A 3-D variational assimilation scheme in coupled transport-biogeochemical models: Forecast of Mediterranean biogeochemical properties

Anna Teruzzi,[1] Srdjan Dobricic,[2] Cosimo Solidoro,[1,3] and Gianpiero Cossarini[1]

Received 16 July 2013; revised 27 November 2013; accepted 2 December 2013; published 13 January 2014.

[1] Increasing attention is dedicated to the implementation of suitable marine forecast systems for the estimate of the state of the ocean. Within the framework of the European MyOcean infrastructure, the pre-existing short-term Mediterranean Sea biogeochemistry operational forecast system has been upgraded by assimilating remotely sensed ocean color data in the coupled transport-biogeochemical model OPATM-BFM using a 3-D variational data assimilation (3D-VAR) procedure. In the present work, the 3D-VAR scheme is used to correct the four phytoplankton functional groups included in the OPATM-BFM in the period July 2007 to September 2008. The 3D-VAR scheme decomposes the error covariance matrix using a sequence of different operators that account separately for vertical covariance, horizontal covariance, and covariance among biogeochemical variables. The assimilation solution is found in a reduced dimensional space, and the innovation for the biogeochemical variables is obtained by the sequential application of the covariance operators. Results show a general improvement in the forecast skill, providing a correction of the basin-scale bias of surface chlorophyll concentration and of the local-scale spatial and temporal dynamics of typical bloom events. Further, analysis of the assimilation skill provides insights into the functioning of the model. The computational costs of the assimilation scheme adopted are low compared to other assimilation techniques, and its modular structure facilitates further developments. The 3D-VAR scheme results especially suitable for implementation within a biogeochemistry operational forecast system.

Citation: Teruzzi, A., S. Dobricic, C. Solidoro, and G. Cossarini (2014), A 3-D variational assimilation scheme in coupled transport-biogeochemical models: Forecast of Mediterranean biogeochemical properties, *J. Geophys. Res. Oceans*, *119*, 200–217, doi:10.1002/2013JC009277.

1. Introduction

[2] The use of data assimilation (DA) in biogeochemical ocean modeling has a more recent history than other geophysical disciplines (e.g., meteorological and physical oceanography) mainly because of the complexity of biogeochemical models and the poor availability of large data sets. In most biogeochemical sequential methods developed from estimation theory based on the Kalman-filter approach are used [*Gregg et al*., 2009]. Examples found in the literature include: optimal interpolation (OI) [*Anderson et al*., 2000; *Popova et al*., 2002], ensemble Kalman filter (EnKF) [*Natvik and Evensen*, 2003; *Simon and Bertino*, 2009; *Ciavatta et al*., 2011], and various implementations of the extended Kalman filter, such as the singular evolutive extended Kalman filter (SEEK) [*Carmillet et al*., 2001; *Triantafyllou et al*., 2007; *Korres et al*., 2012] and the singular evolutive interpolated Kalman filter (SEIK) [*Nerger and Gregg*, 2008]. A recently developed ensemble-based sequential approach is provided by particle filters [*Mattern et al*., 2013].

[3] Variational schemes, the other main class of assimilation methods, are quite common in meteorology and physical oceanography [e.g., *Rabier*, 2005; *Bennett et al*., 2008] but have been less popular in biogeochemical applications, and can be found only for parameter estimations [*Lawson et al*., 1996; *Spitz et al*., 1998; *Friedrichs*, 2001; *Garcia-Gorriz et al*., 2003; *Tjiputra et al*., 2007; *Ward et al*., 2010; *Kane et al*., 2011].

[4] The solution of the assimilation in the variational approaches is based on the minimization of a cost function [*Lorenc*, 1986]. Crucial issues in the implementation of

[1]Istituto Nazionale di Oceanografia e di Geofisica Sperimentale, Trieste, Italy.
[2]Centro Euro-Mediterraneo sui Cambiamenti Climatici, Bologna, Italy.
[3]International Centre for Theoretical Physics, Trieste, Italy.

Corresponding author: G. Cossarini, Istituto Nazionale di Oceanografia e di Geofisica Sperimentale, Borgo Grotta Gigante 42c, Sgonico, Trieste IT-34010, Italy. (gcossarini@ogs.trieste.it)

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

©2013. The Authors. Journal of Geophysical Research: Oceans published by Wiley on behalf of the American Geophysical Union.
2169-9275/14/10.1002/2013JC009277

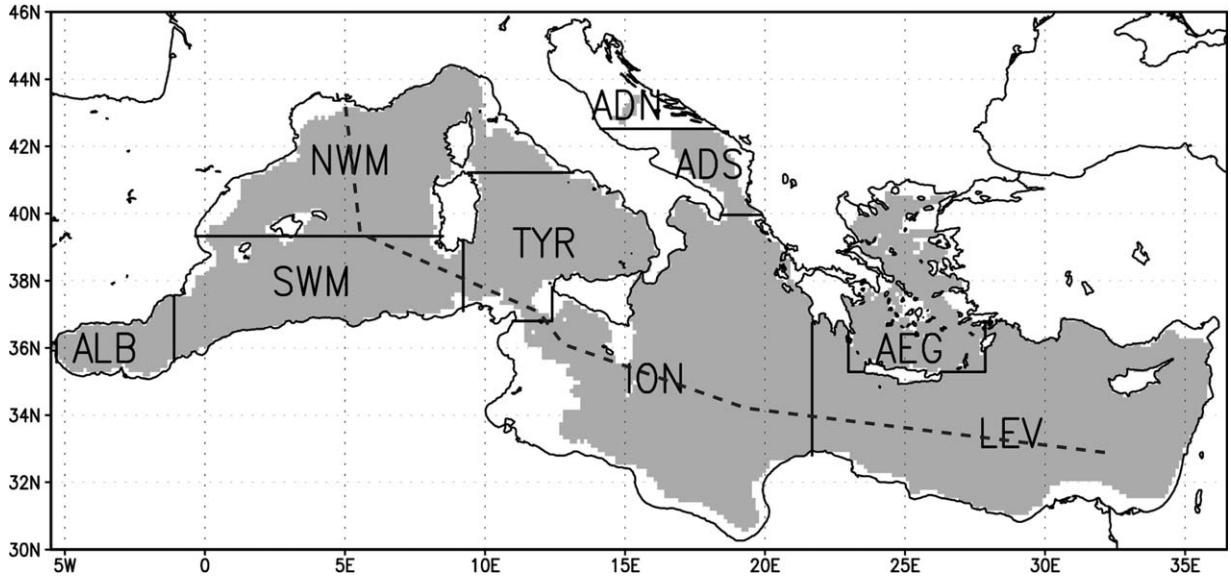

Figure 1. Model domain and subregion subdivision: Alboran Sea (ALB), South Western Mediterranean (SWM), North Western Mediterranean (NWM), Tyrrhenian Sea (TYR), Northern Adriatic Sea (ADN), Southern Adriatic Sea (ADS), Ionian Sea (ION), Aegean Sea (AEG), and Levantine (LEV). Shaded areas are deeper than 200 m. The dashed line indicates the transect of the BOUM cruise.

variational schemes are the definition of the background error covariance matrix, which describes the uncertainty of the model variables and the covariance among them, and the inversion of the background error covariance matrix, which has very large dimensions in geophysical problems (proportional to the spatial size of the model multiplied by the number of state variables).

[5] Using the approach proposed by *Dobricic and Pinardi* [2008], the background covariance matrix can be decomposed into a series of operators. This decomposition makes it possible to define the background error covariance matrix in a modular framework, where each operator describes a different aspect of the correlation between the model variables. The operators can be separately developed and they may be matrices, diagnostic models, or even dynamical models. Decomposition allows to reformulate the cost function in such a way that the inversion of the background covariance matrix is not required. On the other hand, since the operators can be nonlinear, the errors described by the background covariance matrix may be not Gaussian, and the solution of the minimization of the cost function can be not optimal [e.g., *Pires et al*., 1996].

[6] A few recent examples demonstrate that DA techniques are becoming mature for operational forecasting in ocean biogeochemistry [*Fontana et al*., 2010; *Ford et al*., 2012]. Both of the cited works explore the possibility of assimilating ocean color data into a biogeochemical model. *Fontana et al*. [2010] apply the SEEK filter for DA in a coastal area, showing that assimilation improves the skill of a chlorophyll forecast. In the work by *Ford et al*. [2012], a multivariate assimilation is carried out, and the results show that assimilation not only improves the chlorophyll concentration estimates but also that the other model variables are not degraded by assimilation.

[7] Operational systems can benefit from assimilation approaches which, although not optimal, are computationally efficient and can be easily developed. In the present application, a 3-D variational (3D-VAR) approach developed from the one applied by *Dobricic and Pinardi* [2008] is used to constrain the three-dimensional fields of the four phytoplankton functional types in the OPATM-BFM medium complexity biogeochemical model [*Lazzari et al*., 2010]. The observations used in the present assimilation experiment are satellite surface chlorophyll maps [*Volpe et al*., 2007], which are presently the only source of data with sufficient temporal and spatial coverage to constrain a basin-scale three-dimensional ocean biogeochemical model.

[8] The paper is organized as follows. Section 2 describes the study site of the assimilation experiment. In section 3, the biogeochemical model is presented. Section 4 describes the data used for the assimilation. The formulation of the operators of the variational scheme, which constitute the core of the present assimilation experiment, is discussed in section 5. The implementation of the assimilation in the forecast system is then presented in section 6. Section 7 shows the results and their validation, and section 8 contains a discussion and final remarks.

2. Study Site

[9] The present assimilation experiment is applied to the Mediterranean Sea (MS) from the Gibraltar strait to the Dardanelles strait (Figure 1). The domain is subdivided into nine subregions which are used in the definition of the assimilation scheme and in the analysis of the results. This subdivision, similar to that used by *Lazzari et al*. [2012], follows the results obtained for the Western Mediterranean by *D'Ortenzio and Ribera d'Alcalà* [2009], based on a cluster analysis of satellite data, and the provinces proposed by *Morel and André* [1991]. In the Eastern Mediterranean, the adopted subdivision is very close (except for the north-south Levantine division) to that proposed by *Antoine et al*. [1995], which is based on considerations of circulation characteristics and color scanner observations.

21699291, 2014, 1, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2013JC009277 by Ogs Trieste Istituto Nazionale, Wiley Online Library on [18/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3. OPATM-BFM Model

[10] The OPATM-BFM (Océan Parallelisé Tracer Model-Biogeochemical Flux Model) is the product of long-term research on the biogeochemical characteristics of the MS [*Crise et al*., 1998; *Crispi et al*., 2001, 2002; *Solidoro et al*., 2003; *Lazzari et al*., 2012]. The OPATM-BFM model is the coupling of a properly modified version of the OPA transport model [*Lazzari et al*., 2010] and of the BFM biogeochemical model [*Vichi et al*., 2007; *Lazzari et al*., 2012, www.bfm-community.eu]. The model describes the MS with a horizontal resolution of 1/8° and 72 nonuniform vertical levels.

[11] The OPATM transport model, which solves the terms for advection and diffusion, vertical mixing, and sinking, is coupled offline with the Mediterranean ocean forecasting system (MFS) model [*Oddo et al*., 2009], which includes a variational assimilation scheme. As part of the MyOcean infrastructure, MFS model provides the daily fields of current velocity components, temperature, salinity, eddy diffusivity, wind stress, and solar radiation to the OPATM-BFM model. The MFS products are linearly interpolated to obtain the required physical forcings at the temporal resolution of the OPATM-BFM model (30 min). Further, the MFS forcings are spatially interpolated from 1/16° to 1/8°, which is the horizontal resolution of the OPATM-BFM model [*Lazzari et al*., 2010].

[12] The BFM resolves the biological relationships and chemical reactions between four phytoplankton and four zooplankton functional types, a bacterial pool, four macronutrients and particulate and dissolved matter pools. Carbon, nitrogen, phosphorous, and silicon are considered in the formulation. In the BFM model, the phytoplankton chlorophyll content is parameterized using a dynamic quota. The equation which regulates the rate of chlorophyll content in the phytoplankton groups is composed of two terms: net chlorophyll synthesis and losses due to zooplankton grazing. Net chlorophyll synthesis is regulated by environmental conditions (light, temperature, and availability of nutrients) and the phytoplankton internal content of nitrogen, phosphorous, and silicon (the latter is relevant only for the diatoms group). Details concerning the model formulation of chlorophyll are provided in Appendix A.

[13] The boundary conditions for biogeochemical variables at the Gibraltar strait are reproduced by a nudging scheme using seasonally varying climatological values. The rivers and coastal input formulation follows the method proposed by *Ludwig et al*. [2009], with inputs which vary seasonally for the major rivers (Po, Rhone, and Ebro) and are considered constant for the other sources. The atmospheric deposition rates of nutrients are constant during the year, and their values are based on the *Ribera d'Alcalà et al*. [2003] synthesis. The light extinction coefficient (necessary for the evaluation of light attenuation in the water column) varies in space and time and is derived from satellite observations [*Lazzari et al*., 2012].

[14] The OPATM-BFM model has been successfully used for both multiannual and scenario simulations [*Lazzari et al*., 2012, 2013] and for short-term simulations in an operational framework [*Teruzzi et al*., 2011], and the results of the model are consistent with the main known biogeochemical features at seasonal and basin scales [*Lazzari et al*., 2012].

