Assimilation of remotely-sensed optical properties to improve marine biogeochemistry modelling

CrossMark

Stefano Ciavatta *, Ricardo Torres, Victor Martinez-Vicente, Timothy Smyth, Giorgio Dall'Olmo, Luca Polimene, J. Icarus Allen

Plymouth Marine Laboratory, Prospect Place, West Hoe, Plymouth PL1 3DH, United Kingdom

ARTICLE INFO

ABSTRACT

In this paper we evaluate whether the assimilation of remotely-sensed optical data into a marine ecosystem model improves the simulation of biogeochemistry in a shelf sea. A localized Ensemble Kalman filter was used to assimilate weekly diffuse light attenuation coefficient data, $K_d(443)$ from SeaWiFs, into an ecosystem model of the western English Channel. The spatial distributions of (unassimilated) surface chlorophyll from satellite, and a multivariate time series of eighteen biogeochemical and optical variables measured *in situ* at one long-term monitoring site were used to evaluate the system performance for the year 2006. Assimilation reduced the root mean square error and improved the correlation with the assimilated $K_d(443)$ observations, for both the analysis and, to a lesser extent, the forecast estimates, when compared to the reference model simulation. Improvements in the simulation of (unassimilated) ocean colour chlorophyll were less evident, and in some parts of the Channel the simulation of this data deteriorated. The estimation errors for the (unassimilated) *in situ* data were reduced for most variables with some exceptions, e.g. dissolved nitrogen. Importantly, the assimilation adjusted the balance of ecosystem processes by shifting the simulated food web towards the microbial loop, thus improving the estimation of some properties, e.g. total particulate carbon. Assimilation of $K_d(443)$ outperformed a comparative chlorophyll assimilation experiment, in both the estimation of ocean colour data and in the simulation of independent *in situ* data. These results are related to relatively low error in $K_d(443)$ data, and because it is a bulk optical property of marine ecosystems. Assimilation of remotely-sensed optical properties is a promising approach to improve the simulation of biogeochemical and optical variables that are relevant for ecosystem functioning and climate change studies.

© 2014 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/3.0/).

Introduction

Assimilation of satellite ocean colour products into marine ecosystem models is a promising approach towards improving the simulation of biogeochemical variables that are relevant for management and climate change studies (see, e.g., the review by Gregg et al., 2009).

Data assimilation merges models and observations, with the goal of improving the estimates obtained separately from simulation and monitoring efforts. For example, assimilation of satellite chlorophyll improved the simulation of nutrients, plankton types and total particulate nitrogen and carbon in a shelf-sea (Ciavatta et al., 2011); surface nitrate in the North Atlantic ocean (Fontana et al., 2013); and primary production, nitrogen and zooplankton in global ocean models (Gregg, 2008; Ford et al., 2012). Ocean colour assimilation can also have negligible or deleterious effects on

the estimates of unassimilated data, as a consequence of limitations in the assimilation scheme and its tuning, deficiencies in the model, and errors in the assimilated data. For example, Triantafyllou et al. (2007) showed the high sensitivity of chlorophyll forecasts to the inflation parameter which partially accounts for the model error in the Singular Evolutive Extended Kalman (SEEK) scheme. Ciavatta et al. (2011) argued that a defective parameterization of the phytoplankton functional types (PFTs) may limit assimilation improvements of the PFT simulation in a shelf-sea. Gregg (2008) showed that the bias of satellite chlorophyll data induced by other dissolved optical compounds (e.g. coloured dissolved organic matter) influenced the assimilation performance in regions close to large riverine discharges in the global ocean.

Modelling optical properties of seawater may be useful for improving the simulation of biogeochemical variables in oceanic and coastal areas (see e.g. Fujii et al., 2007). Bio-optical modelling exploits the causal links between optical properties and key biogeochemical variables and processes, enhancing the simulation

realism for marine ecosystems (Fujii et al., 2007). In addition, satellite optical properties can be retrieved with lower errors than chlorophyll, particularly in shelf-seas which are often affected by terrestrial inputs of coloured matter (see, e.g., Lee et al., 2005b; Saba et al., 2011; Zhao et al., 2013).

Ecosystem models that include the description of water optical properties have been proposed and successfully applied in marine studies (e.g. Gregg and Walsh, 1992; Bissett et al., 1999, 2005; Manizza et al., 2005; Fujii et al., 2007; Mouw et al., 2012). However, the assimilation of satellite-derived optical data into ecosystem models is still in its infancy. Shulman et al. (2013) demonstrated that the assimilation of phytoplankton absorption can improve the short-term predictions of chlorophyll and PFT ratios, in a five-day assimilative simulation in the Monterey Bay. However, to the authors' knowledge, the benefits of assimilating satellite-derived optical data for long-term biogeochemical simulations have yet to be tested.

The objective of this work is to assess the skill of assimilating remotely-sensed diffuse attenuation coefficient for the yearly simulation of biogeochemical variables in a shelf-sea.

The study area is the western English Channel, a shelf-sea region that includes both case I and case II waters (Groom et al., 2009). Diffuse attenuation coefficient in the blue band, $K_d(443)$, measured by the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) was assimilated into a coupled optical–ecosystem model of the western English Channel. The model couples the biogeochemical European Regional Seas Ecosystem Model (ERSEM, Blackford et al., 2004, as applied to the study region by Lewis and Allen, 2009), with a bio-optical module, based on Lee et al. (2005a). The assimilation scheme is the Ensemble Kalman filter (EnKF) (Evensen, 1994, 2003), which was implemented for the assimilation of chlorophyll in the western English Channel model by Ciavatta et al. (2011) and has been adapted to the assimilation of the diffuse attenuation coefficient in the present work. The outputs of the reference model run (i.e. a run without assimilation) and the output of assimilation for year 2006 were compared quantitatively with unassimilated chlorophyll distributions from SeaWiFS, and with a multivariate time series of eighteen optical and biogeochemical variables measured *in situ*, weekly, at one sampling site of the English Channel. We also ran a supplementary assimilation experiment with SeaWiFS chlorophyll data, based on the assimilation system by Ciavatta et al. (2011), to provide a comparison of the simulation skills of assimilating the two different ocean colour products. It is worth mentioning that the biogeochemical model ERSEM has been used in several previous assimilative works, including the use of the Ensemble Kalman filter (Allen et al., 2003; Torres et al., 2006; Ciavatta et al., 2011), as well as the Singular Evolutive Interpolated Kalman filter (SEIK, Triantafyllou et al., 2003, 2013; Korres et al., 2012), the Singular Evolutive Extended Kalman filter (SEEK, Hoteit et al., 2003, 2004; Triantafyllou et al., 2005; Korres et al., 2012) and its fixed basis and partially-local variants (Hoteit et al., 2005). However, the present work is the first application of ERSEM coupled with a bio-optical module for the assimilation of optical data.

The paper is structured as follows. In Section 'Material and methods', we describe the western English Channel model, the set-up of the EnKF assimilation scheme, the data, and the metrics applied for skill assessment. In Section 'Results and discussion', the results are presented and discussed. The simulation of remotely sensed $K_d(443)$ and chlorophyll data is first assessed in space by investigating the spatial distributions of the skill metrics (Section 'Skill in simulating $K_d(443)$ and chlorophyll spatial distributions'), and then assessed in time by comparing the temporal evolution of the satellite data and model output at three representative sites (Section 'Skill in simulating the satellite data evolution in representative areas'). Subsequently, the simulation of the independent *in situ* data is evaluated (Sections 'Simulation of the

optical and biogeochemical data at L4'), and the impact of data assimilation on the simulated biogeochemical processes is discussed (Sections 'Assimilation impact on ecosystem process modelling'). A comparison between the performance of $K_d(443)$ and chlorophyll assimilation is presented in Section 'Comparing $K_d(443)$ and chlorophyll assimilation skill'. Concluding remarks are pointed out in Section 'Summary and conclusions'.

Material and methods

The western English Channel model

The coupled hydrodynamic–biogeochemical model

The coupled model (Fig. 1) (Lewis and Allen, 2009) describes the hydrodynamics of the western English Channel using the Proudman Oceanographic Laboratory Coastal Ocean Modelling System (POLCOMS) (Holt and James, 2001), which is a three-dimensional, baroclinic, finite difference primitive equation model formulated on an Arakawa B-grid. Temperature and salinity are prognostic variables. The hydrodynamic module includes a vertical–turbulence model and calculations of horizontal pressure gradients (Holt and James, 2001).

The biogeochemical dynamics are described by the European Regional Seas Ecosystem model (ERSEM) (Baretta et al., 1995), and we refer to Blackford et al. (2004) for an extensive description of the model version used in this work (ERSEM-2004). The model has 49 biogeochemical state variables and it applies a functional type approach. Primary producers are split into four size-based phytoplankton functional types (PFTs): diatoms, phytoflagellates, picophytoplankton and dinoflagellates. Each of these is defined in terms of its content of chlorophyll, carbon, nitrogen, phosphate, and (for diatoms only) silicate. In the PFTs, the stoichiometric ratios of chlorophyll-to-carbon and nutrients-to-carbon ratios are time variable (Geider et al., 1997; Baretta-Bekker et al., 1997). Three functional types of zooplankton, i.e. mesozooplankton, microzooplankton, and heterotrophic nanoflagellates (HNAN), prey on the different PFTs, as a function of their size. Besides the PFTs, the model describes the dynamics of other optically-active compounds: particulate matter (distinguished in three classes: small, medium and large) and labile and semi-labile dissolved organic matter. The model also includes the dynamics of one bacterial functional group, dissolved inorganic nutrients, oxygen and carbon dioxide. A benthic return module was used to simulate the nutrient fluxes from sediments to the water column.

Bio-optical module

The bio-optical module describes the spatio-temporal changes of the spectral diffuse attenuation coefficients, $K_d(\lambda)$, at three optical bands (blue: $\lambda = 443$ nm, green: $\lambda = 555$ nm, red: $\lambda = 670$ nm). $K_d(\lambda)$ is a function of the inherent optical properties which, in turn, are related to the concentrations of optically active compounds provided by the ecosystem model (Appendix A). The following equation, based on Lee et al. (2005a), was applied:

$$K_d(\lambda) = (1 + 0.005\theta)a_{TOT}(\lambda) + 4.18(1 - 0.52e^{-10.8a_{TOT}(\lambda)})b_{b,TOT}(\lambda) \quad (1)$$

where θ is the solar zenith angle. The total absorption $a_{TOT}(\lambda)$ was decomposed in the contributions of the absorptions (a_i) by phytoplankton (*phy*), non-algal particles (*NAP*), coloured dissolved organic matter (*CDOM*), and seawater (*sw*) (see, e.g., Kirk, 1983):

$$a_{TOT}(\lambda) = a_{phy}(\lambda) + a_{NAP}(\lambda) + a_{CDOM}(\lambda) + a_{sw}(\lambda) \quad (2)$$

The total backscattering $b_{b,TOT}(\lambda)$ in Eq. (1) is a function of the backscattering ($b_{b,i}$) by seawater (*sw*) and particles (*p*), subdivided in small organic particles (*POC1*), large organic (*POC2*), and particulate inorganic matter (*PIM*) (Fujii et al., 2007):
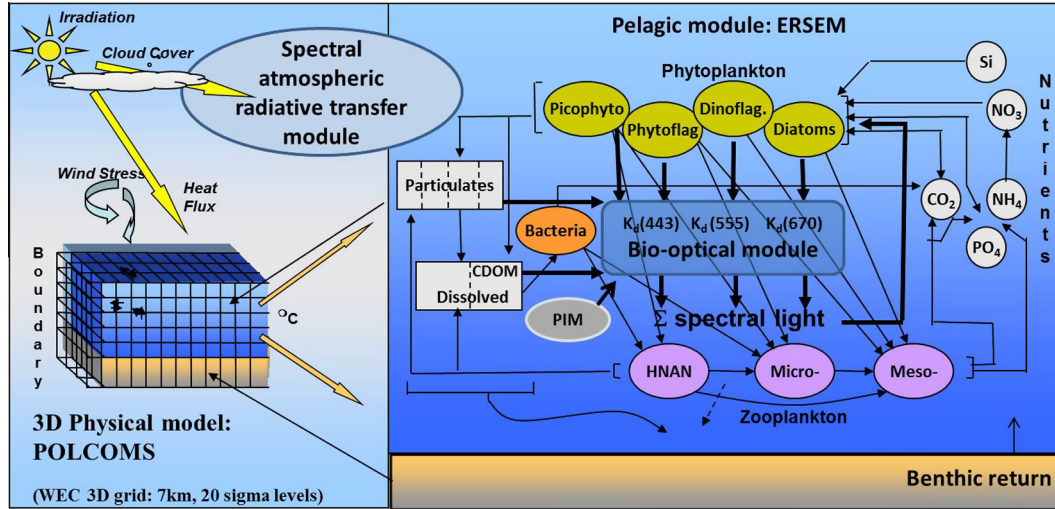
Fig. 1. Schematic diagram of the coupled biogeochemical–optical–physical model of the western English Channel.

$$b_{b,TOT}(\lambda) = b_{b,p}(\lambda) + b_{b,sw}(\lambda)$$
$$= b_{b,POC1}(\lambda) + b_{b,POC2}(\lambda) + b_{b,PIM}(\lambda) + b_{b,sw}(\lambda) \tag{3}$$

The contributions a_i and $b_{b,i}$ in Eqs. (2) and (3) were modelled using the equations and parameterization in Appendix A.

In the coupled optical–ecosystem model, the spectral $K_d(\lambda)$ in Eq. (1) determines the underwater spectral light field according to the classic Beer–Lambert formulation. The spectral downward irradiance at the sea surface was computed at each λ by using the spectral atmospheric radiative transfer model by Gregg and Carder (1990), modified according to Reed (1977) to include the effects of clouds. The spectral irradiances are then integrated into three broad bands that are representative of the blue (400–500 nm), green (500–600 nm) and red (600–700 nm) regions of the visible spectrum (see Manizza et al. (2005) for a comparable approach with a global ecosystem model).

Assimilation scheme

The western English Channel assimilation system for ocean colour (Ciavatta et al., 2011) is based on the Ensemble Kalman filter (EnKF) (Evensen, 1994), in its localized version by Evensen (2003). The EnKF is a Monte Carlo method (Evensen, 1994), where an initial ensemble of N state vectors $\mathbf{x}_0^{a(l)}$ ($l = 1, 2, \ldots, N$) is randomly sampled from an initial probability density function. Each ensemble member, i.e. state vector, is propagated in time using the full model equations during the "forecast step", that provides the EnKF "forecasts":

$$\mathbf{x}_i^{f(l)} = M_{i,i-1}[\mathbf{x}_{i-1}^{a(l)}] + \eta_i^{(l)} \quad \eta_i^{(l)} \propto N(\mathbf{0}, \mathbf{Q}_i) \tag{4}$$

where $\mathbf{x}_i^{f(l)}$ is the lth vector of model variables ($l = 1, 2, \ldots, N$), i.e. a forecasted ensemble member at time i. The forecast is carried out by means of the non-linear model operator M, starting from the "analysis" state vector $\mathbf{x}_{i-1}^{a(l)}$ defined at time $i - 1$. Each state member is subject to a model error defined by the vector $\eta_i^{(l)}$, which is assumed to be Gaussian noise with zero mean and covariance matrix $\mathbf{Q}_i$.

At time i, the forecast state $\mathbf{x}_i^f$ and the forecast covariance matrix $\mathbf{P}_i^f$ are defined from the mean value and the covariance matrix of the N forecasted members:

$$\mathbf{x}_i^f = \frac{1}{N} \sum_{l=1}^N \mathbf{x}_i^{f(l)} \tag{5}$$

$$\mathbf{P}_i^f = \frac{1}{N-1} \sum_{l=1}^N \left(\mathbf{x}_i^{f(l)} - \mathbf{x}_i^f \right) \left(\mathbf{x}_i^{f(l)} - \mathbf{x}_i^f \right)^T \tag{6}$$

Supposing that at time i a vector $\mathbf{y}_i^o$ of observations related to the model variables becomes available, an ensemble of N random observation vectors is generated: $\mathbf{y}_i^{o(l)}$ ($l = 1, 2, \ldots, N$), assuming that the observation errors are Gaussian and have covariance matrix $\mathbf{R}_i$.

The mismatches between the ensemble of observations and model outputs are collected in the "innovation vector":

$$\mathbf{d}_i^{(l)} = \mathbf{y}_i^{o(l)} - H_k\left[\mathbf{x}_i^{f(l)}\right] \tag{7}$$

where H_k is the (possibly non-linear) measurement operator that relates the observations to the "true" values of the variables in the system.

Given the observations $\mathbf{y}_i^{o(l)}$ and the covariance matrix $\mathbf{P}_i^f$ at time i, the assimilation scheme updates (i.e. "corrects") the forecasted states $\mathbf{x}_i^{f(l)}$, in the EnKF "analysis" step. This provides the members $\mathbf{x}_i^{a(l)}$ of the analysis ensemble:

$$\mathbf{x}_i^{a(l)} = \mathbf{x}_i^{f(l)} + \mathbf{K}_i \mathbf{d}_i \tag{8}$$

$$\mathbf{K}_i = \mathbf{P}_i^f \mathbf{H}_i^T \left(\mathbf{H}_i \mathbf{P}_i^f \mathbf{H}_i^T + \mathbf{R}_i \right)^{-1} \tag{9}$$

where $\mathbf{H}_i$ denotes the linearization of H_k defined in Eq. (7). The matrix $\mathbf{K}_i$ in Eq. (9) is the Kalman gain. It scales the model-data mismatches, $\mathbf{d}_i^{(l)}$, by balancing the uncertainty in the model, $\mathbf{P}_i^f$, and in the observations, $\mathbf{R}_i$ (Eq. (9)).

The analysis state $\mathbf{x}_i^a$ and the analysis covariance matrix $\mathbf{P}_i^a$ are defined from the analysis ensemble $\mathbf{x}_i^{a(l)}$, using equations similar to Eqs. (5) and (6) (Evensen, 2003). The analysis ensemble estimated in Eq. (8) at time i is then used to produce a new state forecast for time $i + 1$ (Eq. (4)), in a sequential procedure that estimates the evolution of the model variables over the time window spanned by the assimilated observations.

