 2003. For the second assimilation experiment performed within 2006, the ecosystem model was integrated for the period 2004–2006 in order to produce a historical sequence of ecosystem model states sampled every 2 days. Finally, for both cases, a multivariate EOF analysis was applied on the ensemble of 365 states extracted from the last two years (2001–2002 and 2005–2006 respectively) of model integration. Since the state variables are of different nature, a metric in the state space was introduced in order to make the distance between state vectors independent from the unit of measure. This metric was defined as a diagonal matrix with elements being the inverse of the spatial variance of each state variable averaged over all grid points. For the EOF analysis performed for time period 2001–2002, the first 12 modes explain almost 90% of the total variance while for the second period (2005–2006) the first 15 modes explain approximately 88% of the total variance.

The observational covariance matrix \( R \) needs to be specified in the filters algorithm. It is common to consider a diagonal observation error covariance matrix \( R \), which means that the observations are assumed to be spatially uncorrelated, but this can be partly accounted for by overestimating the diagonal coefficients of \( R \), which were set as a fraction of the variance of the satellite observations at each grid point. In our case a uniform error of 0.05 mg/m³ is assumed.

Model errors originate from the various biogeochemical and physical parameterizations (bearing in mind that the hydrodynamics act as a forcing term on the ecology), the numerical discretization schemes, the boundary conditions (i.e. atmospheric deposition, river input, properties advection at the open boundary), imperfections in
the model forcings (atmospheric forcing for example), and may be characterized by a wide spectrum of space scales. Kalman filters allow for models deficiencies and include model errors as a stochastic forcing term. This is then accounted for as an additional term in the forecast error covariance matrix that increases uncertainties in the forecast state. This is certainly not the optimal way to account for errors in the atmospheric forcings and better treatment would to estimate these errors as part of the state vector. The present work only estimated the state vector but this does not mean that the model errors were not accounted for. Estimating model deficiencies would improve the filters forecast, but this was not critical for our study as shown by the good model forecasts we obtained in our experiments. Future studies will consider improving the skills of our system forecast through the estimation of model errors.

4. Assimilation of satellite remote sensing data during 2003

Two satellite remote sensing Chl-a datasets were used in the assimilation and validation of the system solution. It is important to note that the algorithms used to produce the final datasets may be highly influenced by several parameters such as suspended substances and/or sediments especially in the coastal waters of Pagasitikos Gulf. Thus, a proper validation of this parameter should involve a ground-truth comparison with large number of in situ data of approximately equal spatio-temporal distribution. Because of the lack of these data, several different sensors and algorithms were tested here in order to determine the most appropriate datasets.

The MODIS (Moderate-resolution Imaging Spectroradiometer) and SeaWiFS (Sea-viewing Wide Field-of-view Sensor) data of the near surface Chlorophyll-a (mg m$^{-3}$) were obtained from the NASA Oceancolor website (http://oceancolor.gsfc.nasa.gov/) in HDF format. The analysis involved the processing of the global Level-3 8-day composite products (4 and 9 km$^2$ resolution respectively) for 2003 for Pagasitikos Gulf (Greece). After processing, the data were imported in a GIS software to produce weekly maps, as well as producing monthly and weekly averaged time-series. The Chl-a product was also produced using GlobColour datasets, which briefly are the outcome of a merged product of SeaWiFS, MODIS and MERIS information, literature as well as the GlobColour datasets can be found at: http://www.globcolour.info/. Here we processed the original GlobColour Chl-a data as well as implementing the Mediterranean algorithm MED-OC4 (Volpe et al., 2007).

A series of sensitivity experiments were carried out in order to test the performance of the different filters, SFEK, SEEK and SEIK for the assimilation of SeaWiFS data into the Pagasitikos ecosystem model. All the experiments were performed with an error covariance matrix of rank 7 to limit the computational burden. The forgetting factor was set to 0.5 and the observational error to 0.05 mg/m$^3$. To intercompare the three filters performance, the model derived Chl-a was chosen as this quantity can be directly evaluated against satellite observations and at the same time is closely related to the dynamics of the ecosystem. Fig. 3 shows the time evolution of the Chl-a RMS error for the forecast state (upper panel) and the analysis state (lower panel) for the SFEK, SEEK and SEIK filters and compares their performances to the model free-run (without assimilation). It can be clearly seen that the SFEK filter provides the worst results while the SEEK and SEIK filters performances are quite similar, except at the 40 to 45 8-day assimilation cycles were the SEEK filter provides a better solution. This is probably due to the stochastic nature of the SEIK filter which usually requires larger ensemble size to efficiently operate. In the rest of this section, we only discuss results from the sensitivity experiments with the SEEK filter.

