Journal of Marine Research, 58, 675–707, 2000

Journal of
MARINE RESEARCH

Volume 58, Number 5

Selecting a global optimization method to estimate the oceanic particle cycling rate constants

by Véronique Athias[1], Pierre Mazzega[1] and Catherine Jeandel[1]

ABSTRACT

The objective is to select an inverse method to estimate the parameters of a dynamical model of the oceanic particle cycling from *in situ* data. Estimating the parameters of a dynamical model is a nonlinear inverse problem, even in the case of linear dynamics. Generally, biogeochemical models are characterized by complex nonlinear dynamics and by a high sensitivity to their parameters. This makes the parameter estimation problem strongly nonlinear. We show that an approach based on a linearization around an *a priori* solution and on a gradient descent method is not appropriate given the complexity of the related cost functions and our poor *a priori* knowledge of the parameters. Global Optimization Algorithms (GOAs) appear as better candidates. We present a comparison of a deterministic (TRUST), and two stochastic (simulated annealing and genetic algorithm) GOAs. From an exact model integration, a synthetic data set is generated which mimics the space-time sampling of a reference campaign. Simulated optimizations of two to the eight model parameters are performed. The parameter realistic ranges of values are the only available *a priori* information. The results and the behavior of the GOAs are analyzed in details. The three GOAs can recover at least two parameters. However, the gradient requirement of deterministic methods proves a serious drawback. Moreover, the complexity of the TRUST makes the estimation of more than two parameters hardly conceivable. The genetic algorithm quickly converges toward the eight parameter solution, whereas the simulated annealing is trapped by a local minimum. Generally, the genetic algorithm is less computationally expensive, swifter to converge, and has more robust procedural parameters than the simulated annealing.

1. Introduction

There are four principal sources of oceanic particles. Most of them are biologically produced within the ocean, others are brought as suspended matter by the river load, as

1. Laboratoire d'Etudes en Géophysique et Océanographie Spatiales, UMR5566 CNRS/CNES/UPS, 14 Avenue Edouard Belin, 31401 Toulouse Cedex 4, France. ***email: veronique.athias@cnes.fr***

airborne dust or as cosmic material accumulated by the Earth (Lal, 1977). The magnitude and the composition of the particle fluxes through the ocean are controlled by continual exchanges between the dissolved and particulate phases. These exchanges are the result of complex couplings between biological (e.g., phytoplankton aggregation induced by nutrient depletion, disaggregation due to zooplankton grazing, remineralization by bacteria, active absorption of chemical elements by the phytoplankton), chemical (e.g., dissolution, precipitation, adsorption, desorption) and physical (e.g., aggregation/disaggregation due to advection and turbulent transports) processes. The resulting fluxes are one of the vectors for the transfer of chemical elements toward the deep ocean. Assessing quantitatively the relative importance of these processes is a critical environmental and scientific issue. In particular, it would allow a better evaluation of the ability of the seas to sequester carbon dioxide from the atmosphere and a better understanding of what controls the chemical composition of the sea water (Whitfield and Turner, 1987).

The modeling of the dissolved-particulate exchanges was initiated in the seventies (Craig, 1974), from the study of particle-reactive trace metals and radioisotopes, especially thorium (Th) (Nozaki *et al.,* 1981). One purpose was to study the penetration of reactive pollutants in the deep ocean and the suitability of the deep ocean floor as a dump site for wastes. More generally, Th isotopes are considered as good tracers of the particle cycling because of their range of half-lives and sources and their reactivity with particles (Bacon and Anderson, 1982; Honeyman *et al.,* 1988; Nozaki *et al.,* 1987; Roy-Barman *et al.,* 1996). The studies on Th were at the root of the development of models describing the particulate organic, inorganic and inert detrital matter cyclings (Cochran *et al.,* 1993; Murnane *et al.,* 1996; 1994), with the intention of introducing into ocean general circulation models more explicit schemes for nutrient remineralization and particle dynamics (Henderson *et al.,* 1999). In fact, models have developed very little since the work of Bacon *et al.* (1985).

Most models simulate the temporal evolution of the composition of a 1D water column, with three reservoirs of chemical elements: the dissolved phase (material that passes through a 0.4 to 0.65 μm filter, collected with bottles), the small particle phase (suspended material retained on a 0.4 to 0.65 μm filter, collected by filtration), and the large particle phase (sinking material collected by sediment traps). Honeyman and Santschi (1989) proposed to add a colloidal phase (material that passes through a 0.4 μm filter, and retained by a 10,000 Dalton ultrafiltration membrane) to take into account the role of colloidal aggregation on the scavenging of trace elements and radionuclides. However, dissolved-colloidal exchanges are not elucidated yet (Roy-Barman *et al.,* 2000).

As a general rule, those models are characterized by a very aggregated structure, which is constrained by the kind of *in situ* measurements presently available. The exchanges between the reservoirs are often crudely parameterized, with simple (first-order or second-order) kinetics, which do not necessarily have a mechanistic basis. In fact, each of

these exchanges integrates numerous processes, whose nature, temporal and spatial scales are very different. For instance, the aggregation of suspended particles into sinking particles can be due to physical, chemical and biological processes (Alldredge and Jackson, 1995; Jackson and Lochmann, 1993). As a consequence, the parameters of such aggregated models are hard to deduce directly from *in situ* or laboratory experiments. To quantitatively describe the oceanic dissolved-particulate system, we need to solve the inverse problem which consists of estimating the particle cycling model parameters from *in situ* data.

In our opinion, this complex problem requires two steps to solve it. First, we need to identify an inverse method that is suited to the peculiar characteristics of this inverse problem. Then, we have to check if the available *in situ* data effectively allow constraining of the model parameters. This depends on the nature of the measured variables, on their connection with the model dynamical variables, on the duration, the spatial and temporal resolutions of the sampling, and on data uncertainties. The second step will be presented in a following paper which will be based on data collected in the tropical northeast Atlantic Ocean during the French JGOFS-EUMELI program. This paper focuses on the first step. Note that its conclusions may be valid for the inversion of other biogeochemical models, like the pelagic ecosystem models. The fundamental characteristics of the inversion of pelagic ecosystem and particle cycling models are indeed similar.

Several publications have already been dedicated to the application of assimilation/inverse methods to the estimation of dynamical variables or parameters of biogeochemical models, but the use of a particular method is usually not justified (Evans, 1999; Fasham and Evans, 1995; Matear, 1995; Murnane, 1994; Prunet *et al.,* 1996; Spitz *et al.,* 1998; Vézina and Platt, 1988). It is first advisable to define what must be estimated. Estimating the dynamical variables of a dynamical (that is, time-dependent) model is a linear or nonlinear inverse problem, depending on the linear or nonlinear model dynamics. On the other hand, estimating the parameters of a dynamical model is always a nonlinear inverse problem, even in the case of linear dynamics (Evensen *et al.,* 1998). For instance, the composition of the water column at a given time is a function of all the processes that occurred at the moment just before. As a consequence, the measured variables used to constrain the process parameters depend on those parameters. In addition, biogeochemical models, and among them particle cycling models, are characterized by complex nonlinear dynamics (Athias *et al.,* 1998; Edwards and Brindley, 1996). Therefore, estimating the parameters of biogeochemical models is a highly nonlinear inverse problem. Given this strong nonlinearity, classical linear or quasi-linear inverse methods may not be the most suitable to solve them. Yet, here we show that Global Optimization Algorithms (GOAs) are good candidates.

We chose to compare three of the best known GOAs. Previously, simulated annealing has been used to solve inverse and data assimilation problems in both physical and biogeochemical oceanography (Barth and Wunsch, 1990; Krüger, 1993; Matear, 1995; Vallino, 2000). To our knowledge, the genetic algorithm has only been used once in

Table 1. The state variables and parameters of the PSyDyn model. *e* is a particle reactive chemical element.

Symbol	Description	Unit
State variables		
Fm	mass flux	mg/m^2/d
Fe	flux of *e*	mg/m^2/d
Cpe	concentration of *e* in the small particle phase	mg/m^3
Cde	concentration of *e* in the dissolved phase	mg/m^3
Biogeochemical parameters		
K_{ad}	adsorption rate	/d
K_{ag}	aggregation rate	m^3/mg/d
K_{des}	disaggregation rate	/d
K_{rp}	small particle remineralization rate	/d
K_{rg}	large particle remineralization rate	/d
Physical parameters		
K_z	vertical mixing coefficient	m^2/d
V_p	small particle sinking velocity	m/d
V_g	large particle sinking velocity	m/d

biogeochemical oceanography (Vallino, 2000). Finally, the TRUST (Terminal Repeller Unconstrained Subenergy Tunneling) has never been applied to our field yet, but it would be the fastest (Cétin *et al.,* 1993). The comparison is based on simulated inversion experiments: a pseudo-data set is calculated by a forward integration of our particle cycling model, PSyDyn (PSyDyn: Particulate System Dynamics), with a known set of parameters. We tested and compared the algorithm abilities to recover this set.

The next section starts with a description of the conceptual basis and of the equations of PSyDyn. After mentioning the classical investigations performed in the context of linear inverse problems, we present the peculiar difficulties related to nonlinear inverse problems. This naturally leads to the necessity to resort to global optimization methods. In the third section, we describe the pseudo-data computation and define the cost function. After demonstrating that a gradient descent is unable to simply recover two model parameters, we present each of the GOAs. The experiment results are set out in section four. In section five, we discuss the advantages and drawbacks of each algorithm, with respect to the constraints imposed by the nature of the inverse problem. It leads to the selection of the genetic algorithm to invert PSyDyn with *in situ* data. This will be the subject of another paper.

2. Characterization of the inverse problem

a. The PSyDyn model

The PSyDyn model was initially developed by Ruiz-Pino (1994). It rests on the conceptual scheme proposed by Bacon *et al.* (1985) for the oceanic dissolved-particulate exchanges. It describes the transfers of matter in a 1-D water column that extends from the

base of the euphotic zone to the sediments. The chemical elements are distributed among three reservoirs: the dissolved phase, the suspended and sinking particle phases. This model does not explicitly simulate colloids. The reservoir composition is described by $N_v = 4$ state variables which explicitly depend on time and depth (Table 1): the mass flux (*Fm*), the flux of the studied chemical element (*Fe*), its concentration in the suspended and dissolved phases (*Cpe* and *Cde,* respectively). They exactly correspond to the variables which are measured at sea by classical sampling tools like bottles, filtration pumps and sediment traps. The model dynamics are described by a set of four equations. It is qualified as a dynamical model as each equation describes the *temporal* evolution of a given state variable at a given depth:

$$\frac{\partial Fm}{\partial t} = -V_g \cdot \frac{\partial Fm}{\partial z} - K_{des} \cdot Fm \tag{1}$$

$$\frac{\partial Fe}{\partial t} = -V_g \cdot \frac{\partial Fe}{\partial z} - K_{des} \cdot Fe - K_{rg} \cdot Fe + K_{ag} \cdot Fm \cdot Cpe \tag{2}$$

$$\frac{\partial Cpe}{\partial t} = K_z \cdot \frac{\partial^2 Cpe}{\partial z^2} - V_p \cdot \frac{\partial Cpe}{\partial z} - K_{rp} \cdot Cpe - K_{ag} \cdot \left(\frac{Fm}{V_g}\right) \cdot Cpe + K_{ad} \cdot Cde + K_{des} \cdot \left(\frac{Fe}{V_g}\right) \tag{3}$$

$$\frac{\partial Cde}{\partial t} = K_z \cdot \frac{\partial^2 Cde}{\partial z^2} - K_{ad} \cdot Cde + K_{rp} \cdot Cpe + K_{rg} \cdot \left(\frac{Fe}{V_g}\right) \tag{4}$$

Eqs. 1 and 2 describe the large particle reactions, Eq. 3 the small particle reactions, and Eq. 4 the dissolved phase reactions. The exchanges between the reservoirs are parameterized by the right-hand terms and are summarized in Figure 1. The small particles remove a chemical element *e* from the dissolved phase by chemical adsorption and/or biological active absorption at a rate K_{ad}, according to a first-order kinetics (Eqs. 3 and 4). *e* is added to the dissolved phase from the small and large particle phases by chemical and/or biological remineralization at rates K_{rp} and K_{rg} respectively, and also according to first-order kinetics (Eqs. 2 to 4). The exchange rates of *e* between the small and large particles, as a result of physical, biogeochemical aggregation and disaggregation are K_{ag} and K_{des}. Disaggregation is supposed to follow a first-order kinetics (Eqs. 1 to 3). Aggregation is described as a second-order process: it is supposed to be proportional to the large particle concentration (Athias *et al.,* 1998; Jackson and Lochmann, 1993) and, therefore, to the mass flux (Eqs. 2 and 3). The downward transport of matter also depends on the small and large particle fall, at speeds V_p and V_g, respectively (Eqs. 1 to 3). Finally, the vertical mixing, characterized by the coefficient K_z, influences the dissolved and small