In our application of the EnKF, we used the approach by Evensen (2003), which computes the innovations and observation error covariance matrix (Eq. (7) and $\mathbf{R}$ in Eq. (9)) by using an ensemble of perturbed observations. This approach has been applied with large dimensional ocean models (e.g. Natvik and Evensen, 2003; Ciavatta et al., 2011; Hu et al., 2012; Storto et al., 2013), also because of its computational efficiency (Nerger et al., 2005). However, it is known that this approach introduces

additional sampling errors and may induce loss of rank in the analysis ensemble, in particular when the number of independent observation is larger than the ensemble size (Whitaker and Hamill, 2002; Kepert, 2004; Nerger et al., 2005). These issues can be exacerbated by temporal/spatial interpolation of the assimilated data, which can enhance dependencies among observations. In applications where these issues become critical, deterministic variants of the EnKF that do not require perturbed observations may be applied (see the reviews by Nerger et al. (2005) and Sun et al. (2009)).

Set-up of the assimilation scheme

Following Ciavatta et al. (2011), we used the EnKF with an ensemble size of $N = 100$ and the analysis was carried out on each grid point applying a "local analysis" (Evensen, 2003), i.e. by accounting only for the observations and the states within a pre-defined distance (radius) from the grid point. The radius varies from 25 km, at depths greater than 40 m, to 10 km at shallower locations. The analyzed state vector included forty-seven out of the forty-nine biogeochemical state variables (the two gas components were excluded in analogy with Ciavatta et al., 2011), and it was augmented with $K_d(443)$ as diagnostic variable directly comparable with the observations (Evensen, 2003). The other bio-optical components are diagnostic variables computed in the forecast step, using Eqs. (A.1)–(A.9) in Appendix A.

Model variables and observations were log-transformed prior to the analysis, to guarantee positivity of the solutions (Janjić et al., 2014; and see Torres et al., 2006; Nerger and Gregg, 2008; and Ciavatta et al., 2011, for analogous approaches in biogeochemical assimilation).

The observational error of the satellite data was set $_{rel}$ = 20% of the $K_d(443)$ values. This is a conservative error, larger than 11% computed by Lee et al. (2005b) and 13% for $K_d(490)$ computed by Zhao et al. (2013), in both oceanic and coastal waters. Spatially correlated observation errors were generated by scaling two-dimensional pseudo-random fields by $_{rel}$ (Evensen, 1994), with order 10^2 km decorrelation, to account for spatial patterns in ocean colour uncertainty (Maritorena et al., 2010; and see also Ford et al., 2012).

Model error in the EnKF forecast is generated by perturbing the diffuse attenuation coefficients, thus inducing fluctuations in the underwater light field that drives photosynthesis (Torres et al., 2006; and see comparable perturbation of forcing in, e.g., Natvik and Evensen, 2003; Simon and Bertino, 2009). Specifically, pseudo-random Gaussian errors with standard deviation equal to 20% of the $K_d(443)$ values were applied to each ensemble member, at every model time-step, during the forecast phase. Model error was also added to all the remaining model variables at the end of the forecast phase, prior to the analysis, as pseudorandom Gaussian fields with standard deviation equal to 10% of the value of the variables. This error was lowered to 5% for variables defining the dissolved and particulate organic concentrations, to avoid divergences in the concentrations of the largest pool in the model, as suggested by Torres et al. (2006). This "additive error" is applied in order to guarantee sufficient spread in the EnKF ensemble, thus avoiding possible filter divergence (Hamill and Whitaker, 2005). An alternative approach is the "covariance inflation", where ensemble member's deviations about their mean are inflated by a constant factor (e.g. Hamill and Whitaker, 2005; Hu et al., 2012), or where adaptive factors can be used to introduce variable inflation in the EnKF covariance matrix (see e.g. Luo and Hoteit (2011) and Triantafyllou et al. (2013) for an application with a marine ecosystem model and the SEIK filter).

Following Evensen (2003), the ensemble initial conditions for January 2006 were defined by running a spin-up ensemble simulation for winter 2006, where the underwater light fields were stochastically perturbed, as in the EnKF forecast step described above (see similar applications in Natvik and Evensen (2003) and Torres et al. (2006)). In turn, the spin-up ensemble simulation was initialized by perturbing the initial conditions of the reference simulation, which had been spun-up for five years using the forcing and boundaries of year 2006.

To provide comparative results between the performances of $K_d(443)$ *versus* chlorophyll assimilation in estimating the *in situ* data, we ran a yearlong assimilation of SeaWiFS chlorophyll data for year 2006. This run used the same model configuration and assimilation set-up applied for $K_d(443)$, but it accounted for the higher error of satellite chlorophyll retrievals. We used the empirical error equation proposed by Ciavatta et al. (2011), which defines the error at ∼60% of the chlorophyll concentrations in coastal grid points and 35% in the open-shelf (see, e.g., Saba et al., 2011). Finally, to better understand the relevance of the different errors of $K_d(443)$ and chlorophyll observations, we performed a one year-long sensitivity test for chlorophyll assimilation, by setting the percentage error to 20% of the data value, i.e. as in $K_d(443)$ assimilation. The results of the test are presented in Appendix C.

Data

Global, eight-day composites of $K_d(443)$ at 9 km resolution, derived from SeaWiFS ocean colour data using the method of Lee et al. (2005b), were downloaded from the OceanColor Web site (oceancolor.gsfc.nasa.gov; reprocessing version R2010.0). The data was downscaled to the 7 km model grid, and the eight day composites were interpolated in time to compute daily products. Then the daily products were merged, providing the five-day assimilated composites. The date of each composite was set equal to the middle day of the composite. Eight-day composites of SeaWiFS chlorophyll data, derived using the OC4v4 algorithm at 9 km resolution (oceancolor.gsfc.nasa.gov; reprocessing version R2010.0), were processed in the same way, and applied for skill assessment and in the chlorophyll assimilation experiment. The satellite $K_d(443)$ and chlorophyll data are presented in Figs. 2 and 3, respectively, which show the annual average values of the data and their coefficient of variations (CV, the ratio among the annual standard deviation and average), at each point of the model domain.

We note that the merging and spatial regridding of satellite data described above are common practice in ocean colour assimilation (see e.g. Triantafyllou et al., 2007; Fontana et al., 2013; Mattern et al., 2013; Teruzzi et al., 2014). On the one hand, daily products may suffer from large number of missing data due to cloud cover. On the other hand, the spatial regridding through off-line projection of data onto the model grid is a relatively simple procedure to match-up satellite observations with model output. However, the straight use of data and the interpolation of model output on the observation grid through the operator H in Eq. (7) is in principle recommendable (Brasseur, 2006; Ford et al., 2012). We accounted in part for error patterns introduced by data regridding through setting spatially correlated observation errors (see (Bertino et al., 2003), although increased temporal structures of the error remained unrepresented (see Section "Set-up of the assimilation scheme").

The biogeochemical data used to evaluate the assimilation skill were measured weekly at the long-term monitoring station "L4" of the Western Channel Observatory, situated ∼10 km south of Plymouth (50°15'N, 04°13'W, water depth ∼55 m). L4 is a hydrodynamically complex site that is influenced by the outflows from the river Tamar and exhibits seasonal stratification in summer (Smyth et al., 2010). Concerning the optical properties, L4 can be classified as a case I or case II site depending upon the season, the optical parameter and the wavelength (Groom et al., 2009).
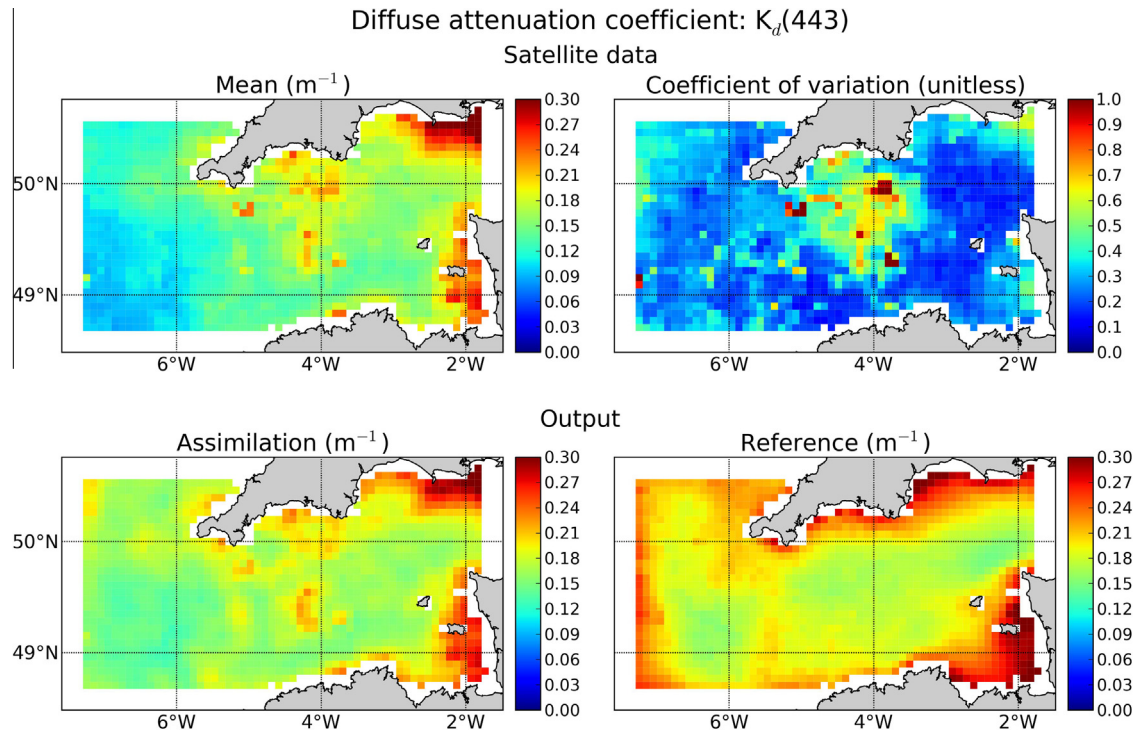
Fig. 2. Spatial distribution of $K_d(443)$ in the western English Channel (year 2006). The upper panels show the mean annual values and the annual coefficient of variation computed from the satellite composites. The lower panels show the average annual values computed from the output of $K_d(443)$ assimilation (analysis), and from the reference simulation.
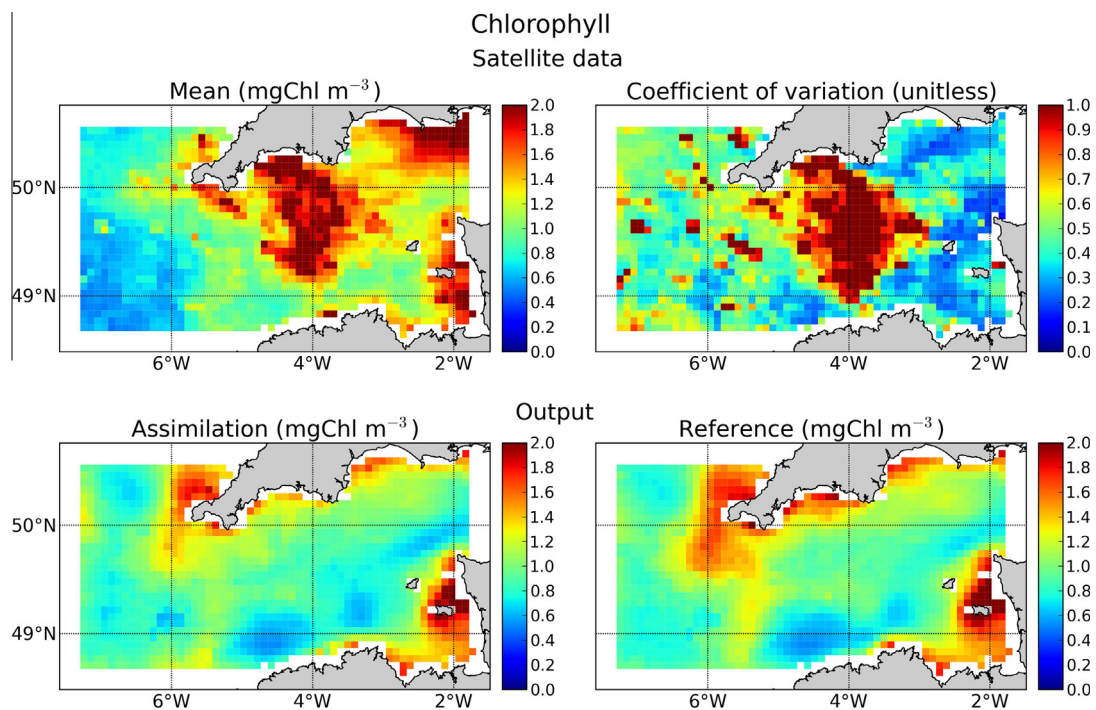
Fig. 3. Spatial distribution of chlorophyll in the western English Channel (year 2006). The upper panels show the mean annual values and the annual coefficient of variation computed from the satellite composites. The lower panels show the average annual values computed from the output of $K_d(443)$ assimilation (analysis), and from the reference simulation.

The *in situ* data and the details of the sampling protocols and analytical methods are available at the western Channel observatory website (www.westernchannelobservatory.org.uk) and, for the optical properties, in Groom et al. (2009) and Tilstone et al. (2012).

Skill metrics

The skill of the reference and assimilation runs were evaluated using three skill metrics (see, e.g., Ciavatta et al. (2011) for

formulas): (1) the root mean square error (*RMSE*); (2) the percentage model bias (*Pbias*); and (3) the Pearson correlation coefficient (*r*).

RMSE evaluates the magnitude of the overall mismatches between the model estimates and the observations. *Pbias* is the ratio between the model-to-data mismatch normalized by the data: it quantifies unidirectional deviations of the estimates. The lower the *RMSE* and *Pbias* (in absolute value), the closer are the estimates to the observations (high skill). The correlation *r* measures the model ability to reproduce the main intra-annual patterns of the data. When the data show no seasonality, the correlation measures the model performance in resolving short-term fluctuations. Values of *r* equal to one indicate a linear, positive correspondence between the fluctuations of data and estimates (high skill). We applied Pearson's correlation to assess linear relationships among model and data of a single variable, i.e. as a skill metric (see, e.g., Gregg et al., 2009). The Spearman rank correlation was used in Section 'Skill in simulating the satellite data evolution in representative areas' to assess monotonic correspondence between *in situ* data of different variables, i.e. as a non-parametric exploratory statistic (see, e.g., Sheskin, 2011).

When comparing the assimilation (DA) and reference (ref) skill in estimating the satellite or *in situ* data, differences of the metrics values were computed as Δmetric = metric$_{DA}$ − metric$_{ref}$, using absolute values for *Pbias*, and normalizing by metric$_{ref}$ in the case of RMSE differences. We assumed that a positive difference of correlation indicates that assimilation has a better skill, because its correlation is "closer to 1", when compared with the reference; negative differences indicate a worse skill. When comparing $K_d(443)$ and chlorophyll (chl) assimilation, differences in metrics were computed as Δmetric = metric$_{Kd}$ − metric$_{chl.}$, and normalized by metric$_{chl}$ in the case of RMSE differences.

The metrics for the EnKF output *versus* the satellite observations were computed before the assimilation of the data (i.e. for the EnKF "forecast", Eq. (4)), as well as after assimilation (i.e. for the EnKF "analysis", Eq. (8)).

Set-up of the simulations

The reference model run and data assimilation were applied to simulate year 2006. The lateral boundary data for the biogeochemical and physical variables were extracted from POLCOMS-ERSEM model simulations of the North-West European Continental Shelf (MRCS) domain, run operationally at the UK Met Office (Lewis and Allen, 2009; Siddorn et al., 2007). In turn, the MRCS model is forced at the boundaries with temperature, salinity, barotropic velocity, and sea-surface height from the Met Office Atlantic Margin Model (AMM) (Siddorn et al., 2007).

The simulation was forced with the operational meteorological data of the European Centre for Medium-Range Weather Forecasts (ECMWF), downloaded from the British Atmospheric Data Centre (BADC, www.badc.nerc.ac.uk). The data set includes six-hourly surface values of wind, relative humidity, pressure, air temperature and cloud cover. The solar irradiance and zenith angle are calculated as a function of the latitude, day of the year and time. The freshwater and nutrient loads from eight rivers were input as climatological daily means (Lewis and Allen, 2009).

ERSEM-2004 does not describe the dynamics of the refractory dissolved organic carbon (DOC2 in Eq. (A.8)) and particulate inorganic minerals (PIM in Eqs. (3) and (A.2)). As a first approximation, these variables were derived from the simulation of a passive tracer that is discharged from rivers, transported in the western English Channel and exchanged through the open boundary. In particular, DOC2 values at the open boundary were fixed to the concentration of 40 µM (i.e. 480 mgC m^{-3}) for oceanic waters (Ogawa and Tanoue, 2003), while the value of the riverine concentration (25 mgC m^{-3})

was tuned to match the data of CDOM absorption at L4 from January to March (included). Therefore, the winter a_{CDOM} data were not used to compute the skill metrics. PIM was parameterized by scaling the passive tracer to meet the annual average of 560 mg m^{-3} measured at L4 in the years 2003–2004 (Groom et al., 2009).

Results and discussion

Skill in simulating $K_d(443)$ and chlorophyll spatial distributions

The reference and assimilative estimates of $K_d(443)$ and chlorophyll are compared with the annual averages of the satellite data in Figs. 2 and 3, respectively, while Figs. 4–7 show the skill in simulating the evolution of the weekly observations at each grid point of the western English Channel.

Assimilation of satellite $K_d(443)$ improved the model simulation of the light attenuation coefficient. Fig. 2 illustrates that, in general, assimilation analysis mitigated the model overestimation of the satellite data along the coast as well as at the open-shelf, where negative boundary effects are evident in the reference model (see to this regard Ciavatta et al., 2011). However, $K_d(443)$ remained overestimated in the south-west boundary of the study area. Fig. 4 confirms that the EnKF analysis (Eq. (8)) has a lower RMSE and higher correlation with the satellite data, with respect to the reference run, for most of the model domain. In particular, the assimilation improved the simulation in areas where the model skill was poor, e.g. in coastal and deeper areas where the correlation with data was in some cases negative (see Fig. 4 and discussion in Section 'Skill in simulating the satellite data evolution in representative areas').