A number of simulations were performed with the SEEK filter to study its sensitivity to different values of the forgetting factor. Fig. 4 shows the time evolution of the forecast (upper panel) and analysis (lower panel) RMS errors for Chl-a as obtained from three assimilation experiments with three different values of the forgetting factor $\rho$ 0.3, 0.5 and 0.9. Results of these experiments reveal that $\rho = 0.5$ was the most appropriate, effectively reducing the RMS
error of Chl-a over all the assimilation window. Note that the simulation with a forgetting factor of 0.9 could not be completed as the assimilation system diverged after 29 assimilation cycles. Additional sensitivity experiments were performed to determine the most appropriate rank of the filter error covariance matrix. Three choices of covariance matrix rank \( r = 5, 7 \), and 9 were examined with the SEEK filter. Fig. 5 presents the time evolution of the Chl-a RMS error before (forecast — upper panel) and after the insertion of observations (analysis — lower panel) steps in comparison with the Chl-a RMS error corresponding to the model free-run. It can be seen that more than five correction directions (\( r = 5 \)) are needed in order to stabilize the filter performance. Results obtained with filter ranks 7 and 9 were not significantly different. It worth noting that the assimilation system was not quite stable by using more than 9 EOFs, suggesting that the last EOFs where spoiled by noise. To limit the influence of these spurious correlations, localization techniques can be used as done in the context of the ensemble Kalman filters by Houtekamer and Mitchell (2001).

The comparison between the estimates of the assimilation system and the satellite data performed using both Chl-a two dimensional maps and 8-day mean time series. Overall, as it can be seen from Fig. 6, there is a reasonable agreement between the unconstrained model solution (model free run) and the various observational data sets. Specifically, the best correlation can be seen with the GlobColour Med data. There are however periods, such as the period between the 7th and 15th weeks, during which the two datasets did not follow similar patterns. Fig. 7 illustrates the effect/contribution of GlobColour Med data on the model solution through the data assimilation process. The data assimilation is performed with SEEK filter using 7 correction directions and a forgetting factor of 0.5. The unconstrained model solution is shown in red while the SEEK filter forecast and analysis time series are denoted with the blue and green lines respectively. It can be clearly seen that the assimilation of GlobColour data into the model significantly enhances the correlations between the satellite data and the model output. Additionally, a strong indication that the filter solution is compatible with the biogeochemical model dynamics is the fact that the forecast and analysis time series of the assimilation system are showing quite a similar behavior.

In addition, to the time series comparison, mean Chl-a 2-D maps for January and November 2003 are presented in Fig. 8, indicating that the impact of assimilating GlobColour data was mainly positive on the final estimates of modeled Chl-a. This can be particularly seen in both months as the Chl-a analyses were in better agreement with the data. Note that the missing points are probably due to cloud cover.

5. Assimilation of satellite ocean color data for the period 2006

This section presents and discusses the results of four assimilation experiments in which the 2006 SeaWiFS data were assimilated into the Pagasitikos ecosystem model using the SEIK, SEEK and SFEK filters.

