



**HAL**  
open science

## Improving the parameters of a global ocean biogeochemical model via variational assimilation of in situ data at five time series stations

Abdou Kane, Cyril Moulin, Sylvie Thiria, Laurent Bopp, Mohamed Berrada, Alessandro Tagliabue, Michel Crépon, Olivier Aumont, Fouad Badran

► **To cite this version:**

Abdou Kane, Cyril Moulin, Sylvie Thiria, Laurent Bopp, Mohamed Berrada, et al.. Improving the parameters of a global ocean biogeochemical model via variational assimilation of in situ data at five time series stations. *Journal of Geophysical Research*, 2011, 116, pp.C06011. 10.1029/2009JC006005 . hal-00741205

**HAL Id: hal-00741205**

**<https://hal.science/hal-00741205>**

Submitted on 28 Oct 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Improving the parameters of a global ocean biogeochemical model via variational assimilation of in situ data at five time series stations

A. Kane,<sup>1,2</sup> C. Moulin,<sup>1</sup> S. Thiria,<sup>3</sup> L. Bopp,<sup>1</sup> M. Berrada,<sup>3</sup> A. Tagliabue,<sup>1</sup> M. Crépon,<sup>3</sup> O. Aumont,<sup>4</sup> and F. Badran<sup>5</sup>

Received 25 November 2009; revised 6 February 2011; accepted 23 March 2011; published 18 June 2011.

[1] The global ocean biogeochemical models that are used in order to assess the ocean role in the global carbon cycle and estimate the impact of the climate change on marine ecosystems are getting more and more sophisticated. They now often account for several phytoplankton functional types that play particular roles in marine food webs and the ocean carbon cycle. These phytoplankton functional types have specific physiological characteristics, which are usually poorly known and therefore add uncertainties to model results. Indeed, this evolution in model complexity is not accompanied by a similar increase in the number and diversity of in situ data sets necessary for model calibration and evaluation. Thus, it is of primary importance to develop new methods to improve model performance using existing biogeochemical data sets, despite their current limitations. In this paper, we have optimized 45 physiological parameters of the PISCES global model, using a variational optimal control method. In order to bypass a global 3-D ocean variational assimilation, which would require enormous computation and memory storage, we have simplified the estimation procedure by assimilating monthly climatological in situ observations at five contrasted oceanographic stations of the JGOFS program in a 1-D version of the PISCES model. We began by estimating the weight matrix in the cost function by using heuristic considerations. Then we used this matrix to estimate the 45 parameters of the 1-D version of the PISCES model by assimilating the different monthly profiles (observed profiles at the five stations) in the same variational procedure on a time window of 1 year. This set of optimized parameters was then used in the standard 3-D global PISCES version to perform a 500 year global simulation. The results of both the standard and the optimized versions of the model were compared to satellite-derived chlorophyll-a images, which are an independent and global data set, showing that our approach leads to significant improvements in simulated surface chlorophyll-a in most of the regions of the world ocean. Besides demonstrating that we have improved the accuracy of the PISCES model, this study proposes a sound methodology that could be used to efficiently account for in situ data in biogeochemical ocean models.

**Citation:** Kane, A., C. Moulin, S. Thiria, L. Bopp, M. Berrada, A. Tagliabue, M. Crépon, O. Aumont, and F. Badran (2011), Improving the parameters of a global ocean biogeochemical model via variational assimilation of in situ data at five time series stations, *J. Geophys. Res.*, *116*, C06011, doi:10.1029/2009JC006005.

### 1. Introduction

[2] Every year, diverse oceanic and terrestrial processes contribute to removing about half the anthropogenic carbon dioxide emissions from the atmosphere. In this context marine autotrophic productivity, which is estimated to be comparable to that of the land ecosystems, plays a crucial role

in converting dissolved inorganic carbon into organic forms. Marine phytoplankton use inorganic carbon during photosynthesis in surface waters and a fraction of the organic matter produced is “exported” through sinking processes (by fecal pellets for example) to the deep ocean and finally some fraction reaches the sediment. The efficiency of this export is

<sup>1</sup>Laboratoire des Sciences du Climat et de l'Environnement/IPSL, CEA-CNRS-UVSQ, Gif-sur-Yvette, France.

<sup>2</sup>ACRI-ST, Sophia Antipolis, France.

<sup>3</sup>Laboratoire d'Océanographie et du Climat: Expérimentation et Approches Numériques/IPSL, CNRS-IRD-UPMC, Paris, France.

<sup>4</sup>Laboratoire de Physique des Océans, IUEM, CNRS-IRD-UBO, Brest, France.

<sup>5</sup>Centre d'Etudes et de Recherche en Informatique du CNAM/CNAM, Paris, France.

partly controlled by the related phytoplankton species, which explains why most biogeochemical models now include several phytoplankton functional types (PFTs) [Le Quéré *et al.*, 2005]. Recently, an intercomparison between different methods and models [Carr *et al.*, 2006] estimated global marine productivity by phytoplankton to be approximately 35–70 PgC/yr. This uncertainty is even larger for carbon export, due to the specific efficiency of the various PFT and the biological processes at work. Reducing uncertainties on species role and biological processes in the ocean appears as an essential step for realistic climate change studies.

[3] Biological effects on carbon export are difficult to assess in climate models because they require the use of a sophisticated biogeochemical model coupled to an ocean general circulation model representing the ocean's physical processes. The most evolved global biogeochemical models currently used to represent the ocean carbon cycle now include several groups of phytoplankton, zooplankton and bacteria; they can also be used to assess the effects of climate and environmental changes on marine ecosystems. The main drawback of these models is that they rely on many physiological parameters (e.g., mortality rates, photosynthetic parameters, affinities for different nutrients etc.) that are difficult to measure. However, although the number of parameters is larger than in more simplistic models (such as NPZD-type models in which only one generic phytoplankton type is represented), they offer the advantage of representing more representative marine organisms and can theoretically be more suitable for modeling carbon export. Some parameters can be assessed directly by using observations in sophisticated laboratory studies, others indirectly through their impact on biogeochemical variables and thus by conducting field campaigns or processing observation and climatology databases.

[4] Very few pertinent observations of phytoplankton physiological processes are available due to the enormous diversity of marine organisms and the related processes for which observations are required. Even if modelers try to ensure that the physiological parameters employed are within a range of realistic values, such as those provided by observations, most parameters in ocean biogeochemical models are tuned in a somewhat arbitrary fashion, so as to best reproduce available observational data sets such as global chlorophyll maps from satellites or the climatology of nutrient concentrations. For instance, Fasham *et al.* [1990] showed that 200 individual model runs were needed for their model to get a set of parameters in agreement with the observations. In this context, data assimilation and the use of variational control theory is a more objective and quantitative way of improving model performance [Lions, 1968; Anderson and Moore, 1971, 2007].

[5] In variational control theory, the parameters or the initial conditions of the model, which govern the model global trajectory, are adjusted to obtain the best agreement between the mean trajectory of the model and the data [Le Dimet and Talagrand, 1986]. A cost function represents the misfits of the model to the observations, which is often taken as a quadratic distance (mean square error). An iterative procedure is then used to minimize this function through an algorithm of gradient, which can be computed by using the adjoint of the model. Oschlies and Schartau [2005] summarized a large number of studies performed with different assimilation algorithms for various ocean biogeochemical models. Schartau and Oschlies [2003], Friedrichs [2001] and

Faugeras *et al.* [2003, 2004] tried to calibrate the parameters of a 1-D biogeochemical model, using stations in the North Atlantic, the equatorial Pacific and the Mediterranean, respectively. In all these studies, the model errors are significantly reduced at the station locations but a major and common drawback is that the set of optimized parameters could not be used elsewhere with the same success.

[6] The difficulty in generalizing the results of 1-D assimilation experiments to 3-D global ocean is mostly due to the variability in biological conditions at the global scale. The optimal solution would be to optimize the parameters at the global scale in a 3-D model using all pertinent available data [Schlitzer, 2000; Tjiputra *et al.*, 2007], but this currently remains out of reach because of the computer resources requirement, especially for high-resolution models, for long assimilation temporal windows or for more complex models, such as the PISCES model [Aumont and Bopp, 2006].

[7] In this work, we have developed a variational approach to estimate a set of parameters for the biogeochemical model PISCES from in situ observations taken at oceanographic stations located in contrasted regions. We examined the suitability of the parameters obtained at global scale. To do so, we performed a long 3-D global simulation using the optimized parameters and compared them to the standard simulations to examine their performance using global maps of satellite-derived chlorophyll.

[8] The paper is organized as follows. The introduction is devoted to a presentation of the need to better estimate the parameters of the ocean biogeochemical models. Part 2 presents a brief overview of the PISCES biogeochemical model and the in situ data used to control the model parameters. Part 3 is dedicated to a description of the variational assimilation formalism developed with the YAO software for estimating the parameters of PISCES. In that part, we show the procedure we have applied to choose the parameters to be optimized, and also this to estimate the matrix coefficients weighing the different observations in the cost function of the variational method. In section 4, we applied the variational method to optimize 45 parameters of a PISCES 1-D multisite version (so-called pseudo-3-D) using five contrasted JGOFS stations. In section 6, we evaluated the performances of a 500 year 3-D global simulation using the parameters optimized at the five JGOFS stations, relative to similar 3-D simulation using the standard parameter set. We conclude with a discussion of the method and some perspectives about future work.

## 2. Model and Data

### 2.1. The PISCES Model

[9] PISCES is a 3-D global ocean biogeochemical model [Aumont and Bopp, 2006], which simulates marine primary production and also describes the cycles of carbon, oxygen and the main macro and micro nutrients (phosphate, nitrate, ammonium, iron and silicic acid). Two Phytoplankton species (nanophytoplankton and diatom) and two zooplankton species (mesozooplankton and microzooplankton), as well as their respective detrital compartments, are represented in PISCES. PISCES has been used for a wide variety of studies, concerned with paleoceanographic [Bopp *et al.*, 2003] and future climate questions [Schneider *et al.*, 2008], as well as process studies which examine, for example, the impact of

dust deposition variability on marine biogeochemistry [e.g., *Aumont et al.*, 2008; *Tagliabue et al.*, 2008]. Therefore PISCES is a good candidate model through which the improvement of parameters obtained from data assimilation can be examined.