One-week EnKF forecasts of $K_d(443)$ (Eq. (4)) had better skill than the reference run in several areas, including the internal part of the Channel and the British coast (Fig. 5). However, the estimates deteriorated in other parts of the region, e.g. in the Celtic Sea (north-west boundary) and in an area close to point S1 in Fig. 5.

Assimilation of $K_d(443)$ brought less evident benefits to the analysis of the (unassimilated) surface chlorophyll. Fig. 3 shows that $K_d(443)$ assimilation slightly decreased the model overestimation of chlorophyll in part of the western region and off the Cornish coast, but the high chlorophyll values in the middle of the channel remained underestimated. Fig. 6 confirms that assimilation decreased the RMSE at the edge of the English Channel, along ∼6°W. Assimilation reduced a model overestimation of chlorophyll, which was due to an unobserved phytoplankton bloom in summer in the reference simulation (not shown). However, deteriorations were also evident. The RMSE increased in the Celtic Sea at north-west and at the south boundary, while correlation decreased in the central part of the channel, along ∼4°W. Changes were less relevant in the remaining areas of the western English Channel.

Finally, the maps of skill changes for the chlorophyll forecasts (Fig. 7) reflect smoothly the values for the chlorophyll analysis (Fig. 6), indicating that, on average, the system keeps some memory of the analysis corrections of the chlorophyll fields, throughout the one-week forecast steps. However, as for the analysis, these forecasts do not represent a significant improvement of the reference estimates of chlorophyll distribution.

While improvements in the analysis of the assimilated variable shown in Fig. 4 are expected from a properly tuned assimilation system (Gregg et al., 2009), forecasting the not-yet-assimilated data (Fig. 5) or the not-assimilated-variables (Fig. 7) are challenging tasks for state-of-the-art operational forecasting systems (e.g. Ford et al., 2012; Teruzzi et al., 2014). In principle, the re-initialization of the assimilated variable closer to the data should improve also the forecast of the next available data, with respect to the
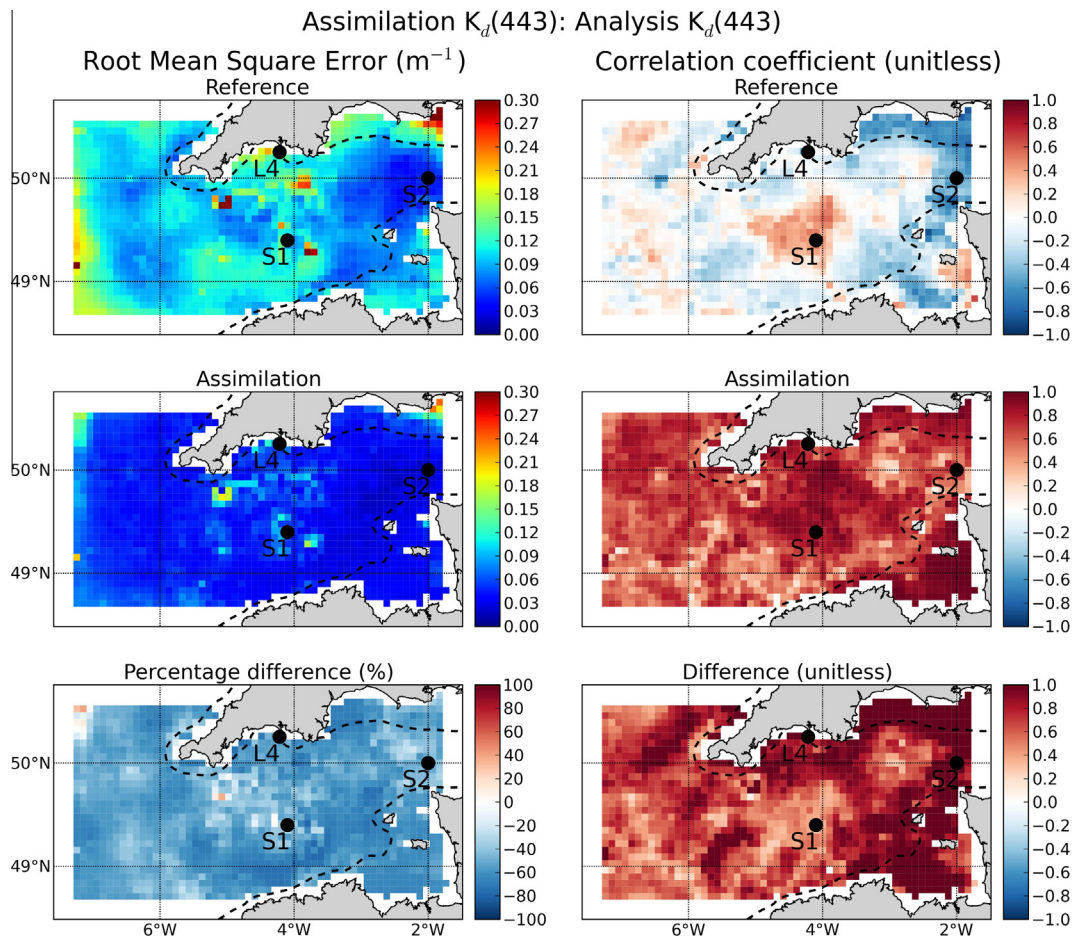
Fig. 4. Skill of reference simulation and $K_d(443)$ assimilation (analysis) in estimating the satellite $K_d(443)$ composites in 2006. The panels show the root mean square error and correlations between weekly satellite observations and the reference output (top panels), and assimilation output (central panels). The bottom panels show the difference between the reference and assimilation metrics. The dotted line indicates the 50 m isobaths; S1, S2 and L4 are representative sites.
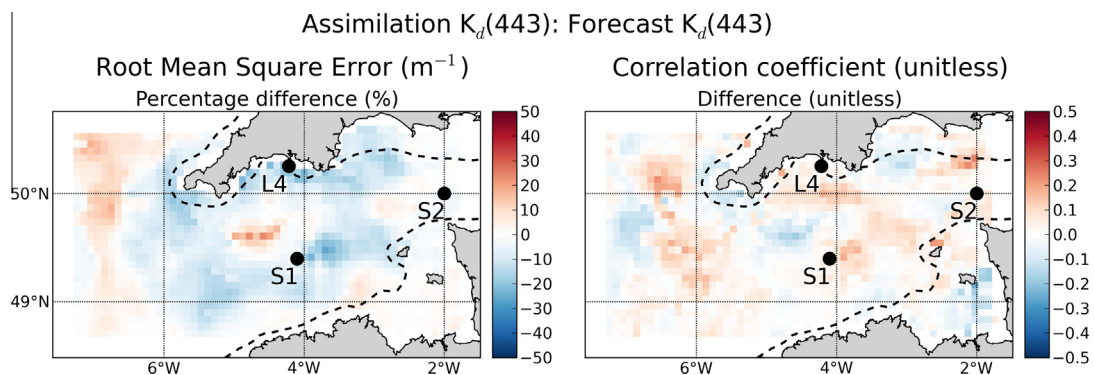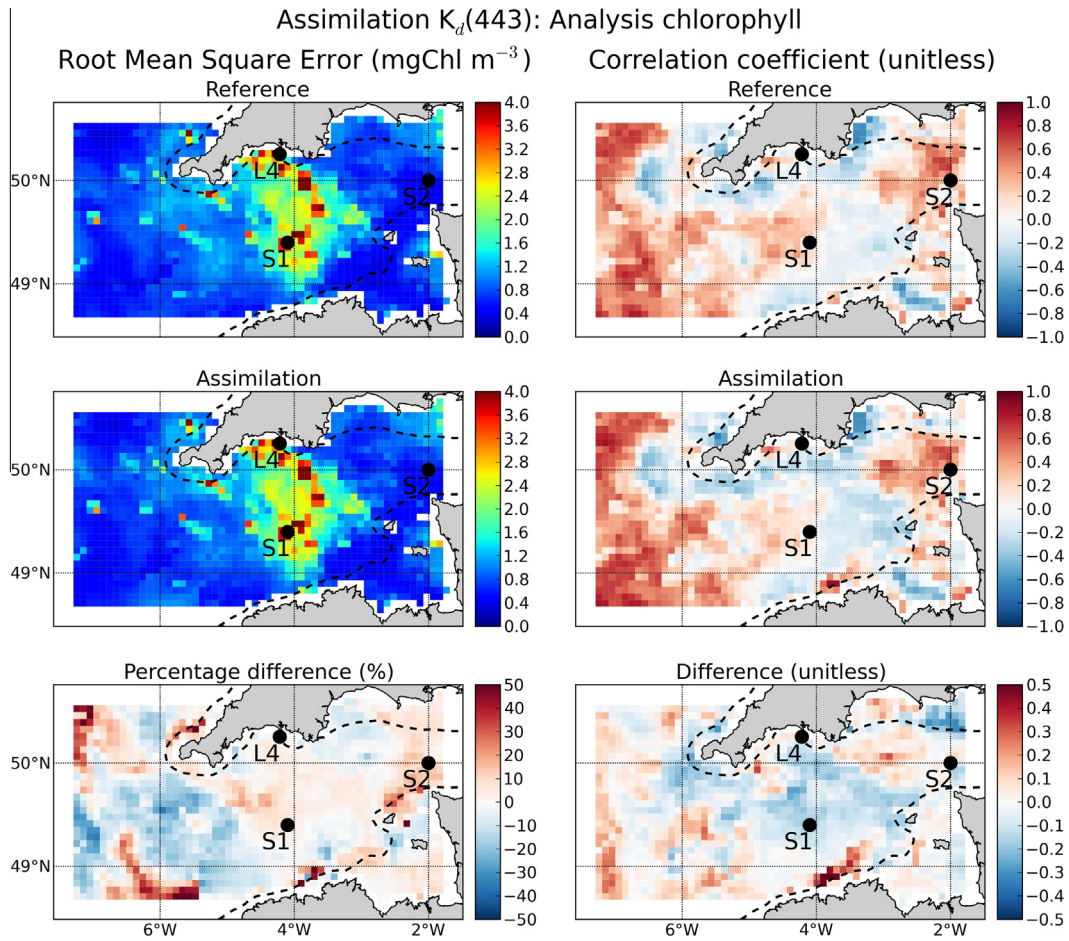

Fig. 5. Comparison of the skill of reference simulation and $K_d(443)$ assimilation (forecast) in estimating the satellite $K_d(443)$ composites in 2006. The panels show the difference between the reference and assimilation metrics. The dotted line indicates the 50 m isobaths; S1, S2 and L4 are representative sites.

reference run. However, re-initialized biogeochemical simulations often tend to converge back to the reference simulation (Allen et al., 2003; Gregg, 2008; Hu et al., 2012) because of the effect of hydrodynamics, forcing and boundaries values (see Friedrichs et al., 2006 for a discussion on the relevance of hydrodynamics in the context of biogeochemical data assimilation). Furthermore, the analysis can produce changes in some variables (e.g. generate a phytoplankton bloom) that are not sustained by changes in the un-assimilated variables (e.g. by nutrient increase), thus that the biogeochemical dynamics progressively counteract the assimilation corrections (e.g. the bloom declines) (Teruzzi et al., 2014).

On the other hand, multivariate analysis can produce values that are not consistent with the simulated model dynamics, e.g. outlier nutrient values, thus developing simulation instabilities that can deteriorate the forecasts of both the assimilated and un-assimilated variables (Ciavatta et al., 2011).

Skill in simulating the satellite data evolution in representative areas

The spatial distribution of the model and assimilation skill in simulating satellite data is related to the optical characteristics of the different areas of the western English Channel.

Fig. 6. Skill of reference simulation and $K_d(443)$ assimilation (analysis) in estimating the un-assimilated satellite chlorophyll composites in 2006. The panels show the root mean square error and correlations between weekly satellite observations and the reference output (top panels), and assimilation output (central panels). The bottom panels show the difference between the reference and assimilation metrics. The dotted line indicates the 50 m isobaths; S1, S2 and L4 are representative sites.
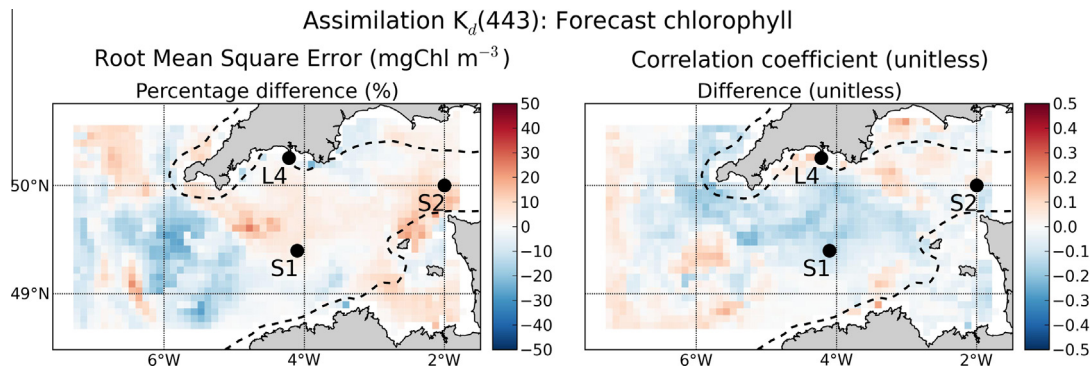

Fig. 7. Comparison of the skill of reference simulation and $K_d(443)$ assimilation (forecast) in estimating the un-assimilated satellite chlorophyll composites in 2006. The panels show the difference between the reference and assimilation metrics. The dotted line indicates the 50 m isobaths; S1, S2 and L4 are representative sites.

In general, the model had better skill in simulating $K_d(443)$ in areas where phytoplankton was the main driver of the optical dynamics (i.e., in Case I waters), than in areas where other processes were more relevant, e.g. vertical mixing and riverine discharges of optical active compounds (Case II waters). Furthermore, the model had the poorest correlation skill where $K_d(443)$ had low seasonal variability.

The above findings are illustrated in Fig. 8, showing the time evolution of the outputs and satellite data at three representative sites (S1, S2 and L4 indicated in Figs. 4–7). At site S1, phytoplankton

absorption appears to be the main driver of $K_d(443)$, as suggested by the correlation among the chlorophyll and K_d data (Spearman rank correlation 0.94). There, the annual variability was relatively high for both $K_d(443)$ and chlorophyll data: the coefficient of variations, CV, were ~50% and 100%, respectively. The seasonal cycles of both variables were characterized by higher values in spring and summer. The model had good skill in simulating the seasonal cycle of $K_d(443)$ and assimilation improved the estimates in August and September. However, in these months, the analysis also drove a decrease of chlorophyll concentration, which explains the
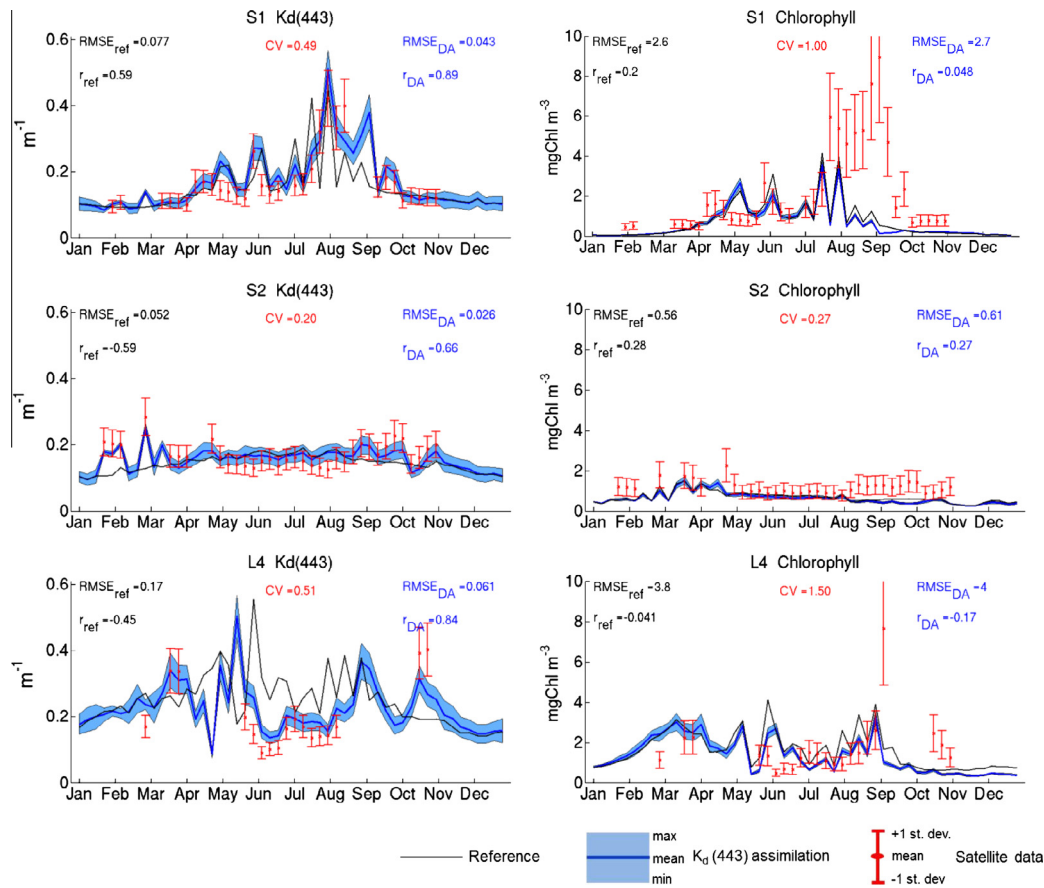
Fig. 8. Simulation of $K_d(443)$ (graphs on the left) and chlorophyll concentration (right) at the sites S1, S2 and L4 indicated in Figs. 4–7. The graphs compare the reference output, the mean value and range of the assimilation output, and the satellite data with their standard deviation. The graphs report the root mean square error (RMSE) and the Pearson correlation (r) between satellite data and both the reference (ref) and data assimilation (DA) outputs. The coefficient of variation (CV) of the data is also reported, from Figs. 2 and 3.

deterioration of the annual chlorophyll RMSE and correlation at S1 (Fig. 8), as well as along ∼4°W (Fig. 5). In this area, the late-summer peaks of chlorophyll are due to massive blooms of vertically mobile, mixotrophic dinoflagellates (Smyth et al., 2010), which the western English Channel ecosystem model does not explicitly describe (Lewis and Allen, 2009).