4. Satellite Data and Observation Error

[15] The observations used in the assimilation consist of surface chlorophyll maps obtained from data collected by the MODIS-aqua satellite. Chlorophyll values are computed by means of an algorithm developed for the MS that accounts for its specific bio-optical characteristics [*Volpe et al*., 2007]. The surface chlorophyll maps are provided at a daily frequency and a spatial resolution of 1 km. Since 2009 the observations have been available as an operational product of the MyOcean project [*Volpe et al*., 2012]. The satellite chlorophyll data are spatially interpolated to the Mediterranean model resolution (1/8°) by means of a bilinear interpolation, and a 5 day average centered on the assimilation dates is performed, using the data provided as log-transformed chlorophyll concentrations [*Volpe et al*., 2012]. The 5 day averages are transformed back to chlorophyll concentrations and finally used in the assimilation. The 5 day aggregation increases the coverage of the satellite maps significantly. Indeed, the mean coverage is nearly equal to 90% of the model grid points, and none of the maps has less than 40% of coverage. Thus, the aggregation ensures sufficient data coverage for the assimilation, which would not be provided by satellite maps at a higher frequency. The observations in regions with water depths of less than 200 m (the white areas in Figure 1) are not used for the assimilation because of the high uncertainty of ocean color data in coastal water [*Saba et al*., 2011; *Odermatt et al*., 2012].

[16] The observation error covariance matrix (see section 5) provides an estimate of observation accuracy, which is one of the elements that constrain the 3D-VAR solution. In the present application, the observation error covariance matrix is a diagonal matrix, where each value of the diagonal provides the variance of a grid point. The observation error covariance varies in time: a diagonal matrix has been evaluated for each of the 12 months using all the available maps of the SeaWiFS time series for the period 1997–2010.

5. 3D-VAR Assimilation and Background Error Covariance

[17] The variational approach provides the assimilated field (the analysis) through the minimization of a cost function, which is defined on the basis of Bayes' theorem [*Lorenc*, 1986]. In the 3D-VAR assimilation, the analysis obtained by minimization provides the new initial condition for the model forecast. The cost function relies on the misfit between the model results and the observations, weighted according to their accuracy estimation (the respective error covariance matrices). Defining $\mathbf{x}$ as the state vector of the OPATM-BFM and $\mathbf{y}$ as the observations available, at each assimilation step k the cost function J is:

$$J(\mathbf{x}_k^a) = \frac{1}{2}\left(\mathbf{x}_k^a - \mathbf{x}_k^f\right)^{\mathrm{T}} \mathbf{B}_k^{-1}\left(\mathbf{x}_k^a - \mathbf{x}_k^f\right) + \frac{1}{2}\left(\mathbf{y}_k - H_k(\mathbf{x}_k^a)\right)^{\mathrm{T}} \mathbf{R}_k^{-1}\left(\mathbf{y}_k - H_k(\mathbf{x}_k^a)\right), \quad (1)$$

where $\mathbf{x}_k^f$ is the model forecast before the assimilation, $\mathbf{x}_k^a$ is the analysis, which is the initial condition for the forecast

21699291, 2014, 1, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2013JC009277 by Ogs Trieste Istituto Nazionale, Wiley Online Library on [18/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

in the following step, and $\mathbf{x}_k^a - \mathbf{x}_k^f$ is the correction made by the assimilation to the model estimation. The matrices $\mathbf{B}_k$ and $\mathbf{R}_k$ are the model and observation error covariance matrices, respectively, and the term H_k is the observational operator, which transforms the state vector $\mathbf{x}$ into the observation space.

[18] In the present application, the dimensions of the covariance matrix $\mathbf{B}_k$ are N-by-N, where N is the size of the state vector equal to the number of the model variables (51) times the number of grid points (nearly equal to 10^6). The size of observation vectors $\mathbf{y}_k$ (surface chlorophyll maps) is approximately equal to 10^3, and the dimensions of $\mathbf{R}_k$ are 10^3-by-10^3. However, assuming that the observation errors are uncorrelated, the observation error matrix $\mathbf{R}_k$ can be reduced to a diagonal matrix (section 4), and thus its inversion is easily achievable.

[19] Assuming that the observational operator is linearized, the cost function can be written as:

$$J(\delta\mathbf{x}_k) = \frac{1}{2}\delta\mathbf{x}_k{}^{\mathrm{T}}\mathbf{B}_k^{-1}\delta\mathbf{x}_k + \frac{1}{2}(\mathbf{d}_k - \mathbf{H}\delta\mathbf{x}_k)^{\mathrm{T}}\mathbf{R}_k^{-1}(\mathbf{d}_k - \mathbf{H}\delta\mathbf{x}_k), \quad (2)$$

where $\mathbf{H}_k$ is the linearized observational operator, the misfit vector is defined as $\mathbf{d}_k = \mathbf{y}_k - H_k(\mathbf{x}_k^f)$, and the innovation vector $\delta\mathbf{x}_k$ is equal to $\mathbf{x}_k^a - \mathbf{x}_k^f$.

[20] According to the scheme implemented by *Dobricic and Pinardi* [2008], background error covariance $\mathbf{B}$ can be approximated in terms of a matrix product $\mathbf{B} = \mathbf{V}\mathbf{V}^{\mathrm{T}}$ (the subscript k is omitted for simplicity), where the matrix $\mathbf{V}$ has dimensions of N-by-n where n is much less than N. A new control variable $\mathbf{v}$ (a vector of size n) can then be related to the innovation $\delta\mathbf{x}$ by the matrix $\mathbf{V}$:

$$\delta\mathbf{x} = \mathbf{V}\mathbf{v}, \quad (3)$$

and by means of opportune substitutions, the cost function can be rewritten as a function of $\mathbf{v}$:

$$J(\mathbf{v}) = \frac{1}{2}\mathbf{v}^{\mathrm{T}}\mathbf{v} + \frac{1}{2}(\mathbf{d} - \mathbf{H}\mathbf{V}\mathbf{v})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{d} - \mathbf{H}\mathbf{V}\mathbf{v}), \quad (4)$$

where the inversion of matrix $\mathbf{B}$ is no longer required and the solution of the minimization is a vector $\mathbf{v}$ of size n. It is worth noting that the dimension, n, of the control variable depends on the definition of $\mathbf{V}$, as discussed below.

[21] The solution $\mathbf{v}$ is iteratively computed using the quasi-Newton L-BFGS minimizer [*Byrd et al.*, 1995], which requires the computation of the cost function gradient:

$$J'(\mathbf{v}) = \mathbf{v} - \mathbf{V}^T\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{d} - \mathbf{H}\mathbf{V}\mathbf{v}). \quad (5)$$

[22] Once the solution $\mathbf{v}$ is found, the innovation $\delta\mathbf{x}$ is evaluated by means of equation (3). The definition of the matrix $\mathbf{V}$, which is a core issue of the approach, should be able to properly describe the model covariance $\mathbf{B}$, and to this end it can be based on a priori knowledge of $\mathbf{B}$. In particular, $\mathbf{V}$ can be decomposed into a sequential linear product of a number of operators $\mathbf{V}_{\mathrm{i}}$ [*Weaver et al.*, 2003; *Dobricic and Pinardi*, 2008] which account for different components of $\mathbf{B}$. For the surface chlorophyll assimilation, the decomposition of $\mathbf{V}$ has been defined as:

$$\mathbf{V} = \mathbf{V}_{\mathrm{b}}\mathbf{V}_{\mathrm{H}}\mathbf{V}_{\mathrm{V}}, \quad (6)$$

which is designed to describe the vertical error covariance of the chlorophyll fields ($\mathbf{V}_{\mathrm{V}}$), the horizontal error covariance ($\mathbf{V}_{\mathrm{H}}$), and the state variable error covariance ($\mathbf{V}_{\mathrm{b}}$). Due to the decomposition of $\mathbf{V}$ defined in equation (6), the solution $\mathbf{v}$ provides information on the surface chlorophyll innovation. Then the three-dimensional innovation for the biogeochemical variables (equation (3)) is obtained through the sequential application to $\mathbf{v}$ of: $\mathbf{V}_{\mathrm{V}}$, which propagates the information along the vertical profile; $\mathbf{V}_{\mathrm{H}}$, which acts as an a horizontal filter propagating the innovation to grid points where observations are not available; and $\mathbf{V}_{\mathrm{b}}$, which provides innovations for the four types of phytoplankton in terms of their chlorophyll, nitrogen, phosphorous, and carbon content. Details on the formulation of the $\mathbf{V}_i$ operators are provided hereafter.

[23] The development of the operator $\mathbf{V}_{\mathrm{b}}$ is an entirely original result of the present work, while the formulations of the operators $\mathbf{V}_{\mathrm{H}}$ and $\mathbf{V}_{\mathrm{V}}$ have been developed from those already used for assimilation of temperature and salinity [*Dobricic and Pinardi*, 2008]. The evaluation of the cost function and its gradient in the 3D-VAR algorithm requires that the adjoint of $\mathbf{V}$ is defined (equations (4) and (5)), and since $\mathbf{V}$ is decomposed in a sequence of operators, the adjoint of each operator is needed. The $\mathbf{V}_{\mathrm{b}}$ adjoint operator was hand-coded similarly to the adjoint operators of $\mathbf{V}_{\mathrm{H}}$, $\mathbf{V}_{\mathrm{V}}$, and $\mathbf{H}$ following the rules of *Giering and Kaminski* [1998]. In particular, because each of the $\mathbf{V}_i$ operators is computed by an independent program subroutine, the adjoint of $\mathbf{V}$ is composed of the sequence of the adjoint program subroutines of each operator.

5.1. The Vertical Covariance Operator $\mathbf{V}_{\mathrm{V}}$

[24] The vertical component of $\mathbf{B}$ is given by a set of chlorophyll profiles that are evaluated for each grid point by means of an empirical orthogonal function (EOF) decomposition [*Davis*, 1976]. Given a set of EOFs of chlorophyll profiles for each grid point, $\mathbf{v}$ has the dimension of the number of the EOFs ($\mathrm{N_{EOF}}$) times the number of surface grid points for which an observation is available. The application of $\mathbf{V}_{\mathrm{v}}$ to $\mathbf{v}$ provides a 3-D field of chlorophyll innovation.

[25] Assuming that the vertical error covariance varies in time and in space, the EOF decomposition is applied separately to each of the nine Mediterranean subregions (Figure 1) and to each month of the year. The EOF decomposition considers the anomalies of the profiles with respect to the average profile $\overline{\mathbf{p}}_m^r$ for a given month m and subregion r:

$$\overline{\mathbf{p}}_m^r = \frac{1}{N_m N_r}\sum_i^{N_m}\sum_j^{N_r}\mathbf{p}_i^j, \quad (7)$$

[26] where N_m is the number of model outputs (10 day average) for month m, N_r is the number of profiles for a given subregion r, and $\mathbf{p}_i^j$ is the profile for the ith date of the month m and the jth model grid point of the subregion r.

21699291, 2014, 1, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2013JC009277 by Ogs Trieste Istituto Nazionale, Wiley Online Library on [18/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

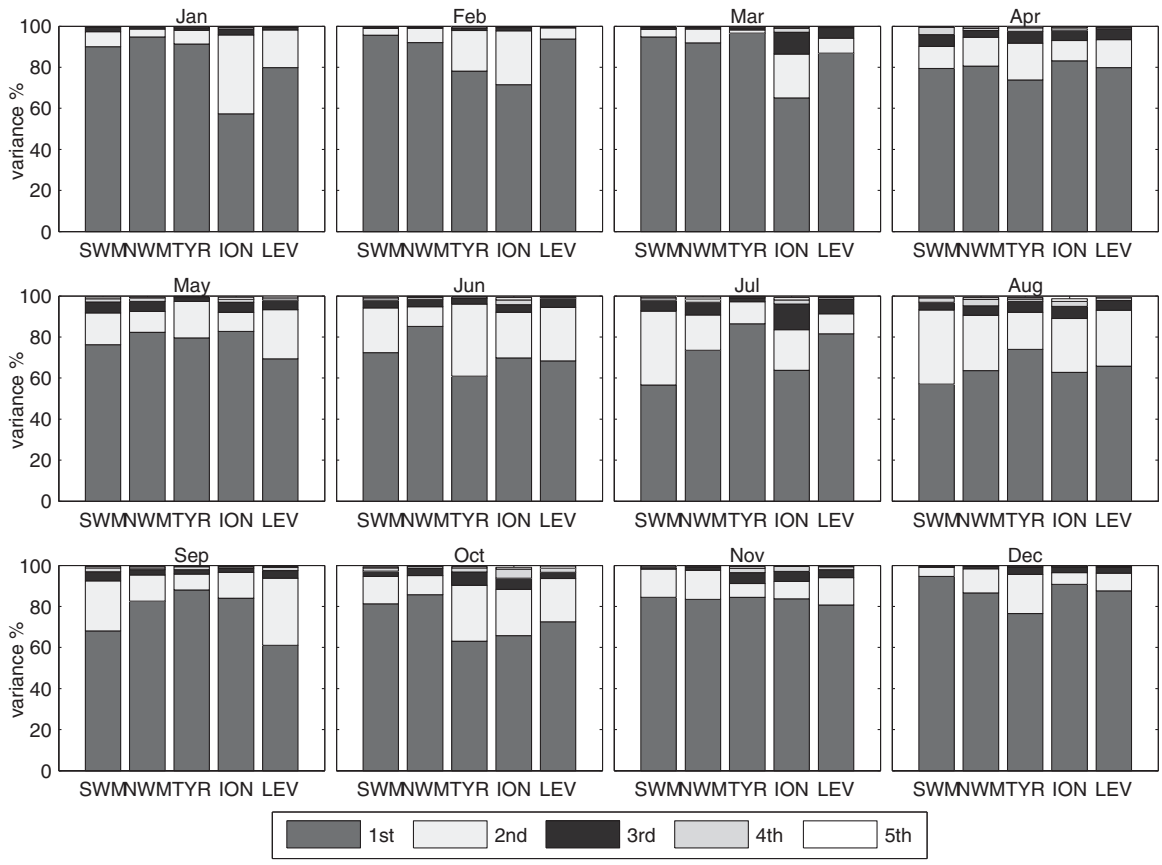

Figure 2. Percentage of the variance accounted for by the first five EOF modes for selected subregions.

