Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Marine Systems 40–41 (2003) 79–97
JOURNAL OF MARINE SYSTEMS
www.elsevier.com/locate/jmarsys

An a priori approach to assimilation of ecological data in marine ecosystem models

C. Solidoro[a,*], A. Crise[a], G. Crispi[a], R. Pastres[b]

[a] *Istituto Nazionale di Oceanografia e di Geofisica Sperimentale, Borgo Grotta Gigante 42/c, Sgonico, Trieste 34016, Italy*
[b] *Department of Physical Chemistry, Università di Venezia, Dorsoduro 2137, Venice, Italy*

Received 26 September 2001; accepted 2 December 2002

Abstract

This study illustrates a methodology that deals with three basic problems that concern the calibration of marine ecosystem models: (a) How many parameters can be calibrated? (b) Which subsets of parameters can be calibrated ? (c) How can the uncertainty in a given model output be estimated?

The methodology is based on an a priori approach: this means that the results depend only on the structure of the model and on the set of variables which forms the data available for assimilation. The methodology is based on the computation of the sensitivities of the state variables to the model parameters, and enables one to analyze the role additional observations could play in constraining the model parameters. The methodology was applied to a one-dimensional (1D) primary production multinutrient model that describes the dynamics of pelagic ecosystem. The model describes nitrogen, phosphorus, and carbon cycles by means of a trophic chain constituted by two phytoplanktonic functional groups, one zooplanktonic pool and the detritus compartment. Nitrates, phosphates, and ammonia are considered as inorganic dissolved nutrients.

Results show that the sensitivities of the majority of the parameters are strongly correlated and, therefore, only 5 of 43 parameters of our model could be accurately calibrated, even if daily measurements of nutrients and chlorophyll *a* were available for 1 year and at three different depths. Most of the state variables show the highest sensitivity to parameters related to the water temperature, phytoplankton growth, and phytoplankton mortality. The analysis of this case study, which, in our opinion, is representative of oligotrophic mid-latitude environments, suggests that water shading coefficient, optimal temperature coefficients for small phytoplankton, optimal temperature coefficient for large phytoplankton, a grazing parameter, and a parameter that describes the influence of water temperature on biological and chemical kinetics could be simultaneously and efficiently calibrated.

Finally, indications about observational strategies are also given.

Keywords: A priori; Assimilation; Marine ecosystem models

1. Introduction

The growing interest of the marine scientific community in data assimilation (DA) is testified by the increasing number of papers and workshops devoted

* Corresponding author. Tel.: +39-40-2140315; fax: +39-40-2140266.
E-mail address: csolidoro@ogs.trieste.it (C. Solidoro).

0924-7963/03/$ - see front matter © 2003 Elsevier Science B.V. All rights reserved.
doi:10.1016/S0924-7963(03)00014-9

to this subject, which is today recognized as a timely, important, and exciting research topic for modelers, as well as for all marine scientists (Robinson and Lermousiax, 2000).

Such an interest arises as a consequence of the facts that: (a) mathematical models are now commonly used both as research and management tools; (b) the reliability of model predictions depends upon the quality of the model used (i.e., upon its capability of reproducing interactions and flows between its major compartments); and (c) in order to improve such reliability, a model can be adjusted to fit the observations, and the constrained model can then be used for making assessments. This adjustment is generally called DA. A variety of DA techniques have been used to improve numerical models and their forecasts. These numerical techniques, based on inverse theory, allow one to use the available experimental information for estimating the parameters of the model and/or constraining the model output.

DA is a research topic quite well known from a theoretical point of view (Gelb, 1979; Luenberg, 1968), but relatively new in marine ecology. Its application to marine ecosystem models is limited because the data are often very scarce and/or uncertain, there are many unknown parameters to be estimated, and the standard calibration algorithms may not converge or may lead to optimum estimates, which are ecologically unrealistic.

In marine ecosystem studies, the technique most frequently used is referred to as data fitting, or parameter estimation, or parameter calibration (Lawson et al., 1996; Matear, 1995; Prunet et al., 1996a,b; Spitz et al., 1998; Vallino, 2000). This technique seeks the values of the model parameters that minimize a given ‘‘cost function.’’ In fact, parameter estimation is an optimization problem that consists of (a) defining a cost function, which is as a measure of the misfit between observed data and model prediction, and may include some penalty terms (Matear, 1995), and (b) seeking for the set of parameter values that minimizes the cost function. The procedure, however, does not always yield meaningful results: in fact, in marine ecosystem model, as well as in water quality models, the overriding difficulty is the lack of parameters identifiability (Beck, 1987).

Even though the failure in the determination of a unique set of optimal parameters can be justified by many reasons, in this paper we focus on the lack identifiability due to the shape of the cost function in the neighborhood of the initial estimates of the parameters, which will be called the ‘‘nominal’’ set.

In marine ecosystem literature, parameter estimation is usually dealt with by adopting an a posteriori approach: given a set of observations, it is assumed that the discrepancies between model predictions and actual data can be accounted for by calibrating the values of a predefined subset of the parameters. The a posteriori approach is quite useful for defining the number of parameters that one can calibrate against a given data set, and also for estimating the confidence in the results. However, the choice of the subset of the parameters is arbitrary and the results depend on the given set of experimental data. Therefore, doubts always persist on the generality of the results thus achieved: Would the conclusion be the same if a different set of parameters had been selected to begin with? Would the estimates be different if more data, or more accurate data, had been available? To answer these questions, in this paper, we present an a priori method, which allows one to investigate the behavior of the model in the neighborhood of the nominal set of parameters. In this way, the whole set of parameters can be taken into consideration and the results do not depend on any assumption on data quality and number, but only on the intrinsic properties of the model structure (i.e., the way in which the parameters are mutually related in the conceptual representation of the phenomena).

The a priori analysis was performed with the purposes of: (1) determining the maximum number of parameters that can be simultaneously estimated from a given data set; (2) providing suggestions on which subset of parameters can be more accurately estimated; (3) comparing the magnitude of the standard errors of the parameters; and (4) obtaining an initial estimation of the uncertainty in a given model output. Furthermore, the method provides useful indications about the efficiency of alternative sampling strategies in relation to the above four questions. The analysis is based on the computation of local sensitivities, which are then organized in sensitivity matrices. The methodology is computationally efficient,

and can be implemented also in the analysis of complex models, such as three-dimensional (3D) coupled models at a basin scale. The analysis was applied on a one-dimensional (1D) primary production multinutrient model for the upper layer of a pelagic ecosystem.

2. Methods

Let the dynamics of an ecosystem be described by the model:

$$\dot{x} = f(x, \tilde{p}, t) \tag{1}$$

where $x \in R^{nv}$ is the state vector, which describes the state of the system, and $\tilde{p} \in R^{np}$ is the nominal parameter vector. Let us denote the nominal solution of Eq. (1) for the initial condition x_0 by:

$$\tilde{x}(\tilde{p}, t) = F(x_0, \tilde{p}, t) \tag{2}$$

If a parameter $\tilde{p}_j$ is slightly changed in $p_j = \tilde{p}_{jj} + \Delta p_j$, the new trajectory $x(p,t)$ can be approximated by a Taylor series expansion around the nominal trajectory $\tilde{x}(\tilde{p},t)$:

$$x(p, t) \cong \tilde{x}(\tilde{p}, t) + \frac{\partial F(x_0, \tilde{p}, t)}{\partial p_j} \Delta p_j \tag{3}$$

The first-order sensitivity of the vector x to a parameter p_j, or simply sensitivity (Turànyi and Rabitz, 2002), is the vector $S_j \in R^{nv}$ defined as:

$$S_j(t) \equiv \frac{\partial \tilde{x}(\tilde{p}, t)}{\partial p_j} \tag{4}$$

These coefficients are local, in the sense that their numerical value depends on the point in the space of the parameter around which the linearization (Eq. (2)) is performed. Local Sensitivity Analysis (LSA) can be used for producing the ensemble of the trajectories corresponding to all the possible (small) variations around the nominal vector $\tilde{p}$:

$$x \cong \tilde{x}(\tilde{p}, t) + \sum_i S_i \Delta p_i \tag{5}$$

LSA provides important information about parameter identifiability. In fact, if two or more trajectories of this ensemble are 'equal' (i.e., if different combinations of the parameters give the same output x), then:

$$\begin{aligned} x(\tilde{p} + \mathrm{d}p, t) &= \tilde{x}(\tilde{p}, t) + \sum_j S_j(\tilde{x}, \tilde{p}, t)\mathrm{d}p_j \\ &= x(\tilde{p} + \mathrm{d}p', t) = \tilde{x}(\tilde{p}, t) + \sum_j S_j(\tilde{x}, \tilde{p}, t)\mathrm{d}p'_j \\ \sum_j S_j(\tilde{x}, \tilde{p}, t)\mathrm{d}p_j &= \sum_j S_j(\tilde{x}, \tilde{p}, t)\mathrm{d}p'_j \\ \sum_j S_j(\tilde{x}, \tilde{p}, t)(\mathrm{d}p_j - \mathrm{d}p'_j) &= \sum_j S_j(\tilde{x}, \tilde{p}, t)\alpha = 0 \end{aligned}$$

that is, the sensitivities S_j are not linearly independent.

Conversely, if two or more sensitivities are not linearly independent, different combinations of the parameters yield the same trajectory. In fact, if $S_1 = \alpha S_2$, the following equalities hold in a small-enough neighborhood of $(\tilde{p}_1, \tilde{p}_2)$:

$$\begin{aligned} &x(\tilde{p}_1 + \Delta p_1, \tilde{p}_2 + \Delta p_2, t) \\ &\quad = \tilde{x}(\tilde{p}_1, \tilde{p}_2, t) + S_1 \Delta p_1 + S_2 \Delta p_2 \\ &\quad = \tilde{x}(\tilde{p}_1, \tilde{p}_2, t) + \alpha S_2 \,\mathrm{d}p_1 + S_1/\alpha \,\mathrm{d}p_2 \\ &\quad = \tilde{x}(\tilde{p}_1, \tilde{p}_2, t) + S_1(\mathrm{d}p_2/\alpha) + S_2(\alpha \mathrm{d}p_1) \\ &\quad = \tilde{x}(\tilde{p}_1, \tilde{p}_2, t) + S_1 \Delta p'_1 + S_2 \Delta p'_2 \end{aligned}$$

that is, given any trajectory $x(\tilde{p}_1 + \Delta\tilde{p}_1, \tilde{p}_2 + \Delta p_2)$ in the neighborhood of $\tilde{p}$, there exist at least two different couples of parameters, which produce this trajectory:

$$\begin{aligned} x &= x(\tilde{p}_1 + \mathrm{d}p_1, \tilde{p}_2 + \mathrm{d}p_2, t) \\ &= x(\tilde{p}_1 + \mathrm{d}p_2/\alpha, \tilde{p}_2 + \alpha \mathrm{d}p_1, t) \end{aligned}$$

This is true also for the nominal trajectory, since if one chooses $\Delta p_2 = -\alpha \Delta p_1$,

$$x(\tilde{p}_1 + \Delta p_1, \tilde{p}_2 + \Delta p_2, t) = \tilde{x}(\tilde{p}_1, \tilde{p}_2, t)$$

and therefore there exist at least two different couples of parameters, which produce the trajectory $\tilde{x}$.