At S2, we found the lowest, negative model correlation with $K_d(443)$ data in the western English Channel ($r \sim -0.6$); this result is attributable to the low seasonality in the observations (CV equal to 20% and 27% for K_d and chlorophyll, respectively). The data fluctuated around the mean annual value. The reference simulation underestimated the winter and autumn K_d data, but laid within the error bounds of the observations in most cases, as confirmed by the relative low RMSE. Extending these considerations to the deeper waters of the western English Channel (>50 m), we found that in 80% of the cases where the reference correlation skill was negative, the seasonal cycle of the attenuation coefficient was relatively low (CV < 35%: this threshold was chosen arbitrarily as the average value between CV of K_d at S1 and S2). Where CV is low, a negative correlation skill indicates poor performance in simulating weekly fluctuations, rather than seasonal cycles. Skilled simulation of short-term fluctuations is still a challenging task for complex marine ecosystem models (see, e.g., Allen et al., 2007; Stow et al., 2009; Shulman et al., 2013).

At S2, the low seasonality of $K_d(443)$ was linked to the low seasonality of chlorophyll, as also observed in other parts of the domain. The coefficient of variation of $K_d(443)$ and chlorophyll data are correlated in the western English Channel (Figs. 2 and 3, Spearman rank correlation = 0.59). Where both seasonalities are low,

processes other than plankton blooms can be important drivers of the ecosystem and optical dynamics. In the S2 area, the inflows of optical active matter from the eastern channel (see the high $K_d(443)$ values at the east boundary in Fig. 2), as well as the vertical mixing of the water column throughout the year (Southward et al., 2005), may limit the seasonal phytoplankton dynamics.

At the coastal site L4 (Fig. 8), processes such as primary production, riverine discharges and seasonal stratification all contribute to the annual cycle of the optical properties (Groom et al., 2009). The cycle of the $K_d(443)$ satellite data is characterized by higher values in winter and autumn than in summer, and is correlated with the cycle of the chlorophyll data (Spearman rank correlation = 0.92). However, we note that the high values of $K_d(443)$ cannot represent just phytoplankton absorption in October, since in this month the chlorophyll concentrations were relatively low (Fig. 8). The model underestimated the peak $K_d(443)$ values and overestimated the summer ones at L4. Similar evolutions were observed at other sites in the coastal shallow areas. In Section 'Evolution of the assimilation performance' we argue that the suboptimal model performance in coastal areas can be related to the use of climatological and parameterized riverine inputs (see Section 'Set-up of the simulations'), as well as to the missing representation of sediment resuspension in the model. However, here we suggest that the mismatches between simulations and observations can be due in part also to large errors in the satellite data in coastal areas. This holds in particular for chlorophyll, since terrestrial CDOM can affect the retrieval algorithms (IOCCG, 2000), while $K_d(443)$ is a bulk measurement that explicitly includes the contribution of a_{CDOM}, amongst the other optically active components (Lee et al., 2005b).

Simulation of the optical and biogeochemical data at L4

Annual skill metrics

The skill metrics of the reference and assimilative simulations of the *in situ* data are listed in Tables 1, B.2 and B.3 (in Appendix B) and they are compared in Fig. 9. Assimilation of satellite $K_d(443)$ improved, in general, the simulation of the optical and biogeochemical variables observed at L4. However, errors slightly increased for nutrients, as well as for non-algal particle absorption (a_{NAP}).

Table 1 shows that the mean annual value of the data was in some cases lower than the standard deviation (in particular for the plankton component) indicating a remarkable temporal variability of the biogeochemical variables. The reference RMSE had in general comparable or higher values than the standard deviation (e.g. for TPC). Assimilation decreased the reference RMSE and bias, and increased the correlation for many simulated variables: twelve (out of eighteen) for RMSE, eleven for PBias and ten for correlation. However, changes were negligible for some variables, with respect the whole set of metrics, e.g. the carbon biomass of dinoflagellates (P4C) and microzooplankton (Z5C), while the simulation of other variables was deteriorated, e.g. ammonia.

The overall improvement of the biogeochemical simulation is illustrated by the total particulate nitrogen and carbon (TPC and TPN, respectively). For these variables, the reference correlation increased and RMSE decreased by ~10%, although the bias for TPN augmented slightly. TPC and TPN are the sum of eleven model variables (including all the plankton and detritus types), thus suggesting an improvement of the simulation throughout the whole model structure.

Data assimilation indicates that phytoplankton and CDOM were critical components of the simulated ecosystem at the L4 site, by imposing strong corrections to the optical properties associated to these two components (Fig. 9). The RMSE and bias of the $a_{phy}(443)$ estimates were reduced by 20%, and the correlation of $a_{CDOM}(443)$ was increased by ΔCorr = +0.16. However, we note that the correlation of a_{CDOM} remained low despite the increase, highlighting that assimilation can mitigate but not "fix" relevant model deficiencies, such as our simplistic simulation of riverine inputs of CDOM.

Evolution of the assimilation performance

In general, the reference simulation and $K_d(443)$ assimilation outputs both captured the average values of the *in situ* observations (see representative variables in Fig. 10), as well as the main seasonal features of the system (e.g. high-nutrients/low-biomasses in winter and autumn, low-nutrients/high-biomasses in spring and summer). The model also resolved the seasonal pattern of phytoplankton absorption, while $K_d(443)$ as well as absorptions by CDOM and non-algal particles had less evident seasonal cycles.

Looking at the phytoplankton functional types (PFTs), the reference simulation had in general good skill in simulating the evolution of phytoflagellates, apart from the unobserved spring peaks (Fig. 10). However, the model performed rather poorly for diatoms and dinoflagellates (see the high values of the annual RMSE and bias, and the low correlations in Table 1). Simulated diatoms were overestimated in winter and summer (not shown), fuelled by overestimated concentrations of dissolved inorganic nitrogen (Fig. 10). The reference simulation was not capable of capturing dinoflagellates blooms in September and October (not shown), and this might have contributed to the underestimation of TPC, TPN, and phytoplankton-excreted CDOM in those months (Fig. 10). Total chlorophyll and phytoplankton absorption were overestimated in summer and underestimated in autumn (Fig. 10). This outcome reflects the overestimation of diatoms in summer and the underestimation of dinoflagellates in autumn mentioned above. The seasonal biases in a_{phy}, contributed to the overestimation of $K_d(443)$ from July to August and underestimation in September, and this led to the annual positive bias for $K_d(443)$ in Table 1.

Assimilation impacted the simulated evolution of the phytoplankton community structure. In spring-summer, unobserved peaks of diatoms and phytoflagellates were pulled down. Consequently, the estimates of total chlorophyll, phytoplankton absorption, $K_d(443)$ and TPC were closer to the observations from April to August (Fig. 10). This explains the reduction of the annual RMSE and bias, and the increase of correlation for all the above-mentioned variables (see Fig. 9). However, assimilation could not improve the simulation of the dinoflagellates blooms in summer and autumn. As a consequence, the observed peaks of TPC, TPN,

Table 1

Statistics of the *in situ* data and skill metrics of the reference model run. Number (*N*), annual mean (*Mean*) and standard deviation (*Std dev*) of the biogeochemical and optical data measured weekly at station L4 in 2006; root mean square error (*RMSE*), percentage bias (*Pbias*) and correlation (*r*) of the reference estimates of the data. The notation of the variables is explained in Table B.1 of Appendix B.

Variables	N data	Mean data	Std dev data	RMSE reference	Pbias reference	r Reference
N1P (mmol m^{-3})	36	0.33	0.20	0.21	−41.80	0.56
N3 N (mmol m^{-3})	36	2.85	2.99	1.95	0.03	0.76
N4 N (mmol m^{-3})	36	0.58	0.60	0.65	21.06	0.15
N5S (mmol m^{-3})	34	2.52	2.10	1.83	−7.49	0.48
P1C (mgC m^{-3})	40	14.20	26.88	35.86	87.27	−0.04
P2C (mgC m^{-3})	40	12.27	8.18	9.50	−7.08	0.56
P4C (mgC m^{-3})	40	10.57	22.97	24.08	−81.63	0.15
Z5C (mgC m^{-3})	40	9.14	28.82	29.19	−57.17	0.04
Chl (mgChl m^{-3})	30	1.40	0.91	1.42	29.59	−0.02
PtC (mgC m^{-3})	40	37.03	37.19	52.05	52.59	0.17
TPC (mgC m^{-3})	38	240.36	103.30	143.02	−37.18	0.25
TPN (mmol m^{-3})	38	2.36	1.04	1.15	4.98	0.38
$a_{phy}(443)$ (m^{-1})	35	0.10	0.06	0.09	27.30	0.17
$a_{NAP}(443)$ (m^{-1})	35	0.04	0.02	0.02	15.52	0.31
$a_{CDOM}(443)$ (m^{-1})[a]	23	0.08	0.03	0.04	−29.92	−0.16
$K_d(443)$ (m^{-1})	24	0.20	0.09	0.15	43.21	−0.06
$K_d(555)$ (m^{-1})	25	0.13	0.06	0.07	24.65	−0.25
$K_d(670)$ (m^{-1})	20	0.57	0.10	0.12	5.63	0.03

[a] The values of a_{CDOM} from January to March (included) were not used to compute the skill metrics (Section 'Set-up of the simulations').
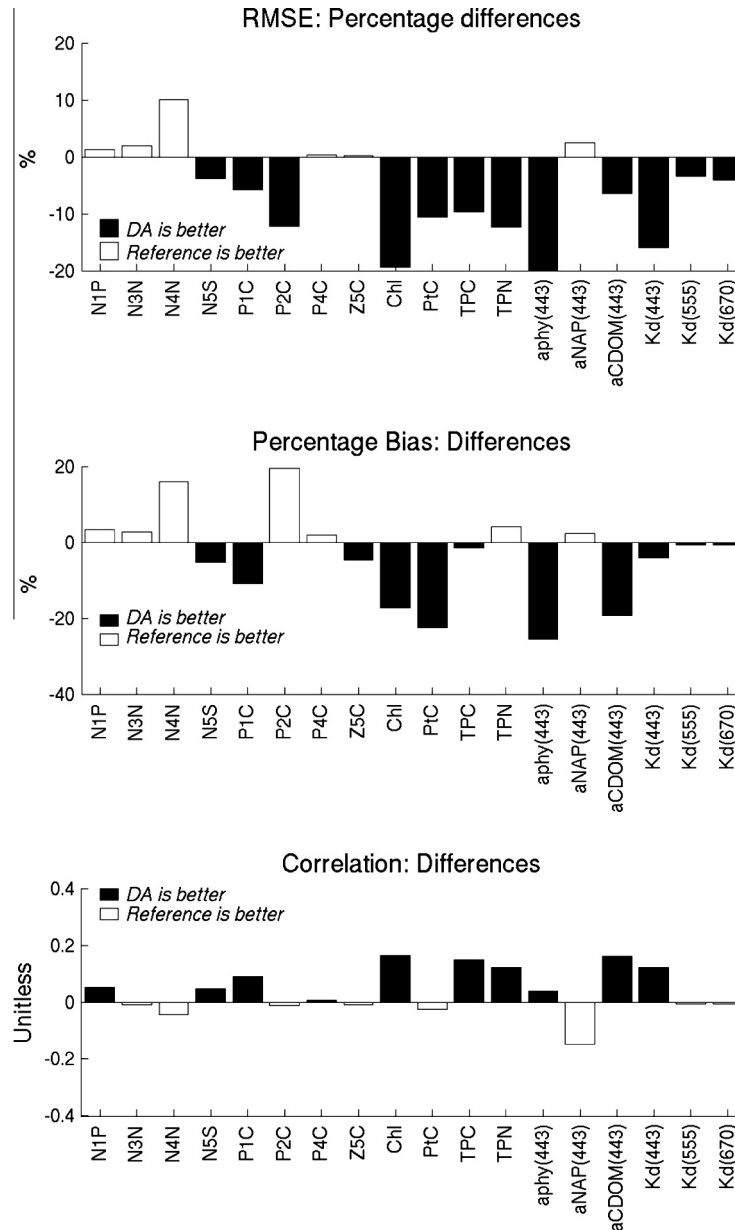
Fig. 9. Comparison of the skill metrics for $K_d(443)$ assimilation (forecast) and reference simulation in estimating the *in situ* data at L4. The bars represent the differences between the root mean square error (RMSE), percentage bias and correlation from the assimilation (DA) *versus* the reference simulation. Percentage values of RMSE were obtained by normalizing the differences with respect to the RMSE from the reference simulation. Black bars highlight the metrics improved by assimilation, i.e. decrease (negative differences) of RMSE and Pbias and increase (positive difference) of correlation.

$a_{phy}(443)$ and $K_d(443)$ remained substantially underestimated in September and October.

The defective simulation of the PFTs' evolution at L4 is explained by model limitations such as (Lewis and Allen, 2009; Ciavatta et al., 2011): (i) each model PFT includes species that actually may respond very differently to the environmental forcing (e.g. light, Polimene et al., 2014); (ii) dinoflagellates are aggregated in a PFT that does not describe the motility and the mixotrophic behaviour widely observed within this taxa (see e.g. Flynn et al., 2013); and (iii) the different PFTs are poorly differentiated in the parameterization by Blackford et al. (2004) applied in this study, and in particular they all have the same chlorophyll and nutrient reference quotas (see Table 3 in Blackford et al., 2004). Ciavatta et al. (2011) previously argued that the last mentioned model limitation, in particular, constrained the capability of chlorophyll

assimilation to improve the simulated PFTs succession. This most likely applies to $K_d(443)$ assimilation as well; here also the uncertainty in the parameterization of the specific absorption coefficients of the different PFTs (Appendix A1) might have limited the capability of the analysis in partitioning more effectively the EnKF correction among PFTs.

The reference simulation overestimated the non-algal particle absorption (a_{NAP}) in summer (Fig. 10). This can be due to the confinement of sediment down in the water column typically observed at L4 during summer stratification (Groom et al., 2009). This confinement cannot be simulated by the model, since it does not include the representation of sedimentation/resuspension of inorganic minerals at the water/sediment interface (Section 'Set-up of the simulations'). Assimilation could not mitigate this model deficiency and caused a slight increase of the a_{NAP} estimates in
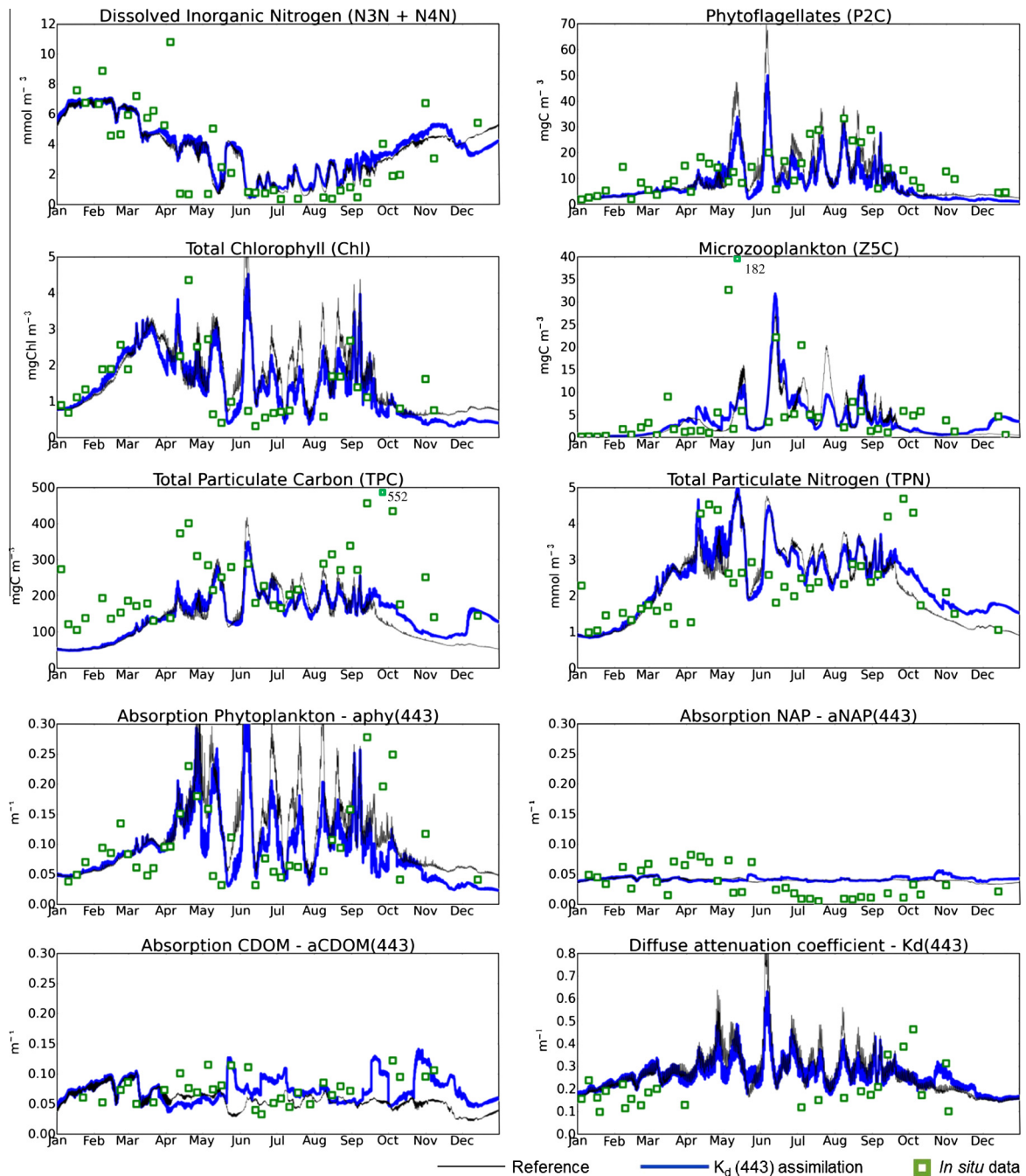
Fig. 10. Comparison of the reference and assimilation outputs with selected *in situ* data at L4. Note the out-of-scale values in the plots of microzooplankton (Z5C) and Total Particulate Carbon (TPC).

autumn, which explains the worsened metrics for this variable in Fig. 9.

