



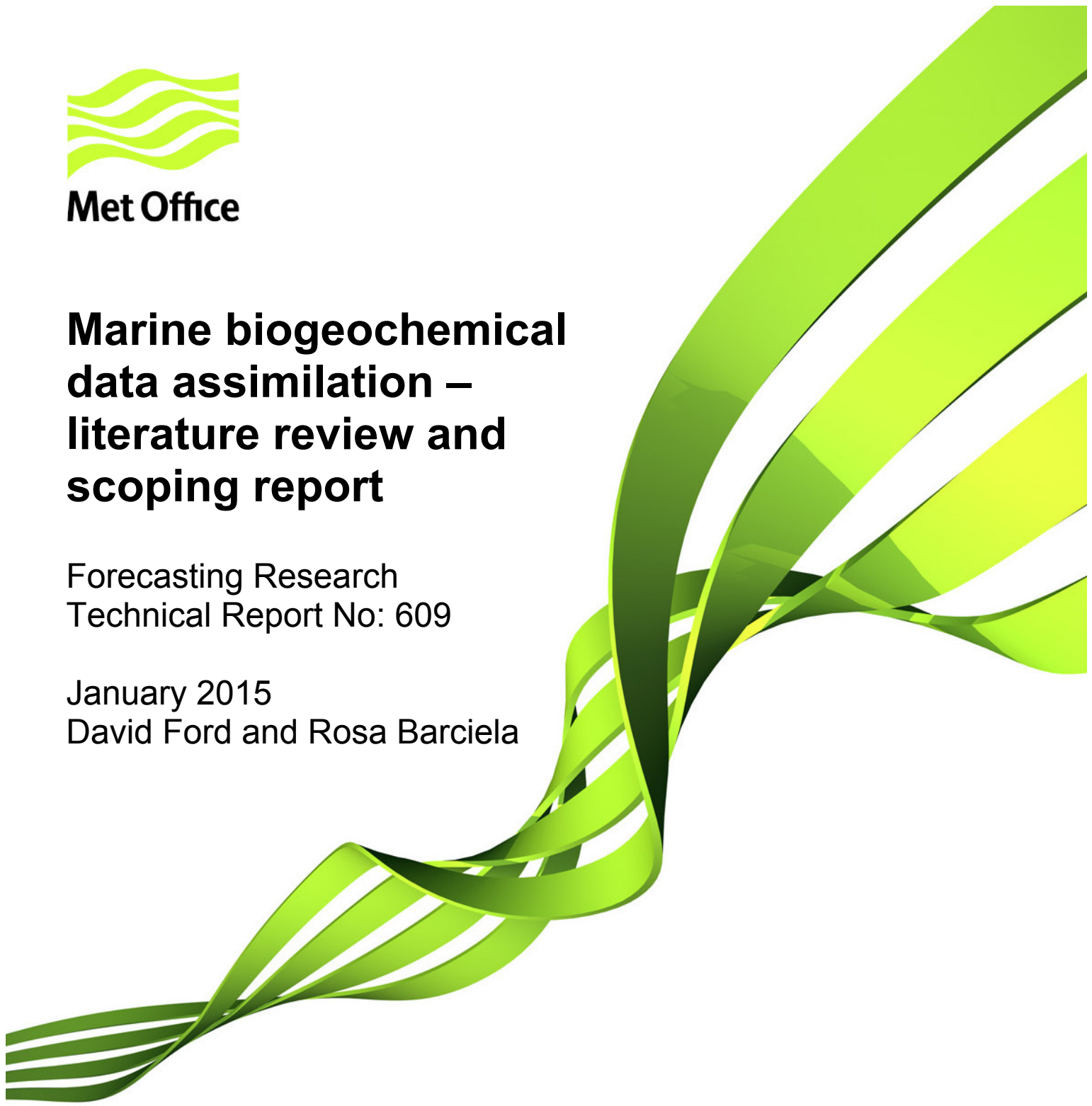
Met Office

Marine biogeochemical data assimilation – literature review and scoping report

Forecasting Research
Technical Report No: 609

January 2015

David Ford and Rosa Barciela



Contents

1. Introduction.....	2
2. Brief overview of data assimilation methods.....	3
3. Current Met Office approach	5
4. Literature review – parameter estimation.....	7
5. Literature review – state estimation.....	9
5.1 Methods	9
5.1.1 OI and “simple” approaches.....	9
5.1.2 SEEK and SEIK filters.....	11
5.1.3 EnKF and other Monte Carlo ensemble methods	14
5.1.4 Variational methods	17
5.2 Dealing with non-Gaussian distributions.....	18
5.3 Generating multivariate increments	19
5.4 Formulation of error covariances	20
6. Impact of physical data assimilation on biogeochemistry	21
7. Discussion and recommendations	25
References	31

1. Introduction

Marine biogeochemical data assimilation is increasingly used as a tool to improve model predictions, and now forms part of operational systems within MyOcean. It remains, however, a maturing subject, with many challenges yet to be overcome. This report aims to give an overview of the current state-of-the-art in marine biogeochemical data assimilation, and to place current and future Met Office activities in the context of this.

There are two main applications of data assimilation – parameter estimation and state estimation. Parameter estimation aims to adjust model parameters so that a resulting model run better fits a given set of observations. It is typically applied offline as a tuning exercise. State estimation updates the model fields based on observations during a model run, and is often applied in reanalysis and forecasting systems. Parameter and state estimation can also be combined to adjust model parameters during a run. Whilst this review will cover both applications, the larger focus will be on state estimation, as this is most relevant to the reanalysis and forecasting activities currently undertaken by the Met Office.

This report aims to give an overview of how data assimilation has been applied to marine biogeochemistry in the scientific literature, and as such, a certain degree of knowledge of data assimilation theory is assumed. A brief explanation of different methods will be provided, but for full details the reader will be left to follow the appropriate references.

A related topic is the impact that assimilating physical data can have on biogeochemical model fields. Counter-intuitively, this has widely been found to have a negative impact, which is generally considered to be due to the creation of spurious vertical velocities. This report will also consider this subject, including recent Met Office work.

The layout of the report is as follows: section 2 will give a brief overview of common techniques in data assimilation. Section 3 will provide an overview of current Met Office activities relating to marine biogeochemical data assimilation, and place this within the wider context of data assimilation in the Met Office. A literature review of parameter estimation work will be given in section 4, and of state estimation in section 5. This will

group together different classes of methods, as well as work towards some of the main challenges facing the community. Section 6 will look at the impact of physical data assimilation, and section 7 will draw some conclusions based on the literature review, and finally make some recommendations for future approaches to be taken by the Met Office. These recommendations should be seen as suggestive rather than prescriptive, and include the potential for collaboration.

Other reviews of marine biogeochemical assimilation exist in the literature. These will be referenced where appropriate, but are noted here for convenience. Robinson and Lermusiaux (2002) gave a forward-looking discussion of different applications of data assimilation, and progress to date. Gregg et al. (2009) tabulated a large number of previous studies, and reviewed and discussed the approaches to skill assessment. In many ways this acted as an extension to the review provided in the introductory sections of Gregg (2008). Gregg et al. (2009) includes a section on fisheries data assimilation, which will not be discussed in this report. McClain (2009) reviewed ocean colour data and their usage, including a section on data assimilation. Recently, Dowd et al. (2014) reviewed the subject from a statistical perspective. Reviews have also emerged from the GODAE (Global Ocean Data Assimilation Experiment) community, with a focus on reanalysis and operational forecasting. These have been provided by Brasseur et al. (2009), Matear and Jones (2011) and Gehlen et al. (2015).

2. Brief overview of data assimilation methods

Data assimilation combines information from models and observations in order to produce an analysis – the best estimate of the state of a system at a given time. Since both models and observations contain errors, these must be estimated and accounted for. When applied to the Earth's ocean or atmosphere, data assimilation becomes a computationally expensive problem, and so a number of approaches, of varying complexity, have been developed and implemented. Detailed introductions to data assimilation theory, as applied to meteorology and oceanography, are provided by Talagrand (1997), Kalnay (2003) and Daley (1991). Furthermore, ECMWF (European Centre for Medium-Range Weather Forecasts) runs an annual course on data assimilation, and comprehensive lecture notes are publically available, linked from <http://www.ecmwf.int/en/learning/education-material>.

There are two main classes of data assimilation methods: sequential and variational (Talagrand, 1997). In sequential assimilation, the model is integrated forward in time until an observation is available, at which point an update to the model occurs. In variational assimilation, the model is updated based on all observations in a given time window, through the minimisation of a cost function. Most methods within these two classes stem from two equivalent formulations of the least squares analysis problem (Lorenc, 1986). Sequential approaches based on statistical estimation theory aim to solve the following equation:

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}(\mathbf{y} - H(\mathbf{x}_b)) \quad (\text{Eq. 1})$$

Variational approaches based on statistical estimation theory aim to calculate the analysis \mathbf{x}_a which minimises the following cost function:

$$J(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) + (\mathbf{y} - H(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y} - H(\mathbf{x})) \quad (\text{Eq. 2})$$

In each case:

- \mathbf{x}_a is the analysis state vector
- \mathbf{x}_b is the background state vector (a prior estimate of the system state)
- \mathbf{y} is the observation state vector
- \mathbf{B} is the background error covariance matrix
- \mathbf{R} is the observation error covariance matrix
- H is the observation operator (which converts from model to observation space)
- \mathbf{H} is the linearised observation operator

There are two main approaches to variational assimilation: 3D-Var and 4D-Var. 3D-Var (Sasaki, 1958) assumes the observations to be static in time, whereas 4D-Var (Talagrand and Courtier, 1987; Courtier et al., 1994) accounts for the distribution of observations in time. 4D-Var requires the model adjoint to be run, making it more complex and computationally expensive to implement than 3D-Var. An alternative is to use 3D-Var (or other methods) with an FGAT (first guess at appropriate time) technique, which compares model and observation values at the observation times, prior to assimilation. This approximates the time dimensionality of 4D-Var, without requiring a full 4D-Var system to be run.

Within sequential assimilation, a number of approaches have been employed. For a linear problem, Eq. 1 can be solved using the Kalman Filter (Kalman, 1960), which evolves the full error covariances at each time step using the forecast model. However, most meteorological and oceanographic applications are non-linear, and of too large an order to make such an evolution of the error covariances computationally feasible. The extended Kalman filter (Jazwinski, 1970) extends the Kalman filter to non-linear problems by linearising about the current model state, and variations (e.g. Pham et al., 1998a) have been developed which reduce the computational cost. A simpler, but widely used, approach is optimal interpolation (OI; Gandin, 1963). This does not evolve the error covariances, and makes the assumption that for each element of the model state vector, only a small number of observations are important in determining the assimilation increment. Use can also be made of ensembles to approximate the evolution of the error covariances. Popular implementations include the ensemble Kalman filter (EnKF; Evensen, 1994; Evensen, 2003), ensemble transform Kalman filter (ETKF; Bishop et al., 2001) and ensemble OI (EnOI; Oke et al., 2002; Evensen, 2003). Recently, research has been conducted into the use of particle filters (Ristic et al., 2004), as a potentially efficient method of using ensemble information to tackle highly non-linear problems. Ensembles can also be used to generate the error covariances used in variational schemes, an approach known as hybrid data assimilation (e.g. Clayton et al., 2013).

