

# An Ensemble Kalman Filter with a complex marine ecosystem model: hindcasting phytoplankton in the Cretan Sea

J. I. Allen<sup>1</sup>, M. Eknes<sup>2</sup>, and G. Evensen<sup>2</sup>

<sup>1</sup>Plymouth Marine Laboratory, Prospect Place, West Hoe, Plymouth, PL1 3DH, UK

<sup>2</sup>Nansen Environmental and Remote Sensing Centre, Edvard Griegsvei 3a, N-5037 Solheimsviken, Norway

Received: 6 September 2001 – Revised: 21 March 2002 – Accepted: 27 March 2002

**Abstract.** The purpose of this paper is to examine the use of a complex ecosystem model along with near real-time in-situ data and a sequential data assimilation method for state estimation. The ecosystem model used is the European Regional Seas Ecosystem Model (ERSEM; Baretta et al., 1995) and the assimilation method chosen is the Ensemble Kalman Filter (EnKF). Previously, it has been shown that this method captures the nonlinear error evolution in time and is capable of both tracking the observations and providing realistic error estimates for the estimated state. This system has been used to assimilate long time series of in-situ chlorophyll taken from a data buoy in the Cretan Sea. The assimilation of this data using the EnKF method results in a marked improvement in the ability of ERSEM to hindcast chlorophyll. The sensitivity of this system to the type of data used for assimilation, the frequency of assimilation, ensemble size and model errors is discussed. The predictability window of the EnKF appears to be at least 2 days. This is an indication that the methodology might be suitable for future operational data assimilation systems using more complex three-dimensional models.

**Key words.** Oceanography: general (numerical modelling; ocean prediction) – Oceanography: biological and chemical (plankton)

## 1 Introduction

Data scarcity and model inaccuracies are at the base of the limited predictability of oceanic systems. The skill of the forecast depends on the quality of the initial conditions and the model. Hence, observed data can be used to either validate the output of models or to set the initial conditions and

provide regular updates to state variables. A data assimilation system allows for the creation of a suitable initial condition for forecasts, for example, an initial condition, which contains the “optimal representation” of a state variable field using both observational data and models.

The application of data assimilation that we are concerned with is state estimation. In this case, one considers the problem of estimating the model state over some time period by simultaneously extracting a maximum amount of information from both observations and a dynamical model. Data assimilation methods have been developed previously for state estimation, both in physical oceanography and meteorology, for use in operational forecasting and monitoring systems. Recently, the focus has begun to turn towards biological state estimation. The number of publications and applications of data assimilation in physical oceanography is huge, and there are several established techniques being used with models of different complexity. A comprehensive review of data assimilation methods is beyond the scope of this paper. However, we can distinguish between the methods traditionally implemented for linear model dynamics, e.g. the Kalman Filter (see Kalman (1960) and Kalman and Bucy (1961) for the original works) and the adjoint technique, first properly introduced to oceanography and meteorology by Talagrand and Courtier (1987) and Courtier and Talagrand (1987). The application of these techniques to work with nonlinear model dynamics has met with variable success. For example, the Extended Kalman Filter applies a linearisation for the error covariance evolution, which may become unstable for some problems (Evensen, 1992). A similar problem arises with the adjoint method, since a tangent linear approximation constrains the length of the assimilation time interval which can be used (Miller et al., 1994). More recently, there has been the development of significant new assimilation formulations and techniques which have been specifically tailored towards

nonlinear dynamical models. The Ensemble Kalman Filter (EnKF) is one such method and is used in this study. It has been chosen based on its ability to predict error statistics for strongly nonlinear systems (see Evensen (1994; 1997a)) and for its simplicity and numerical efficiency. The model parameters will be kept fixed with predefined values. Thus, the model serves as a non-perfect estimate of the true evolution of the biological system (in some statistical sense), and thus, contains errors which are allowed for in the EnKF. Only the prognostic variables, i.e. concentrations of nutrients, phytoplankters, bacteria, detritus and zooplankters, will be estimated. As far as we know, this is one of only a handful of papers discussing state estimation in biological models using data assimilation. Natvik et al. (2001) used a weak constraint inverse method with the 0-dimensional model of Evans and Parslow (1985). Eknes and Evensen, 2000 demonstrated that use of the EnKF in data assimilation experiments with a one-dimensional three-component ecosystem model might be suitable for future operational data assimilation systems using more complex three-dimensional models. Anderson et al. (2000) have made successful state estimations via the assimilation of both physical and biological data into a time-evolving, mesoscale-resolution three-dimensional (3-D) ocean model using optimal interpolation. Ocean colour has been assimilated into a 3D biogeochemical model of the North Atlantic using the EnKF (Natvik and Evensen submitted, a, b) and with a Singular Evolutive Extended Kalman (SEEK) filter (Carmillet et al., 2002). In each of these cases, the biological model used is derived from the nitrogen-based plankton dynamics model of Fasham et al. (1990).

A major aim of the EC MAST-III funded Mediterranean Forecast System Pilot Project (MFSP) was to use in-situ ecological data in conjunction with deterministic hydrodynamic and ecosystem models, to develop predictive capability for phytoplankton biomass in the Mediterranean Sea. The aim of this paper is to develop a data assimilation system for an ecosystem model, which is suitable for use in an operational context, and hence, to assess its forecast capability and predictability window. In this paper, a coupled hydrodynamic/ecosystem water column model of the Cretan Sea (Allen et al., 2002; Tryantafyllou et al., 2002) has been used along with the EnKF in data assimilation experiments. One reason for choosing the EnKF to assimilate data into an ecosystem model is its ability to control the evolution of the whole model state when only one variable has been assimilated (Eknes and Evensen, 2002). This property is a consequence of the EnKF providing a consistent, multivariate analysis scheme based on ensemble statistics that include information about cross correlation between different model variables.

In the following sections, the ecosystem model is described and then the EnKF and the generic algorithm for the EnKF method are briefly explained. We then present results from hindcast experiments where in-situ chlorophyll and nitrate data from a buoy deployed in the Cretan Sea during late winter and spring 2000 are used in assimilation experiments. Finally, sensitivity analysis regarding the ensemble size, as-

simulation frequency and levels of model errors is performed.

### 1.1 Brief description of the ecosystem dynamics of the Cretan Sea

The hydrological structure in the Cretan Sea is characterised by intense mesoscale activity (Georgopoulos et al., 2000), which is not necessarily seasonally driven. Its circulation is determined by the combined effect of two gyral features, an anticyclonic eddy in the west and a cyclonic eddy in the east (Georgopoulos et al., 2000; Theocharis et al., 1999). The surface waters are dominated by Modified Atlantic Waters (MAW). Beneath the MAW lies the Cretan Intermediate Water (CIW) and the Transient Mediterranean Water (TMW). TMW is an important water mass and is characterised by low temperature (14°C) and salinity (38.7 psu). It intrudes into the Cretan Sea via the eastern Kassos and western Antikithira straits and occupies the intermediate layer 200–600 m (Balopoulos et al., 1999). The presence of the TMW is rather variable and is associated with the outflow of Cretan Deep Water (CDP) into the eastern Mediterranean (Souvermezoglou et al., 1999). Since TMW is old water, it is characterised by high nutrient and low oxygen concentrations (Souvermezoglou et al., 1999).

The intensity of the eddy dipole in the Cretan Sea increases in late winter and the TMW downwells west of the domain (to 500–900 m) and upwells nutrients into the euphotic zone in the center of the cyclone (from 20–600 m). This process can be very important ecologically as it initiates small-scale phytoplankton blooms. Intense mixing occurs in early spring and the euphotic zone is re-supplied with nutrients from deep waters. However, phytoplankton biomass remains at relatively low levels due to phosphate limitation (Krom et al., 1991; Thingstad and Rassoulzadegan, 1995). During the spring, summer and autumn, the Cretan Sea is stratified. It exhibits an oligotrophic ecosystem which is characterized by a food chain composed of very small phytoplankton cells and a microbial loop, both of which have a negative effect on energy transfer (carbon and nutrients) to the deeper water layers and the benthos. The very low nutrient concentrations found in the Eastern Mediterranean, in conjunction with the prevailing hydrographic circulatory patterns, are the main factors responsible for maintaining low phytoplankton standing stock and surface primary production levels observed (Azov, 1986; Becacos-Kontos, 1977; Berman et al., 1984).

The Cretan Sea exhibits substantial interannual variability. It undergoes periodically events of intense mixing, which causes higher nutrient availability and may effect dramatic changes in the productivity of the area. This may cause the entire system to respond by shifting its food web structure from a microbial ecosystem to a mesotrophic one, which is characterised by larger phytoplankton and higher sedimentation rates and, therefore, increases energy transfer to the deeper water layers and the benthos.