[27] Given the N_m spatial average profiles for the subregion r,

$$\hat{\mathbf{p}}_i^r = \frac{1}{N_r} \sum_J^{N_r} \mathbf{p}_i^j, \tag{8}$$

the EOF decomposition is applied to the anomalies $\tilde{\mathbf{p}}_i^r = \hat{\mathbf{p}}_i^r - \overline{\mathbf{p}}_m^r$ obtaining a set of N_{EOF} profiles $\mathbf{e}_{m,k}^r$ and eigenvalues $\lambda_{m,k}^r$ for each month and for each subregion. Each pair of an EOF profile and its eigenvalue forms an EOF mode of the innovation in the control space of the variational scheme. The solution of the minimization $\mathbf{v}$ is, therefore, a set of N_{EOF} coefficients, i.e., one for each EOF mode, for each grid point for which an observation is available.

[28] Since below 200 m no significant chlorophyll trace can be detected, in equations (7) and (8) the profiles are defined for depths of 0–200 m. Horizontally, we considered only the profile $\mathbf{p}_i^j$ in grid points where the bottom is deeper than 200 m, because of the high uncertainty of remote sensing observations in coastal areas. The EOF profiles are, therefore, computed on a data set of homogeneous dimensions.

[29] The sets of profiles obtained from the EOF decomposition have been precomputed, and they account for the structure of the chlorophyll error vertical covariance in the model. The EOF decomposition has been applied to a validated multiannual OPATM-BFM run [*Lazzari et al*., 2012]. The number of EOF modes (N_{EOF}) used in the $\mathbf{V}_v$ operator has been chosen so that the first N_{EOF} modes accounts at least for the 95% of the variance.

[30] The results of the EOF decomposition show that, in most regions, the first mode alone accounts for more than 80% of the variance in winter months and for more than 60% in summer months (Figure 2). In the winter months, the chlorophyll profiles for all the subregions are characterized by a prevalent and common shape but with high variability (mean total variance nearly equal to 0.030 (mg chl/m^3)^2). In summer, at least one additional mode is necessary to account for 80% of the variance. The chlorophyll profiles are typically characterized by a deep chlorophyll maximum (DCM) in summer, and variations with respect to the mean profile are not particularly significant (mean total variance nearly equal to 0.002 (mg chl/m^3)^2 with slight variations in DCM positions and intensities). These variations are not characterized by a prevalent shape, and several EOF modes are necessary to describe the variance. Figure 2 highlights that there are few exceptions to this general picture. The Ionian subregion shows the highest number of EOF modes needed to explain at least 80% of the variance for the winter months. Indeed, the ION winter profiles are characterized by significant variability in the presence, depth and

21699291, 2014, 1, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2013JC009277 by Ogs Trieste Istituto Nazionale, Wiley Online Library on [18/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

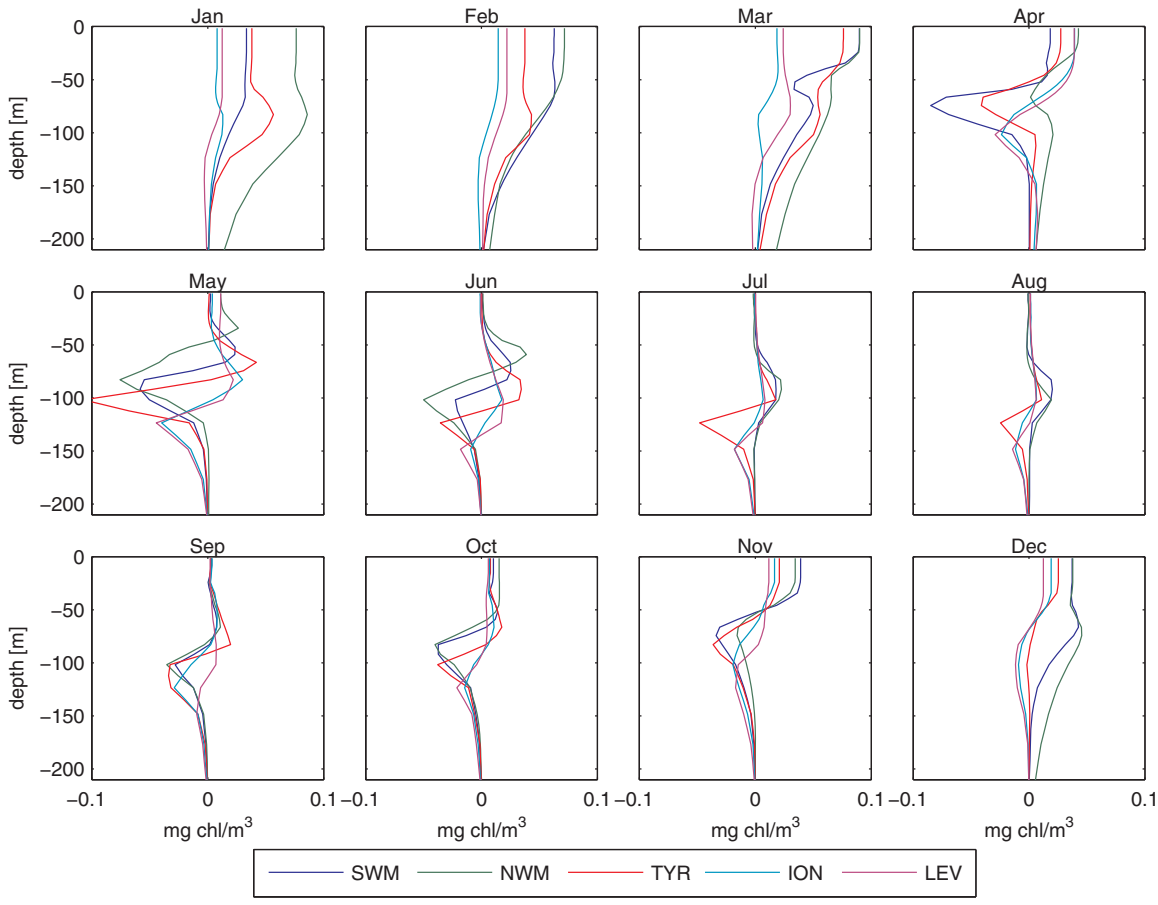

Figure 3. First EOF-mode of anomaly profiles for selected subregions.

intensity of DCM, because they are affected in a complex way by interaction with the Levantine, Adriatic, and West Mediterranean dynamics.

[31] Figure 3 shows the first mode profiles $\mathbf{e}^{r}_{m,1}$ for every month m and for selected subregions. From January to March, when most of the information about the profile variance is described by the first mode profile, the shape of which is very similar for the subregions (S-shape). The first mode profile is nearly constant from the surface to a certain depth (roughly 70 m), because of the vertical mixing which makes the water column characteristics uniform down to the MLD (mixed layer depth), while it tends asymptotically to zero below 100–150 m, where the biogeochemical signal does not vary because of low winter irradiance.

[32] For all of the first mode profiles for April, November, and December and for most of the profiles for October (Figure 3), the shape can be defined as unimodal, i.e., characterized by a unique subsurface minimum or maximum (1M-shape). During these months, the stratification condition changes, becoming established in April and breaking down in October to December. Therefore, the first EOF mode accounts for the transition from homogeneous conditions to those characterized by the presence of a DCM (or vice versa).

[33] For all the other months (late spring and summer), the shape is mainly bimodal, i.e., with a maximum and a minimum peak (2M-shape). During these months, the water column is stratified and characterized by the presence of a stable DCM, and the information contained in the first mode profiles accounts for the variations in position and intensity of the DCM.

[34] For each grid point, the EOF modes obtained in each subregion are multiplied by the ratio of the grid point surface chlorophyll variance and the subregion variance as obtained from the EOF decomposition. Through this procedure a different set of EOF modes has been obtained for each grid point. A different strategy would have been to compute the EOF modes for each model grid point. However, the statistical significance of the modes at each grid point would have been affected by the inaccuracy of the model results at the local scale and by local variability.

5.2. The Horizontal Covariance Operator $\mathbf{V_H}$

[35] The $\mathbf{V_H}$ operator smoothes the preliminary innovation and also propagates the innovation to grid points for which the satellite observations are not available or where the water column depth is less than 200 m.

[36] The horizontal operator is built using a Gaussian parameterization whose correlation radius modulates the smoothing intensity and the horizontal spatial areas influenced by the operator. Details of the numerical definition

21699291, 2014, 1, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2013JC009277 by Ogs Trieste Istituto Nazionale, Wiley Online Library on [18/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 1. Evaluation of $\mathbf{V}_H$ Effect for Different Correlation Radii[a]

Correlation Radius (km)	Relevant Innovation Introduced by V_H	Ratio Between Analysis and Satellite of the Mean Gradient Intensity
Summer		
10	21%	1.82
15	37%	1.09
20	48%	1.07
25	56%	1.06
Winter		
10	29%	1.17
15	51%	1.13
20	65%	1.16
25	75%	1.21

[a]The threshold for relevant innovation is 0.005 mg chl/m^3. The gradient is calculated on the surface chlorophyll field.

of the $\mathbf{V}_H$ operator are given by *Dobricic and Pinardi* [2008].

[37] To avoid the possibility that the application of $\mathbf{V}_H$ excessively smooth the mesoscale structures, the correlation radius applied should be in the order of the Rossby radius in the MS, which has been evaluated as approximately 10 km [*Pinardi and Masetti*, 2000]. Different correlation radii were tested and an example for a summer and a winter condition (26 August and 30 December, respectively) is presented in Table 1. Results are shown for radii of 10, 15, 20, and 25 km. A radius smaller than 10 km has not been taken into account, because it would be too small with respect to the spatial resolution of the model and the application of $\mathbf{V}_H$ would be negligible. On the other hand, a radius greater than 25 km would excessively smooth the mesoscale structures.

[38] Table 1 shows the percentages of the significant innovation (i.e., those with absolute values greater than 0.005 mg chl/m^3, which is half of the minimum chlorophyll concentration detectable by the satellite [*Volpe et al.*, 2007]), introduced by $\mathbf{V}_H$ at grid points that would have had no innovation in the case of no $\mathbf{V}_H$ application. Using the smallest radius (10 km), the presence of significant innovation extends to 21% and 29% of the grid points that would have had no innovation in the summer and winter conditions, respectively. These percentages increase to up to 56% and 75% for the largest radius (25 km). In order to evaluate the smoothing effect of $\mathbf{V}_H$, the spatial mean of the magnitude of the gradient of the surface chlorophyll concentration has been calculated. The lower the mean magnitude is, the smoother the surface chlorophyll field is. Table 1 reports the ratio between the mean magnitude of the analysis field and of the satellite field for the different correlation radii. When the ratio is equal to 1 it is expected that the application of $\mathbf{V}_H$ preserves the gradient magnitude in the analysis field with respect to the satellite field. In all of the cases, the mean gradient magnitude is always greater than 1 (the satellite field has smoother gradients than the analysis field). In the summer case, the ratio is highest for 10 km and nearly constant for all of the radii greater than 15 km (approximately equal to 1.1). In the winter case, the ratio is closest to 1 for 15 km. Based on these considerations the 15 km radius was preferred, because it is a good compromise between the two criteria.

5.3. The Biological Operator $\mathbf{V_b}$

[39] The $\mathbf{V}_b$ operator calculates the innovation for a subset of the biogeochemical model variables. First V_b calculates the chlorophyll innovation for the four phytoplankton functional types of OPATM-BFM model. Then, through the dynamical relationship of chlorophyll to carbon and nutrients content of the four phytoplankton functional types, $\mathbf{V}_b$ spreads the innovation to all of the components that constitute the phytoplankton (carbon, nitrogen, phosphorous, and silicon).

[40] Synthesis and degradation of chlorophyll depend on light availability, on nutrient uptake and on the nutrient content within the cell (details in Appendix A). In particular, the chlorophyll synthesis and the primary production depend on the internal chlorophyll-to-carbon and carbon-to-nutrient ratios (equations (A2) and (A3)), which indicate how far the phytoplankton is from the optimal growth condition (equation (A5)).

[41] Based on the assumption that the physiological status of phytoplankton is determined by the model dynamics through acclimation mechanisms during the forecast, the $\mathbf{V}_b$ operator is designed so that it keeps constant the values of the specific rate of growth, synthesis, and degradation of phytoplankton. Therefore, the $\mathbf{V}_b$ operator applies proportionally the innovation to all of the components of the phytoplankton functional types. Given that the innovation for chlorophyll obtained by the operators $\mathbf{V}_V$ and $\mathbf{V}_H$ is equal to $\delta P_l = V_H \cdot V_v \cdot v$, the innovation for carbon δP_c is determined by $\mathbf{V}_b$ as equal to:

$$\delta P_c = \frac{1}{\theta_{chl}} \delta P_l, \tag{9}$$

where θ_{chl} is the internal chlorophyll-to-carbon ratio of the phytoplankton as provided by the forecast before the assimilation. The $\mathbf{V}_b$ operator for the other phytoplankton components (nitrogen, phosphorous, and silicon) is formulated similarly using the corresponding internal ratios provided by the forecast.