[10] The processes that control the temporal evolution of the concentration of a given state variable are a function of the various sources and sinks affecting it, and are represented by a set of coupled classical differential equations. As an example, equation (1) shows the differential equation for one of the phytoplankton groups in PISCES:

$$\frac{dP}{dt} = (1 - \delta p)\mu_P P - m_P \frac{P}{K_P + P} - \omega_P P^2 - g_Z(P)Z - g_M(P)M \quad (1)$$

where  $P$  is the concentration of a given phytoplankton group (nanophytoplankton or diatoms),  $\delta$  is the dissolved organic carbon (DOC) exudation,  $\mu$  is the specific primary production rate which depends on the Photosynthetically Available Radiation (PAR) and on the amount of nutrients available (phosphate, ammonium, nitrate and iron for both species, as well as silicic acid for diatoms),  $m$  is the mortality rate,  $K$  is a half-saturation constant for mortality,  $\omega$  is the quadratic mortality,  $Z$  is the concentration of microzooplankton,  $M$  is the concentration of mesozooplankton and  $g(P)$  is the maximum grazing rate of a zooplankton group for the phytoplankton group  $P$ . Most of these parameters ( $\delta$ ,  $\mu$ ,  $m$ ,  $K$ ,  $\omega$ ,  $g$ ) are related to the physiology of the phytoplankton. The other state variables (zooplankton, nutrients, etc.) are modeled using similar source-sink equations.

[11] More than 60 fixed parameters are used within PISCES to simulate the temporal evolution of the 24 state variables. Although many of them can be measured or estimated from laboratory experiments, it is important to note that in such a complex model as is used for global studies, some of these parameters have been chosen because they permit realistic simulations of the main biogeochemical state variables, rather than for their particular physiological significance. Given the diversity of phytoplankton and zooplankton species in the global ocean, it is indeed difficult to define the main physiological parameters from laboratory studies. This justifies the optimization approach proposed in the present work, which aims at finding a set of parameters that provides the best global simulation with regards to existing validation data sets.

[12] As assimilation techniques require significant computing resources, we have employed a 1-D offline version of PISCES to optimize the physiological parameters at several stations. We have used two steps to achieve this. First we optimized the model parameters by assimilating the observations of each data profile separately. This permitted us to build an error matrix weighing the importance of the different data. We then optimized the model parameters by assimilating the whole data in the same run. In that assimilation run (defined as “pseudo-3-D”), we used the above weight matrix in the cost function. All the 1-D simulations or optimizations were performed with the 1-D offline version of PISCES. Second, we used the 3-D global version of PISCES to conduct global simulations with the optimized parameters (derived from the pseudo-3-D experiments), which permitted us to evaluate the global response of the model output (CHL) with respect to satellite data sets.

[13] The pseudo-3-D model does not have any advection nor any lateral diffusion; only vertical diffusion operates. However, in the 3-D model all these transport processes ( $u$ ,  $v$ ,  $w$ ) are present. The initial biogeochemical conditions were given by a 3-D online reference run of OPA/PISCES that has already been described and used in several studies [*Aumont and Bopp*, 2006].

## 2.2. JGOFS Data

[14] The Joint Global Ocean Flux Study (JGOFS) was an international and multidisciplinary program launched in the late 1980s. Its goal was to assess more accurately the processes that control carbon fluxes between the atmosphere, ocean surface and ocean interior, as well as their sensitivity to climate change, from regional to global and from seasonal to interannual scales. One of the major achievements of JGOFS was to maintain a network of marine stations for a decade or more. During that program, the monthly vertical profiles of several biogeochemical variables were measured at several stations around the global ocean, as described and defined by *Kleypas and Doney* [2001].

[15] Here we have used five contrasted stations which have been continuously operating for several years, except NABE, as shown in Table 1. Both BATS and HOT stations are located in oligotrophic regions of the Atlantic and Pacific subtropical gyres, whereas KERFIX, DYFAMED and NABE are located in more productive regions characterized by a strong seasonal cycle of biological productivity. In our optimization approach, which is based on the assimilation of in situ measurements, we decided to use the most classical biogeochemical variables only, i.e., CHL (chlorophyll-a concentration),  $\text{NO}_3$  (nitrates concentration), POC (Particulate Organic Carbon concentration) and SIL (silicates concentration). Even with such a limited number of biogeochemical variables, a strong disparity persists between the five stations, from BATS where the four biogeochemical variables have been measured monthly for 10 years to NABE where only three months of data are available. These monthly observations constitute our observation data set. In the present study, we have focused on modeling the seasonal biogeochemical activity. We have thus created a monthly climatology of the observed biogeochemical variables. For that we averaged each January (February... ..December) data profile for the  $n$  years of the observation period. The observational database consists of 15 monthly climatology profiles of the different observed state variables (4 monthly state variable profiles at BATS station, 2 monthly state variable profiles at HOT, 4 at DYFAMED, 2 at KERFIX and 3 at NABE; see Table 1). In the following, we denote  $i$  the index of these data profiles ( $i = 1 \dots 15$  from CHL at BATS station to POC at NABE station; see Table 1, last row). Each observed variable is defined at depth  $k$  ( $k = 1 \dots 18$ ) by  $Y_{i,k}$ . This value is a component of the monthly observation vector  $Y_{obs}^m$  ( $m = 1 \dots 12$ ).

## 3. Development of the Variational Assimilation

### 3.1. Principle: Using the YAO Software

[16] Data assimilation based on a variational approach aims at adjusting the parameters of a model to obtain the best agreement with observations. This is achieved by minimizing a cost function that represents the misfit between model output (the so-called direct model) and a set of available observations. An efficient way for minimizing this cost

**Table 1.** Characteristics of the Data Sets at the Five Stations Used in This Study<sup>a</sup>

Station	BATS	HOT	DYFAMED	KERFIX	NABE
Location	Bermuda (Atlantic)	Hawaii (Pacific)	Mediterranean	Austral	North Atlantic
Coordinates	31°40'N–64°10'W	22°45'N–158°W	43°25'N–07°52'E	50°40'S–62°25'E	47°N–20°W
Number of years available	10, 1988–1999	10, 1988–1998	16, 1991–2007	5, 1990–1995	0.25, 1989
Number of climatological months <sup>b</sup>	12/12/12/12	12/0/0/12	12/12/12/12	12/12/0/0	3/3/3/0
Integer index of the variables at the stations	1, 2, 3, 4	5, -, -, 6	7, 8, 9, 10	11, 12, -, -	13, 14, 15, -
$R_i^{-1}(\text{e-4})$	20, 1, 3, 2	50, -, -, 1	8, 0.5, 5, 2	4, 0.8, -, -	80, 2, 10, -

<sup>a</sup>The numbers in the fifth row show how many climatologic profiles were available for each variable at a given station. The sixth row show the integer indices (i) for each variable at a station (i = 1...15). The last row shows the values of  $R_i^{-1}$ .

<sup>b</sup>Format is CHL/NO<sub>3</sub>/POC/SIL.

function is to use a gradient descent method. The gradients of the cost function, with respect to the parameters we seek to optimize, indicate the direction of search toward the minimum of the cost function where the derivatives should be equal to zero. Beyond the presence of local minimum for nonlinear models, the estimation of the gradients is not a trivial problem, especially with ecosystem models that are often very complex and have many parameters. The adjoint method offers a rigorous and fundamental framework within which we can systematically estimate the gradients of any model using some prior formalism [Talagrand and Courtier, 1987].

[17] We used the YAO software [Thiria et al., 2006] to generate the adjoint of PISCES since YAO is designed to facilitate the construction of an adjoint model. YAO relies on the theory of graphs [Berge, 1958] and has the advantage of automatically generating the adjoint of any direct model that has been coded within it. To do so, YAO uses a methodology that is based on the general formalism of a complex system decomposition into modular graphs [Badran et al., 2008], which is a recurrent procedure in graph theory. YAO then permits to perform data assimilation experiments and is a good platform for their development. This modular approach provides an object structure to the code, allowing a quasi-automatic generation of the adjoint to compute the gradient of the cost function.

[18] An interface with a Quasi Newton optimizer [Gilbert and Lemaréchal, 1989] is used to minimize the cost function. Two terms are needed to minimize this cost function, (1) its value and (2) its partial derivatives, which are both given by YAO using a forward and a backward integration of the model and its adjoint, respectively. YAO also provides a full suite of tools (e.g., cost functions, linear tangent, adjoint test, objective function test etc.) to allow the user to easily perform assimilation experiments.

### 3.2. General Variational Assimilation Algorithm Description

[19] In order to avoid performing a global ocean variational assimilation, which would lead to enormous computation and memory storage, we simplified the estimation procedure by assimilating monthly climatological in situ observations at five contrasted oceanographic stations of the JGOFS program in the 1-D version of the PISCES model. We assimilated the five station observations simultaneously using a temporal window of twelve months and appropriate weights dedicated to each observed state variable profile in the cost function.

[20] The cost function quantifies the misfit between the model output and the data, and has the following general form:

$$J(\mathbf{Xo}) = (\mathbf{Xo} - \mathbf{Xb})^T \mathbf{B}^{-1} (\mathbf{Xo} - \mathbf{Xb}) + \sum_{m=1}^M [\mathbf{Yobs}^m - \mathbf{HM}^m(\mathbf{Xo})]^T \cdot \mathbf{R}^{-1} [\mathbf{Yobs}^m - \mathbf{HM}^m(\mathbf{Xo})] \quad (2)$$

where the first term of (2) is the background term ( $\mathbf{Jb}$ ) and the second the misfit relative to the observations ( $\mathbf{Jobs}$ ).  $\mathbf{Xo}$  is the vector of the parameters of the PISCES model which have to be estimated and  $\mathbf{Xb}$  is our best a priori knowledge of the solution (prior).