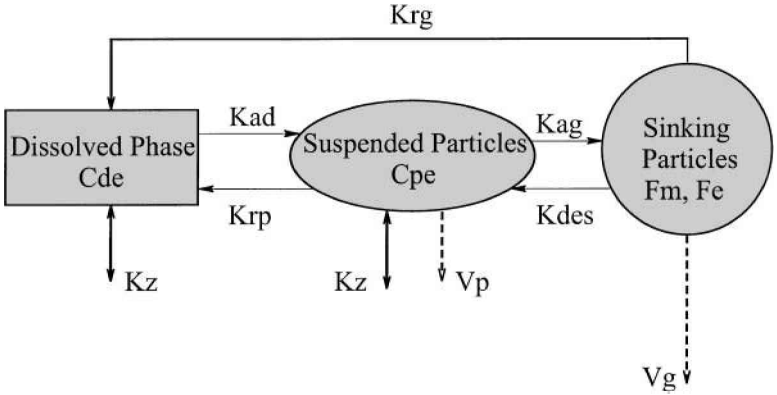

Figure 1. Schematic diagram of the PSyDyn model. The dissolved phase is the material that passes through a 0.65 μm filter. The suspended particle phase gathers particles whose size ranges between 0.65 and 10 μm, and the sinking particle phase particles larger than 10 μm. The state variable and parameter symbols are listed in Table 1.

particle phases (Eqs. 3 and 4). We suppose that the eight model parameters depend neither on time nor on depth.

This model has an aggregated structure: the three model reservoirs are a very simplified representation of the continuous particle spectrum (Jackson *et al.,* 1995). As a consequence, the description of the processes that control the sea water composition is also very crude and is summarized by five biogeochemical exchange reactions. In fact, each of them integrates a whole set of processes, of different nature, spatial and temporal scales. Let us notice that the mass flux equation (Eq. 1) is incomplete: *Fm* is affected neither by aggregation nor by remineralization. This will be discussed and improved in a future paper by comparing PSyDyn outputs to *in situ* data.

The three last equations are said to be coupled, as a given state variable like *Fe* influences its own temporal evolution (Eq. 2), but also the temporal evolution of other state variables, like *Cpe* and *Cde* (Eqs. 3 and 4). This simply represents the couplings between the biogeochemical reservoirs. In addition, the set of equations is nonlinear, as the aggregation term depends on the product of two state variables (Eqs. 2 and 3). Most natural biogeochemical processes are indeed nonlinear. As for most biogeochemical models, PSyDyn rests on a set of Nonlinear Coupled Partial Differential Equations (NCPDE; Athias *et al.,* 1998).

The inverse problem we want to solve consists of estimating the eight model parameters from *in situ* data (Table 1).

b. Linear versus nonlinear inverse problems

We present here a brief overview of the classical investigations and approaches linked to linear inverse problems. Based on this review, we underscore the specificity of nonlinear inverse problems.

In the case of a linear inverse problem, the vector of the N_p true parameters, $\mathbf{x^t}$, is linearly related to an observation vector, $\mathbf{y^o}$, by an operator $\mathbf{H}$ called the data kernel (Menke, 1989):

$$\mathbf{H} \cdot \mathbf{x^t} = \mathbf{y^o}. \tag{5}$$

All inverse methods are based on the minimization of a cost function which depends on the vector of the unknowns, $\mathbf{x}$, and which, for any solution of the model equations $\mathbf{H} \cdot \mathbf{x}$ measures the distance between that solution and the observations. In case the problem is solved according to the least squares criterion (L2 norm), the simplest formulation of the cost function is:

$$J(\mathbf{x}) = (\mathbf{y^o} - \mathbf{H} \cdot \mathbf{x})^T(\mathbf{y^o} - \mathbf{H} \cdot \mathbf{x}). \tag{6}$$

More complex formulations can be defined, by adding *a priori* information on the unknowns, or dynamical constraints like the model stationarity on the annual scale (Matear, 1995). In the ideal case, that is when Eq. 5 provides exactly enough information to determine all the unknowns (even-determined problem), all the sections of $J(\mathbf{x})$ along linear combinations of the x_i components are bowl-shaped ($i = 1, \ldots, N_p$). The solution of the inverse problem minimizes J. It is the unique vector $\mathbf{x^t}$ for which the gradient of the cost function, $\nabla_{\mathbf{x}} J(\mathbf{x})$, is a zero vector. Its components can be independently computed by using elementary linear algebra (Tarantola, 1987). When Eq. 5 does not provide enough information to uniquely determine all the unknowns, the problem is said to be under-determined. On the contrary, when Eq. 5 contains too much information for it to possess an exact solution, the problem is said to be over-determined. In fact, most inverse problems are mixed-determined: some of the unknowns are not constrained by the observations at all (they are said to be nonobservable), others are over-determined and no value will satisfy all the observations exactly. When the problem is linear, the dimensions and rank of $\mathbf{H}$, which can be computed by singular value decomposition, govern whether the system is even-, over-, or under-determined (Menke, 1989). It is possible to identify the linear combinations of the unknowns which are determined by the system, and the linear combinations of the observations which are satisfied by the adjustment. In particular, an unknown x_i for which $\nabla_{x_i} J(\mathbf{x})$ equals zero for $\mathbf{x} \neq \mathbf{x^t}$ is considered nonobservable (the problem is said to be ill-posed). Hence, studying the sensitivity of $J(\mathbf{x})$ to $\mathbf{x}$ allows to limit the search to the components of $\mathbf{x}$ which are observable.

Estimating the parameters of a model becomes a linear inverse problem in the very particular case (i) the equation terms are linear functions of these parameters, and (ii) the model is time-independent. Eqs. 1 to 4 show that the first condition is verified for the dissolved-particulate exchanges. Many publications dedicated to the estimation of particle cycling rates are based on the assumption that the water column is at steady-state (Nozaki *et al.,* 1981; 1987; Bacon and Anderson, 1982; Clegg *et al.,* 1991; Clegg and Whitfield, 1991; Cochran *et al.,* 1993). They rest on time-independent models so as to reproduce the mean features of the chemical element distribution. Provided that the steady-state hypothesis is verified by the assimilated data, the resort to well-known linear inverse techniques is

valid. But as a general rule, biogeochemical *in situ* data time series exhibit time-dependent features on a wide range of timescales (Lampitt and Antia, 1997; Fig. 2). They do not correspond to noise, but rather reflect the process dynamical characteristics. As a consequence, the resort to time-independent models is not valid, and the parameters that were estimated from those models should not be used (Athias *et al.,* 2000).

In the case of nonlinear inverse problems, the data kernel **H** is a function of the unknowns **x**. Some of them are said to be only weakly nonlinear, like the estimation of the parameters of linear dynamical models (Eknes and Evensen, 1997). In this situation, the 1-D slices of $\mathbf{J}(\mathbf{x})$ along linear combinations of the x_i components may have the shape of deformed bowls. A classical approach is based on the determination of an *a priori* solution, that must lie close enough to the solution (that is, inside its attraction basin) for the linearized equations to hold (tangent linear model). The nonlinear inverse problem is then processed by iteratively solving linear inverse problems, from the *a priori* solution. This amounts to progressively going down the local gradients of the cost function, as far as to find the vector $\mathbf{x}^t$ for which $\nabla_{\mathbf{x}} J(\mathbf{x})$ is a zero vector (it is unique in the case of an even-determined problem). Several works in oceanography applied this approach. They are based on various linear inverse techniques. Eknes and Evensen (1997) used the representer method to calculate some parameters of a 1-D Ekman model (Bennett, 1992). Murnane (1994) inverted a time-dependent 1-D particle cycling model by the use of linear algebra. Gradient descent methods, like the steepest descent, and the conjugate gradient methods were also used to estimate the parameters of pelagic ecosystem models. In particular, Spitz *et al.* (1998) and Gunson *et al.* (1999) resorted to the adjoint method to evaluate the local gradients of the cost function.

In fact, estimating the parameters of a biogeochemical dynamical model is generally a strongly nonlinear inverse problem. Athias *et al.* (2000) showed, based on the estimation of oceanic particle cycling rate constants, that it can be related to the model dynamical characteristics. Indeed, the solutions of systems of NCPDE, on which most biogeochemical models are based, are characterized by a rich spectrum of qualitative dynamical behaviors, from simple (fixed points, limit cycles) to more complex ones (chaos, intermittences). The stability of these dynamical regimes and the bifurcation sequences are very sensitive to the parameters which one wants to estimate (Edwards and Brindley, 1996). The cost functions associated to such inverse problems exhibit several local extrema and saddle points, for which the cost gradient is a zero vector. In addition, because of the development of solution instabilities as time goes by, the cost function shape may be all the more complicated and spiky as the length of the assimilation time interval increases (Evensen and Fario, 1997; Mazzega, 2000).

Linearizing this kind of problem and applying gradient descent algorithms requires that one has a sufficiently good knowledge of the real system to be able to define an *a priori* solution that lies inside the attraction basin of the true solution. By definition, those algorithms only allow a search for the minimum nearest to the *a priori* solution. Unfortunately, our knowledge of the oceanic biogeochemistry is quite poor. In addition,

given the sensitivity of the models to their parameters, taking as *a priori* solution parameter values used in other models proves dangerous, even if the model equations seem to be only slightly different. As models are strongly aggregated, it may even be hazardous to use experimentally measured parameters. Indeed, laboratory experiments usually isolate one of the processes that contribute to a model. To get around this difficulty, Spitz *et al.* (1998) repeat their minimization procedure from different *a priori* solutions. But no criterion exists on the number of *a priori* values that must be tried. Such an approach may tend to look like a systematic exploration of the cost function. Moreover, assigning *a priori* statistics to an unknown and including them as an extra term in the cost function is only valid when the model formulation allows that the values of the unknown may be normally distributed. This may be acceptable when the unknown is a model state variable, if it can be shown from observations to be normally distributed. But an unknown, such as a biogeochemical process rate constant, depends strongly on the model formulation and thus cannot be claimed to be normally distributed. Besides, determining the local gradients of very irregular functions requires accurate methods like the adjoint one, which are usually expensive programming methods.

Athias *et al.* (2000) showed that contrary to linear inverse problems, the solutions of highly nonlinear inverse problems cannot be computed by alternating cost minimizations along 1-D directions of the parameter space: the minimization must be performed in the full N_p-dimensional parameter space. Finally, studying whether a problem is even-, under-, or over-determined, as well as studying the observability of parameters are nontrivial tasks in the case of strongly nonlinear problems. Indeed, those properties are local properties of the parameter space and they depend on the set of observations: they can only be studied, for instance, in the neighborhood of the global minimum.

GOAs appear as good candidates to solve highly nonlinear inverse problems, as they are designed for the minimization of cost functions with local minima, and do not require any accurate *a priori* information on the true solution. In view of the previous analysis, we can expect that deterministic algorithms, based on evaluations of cost function derivatives, might be less appropriate than stochastic algorithms based only on cost function calculations. Even if biogeochemical parameters are poorly known, we often have information about their realistic range of values. These constraints will define a bounded search space. Hence, the choice of a GOA will also depend on its ability to manage boundaries. Finally, it must be able to simultaneously estimate numerous parameters.

3. Context of the simulated inversion experiments

a. Computation of an artificial data set

Our ultimate goal is to determine the parameters that control the biogeochemical fluxes through the water column in the northeast tropical Atlantic Ocean, from data collected during the French JGOFS-EUMELI program. From the beginning of 1991 to the end of 1992, *in situ* data characterizing the distribution of biogeochemical tracers between the dissolved, the suspended and sinking particle phases have been collected at an eutrophic

Table 2. Values of the PSyDyn parameters used to compute the synthetic data set ($\mathbf{x}^t$) and limits of the parameter search intervals.