[42] Furthermore, when the internal nutrient concentrations are greater than four times the optimal nutrient-to-chlorophyll ratio, a condition in which BFM considers a phytoplankton organism as very stressed and in a severe declining growth phase (negative net chlorophyll synthesis SYN_N, see Appendix A), the positive innovation prescribed by the assimilation is set equal to zero. In this way, it is avoided to apply a positive innovation which would be in strong opposition to the present phytoplankton dynamics and which would result in relevant nutrients release from phytoplankton. The analysis of results from the OPATM-BFM model shows that this condition occurs at a depth just below the photic layer, where the phytoplankton is transported by sinking, the light does not support new photosynthesis and the phytoplankton loses chlorophyll and carbon and then nutrients through lysis. In addition, limiting growth conditions rarely occur at the surface, where ultraoligotrophic conditions may severely limit the nutrient uptake process of phytoplankton which is acclimated to high irradiance conditions (very low θ_{chl} values).

21699291, 2014, 1, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2013JC009277 by Ogs Trieste Istituto Nazionale, Wiley Online Library on [18/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

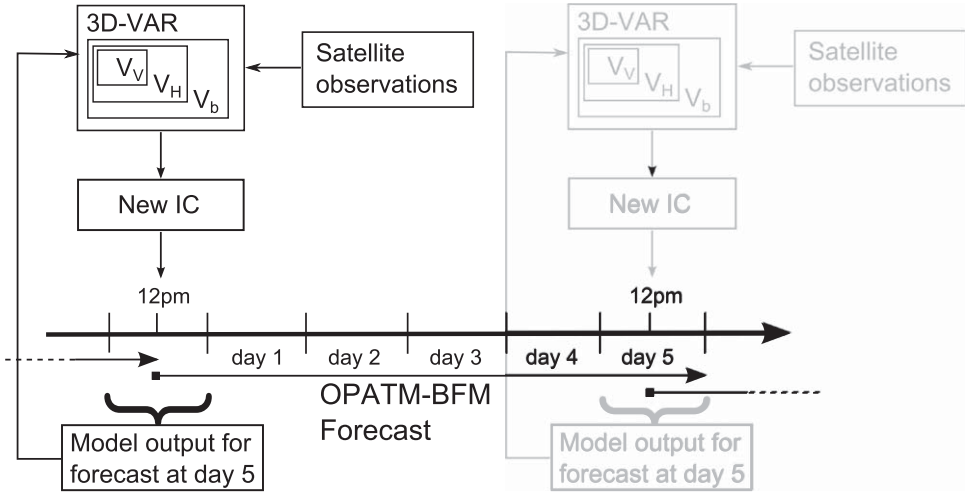

Figure 4. Flow diagram of the assimilation implementation.

6. Data Assimilation Implementation

[43] The assimilation scheme has been implemented as part of the biogeochemical short-term forecast system at OGS (part of the European MyOcean infrastructure) providing updated initial conditions every five days (Figure 4). The assimilation run (AR) consists of a sequence of five steps occurring every 5 days:

[44] (1) five daily satellite maps of chlorophyll concentration are downloaded, temporally averaged and spatially interpolated, and the observation error is provided by the variance map for the corresponding month of the assimilation date;

[45] (2) the misfit is evaluated as the difference between the satellite chlorophyll and the daily mean value of the sum of the four phytoplankton chlorophylls produced by the previous model forecast for the forecast day 5;

[46] (3) compute the 3D-VAR innovation;

[47] (4) the analysis is produced for 12 pm on the date of assimilation, and the forecast simulation for the next 5 days is prepared;

[48] (5) five days of forecast are simulated using the new initial conditions provided by the assimilation and the physical and boundary conditions described in section 2.

[49] The AR covers the period from June 2007 to October 2008, and for the same period, a control run (CR) without assimilation was carried out using the same forcings and boundary conditions. The monthly mean values of the CR outputs are available in the MyOcean catalogue (http://www.myocean.eu/web/24-catalogue.php).

[50] The forecast system is run on a parallel PLX machine at the Italian high-performance computing centre, CINECA. The OPATM-BFM model is run on 36 parallel

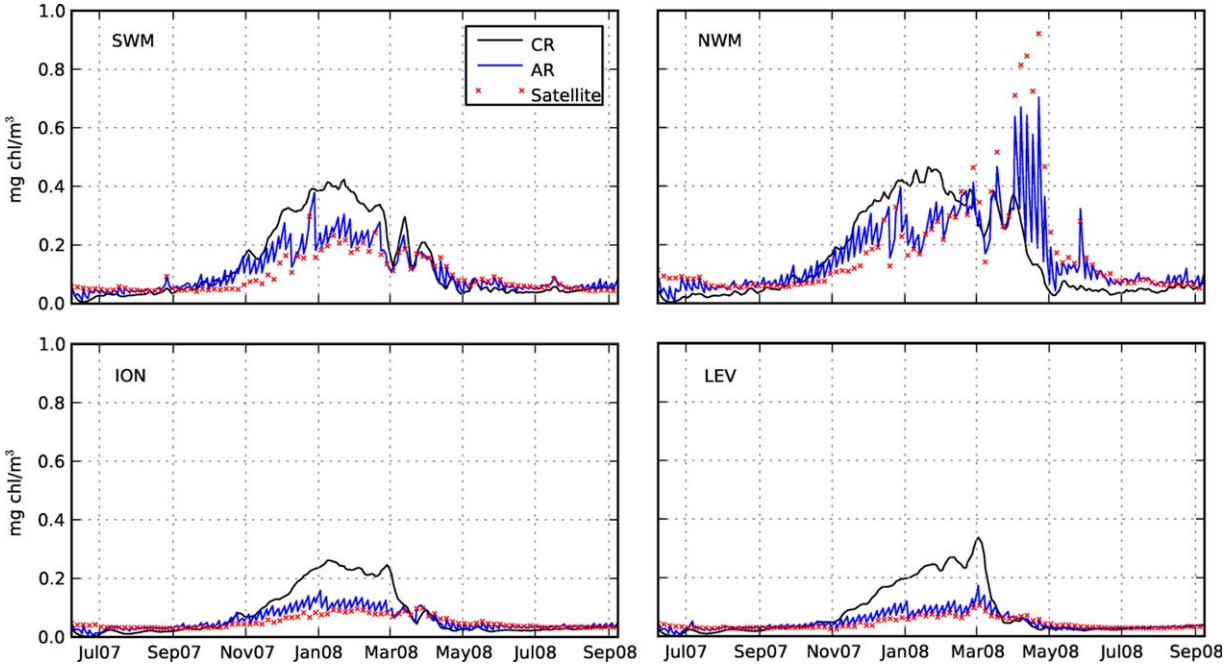

Figure 5. Time series of the mean surface chlorophyll concentrations.

21699291, 2014, 1, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2013JC009277 by Ogs Trieste Istituto Nazionale, Wiley Online Library on [18/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 2. Annual Mean of Skill Metric Statistics and Their Percentage Differences[a]

	|BIAS|		|RMS|			
Sub-region	CR	AR	CR	AR	$\frac{\lvert BIAS\rvert_{AR}-\lvert BIAS\rvert_{CR}}{\lvert BIAS\rvert_{CR}}$ %	$\frac{\lvert RMS\rvert_{AR}-\lvert RMS\rvert_{CR}}{\lvert RMS\rvert_{CR}}$ %
ALB	0.115	0.114	0.234	0.234	−1.0	−0.4
SWM	0.055	0.048	0.0753	0.0754	−15.3	−5.9
NWM	0.124	0.122	0.224	0.222	−1.0	−0.1
TYR	0.044	0.041	0.063	0.059	−12.2	−7.2
ADN	0.063	0.055	0.070	0.061	−11.8	−11.3
ADS	0.055	0.059	0.10	0.078	−28.3	−23.8
AEG	0.042	0.042	0.090	0.075	−23.6	−14.5
ION	0.028	0.028	0.054	0.043	−30.0	−21.9
LEV	0.028	0.029	0.053	0.037	−35.6	−31.5

[a]The skill metric statistics are calculated at the forecast day 5 and then the annual mean is evaluated. The bars indicate the mean values evaluated over the subregion and over the considered time period.

processors with a memory and computational requirement of 36 GB. The simulation of one day requires nearly 4 min, and 500 MB are used to store the model's output. The 3D-VAR scheme is not parallelized, and requires approximately 10 min on a single processor and 2 GB of memory.

7. Results

7.1. Effects of Assimilation on Forecast Performances

[51] The performance of the 3D-VAR scheme has been assessed by comparing the skill of the forecast of the AR and CR against the chlorophyll satellite data and in situ measurements.

[52] Figure 5 shows the time series of the mean surface chlorophyll concentration of the CR, AR results and of the satellite observations for four selected subregions. For all the subregions, the assimilation improvements are more significant from December to March, when phytoplankton blooms occur. Figure 5 shows that the effect of the assimilation is stronger in the eastern subregions, mainly because of the higher error of CR in LEV during winter. During summer, although the concentration values and errors are very low, the AR reports a general improvement with respect to the CR in the four subregions. The time series illustrated in Figure 5 show that in some cases the model trajectory rapidly departs from the conditions introduced by the assimilation. This issue occurs in a significant way in NWM in April, but it can be relevant also during winter both in the eastern and in the Western subregions. The reasons for this behavior will be discussed in section 7.2

[53] The mean absolute bias and the RMSs of the residual values (root mean squares of the differences between the model and satellite maps) have been computed for forecast day 5 (Table 2) for the areas with a water column depth greater than 200 m. The average over one year of simulation (August 2007 to July 2008) of the skill metrics and their percentages of variation are reported in Table 2. In all of the subregions, the forecast improved in the AR compared with the CR both in terms of absolute bias and the RMS of the residual values. The largest improvement was observed for LEV (with a 36% and 31% reduction in absolute bias and RMS, respectively). However, it is worth noting that the assimilation may produce less accurate results locally and for short periods of time, due to the interaction between the assimilation scheme and the model dynamics. As an example, the large bias of the bloom event in the NWM in April is only poorly corrected by the AR (Figure 5), due to the lack of nutrients necessary to sustain the bloom (as will be discussed in section 7.2). Further, Figure 5 shows that in October the forecast at day 5 of the AR performs worse than the CR forecast. Indeed, during a phase of continuous (August to September) and spatially wide positive innovation of the surface chlorophyll, the increase in the concentration of all of the phytoplankton groups in the surface layers creates a pool of nutrients within the phytoplankton cells that is then eventually released to inorganic nutrients (through the mechanisms of phytoplankton excretion and subsequent mineralization), and surplus nutrient availability stimulates local blooms during October more than is observed in the satellite maps and simulated in the CR.

[54] Since the DA acts as a reinitialization every 5 days, the AR is expected to produce forecasts that better reproduce satellite-based local chlorophyll patterns (e.g., local blooms and effects of fronts and gyres) than the CR. As an example, Figure 6 shows a sequence of maps of the Eastern Mediterranean for the period of 3–13 March 2008. For the first two dates, the AR accurately reproduces the local high chlorophyll patch east of Crete due to the fertilization effect of the Rhodes gyre and a general low concentration over the rest of the areas. In contrast, the CR overestimates the concentration and simulates a patch of high concentration north of Crete and along the Turkish coast. On the last date (13 March 2008), despite a general overestimation in the Levantine region, the high concentration of chlorophyll along the Turkish coast is reduced in the AR.

[55] The overall comparison of the patterns between satellite and simulation maps is summarized in the Taylor diagrams of Figure 7. The Taylor diagram [*Jolliff et al*., 2009] is a polar coordinate diagram, which provides a visual representation of the linear correlation coefficient R between the model and observation maps (angular coordinate) and normalized standard deviation of the model over the one of data (radial coordinate). The unbiased root mean square difference (RMSD) of the errors between model and observation maps is proportional to the distance between the model points and the reference point. The diagram shows that the AR performs better than the CR in all subregions. The average correlation increases to values higher than

21699291, 2014, 1, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2013JC009277 by Ogs Trieste Istituto Nazionale, Wiley Online Library on [18/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

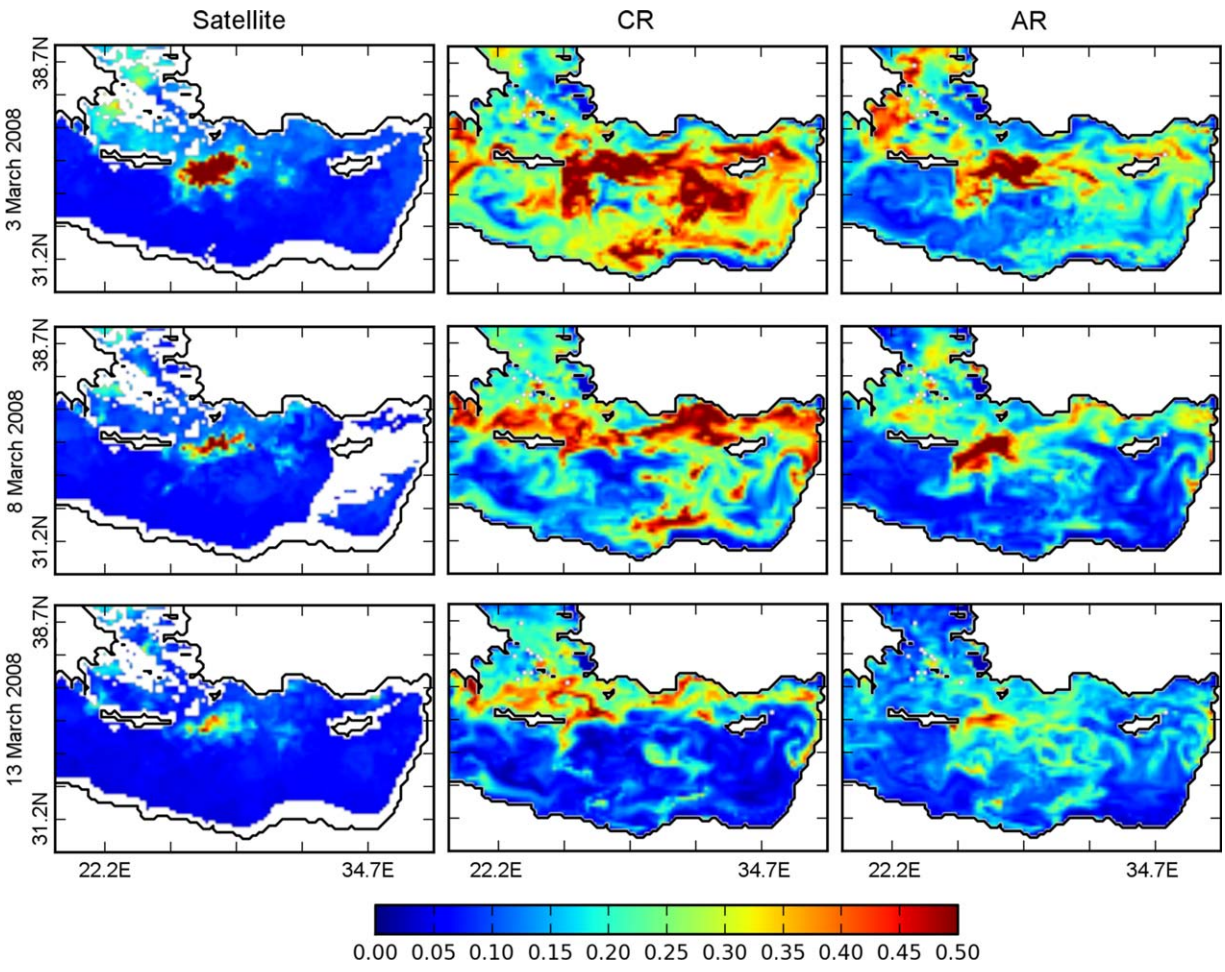

Figure 6. Sequence of maps of surface chlorophyll (mg chl/m^3) for satellite, CR, and AR (forecast day 5) for the period 3–13 March 2008.