5.1. The observational data set

High-Resolution Picture Transmission (HRPT) SeaWiFS Level-1A data, acquired by the receiving station HRROM at ISAC-CNR in Rome, Italy (41.84°N, 12.65°E), for two years (2006–2007) were processed up to Level-3 with the SeaWiFS Data Analysis System (SeaDAS) software package version 4.8 (Baith et al., 2001) available from the NASA website (http://seadas.gsfc.nasa.gov/). Standard flags and Siegel’s atmospheric correction algorithm were applied to Level-1A raw data (McClain et al., 1995; Siegel et al., 2000). Chlorophyll concentrations were computed from Level 2 Remote Sensing Reflectance, using a validated Mediterranean-specific algorithm (MedOC4; Volpe et al., 2007). Volpe et al. (2007) have shown that MedOC4 produces more realistic...
values in the Mediterranean, differing from in situ values by about 35% for CHL < 0.4 mg m$^{-3}$ and 40% for CHL > 0.4 mg m$^{-3}$, in contrast to over 100% when OC4v4 (O’Reilly et al., 2000) is used (relative to the 35% SeaWiFS mission goal). CHL maps were remapped at 1 km spatial resolution at Nadir on an equi-rectangular grid covering the model grid domain (from 22.7 to 23.3°E and from 38.9 to 39.5°N). A daily binning was performed on the base-10 log-transformed chlorophyll to take account of its lognormal distribution (Campbell, 1995). Weekly averages were computed from the daily images.

5.2. Assimilation results

A set of sensitivity experiments were conducted with the SFEK, SEEK and SEIK filters in order to understand the filters behavior and to determine the most appropriate setup for the assimilation of weekly SeaWiFS data into the Pagasitikos ecosystem model for year 2006. The assimilation experiments were performed during the time period extending from 4 January 2006 up to the first week of October 2006.
Fig. 8. GlobColour Med data for January and November 2003 (upper panel), Chl-a model forecast (middle) and analysis (lower panel) for the same months corresponding to the assimilation of GlobColour data with the SEEK filter.
(10 months/40 weeks) because the data were very sparse during the last three weeks of October. Overall four assimilation experiments were performed and these are summarized in Table 1. The filters were implemented with rank-15 (or 15 correction directions). The correction directions of the SEEK and SEIK filters were updated in parallel using MPI programming.

Fig. 9 presents the time evolution of the RMS misfit between the assimilated chlorophyll data and the estimated chlorophyll concentrations as it results from the model free-run (without assimilation), and the SFEK filter run (EXP1) before (forecast) and after (analysis) the filter correction. After a few reasonable corrections cycles and upon entering the spring bloom period, the SFEK filter becomes quite unstable showing worse performance than the model free run. The initiation of this phase is within the spring period when the filter is not able to recover the rapidly changing non-linear ecosystem dynamics. Although the filter correction step was always able to improve the model forecast, the SFEK filter assimilating Chl-a was never able to put the model back on the right trajectory presenting a significant increase in the estimation error after every forecast step. Several different values of covariance matrix rank, forgetting factor, and observational error were tested, but all of these could not prevent the divergence of the filter, suggesting that the SFEK filter cannot efficiently handle rapidly changing regimes that are poorly represented by an invariant set of correction directions. These results also point out to the limitations of Chl-a in controlling the behavior of the ecosystem in the Pagasitikos Gulf.

Fig. 10 shows the time evolution of the RMS misfit between the assimilated chlorophyll data and the estimated chlorophyll concentrations as they result from the model free-run (without assimilation), and the filter forecast and analysis of experiments EXP2, EXP3 and EXP4 (Table 1). The results of these runs demonstrate that the update of the filter correction directions significantly improves the stability of the filter. Moreover, inter-comparison of RMS misfits for experiments EXP2 and EXP3 shows the critical role of the additional 4 correction directions to the performance of the SEIK filter during the spring bloom period (weeks 7–14). The SEIK and SEEK filters using 15 correction directions (EXP3 and EXP4 respectively) show very similar behavior over the whole assimilation period. Overall, both filters behavior is quite satisfactory and clearly improves the model/

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Filter</th>
<th>Number of correction directions</th>
<th>Forgetting factor</th>
<th>Obs. error (mg/m³)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXP1</td>
<td>SFEK</td>
<td>15</td>
<td>0.6</td>
<td>0.05</td>
</tr>
<tr>
<td>EXP2</td>
<td>SEIK</td>
<td>11</td>
<td>0.6</td>
<td>0.05</td>
</tr>
<tr>
<td>EXP3</td>
<td>SEIK</td>
<td>15</td>
<td>0.6</td>
<td>0.05</td>
</tr>
<tr>
<td>EXP4</td>
<td>SEEK</td>
<td>15</td>
<td>0.6</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Fig. 9. Time evolution of the basin mean average Chl-a RMS error corresponding to the model free run and the forecast and analysis of experiment EXP1.