Assimilation improved the reference underestimation of CDOM absorption (a_{CDOM}) in autumn (Fig. 10). The noticeable increase of $a_{CDOM}(443)$ in October was achieved through the increase of both the semi-labile and refractory components of DOC. Thus, assimilation likely mitigated the effects of using climatological river discharges and constant CDOM concentrations for the nearby Tamar river (see Section 'Set-up of the simulations'), which might have caused an underestimation of CDOM input in the rainy winter and autumn seasons. However, we note also that assimilation introduced a bias in the reference estimates of $a_{CDOM}(443)$ in the second half of June. This change was driven by a sudden increase of DOC, which followed a series of heavy corrections to the

phytoflagellates concentration (Fig. 10). Similar side effects of chlorophyll assimilation on unassimilated variables were noted also by Triantafyllou et al. (2007) during highly dynamic blooming events simulated in spring in the Mediterranean Sea.

Finally, we note that assimilation reduced slightly the bias of the microzooplankton estimates (Table 1), by smoothing unobserved summer peaks, in relation to analogous smoothing of the phytoplankton biomass (Fig. 10).

Assimilation impact on ecosystem process modelling

Assimilation of $K_d(443)$ affected the model simulation of the biogeochemical processes. In some instances, this contributed qualitatively to the improvements in the estimates of *in situ* data.
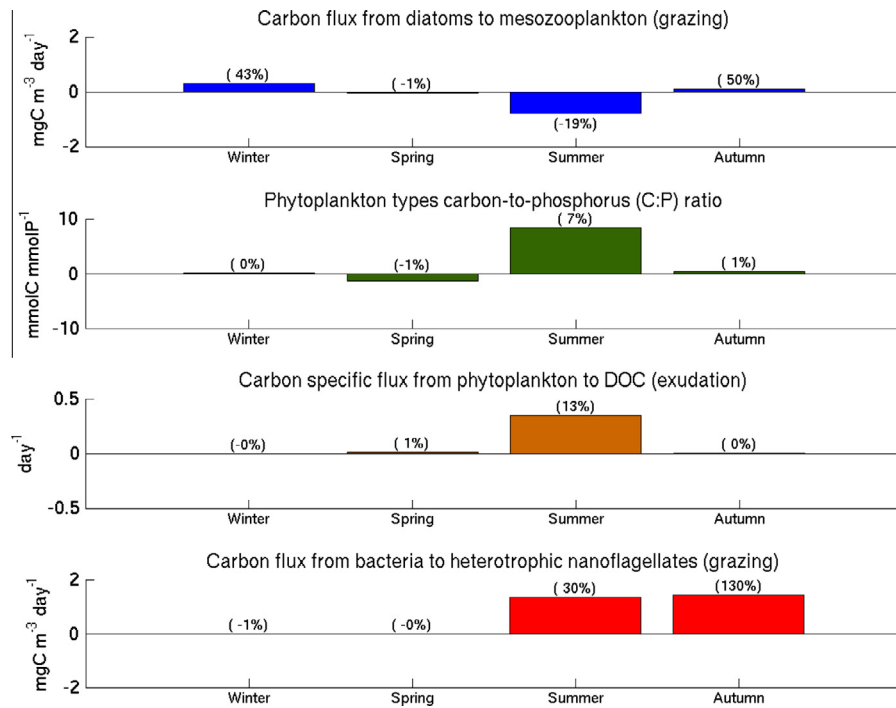
Fig. 11. Differences between the carbon fluxes and carbon-to-phosphorus ratio obtained in the $K_d(443)$ assimilation *versus* the reference simulation. The differences refer to winter (January–March), spring (April–June), summer (July–September) and autumn (October–December) 2006. The percentage values in brackets are the differences normalized by the mean seasonal values from the reference simulation. An interpretation of the changes of these fluxes and of the ratio is given in Fig. 12.
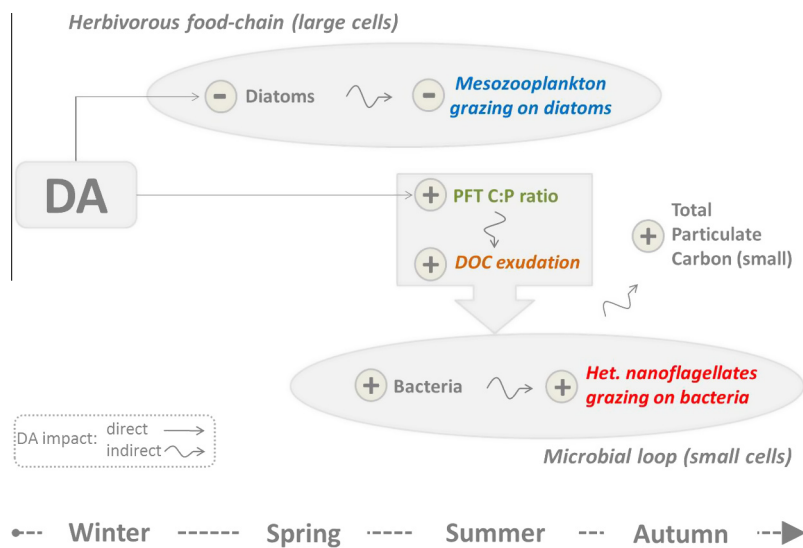
Fig. 12. Schematic of the impacts of $K_d(443)$ assimilation on the trophic web simulated by the reference model. The carbon fluxes and system properties in coloured font are shown in Fig. 11. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

In particular, assimilation impacted the reference simulation of the plankton trophic dynamics (see Figs. 11 and 12), improving the representation of total particulate carbon at station L4 (Fig. 10).

In marine systems, the carbon and nutrient fluxes through the plankton food-web are controlled by a continuum of trophic pathways ranging from the herbivorous-chain (driven by large phytoplankton and mesozooplankton) to the microbial-loop (driven by bacteria, small phytoplankton and heterotrophic nanoflagellates) (Legendre and Rassoulzadegan, 1995). The relative contribution of these two trophic structures is linked to the availability of dissolved nutrients: in general, the herbivorous food web is dominant in eutrophic conditions and the microbial loop in oligotrophic system (Thingstad and Rassoulzadegan, 1999). At L4, surface waters are nutrient-replete in winter (when the water column is fully mixed) and nutrient limited in summer (shallow mixed layer) (Smyth et al., 2010). Thus, one can expect that the herbivorous food web is dominant in winter-spring and that the contribution of the microbial loop increases in summer (Widdicombe et al., 2010; Eloire et al., 2010; Fileman et al., 2010). The seasonal succession of the trophic structures can be evaluated through the evolution

of the grazing of mesozooplankton on diatoms, and grazing of heterotrophic nanoflagellates (HNAN) on bacteria; i.e. by using these carbon fluxes as proxies for the herbivorous food chain and the microbial loop, respectively (Thingstad and Rassoulzadegan, 1999).

Assimilation accentuated the seasonal trophic succession between the herbivorous-chain and the microbial-loop (Fig. 12). The grazing of mesozooplankton on diatoms increased in winter and decreased it in summer, while the grazing of heterotrophic nanoflagellates on bacteria was increased in summer-autumn (Fig. 11).

At the root of the aforementioned changes are the "direct" impacts of assimilation on the model state variables, i.e. the corrections of the EnKF analysis in Eq. (8). Importantly, these impacts triggered "indirect" effects on the simulation of the biogeochemical fluxes, due to the flexible model description of the ecosystem processes. In spring-summer, the direct impacts were a general decrease of the phytoplankton biomass (Fig. 10) and an increase of the carbon-to-limiting nutrient ratio in the PFTs: see the increase of carbon-to-phosphorus (C:P) ratio in summer in Fig. 11. The higher C:P ratio remained within literature ranges (Geider and La Roche, 2002), and had the indirect effect of enhancing the nutrient-stress in phytoplankton, as consequence of the Droop-like kinetics described in the model equations (Blackford et al., 2004). In turn, the model description of nutrient-stress implied an increase of the phytoplankton exudation of dissolved organic carbon (Fig. 11), in order to re-equilibrate the intracellular stoichiometry (Blackford et al., 2004). The excreted DOC was directed to both the semi-labile and labile pool. The semi-labile pool has a relatively low bacterial degradation rate and thus it accumulated in the system increasing the CDOM concentration and absorption in autumn. The increase of labile DOC nurtured an increase of the bacteria biomass (not shown), which in turn enhanced the grazing of heterotrophic nanoflagellates in summer and autumn (Fig. 11). As a further consequence, microzooplankton biomass increased in late autumn (see Z5C in December in Fig. 10); because bacteria and heterotrophic nanoflagellates are the preferred prey of microzooplankton in the model (Blackford et al., 2004). The enhanced microbial loop-related dynamics in summer and autumn produced small detritus (via grazing, cell lysis and death) which, sinking very slowly, increased the suspended total particulate carbon in late autumn. This improved qualitatively the agreement with the observations: see TPC from October to December in Fig. 10.

Continuing the example in Figs. 11 and 12, the increased carbon-to-phosphorus ratio also had the effect of increasing the plankton assimilation of phosphorus and decreasing the assimilation of nitrogen. These changes in the nutrient fluxes contributed to the increase of the bias of the phosphate and nitrogen estimates at L4 (see Figs. 9 and 10). Thus, the assimilation impacts on processes were not necessarily all beneficial.

The model and assimilation estimates of fluxes are related to the complexity of the biogeochemical model in use, which raises the question, are the outputs "reproducing the data for the right reason"? (Friedrichs et al., 2007). Furthermore, energy and mass are often assumed uncertain and are not conserved by sequential assimilative methods (e.g., Hu et al., 2012; but see Hemmings et al., 2008; Losa et al., 2004; and Janjić et al., 2014 for alternative approaches). Thus assimilative changes of fluxes are typically not balanced among the simulated processes, making their interpretation more challenging. As a consequence, improving the match with the observations of state variables only appears a necessary rather than sufficient condition to claim reliability for the assimilative simulation of processes. These considerations lead one to stress once again the relevance of measuring multivariate time series to assess the performance and reliability of marine model and data assimilation systems, including measurements of

processes (e.g. zooplankton grazing, plankton nutrient uptakes) that can constrain biogeochemical flow simulations (Friedrichs et al., 2007).

Importantly, the corrections of biogeochemical fluxes imposed by state variable estimation can also help to identify weaknesses in marine ecosystem models. In our application, for example, the enhancement of the trophic shift was triggered by the systematic assimilative corrections of the reference carbon-to-nutrient ratio (Fig. 10). These corrections point to a "weakness" of the model and suggest that the parameterization and/or equation defining the phytoplankton cellular carbon-to-nutrient ratio needs to be improved in order to increase the plasticity of the modelled PFTs, i.e. their capacity to react to environmental changes. Interestingly, similar conclusions on the improvable representation of internal quotas in PFTs were obtained also when assimilating satellite chlorophyll (Ciavatta et al., 2011; Teruzzi et al., 2014). The more established application of assimilative parameter estimation for model identification and biogeochemical flux analysis in marine systems has been investigated by, e.g., Beck, 1975, 1987; Beck et al., 2002; Losa et al., 2004; Friedrichs et al., 2007; Simon et al., 2012; Xiao and Friedrichs, 2014.
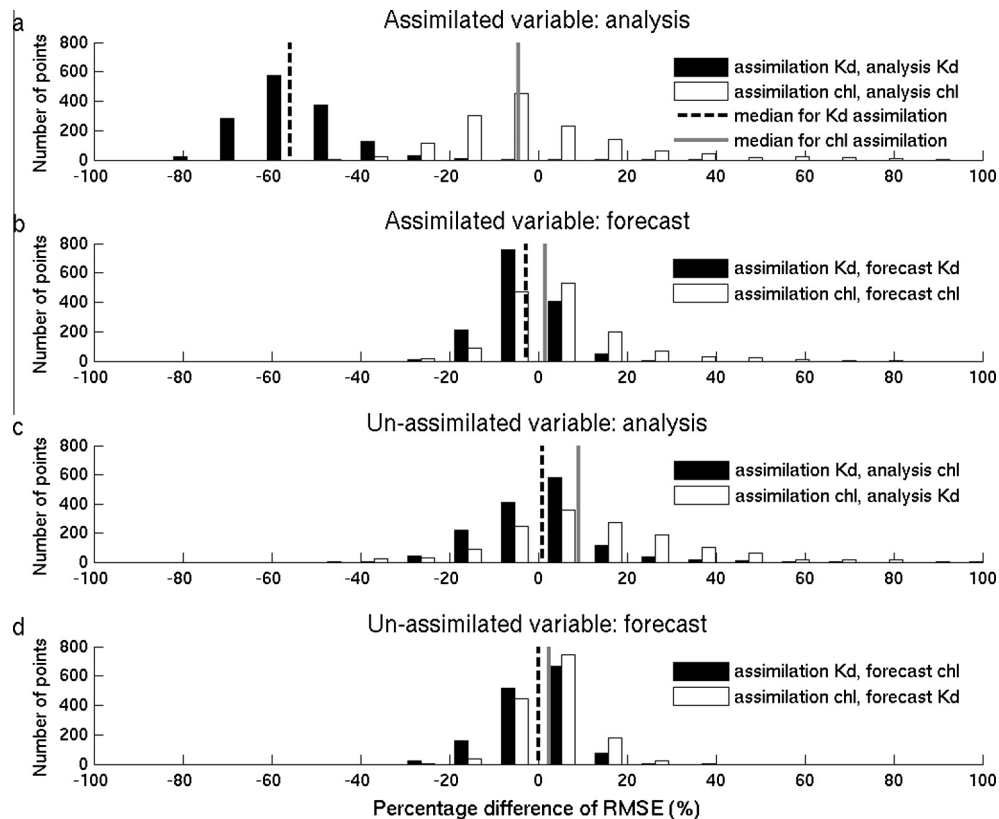
Comparing $K_d(443)$ and chlorophyll assimilation skill

The comparison of $K_d(443)$ assimilation skill with the skill of SeaWiFS chlorophyll assimilation provides further insights in the performance and potential advantages of optical data assimilation. A comprehensive discussion of chlorophyll assimilation skill in a similar model system can be found in Ciavatta et al. (2011).

Assimilation of $K_d(443)$ provided larger improvements of the reference estimates of the assimilated data, than chlorophyll assimilation. However, the skills of the two assimilation systems were comparable in the estimation of the un-assimilated ocean colour variable. Fig. 13 compares the performance of the two assimilation systems in improving the reference simulation of $K_d(443)$ and chlorophyll distributions. The histograms represent frequency distributions of RMSE difference between assimilation and reference. For $K_d(443)$, the distributions were computed from the maps of RMSE differences shown in Figs. 4–7. Fig. 13a shows that the distribution of the skill metric for $K_d(443)$ assimilation lies to the left of the distribution of the skill metric for chlorophyll. In particular, the median values of the two distributions are well far apart: $\sim-60\%$ for K_d *versus* ~-5 for chlorophyll. This means that the K_d analysis, in K_d assimilation, reduced the reference RMSE for K_d data, "more" than the chlorophyll analysis, in chlorophyll assimilation, reduced the reference RMSE for chlorophyll data. "More" here indicates both the overall magnitude of the assimilation improvements (indicated by the median of the distribution), as well as the number of grid points where improvements were obtained (indicated by the mode of the distribution: in ~600 grid points the RMSE decreased by $\sim-60\%$ when assimilating K_d).

Fig. 13b shows that also the forecasts of the assimilated variable (i.e. $K_d(443)$ one-week forecasts in $K_d(443)$ assimilation, and one-week chlorophyll forecasts in chlorophyll assimilation) were typically better for optical than chlorophyll assimilation, as indicated by positive and negative modes, respectively. However, in this case, the median values are comparable, indicating that the difference in the overall skill is not significant.

Finally, Fig. 13c and d show that the performance of the two assimilation systems is similar when considering the analysis and forecasts of the unassimilated variable, i.e. chlorophyll estimates in $K_d(443)$ assimilation, and $K_d(443)$ estimates in chlorophyll assimilation. Interestingly, both the assimilation systems did not improve significantly the reference estimates of unassimilated variables, as already observed for $K_d(443)$ assimilation in Figs. 6

Fig. 13. Comparison of the performance of $K_d(443)$ *versus* chlorophyll assimilation in improving the reference simulation of the satellite data. Frequency distributions of the assimilation-to-reference differences of RMSE, for four different cases: (a) the analysis outputs are compared with data and reference for the assimilated variable; (b) the forecasts outputs are compared with data and reference for the assimilated variable; (c) the analysis outputs are compared with data and reference for the un-assimilated variable; (d) the forecasts outputs are compared with data and reference for the un-assimilated variable. The median values of the frequency distributions are shown as vertical lines.

and 7. In fact, errors were in several cases increased, in particular by chlorophyll assimilation's analysis of $K_d(443)$ (Fig. 13c).

Assimilation of $K_d(443)$ provided analyses that are closer to the assimilated data (Fig. 13a) mainly because the observational error for the remotely sensed optical data was set lower than the error for chlorophyll: 20% *versus* 30–60% of the data values, respectively (Section 'Set-up of the assimilation scheme'). According to a well-known mechanism of Kalman filters (e.g. Brasseur, 2006), the EnKF analysis (Eq. (9)) produces stronger corrections of the assimilated variable towards the data when the observational error is lower (**R** in Eq. (9)), for equal values of the data-to-forecast mismatch (i.e. the innovation **d** in Eq. (8)). Starting from re-initializations closer to the assimilated data, one-week forecasts of the assimilated variable were typically better in $K_d(443)$ than chlorophyll assimilation (Fig. 13b). The relevance of the different observational errors of the two ocean colour products is highlighted also by the sensitivity experiment in Appendix C.

Assimilation of remotely sensed $K_d(443)$ performed better than chlorophyll assimilation in simulating biogeochemical and optical variables measured *in situ*. This is highlighted in Fig. 14, which shows the differences in the skill metrics of the two assimilation systems compared with *in situ* data at L4 (see also Tables B.2–B.4 in Appendix B). The skill metrics were computed removing an outlier analysis in May, that penalized heavily the performance of chlorophyll assimilation with respect to both the reference run and $K_d(443)$ assimilation (Tables B.2–B.4). Fig. 14 shows that $K_d(443)$ assimilation provided RMSE and percentage bias markedly lower than chlorophyll assimilation, with respect to a number of variables that is slightly larger (ten out eighteen), with some

exceptions, e.g. RMSE of phytoflagellates. On the other hand, the performances of the two assimilation systems appear largely comparable when considering the correlation metric: nine better cases for K_d assimilation *versus* eight for chlorophyll, and negligible differences in one case (silicate). Both assimilation systems deteriorated the reference metrics for some variables, e.g. all the metrics for the absorption of non-algal particles a_{NAP}, and RMSE and bias for ammonia. In these cases, Fig. 14 points out the assimilation that performed "less badly" with respect to the reference simulation.