Symbol	Optimal value ($\mathbf{x}^t$)	Search interval	Source
Biogeochemical parameters			
K_{ad} (/d)	0.002	[0; 0.0137]	1
K_{ag} (m^3/mg/d)	0.01597	[0; 0.22]	1
K_{des} (/d)	0.3194	[0; 2.2]	1
K_{rp} (/d)	0.0509	[0; 0.356]	1
K_{rg} (/d)	0.0509	[0; 0.356]	1
Physical parameters			
K_z (m^2/d)	1000	[25; 2000]	2
V_p (m/d)	10	[0; 10]	3
V_g (m/d)	180	[0; 200]	4

1. Athias *et al.* (2000).
2. Zakardjian and Prieur (1994).
3. Whitfield and Turner (1987).
4. Newton (1994), Whitfield and Turner (1987).

site, a mesotrophic site, and an oligotrophic site (Tachikawa *et al.,* 1997, 1999). The objective of this paper is to select an optimization method appropriate for the minimization of the cost functions related to this inverse problem.

With this aim, we generated an artificial data set by a direct integration of PSyDyn, from a known parameter vector, $\mathbf{x}^t$, which will have to be retrieved. This data set and the details of its computation were already presented by Athias *et al.* (2000). The parameter values of $\mathbf{x}^t$ were chosen so as to lie between the extreme values which we found in the literature (Table 2). The outputs simulate the temporal evolution of the mass flux (*Fm*), of the aluminum (Al) flux (*Fe*), of the Al concentrations in the particulate (*Cpe*) and dissolved (*Cde*) phases in the water column of the oligotrophic site, from 1000 to 4200 m (Fig. 2). The model was initialized by a set of four profiles (one for each state variable) computed from *in situ* measurements. We performed a three-years run, and forced the model with boundary conditions at 1000 m. The boundary conditions were generated for one year from *in situ* data, and repeated every year.

We checked that the pseudo-data were of the same magnitude as the *in situ* data. The solution was sampled at $N_d = 2$ depths, and we extracted $N_m = 51$ measurements at each depth: the solution was sampled every 10 days, from the 500th to the 1000th day of the run (Fig. 2). Hence, the artificial data set exactly reproduces the nature of the data collected at sea and mimics their temporal and spatial sampling resolutions. Nevertheless, we did not try to reproduce the gaps in the *in situ* time series. Nor did we add any noise to the pseudo-data to account for measurement or analytical errors. This approach guarantees that at least one solution to the inverse problem exists, and that its cost is zero. We cannot be sure that this solution is unique. In any case, the failure of one of the GOA to find a solution $\mathbf{x}$ so that $J(\mathbf{x}) = 0$ cannot be attributed to noise contamination or model errors. It will

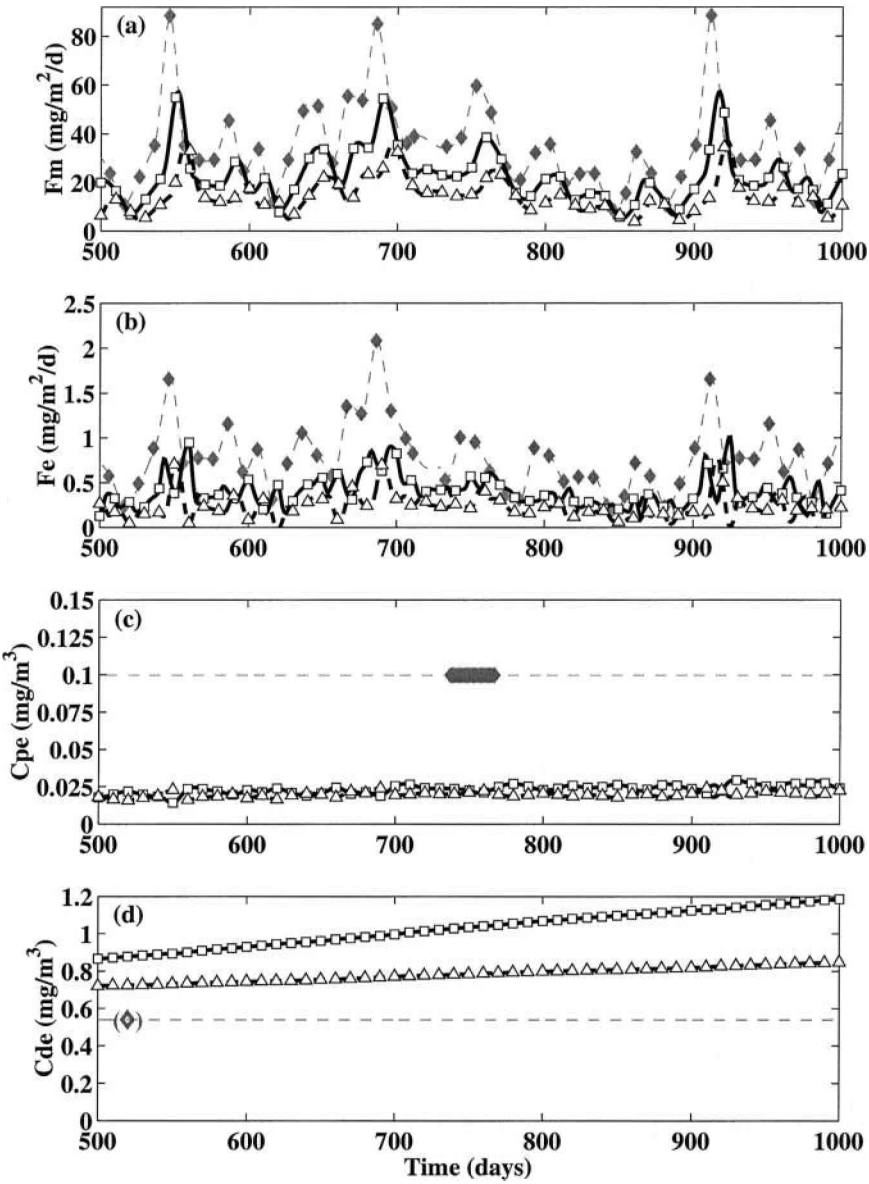

Figure 2. *In situ* data and simulated data computed from the optimal parameter set (Table 2). The gray diamonds represent the *in situ* data used to compute daily boundary conditions at 1000 m (gray dashed line). The black solid and dashed lines are the simulations at 2000 and 3000 m, respectively. The associated symbols (open squares and triangles, respectively) locate the pseudo-data extracted from the simulations. (a) Mass flux and (b) Al flux: data at 1000 m were collected by sediment traps moored at the EUMELI oligotrophic site (21°03′N, 31°10′W) from February 1991 to December 1992 (Bory and Newton, 2000). (c) Particulate Al: data at 1000 m were collected at the same site in June 1992 (Tachikawa *et al.,* 1999). (d) Dissolved Al: data at 1000 m come from the profile of Gelado-Caballero *et al.* (1996) (28°24′N, 15°11′W).

demonstrate a bad adaptation to the inverse problem. The influence of gaps in the time series and of data errors on the parameter estimation will be analyzed in a next paper. We will designate by $\mathbf{W_i^t}$ ($i = 1, \ldots, N_v$) the ($N_d \times N_m$) pseudo-measurement matrices of each state variable, and by V_i the variances of these pseudo-measurements.

b. Definition of the cost function

The simulated inversion experiments are based on the definition of a cost function according to a L2 norm:

$$\mathbf{J(x)} = (N_d \cdot N_v \cdot N_m)^{-1} \cdot \sum_{i=1}^{N_v} (\mathbf{W_i(x)} - \mathbf{W_i^t})^T \cdot \mathbf{C_i^t} \cdot (\mathbf{W_i(x)} - \mathbf{W_i^t}). \quad (7)$$

The N_v matrices $\mathbf{W_i(x)}$ are computed by integrating the model from the trial parameter vector $\mathbf{x}$, with the same initial and boundary conditions and the same sampling strategy as for the computation of the pseudo-data set. The N_v matrices $\mathbf{C_i^t}$ are diagonal, and their elements are the inverse of the variances V_i. They are introduced to make the cost function nondimensional and to give the same importance to all kinds of data, but the pseudo-data are supposed to be independent. We assumed that no *a priori* information on the distribution of the unknowns was available and that we were only able to restrict the parameter space according to the knowledge of their realistic ranges of values.

Figure 3 shows a 2-D section of J. It was calculated from parameter vectors whose values for K_{ad}, K_{rp}, K_{rg}, K_z, V_p and V_g were those of $\mathbf{x^t}$ (Table 2). The values for K_{des} and K_{ag} are regularly distributed in the ranges 0–0.9369/d and 0–0.04684 m^3/mg/d, respectively. These intervals are smaller than the whole range of realistic values for K_{des} and K_{ag}, because of the high computational cost of the mapmaking (Athias *et al.,* 2000). Nevertheless, this small slice allows us to underscore some of the characteristics of nonlinear inverse problems which were mentioned in Section 2b. First, the slice is not bowl-shaped. It exhibits a Global Minimum (GM), whose coordinates (0.3194, 0.01597) equal the K_{des} and K_{ag} values of $\mathbf{x^t}$ (Table 2). It is the solution of the inverse problem, whose cost is zero. The attraction basin of the GM is wide and flat, and its axis is parallel to the K_{ag} axis. This means that K_{ag} is poorly resolved in this region. The slice displays another valley, whose axis is also parallel to the K_{ag} axis, and which is associated to large values of K_{des}($K_{des} \approx 0.6$/d). It is the attraction basin of a secondary minimum, whose coordinates are (0.6175, 0.00319), and whose cost is 38.2468. The point of coordinates (0.5538, 0.00100) is a saddle point. Hence, this restricted 2-D section of J displays three points which obey the stopping criterion of the gradient descent algorithms, that is for which $\nabla_{\mathbf{x}} J(\mathbf{x})$ is a zero vector. Because of numerical instabilities, the cost function cannot be computed for K_{des} values larger than 0.67/d.

We tested the ability of a gradient descent algorithm to minimize this 2-D section of J, that is, to find the values of K_{des} and K_{ag} used to compute the pseudo-data set. Figure 3a shows the minimization path followed by a nonlinear conjugate gradient (Fletcher and Reeves, 1964; Gilbert and Nocedal, 1992) from the point of coordinates (0.6, 0.04684).

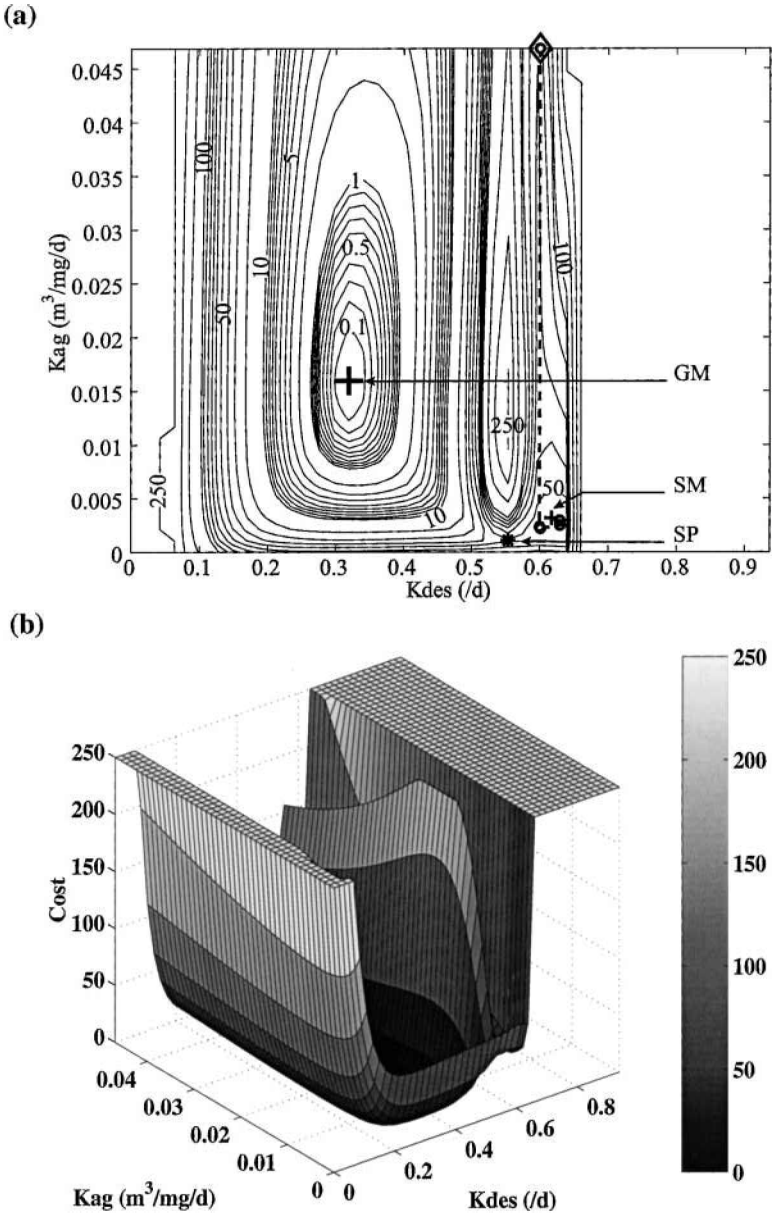

Figure 3. Representations of a 2-D section of the cost function computed from the pseudo-data set, as a function of K_{des} and K_{ag}. The values of K_{ad}, K_{rp}, K_{rg}, K_z, V_p and V_g are fixed to the optimal values (Table 2). (a) 2-D representation. The large cross locates the global minimum (GM), the small one a secondary minimum (SM), and the star a saddle point (SP). The large gray dots and the dashed line indicate a nonlinear conjugate gradient minimization path. The starting point is located by the diamond of coordinates (0.6, 0.04684), and the final point is the secondary minimum. (b) 3-D representation.