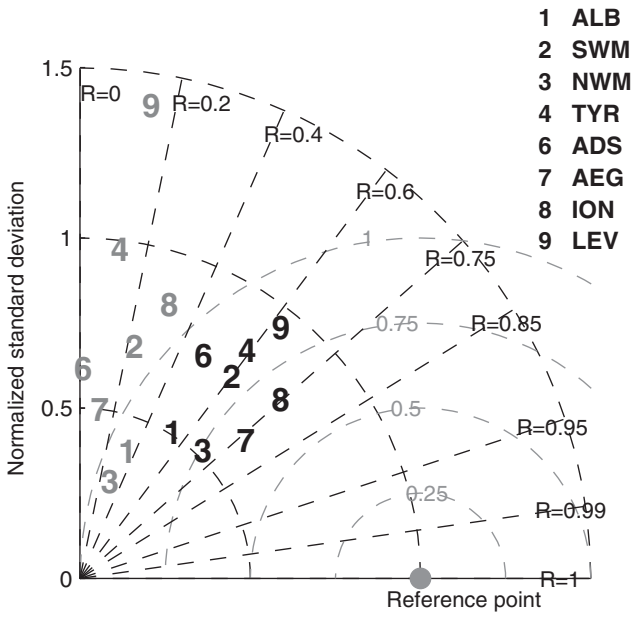

Figure 7. Taylor diagram for the 5 day mean results. The CR and the AR results are in gray and black, respectively. The dashed gray circles indicate the distance form the reference point.

0.60 for most of the subregions. The most significant improvements are those observed for the eastern subregions (in particular ION and LEV), mainly because of a reduction in the normalized standard deviation, a feature which indicates that the AR better represents the range of variability of the surface chlorophyll concentration.

[56] The skill of the model in reproducing the vertical structure of chlorophyll has been evaluated using available in situ data (BOUM 2008 cruise [*Moutin et al*., 2012]) and climatological data [*Manca et al*., 2004]. The comparison with the BOUM data set, which refers to a section from the NWM to the LEV subregions of July 2008, is shown in Figure 8a. The model correctly represents the west-east gradient of the DCM during late spring-summer 2008. The eastward deepening of the DCM is a permanent structure of the MS during spring-summer, which is well reproduced by the OPATM-BFM model [*Lazzari et al*., 2012], and the use of assimilation very slightly improves the performance of the model (Table 3). However, it is worth noting that the vertical structure of the chlorophyll field, which is strongly driven by physical processes (depth of pycnocline and mixed layer depth), is only partly affected by the assimilation.

[57] The comparison with climatological profiles for the western and the eastern subregions (Figure 8b) shows that the model accurately reproduces seasonal changes in the profiles. The assimilation has larger effects in correcting

21699291, 2014, 1, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2013JC009277 by Ogs Trieste Istituto Nazionale, Wiley Online Library on [18/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

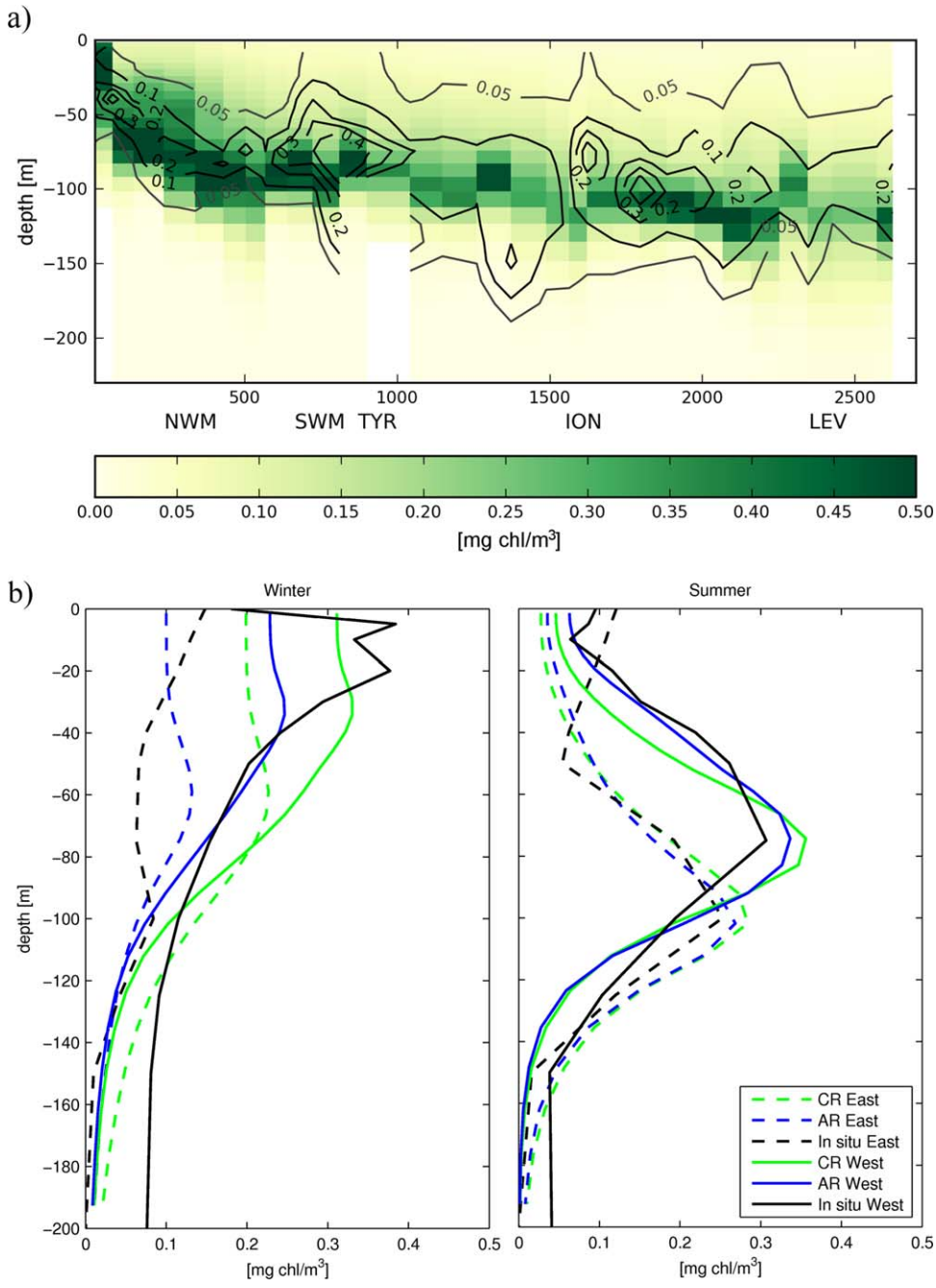

Figure 8. (a) Assimilation run (shaded areas) and in situ (contours) chlorophyll concentration along the BOUM 2008 trans-Mediterranean section from the NWM to the LEV subregions (*x* axis is the distance from the first measurement station). Model results have been interpolated into the geographical coordinates of the observation stations (Figure 1). (b) Winter and summer profiles of chlorophyll of AR (blue), CR (green), and climatological data (black) [*Manca et al*., 2004] for the western (solid lines) and eastern (dotted lines) Mediterranean Sea. The model profiles of the two regions are the averages of the corresponding subregions. The climatological data are the averages of the DS4, DF1, and DF3 regions for the western MS and of the DJ5, DJ7, DJ8, DH3, DL1, DL3, DL4 domains for the eastern MS [*Manca et al*., 2004, Figure 6].

vertical profiles during the winter season, while the effects of the assimilation are weaker in summer.

7.2. Effect of Assimilation Along the Water Column

[58] As described in section 5, through the combination of the **v** solution and the $\mathbf{V}_V$ operator the assimilation scheme provides the innovation profile of chlorophyll along the water column in each grid point where observations are available. In this section, we focus on how the assimilation propagates the information along the vertical dimension and how the biogeochemical model reacts to the new conditions. The analysis profile produced by the assimilation is

21699291, 2014, 1, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2013JC009277 by Ogs Trieste Istituto Nazionale, Wiley Online Library on [18/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 3. Statistics of Comparison Between CR, AR, and the In Situ Data[a]

	BOUM 2008	
	CR	AR
Bias	0.033	0.028
RMS	0.097	0.093
Corr	0.69	0.68

[a]The in situ data are those of the BOUM 2008 cruise. The model results have been interpolated in the observation domain.

the result of a complex combination of several factors involved in the minimization of the cost function: the shape of the most significant vertical EOF modes (S-, 1M-, or 2M-shape, see section 5.1), the intensity and sign of the misfit at the surface, and the status of the chlorophyll profile before the assimilation (i.e., vertically mixed profile or profile with the presence of a DCM).

[59] A cluster analysis of analysis profiles shows that not all of the possible combinations of these factors occur and that the innovation profiles can be grouped into four typical and most-probable behaviors, which are shown in Figure 9. The frequencies of the different cases are shown in Figure 10 for the western and eastern Mediterranean. Most of the subregions present nearly homogeneous frequencies, and only in marginal seas (ADN, ADS, ALB, and AEG) there is a patchier pattern due to the presence of smaller-scale spatial dynamics.

[60] Case A innovation describes little changes to the shape and values of the profile, which leaves the level of the DCM essentially unchanged. This case occurs mainly during the summer months, when a DCM is well established and the surface misfit is slightly positive or negative. In the eastern subregions, this condition is already present at the beginning of spring and persists during the first part of autumn. Case A innovations are characterized by EOF profiles of 1M-shape and 2M-shape with peaks at or around the DCM depth which can cause small changes in the position of the DCM although with very low variation in its intensity. This highlights that when surface values and surface misfit are low, the assimilation hardly produces significant changes in the original profile.

[61] The most intense innovations of the chlorophyll profile are produced when the surface misfit is moderately or largely negative, i.e., Cases B and C. In Case B (Figure 9), the chlorophyll profile before assimilation is characterized by the absence of the DCM. The innovation is mainly S-shaped, and thus is significant in the upper part of water column and vanishes in the deeper layers. Case B occurs mainly in the winter months and in the western subregions in autumn. Case B differs from C because the large S-shaped or 1M-shaped negative innovation occurs in a profile characterized by the presence of a DCM. The intensity of the innovation is large in the upper region and decreases along the column (and may eventually change sign near the DCM). Therefore, the innovation generally causes an intensification of the DCM as shown in Figure 9. This situation is less frequent than Case B, and occurs mainly in the western subregions during the transitional phase when the summer stratification is broken down by the mixed autumnal conditions. In this period, the innovation provided by the assimilation smooths the start of the autumn surface bloom caused by the mixing.

[62] Case D describes a large positive innovation in the surface chlorophyll, which typically occurs in winter-early spring. The frequency of Case D is quite low; it occurs patchily and more frequently in the western subregions, where satellite observation of patchy blooms are more frequent. Due to an S-shaped EOF and vertical mixed profile, the innovation is propagated beyond the upper part of the profile. A variant of Case D (not shown) occurs when the chlorophyll profile is characterized by the presence of a DCM. In this case, the innovation causes a significant increase in chlorophyll concentration at both the upper part of the profile and at the DCM depth. However, the frequency of this variant of Case D is quite low, and it occurs patchily only in late spring or at the beginning of summer.