Therefore, the parameters whose sensitivities are perfectly correlated cannot be calibrated against the same data set because the effects of their variations on the goal function are linearly related. On this basis, it is possible to regroup such parameters, as the results of the calibration are meaningful only if one parameter per subset is optimized. These subsets can be found by means of a Principal Component Analysis (PCA) of the matrix $r = s^T s$, where s

is the normalized local concentration sensitivity (Turànyi, 1990):

$$\mathbf{s}_{v,j}(t) = \frac{\partial x_v(t)}{\partial p_j}\frac{p_j}{x_v} \tag{6a}$$

In Eq. (6a), the index v runs over the nv state variable, and s is a matrix which is described right below.

Alternatively, sensitivities can be normalized by dividing by the average value of the state variables, x_{av}:

$$s_{\mathrm{av}_{v,j}}(t) = \frac{\partial x_v(t)}{\partial p_j}\frac{p_j}{x_{av}} \tag{6b}$$

Definition (6a) will be used throughout the paper, but there is one point where definition (6b) will be preferred.

The sensitivity matrix s is an np column matrix, where np is the number of parameters, made up of nt blocks. Each block is composed by $(nv \times ns)$ rows and contains the sensitivities to all state variables at times $t_1,t_2,\ldots,nt$, where nv is the number of state variables and ns is the number of sampling points. The diagonal elements of the matrix $s^{\mathrm{T}}s$ are then proportional to the variances of the model output that would be obtained if a single parameters would be slightly perturbed, while the off-diagonal ones are proportional to the co-variances induced by the simultaneous perturbation of each couple of parameters (Turànyi and Rabitz, 2002).

If the correlation between sensitivities is low, the off-diagonal elements are much smaller than the diagonal ones: in this case, it would be possible, in principle, to calibrate all the parameters. On the contrary, if the off-diagonal elements are important, this means that the variability of $s^{\mathrm{T}}s$ can be summarized by a fewer factors, which are linear combinations of the parameters. In this case, even the calibration of a "good" model (i.e., a structurally correct model) against a set of "good" data would lead to high variances of the "best" parameters. The PCA represents one of the most common methods for identifying these factors. This method extracts the eigenvectors of the matrix $s^{\mathrm{T}}s$ and gives the weight of the parameters in each eigenvector, which are called loadings. Therefore, it provides a basis for regrouping the parameters because the loadings of the parameters, whose sensitivities are strongly correlated, are very similar, or nearly opposite, in the most important eigenvector.

Once the PCA has been performed, one has still to select the parameters to be calibrated. The analysis of the sensitivity matrix can help in choosing these parameters, by selecting the ones that could be esti-

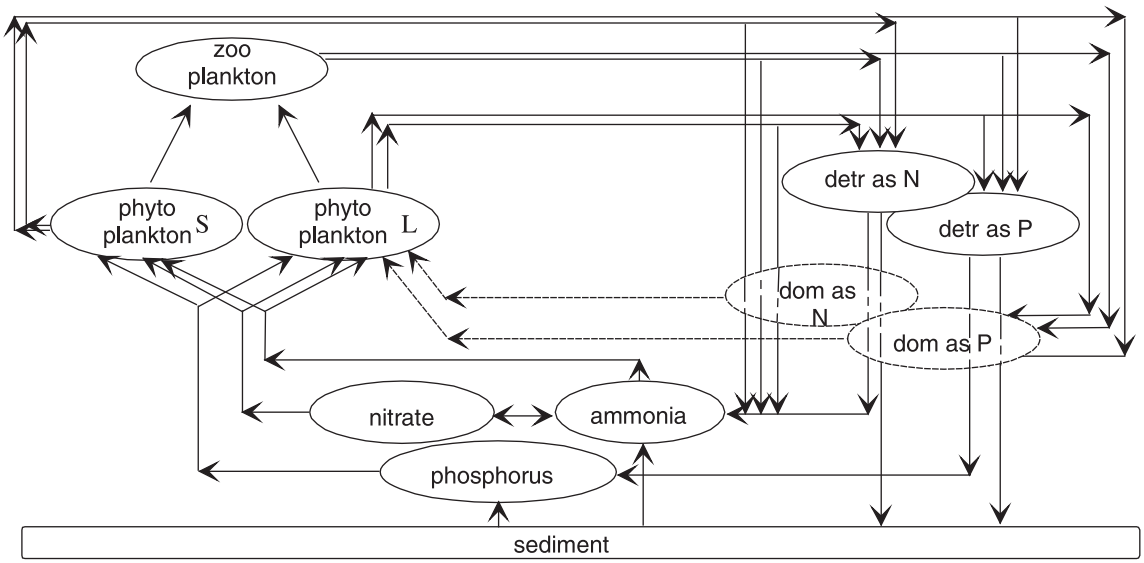

Fig. 1. Conceptual scheme of the interaction between the functional compartments in the ecological submodel.

Table 1
Parameterisation of the ecological submodel

$$dZ/dt = \{k_{grz}[f(P_s) + f(P_l)] - k_{mz} - k^*_{escz}\}Z \quad (T1)$$

$$dP_s/dt = \mu_s \mathrm{flim}_s P_s - (k_{mPs} + k_{sedPs} + k^*_{rPs})P_s - k_{grz} f(P_s)Z \quad (T2)$$

$$dP_l/dt = \mu_l \mathrm{flim}_l P_l - (k_{mPl} + k_{sedPl} + k_{rPl})P_l - k_{grz} f(P_l)Z \quad (T3)$$

$$\begin{aligned} dNH_4^+/dt = {} & -R_{NC}\{\lfloor \mu_s f_s(I) f_s(T) f_s(NH_4^+) f_s(PO_4^{3-}) - k^*_{rPs} \rfloor P_s\} \\ & - R_{NC}\{\lfloor \mu_l f_l(I) f_l(T) f_l(NH_4^+) f_l(PO_4^{3-}) - k^*_{rPl} \rfloor P_l\} \\ & + k_{nit} f_{nit}(O) NH_4^+ - R_{NC}\{-k^*_{dec} f(O) D_C \\ & + \lfloor k^*_{escz} - (1-\varepsilon_l) f(P_l) - (1-\varepsilon_s) f(P_s) \rfloor Z\} \end{aligned} \quad (T4)$$

$$\begin{aligned} dNO_x^-/dt = {} & -R_{NC}\{\lfloor \mu_s f_s(I) f_s(T) f_s(NO_x^-) e^{-\psi_s NH_4^+} f_s(PO_4^{3-}) \\ & - k^*_{rPs} \rfloor P_s\} - R_{NC}\{\lfloor \mu_l f_l(I) f_l(T) f_l(NO_x^-) \\ & \times e^{-\psi_l NH_4^+} f_l(PO_4^{3-}) - k^*_{rPl} \rfloor P_l\} \\ & - k^*_{nit} f_{nit}(O) NH_4^+ - R_{NC}\{-k^*_{dec} f(O) D_C \\ & + \lfloor k^*_{escz} - (1-\varepsilon_l) f(P_l) - (1-\varepsilon_s) f(P_s) \rfloor Z\} \end{aligned} \quad (T5)$$

$$\begin{aligned} dPO_4^{3-}/dt = {} & -R_{PC}\lfloor (\mu_l \mathrm{flim}_l - k^*_{rPl})P_l + (\mu_s \mathrm{flim}_s - k^*_{rPs})P_s \rfloor \\ & + R_{PC}\lfloor k^*_{escz} - (1-\varepsilon_l) f(P_l) - (1-\varepsilon_s) f(P_s) \rfloor Z \end{aligned} \quad (T6)$$

$$\begin{aligned} dD_N/dt = {} & R_{NC}(k_{mPs} P_s + k_{mPl} P_l + k_{mz} Z) \\ & - (k^*_{decN} f_{DN}(O) + k_{sed}) D_N \end{aligned} \quad (T7)$$

$$dD_C/dt = k_{mPs} P_s + k_{mPl} P_l + k_{mz} Z - (k^*_{decC} f_{DC}(O) + k_{sed}) D_C \quad (T8)$$

$$dD_P/dt = R_{PC}(k_{mPs} P_s + k_{mPl} P_l + k_{mz} Z) - (k^*_{decP} f_{DP}(O) + k_{sed}) D_P \quad (T9)$$

$$\begin{aligned} dO/dt = {} & (\mu_s \mathrm{flim}_s - k^*_{rPs})P_s + (\mu_l \mathrm{flim}_l - k^*_{rPl})P_l \\ & - k^*_{decC} f_{DC}(O) D_C - k_{nit} f'(O) NH_4^+ R_{ON} \end{aligned} \quad (T10)$$

Grazing limitation

$$f(P_s) = \alpha P_s/(\alpha P_s + P_l + k_{fz}) \quad (T11)$$

$$f(P_l) = P_l/(\alpha P_s + P_l + k_{fz}) \quad (T12)$$

Growth limitation (x = s,l)

$$\mathrm{flim}_x = f_x(I) f_x(T) \lfloor f_x(NH_4^+) + f_x(NO_x^-) e^{-\gamma_x NO_x^-} \rfloor f_x(PO_4^{3-}) \quad (T13)$$

Light limitation (x = s,l)

$$f_x(I) = \frac{I}{I_{Ox}} \exp\left[1 - \frac{I}{I_{Ox}}\right] \quad (T14)$$

Light absorption and phytoplankton self-shading

$$I = I_i \exp \int_1^z [-k_w z - k_p(P_s + P_l) - k_d D_c] dz \quad (T15)$$

Table 1 (*continued*)

Temperature limitation (x = s,l)

$$\begin{aligned} f_x(T) = {} & [(T_x\max - T)/(T_x\max - T_x o)]^{a_x(T_x\max - T_x o)} \\ & \times \exp[a_x(T - T_x\max)] \end{aligned} \quad (T16)$$

Nutrient limitation (x = s,l; Y = NH_4^+, NO_x^-, PO_4^{3-})

$$f_x(Y) = Y/(Y + k_{Yx}) \quad (T17)$$

Nitrification and decomposition (x = nit, DP, DN, DC)

$$f_x(O) = \mathrm{DO}/(\mathrm{DO} + k_x) \quad (T18)$$

Arrhenius formulation (x = escz, dec, rP_l, rP_s)

$$k^*_x = k_x \vartheta_x^{(T - T_x)} \quad (T19)$$

mated with the highest degree of accuracy or those whose accurate determination minimizes the uncertainty in a given model output.