After 171 cost calculations, the algorithm reached the nearest minimum to the starting point, that is the secondary minimum. This experiment demonstrates that local minimization methods are not suitable to solve our parameter estimation problem.

We first computed three simulated inversions to compare the abilities of the three GOAs to minimize this section. This preliminary 2-D study is essential to graphically analyze the algorithm behaviors.

c. Presentation of the global optimization algorithms

i. The TRUST. The TRUST algorithm was developed by Cétin *et al.* (1993). The solution of the optimization is formulated as the solution of a deterministic dynamical system incorporating terminal repellers and subenergy tunneling functions. Schematically, the algorithm alternates phases during which it goes down the gradient of the cost function, with phases during which it digs tunnels through the cost function bumps. Good presentations of the TRUST can be found in Barhen and Protopopescu (1996) and in Barhen *et al.* (1997). As the TRUST is still poorly known, we recall its principles.

Let us suppose first that J is a function of only one variable, x, which ranges from x_{min} to x_{max}. To begin with, one of the search space limits, x^*($x^* = x_{min}$ or x_{max}), is used to define a subenergy tunneling function $E_{sub}(x,x^*)$ by a nonlinear transformation of J:

$$E_{sub}(x,x^*) = \log\left[\frac{1}{1 + \exp\left(-(\hat{J}(x) + a)\right)}\right]. \tag{8}$$

In this expression, $\hat{J}(x) = J(x) - J(x^*)$ and a is a constant. $E_{sub}(x,x^*)$ has the same extrema as the original cost function $J(x)$, and their relative order is the same. In case $J(x) > J(x^*)$, $E_{sub}(x,x^*)$ tends toward zero. Thus, $E_{sub}(x,x^*)$ has the same shape as $J(x)$, but the segments where $J(x)$ is higher than $J(x^*)$ are flattened. In the case where x^* is not a local minimum of J, the search space $[x_{min},x_{max}]$ is explored by integrating a dynamical system ruled by the equation:

$$\dot{x} = \frac{\partial x}{\partial t} = -\frac{\partial E_{sub}(x,x^*)}{\partial x} \tag{9}$$

from the initial condition $x^* + \varepsilon$ where ε is a small perturbation which drives the system into the search domain. The dynamical system goes down the gradient of $E_{sub}(x,x^*)$, until it detects a first local minimum. It is called x^* in turn, and is used to define a new subenergy tunneling function (Eq. 8). But this time, the subenergy function is flat in the neighborhood of x^*. This may also happen at the beginning, if x_{min} or x_{max} is a local minimum of J. In these cases, a repeller term is added to the subenergy function that allows to move the dynamical system away from x^*:

$$E_{rep}(x,x^*) = -(3/4)b(x - x^*)^{4/3}u(\hat{J}(x)) \tag{10}$$

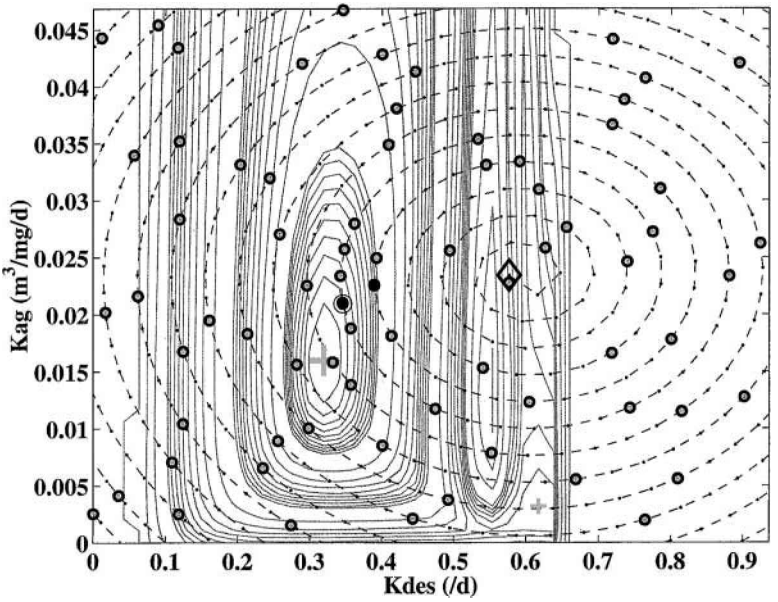

Figure 4. Optimization of the 2-D cost section by the TRUST. The TRUST minimizes the 1-D function $\varphi(\alpha)$ obtained by sampling the section by a discrete Archimedes spiral (dashed and dotted line) from the diamond of coordinates (0.6, 0.0234). The algorithm evaluated $\varphi(\alpha)$ and its gradient on the large gray dots. The black dots are local minima of the successive subernergy functions. The last one is encircled and represents the estimate of the global minimum. The true global minimum is located by the large gray cross, and the secondary minimum by the smaller one.

where b is called the repeller force and u is the Heaviside step function ($u(y) = 0$ if $y < 0$ and $u(y) = 1$ if $y \geq 0$). Now, the following dynamical system is integrated:

$$\dot{x} = \frac{\partial x}{\partial t} = -\frac{\partial E_{sub}(x,x^*)}{\partial x} - \frac{\partial E_{rep}(x,x^*)}{\partial x} \tag{11}$$

from the point located on $x^* + \varepsilon$. The algorithm is first repelled from x^*, and goes through a cost bump using a tunnel. Then, it goes down the gradient of $E_{sub}(x,x^*)$ in areas where $J(x) < J(x^*)$, until it reaches a new local minimum, which will replace the previous one. Hence, the exploration of the cost function consists of alternating descent gradient phases, with tunneling phases, until the whole search space has been explored. The last local minimum found by the TRUST is the global one.

To apply this algorithm to the minimization of a function of several variables, this function can be approximated by a function of only one variable, by sampling the original function by a hyperspiral (Ammar and Cherruault, 1993). To minimize the 2-D slice of J, we first normalized the search space, so that the two parameters vary between -1 and 1. Then, the section was sampled along a discretized Archimedes spiral. It provides an approximate function φ, which only depends on the angle α (Fig. 4). The spiral is centered

in the attraction basin of the secondary minimum, on the point of coordinates (0.6, 0.04684). The distance between the spires in each direction, as well as the arc between two points on the spires is 0.1. The constant a (Eq. 8) was set to 2, as recommended by Cétin *et al.* (1993). The repeller force (Eq. 10) was computed so that the dynamical system is moved 2 spire steps away from all the repellers, which led to $b = 6.84$. The successive dynamical systems were integrated using the finite difference method. The arc of the initial perturbation ε was set to 0.025. A point of the parameter space is qualified as a local minimum if the local gradient norm $|d\varphi/d\alpha|$ is lower than a threshold, which was initially set to 0.5. The algorithm stops when the search space is entirely explored, that is after covering 15.75 spires. Some parts of the more peripheral spires are found out of the domain. For the optimization to be still limited to the preset domain, the cost function was projected on both sides of each boundary.

ii. The Simulated Annealing. The Simulated Annealing (SA) is a stochastic exploration algorithm based on the statistical model of the thermodynamic process for growing crystals. A perfect homogeneous crystal lattice corresponds to the global minimum state of energy of a material. To obtain such a state, one can first heat the material up to an initial temperature T_o until it reaches an amorphous liquid state, and then cool it very slowly. As the material gets cooler, atoms rearrange themselves randomly. Each new configuration at a lower energy level than the previous one is unconditionally accepted by the system. But there is a nonvanishing probability for a configuration at a higher energy level to be accepted. This probability is proportional to the Boltzmann factor $\exp(-\Delta E/(k_B T))$, where ΔE is the difference of energy between two configurations, T is the temperature, and k_B is the Boltzmann constant. The adaptation of this model to optimization problems was initiated by Kirkpatrick *et al.* (1983). The cost function J plays the role of the energy, and the temperature is replaced by a control parameter, which is still called T. The Metropolis function is used to define the probability of accepting a higher cost:

$$f(J(\mathbf{x_1}),J(\mathbf{x_2}),T) = \exp\left[-\frac{(J(\mathbf{x_2}) - J(\mathbf{x_1}))}{T}\right] \tag{12}$$

where $\mathbf{x_1}$ is the current parameter set, $\mathbf{x_2}$ is a new parameter set, so that $J(\mathbf{x_2}) > J(\mathbf{x_1})$. The parameter set $\mathbf{x_2}$ is accepted if the Metropolis function is greater than a random number between 0 and 1. Good presentations of the SA can be found in Barth and Wunsch (1990) and in Krüger (1993).

We used the SA code developed by Allen *et al.* (1996). The algorithm must be supplied with the extreme values between which each parameter can lie. Then, this search domain is normalized, so that each parameter varies between -1 and 1. Following Krüger (1993), eight algorithmic parameters must be tuned. To calculate the initial temperature T_o, 1000 simulations are performed by randomly perturbing the initial parameter set provided by the

user. Then, the mean of the cost function increases, $<\Delta J^+>$, is computed and T_o is given by:

$$T_o = -\frac{<\Delta J^+>}{\ln \chi} \tag{13}$$

where χ is the probability of parameter sets with higher costs to be accepted. It was set to 0.9. The cost function exploration starts from the initial parameter set, and the maximum random step size σ_o is set to 1. The annealing steps, during which T is constant, last until 200 parameter vectors are assessed, or until 100 parameter vectors are accepted. Then T and σ_o are decreased, by a factor of 0.9. The exploration continues until an annealing step verifies two stopping conditions: it must not have led to an absolute cost reduction, and the ratio of the number of accepted cost increases to the number of total cost increases must be lower than 1%.

iii. The Genetic Algorithm. The Genetic Algorithm (GA) is a quasi-stochastic exploration algorithm. It is based on operators that simulate the natural selection and genetics mechanisms. Each vector of model parameters **x** is regarded as the genotype of an individual, and is coded by a string of values in binary form, which is compared with a chromosome. The model prediction generated from **x** is the analogue of the individual phenotype. His fitness, or adaptation to his environment is positive, and defined as the opposite of $J(\mathbf{x})$. The initialization of the canonical GA begins by generating a random population of individuals. In each cycle of genetic operation, a subsequent population is created from the genes of the current one. Couples of parents are first selected from the current population. Then, their genes are mixed and recombined, by the mutation and crossover processes, to produce the offspring of the next generation. It is expected that with this manipulation of the genes, the best individual will create a larger number of offspring, according to Darwin's principles of the survival of the fittest in nature. Good presentations of the GA can be found in Goldberg (1989), Allen *et al.* (1996) and in Man *et al.* (1997).

We used the GA driver developed by Carroll (1996). According to Carroll (1996), most of the algorithmic parameters are robust, so we mainly followed his recommendations. We opted for the micro-GA driver, which is slightly different from the canonical GA. It has the advantage to assess small populations, which limits the calculation costs. Regardless of the number of unknown parameters, each population has 5 individuals, which is 20 to 40 times smaller than for the classical GA. At each generation, 5 couples of parents are defined (some individuals may belong to several couples, and others may not belong to any couple) and a child will be born from each couple. The micro-GA option supposes that the recombinations of the parents' genes are based only on crossover processes. The crossover probability was set to 0.5, and crossovers are uniform: the crossover mask is randomly generated, which leads to a mixture of the parents' genes. In addition, the micro-GA option includes an evaluation of the population convergence. At each generation, the number of the chromosome bits of the whole population that are different from those of the best

Table 3. Results and computational requirements of the optimization tests. The pseudo-data are supposed to be free of error.