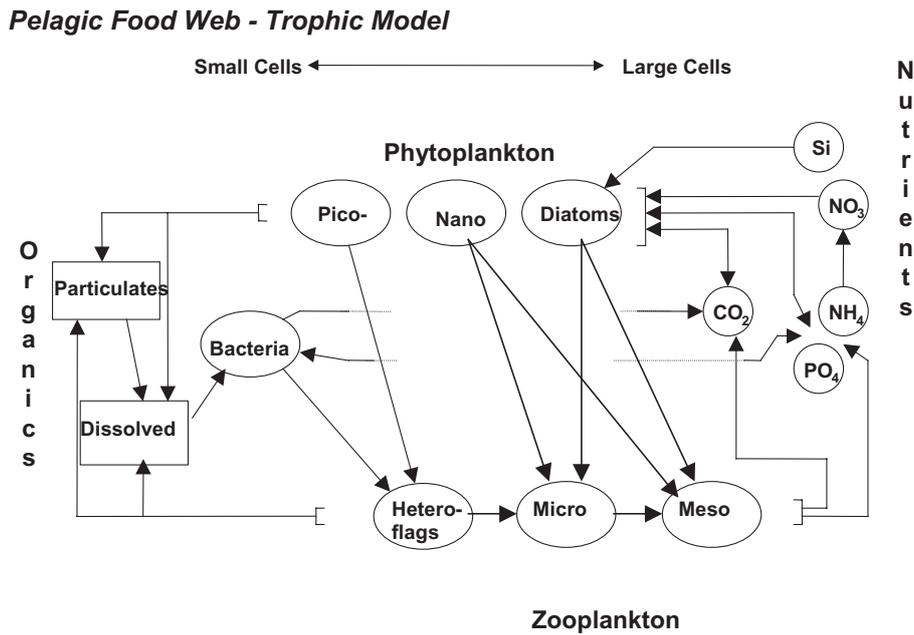


Fig. 1. Schematic diagram of the pelagic food web of the ERSEM model.

## 2 Description of the model

The model used in this study is derived from the highly portable coupled physical biogeochemical water column model described in Allen et al. (1998). It is a synthesis of a one-dimensional version of the Princeton Ocean Model (Blumberg and Mellor, 1988) and the European Regional Seas Ecosystem Model (ERSEM; Baretta et al., 1995). The portability of the model comes from the generic nature of ERSEM, whereby the chemical and biological processes included are non-site specific. All site dependence comes from the physical environment in which they are placed.

### 2.1 The physical model

A turbulence closure model (Mellor and Yamada, 1992) determines the vertical turbulent kinetic energy, temperature and diffusion coefficient profiles generated by a surface wind stress, and heat flux. Vertical diffusion processes are assumed to control heat transfer. A background viscosity is used to parameterise other mixing processes (e.g. internal waves).

### 2.2 ERSEM

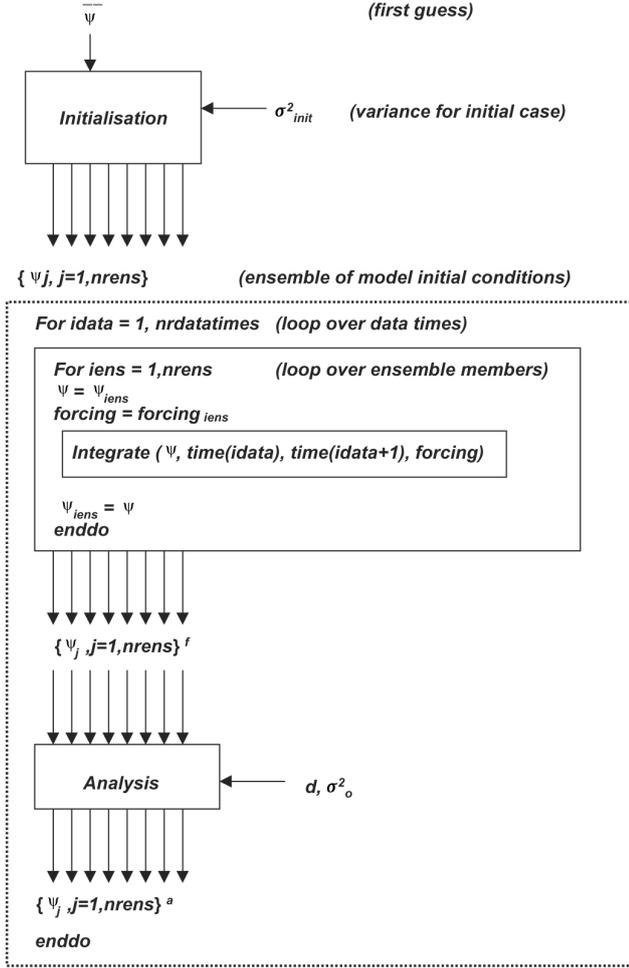
The ecosystem described in ERSEM is considered to be a series of interacting physical, chemical and biological processes that together exhibit coherent system behaviour (Baretta et al., 1995). State variables have been chosen in order to keep the model relatively simple without omitting any component that exerts a significant influence upon the energy balance of the system. ERSEM uses a “functional” group approach to describe the ecosystem, whereby biota is grouped together according to trophic level (subdivided ac-

ording to feeding method or size). Biological functional group dynamics are described by both physiological (ingestion, respiration, excretion and egestion) and population processes (growth, migration and mortality).

A schematic diagram of the pelagic food web is shown in Fig. 1. A detailed description of ERSEM and its sub-models is beyond the scope of this paper. Further details can be found in Baretta et al., 1995; Baretta-Bekker et al., 1997; Broekhuzien et al., 1995 and Ebenhöh et al., 1997. The phytoplankton pool is described by three functional groups based on size and ecological function as follows; diatoms (P1), size class 20–200  $\mu\text{m}$ , eaten by micro and mesozooplankton; autotrophic flagellates (P2), size class 2–20  $\mu\text{m}$ , eaten by micro and mesozooplankton; picoplankton (P3), size class 0.2–2  $\mu\text{m}$ , eaten by heterotrophic nanoflagellates. The potential carbon assimilation rate is not dependent on the nutrient limitation, due to intracellular N and P shortage or external nutrient concentrations. Rather, in the case of intracellular nutrient shortage, the unutilized assimilation products are excreted as dissolved organic carbon. The uptake and loss of carbon (C) by a phytoplankton class is described by the following word equation

$$\frac{\partial C_{i=1-3}}{\partial t} = (\text{Assimilation} - \text{Excretion} - \text{Respiration} - \text{Lysis} - \text{Sedimentation}) - \text{Grazing} + \text{Vertical transport} \quad (1)$$

where the assimilation of carbon is a function of temperature and light limitation. The organic carbon specific photosynthesis rate ( $P_c$ ) is expressed as a function of the irradiance (I), and the ratio  $\theta$  (chlorophyll-a/carbon) as follows (Geider et al., 1995; Taylor et al., 1997). The carbon is determined



**Fig. 2.** Schematic figure showing how the EnKF methodology is in relationship to an arbitrary model.  $Nrens$  is the ensemble size,  $iens$  is the number of the ensemble being integrated and is the data used for assimilation.

from Eq. (1) and the chlorophyll-a from Eq. (3)

$$P_c = P_m \left[ 1 - \exp\left(-\frac{I\theta\alpha_{chl}}{P_m}\right) \right], \quad (2)$$

where  $\alpha_{chl}$  is the initial slope of the photosynthesis irradiance curve normalised to chlorophyll a ( $\alpha_{chl} = 0.0000075 \text{ g C (g Chl)}^{-1} (\mu\text{mol phot})^{-1}$ ), and  $P_m$  is the light saturated rate of photosynthesis normalised to carbon and is a multiplicative function of the maximum photosynthetic rate and the temperature dependence. The uptake and loss of chlorophyll-a ( $Chl$ ) is given by

$$\frac{\partial Chl_{i=1-3}}{\partial t} = \alpha_{chl} P_c Chl_{i=1-3} - (\text{Excretion} + \text{Respiration} + \text{Lysis} + \text{Sedimentation}) Chl - \text{Grazing} + \text{Vertical transport} \quad (3)$$

$\alpha_{chl}$  is assumed to be regulated by the ratio of achieved to

maximum potential photosynthesis (Gieder et al., 1995)

$$\alpha_{chl} = \theta_m \left( \frac{P_c}{\alpha_{chl} I \theta} \right), \quad (4)$$

where  $\theta_m$  is the maximum chlorophyll-a: carbon ratio observed in cells acclimatized to extremely low light ( $\theta_m = 0.05$ ).

The loss terms due to excretion, lysis and sedimentation are enhanced when the plankton is nutrient limited. Thus, the production of dissolved organic carbon (DOC) is increased under these conditions. The uptake of nutrients ( $\text{NO}_3$ ,  $\text{NH}_4$  and  $\text{PO}_4$ ) has been de-coupled from the carbon assimilation processes by including dynamic nutrient kinetics according to Droop, 1974 and Nyholm, 1977. Nutrient uptake is dependent on the external nutrient concentrations and on the level of intercellular storage.