[21] The background terms prevent the solution from taking unrealistic values far from the commonly considered values from previous studies.  $\mathbf{Yobs}^m$  represents the observation state vector and  $\mathbf{M}^m(\mathbf{Xo})$  the model output state at month  $m$ ,  $\mathbf{H}$  is the observation operator projecting the model output in the observation space. The role of  $\mathbf{B}^{-1}$  and  $\mathbf{R}^{-1}$  matrices is to weight the background and observation components in the cost function (2), respectively. This formulation of the cost function implies that the background and observation errors are normally distributed. In the assimilation theory,  $\mathbf{B}$  and  $\mathbf{R}$  are the error covariance matrices of the background and of the observations, respectively. We have developed an empirical determination of these matrices, which can be viewed as an original contribution to the present research.

[22] In Figure 1, phase 1 and 2 consist of the initialization of the controlled parameters and the prior vector  $\mathbf{Xb}$ ; and the computation of  $\mathbf{Jb}$ , which is the first term of the global cost function  $\mathbf{J}$ . A forward computation of the direct model is used in phase 3 to generate model outputs at the points where there are observations. Comparison of these outputs with the observations allows us to calculate  $\mathbf{Jobs}$ , the second term of  $\mathbf{J}$ . In phase 4, a computation of the derivative of  $\mathbf{J}$ , with respect to  $\mathbf{Xo}$ , is performed. The model outputs are used to initialize the gradients used in the backward computation. Phase 5 is the retropropagation of these gradients, computed from the adjoint from the final state to the initial state. It determines the partial derivatives of  $\mathbf{J}$  with respect to each component of the control vector  $\mathbf{Xo}$ .

[23] The gradient of the cost function is

$$\nabla J_{\mathbf{Xo}} = \mathbf{B}^{-1} (\mathbf{Xo} - \mathbf{Xb}) + \sum_{m=1}^M (\mathbf{M}^m)^T (\mathbf{Xo}) \mathbf{H}^T \mathbf{R}^{-1} \cdot [\mathbf{Yobs}^m - \mathbf{HM}^m(\mathbf{Xo})] \quad (3)$$

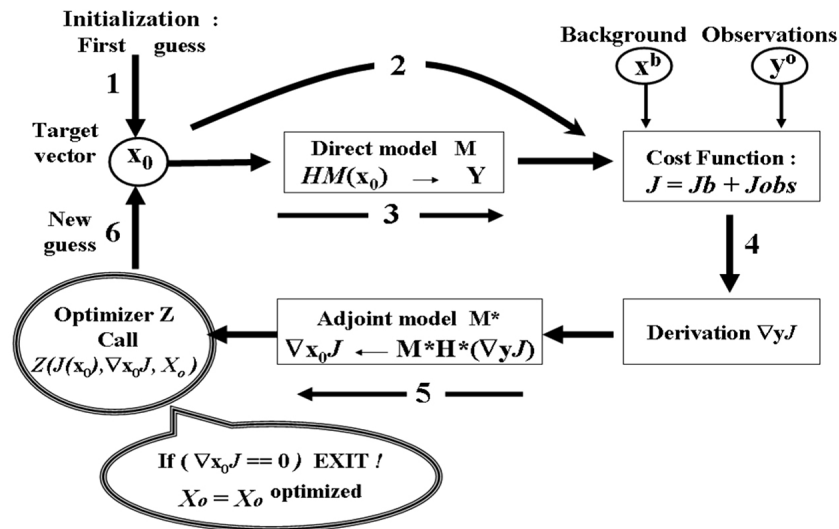


Figure 1. Schematic representation of the optimization algorithm.

where  $M^m$  denotes the derivative of  $M^m$  with respect to  $X_0$  (the so-called linear tangent) and  $(M^m)^T$  the adjoint of  $M$ .

[24] Phase 6 is devoted to the minimization and the update of the parameters. The global cost function  $J$  and its derivatives with respect to the control parameters  $\partial J / \partial X_0$  are given as input to the M1QN2 optimizer [Gilbert and Lemaréchal, 1989]. The control parameters are then updated in an iterative procedure until the convergence to a minimum is reached. After this brief description of the method we are going to show how to adapt this formalism to our particular problem.

### 3.3. Determination of the Size of the Control Vector, Its First Guess, and the Background Constraints

[25] The construction of an adequate cost function is a necessary step for improving the quality of the optimization and preventing a physically irrelevant local minimum. To that end, we have to use all available pertinent a priori information to better constrain the minima search. Additionally, a recurrent question that often arises in an inversion problem, is the dimension of the parameter control vector. If this dimension is too small, the system does not have enough degrees of freedom nor the necessary flexibility to fit the dynamics of the observations. On the other hand, if it is too large and consequently poorly constrained, the model output fits the observations well, but also the noise and the errors that are not linked to the parameterization. Then the model output may easily become erratic in places where no observation is available. In order to overcome these difficulties, we performed several assimilation sensitivity tests using data provided by twin experiments. In these sensitivity tests, the observations  $Y_{obs}$  consist of model output instead of actual observations. These experiments allowed us to understand how the assimilation works in the PISCES model, without having to face the complexity of actual measurements (e.g., noise, missing data). These twin experiments showed that the efficiency of our assimilation system was highly dependent on the quality of the first guesses. As our priori first guesses, we chose the standard 3-D PISCES parameters which proved to perform well in their representation of global ocean biogeochemical properties [Aumont and Bopp, 2006].

[26] Second to determine the size of the control vector  $X_0$  we investigated the respective roles of the 60 parameters of PISCES in the variational procedure for each observation. We estimated all the parameters of the PISCES model by assimilating each observed biogeochemical profile separately (15 profiles; see Table 1, last row for their definition), the first guess of the parameters being the standard values of the 3-D PISCES model. We then obtained 15 optimized sets of the 60 parameters. We chose to only retain the parameters that led to an absolute variation greater than 1% in at least one of the 15 experiments. As a result of this procedure, only 45 parameters were retained for the optimization run, while the other 15 were set to the values of the standard 3-D PISCES model (Table 2) because of their relative lack of sensitivity to our set of observations.

[27] Third, another specificity of our problem is that although we have a good knowledge of the range for which each parameter leads to a realistic simulation, the distribution of their a priori error is unknown. Therefore, we have decided not to include an explicit background on the parameter values in the cost function, but to arbitrarily constrain the parameter fluctuations within a  $\pm 100\%$  range, which is a range in which we do not have unrealistic values in the simulations of our five stations. The optimization of the control vector  $X_0$  (parameters to optimize) has been constrained to range between  $[X_{min}, X_{max}]$ . The optimum value of the control vector ( $X^{opt}$ ) we are searching for, are then defined by minimizing the cost function  $J$ :

$$X^{opt} = \operatorname{argmin}_{X_0 \in [X_{min}, X_{max}]} \left( J(X_0) \right) \\ = \sum_{m=1}^M \left( [Y_{obs}^m - HM^m(X_0)]^T R^{-1} [Y_{obs}^m - HM^m(X_0)] \right) \quad (4)$$

### 3.4. Empirical Estimation of the Observation Weight Matrix

[28] The computation of the observation weight matrix, usually denoted  $R^{-1}$ , is a major issue in the design of our assimilation method since it provides the weights associated

**Table 2.** Standard Values of the 45 Model Parameters Taken as the Components of the Controlled Vector (See Text) and the Percentage of Change After Optimization

Parameters	PISCES	Assimilation Increments	Parameter Definition
Conc0	2.e-6	0.00	phosphate half saturation (umolP $\Gamma^{-1}$ )
Conc1	10E-6	3.32	phosphate half saturation for diatoms (umo P $\Gamma^{-1}$ )
Conc3	0.1E-9	-3.65	iron half saturation for diatoms (nmo Fe $\Gamma^{-1}$ )
Grosip	0.151	79.14	mean Si/C ratio (?)
Pislope	4	-76.92	P-I slope for nano ( $(Wm^{-2})^{-1} d^{-1}$ )
Pislope2	4	-16.90	P-I slope for diatoms ( $(Wm^{-2})^{-1} d^{-1}$ )
Excret	0.05	-5.79	excretion ratio of nano-
Excret2	0.05	-6.32	excretion ratio of diatoms-
Wsbio	3	27.75	POC sinking speed ( $m d^{-1}$ )
Wsbio2	50	-63.85	big particles sinking speed ( $m d^{-1}$ )
Wchl	0.001	0.00	maximum aggregation rate for nano ( $d^{-1} molC^{-1}$ )
Wchld	0.02	26.72	Maximum aggregation rate for diatoms ( $d^{-1} molC^{-1}$ )
Resrat	0.03	69.30	exsudation rate of zooplankton-
Resrat2	0.008	22.82	exsudation rate of mesozooplankton-
Mprat	0.01	-6.82	Mprat phytoplankton mortality rate ( $d^{-1}$ )
Mprat2	0.01	-36.62	diatoms mortality rate ( $d^{-1}$ )
Grazrat	4	53.48	grazrat maximal zoo grazing rate ( $d^{-1}$ )
Grazrat2	0.7	19.27	maximal mesozoo grazing rate ( $d^{-1}$ )
Mzrat2	0.03	2.86	mesozooplankton mortality rate ( $umolC \Gamma^{-1} d^{-1}$ )
Xprefc	1	20.40	zoo preference for phyto-
Xprefp	0.	-32.79	zoo preference for POC-
Xprefz	1	-45.19	zoo preference for zoo-
Unass	0.3	-8.26	nonassimilated fraction of phyto by zoo (?)
Unass2	0.3	8.19	nonassimilated fraction of P by mesozoo (?)
Xkgraz	20.E-6	-9.55	half saturation constant for grazing ( $umolC \Gamma^{-1}$ )
Xkgraz2	20.E-6	59.72	half saturation constant for grazing 2 ( $umolC \Gamma^{-1}$ )
Xkmort	1.E-7	14.89	half saturation constant for mortality ( $umolC \Gamma^{-1}$ )
Xksil	2.E-6	-24.24	half saturation constant for Si uptake ( $umolSi \Gamma^{-1}$ )
Xksic	3.33E-6	0.00	half saturation constant for Si/C ( $umolSi \Gamma^{-1}$ )
Xremip	0.025	-46.56	remineralization rate of POC ( $d^{-1}$ )
Xremik	0.3; 0.25	-0.00011	remineralization rate of DOC ( $d^{-1}$ )
Xsirem	0.015	-1.08	remineralization rate of Si ( $d^{-1}$ )
Xkdoc2	417.E-6	54.84	second half-sat. of DOC remineralization ( $umolC \Gamma^{-1}$ )
Xprefpoc	0.2	-15.61	zoo preference for POC-
Concnh4	1.E-7	23.57	NH4 half saturation for phyto ( $umol P \Gamma^{-1}$ )
Concdnh4	5.E-7	-0.57	NH4 half saturation for diatoms ( $umol P \Gamma^{-1}$ )
Nitrif	0.05	-3.52	NH4 nitrification rate ( $d^{-1}$ )
Epshe2	0.33	52.36	efficiency of mesozoo growth-
Epshe	0.33	93.14	efficiency of microzoo growth-
Sigma1	0.6	-2.44	fraction of microzoo excretion as DOM-
Sigma2	0.6	-0.007	fraction of mesozoo excretion as DOM-
Zprefp	0.6	-28.58	microzoo preference for nanophyto-
Zprefd	0.5	26.68	microzoo preference for diatoms-
Chlcnm	0.033	50.29	minimum Chl/C in nanophytoplankton ( $mgChl/mgC$ )
Chlcdm	0.05	46.01	minimum Chl/C in diatoms ( $mgChl/mgC$ )