3. Current Met Office approach

Operational ocean forecasting and reanalysis at the Met Office is performed using the FOAM (Forecasting Ocean Assimilation Model) system (Blockley et al., 2014; O'Dea et al., 2012), which is built on the NEMO (Nucleus for European Modelling of the Ocean) hydrodynamic model (Madec, 2008). There is a global configuration, run operationally at $1/4^\circ$ horizontal resolution, three $1/12^\circ$ regional configurations covering the North Atlantic, Mediterranean Sea and Indian Ocean, and two shelf seas configurations covering the North-West European Shelf and the Persian Gulf. The North-West European Shelf configuration is coupled with the ERSEM (European Regional Seas Ecosystem Model) biogeochemical model (Edwards et al., 2012), providing operational forecasts of biogeochemical variables. ERSEM can also be coupled with the Persian Gulf configuration, but this functionality is currently pre-operational. The global configuration can be coupled with the HadOCC (Hadley Centre Ocean Carbon Cycle Model)

biogeochemical model (Palmer and Totterdell, 2001) at either $1/4^\circ$ or 1° resolution, again on a pre-operational basis.

In the Met Office, ocean data assimilation is currently only used for state estimation. For assimilation of physical ocean variables, FOAM uses a 3D-Var implementation of the NEMOVAR assimilation framework (Waters et al., 2015). This uses an FGAT technique to approximate the time dimensionality of 4D-Var. NEMOVAR also makes use of a balance operator, to account for multivariate relationships. The adoption of NEMOVAR by the Met Office is a relatively recent development. Previously, the analysis correction scheme of Martin et al. (2007) was used. Analysis correction (Lorenc et al., 1991) is a form of OI which uses an iterative method. For atmospheric data assimilation, the Met Office global numerical weather prediction (NWP) system currently uses a hybrid ensemble/4D-Var scheme (Clayton et al., 2013).

In terms of marine biogeochemistry, the Met Office has developed data assimilation schemes for sea surface chlorophyll (which can be derived from remotely sensed ocean colour) and sea surface $p\text{CO}_2$ (partial pressure of carbon dioxide), for use with FOAM-HadOCC. Detailed descriptions and validation of the schemes are given in Ford et al. (2012) for chlorophyll, and in While et al. (2012) for $p\text{CO}_2$.

The first step in each case is to calculate a set of univariate surface increments of the observed variable. In published work to date, this has been performed using the analysis correction scheme, but this capability has recently been transitioned to NEMOVAR. For chlorophyll, the analysis is performed for $\log_{10}(\text{chlorophyll})$, as chlorophyll follows a logarithmic distribution in nature (Campbell, 1995) – this point will be discussed further in section 5.2. Multivariate balancing schemes are then run, which use the univariate increments to calculate multivariate increments to be applied to the model.

In the case of chlorophyll, the nitrogen balancing scheme of Hemmings et al. (2008) is used. This was developed as a collaboration between the Met Office and NOC (National Oceanography Centre). The scheme ingests the $\log_{10}(\text{chlorophyll})$ increments, and converts them to phytoplankton nitrogen biomass. Using a principle of conservation of nitrogen and carbon, the scheme then calculates increments to all the HadOCC state variables, at all depths. This is performed statistically, based on the model state, and whether growth or loss errors are more likely to be dominant.

In HadOCC, $p\text{CO}_2$ is not a state variable, but a function of temperature, salinity, DIC (dissolved inorganic carbon) and alkalinity. The balancing scheme employed for the $p\text{CO}_2$ data assimilation assumes temperature and salinity to be accurate, and calculates increments to DIC and alkalinity accordingly. These are applied throughout the mixed layer, as detailed in White et al. (2012).

4. Literature review – parameter estimation

The majority of marine biogeochemical data assimilation studies to date have been for the purpose of parameter estimation. This review will not attempt to detail all studies in the literature, but to summarise standard and emerging techniques. The next paragraph will give a long but by no means exhaustive list of parameter estimation studies, and the rest of the section will be spent synthesising the most salient points from these.

Parameter estimation has been performed by, amongst others: Fasham and Evans (1995), Matear (1995), Matear and Holloway (1995), Semovski and Woźniak (1995), Hurtt and Armstrong (1996), Prunet et al. (1996a,b), Harmon and Challenor (1997), Spitz et al. (1998), Fasham et al. (1999), Hurtt and Armstrong (1999), Athias et al. (2000), Schlitzer (2000), Vallino (2000), Fennel et al. (2001), Friedrichs (2001), Schartau et al. (2001), Spitz et al. (2001), Friedrichs (2002), Schlitzer (2002), Dowd and Meyer (2003), Evans (2003), Faugeras et al. (2003), Garcia-Gorriz et al. (2003), Hemmings et al. (2003), Losa et al. (2003), Pastres et al. (2003), Schartau and Oschlies (2003), Faugeras et al. (2004), Hemmings et al. (2004), Kuroda and Kishi (2004), Losa et al. (2004), Schlitzer et al. (2004), Oschlies and Schartau (2005), Weber et al. (2005), Zhao et al. (2005), Dowd (2006), Friedrichs et al. (2006), Kwon and Primeau (2006), Friedrichs et al. (2007), Huret et al. (2007), Rose et al. (2007), Tjiputra et al. (2007), Zhao and Lu (2008), Cossarini et al. (2009), Ward et al. (2010), Ciavatta and Pastres (2011), Doron et al. (2011), Kane et al. (2011), Kidston et al. (2011), Smith and McGillicuddy (2011), Hemmings and Challenor (2012), Mattern et al. (2012), Pelc et al. (2012), Simon and Bertino (2012), Wang et al. (2012), Doron et al. (2013), Kidston et al. (2013), Mattern et al. (2013), Prieß et al. (2013), Ward et al. (2013), Yao and Schlitzer (2013), Hemmings et al. (2014), Mattern et al. (2014), and Xiao and Friedrichs (2014a,b).

Most parameter estimation studies have used a variational technique, normally assimilating *in situ* time series data with a 0D or 1D model. An optimal parameter set is found by altering the model parameters until the resulting model integration gives the best fit to the assimilated observations, usually in terms of minimising a cost function. This is typically performed in a 0D or 1D framework, partly because a large number of model integrations is required and the computational cost is generally prohibitive otherwise, and partly because suitable co-located and seasonally-varying multivariate observations are mostly only available at a small number of time series stations.

Variational techniques used in 0D and 1D tend to vary in the way the cost function is defined and minimised, and include the conjugate gradient method (e.g. Fasham et al., 1995), simulated annealing (e.g. Matear, 1995), the adjoint method (e.g. Prunet et al., 1996a,b), and micro-genetic algorithms (e.g. Schartau and Oschlies, 2003). Another major difference between studies is the way in which 3D processes are accounted for in the 0D or 1D simulations, as discussed by Hemmings and Challenor (2012). Usually parameters are optimised at a single site, but the optimisation procedure can also be run for multiple sites simultaneously (e.g. Hurtt and Armstrong, 1999). Parameter estimation can also be performed using a 3D model (e.g. Garcia-Gorriz et al., 2003; Tjiputra et al., 2007; Doron et al., 2013), but computational constraints require the simulations to be limited in time and/or space. This has prompted the use of emulators as a way to reduce the number of simulations required (e.g. Mattern et al., 2012; Hemmings et al., 2014). In order to consistently and thoroughly account for the different factors which impact parameter estimation, in particular the treatment of the 3D environment, a generic framework has recently been developed by Hemmings and Challenor (2012) and Hemmings et al. (2014), which is available for community use.

As well as estimating a static set of model parameters, data assimilation has been used to combine state and parameter estimation by estimating time-dependent parameters (e.g. Losa et al., 2003; Mattern et al., 2012; Mattern et al., 2014). This has shown promise compared with using static parameter values, and can also feed into model development. A more comprehensive method has been proposed for operational oceanography by Matear and Jones (2011), in which daily data assimilation is used to update both the model state, and space- and time-varying model parameters. A further application of parameter estimation is the investigation of model complexity, by using data assimilation to optimise a series of models at a particular site, and then assessing

the models' fit to different observations (e.g. Friedrichs et al., 2007; Ward et al., 2013; Xiao and Friedrichs, 2014b).

5. Literature review – state estimation

This section provides a literature review of the use of marine biogeochemical data assimilation for state estimation. The first sub-section groups together studies by method, whilst the remaining sub-sections detail the approaches taken to three general challenges relating to the assimilation of marine biogeochemical data. The review intends to be relatively concise but comprehensive – nonetheless, there may inevitably be studies of which the authors are unaware.

5.1 Methods

Whilst a number of different techniques have been employed, the vast majority of methods used have been sequential, and two in particular have dominated in the literature: the singular evolutive extended Kalman (SEEK) filter (Pham et al., 1998a; Brasseur and Verron, 2006) and the ensemble Kalman filter (EnKF; Evensen, 1994). This review groups studies into four categories: 1) relatively simple methods, including OI; 2) the SEEK filter and the related singular evolutive interpolated Kalman (SEIK) filter; 3) Monte Carlo ensemble methods such as the EnKF; 4) variational methods.

5.1.1 OI and “simple” approaches

The simplest of all data assimilation techniques is data insertion, which just involves replacing the model field with observation data. This was the method used in the first ocean colour assimilation study, conducted by Ishizaka (1990). Chlorophyll from the CZCS (Coastal Zone Color Scanner) sensor was inserted into a model of the South-eastern United States continental shelf, and the model run forward. It was found that the assimilation had little memory, and it took only a few days for the model results to converge to those of a free run.

Nudging (Hoke and Anthes, 1976) is a simple empirical method of data assimilation, which involves adding a term to the model equations that nudges the model solution towards data that has been interpolated to the model grid. Armstrong et al. (1995) nudged monthly CZCS chlorophyll data into a model of the Atlantic, and found a great deal of regional variation in the results, which they used to draw conclusions about

model performance and complexity. Nudging can also be used to relax towards climatological data (e.g. Najjar et al., 1992; Anderson and Sarmiento, 1995; Moisan et al., 1996; Lagman et al., 2014).

Semovski and Woźniak (1995) compared two methods of assimilating monthly CZCS data into a model of the North Atlantic and Baltic Sea. The first was a variational method used for parameter estimation. The second was a simple linear interpolation between observed and modelled values, with depth- and time-varying coefficients, used for state estimation. This was essentially a simplified form of OI. The study noted that the CZCS data were not fully representative of the case II (Morel and Prieur, 1977) waters which made up much of the model domain. Nonetheless, improvements in the magnitude and seasonal cycle of phytoplankton were noted, with the state estimation procedure producing the best fit to phytoplankton biomass constructed from the CZCS data. However, this was accompanied by a large and spurious increase in zooplankton biomass.