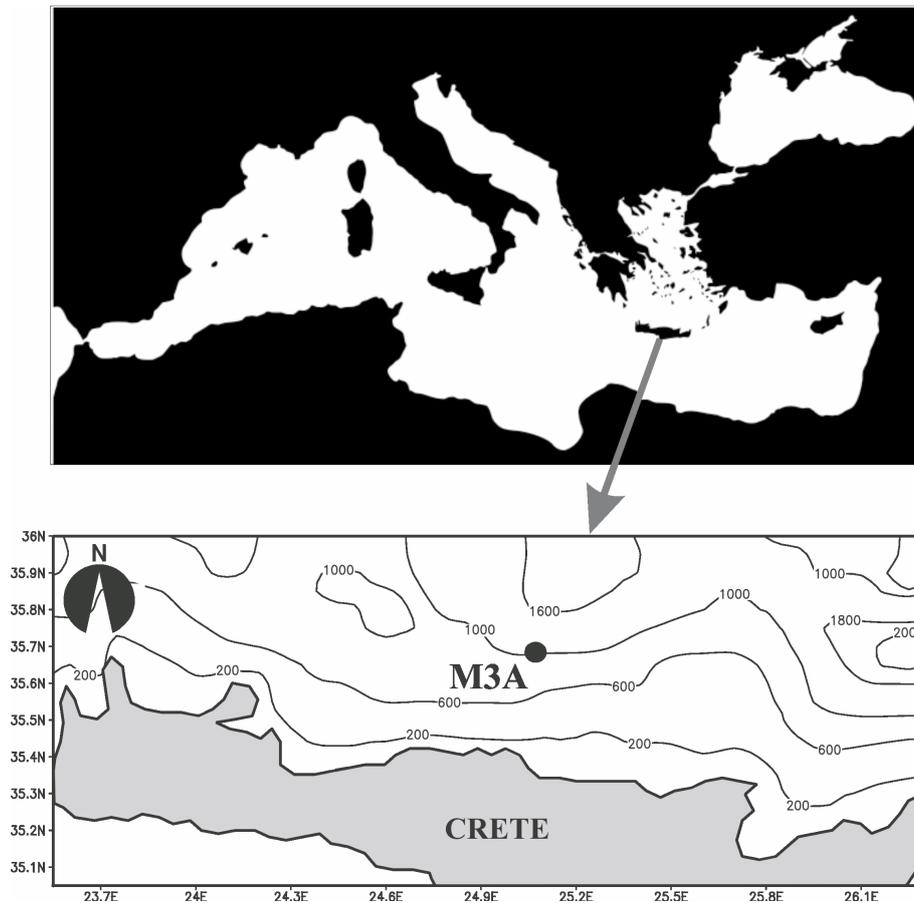
The heterotrophic (bacteria and grazers) components of the model are defined as follows (Baretta-Bekker et al., 1997; Broekhuizen et al., 1995): Bacteria act to consume DOC, decompose detritus and can compete for inorganic nutrients with phytoplankton. Heterotrophic flagellates (Z6) feed on bacteria and picoplankton, are grazed by microzooplankton and are cannibalistic. Microzooplankton (Z5) consume autotrophic, heterotrophic flagellates, and diatoms, are grazed by mesozooplankton and are cannibalistic. Mesozooplankton (Z4) consume diatoms, autotrophic flagellates and microzooplankton and are cannibalistic.

### 2.3 The Ensemble Kalman Filter formulation

The Ensemble Kalman Filter (EnKF) method was formulated with nonlinear dynamics in mind and the emphasis was focused on deriving a method which could properly handle the error covariance evolution in nonlinear models (Evensen, 1994). The method has been used successfully with a number of different dynamical models, from the simple but highly nonlinear and chaotic Lorenz equations (Evensen, 1997a), to ocean circulation models (Evensen and van Leeuwen, 1996; Evensen, 1997b). Statistical noise dominates the errors in the EnKF, and there are no closure problems or unbounded error variance growth, as have been found with assimilation methods relying on the use of a tangent linear model.

The EnKF uses an ensemble of model states, which are generated by integrating the model equations forward in time. The ensemble size is normally of the order of 100 members. Each individual member is integrated as a stochastic differential equation, i.e. forced with a random noise component, which represents the model errors. It can be shown that such ensemble integration becomes identical to a Markov Chain Monte Carlo (MCMC) method for solving the Fokker Planck equation for the evolution in time of the probability density of the model state. Since the full nonlinear dynamical model is used, the only approximation associated with this approach is that a finite number of members are used in the ensemble.

The ensemble is integrated forward in time until measurements are available. At these points, an analysis scheme is



**Fig. 3.** Map showing the Mediterranean Sea and the site of the M3a buoy.

used to update or correct the model state in a statistically consistent way. The updated state can be considered as the model forecast plus a number of influence functions, one for each of the measurements. These are multivariate statistical functions computed from the ensemble statistics, i.e. the cross correlations between the different variables in the model are included. Thus, a change in one of the model variables will influence the other variables.

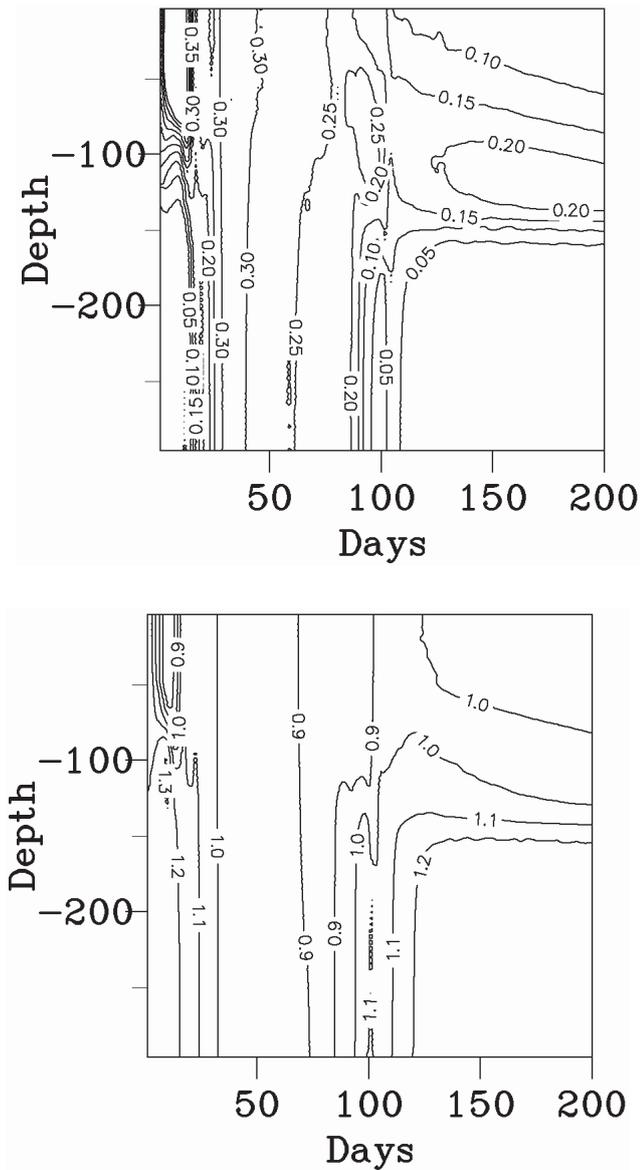
This analysis is based on estimates of the error statistics for the model forecast and the measurements, and it minimises the error variance of the analysed estimate in a least-squares sense. A consequence of this is that the analysis is repeated for each of the members in the ensemble, and the resulting analysed ensemble then has the correct error statistics for the analysed state. Thus, there is no need for resampling to create a new ensemble for the continuing integration. A schematic illustration of the algorithm is given in Fig. 2 and a detailed presentation of the EnKF implementation is given in Burgers et al. (1998). The figure defines three boxes, i.e. an initialisation box, an integration box and an analysis box. In the initialisation box, a first guess model state is used to create an ensemble of size  $n_{\text{ens}}$  by adding pseudo-random noise with prescribed statistics to the first guess initial state. These perturbed model states are then integrated forward in time

until the time when the first observation set becomes available. This is done in the integration box, which integrates each model state forward in time over a specified time interval. Then each of the members are input to the analysis box, which computes a new ensemble of analysed model states based on the ensemble statistics, the observations and their error statistics. The analysed ensemble is then integrated forward to the next data time and the process is repeated. The EnKF estimate is usually defined to be the ensemble mean.

By using object-oriented programming, a generic EnKF has been developed. Thus, one has to specify an object which defines a full model state and a subroutine for the model integration, and most models can be included in the generic implementation, since the analysis scheme operates on objects of full model states.