with each specific set of observations in the optimization process. Since  $\mathbf{R}^{-1}$  directly controls the efficiency and accuracy of the estimation, it is necessary to find a parameterization that best represents this effect. A simple way of defining  $\mathbf{R}^{-1}$  would be to use an identity matrix, implying that the confidence for each type of observation are the same in the assimilation process, which is far from being the case in nature and indeed in the model. A more elaborated representation of the matrix  $\mathbf{R}$  is therefore necessary: several simplified formulations have been proposed using heuristic considerations [Weaver and Courtier, 2001; Weaver et al., 2005]. In the present study, owing to the difficulty in estimating the matrix  $\mathbf{R}$  due to the lack of observations, we have decided to simply model its inverse  $\mathbf{R}^{-1}$  as a diagonal weight matrix, whose determination is described below.

[29] 1. We first calculated the annual mean quadratic errors of the profiles of the observed state variables of the 1-D PISCES model as the difference between the observed and computed variables. We did this for each profile of the

observed state variable  $i$ . The state variables were computed using the standard parameters of the model as defined in section 2. We then averaged the quadratic error computed at different depths  $k$  and at different months  $m$  for each variable  $i$ . We obtained a 15 component vector denoted  $\mathbf{Ref}$ , whose components are of the form

$$Ref_i = \frac{1}{K_i M_i} \sum_{k,m} [Y_i, k, m - \mathbf{HM}i, k, m]^2 \quad (5)$$

where  $\mathbf{HM}i, k, m$  represents the variable  $i$  computed by the 1-D PISCES model, at depth  $k$  and at month  $m$ .  $K_i$  is the number of vertical levels, and  $M_i$  is the number of months ( $K_i$  and  $M_i$  depend on  $i$ ). The component  $Ref_i$  represents the errors of the 1-D PISCES model for the observed state variable  $i$  without any assimilation.

[30] 2. We next estimated the 45 parameters of the model 15 times by performing 15 different assimilation experi-

ments with the 1-D PISCES model. Each assimilation experiment consists of assimilating a monthly profile of the observed variable  $i$  among the 15 monthly profiles in the 1-D PISCES model, the time window being a year. In the assimilation procedure, we used the climatological forcing at each station (temperature, salinity, wind, PAR, mixed layer depth). We thus obtained 15 optimized sets of the 45 parameters of the PISCES model, each set corresponding to one of the 15 observed profiles, which were assimilated independently. We then computed the residual quadratic errors for each assimilation run and for each state variable; we obtained a  $15 \times 15$  square matrix  $\zeta$  whose elements are of the form

$$\zeta_{i,l} = \frac{1}{K_i M_i} \sum_{k,m}^{K_i M_i} [Y_{i,k,m} - (HMI)^{m,i,k}]^2 \quad (6)$$

where  $(HMI)^{m,i,k}$  represents the state variable  $i$  computed by the PISCES model, at depth  $k$  and at month  $m$  with the parameters  $\mathbf{X}_o$  of PISCES obtained by assimilating the profile of the observed variable  $l$  using a 12 month time window. The term  $\zeta_{i,l}$  is an element of a square matrix whose lines are indexed by  $i$  and column by  $l$ ,  $i$  and  $l$  denoting integers such as  $(i, l = 1 \dots 15)$ .  $\zeta_{i,l}$  represents the annual depth averaged quadratic errors of the state variable profile  $i$  (as an example,  $i = 2$  designs  $\text{NO}_3$  at BATS), in the case where the state variable profile  $l$  has been assimilated (as an example,  $l = 10$  designs SIL at DYFAMED). The diagonal terms of this matrix are a measure of the residual quadratic assimilation error with respect to the state variable  $i$  (error between the observed and computed state variable  $i$  after assimilating in the model the state variable profile  $i$  itself). These terms were used to statistically normalize each biogeochemical variable in the cost function.

[31] 3. Finally, we subtracted and normalized each column of  $\zeta$  by the components of the vector  $\mathbf{Ref}$ , element per element. This new square matrix is denoted  $\mathbf{I}$  and it provides an index characterizing the efficiency of the assimilation. This index tells us if the assimilation of a chosen biogeochemical variable  $l$  improves or deteriorates the performances of the model with respect to the standard model for the variables  $i$  (line of the column  $l$ ) computed by the model.

$$\Gamma_{i,l} = \frac{\zeta_{i,l} - \text{Ref}_i}{\text{Ref}_i} \quad (7)$$

We then computed the mean value of the elements of each column of this matrix without the diagonal elements. We obtained a 15-component vector  $\mathbf{Det}$ , whose components are

$$\text{Det}_l = \frac{1}{14} \sum_{i(i \neq l)}^{15} \Gamma_{i,l} \quad (8)$$

$\text{Det}_l$  represents the normalized square bias between the mean of the modeled observation profiles computed with the optimized model  $\mathbf{M}$  and the mean profiles computed with the standard model  $\mathbf{M}$ . We denoted this term as the mean relative deterioration factor of the assimilation of the observed state variables  $l$  with respect to the standard model. The diagonal terms of the matrix  $\mathbf{I}$  were excluded from the computation of  $\mathbf{Det}$ , which just takes into account the cross

impacts of the assimilation of a biogeochemical variable with the other state variables. Each mean relative deterioration factor  $\text{Det}_l$  varies from  $-1$  in the best assimilation case ( $\sum \zeta_{i,l} = 0$ ) to a large positive number in the worst case, corresponding to the case in which the assimilation of the profile of the biogeochemical state variable  $l$  has a damaging impact on the performances of the modeling of the other state variables. We then defined a weight factor of damage  $\text{Fact}_l$  equal to

$$\text{Fact}_l = \frac{1}{1 + \text{Det}_l} \quad (9)$$

$\text{Fact}_l$  is a component of a 15 component vector  $\mathbf{Fact}$ ; a component  $\text{Fact}_l$  is a nondimensional quantity that tends toward 0 in the worst assimilation case and plus infinity in the ideal case. We used  $\text{Fact}_l$  to parameterize the diagonal of the observation weight matrix  $\mathbf{R}^{-1}$  in the cost function (equation (2)). In order to have a quantity whose dimension is the inverse of the variance, we divided  $\text{Fact}_l$  by  $\zeta_{l,l}$ , which represents the residuals errors on the state variable  $l$  computed by the model and resulting from the assimilation of the state variable  $l$  itself into the model. The quantity  $(1/\zeta_{l,l})$  can be seen as a normalization coefficient and also as an index of the flexibility of the model to fit the state variable  $l$ . Finally we calculated an approximation of the weight matrix  $\mathbf{R}^{-1}$ , which favors the data profile  $l$  whose assimilation improves the other state variables computed by the model (state variables  $i \neq l$ ) with respect to the observed state variables. The diagonal of the weight matrix  $\mathbf{R}^{-1}$  is then defined for each state variable  $l$  as the ratio