The choice can be addressed by the determination of the "tuning importance" of each parameter (Turànyi, 1990), which allows one to rank the parameters in respect to their effect on model output. Sensitivities to different parameters are compared using relative sensitivities (Eq. (6a)), which are dimensionless.

Tuning Importance ($TI_{v,j}$) of the variable v in respect to the parameter p_j is defined as the sum of squares of relative sensitivities (Turànyi, 1990):

$$TI_{v,j} = \sum_{it=1}^{nt} s^2 j, v, it \quad (7)$$

If the effect of the parameters on more than one variable needs to be considered, the Global Tuning Importance GTI (Turànyi, 1990) is used:

$$GTI_j = \sum_v \sum_{it=1}^{nt} s^2 j, v, it \quad (8)$$

The indexes GTI_j are the diagonal terms of the matrix $(s^T s)$. Therefore, if the variables considered in GTI are the output variables, high values of GTI_j are associated with low standard errors on the parameter p_j, since, in the linear hypothesis, these are proportional to the diagonal elements of the inverse matrix $(s^T s)^{-1}$. The TI gives an indication about which variable contributes more to GTI and therefore in reducing the standard errors, since the TI is a contribution of each state variable to GTI. Once the standard errors are determined, the uncertainty in a given linear function

Table 2
List of the parameters of the ecological submodel

Z	Density of zooplankton	*zoo*
P_s	Density of phytoplankton small	*phys*
P_l	Density of phytoplankton large	*phyl*
NH_4^+	Water concentration of reduced inorganic nitrogen	*nh4*
NO_x^-	Water concentration of oxidized inorganic nitrogen	*nox*
PO_4^{3-}	Water concentration of reactive phosphorous	*po4*
DC	Carbon in the detritus	
DN	Nitrogen in the detritus	
DP	Phosphorous in the detritus	
O	Dissolved oxygen	
μ_s	Max growth rate phyto small	*ta*
T_{smax}	Inhibition temperature for phyto small	*tmaxs*
T_{so}	Optimal temperature for phyto small	*topts*
a_s	Coefficient for temperature function for phyto small	*cls*
I_{os}	Optimal light for phyto small	*ios*
k_{nhs}	Ammonia halfsaturation for phyto small	*knhs*
k_{nos}	Nitrogen halfsaturation for phyto small	*knos*
ψ_s	Inhibition nitrogen uptake coefficient phyto small	*psis*
k_{po}	Phosphorus halfsaturation for phyto small	*ksps*
k_{rPs}	Respiration rate phyto small	*krs*
k_{exuPs}	Exudation rate phyto small	*kexs*
k_{mPs}	Mortality rate phyto small	*kds*
μ_l	Max growth rate phyto large	*mul*
T_{lmax}	Inhibition temperature for phyto large	*tmaxl*
T_{lo}	Optimal temperature for phyto large	*topl*
A_l	Coefficient for temperature function for phyto large	*cll*
I_{ol}	Optimal light for phyto large	*iol*
k_{nhl}	Ammonia halfsaturation for phyto large	*knhl*
k_{nol}	Nitrogen halfsaturation for phyto large	*knol*
ψ_l	Inhibition nitrogen uptake coefficient phyto large	*psil*
k_{pol}	Phosphorus halfsaturation for phyto large	*kpol*
k_{rPl}	Respiration rate phyto large	*krpl*
k_{exuPl}	Exudation rate phyto large	*kexupl*
k_{mPl}	Mortality rate phyto large	*kmpl*
k_{grz}	Max grazing rate zooplankton	*kgrz*
k_{pz}	Grazing halfsaturation	*kpf*
α	Grazing preference	*alpha*
ε_l	Efficiency yield phyto large to zooplankton	*effs*
ε_s	Efficiency yield phyto small to zooplankton	*effl*
k_{mz}	Mortality rate zooplankton	*kmz*
k_{escz}	Excretion rate zooplankton	*kescz*
k_{nit}	Nitrification rate	*knit*
k_{onit}	Oxygen halfsaturation for nitrification	*konit*
k_{decN}	Decay rate nitrogen	*kdecn*
k_{decP}	Decay rate phosphorus	*kdecp*
k_{sedPl}	Sinking rate phyto large	*ksedpl*
k_{sed}	Sinking rate detritus	*ksedd*
k_{rear}	Superficial recreation rate	*krear*
k_{ps}	Shading coefficient for phyto small	*kps*
k_{pl}	Shading coefficient for phyto large	*kpl*
k_w	Shading coefficient for water	*kw*
k_d	Shading coefficient for detritus	*kd*
θ	Arrhenius coefficient	*teta*
T_a	Arrhenius reference temperature	*ta*

$L(x)$ of a model output can be estimated by means of Eq. (5). For example, the variance of the (time) average value of the i^{th} state variable, as a function of the standard error δp_j, can be estimated as:

$$\mathrm{Var}(x_{av}) \approx \left[\sum_t S_{j,i,t}\right]^2 (\delta p_j)^2 \tag{9}$$

However, care must be taken when using Eq. (9) for ranking the parameters in respect to their contribution to uncertainty in model output, because the linear approximation may be violated if δp_j is too large. For this reason, if the analysis is focused on estimation of uncertainty, the validity of the ranking provided by this local analysis should be checked by using also global methods (Pastres and Ciavatta, 2002; Pastres et al., 1999).

The above equations can be extended to time- and space-dependent models, like the one which is analyzed in this paper. In this case, the state equation becomes:

$$\partial x_i/\partial t = \nabla \cdot (K \nabla x_i) + (\vec{v} \cdot \nabla) x_i + f(x,p) \tag{10}$$

The sensitivities can be computed by solving, together with the state equation, a set of additional vector partial differential equations:

$$\partial S_i/\partial t = \nabla \cdot (K \nabla S_i) + (\vec{v} \cdot \nabla) S_i + J\, S_i + \partial f/\partial p \tag{11}$$

where J is the Jacobian matrix, not to be confused with the cost function indicated in this paper by Ψ. There are several ways of solving Eq. (11). In our work, the *direct method* (Koda et al., 1979), which is described in detail in Appendix A, was applied.

3. Results

The methods outlined above were applied to a 1D water column model, as part of a general research project concerning the modelling of the Mediterranean Sea (http\\:www.ogs.trieste.it). The model equation reads as:

$$\dot{x}(z) = \underbrace{\nabla_z k \nabla_z x(z)}_{\text{diffusion}} + \underbrace{f(x(z),p)}_{\text{reaction}} + \underbrace{v_s \nabla_z x(z)}_{\text{sinking}} \qquad (12)$$

The application of Eq. (11) to the state equation leads to (see Appendix A):

$$\partial S_i/\partial t = \underbrace{\nabla_z(k\nabla_z S_i)}_{\text{diffusion}} + \underbrace{J S_i + \partial f/\partial p}_{\text{reaction}} + \underbrace{v_s \nabla_z S}_{\text{sinking}} + \underbrace{h_{\text{shad}}}_{\text{shading}} \qquad (13)$$

where a new term, h_{shad}, appears explicitly (see also Appendix A). This term takes into account the shading of the incident light, which depends on the depth but cannot be assimilated to the transport terms nor to the reaction one. The reaction term of Eq. (12) is described in detail in Crispi et al. (1998). Both nitrogen and phosphorus were taken into account, since each nutrient is potentially limiting. Two different species of primary producers were considered, which are representative of large and small phytoplanktonic species, respectively, such as diatoms and autotrophic microflagellates. The main interactions among the ecological compartments are visualized in Fig. 1. The dependence of phytoplankton growth on water temperature was modelled by using the empirical formulation proposed by Lassiter and Kearns (1974), which is characterized by an optimum temperature and a cut-off temperature. The limitations due to other factors were described by means of a multiplicative model: the effect of nutrient was described by a Monod kinetic, while that of light was modelled using

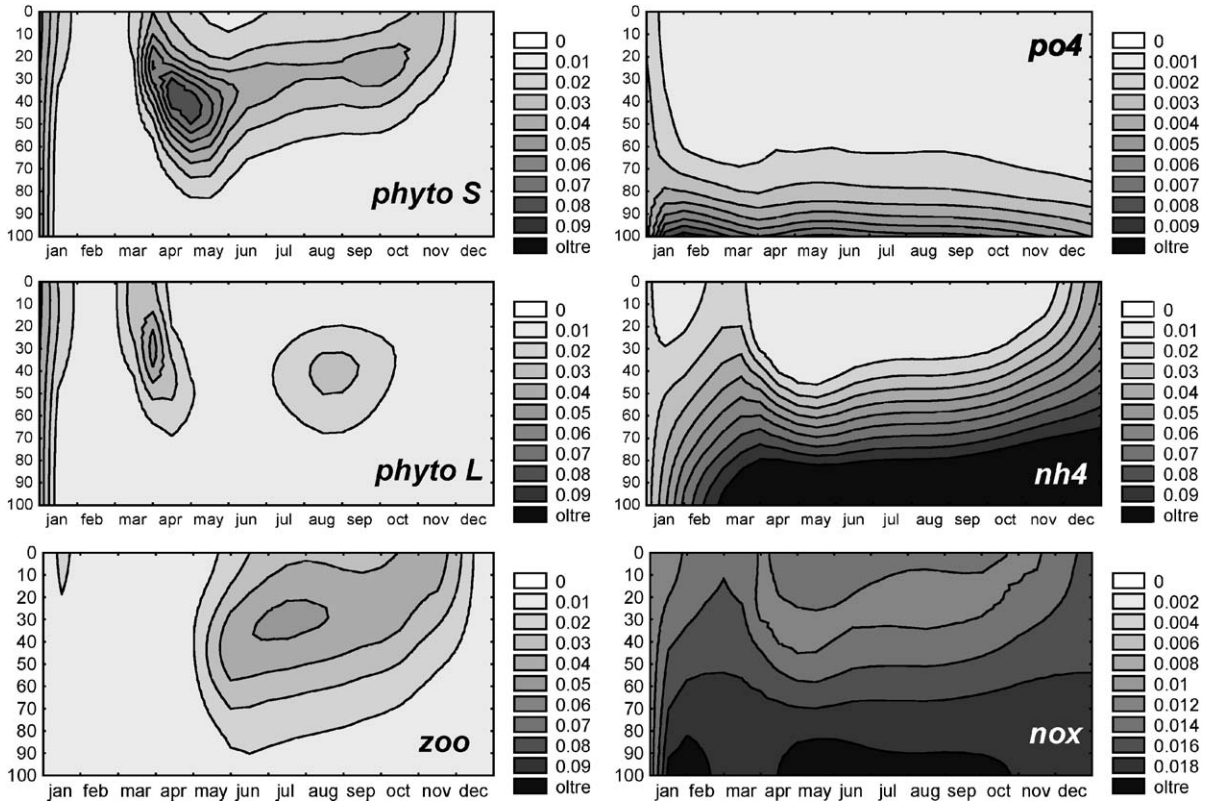

Fig. 2. Time evolution of the simulated concentrations of primary producers (mg/l C), zooplankton (mg/l C), and nutrients (mg/l N or P) along the water column.