## 2.4 Data

The data used for assimilation comes from the M3a buoy which was deployed in the Cretan Sea at 35° N 26° E (Fig. 3) from January to July 2000 as a part of the EU MFSPP program. The M3a was designed as an open-ocean monitoring system capable of providing continuous information for physical and biochemical parameters in the upper thermocline and air-sea interaction parameters at the sea-surface



**Fig. 4.** Time evolution of (a) chlorophyll ( $\text{mg m}^{-3}$ ) and (b) nitrate ( $\text{mmol m}^{-3}$ ) with depth for the reference run.

level. The measurements are grouped in the following categories: physical parameters in the upper 500 m of the ocean: temperature and salinity at  $-1$ ,  $-30$ ,  $-50$ ,  $-75$ ,  $-100$ ,  $-150$ ,  $-250$ ,  $-350$ ,  $-500$  m, current speed and direction:  $0$ – $500$  m with approximately a  $30$  m spacing; biochemical parameters in the upper  $100$  m of the ocean: dissolved oxygen, chlorophyll, PAR, turbidity at  $-1$ ,  $-30$ ,  $-50$ ,  $-75$ ,  $-100$  m, nitrate at  $-40$  m; and air-sea interaction parameters at sea surface: air temperature, wind speed and direction, air pressure, relative humidity, wave height and direction. Measurements were made at three-hourly intervals.

Chlorophyll-a, dissolved oxygen and nutrients were post-calibrated using in-situ bottle measurements. For the calibration, water samples were collected at the same depths as the sensors during each maintenance cruise, once when the

instruments were recovered and once when they were redeployed 2–3 days later. Calibration coefficients were then calculated separately for the period between cruises, using reference data collected during each redeployment when all sensors have been cleaned (Nittis et al., 2002). Depth profiles of nitrate, phosphate, ammonia and silicate were measured during the maintenance of the buoy.

## 2.5 Experimental set up

The model has been set up as the surface  $300$  m of a  $1000$  m deep water-column, with a  $4$  m depth interval for the site of the M3A buoy. The model has been initialised where possible (temperature, salinity, nutrients) with in-situ measurements. The initial conditions for other biological variables were taken from previous simulations of the region (Allen et al., 2002). Due to substantial gaps in the air-sea interaction data sets, the model was forced with 6-hourly ECMWF analysis data taken from the MFSP project website.

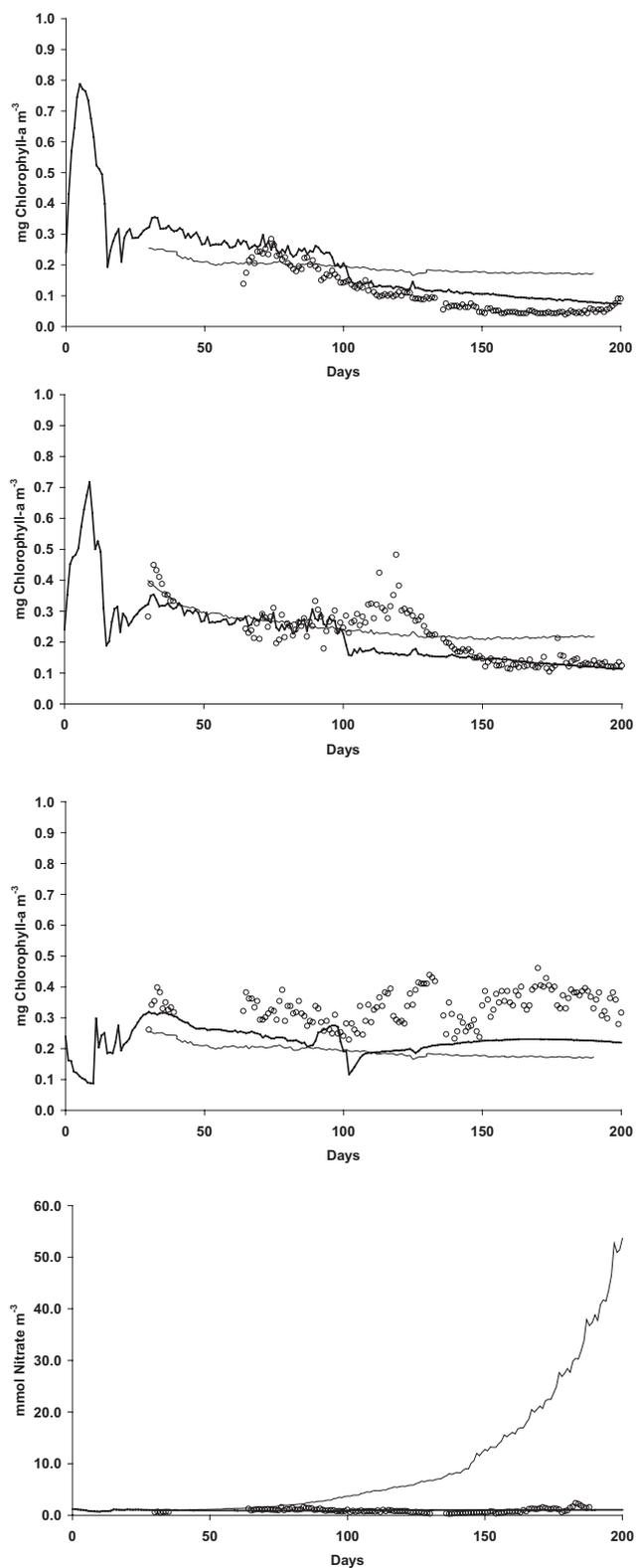
## 3 Results

### 3.1 Reference solution

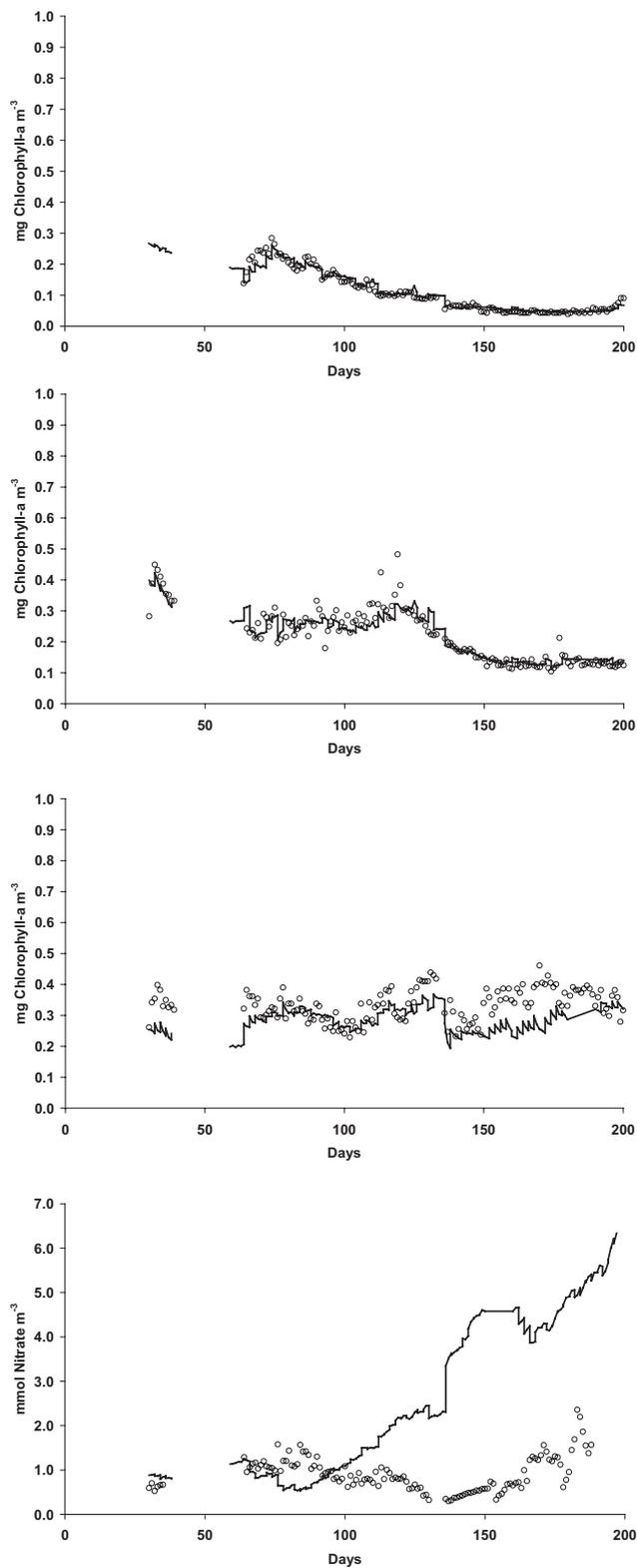
In this case, the reference solution is the standard run of the ERSEM model over the data period ( $0$ – $200$  days). The reference solution for chlorophyll and nitrate is shown in Fig. 4, and the fit with data in Fig. 5. The model system has an initial bloom to a depth of about  $100$  m for the first  $30$  days, followed by a period of deep mixing, where plankton growth is light limited due to high vertical mixing rates. Around day  $90$ – $100$ , the system begins to stratify and a deep chlorophyll maxima (DCM) forms at around  $80$  m. The DCM is maintained for the rest of the period of simulation and deepens to around  $120$  m by day  $200$ . The nitrate simulation shows depletion in the surface layer during the early bloom, followed by an intense period of mixing between days  $30$  and  $90$ . After the onset of stratification (day  $90$ ), a nutricline forms just below the DCM, which is eroded as the DCM deepens. It should be noted that phytoplankton growth is predominantly phosphate limited in the Cretan Sea (Tselephides et al., 2000). The reason we primarily consider nitrate and not phosphate in this paper is because we only have nitrate data available as a high frequency time series.