[63] In the previous paragraphs, we showed how the $\mathbf{V}_V$ and $\mathbf{V}_b$ operators affect the vertical structure of chlorophyll and of other phytoplankton variables (carbon and nutrient contents). The new profiles provide the new concentrations which are used as the initial conditions for the next forecast run. In order to evaluate the response of the model dynamics to the new conditions, we evaluated the terms of the

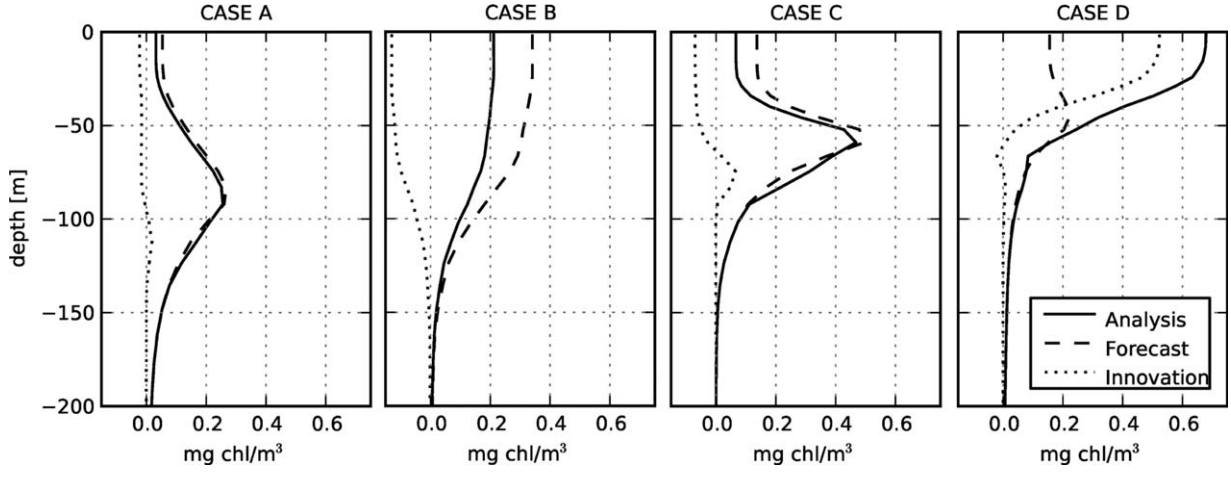

Figure 9. Chlorophyll profiles for forecast day 5, analysis, and innovation.

21699291, 2014, 1, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2013JC009277 by Ogs Trieste Istituto Nazionale, Wiley Online Library on [18/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

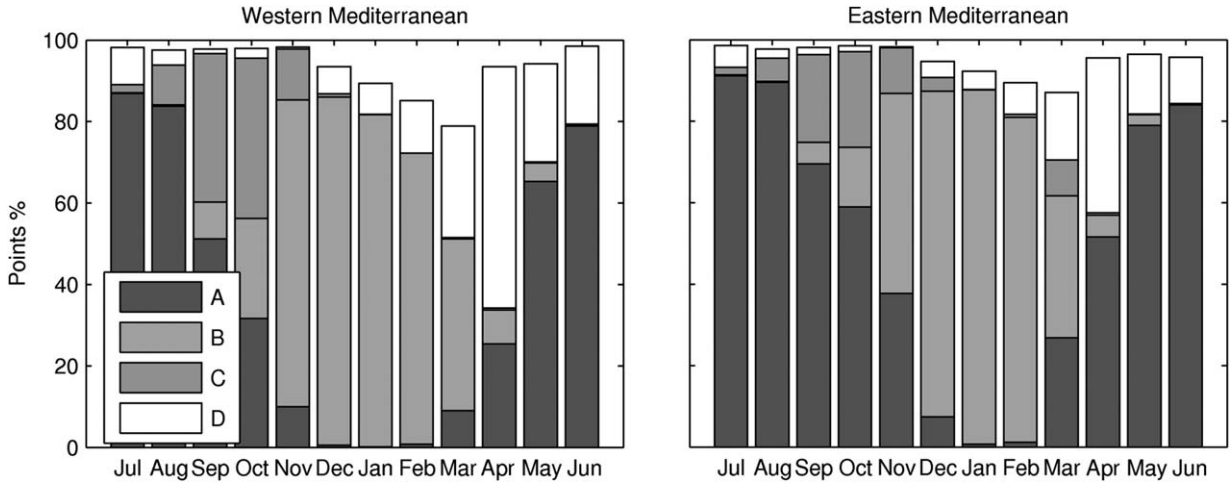

Figure 10. Frequencies of cases of Figure 9 of western and eastern Mediterranean for each month.

chlorophyll equation considering the different cases, comparing the analysis and forecast. The equations which describe the chlorophyll dynamics in the model are provided in Appendix A.

[64] In Case A, the small variations in the phytoplankton variables do not alter the terms of the equation of chlorophyll, and in the next 5 days the model trajectory is not significantly affected by the new initial conditions. In Case C, the innovation reduces the phytoplankton chlorophyll in the surface layer but slightly modifies the deepest part of the profile. In Cases B and D, the phytoplankton variables are significantly modified by the assimilation, causing changes in the model dynamics.

[65] The chlorophyll equation (equation (A1)) consists of two terms. The first term is the net synthesis term, which is regulated by the environmental conditions (light and temperature) and by the internal and external availability of nutrients. Net synthesis is the positive part of the chlorophyll equation, unless net primary production (*npp*) is negative (equation (A2)) or exudation (*exu*) is high (equation (A4)). The second part of the chlorophyll equation represents grazing by zooplankton, which is a negative contribution and is described in equations (A7) and (A8).

[66] For Case B, chlorophyll dynamics were evaluated in NWM in January, when Case B has the highest occurrence and when nutrient concentration is high while temperature and irradiance are low and light is available only in the surface layer. In this case, the innovation causes a reduction in chlorophyll (Figure 11a) and in the other phytoplankton variables up to a depth of 80 m. In both profiles (forecast and analysis), down to this depth the negative part of the chlorophyll equation is due to the grazing term, and positive net synthesis takes place near the surface to a depth of 20 m where nutrients are made available by mixing. Before the assimilation the net synthesis integrated over the water column slightly exceeds the zooplankton grazing term. Depending on the grazing formulation (equation (A8)), the grazing, which is proportional to prey abundance, decreases in the analysis profile (Figure 11a). The net synthesis term decreases as a consequence of the decrease in phytoplankton variables. However, because sugar exudation (exu_{upt}) is equal to zero (equation (A4)), the reduction is lower than the grazing reduction and integrated net synthesis over the profile exceeds grazing by more than it did before the assimilation. Thus, chlorophyll is driven to growth more rapidly than before the assimilation. The model dynamics tends to restore the conditions present before the assimilation, as a result of the vertical mixing processes which supply the surface layer with new nutrients and redistribute the new chlorophyll. The departure of the chlorophyll dynamics produced by the model from those produced by the analysis observed in Figure 5 during winter is thus attributable to mechanisms related to vertical mixing conditions.

[67] For Case D chlorophyll dynamics are evaluated in the NWM in April (when the occurrence of Case D is highest). In this case, the assimilation introduces higher values of phytoplankton biomass down to a depth of 70 m. Before the assimilation, the negative part of the synthesis equation (zooplankton grazing) is significantly higher than net synthesis (Figure 11b). Because of the increased phytoplankton biomass after the assimilation the predation term increases (equation (A8)). In order to maintain the phytoplankton biomass increment, net synthesis should have increased as much as the grazing term with an increase of net primary production (equation (A2)). However, because of the low concentration of external nutrients, the net primary production is low and phytoplankton biomass and chlorophyll are rapidly reduced due also to the effect of sugar exudation (equations (A3) and (A4)). Thus the positive correction introduced by the assimilation to reproduce the bloom observed by satellite in the NWM in April is not maintained (Figure 5). On the other hand, when external nutrients are available or supplied by recycling or transport to the surface layer, net synthesis can offset the higher grazing term and the higher values of biomass propagate forward in time (as, for instance, in May 2007 in NWM and in August 2007 and June 2008 in several subregions: see Figure 5).

8. Discussion and Conclusion

[68] In this work, the implementation of a 3-D variational scheme for the assimilation of satellite surface chlorophyll into a 3-D operational biogeochemical forecast

21699291, 2014, 1, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2013JC009277 by Ogs Trieste Istituto Nazionale, Wiley Online Library on [18/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

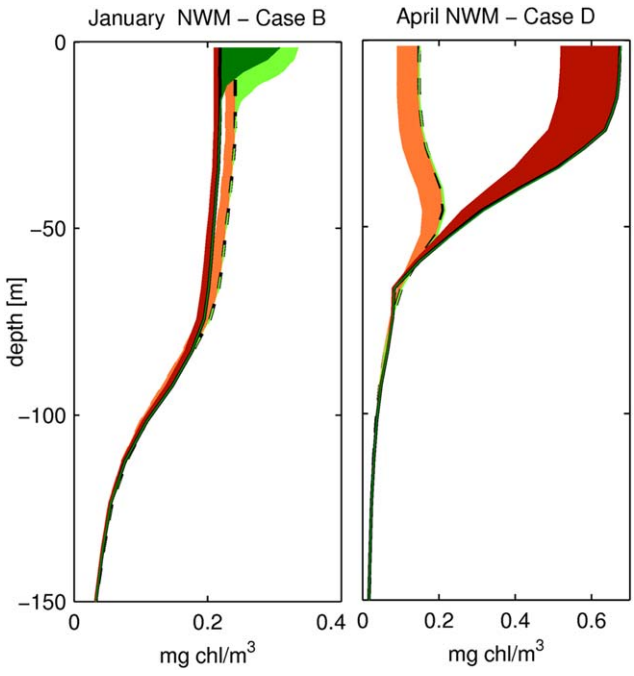

Figure 11. Chlorophyll vertical profiles (mg chl/m^3) before and after the assimilation (dashed and continuous lines, respectively). The shaded areas around each profile represent the positive (green) and negative (red) terms of the chlorophyll equation before and after the assimilation (light and dark shade, respectively).

system for the Mediterranean Sea is presented. Traditionally, Kalman-filter methods are used in biogeochemical oceanographic forecast systems, and the present work represents, to the best of our knowledge, one of the first attempts to use a 3-D variational scheme aimed at state estimation for biogeochemical assimilation.

[69] The 3D-VAR approach is based on the minimization of a cost function, and, since the chlorophyll concentration is generally not normally distributed, the distribution of errors is not Gaussian and the minimization scheme of 3D-VAR may not always ensure that the solution is optimal [*Simon and Bertino*, 2009; *Bocquet et al*., 2010]. In various assimilation applications [e.g., *Bertino et al*., 2003; *Gregg*, 2008; *Ford et al*., 2012], the chlorophyll concentration has been log-transformed in order to better satisfy the condition of Gaussianity of the error distribution. A different approach to ensuring the Gaussian distribution of errors consists in the application of the anamorphic transformation of chlorophyll concentrations, which has been applied in the Kalman-filter assimilation framework [*Lenartz et al*., 2007; *Simon and Bertino*, 2009; *Brankart et al*., 2012]. In our application, log-transformation of chlorophyll concentrations indicated that the EOF modes for log-transformed chlorophyll were less representative of the vertical covariance than those constructed using original, nonlog-transformed chlorophyll. Indeed, the application of the EOF decomposition produces statistical results (the EOF modes) that do not always represent the dynamic modes of the system [*Monahan et al*., 2009]. In addition, the asymmetry of the probability density distribution was not significantly eliminated by the log transformation. Therefore, although not optimal, the solution in the nonlog-transformed space was more representative of the vertical dynamics, preserving the typical structure of chlorophyll profiles observed in the Mediterranean Sea.

[70] Another key issue in the present assimilation scheme is the definition of observation errors. Different approaches have been applied in assimilation applications [*Tjiputra et al*., 2007; *Cossarini et al*., 2009; *Ciavatta et al*., 2011; *Ford et al*., 2012], showing that the definition of observation error is related to the specific application. In addition to the approach defined in section 4, we tested two other different definitions of observation error: a constant time and space variance equal to 0.73 (mg chl/m^3)^2, which is the square of the RMS error provided by *Volpe et al*. [2007], or proportional errors equal to 40% of the satellite estimates (mean absolute percentage difference in *Volpe et al*. [2007]). Results (not shown) indicated that the two approaches are not appropriate to describe the observation error in the whole Mediterranean basin and during the whole year, because of the very large spatial gradients and high temporal seasonal variations in the chlorophyll data. The constant error approach produced ineffective assimilation for large parts of the subregions and periods of the year, whereas the proportional approach produced very high errors in bloom periods, which made it impossible to obtain effective innovations.

[71] The 3D-VAR scheme has been developed within the framework of the pan-European MyOcean project to be implemented in the biogeochemical operational forecast system of the Mediterranean Sea. The implementation of the 3D-VAR assimilation in the forecast system of the Mediterranean Sea framework is in keeping with the requirements of a more accurate estimation of the biological state of the sea, which represents one of the main objectives indicated by the Marine Strategy Framework Directive (EU, 2008) and the European Earth Monitoring Program (GMES).

[72] The skill of the 3D-VAR application has been evaluated using both satellite observations and historical in situ data. It is worth noting that satellite data, which can be delivered routinely [*Volpe et al*., 2012], hitherto have constituted the only data set large enough for validating and constraining through assimilation the operational implementation of biogeochemical forecast systems. The forthcoming availability of profiling floats and other autonomous vehicles equipped with bio-optical and other biogeochemical sensors (Euro-Argo program) will provide additional in situ information to be used in a Mediterranean-wide operational forecast system. The increase of available in situ data in nearly real time might also lead to further development of the 3D-VAR scheme, with a positive impact on the effectiveness of the assimilation in correcting vertical profile characteristics (e.g., DCM and its gradient), as already observed for temperature and salinity profiles [*Dobricic and Pinardi*, 2008]. This may be especially useful in summer, when our results showed that assimilation is less effective in modifying vertical profiles. The forthcoming coastal algorithms for satellite chlorophyll estimates [*Mélin et al*., 2011] will also extend the scope of assimilation applications.