One can argue that the better RMSE and bias skill of $K_d(443)$ assimilation might derive from the lower error of the remotely sensed optical data. The lower error implies that the $K_d(443)$ satellite data are closer to the "true values" in the ecosystem than the chlorophyll data. In principle, this allows the multivariate analysis to correct also the other model states towards their "true values". But this ideal interpretation is surely weakened by existing model bias and deficiencies, besides by un-fulfilled hypothesis of the assimilation scheme, e.g. linearity of the observational operator, which make the filter's estimates sub-optimal. However, interestingly, when testing chlorophyll assimilation with a 20% data error (see Appendix C), the estimates of the *in situ* data deteriorated markedly with respect to both the reference run and optical assimilation (not shown), similarly to what observed in Fig. C1.c for the forecast of the un-assimilated $K_d(443)$. This is a consequence of using an un-realistically low value for the chlorophyll error in case II waters (Saba et al., 2011), which forces un-realistic corrections of the other biogeochemical variables (Ciavatta et al., 2011).
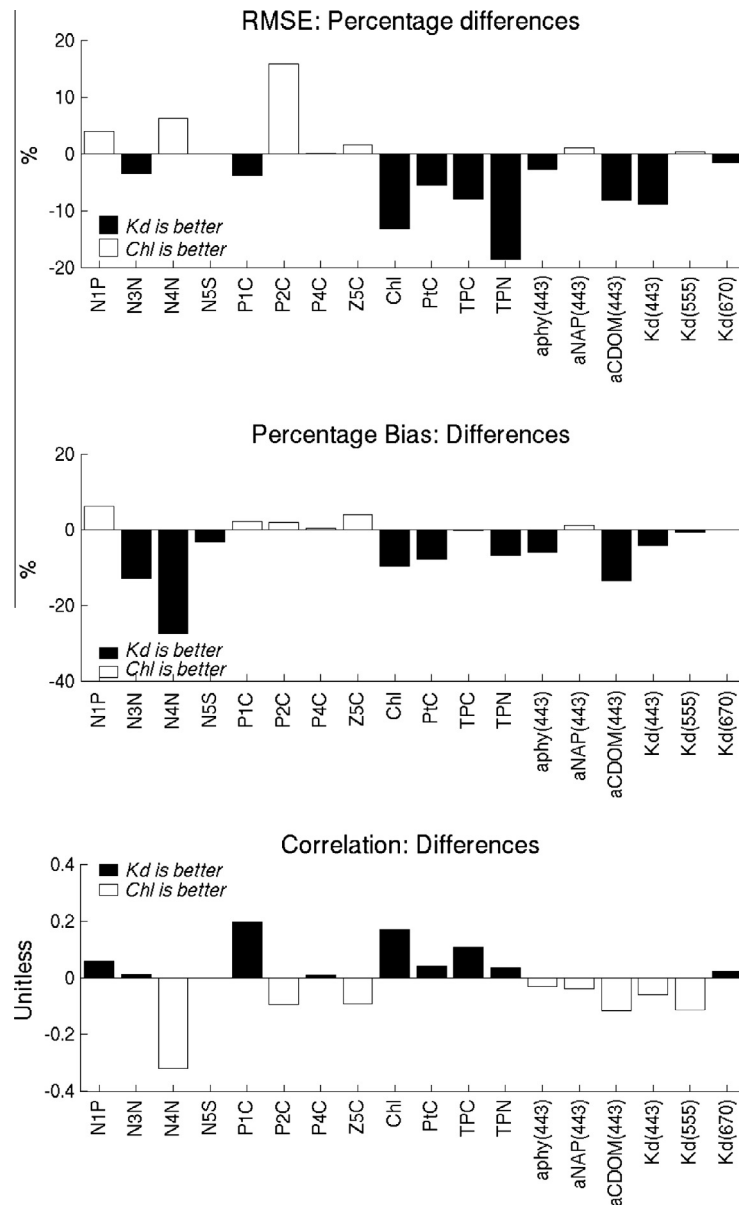
Fig. 14. Comparison of the skill metrics for $K_d(443)$ assimilation and chlorophyll assimilation in estimating the *in situ* data at L4. The bars represent the differences between the root mean square error (RMSE), percentage bias and correlation of the $K_d(443)$ assimilation *versus* chlorophyll (Chl) assimilation. Percentage values of RMSE were obtained by normalizing the differences with respect to the RMSE from chlorophyll assimilation. Black bars highlight when the metric values are better for $K_d(443)$ than for chlorophyll assimilation, i.e. lower RMSE and Pbias (negative differences) and higher correlation (positive differences) for $K_d(443)$ assimilation. The differences were computed using the "not May" metrics in Tables B.3 and B.4.

We believe also that $K_d(443)$ assimilation has a positive impact on a larger number of model variables, compared to chlorophyll assimilation, because $K_d(443)$ is a bulk property of marine ecosystems. In our model, this bulk property integrates the effect on light by seventeen optically active variables (Eqs. (A.1)–(A.12)), ranging from inorganic suspended solids to phytoplankton, while chlorophyll sums-up the contributions of the four phytoplankton types only. This means that the assimilation of K_d data can correct directly a larger number of model variable, crossing the whole trophic structure, when compared to chlorophyll assimilation. This is highlighted in Fig. 15, which compares the annual averages of the corrections to the state-vector imposed by the two assimilation systems. The corrections are more uniformly distributed among state variables in the case of $K_d(443)$ assimilation, than in the case of chlorophyll assimilation. In the latter case, the distribution of

the corrections has an exponential-like shape, with peak values for just the last nine variables, namely biomass and internal nutrient concentrations in diatom and microzooplankton. This confirms that $K_d(443)$ assimilation has a wider impact on the model structure than chlorophyll assimilation, which impacts mostly the plankton levels.

On the other hand, assimilating a bulk property of the ecosystem has potential limits if compared to chlorophyll assimilation. For example, it implies including further complexity and uncertainty in the model (e.g. the equations and parameterization of the bulk property A1–A9), as well it can rise methodological issues for the assimilation scheme. In particular, the observational operator (H in Eq. (7)) becomes non-linear with respect to the model variables when assimilating optical properties, while it is in general linear in chlorophyll assimilation. This would imply
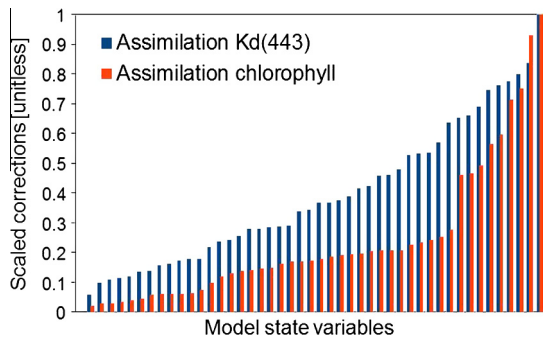
Fig. 15. Distributions of the corrections of the (forty-seven) biogeochemical variables in the state vector when assimilating $K_d(443)$ or chlorophyll. The corrections of the assimilated variables are not represented. Absolute values of the corrections were used in the computation of the annual averages. The averages were rescaled between the maximum and minimum values, separately for the two assimilation state vectors, giving values included in the range [0,1]. This implies that the "shapes" of the distributions of the correction in the two assimilations systems are comparable, but the absolute values of the corrections are not.

the use of approximated linearization in the analysis (Eq. (9)) in the Ensemble Kalman filter. Following Evensen (2003), we addressed this issue by augmenting the model state vector with K_d as a diagnostic variable that is directly comparable with the observations. In our application, this procedure provided satisfactory skill metrics when compared with the linear case of chlorophyll assimilation. However, we note that the procedure was found to perform poorly in other works (e.g. Jardak et al., 2010). This suggests that further benefits of K_d assimilation might be achieved with recently proposed methods that tackle non-linearities in model and observational operators, like for example particle filters (see e.g. van Leeuwen, 2010; Hoiet et al., 2012; Mattern et al., 2013).

Summary and conclusions

This work demonstrated, for the first time, that the assimilation of remotely-sensed diffuse attenuation coefficient into a complex shelf-sea model can improve the simulation of a large range of biogeochemical and optical variables. However, biases were introduced for some variables (nitrogen) at critical stages of the simulation (a trophic web shift). Discernible improvements in the estimates of the short-term (weekly) fluctuations were, in general, not obtained in our application.

We have shown that the assimilative state-estimation affects the simulation of emergent properties, such as the shift of the food web through the continuum pathways proposed by Legendre and Rassoulzadegan (1995). This shift was allowed by the complexity and plasticity of the model here applied, i.e. by the adaptability of the model functional groups to the changes in the variables imposed by the assimilation system. In particular, we recommend modelling variable carbon-to-chlorophyll and carbon-to-nutrients ratios not only for a more realistic representation of complex ecosystem dynamics (e.g. Blackford et al., 2004; Flynn, 2010), but also to exploit assimilation feedbacks for the simulation of ecosystem dynamic and biogeochemical budgeting. Such applications would be further improved by developments in mass balancing assimilative approaches for complex, non-linear systems (Hemmings et al., 2008; Losa et al., 2004; and Janjić et al., 2014), as well as by increased efforts in measuring ecosystem fluxes and processes in the oceans, useful to constrain and validate biogeochemical assimilation systems (Friedrichs et al., 2007).

Our application highlights model limitations that are current challenges in marine optics and ecosystem modelling: (i) the parameterization of the optical properties of the biogeochemical

compounds (see, e.g., Kirk, 1983), (ii) the simulation of the fate of inorganic suspended solids (Allen et al., 2007), (iii) the data scarcity for a reliable representation of terrestrial fluxes of biogeochemical and optical compounds (Bissett et al., 2005), (iv) the definition and parameterization of the phytoplankton functional groups (Anderson, 2005; Flynn et al., 2013). Another challenge is the assessment of subsurface biogeochemical estimates when assimilating surface ocean colour data (Fontana et al., 2013; Shulman et al., 2013). Further modelling and monitoring efforts are needed to cope with the above challenges.

A thorough comparison between $K_d(443)$ and chlorophyll assimilation was out of the scope of this work. However, the juxtaposition of the assimilation skill metrics from our assimilation experiments allows pointing out some advantages of assimilating the bulk optical property. Firstly, the lower error in the satellite K_d product, in particular in coastal case II waters (compare, e.g., Zhao et al., 2013, and Saba et al., 2011), represents an advantage, because it may drive the assimilation corrections closer to the "true values" of the assimilated variable in the system. Comparisons in our application were limited by our assumptions on the error values, but reliable pixel-by-pixel quantification of the uncertainty in the satellite optical products (Sathyendranath, 2014), as well as improvements in optical algorithms for case II waters (IOCCG, 2000), will support future inter-comparisons and data assimilation efforts. Secondly, light attenuation coefficient is a bulk property of marine systems, and its assimilation can drive direct corrections to a larger number of variables, crossing the whole trophic web model. However, where optics is driven mainly by the phytoplankton dynamics (i.e. in case I waters), the assimilation of chlorophyll data still represents a straightforward and effective tool to simulate marine ecosystems (see, e.g., Ford et al., 2012; Teruzzi et al., 2014).

Appendix A. The bio-optical module

The absorption by phytoplankton, $a_{phy}(\lambda)$ in Eq. (2), was computed by summing the contributions of the model phytoplankton types (P_i). These contributions are obtained by multiplying the chlorophyll contents of the PFTs (Chl$_i$, mgChl m^{-3}) by chlorophyll-specific absorption coefficients of each PFT, $a_{P_i}^*(\lambda)$ (Uitz et al., 2008):

$$a_{phy}(\lambda) = a_{P1}^*(\lambda)[Chl1] + a_{P2}^*(\lambda)[Chl2] + a_{P3}^*(\lambda)[Chl3] + a_{P4}^*(\lambda)[Chl4]$$
$$\text{(A.1)}$$

The chlorophyll-specific absorption coefficients of the PFTs (see Table A.1) were approximated by using the coefficients for micro (P1 and P4), nano (P2) and picoplankton (P3) estimated at the surface layer of the water column by Uitz et al. (2008). As a first approximation, Eq. (A.1) accounts for the packaging effect related to the average cell sizes and intracellular chlorophyll

Table A.1
Parameters of the bio-optical module; for each parameter, three values are given at three different wavelengths λ, when applicable.

Notation	Description	Unit	Value	λ (nm)
$a*_{P1}(\lambda)$[a]	Chlorophyll-specific spectral absorption coefficients of diatoms	$m^2 (mg\ Chl)^{-1}$	0.0155	443
			0.089	555
			0.0164	670
$a^*_{P2}(\lambda)$[a]	Chlorophyll-specific spectral absorption coefficients of phytoflagellates	$m^2 (mg\ Chl)^{-1}$	0.0904	443
			0.0084	555
			0.0285	670
$a^*_{P3}(\lambda)$[a]	Chlorophyll-specific spectral absorption coefficients of picophytoplankton	$m^2 (mg\ Chl)^{-1}$	0.1489	443
			0.0141	555
			0.0254	670
$a^*_{P4}(\lambda)$[a]	Chlorophyll-specific spectral absorption coefficients of dinoflagellates	$m^2 (mg\ Chl)^{-1}$	0.0155	443
			0.0089	555
			0.0164	670
$a^*_{det}(\lambda)$[b]	Carbon-specific absorptions coefficient of detritus	$m^2 (mg\ C)^{-1}$	9E−5	440
S_{det}[c]	Spectral slope for detritus absorption	nm^{-1}	0.0117	NA
$a^*_{PIM}(\lambda)$[d]	Mass-specific absorption coefficients of particulate inorganic matter	$m^2\ mg^{-1}$	5.0E−5	443
			2.8E−5	555
			2.1E−5	665
$a^*_{CDOM1}(\lambda)$[e]	Mass-specific absorptions of the semi-labile CDOM	$m^2 (mg\ C)^{-1}$	1.46E−3	410
$a^*_{CDOM2}(\lambda)$[e]	Mass-specific absorptions of the semi-labile CDOM	$m^2 (mg\ C)^{-1}$	5.08E−3	410
r_{CDOM1}[e]	Coloured fractions of the semi-labile CDOM	NA	0.055	NA
r_{CDOM2}[e]	Coloured fractions of refractory CDOM	NA	0.093	NA
S_{CDOM}[c]	Spectral slope for CDOM absorption	nm^{-1}	0.0174	NA
$a_{sw}(\lambda)$[f]	Sea water absorption	m^{-1}	6.96E−3	443
			5.96E−2	555
			4.39E−1	670
$b^*_{b,PIM}(\lambda)$[g]	Mass-specific backscattering of particulate inorganic matter	$m^2\ mg^{-1}$	5.7E−6	443
			8.6E−6	555
			6.4E−6	665
$b_{b,sw}(\lambda)$[h]	Sea water backscattering	m^{-1}	2.16E−3	443
			8.35E−4	555
			3.81E−4	670

NA: not applicable.
[a] Derived from Uitz et al. (2008): see Appendix A.
[b] Derived from Fujii et al. (2007) and Babin et al. (2003a,b).
[c] Babin et al. (2003a).
[d] Bowers and Binding (2006).
[e] Bissett et al. (1999).
[f] Pope and Fry (1997).
[g] Derived from Bowers and Binding, 2006.
[h] Zhang et al. (2009).

concentrations of the different PFTs (Duysens, 1956), but not for the contribution of the dynamic physiological adaptation of the PFTs to light and nutrients (see Fujii et al. (2007) and Shulman et al. (2013) for such an application).

Absorptions of non-algal particles, $a_{NAP}(\lambda)$, was computed by summing the contributions of organic detritus, $a_{det}(\lambda)$, and particulate inorganic material, $a_{PIM}(\lambda)$:

$$a_{NAP}(\lambda) = a_{det}(\lambda) + a_{PIM}(\lambda)$$
$$= a^*_{det}(\lambda)([R4C] + [R6C] + [R8C]) + a^*_{PIM}(\lambda)[PIM] \quad (A.2)$$

where R4C, R6C and R8C represent small, medium and large particulate organic matter ($mgC\ m^{-3}$), respectively, and PIM the particulate inorganic matter ($mg\ m^{-3}$).

The specific absorption coefficient $a^*_{det}(\lambda)$ in Eq. (A.2) was computed as (Fujii et al., 2007):

$$a^*_{det}(\lambda) = a^*_{det}(440) \exp[-S_{det}(\lambda - 440)] \quad (A.3)$$

The carbon-specific absorptions coefficient of detritus at λ = 440, $a^*_{det}(440)$, was computed following Fujii et al. (2007): The value of $a_{NAP}(443)/[SPM] = 0.035 \pm 0.018$ ($m^2\ g^{-1}$) (given for the English Channel in Table 5 in Babin et al., 2003a) was multiplied by a conversion coefficient SPM_{org}:POC = 2.6 g gC^{-1} (Babin et al., 2003b), leading to the value of $a^*_{det}(440)$ in Table A1. In Eq. (A.3), the spectral slope S_{det} was fixed to the value given for the English Channel in Table 4 in Babin et al. (2003a).

To compute the contribution of PIM, $a_{PIM}(\lambda)$, we multiplied the mass specific absorption coefficients for minerals measured in the Irish Sea by Bowers and Binding (2006), $a^*_{PIM}(\lambda)$ in Table A.1, by the PIM concentration approximated through the simulation of a passive tracer, as described in Section 'Set-up of the simulations'.