### 3.2 Generation of the initial ensemble

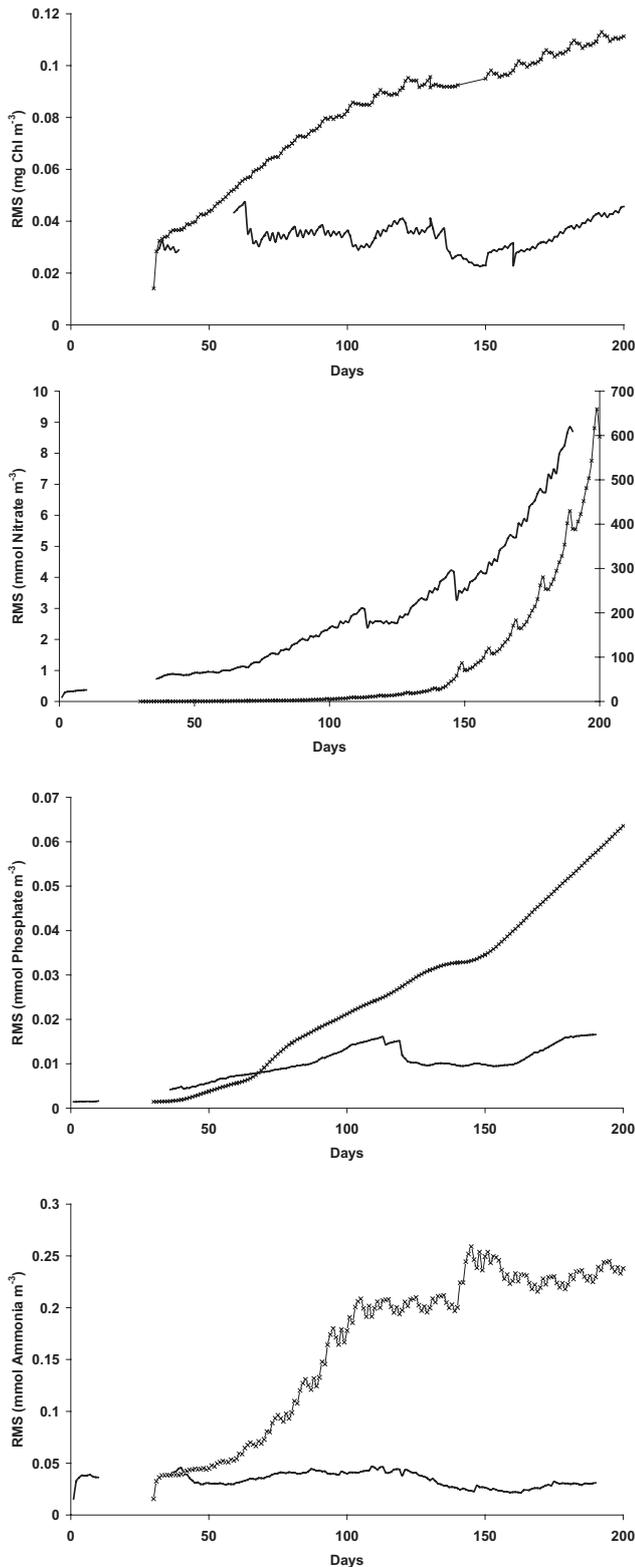
The initial conditions for the ensemble run were taken to be the same as for the reference run and were regarded as a best guess estimate of the truth. An ensemble of model states was then generated by adding a pseudo-random field to the “best guess estimate”. In these experiments, we used an ensemble size of  $120$ . The choice of ensemble size is discussed in Sect. 4.2.



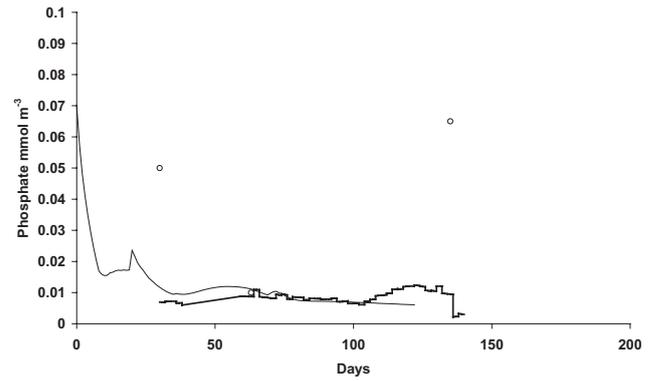
**Fig. 5.** Reference solution (—) for chlorophyll-a at (a) 40 m, (b) 65 m and (c) 115 m and nitrate at 45 m at the M3a site, along with data (o) and the climatology (—).



**Fig. 6.** EnKF estimate (—) for assimilation of *in-situ* chlorophyll-a every two days at (a) 40 m, (b) 65 m and (c) 115 m and nitrate at 45 m at the M3a site, along with data (o).



**Fig. 7.** Root mean square error for the ensemble climatology (x-x) and the 2-day EnKF assimilation (—) for (a) chlorophyll, (b) nitrate, (c) phosphate and (d) ammonium. Nb, the scale for RMS for the nitrate climatology, is on the right side of graph b.



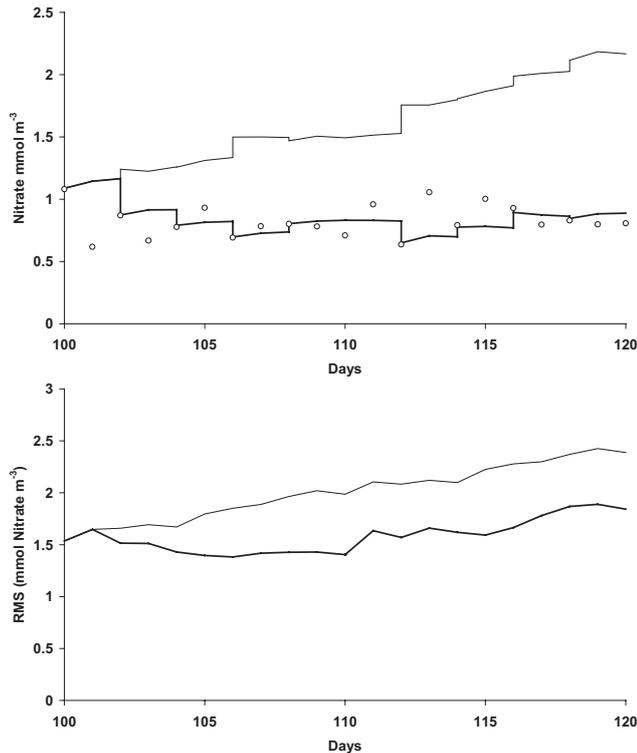
**Fig. 8.** EnKF solution (—) with chlorophyll-a assimilated every two days and the reference solution (—) for phosphate at 65 m at the M3a site, along with in-situ data (o).

### 3.3 Climatology

A pure ensemble simulation with no influence from the data is referred to as the climatology. It is defined as the ensemble mean resulting from the ensemble being integrated forward in time for the period of the simulation. The variance or uncertainty within this climatology is determined by the spread of ensemble members around the mean. This simulation is made to enable us to examine the impact of the assimilation of data by comparison with the EnKF estimate. Figure 5 shows the climatological run for chlorophyll at 40, 65 and 115 m and for nitrate at 45 m, along with the fit with data. For chlorophyll, the climatological ensemble mean is in the vicinity of the data at all three depths, while for nitrate it deviates substantially after day 100. The drift occurred primarily in the surface mixed layer after the onset of stratification. This may be a consequence of the very non-gaussian nature of the ensemble generated for the climatology and that the mean for nitrate is significantly larger than it would be if the ensemble was truly Gaussian. Initially, we assume that the ensemble is Gaussian, but since the model is nonlinear it may become increasingly non-gaussian as the integration proceeds. The most likely cause for the nitrate climatology to greatly deviate from the reference solution, the data and the EnKF estimate, is that the ensemble spreads out during the integration, thereby making the errors in the estimate larger. Why this deviation is larger for nitrate than for chlorophyll-a is not known. It may be that nitrate is more sensitive to perturbations during the integration than the other variables.