21699291, 2014, 1, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2013JC009277 by Ogs Trieste Istituto Nazionale, Wiley Online Library on [18/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

[73] The modular structure of the $\mathbf{V}_i$ operators in the present 3D-VAR application enables continuous improvements to the assimilation scheme, making it particularly suitable for applications in an operational biogeochemical framework. As described in section 5, each operator, which can be separately investigated and developed, reflects a functional knowledge of the biogeochemical state and dynamics. In the present application, a sequence of three **V** operators has been developed: $\mathbf{V}_V$ accounts for the vertical covariance of the chlorophyll fields, $\mathbf{V}_H$ accounts for the horizontal covariance, and $\mathbf{V}_b$ for covariance between the biogeochemical variables. The analysis of section 7.2 shows how 3D-VAR transfers the innovation along the vertical direction through the $\mathbf{V}_V$ operator. Innovation profiles were clustered in a few cases that are defined by the combination of several factors: the vertical covariance, the status of the chlorophyll profiles before the assimilation and the misfit sign and intensity. The vertical covariance was estimated by an EOF decomposition analysis which was performed using a previously validated model simulation. The 3D-VAR assimilation can significantly modify the chlorophyll profiles when the surface misfit is significant and at the same time the vertical covariance is high (i.e., in autumn and winter).

[74] Among the $\mathbf{V}_i$ operators, $\mathbf{V}_b$ can be considered the main candidate for further improvements, because of the relevant complexity and nonlinearity in the relationships between the model biogeochemical variables. The results showed that the hypothesis of the preservation of internal phytoplankton nutrient-chlorophyll and carbon-chlorophyll quotas can generate conditions of inadequacy between the physiological state of the cell and the environmental conditions, which may even lead the model to react in an opposite way with respect to the innovation introduced by assimilation. A possible improvement may be the relaxation of the constraint of the internal nutrient-chlorophyll quota preservation in the case of positive innovation, thus permitting an appropriate decrease in nutrient quotas which should be sufficient to preserve growth conditions for the phytoplankton cells and to reduce the amount of artificially generated nutrients. Additionally, the increase in nutrients in phytoplankton generated by the assimilation could be offset by the consumption of inorganic nutrients, as is done for instance in the nitrogen-balancing assimilation scheme through the conservation of the nutrient mass [*Hemmings et al*., 2008; *Ford et al*., 2012]. In this kind of approach, the assimilation scheme is based on the assumption of mass conservation, which is the principle under which biogeochemical flux models are generally built. However, since the total mass of the system (e.g., total amount of nutrient) is not a known property, the correction of the total mass through the assimilation is an assumption of equal validity, providing that the forecast of other model variables is not degraded. Similarly, in the physical oceanographic or in the atmospheric assimilation, the conservation of mass and energy [*Beuvier et al*., 2010] are not imposed. An alternative to the mass conservation assumption is a multivariate approach which also involves other functional groups of the model (e.g., nutrients and zooplankton), and may be adopted through the opportune definition of the error covariance matrix. In the 3D-VAR scheme, the adequacy between the environmental conditions and the phytoplankton status could be obtained through the development of a $\mathbf{V}_b$ operator which provides a correction with regard to environmental conditions in such a way that the innovation is maintained by the model dynamics. Because of the specific dynamics of BFM, most of the control over the evolution of the innovation depends on the nutrient profiles. Adequate nutrient profiles could prevent the disruption of the conditions introduced by the assimilation on phytoplankton profiles and improve the performance in forecasting bloom events (e.g., the bloom at the end of April only partially corrected in the AR). A further step in this direction may be the use of the biogeochemical assimilation to correct the physical parameters responsible for the nutrient mixing conditions.

Appendix A: Formulation of Chlorophyll Equation

[75] In BFM model, the parameterization of phytoplankton is composed of several constituents (carbon, nitrogen, phosphorus, silica—only for diatoms—and chlorophyll). The concentrations of the constituents evolve accordingly to dynamical equations, such as they do not have a fixed quota but vary, within permitted ranges, driven by environmental external conditions (light, temperature, and dissolved nutrient availability) and internal relationship among the constituents [*Vichi et al*., 2007; *Lazzari et al*., 2010].

[76] The evolution of chlorophyll contents of phytoplankton (P_l) is described by the equation:

$$\frac{dP_l}{dt} = SYN_N - GRZ_l, \tag{A1}$$

where SYN_N is the net chlorophyll synthesis and GRZ_l represents the losses due to the zooplankton grazing.

[77] The net chlorophyll synthesis can be either positive or negative depending on the sign of the phytoplankton net primary production (npp):

$$SYN_N = \begin{cases} npp > 0 & \rightarrow \quad f_P^{p,n}\rho_{chl} npp \\ npp \le 0 & \begin{cases} P_l > \theta^0_{chl} P_c \rightarrow & npp \cdot \left(P_l - \theta^0_{chl} P_c\right), \\ P_l < \theta^0_{chl} P_c \rightarrow & 0 \end{cases} \end{cases} \tag{A2}$$

where $f_P^{p,n}$ and ρ_{chl} account for the internal nutrients status and the light acclimatation, respectively, and they will be defined later, while θ^0_{chl} is the maximum chlorophyll-to-carbon ratio.

[78] The npp is defined as the difference among the gross primary production (gpp) and all the carbon loss terms, which are respiration (rsp) and exudation (exu):

$$npp = \left\{f_T f_E f_P^{p,n} P_c\right\}_{gpp} - \left\{b_P f_T P_c + \gamma_P f_T f_E f_P^{p,n} P_c\right\}_{rsp} - exu_{act} - exu_{upt}, \tag{A3}$$

where gpp and rsp are proportional to the carbon phytoplankton biomass (P_c), and are regulated by $f_P^{p,n}$ function of the internal conditions of phytoplankton (nutrient quotas)

21699291, 2014, 1, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2013JC009277 by Ogs Trieste Istituto Nazionale, Wiley Online Library on [18/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

and by external factors through f_E and f_T for light (with Geider formulation) and temperature (with Q_{10} formulation), respectively. In the *rsp* term b_P and γ_P are the parameters for the rest and activity respiration, respectively.

[79] The *exu* terms account for the processes of carbohydrates excretion. The part exu_{act} accounts for the activity exudation and is proportional to the *gpp*, while the exu_{upt} occurs in case of low concentration of external nutrients, when the phytoplankton cell performs photosynthesis but it is not able to build up new biomass and exudes the excess of carbon as sugar:

$$exu_{upt}=\begin{cases}(gpp-rsp-exu_{act})-\frac{1}{n_{nut}^{\min}}\cdot uptake_{nut}<0 \rightarrow & 0\\ (gpp-rsp-exu_{act})-\frac{1}{n_{nut}^{\min}}\cdot uptake_{nut}>0 \rightarrow & (gpp-rsp-exu_{act})-\frac{1}{n_{nut}^{\min}}\cdot uptake_{nut}\end{cases} \tag{A4}$$

where $\frac{1}{n_{nut}^{\min}}\cdot uptake_{nut}$ is the uptake of nitrogen or phosphorous divided by the minimum internal nitrogen or phosphorous quota, respectively. The minimum uptake between the uptake of nitrogen and phosphorous is chosen, thus considering the most limiting nutrients.

[80] In equation (A2), when *npp* is positive, the synthesis of new chlorophyll, SYN_N, is proportional to *npp* through $f_P^{p,n}$ and ρ_{chl}. In particular, $f_P^{p,n}$ is a function of the minimum between f_P^p and f_P^n, which are the regulating factor related to the internal quotas of phosphorous and nitrogen, respectively. The factors f_P^p and f_P^n have the same formulation, which depends on the optimal and minimum quotas of phosphorous and nitrogen, respectively. For instance, the regulating factor for phosphorous is written as

$$f_P^p=\begin{cases}\frac{P_n}{P_c}<n_P^{\min} & 0\\ n_P^{\min}<\frac{P_n}{P_c}<n_P^{opt} & \frac{P_n/P_c-n_P^{\min}}{n_P^{opt}-n_P^{\min}},\\ \frac{P_n}{P_c}>n_P^{opt} & 1\end{cases} \tag{A5}$$

where P_n is the phytoplankton nitrogen concentration and $n_P^{\min}$ and n_P^{opt} are the minimum and optimal quotas of phosphorous, respectively.

[81] The acclimatation to the light conditions is accounted for by the function ρ_{chl}, which is the ratio between the realized photosynthetic rate pho_R and the maximum potential photosynthesis pho_M multiplied by the maximum chlorophyll-to-carbon ratio θ_{chl}^0:

$$\rho_{chl}=\theta_{chl}^0\frac{pho_R}{pho_M}. \tag{A6}$$

[82] In case of negative *npp*, the net chlorophyll synthesis is equal to 0 when the internal chlorophyll-to-carbon ratio is below the maximum quota. On the other hand, when the chlorophyll-to-carbon ratio exceeds the maximum θ_{chl}^0, the negative synthesis produces a relaxation toward θ_{chl}^0.

[83] The grazing term GRZ_l in equation (A1) is proportional to the grazing on carbon phytoplankton GRZ_c through the chlorophyll-to-carbon ratio:

$$GRZ_l=\frac{P_l}{P_c}GRZ_c. \tag{A7}$$

[84] The grazing GRZ_c is a function of phytoplankton and zooplankton biomass (Z_c) and is regulated by temperature, f^T, a food availability matrix, $\delta_{Zoo,Pc}$, and a capture efficiency parameter, $e_{Zoo,Pc}$:

$$GRZ_c=f^T\cdot r\cdot\frac{\delta_{Zoo,Pc}\cdot e_{Zoo,Pc}\cdot P_c}{totalfood}\cdot\frac{totalfood}{totalfood+h_z}\cdot Z_c. \tag{A8}$$

[85] Since the zooplankton groups can feed on multiple preys, for each zooplankton group the grazing is proportional to the biomass of the phytoplankton and the sum of the biomass of all the available preys (*totalfood*). Details on the formulation and parameters are in *Vichi et al*. [2007] and *Lazzari et al*. [2010].

[86] **Acknowledgments.** This study has been conducted using MyOcean Products. This study was supported by the FP7 EU projects MyOcean2 (grant agreement 283367) and OPEC (grant agreement 283291). Computational efforts were performed at the CINECA high performance computing facility (grants: HP10C1IXMA, 2012; HP10AC2U33, 2013), we thank the CINECA staff for their support. We thank our colleagues from OGS (G. Bolzon, P. Lazzari, and S. Salon) for useful discussions and fruitful technical support. We also thank the reviewers for their many useful comments.

References

Anderson, L. A., A. R. Robinson, and C. J. Lozano (2000), Physical and biological modeling in the Gulf Stream region. I: Data assimilation methodology, *Deep Sea Res., Part I*, *47*(10), 1787–1827, doi:10.1016/S0967-0637(00)00019-4.

Antoine, D., A. Morel, and J.-M. Andre (1995), Algal pigment distribution and primary production in the eastern Mediterranean as derived from coastal zone color scanner observations, *J. Geophys. Res.*, *100*(C8), 16,193–16,209.

Bennett, A. F., B. S. Chua, B. L. Pflaum, M. Erwig, Z. Fu, R. D. Loft, and J. C. Muccino (2008), The inverse ocean modeling system. Part I: Implementation, *J. Atmos. Oceanic Technol.*, *25*(9), 1608–1622, doi:10.1175/2008JTECHO519.1.

Bertino, L., G. Evensen, and H. Wackernagel (2003), Sequential data assimilation techniques in oceanography, *Int. Stat. Rev.*, *71*(2), 223–241, doi:10.1111/j.1751-5823.2003.tb00194.x.

Beuvier, J., F. Sevault, M. Herrmann, H. Kontoyiannis, W. Ludwig, M. Rixen, E. Stanev, K. Béranger, and S. Somot (2010), Modeling the Mediterranean Sea interannual variability during 1961–2000: Focus on the Eastern Mediterranean Transient, *J. Geophys. Res.*, *115*, C08017, doi:10.1029/2009JC005950.

Bocquet, M., C. A. Pires, and L. Wu (2010), Beyond Gaussian statistical modeling in geophysical data assimilation, *Mon. Weather Rev.*, *138*(8), 2997–3023, doi:10.1175/2010MWR3164.1.

Brankart, J.-M., C.-E. Testut, D. Béal, M. Doron, C. Fontana, M. Meinvielle, P. Brasseur, and J. Verron (2012), Towards an improved description of ocean uncertainties: Effect of local anamorphic

21699291, 2014, 1, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2013JC009277 by Ogs Trieste Istituto Nazionale, Wiley Online Library on [18/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

transformations on spatial correlations, *Ocean Sci.*, *8*(2), 121–142, doi:10.5194/os-8-121-2012.

Byrd, R. H., P. Lu, J. Nocedal, and C. Zhu (1995), A limited memory algorithm for bound constrained optimization, *SIAM J. Sci. Comput.*, *16*(5), 1190–1208, doi:10.1137/0916069.

Carmillet, V., J.-M. Brankart, P. Brasseur, H. Drange, G. Evensen, and J. Verron (2001), A singular evolutive extended Kalman filter to assimilate ocean color data in a coupled physical-biochemical model of the North Atlantic ocean, *Ocean Modell.*, *3*(3–4), 167–192, doi:10.1016/S1463-5003(01)00007-5.

Ciavatta, S., R. Torres, S. Saux-Picart, and J. I. Allen (2011), Can ocean color assimilation improve biogeochemical hindcasts in shelf seas?, *J. Geophys. Res.*, *116*, C12043, doi:10.1029/2011JC007219.

Crise, A., G. Crispi, and E. Mauri (1998), A seasonal three-dimensional study of the nitrogen cycle in the Mediterranean Sea: Part I. Model implementation and numerical results, *J. Mar. Syst.*, *18*(1-3), 287–312, doi:10.1016/S0924-7963(98)00016-5.