A number of studies have used OI. Anderson et al. (2000) assimilated both physical and biogeochemical data into a model of the Gulf Stream region. Initial conditions for biogeochemical fields were created from *in situ* observations and temperature data using a “biological feature model” in order to ensure consistency with the physics. The biogeochemical data assimilated were *in situ* observations of phytoplankton biomass (derived from chlorophyll) and nitrate. The assessment largely took the form of checking the consistency of features in the physical and biogeochemical fields over a ten-day period, and it was concluded that the best results were achieved when assimilating both physical and biogeochemical data. This result will be discussed further in section 6. Popova et al. (2002a) used OI to assimilate *in situ* observations of a number of physical and biogeochemical variables taken during a cruise in the North-east Atlantic. The model was run for the duration of the cruise (24 days) and data were assimilated on three days near the start. A particular result noted was the ability to capture two separate ecosystem regimes, which would normally not have been possible with their constant set of model parameters. The method was also used by Popova et al. (2002b). Similarly to Anderson et al. (2000) and Popova et al. (2002a,b), Beşiktepe et al. (2003) used OI to assimilate *in situ* observations of both physical and biogeochemical variables, collected over a seven-week period in Massachusetts Bay. The procedure included a balance relationship between the observed (chlorophyll, nitrate, ammonium) and non-observed (phytoplankton, zooplankton, detritus)

biogeochemical variables. The assimilation was found to improve model results, with a predictive capability of up to two weeks. More recently, Zhang et al. (2014) used OI to assimilate sediment data derived from remotely sensed ocean colour into a sediment transport model of Deep Bay, Hong Kong. This reduced model errors compared with *in situ* measurements. A form of OI has also been used at the Met Office for the assimilation of chlorophyll (Ford et al., 2012; Hemmings et al., 2008) and pCO₂ (While et al., 2012) data, as described in section 3.

A further method to have been used is the conditional relaxation analysis method (CRAM; Thomasell and Welsh, 1963; Oort, 1983), also known as blended analysis (Reynolds, 1988). Previously used for combining remotely sensed and *in situ* data (e.g. Gregg and Conkright, 2001), the method was applied to data assimilation by Gregg (2008). Data are inserted into the model field, and the resulting field is adjusted so that its Laplacian matches that of the original field, with a regionally-varying weighting between model and data. The difference between the two fields then gives the assimilation increments. In Gregg (2008), chlorophyll derived from SeaWiFS (Sea-viewing Wide Field-of-view Sensor) ocean colour was assimilated into a global configuration of NOBM (National Aeronautics and Space Administration (NASA) Ocean Biogeochemical Model), for the period 1998-2003. The assimilation improved the model fit to *in situ* chlorophyll observations, and SeaWiFS-based estimates of primary production. Results were found to improve with assimilation frequency. Gregg (2008) concluded though that underlying model biases were not being fully addressed, and so a multivariate approach would be preferred. The methodology was therefore extended by Rousseaux and Gregg (2012) to include an online correction to nutrient variables, using the model nutrient to chlorophyll ratios. This approach was also used by Gregg and Rousseaux (2014), and will be discussed further in section 5.3.

5.1.2 SEEK and SEIK filters

Two implementations of the extended Kalman filter have been widely used in marine biogeochemistry: the SEIK filter and the SEEK filter. The SEEK filter (Pham et al., 1998; Brasseur and Verron, 2006) reduces the computational burden of the full extended Kalman filter by using a low rank approximation to the forecast error covariance matrix. The filter is initialised with a singular error covariance matrix derived from the model, typically using EOF (empirical orthogonal function) analysis. The errors are then evolved according to the model dynamics. The error covariances include relationships between all model state variables, making the method inherently

multivariate. The SEIK filter (Pham et al., 1998b) takes a similar form to the SEEK filter, except that interpolation is used for the forecast step instead of linearisation. This requires an ensemble representation of the error covariance matrix, making it similar to the EnKF; the main distinction from the EnKF being that the ensemble is preconditioned, with the analysis update computed in the error subspace rather than in observation space. A comprehensive comparison of the SEEK filter, SEIK filter and EnKF is provided by Nerger et al. (2005). The computational cost of the SEEK and SEIK filters are often reduced by using a localised version, which only considers nearby observations. A further cost-saving can be made by keeping the error covariance matrix static, although when implemented in this manner the method is little different from OI.

At least 13 studies have used a SEEK filter for marine biogeochemical data assimilation. Carmillet et al. (2001) assimilated pseudo-observations of surface phytoplankton into a model of the North Atlantic, and found that effective multivariate results could be obtained using a filter of order rank 10, provided the error structure is sufficiently well specified. A similar SEEK filter implementation was used by Hoteit et al. (2003) to assimilate *in situ* data into a 1D version of the ERSEM model in the Cretan Sea. This work was explored further by Hoteit et al. (2004) and expanded to 3D by Hoteit et al. (2005). All three studies obtained improved model results, and the sensitivity of the assimilation to various assumptions regarding localisation was tested. Magri et al. (2005) assimilated data into a 1D model of the Ligurian Sea, using EOFs deduced from observations. Assimilation into a 1D model of the Ligurian Sea was also explored by Raick et al. (2007). Triantafyllou et al. (2007) assimilated SeaWiFS chlorophyll data into a model of the Eastern Mediterranean Sea for the year 1999. The biogeochemical model used was BFM (Biogeochemical Flux Model; Vichi et al., 2007), which shares a heritage with the ERSEM model used by the Met Office to model shelf seas biogeochemistry. The assimilation improved the multivariate performance of the model, though was unable to effectively correct sub-surface errors, or follow the rapid changes in ecosystem dynamics observed during the spring bloom. Fontana et al. (2009) and Fontana et al. (2010) assimilated SeaWiFS chlorophyll data into a model of the Gulf of Fos on the French Mediterranean coast. Pseudo-salinity extracted from the model was also assimilated, to try and ensure that the different biogeochemical dynamics of freshwater and open ocean regions were respected. The studies found generally improved representation of chlorophyll compared with *in situ* data, and some improvement in nutrient concentrations. The impact of the assimilation was greatest

nearer the surface. Ourmières et al. (2009) assimilated physical data in conjunction with a nutrient climatology into a North Atlantic configuration of NEMO coupled with the LOBSTER (Lévy et al., 2005) biogeochemical model. Whilst the assimilated biogeochemical data were climatological rather than individual observations, the study served to demonstrate that improving nutrient concentrations through data assimilation can lead to improvements in ecosystem variables such as sea surface chlorophyll. Korres et al. (2012) assimilated both SeaWiFS and merged GlobColour (<http://www.globcolour.info>) chlorophyll data into a 1/100° configuration of ERSEM for the Pagasitikos Gulf, in the Aegean Sea. Comparisons were made between the SEEK filter, SEIK filter and the simplified version of the SEEK filter used by Carmillet et al. (2001), referred to as the SFEK (singular fixed basis extended Kalman) filter. The SFEK filter performed least well, and the SEEK filter performed slightly better than the SEIK filter with the ensemble size used, so the SEEK filter was used for the main experiments. This brought the model closer to observations of chlorophyll and nutrients, though Korres et al. (2012) suggested that assimilation of physics and of sub-surface biogeochemistry would be required for a more accurate representation. Butenschön and Zavatarelli (2012) compared three implementations of the SEEK filter in a 1D configuration of BFM in the Northern Adriatic Sea. It was found that the more sophisticated, and with it more computationally expensive, implementations brought the best results. Fontana et al. (2013) assimilated SeaWiFS chlorophyll data into a North Atlantic configuration of NEMO-LOBSTER for the period 1998-2006, to test a methodology for creating a multivariate reanalysis. Chlorophyll and nitrate concentrations were improved compared with *in situ* observations, though this improvement was limited to surface waters. This was found to be due to a lack of correlation between surface chlorophyll and sub-mixed layer nitrate in the model.

The SEIK filter has been applied to marine biogeochemistry by Triantafyllou et al. (2003), Triantafyllou et al. (2013), Nerger and Gregg (2007) and Nerger and Gregg (2008). In Triantafyllou et al. (2003) and Triantafyllou et al. (2013) pseudo-observations were assimilated into a model of the Cretan Sea. Both studies found the assimilation to be effective, with Triantafyllou et al. (2013) obtaining improved results through the use of an adaptive inflation scheme. Nerger and Gregg (2007) used a univariate localised SEIK filter with static error covariance to assimilate SeaWiFS chlorophyll into the global NOBM model, for the period 1998-2004. The assimilation improved the model fit to *in situ* chlorophyll observations, but led to a slight degradation of nutrients compared to climatology. The method was extended by Nerger and Gregg (2008) to include an

online bias correction scheme, formulated as per Dee and Da Silva (1998). This further improved the fit to *in situ* chlorophyll observations, though no assessment was made of other variables.

A similar approach to Nerger and Gregg (2007) was taken by Shulman et al. (2013), and presented as a “reduced order Kalman filter with a stationary forecast error covariance”. This was used to assimilate chlorophyll concentration and phytoplankton absorption coefficient ($a_{ph(488)}$) derived from MODIS (Moderate Resolution Imaging Spectroradiometer) ocean colour data into a model of Monterey Bay, for the period of a five-day observation campaign. The assimilation served to improve both surface and sub-surface chlorophyll compared to *in situ* observations. It also increased the proportion of diatoms compared with small phytoplankton, better matching observed distributions. Little impact was found on nitrate concentrations.

A method related to the SEIK filter, known as error subspace statistical estimation (ESSE; Lermusiaux and Robinson, 1999), has been used by Cossarini et al. (2009). It too takes an ensemble approach, but is included in this sub-section rather than section 5.1.3 because, like the SEIK filter, it works in the error subspace rather than in observation space. Cossarini et al. (2009) used the technique to assimilate data into a model of the Lagoon of Venice, and investigated a number of aspects of the assimilation’s formulation, and the impact on seasonal ecosystem dynamics.

5.1.3 EnKF and other Monte Carlo ensemble methods

The EnKF was introduced by Evensen (1994) as an alternative to the extended Kalman filter. It is a Monte Carlo method, which uses ensembles to evolve the error covariances. As with the SEEK filter, the EnKF is inherently multivariate. Its practical implementation is discussed further in Evensen (2003), and comparison is made with the SEEK and SEIK filters in Nerger et al. (2005).