Steele formulation. Zooplankton grazing was modelled by a Hollings Type II functional response (Holling, 1965).

The biogeochemical cycles of the carbon and of the macronutrients were closed by introducing the detritus compartment, which receives the nonliving particulate and dissolved organic matter. The latter is produced by mortality processes, excretion, and exudation. The dynamics of dissolved oxygen was also simulated, since this variable, besides being frequently sampled, is an aggregated index of the quality of a water body. Table 1 summarizes the structure of the biological submodel.

The model is driven by light intensity at the surface and a vertical profile of temperature, which were computed for the Tyrrhenian Sea by using the 3D model of the Mediterranean Sea quoted above. Transport processes were described by means of the sinking and diffusion processes, while advective terms were not considered. The nominal values of the 43 parameters of the model are listed in Table 2. Whenever possible, they were chosen according to the suggestions of the specific literature regarding the Mediterranean ecosystem models (Varela et al., 1992, 1994, 1995; Crise et al., 1999; Crispi et al., 1999; Pinazo et al., 1996; Tusseau et al., 1997). When any specific information was lacking, the nominal values were taken from general literature on marine ecosystem models (Fasham et al., 1990; Moisan and Hofmann, 1996). Eqs. (12) and (13) were solved numerically by using a DEC alpha workstation: a yearly simulation took around 20 min.

The nominal trajectory of the model reproduces correctly the formation of the deep chlorophyll maximum (Fig. 2) and the seasonal succession of the two phytoplanktonic functional groups (Crispi et al., 2002). The increases in light and temperature trigger the bloom of the large phytoplankton (*phyl*) in the early spring, followed by a bloom of the small phytoplankton (*phys*), which reaches its maximum

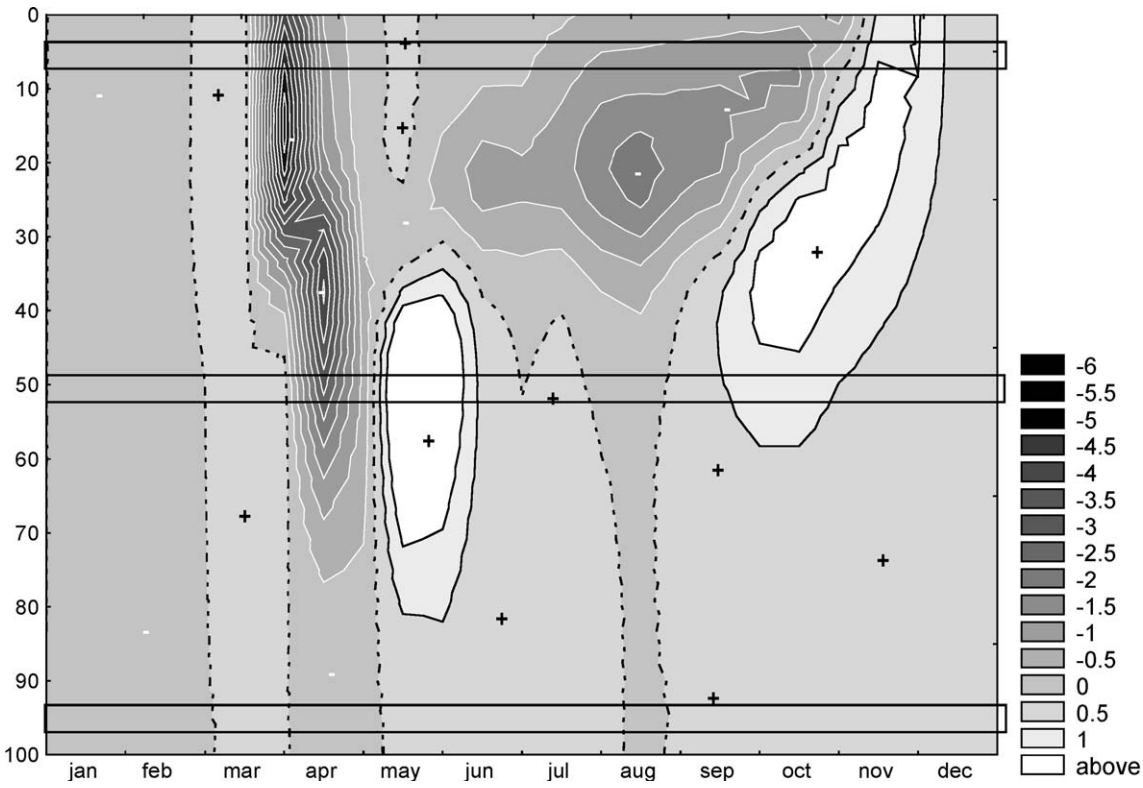

Fig. 3. Time evolution of the relative sensitivity, in the water column, of phytoplankton small density to maximum growth rate of phytoplankton large density.

productivity at higher levels of light intensity and temperature than *phyl*. The blooms cause a rapid depletion of nutrients: this effect is more pronounced near the surface. The grazing activity starts affecting the phytoplanktonic stocks toward the end of the spring, as one can see from the increase in zooplankton density.

An example of the use of LSA for gaining a deeper insight in the mechanism of the model is illustrated in Fig. 3, which shows the sensitivity of *phys* to *mul*, the maximum growth of *phyl*. The sensitivity does not show an appreciable spatial variation with depth until the end of April, when an increase in *mul* causes an increase in *phys* in the upper 50 m and a decrease below that depth. In fact, an increase in this parameter increases the competitiveness of *phyl*, and therefore causes an earlier bloom. As a consequence of this temporal shift, the bloom of *phys* occurs a little bit later and takes place at a lower depth than in the nominal trajectory.

The results of the identifiability analysis depend on the choice of the model output. Our analysis was performed by taking as components of the output vector the concentrations of ammonia, nitrate, and phosphate, and the total phytoplankton biomass, which is the sum of *phys* and *phyl*. Furthermore, we assumed a "sampling scenario" in which each output variable is sampled once a day at the three different depths (5, 50, and 100 m), which are indicated by the stripes in Fig. 3. Therefore, the sensitivity matrix was formed by 43 columns, one per each parameter, and by 365 blocks, one per each day, each one made up of 4 (output variables) × 3 (depths) rows.

The results of the Principal Component Analysis of the matrix $s^{T}s$ are summarized in the scree plot of Fig. 4 and in Table 3, which shows the loadings in the eigenvectors: loadings higher than 0.7 are marked in bold. Table 3 shows the loadings of the factors after a VariMax rotation (Legendre and Legendre, 1998), which allows one to compare the loadings more easily, but unrotated loadings were considered as well. The scree plot of Fig. 4 shows that the first four eigenvalues account for most of the total variance (92%). Parameters whose sensitivities are correlated appear with similar, or opposite, loadings in the most important components, as it happens for the three parameters *ksps*, *krs*, and *kexs*. The inspection of Table 3 indicates that most of the parameters clusterize in agreement with their role in the dynamics of the system. Parameters referring to the two different pools of phytoplankton fall in two distinct factors; therefore, it is possible to calibrate one parameter per pool.

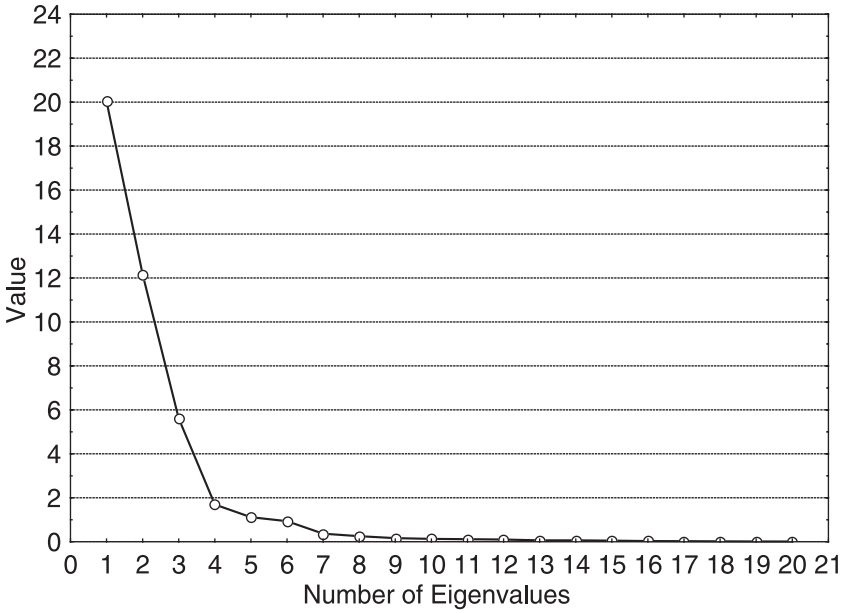

Fig. 4. Scree plot for the PCA performed on the matrix *s*.

Table 3
Loadings (VariMax-normalized) of each parameter in the factors of the PCA performed on the matrix *s*

		Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
Max growth phyto s	*ta*	−0.19939	0.235856	**0.93880**	−0.104926	0.012715	−0.043388	−0.072432
Inhibition temperature phyto s	*tmaxs*	−0.27786	0.040837	**0.94651**	−0.090800	−0.027584	−0.051700	−0.088536
Optimal temperature phyto s	*topts*	0.24052	−0.097462	**−0.94865**	0.104998	0.069136	0.049115	0.089023
Coefficient temperature phyto s	*cls*	0.29373	0.026884	**−0.94186**	0.068737	0.023128	0.052450	0.095551
Optimal light phyto s	*ios*	**−0.83568**	−0.109920	0.01564	0.055422	0.449778	0.033104	0.001652
Ammonia halfsaturation phyto s	*knhs*	−0.32568	−0.293963	**−0.84360**	0.164129	−0.093295	0.103745	−0.176376
Nitrogen halfsaturation phyto s	*knos*	0.09304	−0.308472	−0.18114	−0.019402	**−0.922407**	−0.008128	0.028241
Inhibition nitrogen phyto s	*psis*	**−0.86782**	−0.028520	−0.22389	−0.356308	0.013223	0.001476	0.138945
Phosphorus halfsaturation phyto s	*ksps*	0.11494	−0.229321	**−0.95618**	0.077716	−0.026475	0.045687	0.081338
Respiration rate phyto s	*krs*	0.04041	−0.123799	**−0.96472**	0.066373	−0.149147	0.076918	−0.113587
Exudation rate phyto s	*kexs*	0.04042	−0.123798	**−0.96472**	0.066355	−0.149163	0.076915	−0.113567
Mortality rate phyto s	*kds*	0.44649	0.184346	**−0.86129**	0.044484	−0.012223	0.026525	0.101715
Max growth phyto l	*mul*	**−0.97494**	0.046415	0.19915	0.046462	−0.003716	−0.017745	−0.041874
Inhibition temperature phyto l	*tmaxl*	**−0.91129**	0.222264	0.30361	−0.060130	−0.043858	−0.032954	−0.008435
Optimal temperature phyto l	*topl*	**0.94480**	−0.154373	−0.26507	0.014897	0.028368	0.025981	0.021556
Coefficient temperature phyto l	*cll*	**0.89087**	−0.254735	−0.32156	0.085540	0.052540	0.037553	0.002368
Optimal light phyto l	*iol*	**0.77633**	0.386537	0.13912	−0.414049	−0.101095	−0.003786	0.064401
Ammonia halfsaturation phyto l	*knhl*	**0.93231**	0.183607	0.06907	−0.258864	−0.062180	0.005992	0.030606
Nitrogen halfsaturation phyto l	*knol*	**0.94271**	−0.144388	−0.28108	0.029446	0.024502	0.030914	0.049759
Inhibition nitrogen phyto l	*psil*	**0.87471**	0.277138	0.04426	−0.358103	−0.072298	0.009610	0.056772
Phosphorus halfsaturation phyto l	*kpol*	**0.97086**	−0.060944	−0.21218	−0.025517	0.016451	0.023070	0.040685
Respiratory rate phyto l	*krpl*	**0.98407**	0.056022	0.02425	−0.116799	−0.014003	0.025524	−0.014344
Exudation rate phyto l	*kexupl*	**0.98407**	0.056024	0.02425	−0.116802	−0.014004	0.025524	−0.014341
Mortality rate phyto l	*kmpl*	**0.95536**	−0.075453	−0.26035	0.031180	0.029201	0.038517	0.008021
Max grazing zoo	*kgrz*	−0.00196	**−0.978620**	−0.13083	0.132618	−0.041671	0.031649	−0.010329
Grazing halfsaturation	*kpf*	−0.10326	**0.961184**	0.21441	−0.101971	0.070315	−0.031887	−0.004563
Preferred grazing	*alpha*	0.27143	**0.936256**	0.12430	−0.150204	0.021522	−0.017924	0.027251
Efficient grazing phyto l	*effs*	0.03661	**−0.981946**	−0.02893	0.140366	−0.034339	0.004597	0.036978
Efficient grazing phyto s	*effl*	0.42291	−0.681360	−0.49673	0.247440	0.001786	0.043982	−0.002161
Mortality rate zoo	*kmz*	−0.10950	**0.974329**	0.11927	−0.092954	0.083709	−0.018818	0.052926
Excretion rate zoo	*kescz*	−0.25404	**0.928904**	0.16793	−0.050107	0.115308	0.010589	−0.094111
Nitrification rate	*knit*	**0.71718**	−0.052349	−0.46472	−0.109673	−0.176777	0.238051	0.398982
Oxygen halfsaturation nitrogen	*konit*	**−0.70588**	0.053300	0.44527	0.119436	0.201731	−0.256432	−0.413517
Decay rate nitrogen	*kdecn*	**−0.73537**	−0.546083	0.34362	0.014206	0.032757	−0.043053	−0.095851
Decay rate phosphorus	*kdecp*	**−0.72603**	0.278915	0.50453	0.031167	0.069734	−0.024567	−0.033771
Recreation rate	*krear*	0.17314	−0.027284	−0.28676	0.030855	0.013543	**0.938407**	0.028100
Shading coefficient phyto s	*kps*	0.39705	0.378426	0.32936	**−0.739550**	0.014930	−0.060706	0.049821
Shading coefficient phyto l	*kpl*	0.19196	0.492590	0.31566	**−0.734690**	−0.029170	0.028831	−0.003969
Shading coefficient water	*kw*	**0.79824**	0.315537	0.03302	−0.489172	0.031062	−0.002171	0.039417
Shading coefficient detritus	*kd*	0.11021	0.472714	0.51832	−0.675958	−0.042115	−0.022437	−0.033062
Arrhenius coefficient	*teta*	−0.69007	0.181804	0.53068	0.186081	−0.082887	−0.252828	−0.062841
Arrhenius reference temperature	*ta*	**−0.82985**	0.153409	0.33923	0.199470	0.130553	−0.201210	−0.038933
Sinking rate phyto l		0.52878	0.681859	−0.30048	0.035755	−0.010866	0.065382	0.057867
Explained variance		17.16801	8.427458	10.60362	2.648596	1.275066	1.168131	0.508484
Percent variance total explanation		0.39926	0.195987	0.24660	0.061595	0.029653	0.027166	0.011825

Parameter whose loadings are higher than 0.7 are marked in bold. Last two rows indicate, respectively, the total variance and the percentage of the total variance explained by each of the factors.

Table 4
Groups of parameter that have correlated sensitivities

Group	Parameters	Opt
1	*teta, mul, kmpl, tmaxl, cll, Iol, kpol, knol, kexupl, knhpl, krpl, psil, Ios, psis, ta, kdecn, kedcp, knit, konit, kw, ta*	*teta*
2	*topts, tmaxs, mus, kds, cls, kexs, ksps, krs knhs*	*topts*
3	*alpha, kgrz, kpf, effs, kmz, escz*	*alpha*
4	*kps, kpl, kd*	*kps*
5	*knos*	*knos*
6	*krear*	*krear*
Mix	*teta, effl, ksedph,*	*teta*

The last column gives, for each group, the parameter with the highest TI, and indicates an initial optimal subset of parameters.

Parameters related to zooplankton activity are grouped together in the third factor of the PCA, which still explains 19% of the total variance. The fourth factor, which explains 6% of total variance, is dominated by phytoplankton and detritus shading. Water shading coefficient is instead correlated with the parameters of the first factor. On this basis, the sets of parameters presented in Table 4 may be tentatively identified as independent. These sets include 40 of 43 original parameters. The three left parameters, whose sensitivities are not clearly correlated to those of other sets, are listed in a seventh additional 'mixed' set. Among them, only *teta* has an high GTI, as is shown in Fig. 5.

The relative importance of the parameters was investigated by comparing the GTIs, computed by means of Eq. (8), which are shown in Fig. 5. The inspection of Fig. 5 shows that parameters related to the influence of the temperature on the growth of phyto-s and phyto-l (*topts* and *toptl*) and with the other processes following Arrhenius law (*teta*) have the greatest tuning importance. Parameters related to the other external forcings, nutrient availability, and light intensity are far less important. Moreover, parameters related to maximum phytoplanktonic growth, grazing activity, and shading are of relevance, while the tuning importance of all the other ones is small. The ranking reveals that parameters that play similar roles, such as mortality, respiration, and exudation, have different tuning importance: such a finding would have been hard to foresee. In the figure, different colors indicate

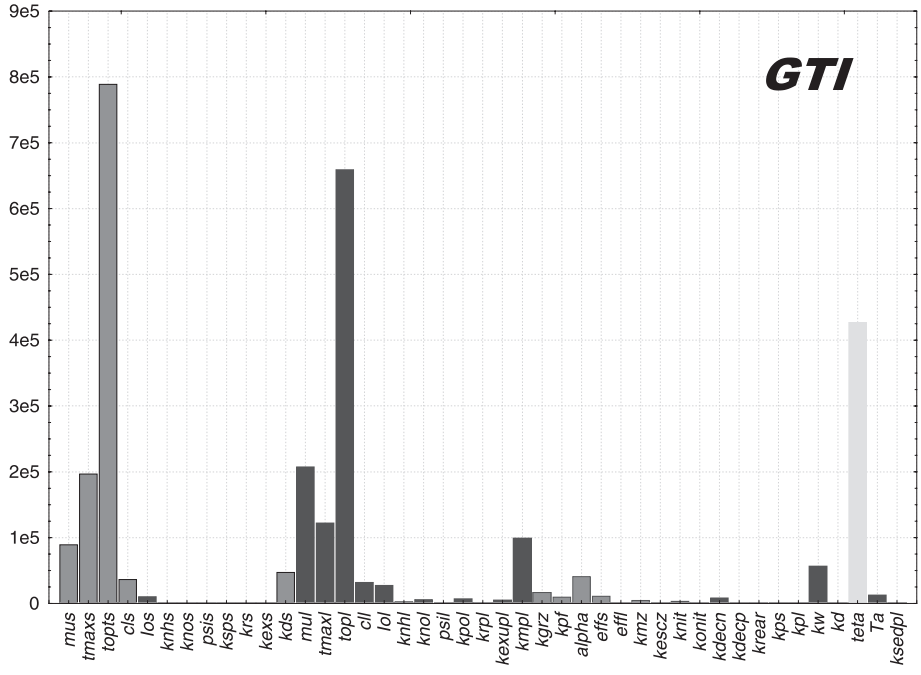

Fig. 5. Global Tuning Importance of the parameters of the model. Different colors indicate parameters belonging to different groups defined in Table 4 (i.e., parameters whose sensitivities are correlated have the same color, and dominate one of the principal components) (TI values are dimensionless).