### 3.4 Assimilation of *in-situ* chlorophyll at three depths every 2 days

To demonstrate the efficacy of the EnKF assimilation system, a hindcast simulation has been made of the M3a buoy for the first 200 days out of 2000. Reference and climatological simulations were made, and then an analysis simulation was performed whereby in-situ chlorophyll-a data was assimilated every 2 days when available. The results of these

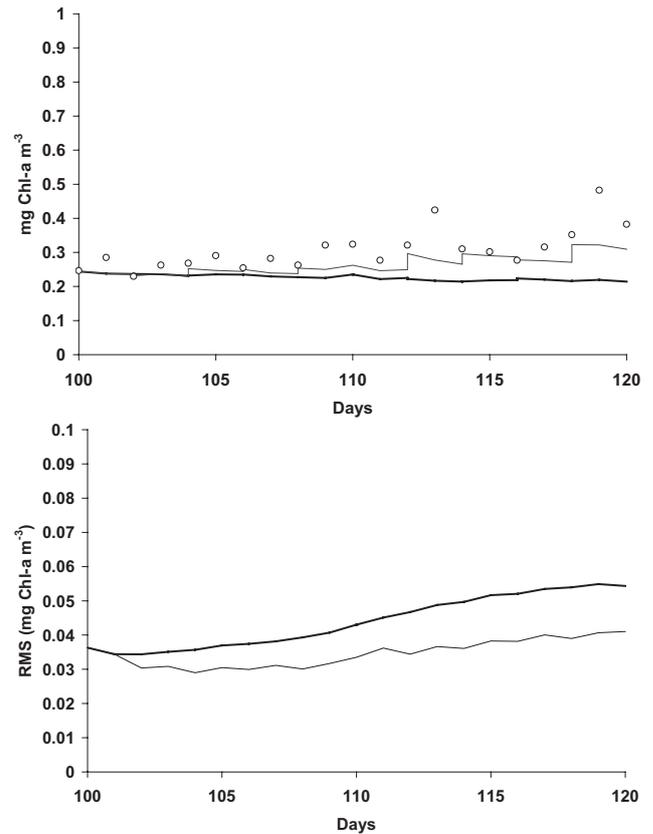


**Fig. 9.** (a) Effect of assimilation of chlorophyll (—) and nitrate (—) compared with data (o), on the nitrate simulation at 45 m, (b) effect of assimilation of chlorophyll (—) and nitrate (—) on the RMS.

analysis simulations for chlorophyll at 40 m, 65 m, 115 m and nitrate at 45 m are shown in Figs. 6 and 7. The observational error used was 10%, the error in the initial conditions 15% and the error in the model was taken to be 15% per day until day 130, then 20% per day (10% for DOC and POC) afterwards. These error values were chosen to optimise the data fit to chlorophyll for the analysis simulation.

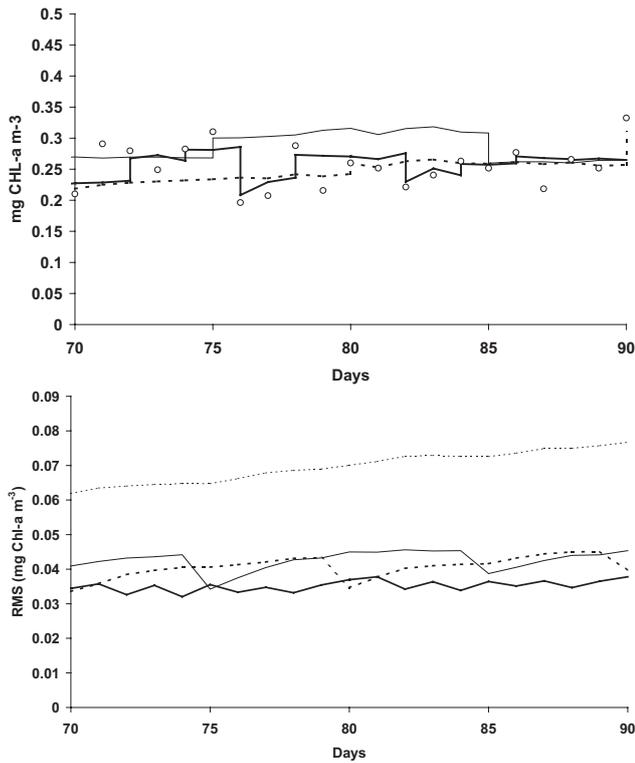
A comparison of the analysis simulations with observations is shown in Fig. 6. The assimilation of chlorophyll into the model results in a marked improvement in the quality of the hindcast compared with the climatology and the reference. In the upper water column (40 m, 60 m), the model tracks the data almost exactly. At 115 m, the simulation is not as good, particularly after day 120, but still represents a marked improvement over the climatology and reference solutions. The effects of chlorophyll assimilation on nitrate is not as good, with the simulation dragged away from the data, but still represents a marked improvement over the climatology (Fig. 5).

The root mean square (RMS) is calculated as the depth integrated root mean square of the different (ensemble) variances at each point in time. RMS estimates for climatological, chlorophyll and nitrate are shown in Fig. 7. The RMS for the chlorophyll climatology increases steadily until day 100, when the ecosystem is under hydrodynamic control and



**Fig. 10.** (a) Effect of assimilation of chlorophyll (—) and nitrate (—) compared with data (o), on the chlorophyll simulation at 65 m, (b) effect of assimilation of chlorophyll (—) and nitrate (—) on the RMS.

then the slope lessens after that. The RMS for the EnKF solution for chlorophyll shows that when data is assimilated, the variance of the ensemble (i.e. the uncertainty in the ensemble) is decreased, and that it increases slowly in between measurement times. The increase is caused by a combination of stochastic model errors and the nonlinear instabilities inherent in the model dynamics. One of the strengths of the EnKF is that it can handle these nonlinear and unstable regimes in a consistent way. The RMS for nitrate shows similar behaviour, with the main difference being that the errors increase faster than for chlorophyll after each measurement, resulting in a net upward trend in the RMS error. The RMS plots for ammonia and phosphate are also shown (Fig. 7) and demonstrate that the EnKF is controlling the evolution of the whole model state when only one variable has been assimilated. The effect of assimilating chlorophyll upon the phosphate simulation is shown in Fig. 8. It can be seen that the assimilation pulls the phosphate closer to the in-situ phosphate observations made when the M3a buoy undergoes maintenance. This further demonstrates that the EnKF has the ability to constrain the evolution of nonassimilated variables. In making this conclusion, it has to be noted that the observed phosphate concentrations are very close to the de-

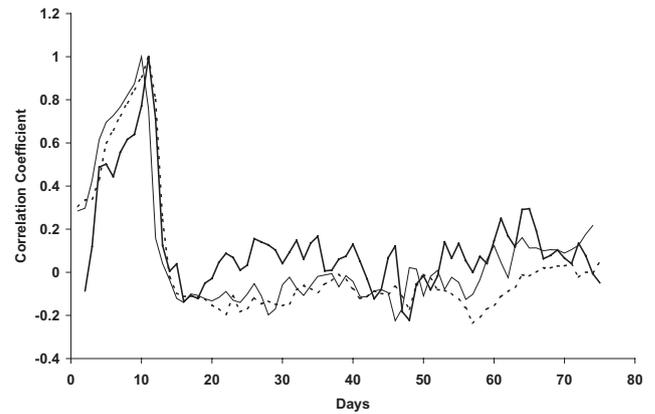


**Fig. 11.** (a) Assimilation of chlorophyll data at 2 (—), 5 (---) and 10 (· · ·) day intervals upon the chlorophyll simulation at 65 m depth compared with data (○), (b) assimilation of chlorophyll data at 2 (—), 5 (---) and 10 (· · ·) day intervals upon the RMS compared with the climatology (· · ·).

tection limit of the sampling method ( $0.03\text{--}0.04\text{ mmol m}^{-3}$ ; M. Woodward pers. com.) and that modelled concentrations are extremely low.

### 3.5 Assimilation of nitrate

The effects of assimilating nitrate every 2 days rather than chlorophyll, on the hindcasting of nitrate and chlorophyll, are shown in Figs. 9 and 10 for a twenty-day period from day 100 to 120. This part of the simulation was chosen because it marks the point at which the climatological simulation for nitrate starts to deviate markedly from the data. The assimilation of nitrate leads to a marked improvement in the fit with data and a substantial reduction in the RMS error between the two simulations. The effect on the chlorophyll simulation is a degradation of the data fit, as expected, compared to that when chlorophyll is assimilated, but it still shows the same trends. The RMS indicates that while the data fit is not as good as when chlorophyll is assimilated, it still represents a substantial improvement in the simulation over the climatology. These results suggest that a nutrient such as nitrate may provide a better overall constraint on the ecosystem dynamics than chlorophyll. This needs to be further investigated to establish which variable or combination of variables is optimal to obtain the best forecast for the whole ecosystem.