	TRUST	Simulated annealing	Genetic algorithm
2-parameter optimizations			
Number of cost function calls	2019	22627	2000
Highest computed cost	3278.3182	9917.5037	7571.6471
Final cost	0.0000	0.0000	0.0000
8-parameter optimizations			
Number of cost function calls	—	17244	1000
Highest computed cost	—	9970.4787	8111.3732
Final cost	—	94.7349	1.3092

individual is calculated. If the proportion of those bits is lower than 5%, the current population is replaced by a new one, made of the best individual plus 4 randomly generated individuals. The reproduction is elitist: the lifetime of the best individual is not limited to one generation. Finally, the algorithm stops when 200 generations have been assessed.

4. Results

We first present the results of the minimization of the 2-D cost slice. As will be discussed in Section 5, the optimization of more than two parameters can be conceived in practice only with the SA and the GA. So, we will also compare their ability to recover all the PSyDyn parameters. It must be emphasized that the results presented in this section partly depend on the choice of the algorithmic parameters. As mentioned above, most of them were set to their recommended values, and we did not try to optimize them. On the contrary, these results do not depend on the norm used to formulate the cost function.

a. Two-parameter optimizations

i. Optimization by the TRUST. The TRUST succeeded in finding the GM of the section after 2019 cost computations (Table 3), although it started inside the basin of the secondary minimum. We performed six successive optimizations on more and more restricted search domains centered on the GM detected at the previous optimization. Figure 4 shows the progress of the algorithm during the first one. Let us note that the TRUST optimizes the 1-D function $\varphi(\alpha)$ which was obtained after sampling the section along the spiral from the center. The large gray dots represent the points where the TRUST evaluated $\varphi(\alpha)$, as well as the gradient of the successive subenergy functions. It detected two local minima, whose cost decreases from the spiral center. The last one is the GM estimate (Fig. 4), and its coordinates are (0.3453, 0.02106). Then, the rest of the function seemed flat to the TRUST; therefore, it left the search domain. The cost of the parameter sets generating model instabilities was fixed to twice the cost of the current repeller, and the cost gradient was fixed to twice the gradient threshold. Hence, the TRUST dug tunnels through the regions of the parameter space where the model cannot be integrated. Figure 5a shows the evolution

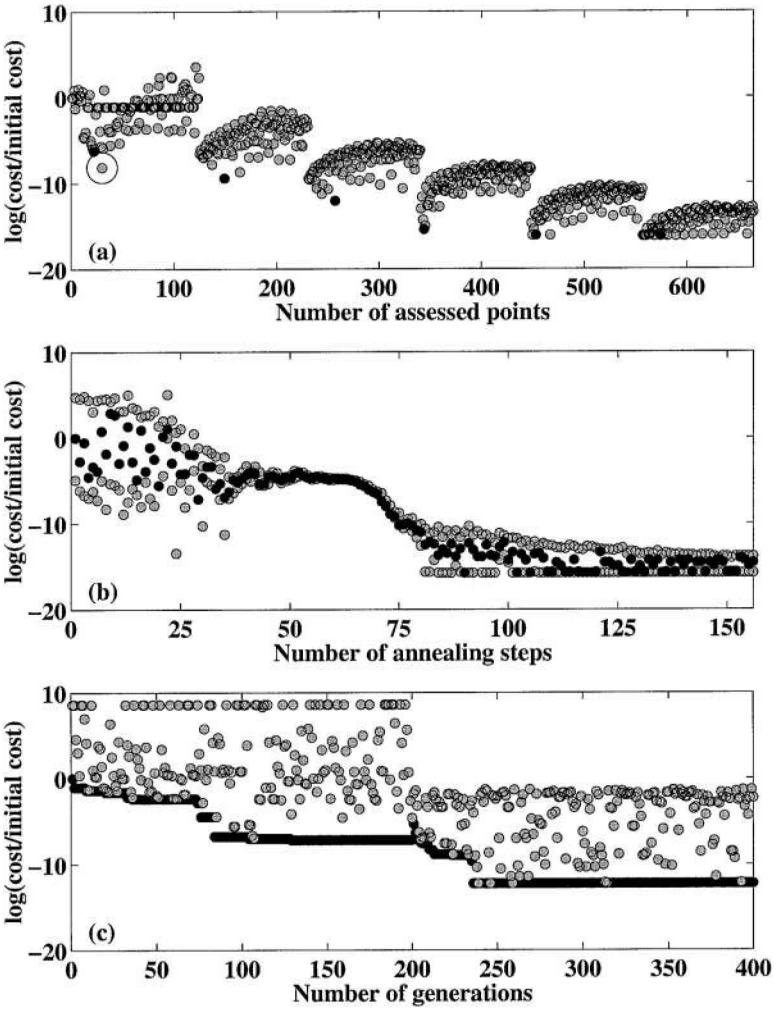

Figure 5. Evolution of the cost during the optimizations of the 2-D cost section. (a) Optimization by the TRUST: the cost of all the assessed points are represented by the gray dots. The black dots designate the global minimum estimates for each of the six optimizations. The encircled gray dot shows that the TRUST missed a better global minimum estimate. (b) Optimization by the simulated annealing. The black dots represent the cost of the last point assessed by the algorithm at the end of each annealing step. The gray ones represent the lowest and the highest costs for each step. (c) Optimization by the genetic algorithm: the black and gray dots indicate the costs of the best (worth) individuals of each generation, respectively.

of the cost during the whole run. Six groups of points can be distinguished which are related to the six successive optimizations. The GM estimates found at each optimization always correspond to the lowest cost, except for the first one, where the TRUST found a parameter set whose cost was lower than the cost of GM estimate. This set can be located

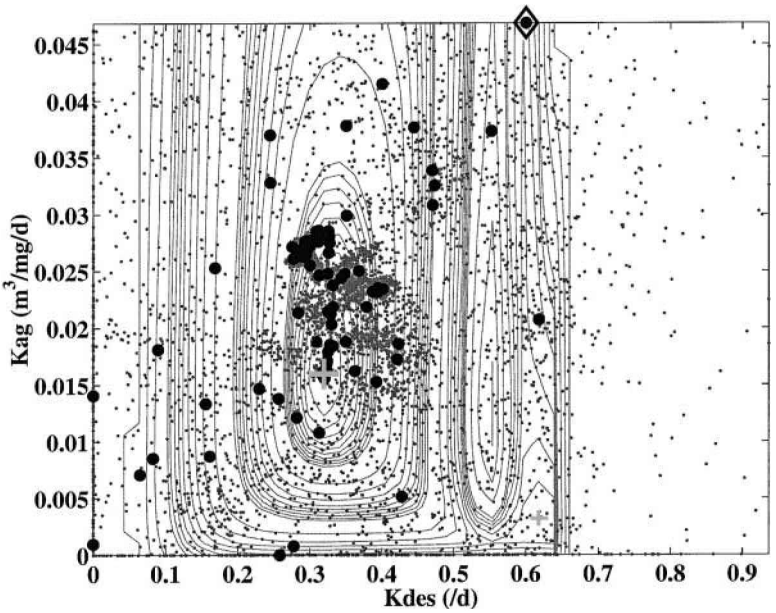

Figure 6. Optimization of the 2-D cost section by the simulated annealing. The diamond of coordinates (0.6, 0.04684) is the starting point. The algorithm evaluated 21627 parameter vectors (small gray dots). The large black dots are the last points assessed at each of the 156 annealing steps. The large gray cross locates the true global minimum, and the smaller one the secondary minimum.

on Figure 4: its coordinates are (0.3318, 0.01581). In fact, the gradient norm $|d\varphi/d\alpha|$ on this point is higher than the threshold, so the TRUST kept going. As the spiral is discrete, the TRUST went over the valley of the GM and missed a better local minimum. Then, we performed a second optimization, on a restricted search domain. The new domain was centered on the first GM estimate, and it corresponded to a contraction of 70% in the K_{des} and K_{ag} directions. The contraction ratio was the same for each new optimization. It takes into account the fact that the TRUST can miss the best GM estimate.

ii. Optimization by the Simulated Annealing. The SA succeeded in finding the global minimum of the 2-D section. It required 156 annealing steps and 22627 evaluations of the cost (Table 3). As recommended, the 1000 first ones were dedicated to the evaluation of the initial temperature, which led to $T_o = 30498$. The progress of the algorithm into the search domain is represented in Figure 6. The SA started inside the secondary valley, on the point of coordinates (0.6, 0.04684), that is from the same point as the nonlinear conjugate gradient in Section 3b (Fig. 3a). It explored the whole parameter space, before heading for the attraction basin of the GM, without being trapped inside the secondary basin. The SA was not disturbed by parameter sets for which the model is unstable: when a random step led the SA on such a set, the new set was refused and the algorithm had to make another

random trial. Each time a random step took it out of the domain, it was brought back where the boundary was crossed. It makes the SA stay inside the search domain, but it also makes it linger over two of the four boundaries. A careful analysis of the assessed points shows that a significant part of the cost evaluations is redundant. Figure 5b displays the evolution of the cost during the optimization. Globally, the cost decreases progressively. It clearly shows that the algorithm accepts some cost increases, but that the number of increases diminishes as T gets lower. Let us notice that the last parameter set of an annealing step does not systematically correspond to the lowest cost. This is related to the criteria controlling the transition from one annealing step to the next one. As a consequence, no information on the parameter set which has provided the lowest cost is kept. From the 50th to the 65th annealing step, all the costs are very close, between 0.4 and 0.6. At this time, the SA was exploring the area near (0.3, 0.027), where the cost gradient is very low. The GM was found during the 90th step: since this step, the first stopping criterion was checked. The algorithm went on oscillating around the solution until the 156th step, when the proportion of accepted increases got lower than 1%.

To test the robustness of the SA with respect to the starting point, we performed two other runs with the same algorithmic parameters, but from the corners of coordinates (0, 0) and (0, 0.04684). We had empirically noticed that the chance for the SA to be trapped by a local minimum was all the lower as the SA started on a search space boundary. The 1000 first cost evaluations provided initial temperatures which were quite equal to the previous one (T_o = 30285 and T_o = 30274, respectively). The SA found the GM again, after performing 18293 and 22721 cost computations, respectively.

iii. Optimization by the Genetic Algorithm. The GA found the GM of the cost section after 2000 evaluations of the cost (Table 3). As for the TRUST, we had to perform two successive optimizations. Two hundred generations of 5 individuals were assessed for each of them. Figure 7 shows how the GA explored the section during the first optimization. The individuals of the first population are scattered. Although one of them is found inside the basin of the secondary minimum, the GA was not trapped. The algorithm visited all the regions. When a new individual is born, who corresponds to a parameter set for which the model is unstable, it is allocated a very low fitness (that is, a very high cost, equal to 10000), to limit the transmission of its genes. In addition, the boundaries of the search domain are indicated to the algorithm from the start, to code the set of parameters as a chromosome, so no child can be born out of the domain. The convergence toward the attraction basin of the GM is very fast, so that it is not possible to distinguish all the 200 best individuals. This optimization led to a first estimate of the GM, whose coordinates are (0.3207, 0.01612). Figure 5c shows the evolution of the costs of the worst and of the best individuals of each generation, during the two optimizations. It shows that the decrease of the cost subsides from the 84th generation of the first one. From the 129th generation, the best individual of the populations remains the same: this represents the elitist nature of the reproduction. We performed a next optimization on a restricted domain, centered on the