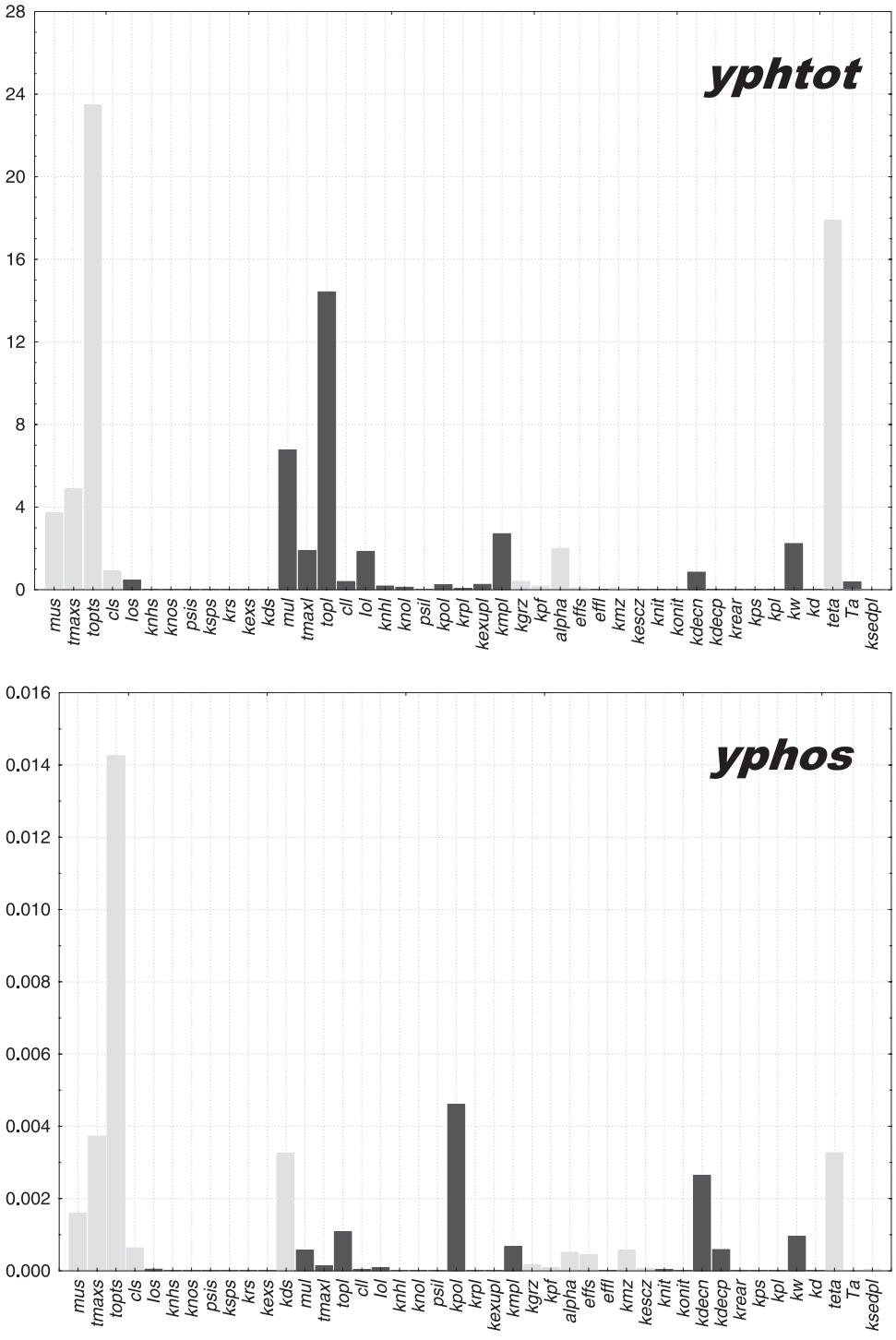

Fig. 6. Comparison of the contribution of different parameters to the uncertainty in the yearly integrated value of the total phytoplanktonic biomass (above) and of the yearly integrated value of phosphorus (below). Different colors indicate parameters belonging to different groups, as in Table 4.

parameters belonging to different groups, defined in Table 4. The last column of Table 4 gives, for each group, the parameter with the highest tuning importance, and therefore indicates an initial optimal set of parameters to be calibrated, namely the subset of parameters that could be calibrated with the lowest standard error on their estimates.

The sensitivity matrix can also be used for obtaining a first estimation of the uncertainty in a linear function $L(x)$ of the state vector, which could be of interest. This application of LSA is here illustrated by considering the uncertainties in the yearly integrated values of the two phytoplanktonic groups *yphtot*, and of the total phosphorous *yphos*. For example, an accurate knowledge of the first quantity may be more important than the accurate knowledge of the actual trajectory if one is dealing with the analysis of the carbon cycle in the context of global changes. If one assumes that the standard deviation of all the parameters is given by the same percentage of their nominal values, the variances of the yearly integrated values can be estimated by using the expression:

$$\mathrm{Var}(\mathrm{yiv}) \propto \sum_{j=1}^{\mathrm{np}} \left[\frac{\sum_{d=1}^{365} x_1 s_{i,j}}{\sum_{d=1}^{365} x_i} \right]^2 \qquad (14)$$

The results are summarised in Fig. 6, which shows that the contribution of a given parameter to the total uncertainty is markedly different. For example, an accurate estimation (i.e., a low standard deviation) of the semisaturation value kpol would markedly reduce the uncertainty in the total phosphorous, whereas it would not improve the accuracy of the estimation of total phytoplankton. In the figure, different colors indicate parameters belonging to different groups, as in Table 4.

4. Discussion

4.1. Parameter identifiability

Most of the applications of DA to marine ecosystem have been focused so far on the parameter estimation problem, which can be reduced to the following question: 'Given a certain model structure, is it possible to estimate accurately, uniquely, and with as little uncertainty as possible the unknown parameters?' Usually, the question is addressed by following an a posteriori approach (i.e., by finding in some way the set of optimal parameter and their standard error). This approach relies on the particular data set used, and implicitly assumes that a solution does exist.

The novelty of this paper is that we suggest to use LSA in order to check a priori if there are different vectors of parameters that would give the same trajectory. In such a case, it would not be possible to discriminate among such vectors, and therefore the solutions of the calibration problem would not be unique, at least in the neighborhood of the nominal set of parameters. As shown in the Methods section, this situation occurs if two sensitivities are linearly dependent. The existence of at least one trajectory, which is produced by two distinct combinations of parameters, is an equivalent condition. The number of parameters or, more precisely, of linear combinations of parameters, which can be actually constrained by a given data set, can be found by applying the PCA to the matrix $s^{\mathrm{T}}s$.

From another point of view, the sensitivity matrix can be used for the creation of a set of error-free synthetic data (Eq. (5)), and the PCA represents an attempt to determine whether or not it is possible to recover the nominal parameters through model calibration and to assess, if this is not possible, the maximum number of parameters that can be recovered. The plus in the a priori approach is that the sensitivity matrix s contains information about *all* the trajectories, which results from any arbitrary, slight variations δp in the parameter set. In this way, the analysis can be seen as an ensemble of an infinite number (one for each possible variation δp) of twin experiments.

The present analysis was performed by assuming that four output variables, nutrients, and total phytoplanktonic biomass were sampled with a very high frequency. The results of the PCA analysis show that the first four eigenvalues account for more than 90% of the total variance, and that 40 parameters can be divided into six sets of correlated parameters, while the three remaining ones identify a seventh group. In addition, it can be seen that the most important parameters can be divided in only five of these groups.

Therefore, a small number of parameters (at most seven) could be simultaneously and efficiently cali-

brated, even if daily data were available. This is in agreement with Prunet et al. (1996b) and Matear (1995), who, by using an a posteriori approach, concluded that, at most, seven to eight parameters could be estimated by calibration against experimental data. In both cases, the complexity of their models is similar to the one we use, but the data set was slightly different. A much greater number of parameters, around 30, were instead identified in Vallino (2000), but his own analysis showed that the estimates were not unique.

Qualitatively similar results were obtained by applying the PCA analysis to a different sampling scenario, in which weekly data were assumed to be collected. This finding agrees with the results presented in Lawson et al. (1996), who analysed the influence of different sampling schemes on parameter identifiability by using an a posteriori approach. Their analysis showed that the frequency of the sampling is a critical factor below a certain threshold, but that above the threshold, an increase in the frequency of the sampling leads to more accurate estimates but does not enhance substantially the possibility to identify the parameters.

4.2. Ranking of the parameters and choice of the subset of parameters to be estimated

Once the maximum number of parameters that can be calibrated has been determined, there still is the necessity to decide which combinations of parameters can be estimated. As remarked in the Methods section, the analysis of sensitivity matrix is useful in choosing parameters that are 'important,' either because they strongly affect model output (i.e., have high TI) or minimize the uncertainty in a given model output. The results of the Principal Component Analysis of the matrix $s^{T}s$ suggest that, at most, seven parameters can be independently calibrated. However, the maximization of the tuning importance, which led to the subset presented in the last column of Table 4, should not be considered as the only criterion for selecting the parameters. In fact, one could be interested also in the minimization of the uncertainty in a given model output, and take into account the results of uncertainty analysis. For example, Table 5 compares the subsets that are obtained by minimizing the uncertainty in yearly average of total phytoplankton (*yphtot*), second row, and yearly average of total phosphorous (*yphos*), third row, with the one identified on the basis of the tuning importance. The comparison among the three subsets shows that the maximization of GTI and the minimization of the uncertainty in phytot leads to the same choice. Furthermore, the five parameters *topts*, *kps*, *knos*, *krear*, and *teta* satisfy also the third criterion, which instead suggests the substitution of *toptl* with *kpol* and *alpha* with *kmz*.

Table 5
Comparison of possible optimal subsets of parameters, obtained by maximization of GTI (A), minimization of uncertainty on *yphtot* (B), and minimization of uncertainty on *yphos* (C)

	1	2	3	4	5	6	7
A (GTI)	*toptl*	*alpha*	*topts*	*kps*	*knos*	*krear*	*teta*
B *(yphtot)*	*toptl*	*alpha*	*topts*	*kps*	*knos*	*krear*	*teta*
C *(yphos)*	*kpol*	*kmz*	*topts*	*kps*	*knos*	*krear*	*teta*

Even though the three subsets include parameters that have an independent effect on the model output, three of them (*knos*, *krear*, *kps*) contribute very little to the GTI and to uncertainties in *yphytot* and *yphos*, as one can see from Fig. 5. Therefore, it seems reasonable to eliminate these parameters and to include tentatively some parameters that are partially correlated with the left four ones, in bold in Table 5.

Two criteria can be taken as guidelines for the final choice: the ranking of the parameters and their degree of correlation with the basic subset, which often depend on the ecological role. For example, the first group of Table 4 includes 21 of 43 parameters: 15 of them are directly related to the dynamics of the two phytoplanktonic groups, while the other six are involved in other processes. Among them, the most important parameter is *kw*, whose loads on the first four eigenvectors are rather different from the ones of *toptl*, which already belong to the basic set "A." The degree of correlation of the subset thus identified can be estimated by computing the inverse of the matrix R, which is obtained by extracting from the matrix $s_{av}^{T}s_{av}$ only the row and columns that correspond to the five candidate parameters. Low values of the condition number (i.e., the ratio of the largest singular value to the smallest) indicate that the matrix R is not ill conditioned and, therefore, can be easily inverted: in turn, this means that the calibration of the parameters leads to meaningful results, as the standard errors are proportional to the diagonal elements of the matrix $(s_{av}^{T}s_{av})^{-1}$. In this case, if we add to the four pa-

rameters (in bold) of group A the parameter *kw*, a condition number of 167.12 is obtained, whereas a condition number of 1726 is computed if we add the parameter *mul*, whose loads are similar to those of *toptl* (see Table 3). Analogous results are obtained by considering the subset B and adding to the parameters marked in bold the same parameter *kw*.

4.3. Choice of the sampling strategy

The importance of a variable in constraining a model parameter (i.e., the contribution of the variable to the gradient of the cost function Ψ) can be assessed by computing the indexes TI. Analogously, the analysis of the evolution of s_j^2 in time provides information on the contribution to the gradient of Ψ of measurement taken at different times. In fact, a high value of s_j^2 implies a high contribution to GTI, which—as remarked in the Methods section—is inversely related to standard errors on the parameters. Therefore, the analysis of the evolution of s_j^2 for the parameters, which one has decided to calibrate, can provide useful indications on optimal sampling strategy. A detailed analysis of this important issue is beyond the scope of this paper, in which we illustrate the potentiality of a preliminary analysis only by means of the following qualitative example. However, the methods could be used to compare the effectiveness of alternative sampling plans, in order to optimize the resources. As an example, the TI values of the output variables (i.e., nutrients and total phytoplankton) of the five candidate parameters *topts*, *toptl*, *alpha*, *teta*, and *kw* are compared in Fig. 7. The inspection of the figure evidence as the measurement of phytoplankton and phosphorus is more valuable than those of nitrogen, and that sampling is more informative in March and August.