Eknes and Evensen (2002) implemented an EnKF in a simple 1D biogeochemical model, as an initial demonstration of its application to marine biogeochemistry. Using twin experiments, it was shown that assimilation of phytoplankton data could also control the evolution of the zooplankton and nutrient variables. The work was extended by Allen et al. (2003) to a 1D configuration of the more complex ERSEM model in the Cretan Sea. *In situ* chlorophyll and nitrate observations were assimilated, and the assimilation was found to have a predictability window of at least two days. An

ensemble size of around 100 was required to give a sufficient description of the errors. The EnKF was first demonstrated for a 3D marine biogeochemical model by Natvik and Evensen (2003a,b), who assimilated SeaWiFS chlorophyll data into a model of the North Atlantic. In Natvik and Evensen (2003a), it was shown that the assimilation resulted in a better model fit to the assimilated data, and qualitatively consistent changes to other variables. At least 60 ensemble members were required, with an ensemble size of 150 suggested for operational applications. Natvik and Evensen (2003b) discussed non-linearity, and introduced diagnostics which could be used to examine the system. Torres et al. (2006) assimilated chlorophyll and nutrient data into a 1D configuration of ERSEM for the Ria de Vigo estuary in Spain. The assimilated variables were all improved, as was diatom biomass, but the impact on other variables was less consistent. Lenartz et al. (2007) implemented an EnKF in a 1D model of the Ligurian Sea, and compared results to the SEEK filter implementation of Raick et al. (2007). The EnKF was concluded to give slightly better results, but at a much higher computational cost, as 100 ensemble members were required, rather than the single evolution performed with the SEEK filter. Simon and Bertino (2009) performed twin experiments with a 3D model of the North Atlantic, in which the EnKF was extended to include a Gaussian anamorphosis transform. This is a technique for dealing with non-Gaussian variables, and will be discussed further in section 5.2. Simon and Bertino (2012) applied a similar technique to a 1D model for combined state and parameter estimation, and these methods are now used as part of the Arctic operational forecasting system which delivers products to MyOcean (Gehlen et al., 2015). Anamorphic transformations are also the subject of Béal et al. (2010) and Brankart et al. (2012). Ciavatta et al. (2011) assimilated MODIS chlorophyll data into a 3D configuration of ERSEM for the Western English Channel. This improved analyses of both chlorophyll and a number of non-assimilated variables compared with *in situ* observations at a time series site, although the impact on forecast skill was small. Ciavatta et al. (2014) added a more sophisticated bio-optical model to ERSEM, and adapted the EnKF to assimilate SeaWiFS light attenuation coefficient ($K_d(443)$) data rather than chlorophyll. This approach resulted in improvements to many variables compared with *in situ* data. It was also demonstrated that the assimilation had changed the balance of ecosystem processes, shifting the simulated food web towards the microbial loop. Hu et al. (2012) assimilated SeaWiFS and MODIS chlorophyll data into a model of the Middle Atlantic Bight, and found improvements compared with satellite-derived chlorophyll and particulate organic carbon (POC) data.

As well as the application to lower-trophic level ecosystem models in the above studies, the EnKF has also been used to reconstruct air-sea CO₂ fluxes in the pre-industrial and industrial eras, in order to study sources and sinks of anthropogenic carbon. This has been performed by Gerber et al. (2009) and Gerber and Joos (2010), by assimilating observation-based reconstructions of anthropogenic carbon into an ocean model with carbon tracers. This was tested as an alternative to techniques such as inversion-based methods.

Recently, other Monte Carlo ensemble methods have been applied to marine biogeochemical data assimilation, with the aim of better handling non-linearities. Smith and McGillicuddy (2011) introduced a strong constraint iterative ensemble smoother, a variation on the ensemble Kalman smoother (EnKS; Evensen and van Leeuwen, 2000). The EnKS differs from the EnKF in that the analysis is updated for all previous observation times, with the EnKF solution having been the first estimate. This allows information to be propagated backward in time, without requiring the use of the model adjoint. Smith and McGillicuddy (2011) used the method to assimilate MODIS chlorophyll data into a model of the Middle Atlantic Bight, for joint state/parameter estimation.

Another class of ensemble methods is particle filters (Ristic et al., 2004; van Leeuwen, 2009). These are similar in principle to the EnKF, except that the EnKF assumes all probability distributions to be Gaussian, whereas particle filters are applicable for the full non-linear system, by weighting the probabilities of ensemble members (particles) according to their fit to observations, rather than updating the state of each ensemble member in the analysis step. This should make particle filters better suited to the marine biogeochemical data assimilation problem, but a careful formulation is required in order to obtain an accurate solution with a small enough ensemble to be computationally feasible. This remains an active area of research. One method proposed is the sequential importance resampling filter (SIR; Rubin, 1988; Gordon et al., 1993; Kivman, 2003), which abandons particles with low weight, instead using multiple copies of particles with high weight. SIR has been used for joint state/parameter estimation in a 0D model by Losa et al. (2003), in a 1D model by Mattern et al. (2010), and in a 3D model of the Middle Atlantic Bight by Mattern et al. (2013). This latter study highlighted a number of issues faced when implementing such a method for biogeochemistry, but also presented potential solutions which allowed effective results to be achieved. Dowd (2007) introduced a Markov Chain Monte Carlo

(MCMC) approach intended to address the problem of filter (or sample) degeneracy (van Leeuwen, 2009) which can arise when using SIR and other particle filters. MCMC had previously been implemented in conjunction with SIR by Dowd (2006) for parameter estimation, and was developed by Dowd (2007) to work stand-alone. This stand-alone approach was tested for state estimation alongside SIR and an EnKF in a 0D model, as an illustration of its implementation, with comparable results obtained with all three methods. This was deemed to be more successful than an earlier attempt made by Dowd and Meyer (2003).

Metref et al. (2014) introduced a further method, the multivariate rank histogram filter (MRHF), intended to sit part way between the EnKF and particle filters: able to handle non-linear systems, whilst reducing the sampling issues of particle filters. The MRHF is an extension of the rank histogram filter (RHF) introduced by Anderson (2010), in which probability densities are represented as piecewise continuous functions. Whilst Metref et al. (2014) did not apply the MRHF to a full marine biogeochemical data assimilation problem, data from a coupled physical-biogeochemical model was used to test the behaviour of the MRHF, as part of the method's ongoing development.

5.1.4 Variational methods

Apart from the recent use of 3D-Var by the Met Office, described in section 2, to the best of the authors' knowledge the only application of a variational method to state estimation for a full 3D lower-trophic level marine biogeochemical model has been by Teruzzi et al. (2014). This described the implementation of a 3D-Var scheme for the Mediterranean Sea forecasting system, which delivers products to MyOcean. A version of NEMO is used for the physical model, but the assimilation does not use the NEMOVAR framework. The biogeochemical model is BFM. The scheme of Teruzzi et al. (2014) assimilates MODIS chlorophyll data, and uses this to update the four phytoplankton functional groups within BFM, using the ratios existing within the model. This improved analyses and forecasts of both surface and sub-surface chlorophyll compared with independent *in situ* data. Other variables were not assessed, but it was noted that nutrient biases could sometimes act against the effect of the assimilation.

The study of Song et al. (2012) used a 1D marine biogeochemical model to demonstrate a formulation of 4D-Var that assumes lognormal error distributions, which will be discussed further in section 5.2. Variational methods have also been used to try and improve estimates of global air-sea CO₂ fluxes, through the use of an ocean-tracer

model without an ecosystem. A form of 4D-Var was used by Valsala and Maksyutov (2010) to assimilate *in situ* observations of $p\text{CO}_2$ into a simplified global model. The model was formulated so as to have a single state variable, DIC, and was forced offline by physical and biological fields. Having just one state variable made the adjoint much simpler to derive. Data were then assimilated with a two-month observation window, and the model was found to better fit the $p\text{CO}_2$ data. Inverse methods have also been used by Mikaloff Fletcher et al. (2006), Gruber et al. (2009) and DeVries (2014), to reconstruct sources and sinks of anthropogenic carbon.

5.2 Dealing with non-Gaussian distributions

Most data assimilation methods assume Gaussian error distributions, but many marine biogeochemical variables are positive-definite and highly non-linear, with non-Gaussian error distributions. This section will discuss approaches taken in the literature to deal with this potential incompatibility. The focus will be on the assimilation of chlorophyll data, but many aspects will be relevant to other variables too. It is also useful to consider approaches taken in the meteorological community for variables such as precipitation (Bauer et al., 2011), and the techniques presented in Bocquet et al. (2010), which reviews progress in non-Gaussian data assimilation in geophysics.

Chlorophyll is generally observed to follow a logarithmic distribution in nature (Campbell, 1995). Therefore a common and simple technique to, at least approximately, normalise chlorophyll (and hopefully therefore its error distribution), is to log-transform it prior to assimilation (e.g. Gregg, 2008; Ciavatta et al., 2011; Ford et al., 2012). This was shown by Bertino et al. (2003) to give better results than using the raw variable, when using an EnKF to assimilate pseudo-observations into a simplified ecological model. However, Teruzzi et al. (2014) decided not to apply a transformation, noting that in their application “the EOF modes for log-transformed chlorophyll were less representative of the vertical covariance”, and “the asymmetry of the probability density distribution was not significantly eliminated by the log transformation”.

Because log-transformation does not guarantee a Gaussian error distribution, some studies have investigated the more sophisticated approach of applying an anamorphic transformation (Bertino et al., 2003; Bocquet et al., 2010; Brankart et al., 2012) instead. Studies to have implemented this approach, either with an EnKF or SEEK filter, include Lenartz et al. (2007), Simon and Bertino (2009), Béal et al. (2010), Simon and Bertino (2012) and Fontana et al. (2013).

An alternative to transforming the assimilated variable is to use an assimilation method which does not assume a Gaussian distribution. One such approach is to use particle filters (van Leeuwen, 2009), as discussed in section 5.1.3. Particle filters are starting to be applied successfully to marine biogeochemistry (e.g. Mattern et al., 2013), but such techniques are very much still under development. Another possibility is to extend the 3D-Var and 4D-Var approaches to work with lognormal distributions, as proposed by Fletcher (2010) and Fletcher and Jones (2014). This was demonstrated with a simple marine biogeochemical model by Song et al. (2012).

5.3 Generating multivariate increments

Currently, the only source of marine biogeochemical observations with widespread coverage is remotely sensed ocean colour. It is therefore important to try and exploit this data set (as well as other, sparser ones) to its fullest. This means developing assimilation methods which propagate the information to non-observed variables, including the sub-surface, by creating multivariate increments from univariate data.

One approach is to use an assimilation method which is inherently multivariate through its use of error covariances, such as the EnKF or SEEK filter, as detailed in sections 5.1.2 and 5.1.3. This approach has the advantage of being rooted in the mathematical derivation of the assimilation methodology, and is easily and consistently applied to all state variables. It can become computationally expensive though if ensembles are used to evolve the error covariances.