$$R_l^{-1} \approx \frac{\text{Fact}_l}{\zeta_{l,l}} \quad (10)$$

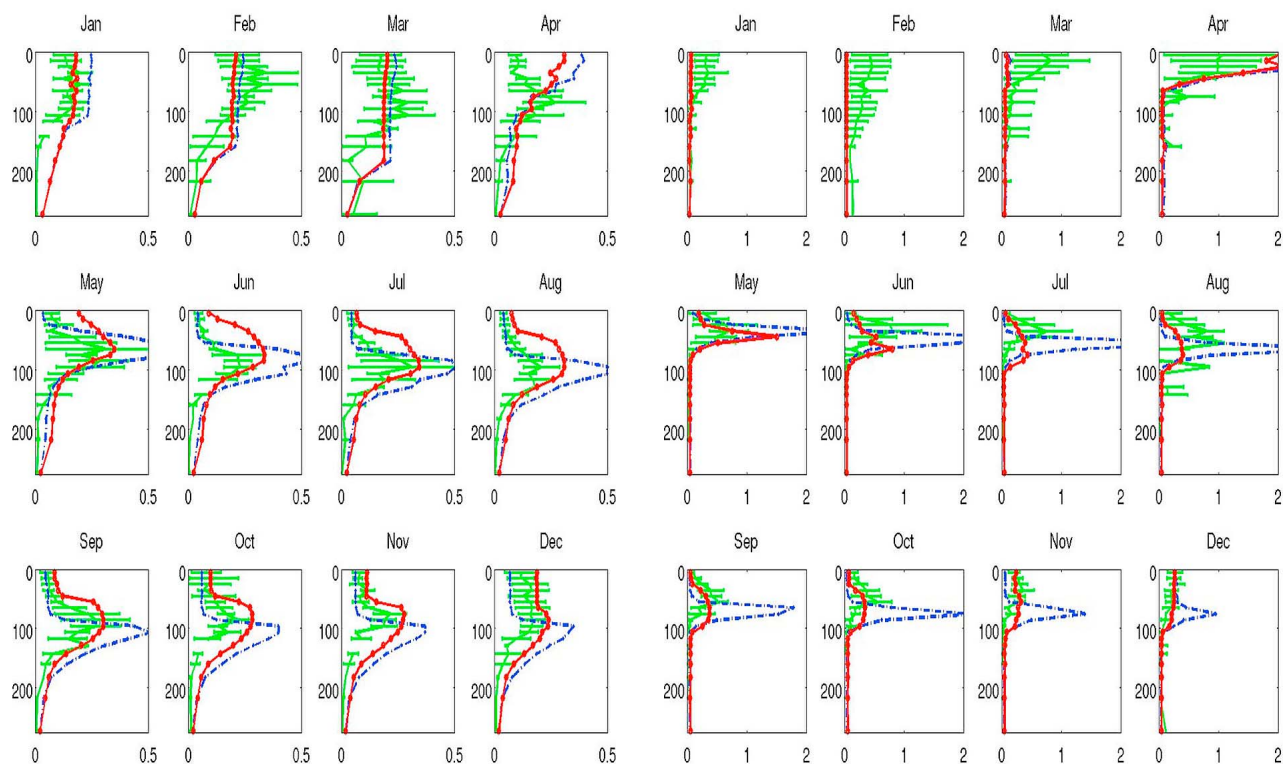
(Table 1) represents the weight given to the observed state variable  $l$ , whatever the depth and the month. This method is designed in such a way that it uncorrelates the biogeochemical observations from the errors arising from other sources; for example, if the physics at a given station where the state  $l$  variable was observed, is poorly modeled or if the external forcing at that station is unrealistic, the PISCES parameters obtained by the assimilation of the state variable  $l$ , may lead to a bad generalization and should be penalized. The matrix  $\mathbf{R}^{-1}$  allows us to weigh each variable  $l$  in the cost function according to its statistical confidence and consistency with the model when the 15 biogeochemical variables are assimilated simultaneously in the 1-D multisite optimization procedure, the so-called pseudo-3-D, which is discussed in section 4.

#### 4. Optimization of the Multisite PISCES Model Parameters

[32] The variational assimilation method described in section 3 was then used to optimize 45 parameters of the PISCES model. The cost function used in this assimilation experiment was

$$\text{Jobs} = \sum_{i=1}^{15} \sum_{m=1}^{M_i} \sum_{k=1}^{K_i} [Y_{i,k,m} - \mathbf{HMI}_{i,k,m}]^T R_i^{-1} [Y_{i,k,m} - \mathbf{HMI}_{i,k,m}] \quad (11)$$





**Figure 2.** Monthly mean climatological vertical profiles of the chlorophyll concentration ( $\mu\text{g/l}$ ) at (left) BATS and (right) DYFAMED. Data are in green, and standard and optimized PISCES 1-D simulations are in blue and red, respectively. Horizontal bars of the in situ profiles show the standard deviation computed on all available years (see Table 1).

The observation vector consisted of the 15 monthly data profiles described above taken as a whole data set. The assimilation time window was 12 months. The cost function (equation (11)) was minimized using the algorithm shown in Figure 1 by adjusting 45 parameters of the PISCES model. The list of these parameters with their changes in value arising from the assimilation, are summarized in Table 2. A brief inspection of the optimization of the parameters deserves some remarks. The objective here was not to provide a full and detailed discussion on the optimized values of all parameters, but rather to explain some of the most significant variations. With the standard set of parameters, the biogeochemical model was simulating too strong and too deep chlorophyll maximum (DCM) in the oligotrophic regions. Figure 2 illustrates this problem at BATS and DYFAMED. After the assimilation, the DCM is in much better agreement with the observations at the two stations. This improvement is due a decrease by almost 80% of the slope of the photosynthesis-irradiance (P-I) curve (pislope) of nanophytoplankton, which is the dominant group in the DCM, reducing the growth rate of this group at low light levels. In comparison, the slope of diatoms P-I curve was hardly modified since diatoms do not contribute much to the phytoplankton biomass in the lower part of the euphotic zone. Furthermore, the microzooplankton grazing rate (grazrat) has been increased by 50% to further decrease the nanophytoplankton biomass in the DCM.

[33] Another defect of the standard model is the underestimated chlorophyll concentrations at the sea surface in oligotrophic regions [Aumont and Bopp, 2006]. With the optimized set of parameters, the agreement with surface chlorophyll data has been significantly improved. The main explanation for this improvement is the increase in the maximum chlorophyll-to-carbon ratio of nanophytoplankton which is 50% greater. The equivalent ratio for diatoms is also significantly increased, but it plays a marginal role, as diatoms represent less than 10% of the biomass in the oligotrophic gyres. Other parameters significantly modified after the assimilation include the sinking speed of big particles (wsbio2) or the exudation rate of microzooplankton (resrat). However, the impact of these changes on the simulated fields cannot be easily inferred, since these changes depend on a multitude of complex nonlinear interactions and feedbacks. This would require a detailed analysis of the model behavior and results. Such analysis is beyond the objectives of the present study.

[34] Figure 2 illustrates the results of the assimilation at two contrasted stations, BATS, which is located in an oligotrophic region, and DYFAMED, which has a well marked seasonal cycle characterized by a spring bloom. Vertical profiles of CHL at BATS show low surface CHL values and a quasi-permanent DCM, which is a characteristic of oligotrophic waters [Morel and Berthon, 1989]. These features are poorly represented by the standard PISCES configuration in terms of

**Table 3.** Comparison of the Root Mean Square Error (RMSE, %) Between in Situ and PISCES 1-D Vertical Profiles of the Four Tracers and the Five Stations, Computed on All Available Monthly Climatological Vertical Profiles (From the Surface to 150 m)<sup>a</sup>

	CHL	NO3	POC	SIL
BATS	13/9	<b>140/156</b>	167/89	29/33
HOT	9/6			<b>98/102</b>
DYFAMED	60/30	245/232	523/388	611//595
KERFIX	60/21	1239/896		
NABE	<b>101/105</b>	266/191	<b>512/668</b>	

<sup>a</sup>In each cell, the left figure is for the standard configuration and the right one is for the optimized configuration. Figures in bold display the lowest RMSE.

both depth and intensity, whereas the use of the optimized parameters corrects most of these discrepancies and leads to a better agreement with data for all months. This improvement is also observed at the DYFAMED station, with a much better representation of the spring bloom and of the autumnal decrease in phytoplankton biomass. The optimization was not successful in improving the too low Chl values modeled during winter, because the vertical mixing in the model was too intense at that station, which results in intense light limitation of phytoplankton growth.

[35] In order to summarize the results at the five stations, the Root Mean Squared Error between the simulated and observed vertical profiles were computed (Table 3). Table 3 shows that a good minimization has been achieved globally. Overall, the optimization procedure results in a significant improvement for Chl (Table 3), except at NABE where the standard model gives the smallest RMSE. This is probably due to the fact that only three months of data were available at that station, which gave these data a smaller weight during the optimization, compared to the weights given to the data used at other stations. The improvement in model performance was also very significant for nitrate, as well as POC, at least at BATS and DYFAMED. At NABE, parameters are poorly constrained because of the limited period covered by measurements. The small change in values for silicate in Table 3 arises from the fact that silicate observations did not have a significant weight in the cost function.