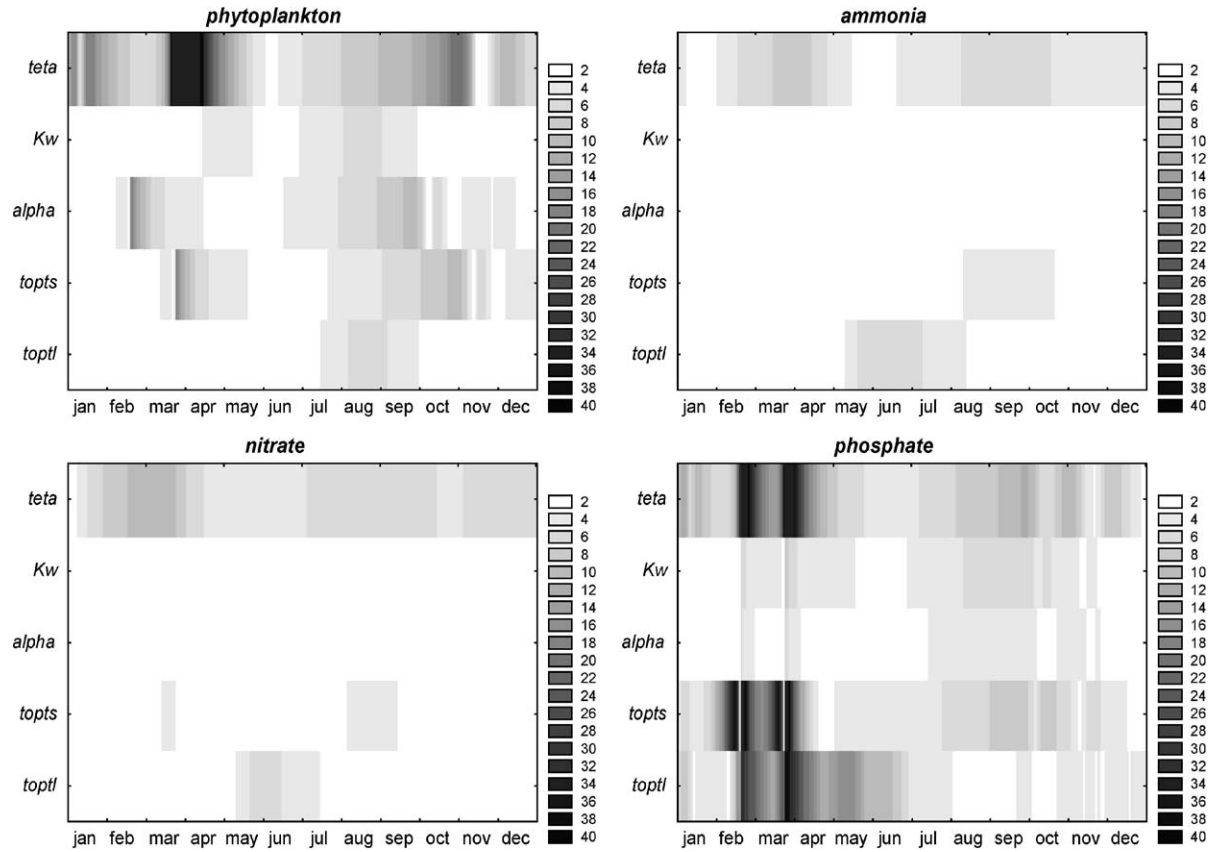

Fig. 7. Time evolution of the TI of nutrients and phytoplankton to the parameters that compose the optimal subset (TI values are dimensionless).

5. Conclusion

The paper illustrates a methodology for selecting the parameters to be estimated when assimilating ecological data in a model of marine ecosystem. The a priori approach here adopted gives one the possibility of exploring the identifiability of the parameters without being constrained by the collection of a given data set. In our opinion, this grants more general validity to the conclusions, in respect to a posteriori analysis based on a specific data set.

The choice of the parameters should be guided by (a) a careful comparison of the *sensitivities*, and sensitivity-derived indexes such as *tuning importance* and *global tuning importance*, (b) the constraints posed by the identifiability problem, and (c) the parameters one is most interested in.

In regard to the identifiability problem (point *b*), a formal proof is given that no data set allows one to discriminate among different parameters if there exists a single data set which could have been produced by two (or more) distinct combinations of such parameters. This is equivalent to say that the sensitivities of the data set to these parameters are linearly correlated. It is then shown how a statistical analysis of the sensitivities helps one in organizing this information, by indicating groups of parameters with correlated effects on model output.

Even though, in principle, each parameter of a given group could be calibrated, a calibration algorithm converges more quickly and the uncertainties of the estimates are lower if one chooses among each group the parameter with the greatest tuning importance (point *a*). In this case, the focus is on sensitivity and sensitivity-derived indexes (TI and GTI) of the variable homologous to the available data eventually used for the assimilation. Furthermore, sensitivities and sensitivities-derived indexes provide information on how valuable the knowledge of a given parameter is in respect to some specific process (point *c*) (i.e., how much the uncertainty on a given parameter contributes to the uncertainty in a given model output). Therefore, there might be different optimal choices for different purposes. Finally, one might like to consider also the amount of information already available for each parameter.

The paper presents in Appendix A also an ad hoc implementation of the so-called ‘direct method’ useful for a systematic computation of the local sensitivities. This method is computationally efficient and can be applied to 3D models. The implementation here described permits one to consider processes, like shading effect of plankton and detritus on the light intensity along the column, that cannot be included in a straightforward implementation of the method.

The specific results of our analysis evidence that only a small number of parameters (i.e., five) can be calibrated on the basis of the sampling scenario here hypothesized. Further, a clear distinction appears among parameters that are related to the two primary producers and with zooplankton. Our analysis indicates that the parameters which relate water temperature to biological activity exert a strong influence on model output: this influence is not considered in other papers, where an important, but less systematic, sensitivity analysis of ecosystem models is performed, such as Varela et al. (1992), where temperature is not considered and light intensity and nitrogen availability are indeed indicated as the key factors in regulating deep chlorophyll maximum. Fasham et al. (1990) investigate the effect of a large number of parameters on annual net primary production and f ratio, but temperature is not considered as a force of the model. In Levy et al. (1998), temperature is considered in the model, but temperature-related parameters are not included in the set of parameters subjected to sensitivity experiments.

The results also suggest that it is possible to extract information on the dynamics of both primary producers and zooplanktonic pools from total phytoplankton biomass data. From a modelling point of view, this motivates a more complex biological submodel than those which consider a single planktonic pool. On the other hand, since it is not possible to rely on calibration for more than very few parameters, it is necessary to estimate most of them from direct (or laboratory or mesocosm) measurements. Since direct measurements of rates of biological processes and multiannual time series are both still rare, model parsimony in building mathematical models of pelagic ecosystems looks appropriate.

The results include the ranking of the parameters, according to TI, s, and GTI. Sensitivities of the trajectories are generally higher than sensitivities of yearly integrated values, indicating that in most cases, the modification of a parameter causes a temporal, or

spatial, shift in the plankton evolution, but does not substantially change the annual standing crop. In general, the parameters that relate biological activity with water temperature (*topts*, *toptl*, *teta*, and, to a lower extent, *tmaxl* and *tmaxs*) are the most important. They are followed by phytoplankton growth rates (*mul* and *mus*) and mortalities (*kdl* and *kds*), and then by grazing parameters (*kgrz* and *alpha*). Best results can be achieved by constraining the set of parameter composed by *topts*, *toptl*, *alpha*, *kw*, and *teta*.

Data can be ranked, too, based on their relative importance in constraining the parameters. Indications are given on optimal observing strategies for the most important parameters: the observation of phytoplankton and phosphorus allows one to estimate the optimal subset of parameters with the greatest accuracy, and March and August are the more valuable sampling moments.

Acknowledgements

The research has been partially funded by EC contract MAST3-CP96-0051 and by CoRiLa. The authors wish to thank E.E. Hofmann for her generous encouragement and for helpful comments on a preliminary draft of the paper.

Appendix A

There are several ways to compute sensitivities. The most frequently used is the 'brute force' method, that is, the comparison of the nominal trajectory, which is the trajectory obtained using the reference values of the parameters, and several trajectories obtained by running the model after perturbing each parameter at a time. The brute force method can be easily understood and implemented; indeed, it has been used also for performing sensitivity analysis referring to marine ecosystem. (Fasham et al., 1990; Varela et al., 1994, 1995). However it is expensive from a computational point of view, and would be hardly applicable for analysis as detailed as the one here presented, especially if these would be extended to basinwide 3D models. However, it is possible to use approaches more sophisticated and efficient than the brute force method.

In fact, if one takes into consideration—to begin with—a spatially homogenous system in which the state variables evolve in agreement with Eq. (1), by differentiating Eq. (1) in respect to a parameter p_j, commuting the order of differentiation, and substituting $\mathrm{d}x/\mathrm{d}p_j$ with S_j, one finds:

$$\dot{S}_j = \frac{\mathrm{d}f(x,p,t)}{\mathrm{d}p_j} = \frac{\partial f(x,p,t)}{\partial p_j} + \frac{\partial f(x,p,t)}{\partial x} S_j \qquad \text{(A1)}$$

The simultaneous integration of Eqs. (1) and (A1) gives the time evolution of $S(x,p,t)$.

The implementation of this method requires additional programming work, since the Jacobian matrix $\partial f/\partial x$ has to be computed, but once this operation is done, any single run enables one to obtain the sensitivities of x to each of the parameters, nearby a given point of the parameter space: if the sensitivities have to be computed around a different nominal trajectory, the computer code does not require any modification.

The analysis can be extended to spatially heterogeneous systems by using the so-called direct method of Koda et al. (1979), which has been applied to a 1D model of a eutrophic channel in Pastres et al. (1997).

In the direct method, one adopts a finite-difference scheme for the space discretization of Eq. (12): in this way, the system is defined by a state vector x, continuous in time, of dimension *nz*nv*, where *nz* is the number of vertical grid points. Analogously, the vector p becomes an *nz*np* vector. The dynamics of this system is then defined by a set of ordinary differential equations, to which Eq. (A1) can be applied. The evolution of a generic variable v in the generic layer z reads as:

$$\dot{x}_{v,z} = \frac{1}{\Delta z^2}\{k^-(x_{v,z-1} - x_{v,z}) + k^+(x_{v,z+1} - x_{v,z})\} + \frac{v_v}{\Delta z}(x_{v,z-1} - x_{v,z}) + f_z(x,p_z) \qquad \text{(A2)}$$

where v ranges from 1 to *nv*, z ranges from 1 to *nz*, and k_- and k_+ are, respectively, the eddy diffusion coefficients between the cell z and $z-1$ and the cells z and $z+1$, while Δz is the spatial (vertical) step. Note that the local term $f_z(x,p)$ may depend on all the components of the vector x, while it depends explicitly only on the parameters defined for the layer z.