A second, computationally cheap, approach is to use a univariate assimilation method, and then apply balance relationships to generate multivariate increments. In most studies this just involves updating phytoplankton biomass (e.g. Teruzzi et al., 2014) and sometimes nutrients (e.g. Rousseaux and Gregg, 2012) based on existing model ratios. In the case of Hemmings et al. (2008) a more sophisticated scheme was developed, which produced increments to all state variables based on a statistical treatment of the univariate increments and the model state, and which aimed to conserve mass (nitrogen and carbon) at each model grid point. The conservation of mass is the principle under which most biogeochemical models are built, but whether data assimilation should adhere to this is a choice to be made – after all, the total mass of the system is unknown, and physical assimilation schemes do not typically conserve mass.

To date, most studies have paid minimal focus to the assessment of non-assimilated variables. However, improvements in such variables have been noted when expanding univariate schemes to be multivariate (Hemmings et al., 2008; Rousseaux and Gregg, 2012), and positive results have been found with both inherently multivariate methods (Ciavatta et al., 2011) and by using a balancing scheme (Ford et al., 2012). Often the impact remains marginal though, with no clear “best” approach. This may be a challenging area of research, with Fontana et al. (2013) noting that a lack of correlation between surface chlorophyll and sub-mixed layer nutrients can make it difficult to effectively correct nutrient errors. A potential approach to this problem is to combine state and parameter estimation, so that parameter values as well as state variables are modified online, as proposed for operational systems by Matear and Jones (2011). Such techniques have begun to be explored in studies such as Mattern et al. (2014).

As well as the above ideas, the question of what ocean colour information to assimilate is also relevant. Typically, chlorophyll concentration is used, as this is most easily related to model variables. However, radiance data can be assimilated instead. This has undergone less processing in its derivation, so observation errors are smaller. Furthermore, it represents a more fundamental property of the marine environment, which can be related to a greater number of variables. However, for this to be the case for a model, a detailed treatment of the optics is required, which many models currently lack (Gehlen et al., 2015). Nonetheless, the recent review paper of Gehlen et al. (2015) advocates the development of this approach, which has already been successfully demonstrated by Shulman et al. (2013) and Ciavatta et al. (2014). It will be particularly relevant for shelf seas, where suspended sediments and coloured dissolved organic matter play an important role.

5.4 Formulation of error covariances

Most assimilation schemes require *a priori* estimates of background and observation error covariances, with their subsequent treatment depending on the scheme. The specification of the background errors is of crucial importance, but effective estimation of these is generally more closely tied to the assimilation method used than the variable(s) it is used for, so the reader is referred to the more general assimilation literature (e.g. Bannister, 2008).

Observation error covariances are typically kept static during the assimilation process and are observation-specific. Since most marine biogeochemical assimilation studies assimilate chlorophyll derived from ocean colour, estimating observation error covariances for this variable is of most relevance to the community, with approaches likely to be generalised for other variables. In keeping with the current implementation of most physical assimilation schemes (Cummings et al., 2009), observation errors are typically assumed to be uncorrelated. This is an assumption which is likely to be often violated, but means that each (surface) model grid point just requires a single value for the error variance to be specified at that point. This will be comprised of both instrument error and representativeness error (that is, error due to the observations representing smaller scale processes than the model is able to resolve).

The simplest method is just to assume a constant (percentage) observation error, for instance 35% based on the pre-launch accuracy target for SeaWiFS (Hooker et al., 1992). This is a common approach (e.g. Natvik and Evensen, 2003a; Fontana et al., 2013). It can be made slightly more sophisticated by introducing a regional weighting (e.g. Nerger and Gregg, 2007; Gregg, 2008). Constant, and constant proportional, observation error variances were tested by Teruzzi et al. (2014), and found to be less effective than a more complex treatment of the errors, varying in space and time. More complex treatments have also been used in other studies (Ciavatta et al., 2011; Ford et al., 2012), but with no consensus on method found in the literature.

6. Impact of physical data assimilation on biogeochemistry

Biogeochemistry is strongly affected by the physical environment, and so the use of physical data assimilation, both on its own and as part of coupled physical-biogeochemical assimilation schemes, is expected to be vital for improving modelled biogeochemistry (Gehlen et al., 2015). However, the fields which are of greatest importance for biogeochemistry, notably vertical velocities, are not always those of greatest concern to developers and (other) users of physical modelling systems. Therefore data assimilation schemes are not necessarily tuned to these variables, and the impact can be minimal or even adverse, with knock-on implications for biogeochemical fields (Barciela et al., 2012; Brasseur et al., 2009). This section will review studies which have assimilated physical observations into a coupled physical-biogeochemical model for state estimation, and the impact found on the biogeochemical

fields. An overview will then be given of current, as yet unpublished, work at the Met Office to address issues found when assimilating physical data into FOAM-HadOCC. This section will focus on the one-way impact of assimilating physical data: whilst a few studies have assimilated physical and biogeochemical data concurrently (e.g. Anderson et al., 2000; Ourmières et al., 2009; Ford et al., 2012), the authors are unaware of any truly coupled assimilation schemes. This is noted as a future challenge for the community in the recent review of Gehlen et al. (2015).

Most studies to have assimilated physical data into a coupled physical-biogeochemical model have done so in a regional model, usually of the North Atlantic. Oschlies and Garçon (1998) assimilated sea surface height (SSH) data into a $1/3^\circ$ resolution model of the North Atlantic, performed in such a way that temperature and salinity were not changed on isopycnals, and the amount of nitrogen in each water column was conserved for each biological variable. The introduction of assimilation increased nutrient supply and primary production at mid-latitudes, which was concluded to highlight the importance of including eddy dynamics. Anderson et al. (2000) assimilated *in situ* physical and biogeochemical data from a particular observing campaign into a model of the Gulf Stream region. Best results were obtained when assimilating both data types, with a misalignment of physical and biogeochemical fronts found when assimilating just physical or just biogeochemical data. Popova et al. (2002a,b) assimilated physical and biogeochemical *in situ* cruise data into models of the North-East Atlantic, but made no comparison to control runs. Similarly, Beşiktepe et al. (2003) assimilated *in situ* physical and biogeochemical data into a model of Massachusetts Bay. Eden and Oschlies (2006) used the semiprognostic method (Sheng et al., 2001) to assimilate physical data into a model of the North Atlantic. The semiprognostic method applies a correction of the horizontal pressure gradient in the horizontal momentum equations, calculated from climatological temperature and salinity, and is designed to be adiabatic and conserve tracer and water mass properties. The assimilation increased ocean CO₂ uptake by 25%, better matching observations, which was attributed in particular to improvements in the representation of the Gulf Stream. Berline et al. (2007) used a SEEK filter to assimilate sea surface temperature (SST) and SSH observations, plus climatological sea surface salinity, into a model of the North Atlantic. Preliminary experiments resulted in “a strong increase of the vertical mixing of biogeochemical tracers in the surface layers”, due to the creation of hydrostatic instabilities. This was addressed for the main experiments by implementing a scheme similar to that of Cooper and Haines (1996) to perform the vertical extrapolation of increments, and by not applying corrections south of

10°N (the domain extended from 70°N to 20°S). In the main experiments the assimilation resulted in a 50% reduction in nutrient input to the euphotic zone due to vertical mixing, but a 400-600% increase due to advection. This varied with latitude, with an overall large increase in surface nutrients in the subtropical gyre, and a moderate decrease at mid to high latitudes. The ecosystem response was concluded to be weak and mixed, but overall a better match for observations. Physical assimilation was therefore concluded to be beneficial. Ourmières et al. (2009) used a version of the same model and assimilation scheme to assimilate physical and nutrient data both separately and concurrently. It was concluded that physical assimilation is beneficial if nutrient concentrations are already generally correct, but can make any existing errors worse. The effect of vertical mixing errors in the model was then investigated by Béal et al. (2010), through perturbations to the wind forcing. Samuelsen et al. (2009) assimilated SST, SSH and sea ice concentration data into a model of the North Atlantic and Arctic Ocean. The impact on the biogeochemical fields was minimal, with the biggest differences resulting from changes in ice extent. Fiechter et al. (2011) assimilated SST, SSH and *in situ* temperature and salinity profile data into a model of the Coastal Gulf of Alaska, and found that improvements to the simulation of mesoscale processes resulted in improvements to the biogeochemical fields. Edwards et al. (2012) validated the Met Office's NEMO-ERSEM system for the North-West European Shelf, which includes assimilation of SST data, but made no comparison with a non-assimilative run.

The most serious issues resulting from physical data assimilation have typically been noted in global models, particularly in tropical regions (Barciela et al., 2012). However, very little discussion of this can be found in the literature. While et al. (2010) assimilated *in situ* temperature and salinity observations into a 2° global model, and found the assimilation resulted in unrealistically increased advective upwelling of nutrients within boundary currents, which spread into the subtropical gyres. A “nutrient increment method” was introduced, which applies a correction to the nutrient profiles to ensure that nutrient-potential density relationships are maintained following assimilation. The method had little impact on the issues in the subtropical gyres, but did result in improvements elsewhere. El Moussaoui et al. (2011) describe how assimilating physical data into the Mercator global forecasting system results in unrealistically high vertical velocities in equatorial regions, leading to unrealistic increases in nutrient input to the euphotic zone in these areas, and over-estimated surface chlorophyll concentrations. Gehlen et al. (2015) note that this chlorophyll bias is reduced in the most recent version of the system, partly due to changes to the assimilation of SSH data, and partly due to changes to the

physical and biogeochemical models. In a Biogeosciences Discussions paper (stated at <http://www.biogeosciences-discuss.net/11/5399/2014/bgd-11-5399-2014-discussion.html> to be unlikely to lead to a final peer-reviewed version) Visinelli et al. (2014) investigated the impact on the carbonate system of assimilating *in situ* temperature and salinity observations into a 2° global model. On a global scale, carbon variables were improved compared with observations. However, issues due to spurious upwelling were also found, similar to those described in While et al. (2010) – this was touched on in the paper, and elaborated on during discussions with reviewers (see above link).

A degradation of biogeochemical fields has also been found when assimilating physical data into the Met Office's global FOAM-HadOCC model, as noted by Ford et al. (2012) and While et al. (2012). This has been found with both the analysis correction scheme of Martin et al. (2007) and the 3D-Var scheme of Waters et al. (2015), with both NEMO and the previous ocean model (Gordon et al., 2000) used by FOAM, and at both 1° and 1/4° resolution. The remainder of this section will give a brief overview of the issues seen, and work being done to address these. This work is being performed by the authors in collaboration with Jennie Waters and Matt Martin, also at the Met Office. In particular, development of the “instantaneous pressure correction scheme” described below has been led by Jennie Waters.

Assimilating physical data into FOAM-HadOCC has been found to substantially increase the input of nutrients into the euphotic zone, particularly in equatorial regions, which results in an increase in surface chlorophyll. This is a degradation compared to observed fields. The issue is only encountered when corrections are made to temperature and salinity profiles – either as a result of the direct assimilation of these variables, or through balances applied when assimilating SSH data. A large increase in vertical velocities is seen in these regions, as well as a degradation of surface currents at the equator. Biogeochemical impacts can also be seen in highly variable regions such as the Gulf Stream and Antarctic Circumpolar Current. The assimilation leads to changes in both advective and diffusive processes, but the relative importance varies with depth and latitude. There is evidence of the creation of hydrostatic instabilities, although impacts occur even when this is not the case, so the problems cannot be solely attributed to this.