**Fig. 12.** Correlation between nitrate at 65m and nitrate in any box, for an ensemble size of 50 (—), 90 (---) and 150 (· · ·) members. Data was assimilated once after 32 days.

## 4 Sensitivity analysis

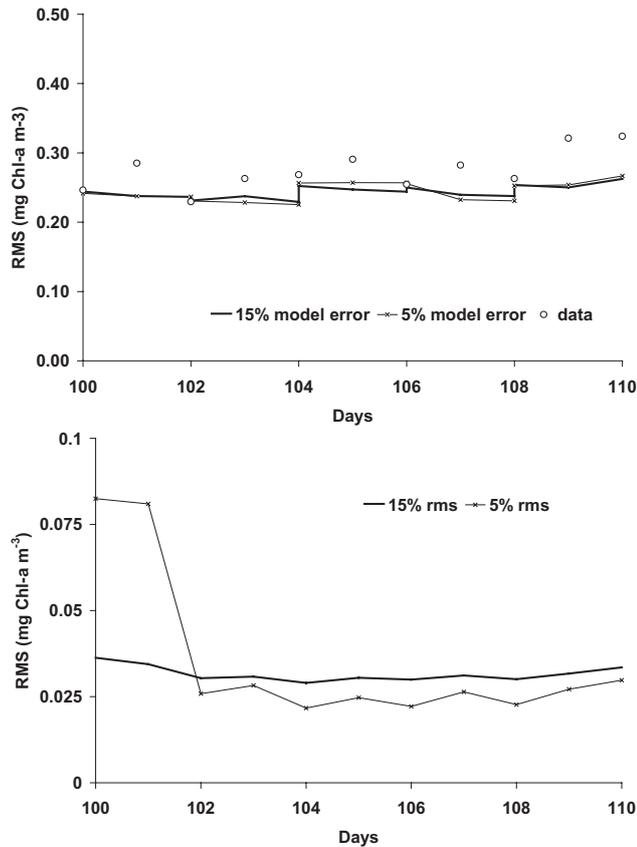
### 4.1 Sensitivity to assimilation frequency

The sensitivity of the EnKF assimilation to the measurement frequency is shown in Fig. 11. In this case, chlorophyll measurements have been assimilated every 2, 5 and 10 days. In all cases, the EnKF estimate gives a reasonable approximation to the data, but the 2-day assimilation follows the trends the best, as expected. The RMS indicates that all three runs represent a marked improvement over the climatology and that the 2-day assimilation has the smallest error propagation. Since the magnitude of the RMS for the 5- and 10-day runs is similar, this would suggest that the predictive window for this particular model system probably lies between 2 and 5 days. The assimilation of nitrate (not shown) shows similar properties.

### 4.2 Sensitivity to ensemble size

The size of the ensemble is important because if it is too small, we will not be able to obtain the correct estimate of the error variances in the predicted model state; if it is too large, we will incur an unnecessary computational overhead. Previous works have found that an ensemble size of around 100 members is generally sufficient, but this has to be evaluated for every new assimilation system (Evensen and van Leeuwen, 1996; Evensen 1997a; Ekenes and Evensen, submitted).

In many cases, only a few members are sufficient to provide reliable estimates of the true mean values and variances. For a proper description of the covariances, however, a considerably larger number is needed. According to EnKF theory, the error of an estimate of the covariance decreases in a rate proportional to  $1/\sqrt{N}$ , where  $N$  is the number of ensemble members (Evensen, 1994). In Fig. 12, the correlation, i.e. the covariance scaled by the individual variances, between nitrate in box 11 (44 m) and nitrate in any box is shown. From this figure, it is clear that 50 members is too



**Fig. 13.** The influence of the size of the error in the model upon (a) chlorophyll assimilation and (b) the RMS.

few to provide a proper representation of the true error covariance, as the correlation fields corresponding to 50 and 90 members are very different. The estimate of the correlations are, however, quite consistent for the 90-member ensemble. Increasing the ensemble size to 150 does not change the correlation estimates drastically. Thus, it seems as if an ensemble of around 100 members (say between 90 and 150) is sufficient to provide a proper description of the correlations and estimates in our case therefore, we chose an ensemble of 120 members for our experiments, and believe that this is a sufficient size to demonstrate that the data assimilation system used with ERSEM works. However, we are aware that the estimates (results) would benefit from a slightly larger ensemble for any operational purpose. Any precision can be achieved by increasing the ensemble size. Thus, the need to improve the quality of the estimates should be balanced by the numerical cost related to the storage and integration of more members.

#### 4.3 Sensitivity to size of the model errors

The EnKF requires a prior knowledge of the variances of the errors in both the model and the data. The choice of the error variances affects the assimilation results. For example, by choosing a small error variance in the data, we are

assuming that the observations are reasonably accurate and contain only small errors. Previously, twin experiments have demonstrated that the EnKF estimate is closer to the reference solution when the errors in the measurements are small and closer to the climatology when these errors are assumed large and/or if the dynamical errors are assumed small (Eknes and Evensen, 2002). In this case, the EnKF estimate will be close to the observations. We have assumed that the errors in the data are consistent at 10%, indicating a reasonable amount of faith in the accuracy of the data. Figure 13 shows the response of the system to a decrease in the error variance in the model (i.e. we assume that our faith in the model has improved). In this case, we compare the results for a 5% error with a 15% error. Reducing the model error marginally increases the fit with the data and reduces the RMS when compared with the model solution using a 15% error.

## 5 Summary

An Ensemble Kalman Filter has been used with a complex vertically resolved hydrodynamic marine ecosystem model ERSEM. This system has been used to assimilate long time series of in-situ chlorophyll taken from a data buoy in the Cretan Sea. The assimilation of in-situ chlorophyll data via the EnKF method results in a marked improvement in the ability of the ecosystem model to hindcast chlorophyll. The predictability window of the EnKF appears to be at least 2 days. This illustrates the importance of having access to high quality, high frequency data sets for assimilation, because without them, a long-term assessment of the hindcast potential of the EnKF could not have been achieved. The system has been shown to be sensitive to the type of data used for assimilation, the frequency of assimilation, and the size of the ensemble and model errors.

A great strength of the EnKF is its ability to control the evolution of the whole model state when only one variable has been assimilated (Eknes and Evensen, 2002). This property is a consequence of the EnKF providing a consistent, multivariate analysis scheme based on ensemble statistics that includes information about cross correlation between different model variables. Aspects of this can be seen when comparing the assimilation of nitrate with the assimilation of chlorophyll; in both cases it is apparent that the simulation of both nitrate and chlorophyll is better when assimilation takes place. It also indicates that the assimilation of nitrate may control the chlorophyll better than the opposite case when the assimilation of chlorophyll controls the nitrate.

Extensive further work is required to test and validate both systems. Ensemble sizes need to be optimised. The number of state variables included in the error covariance matrix needs to be optimised along with the model errors associated with them. In particular, the balance between computational requirements and accuracy needs to be evaluated before this system is implemented in a three-dimensional ecosystem model. The combination of data types for assimilation needs to be addressed, i.e. whether to assimilate physical and bi-

ological data together or separately. Additionally, we need to investigate which biological variable (or combination of variables) gives the best forecasts. This, in turn, requires a proliferation of in-situ data collection, both spatially and in terms of the variables sampled. The choice of variable for assimilation when forecasting phytoplankton may often be a case of identifying the limiting factors for growth (i.e. turbulence, turbidity, nutrients or predation). Additionally, we need to apply these assimilation systems in a spatial context to exploit the information obtainable from satellites (Sea Surface Height, Sea Surface Temperature, Ocean Colour) and to ascertain whether they can be used to significantly improve forecasts in both coastal regions and on a basin scale. The deployment of in-situ data collection buoys, along with data from Voluntary Observing Ship routes, will be invaluable for the validation of such systems.

*Acknowledgements.* This work has been supported by the EC MAST-III MFSPP project (MA53-CT98-0171). We would like to thank P. Ruardij (NIOZ) for his assistance in developing the assimilation system and G. Petihakis (IMBC) for the provision of Fig. 3.

Topical Editor N. Pinardi thanks two referees for their help in evaluating this paper.