After differentiating Eq. (A2) in respect to $p_{j,k}$, one obtains the following equation:

$$\dot{S}_{j,v,z,k} = \frac{1}{\Delta z^2}\{k^-(S_{j,v,z-1,k} - S_{j,v,z,k}) + k^+(S_{j,v,z+1,k} - S_{j,v,z,k})\} + \frac{v_v}{\Delta z} \times (S_{j,v,z-1,k} - S_{j,v,z,k}) + \frac{\partial f_z(x,p)}{\partial p_{j,k}} + \sum_{q=1}^{nz}\sum_{w=1}^{nv}\frac{\partial f_z(x,p)}{\partial x_{w,q}} S_{j,w,q,k} \quad (A3)$$

The sensitivity $S_{j,v,z,k}$ represents an estimate of the effects of the variations of the jth component of p in a layer k, and $p_{j,k}$ on the state variable v at the layer z.

Since, in our hypothesis, p_j is changed simultaneously in the water column, its sensitivity is obtained by summing over all the layers k the sensitivities given by Eq. (A3):

$$\dot{S}_{j,v,z} = \sum_k S_{j,w,z,k} = \frac{k^-}{\Delta z^2}(S_{j,v,z-1} - S_{j,v,z}) + \frac{k^+}{\Delta z^2}(S_{j,v,z+1} - S_{j,v,z}) \frac{v_{s,v}}{\Delta z}(S_{j,v,z-1} - S_{j,v,z}) + \sum_k \frac{\partial f_z(x,p)}{\partial p_{j,k}} + \sum_{w,q}\frac{\partial f_z(x,p)}{\partial x_{w,q}} S_{j,w,q} \quad (A4)$$

The reaction term of Eq. (A2) depends explicitly only on the parameters of the current layer z, that is, $\frac{\partial f_z}{\partial p_{j,k}} = 0$ for $k \neq z$. Therefore, the corresponding term of Eq. (A4) simplifies to:

$$\sum_k \frac{\partial f_z(x,p)}{\partial p_{j,k}} = \frac{\partial f_z}{\partial p_{j,z}} \quad (A5)$$

Further, if the reaction term of Eq. (A2) depends explicitly only on the variables of the current layer z, that is, $\frac{\partial f_z}{\partial x_{w,q}} = 0$ for $q \neq z$, the corresponding term of Eq. (A4) becomes:

$$\sum_{w,q}\frac{\partial f_z(x,p)}{\partial x_{w,q}} S_{j,w,q} = \sum_w \frac{\partial f_z(x,p)}{\partial x_{w,z}} S_{j,w,z} \quad (A6)$$

As a consequence, in this case, the sensitivities can be handled in the code simply as additional state variables: the reaction term of their state equation is given by Eqs. (A5) and (A6), and, like other variables (Eq. (A2)), they are subjected to sinking and diffusion processes:

$$\dot{S}_{j,v,z} = \frac{k^-}{\Delta z^2}(S_{j,v,z-1} - S_{j,v,z}) + \frac{k^+}{\Delta z^2}(S_{j,v,z+1} - S_{j,v,z}) + \frac{v_{s,v}}{\Delta z}(S_{j,v,z-1} - S_{j,v,z}) + \frac{\partial f_z(x,p)}{\partial p_{j,z}} + \sum_w \frac{\partial f_z(x,p)}{\partial x_{w,z}} S_{j,w,z} \quad (A7)$$

In our case, however, as it is shown in Table 1, the simplification expressed in Eq. (A5) cannot be applied, since only a weaker form of it (Eq. (A8)) holds true:

$$\frac{\partial f_z}{\partial x_{v,q}} = 0 \text{ only if } q > z \quad (A8)$$

and Eq. (A4) is rewritten as:

$$\dot{S}_{j,v,z} = \frac{k^-}{\Delta z^2}(S_{j,v,z-1} - S_{j,v,z}) + \frac{k^+}{\Delta z^2}(S_{j,v,z+1} - S_{j,v,z}) + \frac{v_{s,v}}{\Delta z}(S_{j,v,z-1} - S_{j,v,z}) + \frac{\partial f_z(x,p)}{\partial p_z} + \sum_w \frac{\partial f_z(x,p)}{\partial x_{w,z}} S_{j,w,z} + \sum_w \sum_{h=1}^{z-1} \frac{\partial f_z(x_{v,k})}{\partial x_{w,h}} S_{j,w,h} \quad (A9)$$

In this particular case, the last term of Eq. (A9) turns out to be fairly simple, as it depends only on the three variables that are involved in shading (i.e., phyto large, phyto small, and detritus). As a result, the sensitivities can be computed without introducing too much complication in the code, which remains quite efficient. It can be seen as Eq. (9) is the discrete form of Eq. (13).

References

Beck, M.B., 1987. Water quality modelling: a review of the analysis of uncertainty. Water Resour. Res. 23, 1392–1442.

Crispi, G., Crise, A., Solidoro, C., 1998. Three-dimensional oligotrophic ecosystem model driven by physical forcings: the Mediterranean Sea case. Environ. Model. Softw. 13, 483–490.

Crise, A., Allen, I., Baretta, J., Crispi, G., Mosetti, R., Solidoro, C., 1999a. The effect of physical forcing on ecosystem dynamic. Prog. Oceanogr. 44, 219–243.

Crispi, G., Crise, A., Mauri, E., 1999b. A seasonal three-dimensional study of the nitrogen cycle in the Mediterranean Sea: Part II. Verification of the energy constrained trophic model. J. Mar. Syst. 20, 357–379.

Crispi, G., Crise, A., Solidoro, C., 2002. Coupled three-dimensional ecomodel of phosphorus and nitrogen cycles in the Mediterranean Sea. J. Mar. Syst. 33–34, 497–521.

Fasham, M.J.R., Ducklow, H.W., McKelvie, S.M., 1990. A nitrogen based model of phytoplankton dynamics in the oceanic mixed layer. J. Mar. Res. 48, 591–639.

Gelb, A., 1979. Applied Optimal Estimation. The MIT Press (Massachusetts Institute of Technology), Cambridge, MA, p. 373.

Holling, C.S., 1965. The functional response of predators to prey density. Mem. Entomol. Soc. Can. 45, 5–60.

Koda, M., Drogu, A.H., Seinfeld, J.H., 1979. Sensitivity analysis of partial differential equations with applications to reaction and diffusion processes. J. Comp. Phys. 30, 259–292.

Lassiter, R.R., Kearns, D.K., 1974. Phytoplankton population changes and nutrient fluctuations in a simple aquatic ecosystem model. In: Middlebrookers, E.J., Falkenberger, D.H., Maloney, T.E. (Eds.), Modelling the Eutrophication Process. Ann Arbour Science, 131–188.

Lawson, L.M., Hofmann, E.E., Spitz, Y., 1996. Time series sampling and data assimilation in a simple marine ecosystem model. Deep Sea Res. Part 2. Top. Stud. Oceanogr. 43 (2–3), 625–651.

Legendre, P., Legendre, S., 1998. Numerical Ecology. Elsevier, Amsterdam.

Levy, M., Memery, L., Andre, J.M., 1998. Simulation of primary production and export fluxes in the northwestern Mediterranean sea. J. Mar. Res. 56, 197–238.

Luenberg, D.G., 1968. . Optimization by Vector Space Methods. Wiley, New York, p. 326.

Matear, R.J., 1995. Parameter optimization and analysis of ecosystem models using simulated annealing: a case study at Station P. J. Mar. Res. 53, 571–607.

Moisan, J.R., Hofmann, E.E., 1996. Modeling nutrient and plankton processes in the California coastal transition zone: 1. A time- and depth-dependent model. J. Geophys. Res. 101, 22647–22676.

Pastres, R., Ciavatta, S., 2002. Proceedings ISEEM.

Pastres, R., Franco, D., Pecenik, G., Solidoro, C., Dejak, C., 1997. Local sensitivity of a distributed parameters water quality model. Reliab. Eng. Syst. Saf. 57, 21–30.

Pastres, R., Chan, K., Solidoro, C., Dejak, C., 1999. Global sensitivity analysis of a shallow-water 3D eutrophication model. Comput. Phys. Commun. 117, 62–74.

Pinazo, C., Marsaleix, P., Millet, B., Estournel, C., Vehil, R., 1996. Spatial and temporal variability of phytoplankton biomass in upwelling areas of the northwestern Mediterranean: a coupled physical and biogeochemical modelling approach. J. Mar. Syst. 7, 161–191.

Prunet, P., Minster, J., Ruiz-Pino, D., Dadou, I., 1996a. Assimilation of surface data in a one-dimensional physical–biogeochemical model of the surface ocean: 1. Method and preliminary results. Glob. Biogeochem. Cycles 10 (1), 111–138.

Prunet, P., Minster, J., Echevin, V., Dadou, I., 1996b. Assimilation of surface data in a one-dimensional physical–biogeochemical model of the surface ocean: 2. Adjusting a simple trophic model to chlorophyll, temperature, nitrate, and pCO_2 data. Glob. Biogeochem. Cycles 10 (1), 139–158.

Robinson, A.A., Lermousiax, P., 2000. GLOBEC workshop on the assimilation of biological data in coupled physical/ecosystem models. GLOBEC Spec. Contrib. 3.

Spitz, Y.H., Moisan, J.R., Abbott, M.R., Richman, J.G., 1998. Data assimilation and a pelagic ecosystem model: parameterization using time series observation. J. Mar. Syst. 16, 51–68.

Turànyi, T., 1990. Sensitivity analysis of complex kinetic systems. Tool and application. J. Math. Chem. 5, 203–248.

Turànyi, T., Rabitz, H., 2002. Local methods. In: Saltelli, A., et al., (Eds.), Sensitivity Analysis. Wiley, New York, pp. 81–99.

Tusseau, M.H., Lancelot, C., Martin, J.M., Tassin, B., 1997. 1D coupled physical–biological model of the northwestern Mediterranean sea. Deep Sea Res. Part 2. Top. Stud. Oceanogr. 44 (3/4), 851–880.

Varela, R.A., Cruzado, A., Tintore, J., Ladona, E.G., 1992. Modelling the deep-chlorophyll maximum: a coupled physical–biological approach. J. Mar. Res. 50, 441–463.

Varela, R.A., Cruzado, A., Tintore, J., 1994. Simulation analysis of various biological and physical factors influencing the deep-chlorophyll maximum structure in oligotrophic areas. J. Mar. Syst. 5, 143–157.

Varela, R.A., Cruzado, A., Gabaldon, J.G., 1995. Modelling the dynamics of the North Sea's using the European Regional Seas Ecosystem Model. Neth. J. Sea Res. 33 (3/4), 337–361.

Vallino, J.J., 2000. Improving marine ecosystem models: use of data assimilation in mesocosm experiments. J. Mar. Res. 58 (1), 117–164.