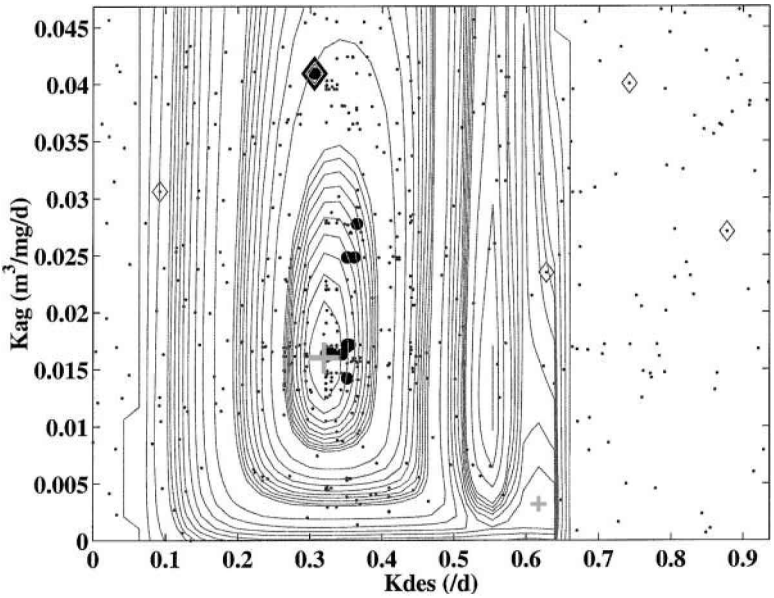

Figure 7. Optimization of the 2-D cost section by the genetic algorithm. The algorithm assessed 1000 parameter vectors (small gray dots). The large black ones represent the best individuals of each of the 200 generations. The diamonds locate the five individuals of the first generation. The large gray cross locates the true global minimum and the small one the secondary minimum.

first GM estimate. The new domain limits were empirically determined. We first identified the generation from which the relative decrease of the best individual fitness is never higher than 30%. It is the 84th one (Fig. 5c). We computed the distance, in the K_{des} and the K_{ag} directions between the GM estimate, and each of the best individuals of the 84th to the 200th generations. The maximal distances in each direction defined the dimensions of the new domain. It led to a contraction of 90 and 98% in the K_{des} and K_{ag} directions, respectively. Figure 5c shows that the decrease of the cost is very slow as soon as the 40th generation of the second optimization. The true GM is found at the 168th generation. Let us emphasize that the contraction of the search domain is not essential for the success of the GA.

b. Delimiting a subspace of acceptable parameters

The 2-D simulated inversions performed to test the GOAs were based on an artificial and noise-free data set. Hence, we knew that at least one solution existed: it is the couple (K_{des},K_{ad}) used to compute the pseudo-data set, and therefore for which the cost is zero.

But when one is using real *in situ* data, which are characterized by measurement and analytical errors, searching for the only cost function GM is not sufficient anymore. Indeed, given the uncertainty of the data, a cost value σ_J^2 can be computed, so that all the parameter sets for which the cost is equal to or smaller than σ_J^2 explain the data in an acceptable way.

Table 4. Results of the optimizations of K_{des} and K_{ag}, in the case of a 20% rms uncertainty on the pseudo-measurements of *Fm* and *Fe* and a 10% rms uncertainty on the ones of *Cpe* and *Cde*. The acceptable intervals are compared on Figure 8.

	TRUST	Simulated annealing	Genetic algorithm
Cost of the GM estimate	0.0077	0.0000	0.0014
Number of cost function calls	699	22627	1000
K_{des} (K^t_{des} = **0.3194/d**)			
Estimate (/d)	0.3203	0.3194	0.3207
Acceptable interval (/d)	[0.3105; 0.3273]	[0.3060; 0.3311]	[0.3090; 0.3336]
K_{ag} (K^t_{ag} = **0.01597 m^3/mg/d**)			
Estimate (m^3/mg/d)	0.01725	0.01597	0.01612
Acceptable interval (m^3/mg/d)	[0.01593; 0.01862]	[0.01398; 0.01863]	[0.01374; 0.01806]

The term "acceptable" is defined with respect to the information brought by the data and to the constraints introduced into the formulation of the cost function. In this case, optimizing a cost function consists not only of estimating the GM, but also of delimiting the subspace of acceptable parameter vectors, E_a. As the three tested GOAs explore the cost function, they should be able to delimit E_a. In particular, when E_a is connected, we are able to compute intervals of acceptable values for each optimized parameter. It provides a first evaluation of the size of E_a.

Still within the framework of the optimization of the 2-D cost section, we compared the acceptable intervals computed by the three GOAs, supposing that the pseudo-data are characterized by realistic uncertainties: 20% rms on the pseudo-measurements of the mass flux (*Fm*) and of the Al flux (*Fe*) (Tachikawa *et al.,* 1997), and 10% rms on the ones of the particulate Al (*Cpe,* Tachikawa *et al.,* 1999) and of the dissolved Al (*Cde,* Gelado-Caballero *et al.,* 1996). Considering that the pseudo-data are Gaussian, and considering that J is normalized both by the number of pseudo-data, and by the variance of the pseudo-measurements (Eq. 7), all the parameter sets whose cost is lower than $\sigma_J^2 = (0.2)^2 = 0.04$ can be considered as acceptable. The acceptable intervals of the optimized parameters were computed in the same way for the three algorithms. First, we assumed that we did not know the solution of the inverse problem. Hence, the optimizations were considered finished as soon as an acceptable estimate of the GM was found. Then, we extracted all the acceptable parameter sets. For each component of these sets, the minimum and the maximum values delimit the acceptable interval. The results are summarized in Table 4, and the accepted intervals are represented on Figure 8.

At the end of the second optimization, after 699 cost computations, the TRUST provided an estimate of the GM whose cost was $0.0077 < \sigma_J^2$, and whose coordinates are (0.3203, 0.01725). The acceptable values of K_{des} range between 0.3105 and 0.3273/d, and those of K_{ag} between 0.01593 and 0.01862 m^3/mg/d. The SA exactly found the true GM in one step. The acceptable values of K_{des} range between 0.3060 and 0.3311/d and those of K_{ag} between 0.01398 and 0.01863 m^3/mg/d. The first optimization of the 2-D section by the

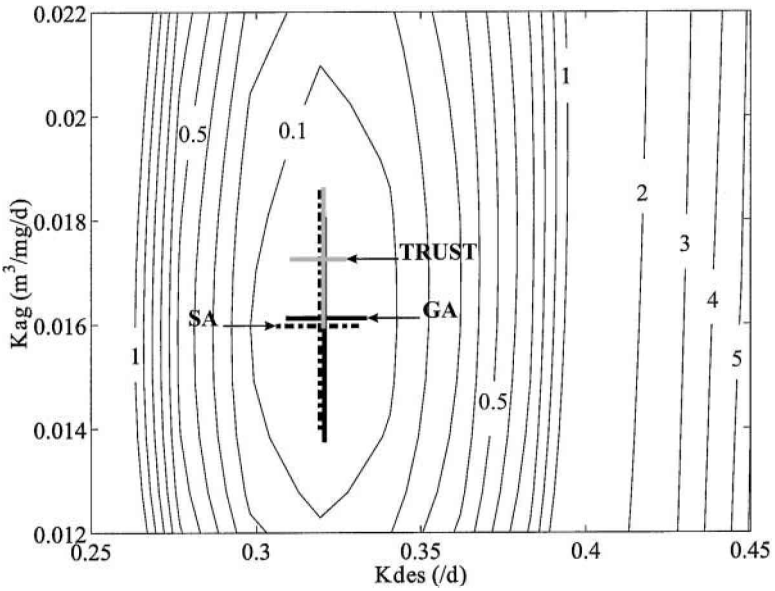

Figure 8. Comparison of the acceptable parameter estimations by the TRUST (gray solid line), by the simulated annealing (SA, black dashed and dotted line) and by the genetic algorithm (GA, black solid line) in the case of a 20% rms uncertainty on the pseudo-measurements of *Fm* and *Fe* and of a 10% rms uncertainty on the ones of *Cpe* and *Cde*. The global minimum estimate of the simulated annealing corresponds to the true one.

GA provided an estimate of the GM whose cost was $0.0014 < \sigma_J^2$, and whose coordinates are (0.3207, 0.01612). The acceptable values for K_{des} range between 0.3090 and 0.3336/d and those for K_{ag} between 0.01374 and 0.01806 m^3/mg/d.

c. Optimizing all the model parameters

As will be emphasized in the next section, it quickly appeared to us that the complexity of the TRUST would make its application to the estimation of more than two parameters very demanding in practice. Moreover, the results of the previous experiments suggest that it may not be worth the effort. Given its deterministic nature, we can expect poor results. On the contrary, without considering possible limitations imposed by their computational requirements, the SA and the GA are both theoretically designed for high dimensional optimizations. We performed two further simulated inversions, with the same pseudo-data set, to compare the ability of the two algorithms to recover the eight parameters of PSyDyn. The algorithm procedural parameters were not altered. The search space was delimited by the intervals of values found in the literature for each parameter (Table 2). The results are summarized in Table 3.

The SA started from the point whose coordinates correspond to the lower limits of the search intervals of each parameter. The computation of the initial temperature led to $T_o =$

64497. The cost slowly decreased from 395.7762 to 94.7349, and stopped after performing 17244 cost evaluations, during 114 annealing steps. Figure 9a shows that as soon as the 70th annealing step, the SA was trapped into the attraction basin of a local minimum. The true solution might have been found by tuning the algorithmic parameters, but this was not our objective.

With the GA, the assessment of 200 generations (that is, 1000 cost computations) allowed us to quickly find a parameter vector with a cost of 1.3092 (Fig. 9b) quite close to the solution (Fig. 9c). From the optimization of the 2-D cost section, we had noticed that the convergence of the GA in the neighborhood of the GM is slow. So, instead of continuing the minimization with the GA, we started the nonlinear conjugate gradient algorithm (Gilbert and Nocedal, 1992) from the best parameter set proposed by the GA for the 200th generation. The GM was found after 741 further cost computations. This means that the GA had succeeded in quickly locating the attraction basin of the true solution. In comparison, continuing the GA minimization up to 1650 generations (that is, 7250 further cost calculations) would have permitted to reduce the cost only to 0.66.

5. Discussion

The tests presented in the previous section show that the three GOAs found the GM of the 2-D cost section when the pseudo-data were supposed to be free of error and with a precision limited only by the precision of the computer (Table 3). When realistic data uncertainties were taken into account, they all computed acceptable intervals that surrounded the true values of K_{des} and K_{ag} (Table 4, Fig. 8). Contrary to what could be expected from classical inverse methods based on gradient descent algorithms, the GOAs have not been trapped by the secondary minimum, although they started from inside its attraction basin. Neither were they trapped by the saddle point. They proved to be able to tackle the low gradient inside the attraction basin of the GM. In addition, they contented themselves with *a priori* information on the realistic ranges of values of K_{des} and K_{ag}. In this section, we compare the behavior of the GOAs during the 2-D and 8-D optimizations. We want an algorithm which gives good results, but which is also easy to implement. We paid attention to the way the GOAs reacted to the constraints mentioned at the end of Section 2b, to their dependence upon the cost function structure, and to their ability to optimize more than two parameters. It must be emphasized that even if afterward it seems that some of the algorithm drawbacks could have been anticipated, testing all of them in the context of simulated inversion experiments was essential. Their drawbacks and advantages interact strongly, so that it was difficult to know in advance which one would be the most efficient and best adapted.