Fig. 10. Time evolution of the basin mean average Chl-a RMS error corresponding to the model free run and experiments EXP2, EXP3, EXP4 forecast (upper panel) and analysis (lower panel).
data consistency. The RMS error for both the forecast and the analysis is always lower than the RMS error of the free-run reaching values comparable to the specified observational error. The filter correction step efficiently improves the forecast after every filtering cycle, suggesting the importance of the data assimilation into the model. The analysis step of the filters successfully provides states that respect the model dynamics, as no significant increase in the estimation error is observed after every assimilation cycle. The spatial distribution of the Chl-a RMS model/observations data differences as obtained from the model free run and experiment EXP4, averaged over the entire assimilation window (04 Jan – 04 Oct 2006), are shown in Fig. 11a-c. As expected, the model free-run/observations differences (Fig. 11a) are the largest, with the worst model performance observed in the north-western part of the gulf and within the elongated channel to the south of the gulf where the water mass exchange with the Aegean Sea is taking place. The filter forecast RMS error (Fig. 11b) is better than that of the free-run over the whole model domain with the exception of the north-western edge of the gulf where the model error remains considerable with respect to the free run. The filter correction step (Fig. 11c) significantly reduces the RMS error over the whole model domain.

To further assess the ecosystem functioning of the assimilation system compared to that of the free run, Fig. 12 top shows the horizontal distribution of Chl-a integrated from surface to 20 m and a north-south cross section of the same ecological variable at longitude 23.02°E (exactly at the middle of the basin) during the late winter period (15 Feb 2006) as they result from the assimilation and the free run. In situ data have shown that there is absence of periodicity with maximum Chl-a concentrations observed at the coastal areas and in particular at the inner part of the gulf close to the city of Volos (3.5 mg/m³), followed by the surface layer of the central–external gulf (Petihakis et al., 2005). Winter is characterized by complete water column mixing with a rather variable circulation pattern and a sporadic presence of an anticyclone in the east part of the gulf. In situ surface Chl-a concentrations during 24/02/1999 were in the range of 0.1–0.5 mg/m³ with the areas of Volos and Almyros at the North West exhibiting higher values (Theodorou and Petihakis, 2000). An analogous E–W gradient was also observed in Trikeri channel (connection with N. Evoikos) exhibiting higher values (up to 0.5 mg/m³) at its western part. The free run during 15/02/2006 tends to overestimate with higher concentrations in the whole domain in contrast with the assimilated run where there is a significant improvement. Thus with the assimilation the internal and central parts of the gulf reach Chl-a values 0.2 to 0.4 mg/m³ with an increase in the external part and in the connection channel from 0.5 to 1.0 mg/m³.

In the vertical, the mixed water column with the subsequent vertically uniform Chl-a concentrations are reproduced by the assimilation filter. Another feature observed in field measurements (Theodorou and Petihakis, 2000) are the slightly higher values at the west part of the gulf (0.4–0.6 mg/m³) compared to the east (0.1–0.3 mg/m³). This is nicely exhibited in the assimilated run as the overestimations produced by the free run are successfully corrected.