Not applying assimilation increments close to the equator considerably reduces the impact on vertical velocities and surface currents, with nutrient profiles also looking more realistic, although still with a small increase at the surface. However, this still leads to a

large increase in surface chlorophyll in nutrient-limited areas, suggesting that the biogeochemistry is extremely sensitive to the assimilation and so a robust solution is required. Widening the latitude band in which observations are not assimilated helps to alleviate the issues further, but simply not assimilating data in the tropics is not considered a satisfactory long-term solution. Changing how baroclinic velocity increments are applied near the equator has also been investigated, but has made minimal difference. Furthermore, the divergence damping scheme of Dobricic et al. (2007) has been tested, which filters the increments to the horizontal velocities to reduce their divergence, with the aim of not adding any further divergence component to the model velocity fields. Whilst this reduced the vertical velocities as intended, the magnitude of the reduction was small, so its overall impact was minimal.

Current investigative work follows three main strands. Firstly, a thorough investigation of the fundamental mechanisms responsible, to aid the development of a targeted solution. Secondly, testing of short-term workarounds to allow at least the partial assimilation of physical data without major degradation of the biogeochemical fields. Thirdly, the development of an instantaneous pressure correction scheme. This aims to correct the imbalance caused by the assimilation altering sub-surface pressure gradients whilst leaving the wind stress unchanged, which can lead to the generation of spurious equatorial waves and therefore vertical velocities. The currently used pressure correction scheme of Bell et al. (2004) addresses this on longer timescales by applying a correction based on accumulated temperature and salinity increments. An imbalance is still observed on shorter timescales though, and so the new instantaneous pressure correction scheme complements this by applying a correction based on temperature and salinity increments from the current analysis step.

7. Discussion and recommendations

Marine biogeochemical data assimilation is receiving an increasing amount of attention, and whilst challenges remain, has matured to the point where it is suitable for operational forecasting and reanalysis applications (Gehlen et al., 2015). A number of techniques have been used for parameter estimation, although computational constraints still limit their use for full 3D models. State estimation is usually performed by assimilating surface chlorophyll data derived from ocean colour. The two most commonly used techniques are the SEEK filter and EnKF, although other methods in

current use include OI, CRAM, 3D-Var, SEIK filter and SIR particle filter. Within MyOcean, an EnKF is used for operational forecasting of the Arctic Ocean (Simon and Bertino, 2009; Simon and Bertino, 2012; Gehlen et al., 2015), and 3D-Var for operational forecasting of the Mediterranean Sea (Teruzzi et al., 2014). More widely, reanalysis capability is being developed for the North Atlantic using a SEEK filter (Fontana et al., 2013), and for the global ocean using CRAM (Gregg, 2008; Gregg and Rousseaux, 2014). Met Office capability for the global ocean is currently being transitioned from OI to 3D-Var.

State estimation methods assimilating chlorophyll data have consistently resulted in improvements to model chlorophyll compared with both the assimilated data, and independent *in situ* observations (Gehlen et al., 2015). Many studies suggest that further improvements to chlorophyll are most likely to be achieved by correcting underlying errors in other variables such as nutrients, rather than simply concentrating on better fitting the assimilated data (although algorithmic improvements are doubtless possible). It should also be remembered that errors in satellite chlorophyll data are relatively large, and so there is a danger of over-fitting to the data. Therefore, the most important challenge, and consideration when deciding between assimilation methods, is the ability to produce multivariate increments to correct errors in other model variables. To date, few studies have paid much attention to validating non-assimilated variables, and the results that have been presented often show quite a small impact, which is sometimes beneficial and sometimes degrading. Whilst encouraging results on this front are beginning to be obtained (e.g. Ciavatta et al., 2014), no one method has clearly shown itself to be superior, despite large variations in computational cost.

Currently, the Met Office has the capability to assimilate surface chlorophyll and pCO₂ data into the global FOAM-HadOCC model, using 3D-Var (recently transitioned from OI) combined with the multivariate balancing scheme of Hemmings et al. (2008). Future development of this capability should account for the current state-of-the-art, as documented in this literature review, alongside constraints on resource. Furthermore, it is planned to transition from using HadOCC to using the MEDUSA (Model of Ecosystem Dynamics, nutrient Utilisation, Sequestration and Acidification) model (Yool et al., 2013). Any planned developments to the biogeochemical data assimilation should also account for developments to the physical data assimilation. Using the same basic framework for all ocean data assimilation within the Met Office is much easier to implement and maintain, enables effective sharing of resources, and allows a consistent approach to

any coupling of physics and biogeochemistry. Therefore, unless there are compelling reasons to take a fundamentally different approach, the default choice should be to align development of the physical and biogeochemical data assimilation systems.

Physical data assimilation in the Met Office uses the 3D-Var NEMOVAR scheme, as implemented by Waters et al. (2015), and the continued use of NEMOVAR is planned for the foreseeable future. Long-term developments are likely to take the form of using ensemble information to update the background error covariances, in a similar manner to the hybrid ensemble/4D-Var scheme the Met Office uses for operational NWP (Clayton et al., 2013). Currently, a single deterministic FOAM run is performed operationally, but in the future an ensemble might be run instead, from which such information could be obtained. This is planned to take the form of a fully coupled ocean-atmosphere forecasting system, envisaged to eventually replace the separate systems the Met Office currently runs. Coupled ocean-atmosphere data assimilation will be included, and is being developed at the moment. Shelf seas data assimilation is likely to follow a similar development path.

As noted above, there are a number of methods in use for biogeochemical state estimation, with no method standing out as the “best”. The approach currently taken by the Met Office gives at least as encouraging results as other state-of-the-art techniques, and is the most sophisticated balancing scheme approach documented in the literature. Furthermore, it is computationally much cheaper than ensemble techniques such as the EnKF, which is an important consideration for an operational centre, and allows resources to be focussed towards other developments such as increasing model resolution. Since the current approach is in line with Met Office requirements, has clear potential for development, and no other methods in the literature have yet demonstrated a clear superiority, the authors propose to maintain and develop the current approach of using NEMOVAR with multivariate balancing. Doing so will also help maintain scientific diversity within the community, although this decision should of course be periodically re-evaluated as other techniques are developed and refined.

Given this, the following development path is suggested for global marine biogeochemical assimilation within the Met Office (shelf seas will be discussed afterwards). Actions are loosely sorted into short-, medium- and long-term, which are deliberately undefined in terms of dates. No direct collaborations are included, but links should be established/maintained with other groups performing data assimilation within

NCOF (National Centre for Ocean Forecasting), MyOcean/Copernicus, and GODAE OceanView. Continuous development based on the latest results, and on developments to the physical assimilation scheme and modelling systems, is implied rather than explicitly included.

Short-term:

- Finalise the transition from OI to NEMOVAR.
- The chlorophyll and $p\text{CO}_2$ assimilation schemes both produce increments to DIC and alkalinity. A method should be developed to balance these increments so that chlorophyll and $p\text{CO}_2$ data can be assimilated simultaneously.
- Continue investigations into the impact of physical assimilation on biogeochemical fields, and the development of techniques to tackle this.
- Whilst maintaining the general approach to the calculation of error covariances taken by Ford et al. (2012), adapt the chlorophyll observation error variances to use the per-pixel error estimates which are included with satellite products.
- When transition is made from HadOCC to MEDUSA, transition the assimilation so that it is performed in the same manner as for HadOCC – univariate increments produced using NEMOVAR, followed by multivariate balancing. The scheme of Hemmings et al. (2008) will need some adaptation because MEDUSA has extra state variables, but the initial implementation can apply increments to these using existing model ratios.
- Once an assimilation capability for MEDUSA has been established, further develop the scheme of Hemmings et al. (2008) to be better tailored for MEDUSA.

Medium-term:

- Incorporate the multivariate balancing schemes within the NEMOVAR framework, as a step towards a fully coupled assimilation capability.
- Investigate the option of assimilating single-sensor along-track chlorophyll products rather than merged products, and whether a bias correction scheme would be required. Develop this capability if deemed appropriate.
- Assess the potential for developing the 3D-Var methodology to account for non-Gaussian distributions, as proposed by Fletcher and Jones (2014), as an alternative to log-transformation of chlorophyll.

Long-term:

- Investigate the use of ensemble information to update the background error covariances.
- Expand the assimilation capability to other variables such as nutrients or oxygen.
- Develop balance relationships between physical and biogeochemical variables, so that information from physical observations is used when calculating increments to biogeochemical variables (and potentially *vice versa*).
- Assess the potential for balance relationships to update model parameters, as well as the model state, as proposed by Matear and Jones (2011).

The Met Office currently lacks a biogeochemical assimilation capability for shelf seas. It is clear that this could be extremely beneficial, and the authors recommend that this capability should be developed as a matter of priority. Where possible, this should be aligned with the global schemes. However, ERSEM is a much more complex model than HadOCC or MEDUSA, so simply adapting the scheme of Hemmings et al. (2008) would not be straightforward. An assimilation capability for ERSEM in the North-West European shelf has been developed by PML (Plymouth Marine Laboratory), with whom the Met Office already collaborate on model development as part of the NCOF consortium. It would therefore seem sensible to collaborate with PML on the development of data assimilation. The approach taken by PML (Ciavatta et al., 2011; Ciavatta et al., 2014) is based on the EnKF, which would be computationally unfeasible for the Met Office to implement operationally, and is not aligned with the Met Office's use of NEMOVAR for physical data assimilation. This means the PML scheme cannot simply be used by the Met Office, a new approach would need to be developed. Due to the differing needs of the organisations, a single scheme to be used by both parties is unlikely to be practical, at least in the short-term, but elements could be shared where appropriate. PML expertise could also be sought if developing the capability to generate ensembles, as is the long-term aim for physical data assimilation (see above).

Any collaboration between the Met Office and PML is obviously subject to agreement between both parties, and suitable funding arrangements. However, a potential development plan for Met Office shelf seas biogeochemical assimilation is as follows:

Short- to medium-term:

- Develop an initial chlorophyll assimilation scheme which uses NEMOVAR to produce surface chlorophyll increments, with the multiple ERSEM phytoplankton types updated using existing model ratios.

- Extend this scheme to produce multivariate balancing increments, in a similar manner to Hemmings et al. (2008). A starting point, if collaborating with PML, could be to base this on a climatological set of correlations obtained through running the EnKF at PML. However, the exact approach should be the subject of scientific discussions between interested parties.

Medium- to long-term:

- Make further use of the information from satellite ocean colour, either by assimilating radiance data, as per Ciavatta et al. (2014), or by assimilating both chlorophyll and SPM (suspended particulate matter) products, or phytoplankton functional type products.
- Extend the assimilation capability to other data types, such as nutrients and $p\text{CO}_2$. Since *in situ* observations are sparse, a compromise will be needed between the desire to use all available information for assimilation, and the desire to maintain independent data sets for validation.
- More general developments to the assimilation methodology, as detailed for the global system.