## 5. Global 3-D Simulation With the PISCES Parameters Estimated With the Pseudo-3-D Assimilation Method

### 5.1. Vertical Profiles at the Five JGOFS Stations

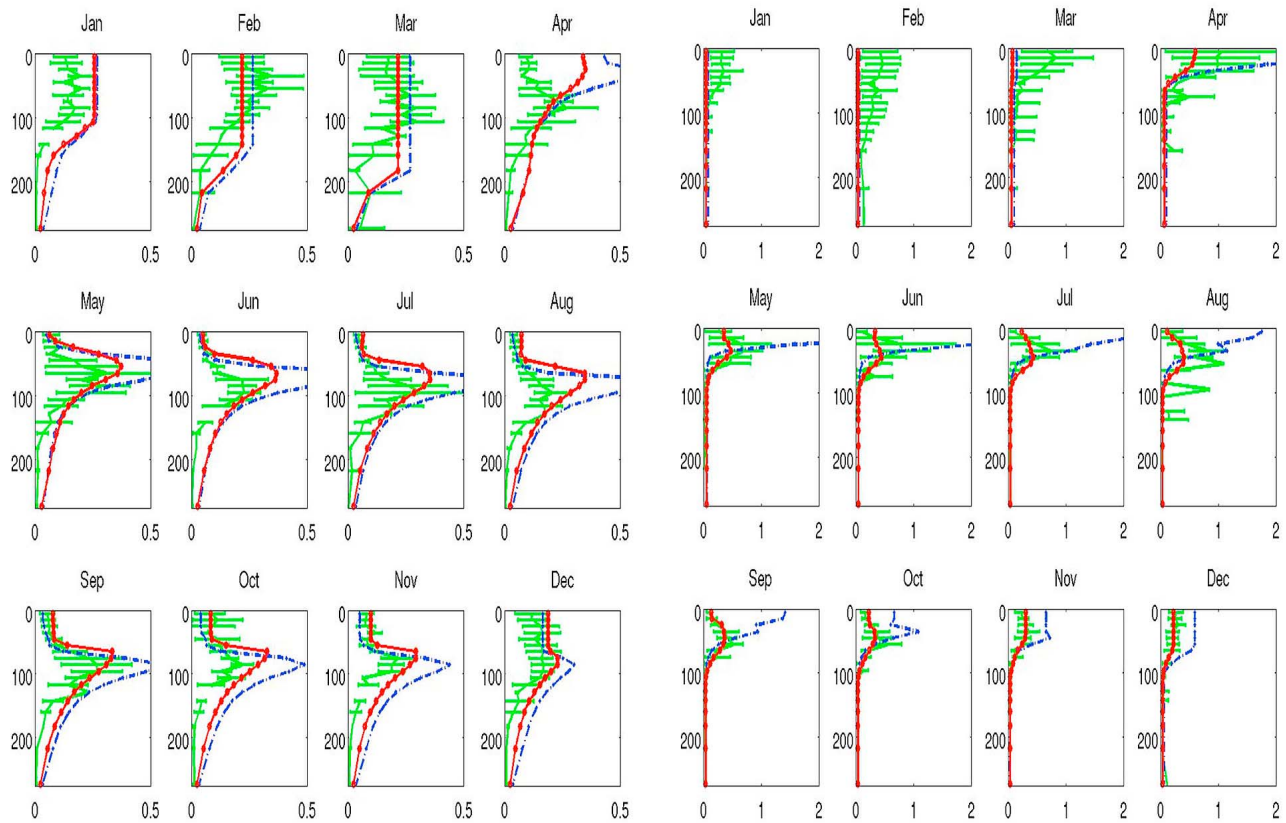
[36] Two global ocean simulations of 500 years each were performed, one using the standard PISCES parameters and the other the 45 optimized PISCES parameters (see Table 2). Before showing a detailed comparison of the two simulations, it is important to verify that the 3-D optimized PISCES model also improves the simulated vertical profiles of the main four biogeochemical variables (CHL, NO<sub>3</sub>, Si, POC) at the five JGOFS stations. It is possible that the assimilation procedure might have compensated for some bias in nutrient vertical profiles, due to the absence of advective processes during the 1-D assimilation, by modifying the parameters of the 1-D PISCES model. If this had happened, such an error compensation would be evident in the 3-D simulation since the 3-D optimized PISCES model would have reduced the agreement

between simulated and measured vertical profiles at the five stations. In addition, this optimized model would produce less realistic global chlorophyll fields, as compared to those of the SeaWiFS satellite climatology.

[37] At the exact location of the five JGOFS stations, we extracted the vertical profiles of the four biogeochemical variables used in section 4 from the two global 3-D simulations in order to verify that the use of the optimized parameters improves the fit with observations. Figure 3 is similar to Figure 2 except that the profiles are extracted from global 3-D simulations. Figure 3 shows that at both BATS and DYFAMED stations, the optimized results from the global 3-D simulation are much more realistic than those from the standard run, particularly in terms of the DCM at BATS and surface chlorophyll concentration at DYFAMED. Similar results are obtained for the entire set of the five stations, as shown in Table 4 in which the RMSE on CHL, NO<sub>3</sub>, and POC from the global 3-D standard model are most of the time larger than those from the global 3-D model with optimized parameters. However, this is not true for Si, which is worse in the optimized simulation and for POC at NABE (Table 4). As suggested in section 4, this problem probably arises from the fact that there is an underrepresentation of diatom-dominated waters at the five stations used here. The contribution of waters with diatoms to our data set is too small to allow us to rigorously optimize the parameters associated with this species. Despite this diatom-specific issue, Figure 2 and Table 3 show that the use of the 1-D-optimized parameters in the 3-D model yields improvements at the five JGOFS stations. Accordingly, our assimilation in 1-D provides a parameter set that improves the global 3-D model simulation (relative to the available observations), despite a more simplistic dynamics at the 1-D sites.

### 5.2. Comparison With the SeaWiFS Climatology

[38] SeaWiFS is a satellite ocean color sensor which was launched in 1997 by NASA. It has provided daily global observations of the chlorophyll concentration in surface waters for more than 12 years [McClain *et al.*, 2004, 2006] and its products are available online (<http://oceancolor.gsfc.nasa.gov>). Here we have used a 10 year monthly climatology (1997–2006) to evaluate our optimization of the PISCES model. The standard simulation has already been compared to SeaWiFS data by *Aumont and Bopp* [2006]. Figure 4 allows a qualitative comparison of the two simulations with the SeaWiFS climatology for the months of June and December. Figure 4 shows that the chlorophyll concentrations in productive regions (North Atlantic and Pacific in June, Southern Ocean in December) are lower in the optimized simulation than those obtained with the standard model. On the contrary, the optimized model yields higher chlorophyll concentrations than the standard model in oligotrophic regions (subtropical gyres essentially). The comparison with the SeaWiFS climatology suggests that in both cases, the changes associated with the optimized model are of the right sign and improve the ability of the global 3-D PISCES model to reproduce the concentration of chlorophyll in the global ocean. These two general defects of the standard PISCES simulation were already identified by *Aumont and Bopp* [2006]. It is important to note that our optimization cannot improve problems due to the low



**Figure 3.** Same as Figure 2 but using the 500 year 3-D simulations of the standard and optimized configurations.

spatial resolution of the model ( $2^\circ \times 2^\circ$ ), which cannot properly represent coastal upwellings (e.g., those adjacent to Mauritania and Angola) or western boundary currents (e.g., the Gulf Stream).

[39] In order to perform a more quantitative analysis of the improvement of the 3-D model due to the optimization of parameters in the 1-D version of the PISCES model, monthly mean surface Chlorophyll concentrations were extracted from the two global 3-D simulations (3-D standard and 3-D optimized) and compared to SeaWiFS data at 15 locations (termed “stations”) shown in Figure 5. Figure 6 and Table 5 compare qualitatively and quantitatively, respectively, the observed climatological seasonal cycle, using the two versions of PISCES (standard and optimized) at these 15 stations. As mentioned above, the seasonal cycle obtained with the optimized set of parameters (3-D global optimized) is in better agreement with measurements in productive regions in both northern (stations 1 and 11) and southern (stations 5, 10 and 15) hemispheres. The improvement is smaller, but still significant, in oligotrophic tropical regions (stations 12, 13, 14). In equatorial regions, the improvement is less marked. The optimization improved the representation of the seasonal cycle in the Indian Ocean (station 3), but has worsened the results in the Pacific Ocean (stations 7, 8, 9). It is important to note that a correct representation of equatorial regions, characterized by a strong upwelling, is particularly challenging for global biogeochemical models because of the sensitivity of the nutrient input to vertical velocities, which are very difficult

to properly simulate in such narrow regions (in terms of latitude) within global models [Aumont and Bopp, 2006].

## 6. Conclusions and Perspectives

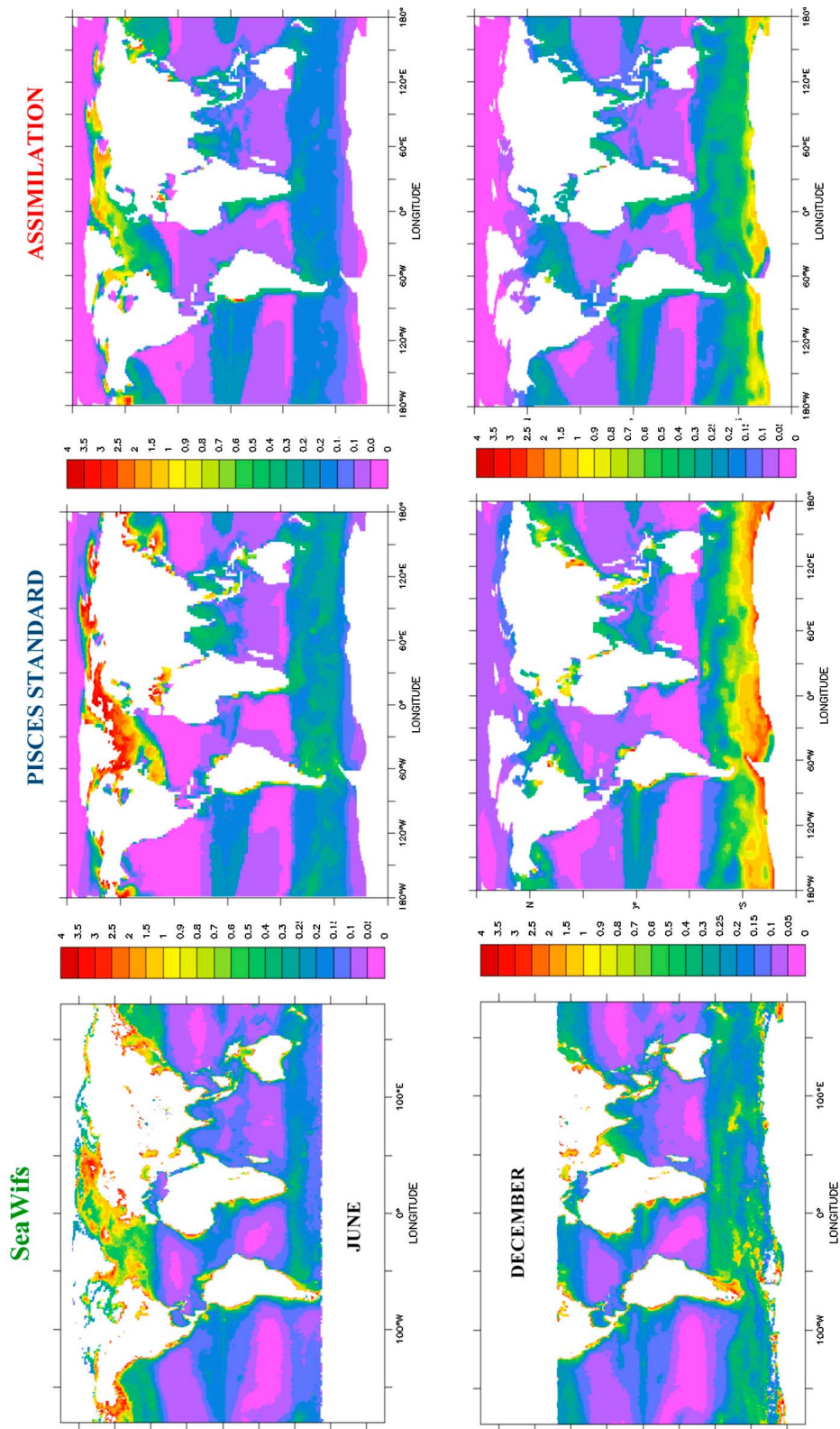
[40] Global ocean biogeochemical models are difficult to improve because they rely on a large number of physiological parameters that are poorly constrained. Moreover the scarcity and inherent limitations present in observation data sets make evaluating the performance of such models a very challenging task. Assimilation techniques provide a promising framework for dealing with such complex situations and for optimizing the parameters of a model, using diverse data sets. Here we have developed an optimization method based on a variational assimilation that improves the parameter values of the global PISCES model [Aumont and Bopp, 2006]. We used a limited data set of observations at five oceanographic stations (BATS,