The absorption of coloured dissolved organic matter, $a_{CDOM}(\lambda)$, was decomposed in the contributions of the semi-labile and refractory components of dissolved organic carbon ($a_{CDOM1}(\lambda)$ and $a_{CDOM2}(\lambda)$, respectively) (Bissett et al., 1999, 2005):

$$a_{CDOM}(\lambda) = a_{CDOM1}(\lambda) + a_{CDOM2}(\lambda)$$
$$= a^*_{CDOM1}(\lambda)[CDOM1] + a^*_{CDOM2}(\lambda)[CDOM2] \quad (A.4)$$

$$a^*_{CDOM1}(\lambda) = a^*_{CDOM1}(410) \exp(-S_{CDOM}(\lambda - 410)) \quad (A.5)$$

$$a^*_{CDOM2}(\lambda) = a^*_{CDOM2}(410) \exp(-S_{CDOM}(\lambda - 410)) \quad (A.6)$$

$$[CDOM1] = r_{CDOM1}([R1C] + [R2C]) \quad (A.7)$$

$$[CDOM2] = r_{CDOM2}[DOC2] \quad (A.8)$$

The spectral, mass-specific absorptions of the semi-labile and refractory components of CDOM ($a^*_{CDOM1}(\lambda)$ and $a^*_{CDOM2}(\lambda)$ respectively) were taken from Bissett et al. (1999) (see Table A.1), while the spectral slope S_{CDOM} was set equal to the value estimated in the English Channel by Babin et al. (2003a). The coloured fractions of the semi-labile and refractory DOC (CDOM1 and CDOM2, respectively), were computed using the ratios r_{CDOM1} and r_{CDOM2} given by Bissett et al. (1999). ERSEM simulates the labile

Table B.1
The variables monitored weekly at the long term monitoring station L4 used for skill assessment in the present work. The unit of measurement and the notation used in the manuscript to refer to the variables are also reported.

Variable	Notation	Unit
Phosphate	N1P	mmol m^{-3}
Nitrate	N3 N	mmol m^{-3}
Ammonia	N4 N	mmol m^{-3}
Silicate	N5S	mmol m^{-3}
Diatoms	P1C	mgC m^{-3}
Phytoflagellates	P2C	mgC m^{-3}
Dinoflagellates	P4C	mgC m^{-3}
Microzooplankton	Z5C	mgC m^{-3}
Total chlorophyll	Chl	mgChl m^{-3}
Total phytoplankton carbon	PtC	mgC m^{-3}
Total particulate carbon	TPC	mgC m^{-3}
Total particulate nitrogen	TPN	mmol m^{-3}
Absorption by phytoplankton	$a_{phy}(443)$	m^{-1}
Absorption by nonalgal particles	$a_{NAP}(443)$	m^{-1}
Absorption by CDOM[1]	$a_{CDOM}(443)$	m^{-1}
Diffuse attenuation coefficient in the blue	$K_d(443)$	m^{-1}
Diffuse attenuation coefficient in the green	$K_d(555)$	m^{-1}
Diffuse attenuation coefficient in the red	$K_d(670)$	m^{-1}

(R1C, mgC m^{-3}) and semi-labile (R2C, mgC m^{-3}) DOC, but not its refractory component DOC2, which was approximated through the simulation of a passive tracer, as described in Section 'Set-up of the simulations'

The backscattering of organic particles in Eq. (3) was decomposed in the contribution of small (POC1) and large (POC2) organic particles (Fujii et al., 2007):

$$b_{b,POC1}(\lambda) = \left(\frac{[POC1]}{476935.8}\right)^{\frac{1}{1.277}} \cdot \left(\frac{\lambda}{510}\right)^{-0.5} \tag{A.9}$$

$$b_{b,POC2}(\lambda) = \left(\frac{[POC2]}{17069.0}\right)^{\frac{1}{0.859}} \tag{A.10}$$

where

$$[POC1] = [P3C] + [R4C] \tag{A.11}$$

$$[POC2] = [P1C] + [P2C] + [P4C] + [R6C] + [R8C] \tag{A.12}$$

where P1C, P2C, P3C and P4C are the carbon biomasses (mgC m^{-3}) of the model PFTs.

Table B.2
Skill metrics of reference model in simulating the biogeochemical and optical data measured weekly at station L4; root mean square error (RMSE), percentage bias (Pbias) and correlation (r) of the model reference estimates of the data. The notation of the variables is explained in Table B.1. The metrics are reported separately for the case where all the 2006 data are included in the computation (left side of the table), and where one model-to-data matchup for 8th May 2006 is excluded from the computation (right side).

Variables	RMSE 2006	Pbias 2006	r 2006	RMSE not May	Pbias not May	r not May
N1P (mmol m^{-3})	0.21	−41.80	0.56	0.21	−40.58	0.58
N3 N (mmol m^{-3})	1.95	0.03	0.76	1.89	3.53	0.79
N4 N (mmol m^{-3})	0.65	21.06	0.15	0.66	22.94	0.15
N5S (mmol m^{-3})	1.83	−7.49	0.48	1.84	−6.29	0.48
P1C (mgC m^{-3})	35.86	87.27	−0.04	36.00	83.97	−0.04
P2C (mgC m^{-3})	9.50	−7.08	0.56	8.71	−12.58	0.60
P4C (mgC m^{-3})	24.08	−81.63	0.15	24.38	−82.42	0.16
Z5C (mgC m^{-3})	29.19	−57.17	0.04	29.55	−58.33	0.04
Chl (mgChl m^{-3})	1.42	29.59	−0.02	1.37	24.12	0.02
PtC (mgC m^{-3})	52.05	52.59	0.17	50.88	47.62	0.19
TPC (mgC m^{-3})	143.02	−37.18	0.25	144.81	−38.49	0.27
TPN (mmol m^{-3})	1.15	4.98	0.38	1.12	2.78	0.40
$a_{phy}(443)$ (m^{-1})	0.09	27.30	0.17	0.08	22.07	0.22
$a_{NAP}(443)$ (m^{-1})	0.02	15.52	0.31	0.02	13.90	0.32
$a_{CDOM}(443)$ (m^{-1})[a]	0.04	−29.92	−0.16	0.04	−30.31	−0.16
$K_d(443)$ (m^{-1})	0.15	43.21	−0.06	0.15	43.21	−0.06
$K_d(555)$ (m^{-1})	0.07	24.65	−0.25	0.07	24.65	−0.25
$K_d(670)$ (m^{-1})	0.12	5.63	0.03	0.12	5.63	0.03

[a] The values of a_{CDOM} from January to March (included) were not used to compute the skill metrics (Section 'Set-up of the simulations').

Table B.3
Skill metrics of $K_d(443)$ assimilation in simulating the biogeochemical and optical data measured weekly at station L4; Root mean square error (RMSE), percentage bias (Pbias) and correlation (r) of the assimilation estimates of the data. The notation of the variables is explained in Table B.1. The metrics are reported separately for the case where all the 2006 data are included in the computation (left side of the table), and where one model-to-data matchup for 8th May 2006 is excluded from the computation.

Variables	RMSE 2006	Pbias 2006	r 2006	RMSE not May	Pbias not May	r not May
N1P (mmol m^{-3})	0.22	−45.15	0.61	0.21	−44.01	0.63
N3 N (mmol m^{-3})	1.99	2.70	0.75	1.94	6.09	0.78
N4 N (mmol m^{-3})	0.72	37.02	0.10	0.73	37.84	0.10
N5S (mmol m^{-3})	1.76	−2.23	0.52	1.77	−0.64	0.54
P1C (mgC m^{-3})	33.76	76.42	0.05	33.92	73.38	0.05
P2C (mgC m^{-3})	8.33	−26.66	0.55	8.31	−29.32	0.57
P4C (mgC m^{-3})	24.16	−83.59	0.15	24.46	−84.43	0.17
Z5C (mgC m^{-3})	29.24	−52.45	0.03	29.60	−54.06	0.03
Chl (mgChl m^{-3})	1.14	12.39	0.14	1.10	7.79	0.19
PtC (mgC m^{-3})	46.50	30.18	0.14	45.81	26.04	0.16
TPC (mgC m^{-3})	129.21	−35.76	0.40	130.92	−36.79	0.42
TPN (mmol m^{-3})	1.01	9.01	0.51	0.97	7.05	0.53
$a_{phy}(443)$ (m^{-1})	0.07	1.88	0.21	0.07	−2.46	0.28
$a_{NAP}(443)$ (m^{-1})	0.02	17.88	0.16	0.02	16.62	0.14
$a_{CDOM}(443)$ (m^{-1})[a]	0.03	−10.55	0.00	0.03	−9.51	−0.01
$K_d(443)$ (m^{-1})	0.13	39.11	0.07	0.13	39.11	0.07
$K_d(555)$ (m^{-1})	0.07	24.03	−0.26	0.07	24.03	−0.26
$K_d(670)$ (m^{-1})	0.11	4.86	0.02	0.11	4.86	0.02

[a] The values of a_{CDOM} from January to March (included) were not used to compute the skill metrics (Section 'Set-up of the simulations').

Table B.4

Skill metrics of chlorophyll assimilation in simulating the biogeochemical and optical data measured weekly at station L4; root mean square error (*RMSE*), percentage bias (*Pbias*) and correlation (*r*) of the model reference estimates of the data. The notation of the variables is explained in Table B.1. The metrics are reported separately for the case where all the 2006 data are included in the computation (left side of the table), and where one model-to-data matchup for 8th May 2006 is excluded from the computation.

| Variables	RMSE 2006	Pbias 2006	r 2006	RMSE not May	Pbias not May	r not May
N1P (mmol m^{-3})	0.21	−39.65	0.54	0.21	−37.92	0.58
N3N (mmol m^{-3})	2.09	14.43	0.72	2.02	19.05	0.77
N4N (mmol m^{-3})	0.68	61.89	0.41	0.68	65.33	0.43
N5S (mmol m^{-3})	1.78	1.58	0.51	1.77	3.97	0.54
P1C (mgC m^{-3})	50.83	110.78	−0.07	35.25	71.18	−0.15
P2C (mgC m^{-3})	7.41	−23.97	0.62	7.17	−27.49	0.66
P4C (mgC m^{-3})	24.16	−83.10	0.14	24.46	−84.11	0.16
Z5C (mgC m^{-3})	29.01	−43.39	0.09	29.13	−50.18	0.12
Chl (mgChl m^{-3})	1.99	37.41	−0.12	1.27	17.50	0.02
PtC (mgC m^{-3})	63.72	51.24	0.04	48.50	33.89	0.12
TPC (mgC m^{-3})	145.79	−33.48	0.24	142.12	−37.05	0.31
TPN (mmol m^{-3})	1.55	20.45	0.39	1.20	13.96	0.49
a_{phy}(443) (m^{-1})	0.08	14.17	0.22	0.07	8.46	0.31
a_{NAP}(443) (m^{-1})	0.02	17.00	0.18	0.02	15.47	0.18
a_{CDOM}(443) (m^{-1})[a]	0.04	−23.68	0.11	0.04	−23.08	0.11
K_d(443) (m^{-1})	0.14	43.35	0.13	0.14	43.35	0.13
K_d(555) (m^{-1})	0.07	24.70	−0.15	0.07	24.70	−0.15
K_d(670) (m^{-1})	0.11	4.91	−0.003	0.11	4.91	−0.003

[a] The values of a_{CDOM} from January to March (included) were not used to compute the skill metrics (Section 'Set-up of the simulations').

The spectral backscattering of particulate inorganic matter, $b_{b,PIM}(\lambda)$ in Eq. (3), was computed by multiplying the mass specific scattering coefficients and backscattering ratio given by Bowers and Binding (2006) for minerals in the Irish sea (see Table A.1), by the PIM concentration.

Finally, literature values were used to define sea-water absorption, $a_{sw}(\lambda)$ in Eq. (2) (Pope and Fry, 1997), and backscattering, $b_{b,sw}(\lambda)$ in Eq. (3) (Zhang et al., 2009): see Table A.1.

Appendix B. Skill metrics for K_d(443) and chlorophyll assimilation at L4

Details of the skill of the reference run, K_d(443) assimilation, and chlorophyll assimilation in simulating the *in situ* data collected at the monitoring station L4 (see list in Table B.1) are given in Tables B.2–B.4. The tables present the values of the root mean square error, percentage bias and correlations when all the *in situ*
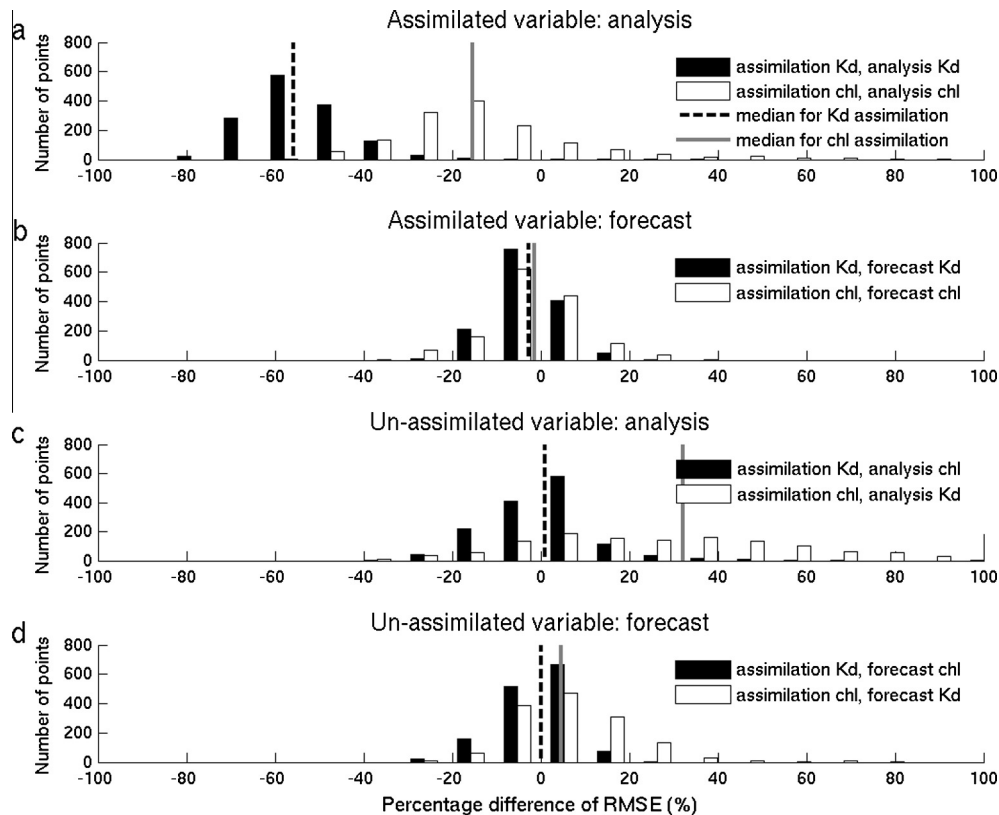
Fig. C.1. Results of the sensitivity experiment for chlorophyll assimilation, using an observational error equal to 20% of the concentration values. Comparison of the performance of K_d(443) *versus* chlorophyll assimilation in improving the reference simulation of the satellite data (analogous to Fig. 13).

data are included in the computation of the metrics (left parts of the tables), as well as when removing the data in correspondence of one chlorophyll analysis (8th May), which produced outliers values for the variables representing the plankton community (right part of the tables). For the sake of comparison, the table of the reference output shows the skill metrics without the May outlier, along with the metrics computed from the whole data set and shown already in Table 1.

Appendix C. Sensitivity experiment for chlorophyll assimilation

A sensitivity experiment was performed for chlorophyll assimilation, by lowering the percentage observational error to 20% of the chlorophyll concentration, i.e. the same percentage error applied in $K_d(443)$ assimilation (see Section 'Set-up of the assimilation scheme'. The results confirmed that the different observational errors of the two ocean colour products have a relevant influence on the performance of the assimilation systems. Lowering the chlorophyll observation error reduced the simulation error for the assimilated chlorophyll data. However, it also deteriorated the simulation of the unassimilated variables. This is shown in Fig. C.1 for the ocean colour data. The decrease of the reference RMSE in the chlorophyll analysis (Fig. C.1a) was more remarkable than in Fig. 13a. Nevertheless, improvements remain lower than the ones imposed by $K_d(443)$ analysis to the assimilated variable. This follows the fact that the absolute errors of $K_d(443)$ data were still lower than the ones of chlorophyll, because the same 20% relative error was applied to data that are different by one order of magnitude, on average. Interestingly, in Fig. C.1c the $K_d(443)$ analysis by chlorophyll assimilation deteriorated the model estimates of the optical data.

Also the simulation of the *in situ* data at station L4 resulted severely deteriorated (not shown), with several variables that diverged to unobserved high values (e.g. ammonia and silicate). Analogous results were obtained in Ciavatta et al. (2011) when setting the chlorophyll observational error to 35%, in a comparable data assimilation system. The deteriorated simulation of the unassimilated variable follows the application of an error that is unrealistically low for chlorophyll in case II waters in general (e.g. Saba et al., 2011) and in the western English Channel in particular (Ciavatta et al., 2011). The low error implies heavy corrections of the state variables, to match chlorophyll data that are actually highly uncertain. This produces outlier values and instabilities in the assimilative simulation, and eventually deteriorates the reference skill for the unassimilated variables (Ciavatta et al., 2011).

References

Allen, J.I., Eknes, M., Evensen, G., 2003. An ensemble Kalman filter with a complex marine ecosystem model: hindcasting phytoplankton in the Cretan Sea. Annales Geophysicae 21 (1), 399–411.

Allen, J., Holt, J.T., Blackford, J., Proctor, R., 2007. Error quantification of a high-resolution coupled hydrodynamic-ecosystem coastal-ocean model: Part 2. Chlorophyll-a, nutrients and SPM. Journal of Marine Systems 68 (3), 381–404.

Anderson, T.R., 2005. Plankton functional type modelling: running before we can walk? Journal of Plankton Research 27 (11), 1073–1081.

Babin, M., Stramsky, D., Ferrari, G.M., Claustre, H., Bricaud, A., Oboloensky, G., Hoepffner, N., 2003a. Variations in the light absorption coefficients of phytoplankton, nonalgal particles, and dissolved organic matter in coastal waters around Europe. Journal of Geophysical Research 108 (C7), 3211.

Babin, M., Morel, A., Fournier-Sicre, V., Fell, F., Stramski, D., 2003b. Light scattering properties of marine particles in coastal and open ocean waters as related to the particle mass concentration. Limnology and Oceanography 48 (2), 843–859.

Baretta, J.W., Ebenhöh, W., Ruardij, P., 1995. The European regional seas ecosystem model, a complex marine ecosystem model. Netherlands Journal of Sea Research 33 (3), 233–246.

Baretta-Bekker, J.G., Baretta, J.W., Ebenhöh, W., 1997. Microbial dynamics in the marine ecosystem model ERSEM II with decoupled carbon assimilation and nutrient uptake. Journal of Sea Research 38, 195–211.