Finally, the Met Office currently makes no use of data assimilation for parameter estimation. Any updates to model parameters are typically made by simply making some changes which might scientifically be expected to result in an improvement, running the model, assessing how model fields compare to large-scale observations, and repeating until satisfied. Given the computational constraints on performing parameter estimation for 3D models, this methodology does still have its place. Nonetheless, improved results would be expected by coupling this with more formal techniques based on data assimilation. A generic framework known as MarMOT (Marine Model Optimisation Testbed) has recently been developed by Hemmings and Challenor (2012) and Hemmings et al. (2014). This allows the user to thoroughly account for uncertainty in the physical model, and is at the forefront of parameter estimation research. Furthermore, the code is open source, and interfaces for HadOCC and MEDUSA have already been developed. An assessment of the potential of MarMOT for parameter estimation within the Met Office is therefore recommended, to aid model development on all time-scales.

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, 399-411.
- Anderson, J. L. (2010). A non-Gaussian ensemble filter update for data assimilation. *Monthly Weather Review*, 138(11), 4186-4198.
- Anderson, L. A., & Sarmiento, J. L. (1995). Global ocean phosphate and oxygen simulations. *Global Biogeochemical Cycles*, 9(4), 621-636.
- Anderson, L. A., Robinson, A. R., & Lozano, C. J. (2000). Physical and biological modeling in the Gulf Stream region: I. Data assimilation methodology. *Deep Sea Research Part I: Oceanographic Research Papers*, 47(10), 1787-1827.
- Armstrong, R. A., Sarmiento, J. L., & Slater, R. D. (1995). Monitoring ocean productivity by assimilating satellite chlorophyll into ecosystem models. In *Ecological time series* (pp. 371-390). Springer US.
- Athias, V., Mazzega, P., & Jeandel, C. (2000). Selecting a global optimization method to estimate the oceanic particle cycling rate constants. *Journal of Marine Research*, 58(5), 675-707.
- Bannister, R. N. (2008). A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characteristics and measurements of forecast error covariances. *Quarterly Journal of the Royal Meteorological Society*, 134(637), 1951-1970.
- Barciela, R., Brasseur, P., & Wilmer-Becker, K. (2012). Marine Ecosystem Prediction Task Team (MEP-TT) report. Available from <https://www.godae-oceanview.org/>.
- Bauer, P., Auligné, T., Bell, W., Geer, A., Guidard, V., Heilliette, S., Kazumori, M., Kim, M.-J., Liu, E. H.-C., McNally, A. P., Macpherson, B., Okamoto, K., Renshaw, R. and Riishøjgaard, L.-P. (2011). Satellite cloud and precipitation assimilation at operational NWP centres. *Quarterly Journal of the Royal Meteorological Society*, 137(661), 1934-1951.
- Béal, D., Brasseur, P., Brankart, J. M., Ourmières, Y., & Verron, J. (2010). Characterization of mixing errors in a coupled physical biogeochemical model of the North Atlantic: implications for nonlinear estimation using Gaussian anamorphosis. *Ocean Science*, 6(1), 247-262.
- Bell, M. J., Martin, M. J., & Nichols, N. K. (2004). Assimilation of data into an ocean model with systematic errors near the equator. *Quarterly Journal of the Royal Meteorological Society*, 130(598), 873-893.
- Berline, L., Brankart, J. M., Brasseur, P., Ourmières, Y., & Verron, J. (2007). Improving the physics of a coupled physical-biogeochemical model of the North Atlantic through data assimilation: Impact on the ecosystem. *Journal of Marine Systems*, 64(1), 153-172.
- Bertino, L., Evensen, G., & Wackernagel, H. (2003). Sequential data assimilation techniques in oceanography. *International Statistical Review*, 71(2), 223-241.
- Beşiktepe, Ş. T., Lermusiaux, P. F., & Robinson, A. R. (2003). Coupled physical and biogeochemical data-driven simulations of Massachusetts Bay in late summer: real-time and postcruise data assimilation. *Journal of Marine Systems*, 40, 171-212.
- Bishop, C. H., Etherton, B. J., & Majumdar, S. J. (2001). Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Monthly weather review*, 129(3), 420-436.
- Blockley, E. W., Martin, M. J., McLaren, A. J., Ryan, A. G., Waters, J., Lea, D. J., Mirouze, I., Peterson, K. A., Sellar, A., & Storkey, D. (2014). Recent development of the Met Office operational ocean forecasting system: an overview and assessment of the new Global FOAM forecasts. *Geosci. Model Dev.*, 7, 2613-2638.

- Bocquet, M., Pires, C. A., & Wu, L. (2010). Beyond Gaussian statistical modeling in geophysical data assimilation. *Monthly Weather Review*, 138(8), 2997-3023.
- Brankart, J. M., Testut, C. E., Béal, D., Doron, M., Fontana, C., Meinvielle, M., Brasseur, P., & Verron, J. (2012). Towards an improved description of ocean uncertainties: effect of local anamorphic transformations on spatial correlations. *Ocean Science*, 8(2), 121-142.
- Brasseur, P., & Verron, J. (2006). The SEEK filter method for data assimilation in oceanography: a synthesis. *Ocean Dynamics*, 56(5-6), 650-661.
- Brasseur, P., Gruber, N., Barciela, R., Brander, K., Doron, M., El Moussaoui, A., Hobday, A. J., Huret, M., Kremer, A.-S., Lehoudey, P., Matear, R., Moulin, C., Murtugudde, R., Senina, I., & Svendsen, E. (2009). Integrating biogeochemistry and ecology into ocean data assimilation systems. *Oceanography*, 22(3), 206-215.
- Butenschön, M., & Zavatarelli, M. (2012). A comparison of different versions of the SEEK Filter for assimilation of biogeochemical data in numerical models of marine ecosystem dynamics. *Ocean Modelling*, 54, 37-54.
- Campbell, J. W. (1995). The lognormal distribution as a model for bio-optical variability in the sea. *Journal of Geophysical Research: Oceans* (1978–2012), 100(C7), 13237-13254.
- Carmillet, V., Brankart, J. M., Brasseur, P., Drange, H., Evensen, G., & Verron, J. (2001). A singular evolutive extended Kalman filter to assimilate ocean color data in a coupled physical–biochemical model of the North Atlantic ocean. *Ocean Modelling*, 3(3), 167-192.
- Ciavatta, S., & Pastres, R. (2011). Exploring the long-term and interannual variability of biogeochemical variables in coastal areas by means of a data assimilation approach. *Estuarine, Coastal and Shelf Science*, 91(3), 411-422.
- Ciavatta, S., Torres, R., Saux-Picart, S., & Allen, J. I. (2011). Can ocean color assimilation improve biogeochemical hindcasts in shelf seas?. *Journal of Geophysical Research: Oceans* (1978–2012), 116(C12).
- Ciavatta, S., Torres, R., Martinez-Vicente, V., Smyth, T., Dall’Olmo, G., Polimene, L., & Allen, J. I. (2014). Assimilation of remotely-sensed optical properties to improve marine biogeochemistry modelling. *Progress in Oceanography*, 127, 74-95.
- Clayton, A. M., Lorenc, A. C., & Barker, D. M. (2013). Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. *Quarterly Journal of the Royal Meteorological Society*, 139(675), 1445-1461.
- Cooper, M., & Haines, K. (1996). Altimetric assimilation with water property conservation. *Journal of Geophysical Research: Oceans* (1978–2012), 101(C1), 1059-1077.
- Cossarini, G., Lermusiaux, P. F. J., & Solidoro, C. (2009). Lagoon of Venice ecosystem: Seasonal dynamics and environmental guidance with uncertainty analyses and error subspace data assimilation. *Journal of Geophysical Research: Oceans* (1978–2012), 114(C6).
- Courtier, P., Thépaut, J. N., & Hollingsworth, A. (1994). A strategy for operational implementation of 4D-Var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society*, 120(519), 1367-1387.
- Cummings, J., Bertino, L., Brasseur, P., Fukumori, I., Kamachi, M., Martin, M. J., Mogensen, K., Oke, P., Testut, C. E., Verron, J., & Weaver, A. (2009). Ocean data assimilation systems for GODAE. *Oceanography*, 22(3), 96-109.
- Daley, R. (1991). *Atmospheric data analysis*. Cambridge University Press.
- Dee, D. P., & Da Silva, A. M. (1998). Data assimilation in the presence of forecast bias. *Quarterly Journal of the Royal Meteorological Society*, 124(545), 269-295.
- DeVries, T. (2014). The oceanic anthropogenic CO₂ sink: Storage, air-sea fluxes, and transports over the industrial era. *Global Biogeochemical Cycles*, 28(7), 631-647.

- Dobricic, S., Pinardi, N., Adani, M., Tonani, M., Fratianni, C., Bonazzi, A., & Fernandez, V. (2007). Daily oceanographic analyses by Mediterranean Forecasting System at the basin scale. *Ocean Science*, 3(1), 149-157.
- Doron, M., Brasseur, P., & Brankart, J. M. (2011). Stochastic estimation of biogeochemical parameters of a 3D ocean coupled physical–biogeochemical model: Twin experiments. *Journal of Marine Systems*, 87(3), 194-207.
- Doron, M., Brasseur, P., Brankart, J. M., Losa, S. N., & Melet, A. (2013). Stochastic estimation of biogeochemical parameters from Globcolour ocean colour satellite data in a North Atlantic 3D ocean coupled physical–biogeochemical model. *Journal of Marine Systems*, 117, 81-95.
- Dowd, M., & Meyer, R. (2003). A Bayesian approach to the ecosystem inverse problem. *Ecological modelling*, 168(1), 39-55.
- Dowd, M. (2006). A sequential Monte Carlo approach for marine ecological prediction. *Environmetrics*, 17(5), 435-455.
- Dowd, M. (2007). Bayesian statistical data assimilation for ecosystem models using Markov Chain Monte Carlo. *Journal of Marine Systems*, 68(3), 439-456.
- Dowd, M., Jones, E., & Parslow, J. (2014). A statistical overview and perspectives on data assimilation for marine biogeochemical models. *Environmetrics*, 25(4), 203-213.
- Eden, C., & Oschlies, A. (2006). Adiabatic reduction of circulation-related CO₂ air-sea flux biases in a North Atlantic carbon-cycle model. *Global Biogeochemical Cycles*, 20(2).
- Edwards, K. P., Barciela, R., & Butenschön, M. (2012). Validation of the NEMO-ERSEM operational ecosystem model for the North West European Continental Shelf. *Ocean Science*, 8(6), 983-1000.
- Eknes, M., & Evensen, G. (2002). An ensemble Kalman filter with a 1-D marine ecosystem model. *Journal of Marine Systems*, 36(1), 75-100.
- El Moussaoui, A., Perruche, C., Greiner, E., Ethé, C., & Gehlen, M. (2011). Integration of biogeochemistry into Mercator Ocean systems. *Mercator Ocean Newsletter #40*, January 2011.
- Evans, G. T. (2003). Defining misfit between biogeochemical models and data sets. *Journal of marine systems*, 40, 49-54.
- Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans* (1978–2012), 99(C5), 10143-10162.
- Evensen, G., & van Leeuwen, P. J. (2000). An ensemble Kalman smoother for nonlinear dynamics. *Monthly Weather Review*, 128(6), 1852-1867.
- Evensen, G. (2003). The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, 53(4), 343-367.
- Fasham, M. J. R., & Evans, G. T. (1995). The Use of Optimization Techniques to Model Marine Ecosystem Dynamics at the JGOFS Station at 47 degrees N 20 degrees W. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 348(1324), 203-209.
- Fasham, M. J., Boyd, P. W., & Savidge, G. (1999). Modeling the relative contributions of autotrophs and heterotrophs to carbon flow at a Lagrangian JGOFS station in the Northeast Atlantic: The importance of DOC. *Limnology and Oceanography*, 44(1), 80-94.
- Faugeras, B., Lévy, M., Mémery, L., Verron, J., Blum, J., & Charpentier, I. (2003). Can biogeochemical fluxes be recovered from nitrate and chlorophyll data? A case study assimilating data in the Northwestern Mediterranean Sea at the JGOFS-DYFAMED station. *Journal of Marine Systems*, 40, 99-125.
- Faugeras, B., Bernard, O., Sciandra, A., & Lévy, M. (2004). A mechanistic modelling and data assimilation approach to estimate the carbon/chlorophyll and carbon/nitrogen ratios in a coupled hydrodynamical-biological model. *Nonlinear Processes in Geophysics*, 11, 515-533.