The TRUST required approximately as much cost computations as the GA to find the GM of the 2-D cost slice (Table 3). When data errors were taken into account, it proved to be the fastest to locate the basin of acceptable parameters (Table 4). When compared to the SA and the GA, the TRUST tends to minimize the size of the acceptable parameter

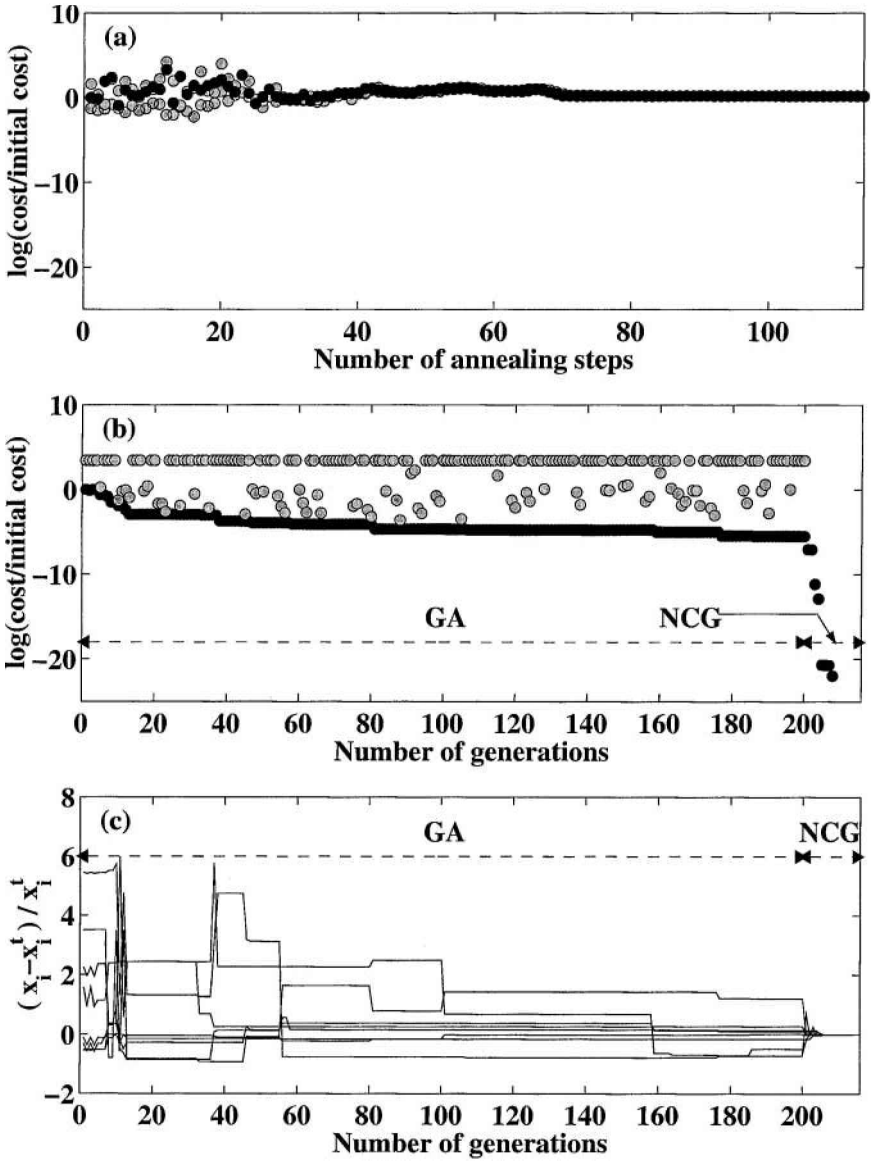

Figure 9. Optimizations of the eight PSyDyn parameters. (a) Cost evolution during the optimization by the simulated annealing. The black dots represent the cost of the last point assessed by the algorithm at the end of each annealing step. The gray ones represent for each step the lowest and the highest costs. (b) Cost evolution during the optimization by the genetic algorithm (GA) over 200 generations, then by a nonlinear conjugate gradient algorithm (NCG) over 16 iterates. The black and gray dots indicate the costs of the best (worth) individuals for each generation, respectively. (c) Recovery of the parameters by the genetic algorithm (GA) coupled to a nonlinear conjugate gradient algorithm (NCG). For each parameter, the difference between the estimated (x_i) and true (x_i^t) values is normalized with respect to the true value.

subspace (Fig. 8). This depends on the sampling resolution of the cost by the 1-D spiral. It may not be a major drawback: it could be envisaged to improve the location of the acceptable space boundaries by a further exploration in the vicinity of the GM estimate. In fact, the TRUST displays more serious disadvantages in the sight of our problem. First, as it is a deterministic algorithm, it not only requires punctual evaluations of the 1-D function φ, but also of its first derivative (Eqs. 9 and 11). In addition, the repeller force b (Eq. 10) rigorously depends on the second derivative of φ with respect to the repellers x^*. We simplified the original algorithm by considering b as a constant, but the validity of this simplification might be problem-dependent. We showed that the convergence of the TRUST is strongly related to the gradient norm threshold used to select local minima (Fig. 5), which depends on the cost function structure. The gradients of φ were estimated by the finite difference method. We expect that this crude method may become insufficient in the case of very irregular functions. It would be necessary to resort to more accurate methods, with high programming requirements, like the adjoint one. This would make the TRUST far less attractive. As the cost function associated to the estimation of biogeochemical parameters commonly display spiky areas, and as we do not have any direct access to the cost function derivatives, we think that the deterministic methods are generally not well adapted to solve these inverse problems (Vallino, 2000).

Second, the convergence of the TRUST is only rigorously demonstrated in the case of the optimization of 1-D continuous functions by the canonical algorithm (Cétin *et al.,* 1993). As we tried to optimized a 2-D function by a simplified version of the original algorithm, the convergence is no longer guaranteed. We had to empirically tune all the algorithmic parameters, except a (Eq. 8). The number of spires, as well as the spire step, depend on the structure of the cost function. The values of b (Eq. 10) and ε are functions of the sampling accuracy, by a relation that was empirically defined.

Third, the TRUST theoretical principles proved complicated. It was developed very recently in comparison to the SA and the GA, and to our knowledge no code is directly available neither on the web nor in the literature. Applying the sampling of the cost function by a hyperspiral (Ammar and Cherruault, 1993) to estimate more than two parameters quickly becomes tricky. Moreover, the spiral geometry is not easily compatible with *a priori* information defined as intervals of realistic values: we resorted to several projections of the cost function, which might become complicated when optimizing numerous parameters. We also showed that using a discrete approximate function can lead the algorithm to miss a local minimum (Fig. 4). This is why we considered it was not worth testing the TRUST on the estimation of all the PSyDyn parameters. More generally, the TRUST can hardly be implement for the inversion of biogeochemical models, which are characterized by a high number of parameters.

Contrary to the TRUST, stochastic algorithms as the SA and the GA require only punctual evaluations of the cost function. Given their conceptual basis, their application to the optimization of numerous parameters does not bring any additional theoretical

difficulty at first sight. As a general rule, stochastic methods should be favored to optimize parameters of biogeochemical models.

In addition, the SA has been studied and developed for a sufficiently long time for codes to be directly available in the literature or on the web (Allen *et al.,* 1996). The optimization test on the 2-D cost section showed that the SA was able to very accurately estimate the GM, regardless of data errors (Table 3, Table 4). It proves that the SA manages very well in regions where the cost gradient is low, and suggests that the stopping criteria and their associated parameters would be appropriate. Given its high sampling resolution, we expect that it provides good evaluations of the acceptable intervals (Fig. 8). Finally, Figure 6 shows that the technique used to manage parameters generating numerical instabilities allows the algorithm not to linger in these regions of the search space, and to concentrate in areas where the model can be integrated.

However, the SA displays a significant number of drawbacks. First, it is computationally very expensive: it required ten times more cost evaluations than the TRUST and the GA to estimate the GM of the cost slice. The SA is purely stochastic, and no information is stored during the optimization. Only the decrease of the temperature (T) and of the random step size (σ) statistically allow it to reach the GM. This is why many evaluations were redundant. Hence, the high computational cost of the SA is above all explained by its theoretical basis. This may strongly restrict its application to the optimization of numerous parameters. Second, we tested several methods to impose the search space boundaries. None of them proved satisfying. The method we chose made the SA stagnate on the boundaries (Fig. 6), which led to further computational costs. Unfortunately, such inequality constraints frequently arise in biogeochemical optimization problems: most parameters are submitted to a positivity constraint. Third, stochastic algorithms do not, on principle offer any guarantee as to their convergence during a given run. We noticed that the SA was strongly problem-dependent, and that, to our knowledge, very few rules (even empirical ones) exist that would guide users in the choice of the procedural parameters. For instance, the convergence highly depends on the annealing schedule (Krüger, 1993), but the methods used to compute T_o are various, and provide very different results, some of which make the optimization fail. The stopping criteria are also various, and we had to chose them empirically. Although stochastic, the convergence slightly depends on the initial parameter set: the chance for it to be trapped by a local minima seems all the more important as it starts from the interior of the search domain. The fact that the SA got trapped by a local minimum when estimating the eight PSyDyn parameters brings another proof of its lack of robustness.

The canonical GA is based on quite simple principles. Even if the present growing interest for this kind of algorithms (globally qualified as Evolutionary Algorithms (EAs)) leads to further complexity, it is still possible to get on the web (Allen *et al.,* 1996; Carroll, 1996) or in the literature (Goldberg, 1989) simple and efficient versions. The GA displays three major advantages with respect to the SA. First, the essentials of the GA theory concentrate on the processing of schemata. A schema $S \in \{0, 1, *\}^l$ is a description of a hyperplane in a l-dimensional bit space (Man *et al.,* 1997). A parameter vector coded by a

binary string results from the juxtaposition of schemata. So the assessment of a parameter vector by the GA does not only bring punctual information on the cost function, but also allows us to simultaneously evaluate several hyperplanes. Evaluating a whole population provides the algorithm with statistical information about the cost geometry. The Schema Theorem is at the root of the "building block hypothesis." Building blocks designate short, low-order (that is, with a small number of fixed positions) and high-performance schemata. According to this hypothesis, the GA seeks optimal solutions through the juxtaposition of building blocks identified during the run. The access to statistical information on the cost function shape and the partial storage of this information explains the reduction of the computational cost from the SA to the GA (Table 3, Table 4). Second, the binary coding of the parameters uses the search space limits, defined from intervals of realistic values of each parameter. The GA is, therefore, directly adapted to the exploration of bounded search spaces. Inside the search domain, we noticed that parameter sets, inducing numerical instabilities, do not perturb the optimization.

Figure 5c shows that the convergence of the GA in the neighborhood of the GM (that is, from the 74th generation) is very slow. It was accelerated by restarting the algorithm (generation 201). However, the convergence permits to quickly locate the basin of acceptable parameters (Table 4). Although the GA performs far less cost computations than the SA, its estimation of the acceptable intervals is equivalent to the estimation by the SA (Fig. 8). Figure 9a-b also prove that the GA is able to very quickly locate the attraction basin of the solution in the 8-D search space. Then, a classical gradient descent provides the exact GM coordinates. As for all the stochastic algorithms, the global convergence of the GA is not guaranteed. The coding, selection, mutation and crossover operators are tuned according to empirical criteria. Nevertheless, it seemed to us, from reading the recommendations of Bäck and Schwefel (1993), Carroll (1996), Goldberg (1989) and Schoenauer and Michalewicz (1997), that some empirical but robust rules exist which form the object of intense research. The success of the GA on the estimation of the eight PSyDyn parameters is an argument in favor of its robustness.

Nevertheless, applying this version of the GA to the optimization of numerous parameters, for instance to invert complicated pelagic ecosystem models, may be limited by its computational cost. A partial solution may be provided by evaluating the individuals of a population in parallel (master-slave model; Schoenauer and Michalewicz, 1997). The GA is one of the four mainstream examples of Evolutionary Algorithms (EAs), which are all based on the simulation of some genetics mechanisms. Another one, still unknown in the ocean science field, is called Evolution Strategies (ESs). The ESs are designed to specifically deal with parameter optimization problems. Bäck and Schwefel (1993) present a comparison of the EAs on the optimization of three theoretical functions: the convergence of the ESs is always much faster than the convergence of the GAs.

To end this section, we mention some difficulties which did not arise during our simulated inversion experiments, but of which one should be aware. In the case of the optimization of the 2-D cost section, the surface of the acceptable values of K_{des} and K_{ag} is

connected, at least inside the limits of the search domain. Its form resembles an ellipse, which is what would be obtained in case of a linear inverse problem. Nevertheless, the problem being strongly nonlinear, the shape of the acceptable parameter basin might be much more complex (Mazzega, 2000). First, it might not be convex. In this case, the computation of acceptable intervals to evaluate the basin size is not sufficient anymore. Second, it might not be connected: several acceptable secondary minima might exist, amongst which the GOAs might be unable to distinguish only from the information the modeller put into the formulation of the cost function. In this case, it may be fruitful to study the prediction families produced by each local minimum and for each model variable, and to submit them to other selection criteria. For instance, one may analyze the dynamical characteristics and stability of the solutions, or check if the predictions verify some empirical biogeochemical algorithm.

6. Conclusion

The results of the simulated inversions of the PSyDyn model allow us to select an inverse method to estimate, from *in situ* data, the parameters controlling the oceanic particle cycling. We demonstrated that given both the strong nonlinearity of this inverse problem, which translates into the high complexity of the related cost functions, and our limited knowledge of the particulate system, classical gradient descent methods are not appropriate. Global Optimization Algorithms (GOAs) were shown to be much more adapted. The TRUST, the Simulated Annealing (SA) and the Genetic Algorithm (GA) are able to recover at least two of the model parameters. However, the gradient requirement of deterministic GOAs like the TRUST is a serious drawback to minimize very irregular cost functions. Moreover, estimating more than two parameters with the TRUST would be very expensive in terms of programming cost. Because of its strong stochastic nature, the SA has very high computational requirements. Its low robustness may explain its incapability to recover the eight model parameters. The GA has much lower programming requirements than the TRUST and much lower computational requirements than the SA. It proved robust and able to quickly converge toward the basin of the true model parameters. As a consequence, we will use this GA in a next paper to estimate particle cycling parameters from *in situ* data collected within the framework of the French JGOFS-EUMELI program.