**Table 4.** Same as Table 3 but Using the 500 Year 3-D Simulations of the Standard and Optimized Configurations<sup>a</sup>

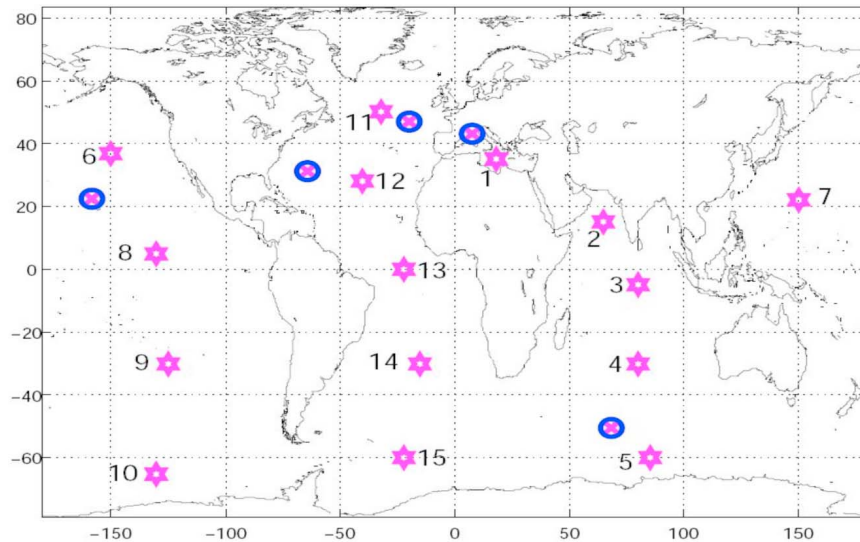
	CHL	NO3	POC	SIL
BATS	16/9	134/84	165/85	30/357
HOT	9/9			98/750
DYFAMED	98/22	185/154	430/414	585/1166
KERFIX	15/17	539/363		
NABE	101/107	211/183	529/708	

<sup>a</sup>Bold values indicate the pixels where the RMSE after assimilation is inferior to the one of the standard model.





**Figure 4.** Comparison for (top) June and (bottom) December of the monthly mean chlorophyll concentration in the surface layer of the (middle) standard and (right) optimized PISCES simulations with that of (left) the SeaWifs climatology (1997–2006).

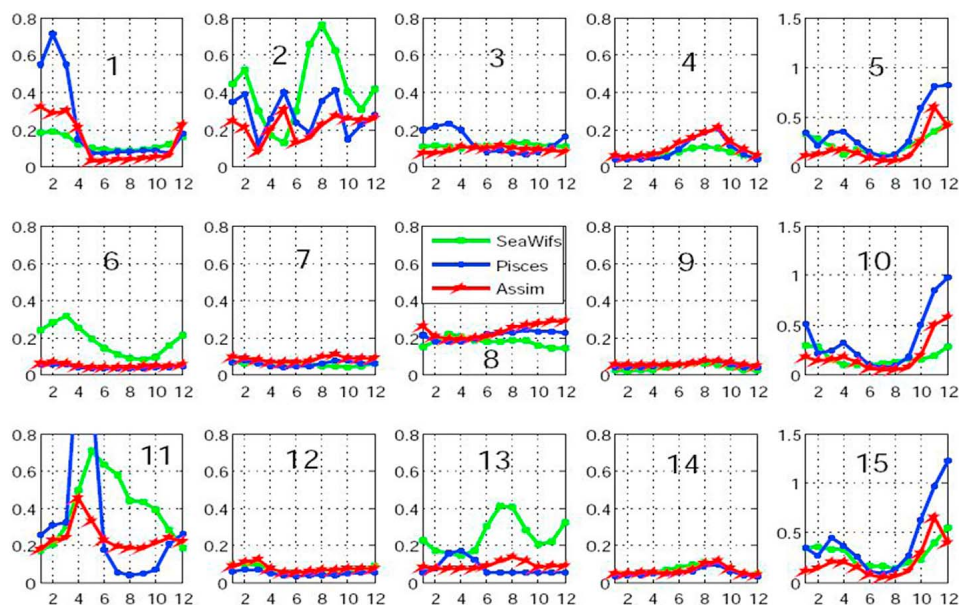


**Figure 5.** Location of the 15 stations (stars) used to study the seasonal cycle of the surface chlorophyll concentration in Figure 6 (see text). The five JGOFS stations are circled.

HOT, DYFAMED, NABE, KERFIX) from the JGOFS program. Monthly vertical profiles of several biogeochemical variables (CHL,  $\text{NO}_3$ , POC, Si), have been measured for decades at these stations, permitting us to build a monthly climatology that has been assimilated in the PISCES model.

[41] The optimization method was developed using the YAO software [Thiria *et al.*, 2006] and has required several original developments in order to derive a cost function that is adapted to this specific application. Applying an assimilation

technique to a global 3-D model (with thousands of grid points) for a whole annual cycle (with a time step of 1.5 h) is very expensive in computation time. This was why we decided to optimize the parameters of the 1-D version of the PISCES model by simultaneously assimilating the measurements taken at five JGOFS stations. But in the cost function, we gave distinct weights to each observation profile at each station according to their statistical confidence. These weights, which drive the efficiency of the



**Figure 6.** Seasonal cycle of the climatological surface Chlorophyll concentration from SeaWiFS data (green) and from the standard (blue) and optimized (red) PISCES 3-D simulations after 500 years for the 15 stations shown in Figure 5. Stations 1 to 5 are located in the Mediterranean and Indian Ocean, stations 6 to 10 are in the Pacific Ocean (from north to south), and stations 11 to 15 are in the Atlantic Ocean (from north to south).

**Table 5.** RMSE for Each Point Indicated in Figure 5

Comparison Dots	Latitude	Longitude	PISCES/SeaWiFS	Assimilation/SeaWiFS <sup>a</sup>
Black Sea (station 1)	35	18	21.5	<b>8.2</b>
North Indian (station 2)	15	65	23.8	28.5
Equatorial Indian (station 3)	-5	80	6.8	<b>2.6</b>
South Indian (station 4)	-30	80	4.3	4.8
Antarctic (ind) (station 5)	-60	85	21.1	<b>11.4</b>
North Pacific (station 6)	37	-150	15.5	<b>15.0</b>
North central Pacific (station 7)	22	150	1.5	3.2
Equatorial Pacific (station 8)	5	-130	5.6	8.4
South Pacific (station 9)	-30	-125	1.5	2.4
Antartic (Pac) (station 10)	-65	-130	31	<b>14.1</b>
North Atlantic (station 11)	50	-32	35.1	<b>22.9</b>
North central Atlantic (station 12)	28	-40	2.2	<b>1.5</b>
Equatorial Atlantic (station 13)	0	-22	21	<b>17.6</b>
South Atlantic (station 14)	-30	-15	1.8	<b>1.4</b>
Antarctic (Atlantic) (station 15)	-60	-22	28.7	<b>14.8</b>
Total mean	-	-	221.4	<b>156.8</b>

<sup>a</sup>Bold values indicate the pixels where the RMSE after assimilation is inferior to the one of the standard model.

assimilation, were computed through an original procedure presented in section 3.4. The assimilation provided a single set of parameters of the PISCES model that minimizes the differences between the model and the data at the five JGOFS stations.

[42] Such an approach was challenging for several reasons. On the one hand, even if the five JGOFS stations were located in contrasted biogeochemical regions, from oligotrophic to eutrophic zones, the underrepresentation of productive regions (and particularly those dominated by diatoms) was clearly an issue. On the other hand, several drawbacks to this approach arise from a few assumptions: (1) the use of a 1-D modeling relies on the hypothesis that advective processes can be neglected (or at least that their different effects compensate each other) at the chosen stations, which may not be true and (2) the fact that the assimilation only relies on biogeochemical variables, does not permit to correct errors or bias in ocean physics (such as the mixed-layer depth). The results presented here demonstrate that our assimilation approach is sound. The optimized PISCES model parameters obtained by assimilating the monthly climatology of observed state variables at the five JGOFS stations in the 1-D version of PISCES permit significant improvements (for almost all the state variables) in the global 3-D version of PISCES. The major improvements, compared to the standard version of PISCES, are a decrease in the eutrophic chlorophyll concentrations and a greater biomass in oligotrophic regions. The seasonal cycle in surface chlorophyll is also significantly improved, particularly in productive high-latitude regions. The main limitation of the approach concerns the diatom parameters, essentially because there were not enough in situ data available during the blooms of this species to make an efficient optimization. Another way of solving this problem in the future would be to combine surface Chl derived from satellites and the climatology [e.g., *Levitus et al.*, 1998, 2009, 1994; *Levitus*, 1982] for the other biogeochemical variables (Si, NO<sub>3</sub>) during the assimilation. This should equilibrate the contribution of diatoms with respect to that of nanophytoplankton even though some difficulty may arise because the areas with diatom dominance (North Atlantic, Southern Ocean) coincide with areas where the ocean transport model that forces the biogeochemical models is often deficient.