## References

- Anderson, L. A., Robinson, A. R., and Lozano, C. J.: Physical and biological modeling in the Gulf Stream region: Data assimilation methodology, *Deep Sea Res.*, 47, 1787–1827, 2000.
- Allen, J. I., Somerfield, P. J., and Siddorn, J. R.: Primary and bacterial production in the Mediterranean Sea: a modelling study, *J. Mar. Sys.* (in press), 2002.
- Allen, J. I., Blackford, J. C., and Radford, P. J.: A 1-D vertically resolved modelling study of the ecosystem dynamics of the middle and southern Adriatic Sea, *J. Mar. Sys.*, 18, 265–286, 1998.
- Azov, Y.: Seasonal patterns of phytoplankton productivity and abundance in nearshore oligotrophic waters of the Levant Basin (Mediterranean), *J. of Plankton Research*, 8, 1, 41–53, 1986.
- Balopoulos, T. E. et al.: Major advances in the oceanography of the southern Aegean Sea – Cretan Straits system (eastern Mediterranean), *Progress in Oceanography*, 44, 109–130, 1999.
- Baretta, J. W., Ebenhöf, W., and Ruardij, P.: The European Regional Seas Ecosystem Model, a complex marine ecosystem model, *N. J. Sea Res.* 33, 233–246, 1995.
- Baretta-Bekker, J. G., Baretta, J. W., Hansen, A. S., and Riemann, B.: An improved model of carbon and nutrient dynamics in the microbial food web in marine enclosures, *Aquat. Microb. Ecol.*, 14, 91–108, 1997.
- Becacos-Kontos, T.: Primary production and environmental factors in an oligotrophic biome in the Aegean Sea, *Marine Biology*, 42, 93–98, 1977.
- Berman, T., Azov, Y. and Townsend, D.: Understanding oligotrophic oceans: can the eastern Mediterranean be a useful model, in: Holm-Hansen, O., Bolis, L., and Giles, R. (Eds.): *Marine Phytoplankton and Productivity*, Springer Verlag, Berlin, 1984.
- Blumberg, A. F., and Mellor, G. L.: A coastal ocean numerical model, In *Mathematical Modelling of Estuarine Physics*, Proceedings of the International Symposium, Hamburg, Aug. 1978 (Sunderman, J. and Holtz, K. P., eds.), Springer-Verlag, Berlin, 203–214, 1988.
- Broekhuizen, N., Heath, M. R., Hay, S. J., and Gurney, W. S. C.: Modelling the dynamics of the North Sea's mesozooplankton. *N. J. Sea Res.* 33, 381–406, 1995.
- Burgers, G., van Leeuwen, P. J., and Evensen, G.: Analysis scheme in the ensemble Kalman filter, *Mon Weather Rev.*, 126, 1719–1724, 1998.
- Carmillet, V., Brankart J.-M., Brasseur, P., Drange, H., and Evensen, G.: A singular evolutive extended Kalman filter to assimilate ocean color data in a coupled physical-biochemical model of the North Atlantic, *Ocean Modelling* (in press), 2002.
- Courtier, P., and Talagrand, O.: Variational assimilation of meteorological observations with the adjoint vorticity equation II: Numerical results, *Q. J. R. Meteorol. Soc.*, 113, 1329–1347, 1987.
- Droop, M. R.: The nutrient status of algal cells in continuous culture, *J. Mar. Biol. Assoc. UK*, 54, 825–855, 1974.
- Ebenhöh, W., Baretta-Bekker, J. G., and Baretta, J. W.: The primary production model in the ecosystem model ERSEM II, *J. Sea. Res.* 38, 173–192, 1997.
- Eknes, M. and Evensen G.: An Ensemble Kalman Filter with a 1-D Marine Ecosystem Model, *J. Mar. Sys.* (in press), 2002.
- Evans, G. T. and Parslow, J. S.: A model of annual plankton cycles, *Biol. Oceanogr.*, 3, 327–347, 1985.
- Evensen, G.: Using the extended Kalman Filter with a multilayer quasi-geostrophic ocean model, *J. Geophys. Res.*, 97, 17905–17924, 1992.
- Evensen, G.: Sequential data assimilation with a non-linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, 99, 10143–10162, 1994.
- Evensen, G.: Advanced data assimilation for strongly non-linear dynamics, *Mon. Weather Rev.*, 125, 1342–1354, 1997a.
- Evensen, G.: Application of ensemble integrations for predictability studies and data assimilation, in *Monte Carlo Simulations in Oceanography Proceedings 'Aha Huliko'a Hawaiian Winter Workshop*, University of Hawaii at Manoa, 14–17 January 1997, edited by Muller, P. and Henderson, D., 11–22, SOEST, 1997b.
- Evensen, G. and van Leeuwen, P. J.: Assimilation of Geosat altimeter data for the Agulas current using the ensemble kalman filter with a quasis-geostrophic model, *Mon. Weather Rev.*, 124, 85–96, 1996.
- Fasham M. J. R., Ducklow, H. W., and McKelvie S. M.: A nitrogen based model of plankton dynamics in the oceanic mixed layer, *J. Mar. Res.*, 48, 591–639, 1990.
- Geider, R. J., MacIntyre H. L., and Kana, T. M.: A dynamic model of phytoplankton growth and acclimation: responses of the balanced growth rate and the chlorophyll-a: carbon ratio to light, nutrient limitation and temperature, *Mar. Ecol. Prog. Ser.*, 148, 187–200, 1997.
- Georgopoulos, D. et al.: Hydrology and circulation in the Southern Cretan Sea during the CINCS experiment (May 1994 – September 1995), *Progress in Oceanography*, 46, 89–112, 2000.
- Kalman, R. E.: A new approach to linear filtering and prediction theory, *J. Basic Eng.*, 82, 35–45, 1960.
- Kalman, R. E. and Bucy, R. S.: New results of linear filtering and prediction theory, *J. Basic Eng.*, 83, 95–108, 1961.
- Krom, M. D., Kress, N., and Brenner, S.: Phosphorus limitation of primary productivity in the eastern Mediterranean Sea, *Limnology Oceanography*, 36, 3, 424–432, 1991.
- Mellor, G. L. and Yamada, T.: Development of a turbulence closure model for geophysical fluid problems, *Review of Geophysics and Space Physics*, 20, 851–875, 1982.

- Miller, R. N., Ghil, M. and Gauthiez, F.: Advanced data assimilation in strongly nonlinear dynamical systems, *J. Atmos. Sci.*, 51, 1037–1056, 1994.
- Natvik, L. J., Eknes, M., and Evensen, G.: A weak constraint inverse for a zero-dimensional four component marine ecosystem model, *J. Mar. Sys.*, 28, 19–44, 2001.
- Natvik, L. J. and Evensen, G.: Assimilation of ocean colour data into a biochemical model of the North Atlantic. Part 1. Data assimilation experiments, Submitted to *Ocean Modelling*, 2001.
- Natvik, L. J. and Evensen, G.: Assimilation of ocean colour data into a biochemical model of the North Atlantic. Part 2. Statistical analysis, Submitted to *Ocean Modelling*, 2001.
- Nittis, K., Tzazaaaiavos, C., Thanos, I., Drakopoulos, P., Cardin, V., Gacic, M., Petihakis, G., and Basana, R.: The Mediterranean Moored Multi-sensor array (M3A): System development and initial results, *Ann. Geophysique* (submitted), 2002.
- Nyholm, N., Kinetics of Nitrogen-Limited Algal Growth, *Progress in Water Technology*, 8, 347–35, 1977.
- Souvermezoglou, E., Krasakopoulou, E., and Pavlidou, A.: Temporal variability in oxygen and nutrient concentrations in the southern Aegean Sea and the straits of the Cretan Arc. *Progress in Oceanography*, 44, 573–600, 1999.
- Talagrand, O. and Courtier, P.: Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory, *Q. J. R. Meteorol. Soc.*, 113, 1311–1328, 1987.
- Taylor, A. H., Geider, R. J., and Gilbert, F. J. H.: Seasonal and latitudinal dependencies of phytoplankton carbon to chlorophyll-a ratios: results of a modelling study, *Mar. Ecol. Prog. Ser.*, 152, 51–66, 1997.
- Theocharis, A., Balopoulos, E., Kioroglou, S., Kontoyiannis, H., and Iona, A.: A synthesis of the circulation and hydrography of the South Aegean Sea and the Straits of the Cretan Arc (March 1994 – January 1995), *Progress in Oceanography*, 44, 469–509, 1999.
- Thingstad, T. F. and Rassoulzadegan, F.: Nutrient limitations, microbial food webs, and “biological C-pumps”: suggested interactions in a P-limited Mediterranean, *Marine Ecology Progress Series*, 117, 299–306, 1995.
- Triantafyllou, G., Petihakis, G., and Allen, J. I.: Assessing the performance (predictability) of the Cretan Sea ecosystem model with the use of high frequency M3a buoy data set, *Ann. Geophysicae* (submitted), 2002.
- Tselepides, A., Zervakis, V., Polychronaki, T., Donavaro, R., and Chronis, G.: Distribution of nutrients and particulate organic matter in relation to the prevailing hydrographic features of the Cretan Sea (NE Mediterranean), *Prog. Oceanogr.*, 46, 113–142, 2000.