Fig. 13 presents the horizontal distributions of phosphates and nitrates at the surface for the same time period resulting from the free and the assimilation runs. Phosphorus is particularly important as it is the most limiting nutrient in large parts of the Mediterranean determining to large extent both the range as well as the spatial pattern of primary and secondary productions. Closed and semi-closed systems are characterized by phosphate limitation because the amount of biologically available phosphorus is small in relation to the quantity required for algal growth (Mason, 1983). Thus, although one would expect Pagasitikos to follow the above trend with nitrogen ions being in excess, looking at the Redfield ratio (16) in the measured data the system of the gulf undergoes alternating periods of nitrogen/phosphorus limitation. The top layer of the central-external gulf exhibits longer periods of phosphate limitation in contrast to the deeper layer which is mostly nitrogen limited. The internal gulf is also mostly nitrogen limited despite a significant period (May 1998–September 1998) during which the N/P ratio is equal to the Redfield ratio (Petihakis et al., 2005). Measured surface layer phosphate concentrations during 24/02/1999 (Theodorou and Petihakis, 2000) exhibited an increasing gradient from coastal areas to the center of the gulf with values being close to detection limits (<0.04 mmol/m³). A prominent characteristic is that maximum nutrient concentrations observed in the deeper part of the gulf. This is due to the action of the dominant anticyclone, transporting organic matter to the benthos where through biodegradation nutrients...
are released into the overlying waters. Although the free model run reproduces the spatial variability, phosphate concentrations are overestimated over the whole model domain. The application of the assimilation filter causes the lowering of phosphate both inside the gulf as well as in the connection channel. The distribution of nitrate in Pagasitikos is determined by a number of processes such as circulation, biology, atmospheric deposition and run-off with the latter playing a significant role due to the extensive agriculture activities in the wider area of Almyros. As with phosphate, minimum nitrate concentrations were observed in the coastal areas and surface layers throughout the year, increasing at the deeper part of the gulf. Additionally nitrate exhibits a much higher variability with concentrations ranging from 0.01 to 2.4 mmol/m$^3$ (Theodorou and Petihakis, 2000). Model results both in the free and the assimilated runs are within the measured ranges with the latter reproducing more efficiently the observed spatial variability.

6. Conclusions and discussion

The assimilation system developed for the Pagasitikos Gulf ecosystem, clearly demonstrates the effectiveness of the Kalman filters in improving the consistency between the biogeochemical ocean model and available observations. Earlier works have shown that the performance of the SFEK, SEEK and SEIK filters are quite similar when applied to weakly nonlinear systems. However, in all cases studied here, the SEEK and SEIK filters performed better in the high resolution Pagasitikos ecosystem model. The behavior of the SFEK filter significantly degraded after the early spring bloom period, failing to follow this rapid change in the state of the ecology. The SEEK and SEIK filters remained remarkably more stable thanks to update of the correction base that allowed the filter to follow the model dynamics. The analysis step of these filters was however not very efficient in the spring bloom. These weaknesses might be related to the likely non-Gaussian character of the ecological state during the bloom period which cannot be properly handled by Gaussian-based filters. The use of the analysis step of the fully optimal nonlinear filter is expected to improve the analysis step. The use of nonlinear filters in realistic ocean applications is still however limited by prohibitive computational burden of these methods.

The benefit of using satellite data to compute accurate estimates of chlorophyll was clear in the experiments presented in this paper. As satellite remote sensing data are freely available and cover areas and periods during which in situ data can be lacking, these provide ‘supplementary’ information which can be proven vital to build a set of accurate analysis. However, it has to be mentioned, particularly for small areas such as the Pagasitikos Gulf, which is surrounded by coastal and CASE II waters, that it is very difficult to find the ideal remote sensing product. Different satellites and algorithms were then used to identify an appropriate product. Our findings suggest that the GlobColour data implemented with the MEDOC4 algorithm are more consistent with the model dynamics, leading to the best estimates of the ecological state in the Pagasitikos Gulf. The results of these preliminary runs also suggest that the ecosystem of the Pagasitikos Gulf is not fully controllable by Chl-a data and in situ data might still be needed to better control the system behavior.

![Figure 12: Spatial distribution and cross sections (along 23.02°E) of Chl-a for 15 Feb 2006 corresponding to the model free run (left) and model experiment EXP4 (right).](image)
Overall, assimilation of real satellite data revealed a clear improvement in the model's behavior with respect to the model free-run without assimilation. Improving the physics of the coupled model through the joint assimilation of physical and ecological observations is currently under development. Such a complete assimilation system is expected to enhance the accuracy of the ecological state estimates by providing better physics to force the ecology. Future work will also focus on incorporating a more realistic model error in the filters algorithms. This should further improve the assimilation system behavior by increasing the efficiency of the filters correction step.

Acknowledgments

This work has been supported by the INSEA EU project. Here we would like to thank Dr. Yaswant Pradhan and Dr. Samantha Lavender for the significant contribution of providing the global GlobColour datasets of Chlorophyll-a. GlobColour is a European Space Agency (ESA) funded project.

References


Fig. 13. Spatial surface distribution of phosphates and nitrates for 15 Feb 2006 corresponding to the model free run (left) and experiment EXP4 (right).