[43] Besides demonstrating that we have improved the accuracy of the PISCES model, this study proposes a sound methodology that could be used to efficiently account for in situ data in biogeochemical ocean models.

[44] **Acknowledgments.** The authors would like to thank all the people who participated in the development of the YAO software at LOCEAN and the implementation of PISCES: Charles Sorrow, Luigi Nardi, Loutfi Bouali, Carlos Mejia, Julien Bradjard, and Rachid Benshila. Abdou Kane acknowledges the French company ACRI-ST for the financial support of his Ph.D. and thanks Antoine Mangin and Odile Hembise Fanton d'Andon for their interest. This work was performed using HPC resources from GENCI-IDRIS (grant 2009-010040).

## References

- Anderson, B. D. O., and J. B. Moore (1971), *Linear Optimal Control*, Prentice-Hall, Englewood Cliffs, N. J.
- Anderson, B. D. O., and J. B. Moore (2007), *Optimal Control: Linear Quadratic Methods*, Dover, New York.
- Aumont, O., and L. Bopp (2006), Globalizing results from ocean in situ iron fertilization studies, *Global Biogeochem. Cycles*, *20*, GB2017, doi:10.1029/2005GB002591.
- Aumont, O., L. Bopp, and M. Schulz (2008), What does temporal variability in aeolian dust deposition contribute to sea-surface iron and chlorophyll distributions?, *Geophys. Res. Lett.*, *35*, L07607, doi:10.1029/2007GL031131.
- Badran, F., M. Berrada, J. Brajard, M. Crépon, C. Sorrow, S. Thiria, J.-P. Hermand, M. Meyer, L. Perichon, and M. Asch (2008), Inversion of satellite ocean colour imagery and geoacoustic characterization of seabed properties: Variational data inversion using a semi-automatic adjoint approach, *J. Mar. Syst.*, *69*, 126–136, doi:10.1016/j.jmarsys.2007.02.018.
- Berge, C. (1958), *Théorie des Graphes et Ses Applications*, Dunod, Paris.
- Bopp, L., K. E. Kohfeld, C. Le Quéré, and O. Aumont (2003), Dust impact on marine biota and atmospheric CO<sub>2</sub> during glacial periods, *Paleoceanography*, *18*(2), 1046, doi:10.1029/2002PA000810.
- Carr, M.-E., et al. (2006), A comparison of global estimates of marine primary production from ocean color, *Deep Sea Res., Part II*, *53*, 741–770, doi:10.1016/j.dsr2.2006.01.028.
- Fasham, M. J. R., H. W. Ducklow, and S. M. McKelvie (1990), A nitrogen-based model of plankton dynamics in the oceanic mixed layer, *J. Mar. Res.*, *48*, 591–639.
- Faugeras, B., M. Levy, L. Memery, J. Verron, J. Blum, and I. Charpentier (2003), Can biogeochemical fluxes be recovered from nitrate and chlorophyll data? A case study assimilating data in the northwestern Mediterranean Sea at the JGOFS-DYFAMED station, *J. Mar. Syst.*, *40–41*, 99–125, doi:10.1016/S0924-7963(03)00015-0.
- Faugeras, B., O. Bernard, A. Sciandra, and M. Levy (2004), A mechanistic modeling and data assimilation approach to estimate the carbon/chlorophyll and carbon/nitrogen ratios in a coupled hydrodynamical-biological model, *Nonlinear Process. Geophys.*, *11*, 515–533, doi:10.5194/npg-11-515-2004.

- Friedrichs, M. A. M. (2001), Assimilation of JGOFS EqPac and SeaWiFS data into a marine ecosystem model of the central equatorial Pacific Ocean, *Deep Sea Res., Part II*, 49, 289–319, doi:10.1016/S0967-0645(01)00104-7.
- Gilbert, J. C., and C. Lemaréchal (1989), Some numerical experiments with variable-storage quasi-Newton algorithms, *Math. Program.*, 45, 407–435, doi:10.1007/BF01589113.
- Kleypas, J. A., and S. C. Doney (2001), Nutrients, chlorophyll, primary production and related biogeochemical properties in the ocean mixed layer—A compilation of data collected at nine JGOFS sites, *Tech. Note NCAR/TN-447+STR*, 55 pp., Natl. Cent. for Atmos. Res., Boulder, Colo.
- Le Dimet, F.-X., and O. Talagrand (1986), Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects, *Tellus, Ser. A*, 38A, 97–110, doi:10.1111/j.1600-0870.1986.tb00459.x.
- Le Quéré, C., et al. (2005), Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models, *Global Change Biol.*, 11, 2016–2040.
- Levitus, S. (1982), *Climatological Atlas of the World Ocean*, NOAA Prof. Pap. 13, U.S. Gov. Print. Off., Washington, D. C.
- Levitus, S., R. Gelfeld, T. Boyer, and D. Johnson (1994), Results of the NODC Oceanographic Data and Archaeology and Rescue Project, 73 pp., *Key to Oceanogr. Rec. Doc. 19*, U.S. Gov. Print. Off., Washington, D. C.
- Levitus, S., T. P. Boyer, M. E. Conkright, T. O'Brien, J. Antonov, C. Stephens, L. Stathopoulos, D. Johnson, and R. Gelfeld (1998), *World Ocean Data Base 1998*, vol. 1, *Introduction*, NOAA Atlas NESDIS, vol. 18, 346 pp., NOAA, Silver Spring, Md.
- Levitus, S., J. I. Antonov, T. P. Boyer, R. A. Locarnini, H. E. Garcia, and A. V. Mishonov (2009), Global ocean heat content 1955–2008 in light of recently revealed instrumentation problems, *Geophys. Res. Lett.*, 36, L07608, doi:10.1029/2008GL037155.
- Lions, J. L. (1968), *Contrôle Optimal de Systèmes Gouvernés par des Équations aux Dérivées Partielles*, Dunod Gauthier-Villars, Paris.
- McClain, C., S. B. Hooker, G. Feldman, and P. Bontempi (2006), Satellite data for ocean biology, biogeochemistry, and climate research, *Eos Trans. AGU*, 87(34), 337–343, doi:10.1029/2006EO340002.
- McClain, C. R., G. C. Feldman, and S. B. Hooker (2004), An overview of the SeaWiFS project and strategies for producing a climate research quality global ocean bio-optical time series, *Deep Sea Res., Part II*, 51, 5–42, doi:10.1016/j.dsr2.2003.11.001.
- Morel, A., and J. F. Berthon (1989), Surface pigments, algal biomass profiles, and potential production of the euphotic layer: Relationships reinvestigated in view of remote-sensing applications, *Limnol. Oceanogr.*, 34, 1545–1562, doi:10.4319/lo.1989.34.8.1545.
- Oschlies, A., and M. Schartau (2005), Basin-scale performance of a locally optimized marine ecosystem model, *J. Mar. Res.*, 63, 335–358, doi:10.1357/0022240053693680.
- Schartau, M., and A. Oschlies (2003), Simultaneous data-based optimization of a 1D-ecosystem model at three locations in the North Atlantic: Part I—Method and parameter estimates, *J. Mar. Res.*, 61, 765–793, doi:10.1357/002224003322981147.
- Schlitzer, R. (2000), Applying the adjoint method for global biogeochemical modeling, in *Inverse Method in Global Biogeochemical Cycles*, *Geophys. Monogr. Ser.*, vol. 114, edited by P. Kasibhatla et al., pp. 107–124, AGU, Washington, D. C.
- Schneider, B., L. Bopp, M. Gehlen, J. Segsneider, T. L. Frölicher, P. Cadule, P. Friedlingstein, S. C. Doney, M. J. Behrenfeld, and F. Joos (2008), Climate-induced interannual variability of marine primary and export production in three global coupled climate carbon cycle models, *Biogeosciences*, 5, 597–614, doi:10.5194/bg-5-597-2008.
- Tagliabue, A., L. Bopp, and O. Aumont (2008), Ocean biogeochemistry exhibits contrasting responses to a large scale reduction in dust deposition, *Biogeosciences*, 5, 11–24, doi:10.5194/bg-5-11-2008.
- Talagrand, O., and P. Courtier (1987), Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory, *Q. J. R. Meteorol. Soc.*, 113, 1311–1328.
- Thiria, S., F. Badran, and C. Sorror (2006), *YAO: Un logiciel pour les modèles numériques et l'assimilation de données (schéma direct adjoint et assimilation variationnelle)*, LOCEAN, Paris.
- Tjiputra, J. F., D. Polzin, and A. M. E. Winguth (2007), Assimilation of seasonal chlorophyll and nutrient data into an adjoint three-dimensional ocean carbon cycle model: Sensitivity analysis and ecosystem parameter optimization, *Global Biogeochem. Cycles*, 21, GB1001, doi:10.1029/2006GB002745.
- Weaver, A. T., and P. Courtier (2001), Correlation modelling on the sphere using a generalized diffusion equation, *Q. J. R. Meteorol. Soc.*, 127, 1815–1846, doi:10.1002/qj.49712757518.
- Weaver, A. T., C. Deltel, E. Machu, S. Machu, and N. Daget (2005), A multivariate balance operator for variational ocean data assimilation, *Q. J. R. Meteorol. Soc.*, 131, 3605–3625.
- O. Aumont, Laboratoire de Physique des Océans, IUEM, CNRS-IRD-UBO, F-29285 Brest, France.
- F. Badran, Centre d'Etudes et de Recherche en Informatique du CNAM/CNAM, F-75141 Paris, France.
- M. Berrada, M. Crépon, and S. Thiria, Laboratoire d'Océanographie et du Climat: Expérimentation et Approches Numériques/IPSL, CNRS-IRD-UPMC, F-75005 Paris, France.
- L. Bopp, A. Kane, C. Moulin, and A. Tagliabue, Laboratoire des Sciences du Climat et de l'Environnement/IPSL, CEA-CNRS-UVSQ, F-91191 Gif-sur-Yvette, France. (abdou.kane@lsce.ipsl.fr)