Acknowledgments. We are grateful to Geoff Evans (Fisheries and Oceans, Canada) and to Marc Schoenauer (Equipe Evolution Artificielle et Appentissage de l'X–URA CNRS 756/URA CNRS 317, France) for fruitful discussions. Constructive comments were also provided by Jim Gunson (The UK Meteorological Office) and an anonymous reviewer. This research was made possible by support from the Institut National des Sciences de l'Univers (France-JGOFS/PROOF and DYTMOAT programs). The optimization runs were performed on the Cray SV1 of the CNES (Toulouse, France).

REFERENCES

Alldredge, A. L. and J. A. Jackson. 1995. Aggregation in marine systems. Deep-Sea Res. II, *42,* 1–7.

Allen, R. C. *et al.* 1996. Computational Science Education Project ebook, Mathematical Optimization chapter (http://csep1.phy.ornl.gov/csep.html). U.S. Department of Energy, 62 pp.

Ammar, H. and Y. Cherruault. 1993. Approximation of a several variables function by one variable function and application to global optimization. Math. Comput. Model., *18,* 17–21.

Athias, V., P. Mazzega and C. Jeandel. 2000. Nonlinear inversions of a model of the oceanic dissolved-particulate exchanges, *in* Inverse Methods in Global Biogeochemical Cycles, P. Kasibhatla *et al.,* eds., Geophysical Monograph. American Geophysical Union, 205–222.

Athias, V., P. Mazzega, D. Ruiz-Pino, R. Arraes and C. Jeandel. 1998. PSyDyn: a semi-spectral model of the dissolved-particulate exchanges. A preliminary stability analysis, *in* European Network for Integrated Marine Systems Analysis, J. Baeyens, F. Dehairs and L. Goeyens, eds., Brussels, 299–312.

Bäck, T. and H.-P. Schwefel. 1993. An overview of evolutionary algorithms for parameter optimization. Evol. Comp., *1,* 1–23.

Bacon, M. P. and R. F. Anderson. 1982. Distribution of thorium between dissolved and particulate forms in the deep sea. J. Geophys. Res., *87,* 2045–2056.

Bacon, M. P., C.-A. Huh, A. P. Fleer and W. E. Deuser. 1985. Seasonality in the flux of natural radionuclides and plutonium in the deep Sargasso Sea. Deep-Sea Res., *32,* 273–286.

Barhen, J. and V. Protopopescu. 1996. Generalized TRUST algorithm for global optimization, *in* State of the Art in Global Optimization: Computational Methods and Applications, C. A. Floudas and P. M. Pardalos, eds., Kluwer Academic Press, 163–180.

Barhen, J., V. Protopopescu and D. Reister. 1997. TRUST: a deterministic algorithm for global optimization. Science, *276,* 1097–1097.

Barth, N. and C. Wunsch. 1990. Oceanographic experiment design by simulated annealing. J. Phys. Oceanogr., *20,* 1249–1263.

Bennett, A. F. 1992. Inverse Methods in Physical Oceanography. Cambridge Monographs on Mechanics. Cambridge University Press, NY, 364 pp.

Bory, A. and P. Newton. 2000. Transport of airborne lithogenic material down through the water column in two contrasting regions of the eastern subtropical North Atlantic Ocean. Global Biogeochem. Cycles, *14,* 297–315.

Carroll, D. L. 1996. Chemical laser modeling with genetic algorithms. AAIA J., *34,* 338–346 (http://www.staff.uiuc.edu/~carroll/ga.html).

Cétin, B. C., J. Barhen and J. W. Burdick. 1993. Terminal Repeller Unconstrained Subenergy Tunneling (TRUST) for fast global optimization. J. Optim. Theory Appl., *77,* 97–126.

Clegg, S. L., M. P. Bacon and M. Whitfield. 1991. Application of a general scavenging model to thorium isotope and particle data at equatorial and high latitude sites in the Pacific Ocean. Deep-Sea Res., *96,* 20655–20670.

Clegg, S. L. and M. Whitfield. 1991. A generalized model for the scavenging of trace metals in the open ocean-II. Thorium scavenging. Deep-Sea Res., *38,* 91–120.

Cochran, J. K., K. O. Buesseler, M. P. Bacon and H. D. Livingston. 1993. Thorium isotopes as indicators of particle dynamics in the upper ocean: results of JGOFS North Atlantic Bloom Experiment. Deep-Sea Res. I, *40,* 1569–1595.

Craig, H. 1974. A scavenging model for trace elements in the deep sea. Earth Planet. Sci. Lett., *23,* 149–159.

Edwards, A. M. and J. Brindley. 1996. Oscillatory behaviour in a three-component plankton model. Dyn. Stab. Syst., *11,* 347–370.

Eknes, M. and G. Evensen. 1997. Parameter estimation solving a weak constraint variational formulation for an Ekman model. J. Geophys. Res., *102,* 12479–12491.

Evans, G. T. 1999. The role of local models and data sets in the Joint Global Ocean Flux Study. Deep-Sea Res. I, *46,* 1369–1389.

Evensen, G., D. P. Dee and J. Schröter. 1998. Parameter estimation in dynamical models, *in* Ocean Modeling and Parameterization, E. P. Chassignet and J. Verron, eds., NATO, ASI Series. Kluwer Acad Publ., 451 pp.

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

Fasham, M. J. R. and G. T. Evans. 1995. The use of optimization techniques to model marine ecosystem dynamics at the JGOFS station at 47°N20°W. Phil. Trans. R. Soc. Lond. B, *348,* 203–209.

Fletcher, R. and C. M. Reeves. 1964. Function minimization by conjugate gradients. Comput. J., *7,* 149–154.

Gelado-Caballero, M. D., M. E. Torres-Padron, J. J. Hernandez-Brito, J. A. Herrera-Melian and J. Pérez-Pena. 1996. Aluminium distribution in central East Atlantic waters (Canary Islands). Mar. Chem., *51,* 359–372.

Gilbert, J. C. and J. Nocedal. 1992. Global convergence properties of conjugate gradient methods. SIAM J. on Optimiz., *2,* 21–42 (http://www.ece.nwu.edu/~rwaltz/CG+.html).

Goldberg, D. E. 1989. Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley, USA, 412 pp.

Gunson, J., A. Oschlies and V. Garçon. 1999. Sensitivity of ecosystem parameters to simulated satellite ocean color data using a coupled physical-biological model of the North Atlantic. J. Mar. Res., *57,* 613–639.

Henderson, G. M., C. Heinze, R. F. Anderson and A. M. E. Winguth. 1999. Global distribution of the ^{232}Th flux to ocean sediments constrained by GCM modelling. Deep-Sea Res. I, *46,* 1861–1893.

Honeyman, B. D., L. S. Balistrieri and J. W. Murray. 1988. Oceanic trace metal scavenging: the importance of particle concentration. Deep-Sea Res., *35,* 227–246.

Honeyman, B. D. and P. H. Santschi. 1989. A brownian-pumping model for oceanic trace metal scavenging: Evidence for Th isotopes. J. Mar. Res., *47,* 951–992.

Jackson, G. A. and S. Lochmann. 1993. Modeling coagulation of algae in marine ecosystems, *in* Environmental Particles, J. Buffle and H. P. van Leeuwen, eds., Lewis Publishers, Boca Raton, FL, 387–414.

Jackson, G. A., B. E. Logan, A. L. Alldredge and H. G. Dam. 1995. Combining particle size spectra from a mesocosm experiment measured using photographic and aperture (Coulter and Elzone) techniques. Deep-Sea Res. II, *42,* 139–157.

Kirkpatrick, S., C. D. Gelatt and M. P. Vecchi. 1983. Optimization by simulated annealing. Science, *224,* 1340–1342.

Krüger, J. 1993. Simulated annealing: a tool for data assimilation into an almost steady model state. J. Phys. Oceanogr., *23,* 679–688.

Lal, D. 1977. The oceanic microcosm of particles. Science, *198,* 997–1008.

Lampitt, R. S. and A. N. Antia. 1997. Particle flux in the deep seas: regional characteristics and temporal variability. Deep-Sea Res. I, *44,* 1377–1403.

Man, K. F., K. S. Tang, S. Kwong and W. A. Halang. 1997. Genetic algorithms for control and signal processing. Springer-Verlag, London, 221 pp.

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.

Mazzega, P. 2000. On the assimilation and inversion of small data sets under chaotic regimes, *in* Inverse Methods in Global Biogeochemical Cycles, P. Kasibhatla *et al.,* eds., Geophysical Monograph. Amer. Geophys. Un., 223–253.

Menke, W. 1989. Geophysical Data Analysis: Discrete Inverse Theory, International Geophysical Series, *45,* Academic Press, 290 pp.

Murnane, R. J. 1994. Determination of the thorium and particulate matter cycling parameters at station P: a reanalysis and comparison of least squares techniques. J. Geophys. Res., *99,* 3393–3405.

Murnane, R. J., J. K. Cochran, K. O. Buesseler and M. P. Bacon. 1996. Least-squares estimates of thorium, particle, and nutrient cycling rate constants from the JGOFS North Atlantic Bloom Experiment. Deep-Sea Res., *43,* 239–258.

Murnane, R. J., J. K. Cochran and J. L. Sarmiento. 1994. Estimates of particle- and thorium-cycling rates in the northwest Atlantic Ocean. J. Geophys. Res., *99,* 3373–3392.

Newton, P. P., R. S. Lampitt, T. D. Jickells, P. King and C. Boutle. 1994. Temporal and spatial variability of biogenic particle fluxes during the JGOFS northeast Atlantic process studies at 47N, 20W. Deep-Sea Res. I, *41,* 1617–1642.

Nozaki, Y., Y. Horibe and H. Tsubota. 1981. The water column distribution of thorium isotopes in the western North Pacific. Earth Planet. Sci. Lett., *54,* 203–216.

Nozaki, Y., H.-S. Yang and M. Yamada. 1987. Scavenging of thorium in the ocean. J. Geophys. Res., *92,* 772–778.

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

Roy-Barman, M., J. H. Chen and G. J. Wasserburg. 1996. ^{230}Th-^{232}Th systematics in the central Pacific Ocean: The sources and the fates of thorium. Earth Planet. Sci. Lett., *139,* 351–363.

Roy-Barman, M., L. Coppola and M. Souhaut. 2000. Thorium isotopes in the western Mediterranean Sea: an insight into the marine particle dynamics. Deep-Sea Res., (submitted).

Ruiz-Pino, D. P. 1994. Modèle Colonne d'Eau au site DYFAMED, Modélisation des cycles biogéochimiques marins; applications aux opérations France/JGOFS, Paris.

Schoenauer, M. and Z. Michalewicz. 1997. Evolutionary Computation. Control and Cybernetics, *26,* 307–338.

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

Tachikawa, K., C. Jeandel and B. Dupré. 1997. Distribution of rare earth elements and neodymium isotopes in settling particulate material of the tropical Atlantic Ocean (EUMELI site). Deep-Sea Res. I, *44,* 1769–1792.

Tachikawa, K., C. Jeandel, A. Vangrisheim and B. Dupré. 1999. Distribution of rare earth elements and neodymium isotopes in suspended particles of the tropical Atlantic Ocean (EUMELI site). Deep-Sea Res. I, *46,* 733–755.

Tarantola, A. 1987. Inverse Problem Theory; Methods for Data Fitting and Model Parameter Estimation. Elsevier, 614 pp.

Vallino, J. J. 2000. Improving marine ecosystem models: use of data assimilation and mesocosm experiments. J. Mar. Res., *58,* 117–164.

Vézina, A. F. and T. Platt. 1988. Food web dynamics in the ocean. I. Best-estimates of flow networks using inverse methods. Mar. Ecol. Prog. Ser., *42,* 269–287.

Whitfield, M. and D. R. Turner. 1987. The Role of Particles in Regulating the Composition of Seawater, Aquatic Surface Chemistry. Stumm-Wiley, NY, 457–464.

Zakardjian, B. and L. Prieur. 1994. A numerical study of primary production related to vertical turbulent diffusion with special reference to vertical motions of the phytoplankton cells in nutrients and light fields. J. Mar. Syst., *5,* 267–295.

Received: *20 December, 1999;* revised: *14 July, 2000.*