Crispi, G., R. Mosetti, C. Solidoro, and A. Crise (2001), Nutrients cycling in Mediterranean basins: The role of the biological pump in the trophic regime, *Ecol. Model.*, *138*(1-3), 101–114, doi:10.1016/S0304-3800(00)00396-3.

Crispi, G., A. Crise, and C. Solidoro (2002), Coupled Mediterranean ecomodel of the phosphorus and nitrogen cycles, *J. Mar. Syst.*, *33-34*, 497–521, doi:10.1016/S0924-7963(02)00073-8.

Cossarini, G., P. F. J. Lermusiaux, and C. Solidoro (2009), Lagoon of Venice ecosystem: Seasonal dynamics and environmental guidance with uncertainty analyses and error subspace data assimilation, *J. Geophys. Res.*, *114*, C06026, doi:10.1029/2008JC005080.

D'Ortenzio, F., and M. Ribera d'Alcalà (2009), On the trophic regimes of the Mediterranean Sea: A satellite analysis, *Biogeosciences*, *6*(2), 139–148, doi:10.5194/bg-6-139-2009.

Davis, R. E. (1976), Predictability of sea surface temperature and sea level pressure anomalies over the North Pacific Ocean, *J. Phys. Oceanogr.*, *6*(3), 249–266, doi:10.1175/1520-0485(1976)006<0249:POSSTA>2.0.CO;2.

Dobricic, S., and N. Pinardi (2008), An oceanographic three-dimensional variational data assimilation scheme, *Ocean Modell.*, *22*(3–4), 89–105, doi:10.1016/j.ocemod.2008.01.004.

Fontana, C., C. Grenz, and C. Pinazo (2010), Sequential assimilation of a year-long time-series of SeaWiFS chlorophyll data into a 3D biogeochemical model on the French Mediterranean coast, *Cont. Shelf Res.*, *30*(16), 1761–1771, doi:10.1016/j.csr.2010.08.003.

Ford, D. A., K. P. Edwards, D. Lea, R. M. Barciela, M. J. Martin, and J. Demaria (2012), Assimilating GlobColour ocean colour data into a pre-operational physical-biogeochemical model, *Ocean Sci.*, *8*(5), 751–771, doi:10.5194/os-8-751-2012.

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*(1–3), 289–319, doi:10.1016/S0967-0645(01)00104-7.

Garcia-Gorriz, E., N. Hoepffner, and M. Ouberdous (2003), Assimilation of SeaWiFS data in a coupled physical–biological model of the Adriatic Sea, *J. Mar. Syst.*, *40–41*, 233–252, doi:10.1016/S0924-7963(03)00020-4.

Giering, R., and T. Kaminski (1998), Recipes for adjoint code construction, *ACM Trans. Math. Software*, *24*(4), 437–474, doi:10.1145/293686.293695.

Gregg, W. W. (2008), Assimilation of SeaWiFS ocean chlorophyll data into a three-dimensional global ocean model, *J. Mar. Syst.*, *69*(3–4), 205–225, doi:10.1016/j.jmarsys.2006.02.015.

Gregg, W. W., M. A. M. Friedrichs, A. R. Robinson, K. A. Rose, R. Schlitzer, K. R. Thompson, and S. C. Doney (2009), Skill assessment in ocean biological data assimilation, *J. Mar. Syst.*, *76*(1–2), 16–33, doi:10.1016/j.jmarsys.2008.05.006.

Hemmings, J. C. P., R. M. Barciela, and M. J. Bell (2008), Ocean color data assimilation with material conservation for improving model estimates of air-sea CO_2 flux, *J. Mar. Res.*, *66*(1), 87–126, doi:10.1357/002224008784815739.

Jolliff, J. K., J. C. Kindle, I. Shulman, B. Penta, M. A. M. Friedrichs, R. Helber, and R. A. Arnone (2009), Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment, *J. Mar. Syst.*, *76*(1–2), 64–82, doi:10.1016/j.jmarsys.2008.05.014.

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.

Korres, G., G. Triantafyllou, G. Petihakis, D. E. Raitsos, I. Hoteit, A. Pollani, S. Colella, and K. Tsiaras (2012), A data assimilation tool for the Pagasitikos Gulf ecosystem dynamics: Methods and benefits, *J. Mar. Syst.*, *94*, S102–S117, doi:10.1016/j.jmarsys.2011.11.004.

Lawson, L. M., E. E. Hofmann, and Y. H. Spitz (1996), Time series sampling and data assimilation in a simple marine ecosystem model, *Deep Sea Res., Part II*, *43*(2–3), 625–651, doi:10.1016/0967-0645(95)00096-8.

Lazzari, P., A. Teruzzi, S. Salon, S. Campagna, C. Calonaci, S. Colella, M. Tonani, and A. Crise (2010), Pre-operational short-term forecasts for Mediterranean Sea biogeochemistry, *Ocean Sci.*, *6*(1), 25–39, doi:10.5194/os-6-25-2010.

Lazzari, P., C. Solidoro, V. Ibello, S. Salon, A. Teruzzi, K. Béranger, S. Colella, and A. Crise (2012), Seasonal and inter-annual variability of plankton chlorophyll and primary production in the Mediterranean Sea: A modelling approach, *Biogeosciences*, *9*(1), 217–233, doi:10.5194/bg-9-217-2012.

Lenartz, F., C. Raick, K. Soetaert, and M. Grégoire (2007), Application of an Ensemble Kalman filter to a 1-D coupled hydrodynamic-ecosystem model of the Ligurian Sea, *J. Mar. Syst.*, *68*(3–4), 327–348, doi:10.1016/j.jmarsys.2006.12.001.

Lewis, K., and J. I. Allen (2009), Validation of a hydrodynamic-ecosystem model simulation with time-series data collected in the western English Channel, *J. Mar. Syst.*, *77*(3), 296–311, doi:10.1016/j.jmarsys.2007.12.013.

Lorenc, A. C. (1986), Analysis methods for numerical weather prediction, *Q. J. R. Meteorol. Soc.*, *112*(474), 1177–1194, doi:10.1002/qj.49711247414.

Ludwig, W., E. Dumont, M. Meybeck, and S. Heussner (2009), River discharges of water and nutrients to the Mediterranean and Black Sea: Major drivers for ecosystem changes during past and future decades?, *Prog. Oceanogr.*, *80*(3–4), 199–217, doi:10.1016/j.pocean.2009.02.001.

Manca, B., M. Burca, A. Giorgetti, C. Coatanoan, M.-J. Garcia, and A. Iona (2004), Physical and biochemical averaged vertical profiles in the Mediterranean regions: An important tool to trace the climatology of water masses and to validate incoming data from operational oceanography, *J. Mar. Syst.*, *48*(1–4), 83–116, doi:10.1016/j.jmarsys.2003.11.025.

Mattern, J. P., M. Dowd, and K. Fennel (2013), Particle filter-based data assimilation for a three-dimensional biological ocean model and satellite observations, *J. Geophys. Res. Oceans*, *118*, 2746–2760, doi:10.1002/jgrc.20213.

Mélin, F., V. Vantrepotte, M. Clerici, D. D'Alimonte, G. Zibordi, J.-F. Berthon, and E. Canuti (2011), Multi-sensor satellite time series of optical properties and chlorophyll-a concentration in the Adriatic Sea, *Prog. Oceanogr.*, *91*(3), 229–244, doi:10.1016/j.pocean.2010.12.001.

Monahan, A. H., J. C. Fyfe, M. H. P. Ambaum, D. B. Stephenson, and G. R. North (2009), Empirical orthogonal functions: The medium is the message, *J. Clim.*, *22*(24), 6501–6514, doi:10.1175/2009JCLI3062.1.

Morel, A., and J.-M. André (1991), Pigment distribution and primary production in the western Mediterranean as derived and modeled from coastal zone color scanner observations, *J. Geophys. Res.*, *96*(C7), 12,685–12,698.

Moutin, T., F. Van Wambeke, and L. Prieur (2012), Introduction to the biogeochemistry from the oligotrophic to the ultraoligotrophic Mediterranean (BOUM) experiment, *Biogeosciences*, *9*(10), 3817–3825, doi:10.5194/bg-9-3817-2012.

Natvik, L.-J., and G. Evensen (2003), Assimilation of ocean colour data into a biochemical model of the North Atlantic. Part 1: Data assimilation experiments, *J. Mar. Syst.*, *40–41*, 127–153, doi:10.1016/S0924-7963(03)00016-2.

Nerger, L., and W. W. Gregg (2008), Improving assimilation of SeaWiFS data by the application of bias correction with a local SEIK filter, *J. Mar. Syst.*, *73*(1–2), 87–102, doi:10.1016/j.jmarsys.2007.09.007.

Oddo, P., M. Adani, N. Pinardi, C. Fratianni, M. Tonani, and D. Pettenuzzo (2009), A nested Atlantic-Mediterranean Sea general circulation model for operational forecasting, *Ocean Sci.*, *5*(4), 461–473, doi:10.5194/os-5-461-2009.

Odermatt, D., A. Gitelson, V. E. Brando, and M. Schaepman (2012), Review of constituent retrieval in optically deep and complex waters from satellite imagery, *Remote Sens. Environ.*, *118*, 116–126, doi:10.1016/j.rse.2011.11.013.

Pinardi, N., and E. Masetti (2000), Variability of the large scale general circulation of the Mediterranean Sea from observations and modelling: A

21699291, 2014, 1, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2013JC009277 by Ogs Trieste Istituto Nazionale, Wiley Online Library on [18/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

review, *Palaeogeogr. Palaeoclimatol. Palaeoecol.*, *158*(3–4), 153–173, doi:10.1016/S0031-0182(00)00048-1.

Pires, C., R. Vautard, and O. Talagrand (1996), On extending the limits of variational assimilation in nonlinear chaotic systems, *Tellus, Ser. A*, *48*(1), 96–121, doi:10.1034/j.1600-0870.1996.00006.x.

Popova, E., C. Lozano, M. Srokosz, M. J. Fasham, P. Haley, and A. Robinson (2002), Coupled 3D physical and biological modelling of the mesoscale variability observed in North-East Atlantic in spring 1997: Biological processes, *Deep Sea Res., Part I*, *49*(10), 1741–1768, doi: 10.1016/S0967-0637(02)00091-2.

Rabier, F. (2005), Overview of global data assimilation developments in numerical weather-prediction centres, *Q. J. R. Meteorol. Soc.*, *131*(613), 3215–3233, doi:10.1256/qj.05.129.

Ribera d'Alcalà, M., G. Civitarese, F. Conversano, and R. Lavezza (2003), Nutrient ratios and fluxes hint at overlooked processes in the Mediterranean Sea, *J. Geophys. Res.*, *108*(C9), 8106, doi:10.1029/2002JC001650.

Saba, V. S., et al. (2011), An evaluation of ocean color model estimates of marine primary productivity in coastal and pelagic regions across the globe, *Biogeosciences*, *8*(2), 489–503, doi:10.5194/bg-8-489-2011.

Simon, E., and L. Bertino (2009), Application of the Gaussian anamorphosis to assimilation in a 3-D coupled physical-ecosystem model of the North Atlantic with the EnKF: A twin experiment, *Ocean Sci.*, *5*(4), 495–510, doi:10.5194/os-5-495-2009.

Spitz, Y. H., J. R. Moisan, M. R. Abbott, and J. G. Richman (1998), Data assimilation and a pelagic ecosystem model: Parameterization using time series observations, *J. Mar. Syst.*, *16*(1–2), 51–68, doi:10.1016/S0924-7963(97)00099-7.

Teruzzi, A., S. Salon, G. Bolzon, P. Lazzari, S. Campagna, F. Ficarelli, C. Solidoro, and A. Crise (2011), Operational forecasts of the biogeochemical state of Mediterranean Sea, *Mercator Ocean Quarterly Newsletter*, *40*, 15–25.

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.

Triantafyllou, G., G. Korres, I. Hoteit, G. Petihakis, and A. C. Banks (2007), Assimilation of ocean colour data into a Biogeochemical Flux Model of the Eastern Mediterranean Sea, *Ocean Sci.*, *3*(3), 397–410, doi:10.5194/os-3-397-2007.

Vichi, M., N. Pinardi, and S. Masina (2007), A generalized model of pelagic biogeochemistry for the global ocean ecosystem. Part I: Theory, *J. Mar. Syst.*, *64*(1–4), 89–109, doi:10.1016/j.jmarsys.2006.03.006.

Volpe, G., R. Santoleri, V. Vellucci, M. Ribera d'Alcalà, S. Marullo, and F. D'Ortenzio (2007), The colour of the Mediterranean Sea: Global versus regional bio-optical algorithms evaluation and implication for satellite chlorophyll estimates, *Remote Sens. Environ.*, *107*(4), 625–638, doi: 10.1016/j.rse.2006.10.017.

Volpe, G., S. Colella, V. Forneris, C. Tronconi, and R. Santoleri (2012), The Mediterranean ocean colour observing system: System development and product validation, *Ocean Sci. Discuss.*, *9*(2), 1349–1385, doi: 10.5194/osd-9-1349-2012.

Ward, B. A., M. A. M. Friedrichs, T. R. Anderson, and A. Oschlies (2010), Parameter optimisation techniques and the problem of underdetermination in marine biogeochemical models, *J. Mar. Syst.*, *81*(1–2), 34–43, doi:10.1016/j.jmarsys.2009.12.005.

Weaver, A. T., J. Vialard, and D. L. T. Anderson (2003), Three- and four-dimensional variational assimilation with a general circulation model of the tropical Pacific Ocean. Part I: Formulation, internal diagnostics, and consistency checks, *Mon. Weather Rev.*, *131*(7), 1360–1378.

21699291, 2014, 1, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2013JC009277 by Ogs Trieste Istituto Nazionale, Wiley Online Library on [18/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License