Beck, M.B., 1975. The identification of algal population dynamics in a freshwater stream. Computer Simulation of Water Resources Systems, G.C, 483–494.

Beck, M.B., 1987. Water-quality modeling – a review of the analysis of uncertainty. Water Resources Research 23 (8), 1393–1442.

Beck, M.B., Stigter, J.D., Lloyd Smith, D., 2002. Elasto-plastic deformation of structure. In: Beck, M.B. (Ed.), Environmental Foresight and Models: A Manifesto. Elsevier, Oxford, UK, pp. 323–350.

Bertino, L., Evensen, G., Wackernagel, H., 2003. Sequential data assimilation techniques in oceanography. International Statistical Review 71 (2), 223–241.

Brasseur, P., 2006. Ocean data assimilation using sequential methods based on the Kalman filter – from theory to practical implementations. In: Chassignet, E.P., Verron, J. (Eds.), Ocean Weather Forecasting: An Integrated View of Oceanography. Springer, Dordrecht, Netherlands, pp. 271–316.

Bissett, W.P., Carder, K.L., Walsh, J.J., Dieterle, D.A., 1999. Carbon cycling in the upper waters of the Sargasso Sea: II. Numerical simulation of apparent and inherent optical properties. Deep-Sea Research I 46, 271–317.

Bissett, W.P., Arnone, R., DeBra, S., Dieterle, D.A., Dye, D., Kirkpatrick, G.J., Schofield, O.M., Vargo, G.A., 2005. Predicting the optical properties of the West Florida Shelf: resolving the potential impacts of a terrestrial boundary condition on the distribution of coloured dissolved and particulate matter. Marine Chemistry 95 (3–4), 199–233.

Blackford, J.C., Allen, J.I., Gilbert, F.J., 2004. Ecosystem dynamics at six contrasting sites: a generic modelling study. Journal of Marine Systems 52 (1–4), 191–215.

Bowers, D.G., Binding, C.E., 2006. The optical properties of mineral suspended particles: a review and synthesis. Estuarine, Coastal and Shelf Science 67 (1–2), 219–230.

Ciavatta, S., Torres, R., Saux-Picart, S., Allen, I.J., 2011. Can ocean colour assimilation improve biogeochemical hindcasts in shelf seas? Journal of Geophysical Research 116, C12043.

Duysens, L.M.N., 1956. The flattering effect of the absorption spectra of suspensions as compared to that of solutions. Biochimica et Biophysica Acta 19, 1–12.

Eloire, D., Somerfield, P.J., Conway, D.V.P., Halsband-Lenk, C., Harris, R., Bonnet, D., 2010. Temporal variability and community composition of zooplankton at station L4 in the Western Channel: 20 years of sampling. Journal of Plankton Research 32 (5), 657–679.

Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics. Journal of Geophysical Research 99 (C5), 10143–10162.

Evensen, G., 2003. The Ensemble Kalman filter: theoretical formulation and practical implementation. Ocean Dynamics 53 (4), 343–367.

Fileman, E., Petropavlovsky, A., Harris, R., 2010. Grazing by the copepods Calanus helgolandicus and Acartia clausi on the protozooplankton community at station L4 in the Western English Channel. Journal of Plankton Research 32 (5), 709–724.

Flynn, K.J., 2010. Ecological modelling in a sea of variable stoichiometry: dysfunctionality and the legacy of Redfield and Monod. Progress in Oceanography 84, 52–65.

Flynn, K.J., Stoecker, D.K., Mitra, A., et al., 2013. Misuse of the phytoplankton–zooplankton dichotomy, the need to assign organisms as mixotrophs within plankton functional types. Journal of Plankton Research 35, 3–11.

Fontana, C., Brasseur, P., Brankart, J.M., 2013. Toward a multivariate reanalysis of the North Atlantic ocean biogeochemistry during 1998–2006 based on the assimilation of SeaWiFS chlorophyll data. Ocean Science 9, 37–56.

Ford, D.A., Edwards, K.P., Lea, D., Barciela, R.M., Martin, M.J., Demaria, J., 2012. Assimilating GlobColour ocean colour data into a pre-operational physical–biogeochemical model. Ocean Science Discussions 9, 687–744.

Friedrichs, M.A.M., Hood, R.R., Wiggert, J.D., 2006. Ecosystem model complexity versus physical forcing: quantification of their relative impact with assimilated Arabian Sea data. Deep Sea Research Part II: Topical Studies in Oceanography 53 (5–7), 576–600.

Friedrichs, M.A.M. et al., 2007. Assessment of skill and portability in regional marine biogeochemical models: role of multiple planktonic groups. Journal of Geophysical Research 112, C08001.

Fujii, M., Boss, E., Chai, F., 2007. The value of adding optics to ecosystem models: a case study. Biogeosciences 4, 817–835.

Geider, R.J., La Roche, J., 2002. Redfield revisited: variability of C:N: P in marine microalgae and its biochemical basis. European Journal of Phycology 37, 1–17.

Geider, R.J., MacIntyre, H.L., Kana, T.M., 1997. Dynamic model of phytoplankton growth and acclimation: responses of the balanced growth rate and the chlorophyll a:carbon ratio to light, nutrient-limitation and temperature. Marine Ecology Progress Series 148 (1–3), 187–200.

Gregg, W.W., 2008. Assimilation of SeaWiFS ocean chlorophyll data into a three-dimensional global ocean model. Journal of Marine Systems 69, 205–225.

Gregg, W.W., Carder, K.L., 1990. A simple spectral solar irradiance model for cloudless maritime atmospheres. Limnology and Oceanography 35 (8), 1657–1675.

Gregg, W.W., Walsh, J.J., 1992. Simulation of the 1979 spring bloom in the mid-Atlantic bight a coupled physical/biological/optical model. Journal of Geophysical Research 97 (C4), 5723–5743.

Gregg, W.W., Friedrichs, M.A.M., Robinson, A.R., Rose, K.A., Schlitzer, R., Thompson, K.R., Doney, S.C., 2009. Skill assessment in ocean biological data assimilation. Journal of Marine Systems 76 (1–2), 16–33.

Groom, S., Martinez-Vicente, V., Fishwick, J., Tilstone, G., Moore, G., Smyth, T., Harbour, D., 2009. The Western English Channel observatory: optical characteristics of station L4. Journal of Marine Systems 77 (3), 278–295.

Hamill, T.M., Whitaker, J.S., 2005. Accounting for the error due to unresolved scales in ensemble data assimilation: a comparison of different approaches. Monthly Weather Review 133 (11).

Hemmings, J.C., Barciela, R.M., Bell, M.J., 2008. Ocean colour data assimilation with material conservation for improving model estimates of air-sea CO_2 flux. Journal of Marine Research 66 (1), 87–126.

Holt, J.T., James, I.D., 2001. An s-coordinate density evolving model of the northwest European continental shelf – 1, Model description and density structure. Journal of Geophysical Research 106 (C7), 14015–14034.

Hoteit, I., Triantafyllou, G., Petihakis, G., Allen, J.I., 2003. A singular evolutive extended Kalman filter to assimilate real in situ data in a 1-D marine ecosystem model. Annales Geophysicae 21, 389–397.

Hoteit, I., Triantafyllou, G., Petihakis, G., 2004. Towards a data assimilation system for the Cretan Sea ecosystem using a simplified Kalman filter. Journal of Marine Systems 45 (3), 159–171.

Hoteit, I., Triantafyllou, G., Petihakis, G., 2005. Efficient data assimilation into a complex, 3-D physical–biogeochemical model using partially-local Kalman filters. Annales Geophysicae 23 (10), 3171–3185.

Hoteit, I., Luo, X., Pham, D.T., 2012. Particle Kalman filtering: a nonlinear bayesian framework for Ensemble Kalman filters, 2012. Monthly Weather Review 140, 2.

Hu, J., Fennel, K., Mattern, J.P., Wilkin, J., 2012. Data assimilation with a local Ensemble Kalman filter applied to a three-dimensional biological model of the Middle Atlantic Bight. Journal of Marine Systems 94, 145–156.

IOCCG, 2000. Remote sensing of ocean colour in coastal and other optically complex waters. In: Sathyendranath, S. (Ed.), Reports of the International Ocean Colour Coordinating Group. No. 3, Dartmouth, Canada, 140p.

Janjić, T., McLaughlin, D., Cohn, S.E., Verlaan, M., 2014. Conservation of mass and preservation of positivity with ensemble-type Kalman filter algorithms. Monthly Weather Review 142 (2).

Jardak, M., Navon, I.M., Zupanski, M., 2010. Comparison of sequential data assimilation methods for the Kuramoto–Sivashinsky equation. International Journal for Numerical Methods in Fluids 62, 374–402.

Kepert, J.D., 2004. On ensemble representation of the observation-error covariance in the Ensemble Kalman filter. Ocean Dynamics 54 (6), 561–569.

Kirk, J.T.O., 1983. Light and Photosynthesis in Aquatic Ecosystems. Cambridge University Press, Cambridge, UK, 509p.

Korres, G., Triantafyllou, G., G. Petihakis, G., Raitsos, D.E., Hoteit, I., Pollani, A., Colella, S., Tsiaras. K, 2012. A data assimilation tool for the Pagasitikos Gulf ecosystem dynamics: methods and benefits. Journal of Marine Systems 94, S102–S117.

Lee, Z.-P., Du, K.-P., Arnone, R., 2005a. A model for the diffuse attenuation coefficient of downwelling irradiance. Journal of Geophysical Research 110, C02016.

Lee, Z.-P., Darecki, M., Carder, K.L., Davis, C.O., Stramski, D., Rhea, W.J., 2005b. Diffuse attenuation coefficient of downwelling irradiance: an evaluation of remote sensing methods. Journal of Geophysical Research 110, C02017.

Legendre, L., Rassoulzadegan, F., 1995. Plankton and nutrient dynamics in marine waters. Ophelia 41 (1), 153–172.

Lewis, K., Allen, J.I., 2009. Validation of a hydrodynamic–ecosystem model simulation with time-series data collected in the western English Channel. Journal of Marine Systems 77 (3), 296–311.

Losa, S., Kivman, G., Ryabchenko, V., 2004. Weak constraint parameter estimation for a simple ocean ecosystem model: what can we learn about the model and data? Journal of Marine Systems 45 (1–2), 1–20.

Luo, X., Hoteit, I., 2011. Robust ensemble filtering and its relation to covariance inflation in the Ensemble Kalman filter. Monthly Weather Review 139, 12.

Manizza, M., Le Quéré, C., Watson, A.J., Buitenhuis, E.T., 2005. Bio-optical feedbacks among phytoplankton, upper ocean physics and sea-ice in a global model. Geophysical Research Letters 32, L05603.

Maritorena, S., d'Andon, O.H.F., Mangin, A., Siegel, D.A., 2010. Merged satellite ocean color data products using a bio-optical model: Characteristics, benefits and issues. Remote Sensing of Environment 114 (8), 1791–1804.

Mattern, J.P., Dowd, M., Fennel, K., 2013. Particle filter-based data assimilation for a three-dimensional biological ocean model and satellite observations. Journal of Geophysical Research: Oceans 118 (5), 2746–2760.

Mouw, C.B., Yoder, J.A., Doney, S.C., 2012. Impact of phytoplankton community size on a linked global ocean optical and ecosystem model. Journal of Marine Systems 89 (1), 61–75.

Natvik, L.J., Evensen, G., 2003. Assimilation of ocean colour data into a biochemical model of the North Atlantic – Part 1. Data assimilation experiments. Journal of Marine Systems 40, 127–153.

Nerger, L., Gregg, W.W., 2008. Improving assimilation of SeaWiFS data by the application of bias correction with a local SEIK filter. Journal of Marine Systems 73 (1–2), 87–102.

Nerger, L., Hiller, W., Schroter, J., 2005. A comparison of error subspace Kalman filters. Tellus Series A 57 (5), 715–735.

Ogawa, H., Tanoue, E., 2003. Dissolved organic matter in oceanic waters. Journal of Oceanography 59 (2), 129–147.

Polimene, L., Brunet, C., Butenschon, M., Martinez-Vicente, V., Widdicombe, C., Torres, R., Allen, J.I., 2014. Modelling a light-driven phytoplankton succession. Journal of Plankton Research 36 (1), 214–229.

Pope, R.M., Fry, E.S., 1997. Absorption spectrum (380–700 nm) of pure water, II. Integrating cavity measurements. Applied Optics 36, 8710–8723.

Reed, R.K., 1977. On estimating insolation over the ocean. Journal of Physical Oceanography 7, 482–485.

Saba, V.S. et al., 2011. An evaluation of ocean colour model estimates of marine primary productivity in coastal and pelagic regions across the globe. Biogeosciences 8 (2), 489–503.

Sathyendranath, S., 2014. Climate Assessment Report (CAR), Ocean Colour Climate Change Initiative (OC_CCI) – Phase One, Technical Document AO-1/6207/09/I-LG, European Space Agency.

Sheskin, D.J., 2011. Handbook of Parametric and Nonparametric Statistical Procedures, fifth ed. CRC Press, 1880p.

Shulman, I., Frolov, S., Anderson, S., Penta, B., Gould, R., Sakalaukus, P., Ladner, S., 2013. Impact of bio-optical data assimilation on short-term coupled physical, bio-optical model predictions. Journal of Geophysical Research: Oceans 118 (4), 2215–2230.

Siddorn, J.R., Allen, J.I., Blackford, J.C., Gilbert, F.J., Holt, J.T., Holt, M.W., Osborne, J.P., Proctor, R., Mills, D.K., 2007. Modelling the hydrodynamics and ecosystem of the North-West European continental shelf for operational oceanography. Journal of Marine Systems 65 (1–4), 417–429.

Simon, E., Bertino, L., 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 Science 5 (4), 495–510.

Simon, E., Samuelsen, A., Bertino, L., Dumont, D., 2012. Estimation of positive sum-to-one constrained zooplankton grazing preferences with the DEnKF: a twin experiment. Ocean Science 8 (4).

Smyth, T.J., Fishwick, J.R., Al-Moosawi, L., Cummings, D.G., Harris, C., Kitidis, V., Rees, A., Martinez-Vicente, V., Woodward, E.M.S., 2010. A broad spatio-temporal view of the Western English Channel observatory. Journal of Plankton Research 32 (5), 585–601.

Southward, A.J. et al., 2005. Long-term oceanographic and ecological research in the western English Channel. Advances in Marine Biology 47, 1–105.

Storto, A., Masina, S., Dobricic, S., 2013. Ensemble spread-based assessment of observation impact: application to a global ocean analysis system. Quarterly Journal of the Royal Meteorological Society 139 (676), 1842–1862.

Stow, C.A., Jolliff, J., McGillicuddy, D.J., Doney, S.C., Allen, J.I., M. Friedrichs, M.A., Rose, K.A., Wallheadg, P., 2009. Skill assessment for coupled biological/physical models of marine systems. Journal of Marine Systems 76 (1–2), 4–15.

Sun, A.Y., Morris, A., Mohanty, S., 2009. Comparison of deterministic Ensemble Kalman filters for assimilating hydrogeological data. Advances in Water Resources 32 (2), 280–292.

Teruzzi, A., Dobricic, S., Solidoro, C., Cossarini, G., 2014. A 3-D variational assimilation scheme in coupled transport–biogeochemical models: forecast of Mediterranean biogeochemical properties. Journal of Geophysical Research: Oceans 119, 200–217.

Thingstad, T.F., Rassoulzadegan, F., 1999. Conceptual models for the biogeochemical role of the photic zone microbial food web, with particular reference to the Mediterranean Sea. Progress in Oceanography 44, 271–286.

Tilstone, G.H., Peters, S.W.M., van der Woerd, H.J., et al., 2012. Variability in specific-absorption properties and their use in a semi-analytical Ocean colour algorithm for MERIS in North Sea and Western English Channel coastal waters. Remote Sensing of Environment 118, 320–338.

Torres, R., Allen, J.I., Figueiras, F.G., 2006. Sequential data assimilation in an upwelling influenced estuary. Journal of Marine Systems 60 (3–4), 317–329.

Triantafyllou, G., Hoteit, I., Petihakis, G., 2003. A singular evolutive interpolated Kalman filter for efficient data assimilation in a 3-D complex physical–biogeochemical model of the Cretan Sea. Journal of Marine Systems 40, 213–231.

Triantafyllou, G., Hoteit, I., Korres, G., Petihakis, G., 2005. Ecosystem modeling and data assimilation of physical–biogeochemical processes in shelf and regional areas of the Mediterranean Sea. Applied Numerical Analysis & Computational Mathematics 2 (2), 262–280.

Triantafyllou, G., Korres, G., Hoteit, I., et al., 2007. Assimilation of ocean colour data into a biogeochemical flux model of the eastern Mediterranean Sea. Ocean Science 3, 397–410.

Triantafyllou, G., Hoteit, I., Luo, X., Tsiaras, K., Petihakis, G., 2013. Assessing a robust ensemble-based Kalman filter for efficient ecosystem data assimilation of the Cretan Sea. Journal of Marine Systems 125, 90–100.

Uitz, J., Huot, Y., Bruyant, F., Babin, M., Claustre, H., 2008. Relating phytoplankton photophysiological properties to community structure on large scale. Limnology and Oceanography 53 (2), 614–630.

van Leeuwen, P.J., 2010. Nonlinear data assimilation in geosciences: an extremely efficient particle filter. Quarterly Journal of the Royal Meteorological Society 136 (653), 1991–1999.

Whitaker, J.S., Hamill, T.M., 2002. Ensemble data assimilation without perturbed observations. Monthly Weather Review 130 (7), 1913–1924.

Widdicombe, C.E., Eloire, D., Harbour, D., Harris, R.P., Somerfield, P.J., 2010. Long-term phytoplankton community dynamics in the Western English Channel. Journal of Plankton Research 32 (5), 643–655.

Xiao, Y., Friedrichs, M.A.M., 2014. Using biogeochemical data assimilation to assess the relative skill of multiple ecosystem models: effects of increasing the complexity of the planktonic food web. Biogeosciences Discussions 11, 481–520.

Zhang, X., Hu, L., He, M.X., 2009. Scattering by pure seawater: effect of salinity. Optics Express 17, 5699–5709.

Zhao, J., Barnes, B., Melo, N., et al., 2013. Assessment of satellite-derived diffuse attenuation coefficients and euphotic depths in south Florida coastal waters. Remote Sensing of Environment 131, 38–50.