- Fennel, K., Losch, M., Schröter, J., & Wenzel, M. (2001). Testing a marine ecosystem model: sensitivity analysis and parameter optimization. *Journal of Marine Systems*, 28(1), 45-63.
- Fiechter, J., Broquet, G., Moore, A. M., & Arango, H. G. (2011). A data assimilative, coupled physical–biological model for the Coastal Gulf of Alaska. *Dynamics of Atmospheres and Oceans*, 52(1), 95-118.
- Fletcher, S. J. (2010). Mixed Gaussian-lognormal four-dimensional data assimilation. *Tellus A*, 62(3), 266-287.
- Fletcher, S. J., & Jones, A. S. (2014). Multiplicative and Additive Incremental Variational Data Assimilation for Mixed Lognormal and Gaussian Errors. *Monthly Weather Review*.
- Fontana, C., Grenz, C., Pinazo, C., Marsaleix, P., & Diaz, F. (2009). Assimilation of SeaWiFS chlorophyll data into a 3D-coupled physical–biogeochemical model applied to a freshwater-influenced coastal zone. *Continental Shelf Research*, 29(11), 1397-1409.
- Fontana, C., Grenz, C., & Pinazo, C. (2010). Sequential assimilation of a year-long time-series of SeaWiFS chlorophyll data into a 3D biogeochemical model on the French Mediterranean coast. *Continental Shelf Research*, 30(16), 1761-1771.
- 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(1), 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*, 8(5), 751-771.
- Friedrichs, M. A. (2001). A data assimilative marine ecosystem model of the central equatorial Pacific: Numerical twin experiments. *Journal of Marine Research*, 59(6), 859-894.
- Friedrichs, M. A. (2002). Assimilation of JGOFS EqPac and SeaWiFS data into a marine ecosystem model of the central equatorial Pacific Ocean. *Deep Sea Research Part II: Topical Studies in Oceanography*, 49(1), 289-319.
- Friedrichs, M. A., 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), 576-600.
- Friedrichs, M. A., Dusenberry, J. A., Anderson, L. A., Armstrong, R. A., Chai, F., Christian, J. R., Doney, S. C., Dunne, J., Fujii, M., Hood, R., McGillicuddy Jr., D. J., Moore, J. K., Schartau, M., Spitz, Y. H., & Wiggert, J. D. (2007). Assessment of skill and portability in regional marine biogeochemical models: role of multiple planktonic groups. *Journal of Geophysical Research: Oceans* (1978–2012), 112(C8).
- Gandin, L. S., (1963). Objective analysis of meteorological fields. Leningrad: Gridromet (in Russian) (1963), English translation by the Israel program for scientific translations, 1965, pp 242.
- Garcia-Gorriz, E., Hoepffner, N., & Ouberdous, M. (2003). Assimilation of SeaWiFS data in a coupled physical–biological model of the Adriatic Sea. *Journal of Marine Systems*, 40, 233-252.
- Gehlen, M., Barciela, R., Bertino, L., Brasseur, P., Butenschön, M., Chai, F., Crise, A., Drillet, Y., Ford, D., Lavoie, D., Lehodey, P., Perruche, C., Samuelsen, A., & Simon, E. (2015). Building the capacity for forecasting marine biogeochemistry and ecosystems: recent advances and future developments. *Journal of Operational Oceanography*, 8(sup1), s168-s187.
- Gerber, M., Joos, F., Vázquez-Rodríguez, M., Touratier, F., & Goyet, C. (2009). Regional air-sea fluxes of anthropogenic carbon inferred with an Ensemble Kalman Filter. *Global Biogeochemical Cycles*, 23(1).

- Gerber, M., & Joos, F. (2010). Carbon sources and sinks from an Ensemble Kalman Filter ocean data assimilation. *Global Biogeochemical Cycles*, 24(3).
- Gordon, C., Cooper, C., Senior, C. A., Banks, H., Gregory, J. M., Johns, T. C., Mitchell, J. F. B., & Wood, R. A. (2000). The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Climate Dynamics*, 16(2-3), 147-168.
- Gordon, N. J., Salmond, D. J., & Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE-Proceedings-F*, 140, 107–113.
- Gregg, W. W., & Conkright, M. E. (2001). Global seasonal climatologies of ocean chlorophyll: Blending in situ and satellite data for the Coastal Zone Color Scanner era. *Journal of Geophysical Research: Oceans* (1978–2012), 106(C2), 2499-2515.
- Gregg, W. W. (2008). Assimilation of SeaWiFS ocean chlorophyll data into a three-dimensional global ocean model. *Journal of Marine Systems*, 69(3), 205-225.
- Gregg, W. W., Friedrichs, M. A., 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), 16-33.
- Gregg, W. W., & Rousseaux, C. S. (2014). Decadal trends in global pelagic ocean chlorophyll: A new assessment integrating multiple satellites, in situ data, and models. *Journal of Geophysical Research: Oceans*, 119(9), 5921-5933.
- Gruber, N., Gloor, M., Mikaloff Fletcher, S. E., Doney, S. C., Dutkiewicz, S., Follows, M. J., Gerber, M., Jacobson, A. R., Joos, F., Lindsay, K., Menemenlis, D., Mouchet, A., Müller, S. A., Sarmiento, J. L., & Takahashi, T. (2009). Oceanic sources, sinks, and transport of atmospheric CO₂. *Global Biogeochemical Cycles*, 23(1).
- Harmon, R., & Challenor, P. (1997). A Markov chain Monte Carlo method for estimation and assimilation into models. *Ecological modelling*, 101(1), 41-59.
- Hemmings, J. C., Srokosz, M. A., Challenor, P., & Fasham, M. J. (2003). Assimilating satellite ocean-colour observations into oceanic ecosystem models. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 361(1802), 33-39.
- Hemmings, J. C., Srokosz, M. A., Challenor, P., & Fasham, M. J. (2004). Split-domain calibration of an ecosystem model using satellite ocean colour data. *Journal of Marine Systems*, 50(3), 141-179.
- Hemmings, J. C., Barciela, R. M., & Bell, M. J. (2008). Ocean color data assimilation with material conservation for improving model estimates of air-sea CO₂ flux. *Journal of Marine Research*, 66(1), 87-126.
- Hemmings, J. C. P., & Challenor, P. G. (2012). Addressing the impact of environmental uncertainty in plankton model calibration with a dedicated software system: the Marine Model Optimization Testbed (MarMOT 1.1 alpha). *Geoscientific Model Development*, 5(2), 471-498.
- Hemmings, J. C. P., Challenor, P. G., & Yool, A. (2014). Mechanistic site-based emulation of a global ocean biogeochemical model for parametric analysis and calibration. *Geoscientific Model Development Discussions*, 7(5), 6327-6411.
- Hoke, J. E., & Anthes, R. A. (1976). The initialization of numerical models by a dynamic-initialization technique. *Monthly Weather Review*, 104(12), 1551-1556.
- Hooker, S., Firestone, E., Esaias, W., Feldman, G., Gregg, W., & McClain, C. (1992). SeaWiFS technical report series. Volume 1: An overview of SeaWiFS and ocean color.
- 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, 3171-3185.
- 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.
- Huret, M., Gohin, F., Delmas, D., Lunven, M., & Garçon, V. (2007). Use of SeaWiFS data for light availability and parameter estimation of a phytoplankton production model of the Bay of Biscay. *Journal of Marine Systems*, 65(1), 509-531.
- Hurtt, G. C., & Armstrong, R. A. (1996). A pelagic ecosystem model calibrated with BATS data. *Deep Sea Research Part II: Topical Studies in Oceanography*, 43(2), 653-683.
- Hurtt, G. C., & Armstrong, R. A. (1999). A pelagic ecosystem model calibrated with BATS and OWSI data. *Deep-Sea Research Part I*, 46(1), 27-61.
- Ishizaka, J. (1990). Coupling of coastal zone color scanner data to a physical-biological model of the southeastern US continental shelf ecosystem: 3. Nutrient and phytoplankton fluxes and CZCS data assimilation. *Journal of Geophysical Research*, 95, 20201-20212.
- Jazwinski, A. H. (1970). *Stochastic processes and filtering theory*. Academic Press.
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Fluids Engineering*, 82(1), 35-45.
- Kalnay, E. (2003). *Atmospheric modeling, data assimilation, and predictability*. Cambridge University Press.
- Kane, A., Moulin, C., Thiria, S., Bopp, L., Berrada, M., Tagliabue, A., Crépon, M., Aumont, O., & Badran, F. (2011). Improving the parameters of a global ocean biogeochemical model via variational assimilation of in situ data at five time series stations. *Journal of Geophysical Research: Oceans* (1978–2012), 116(C6).
- Kidston, M., Matear, R., & Baird, M. E. (2011). Parameter optimisation of a marine ecosystem model at two contrasting stations in the Sub-Antarctic Zone. *Deep Sea Research Part II: Topical Studies in Oceanography*, 58(21), 2301-2315.
- Kidston, M., Matear, R., & Baird, M. E. (2013). Phytoplankton growth in the Australian sector of the Southern Ocean, examined by optimising ecosystem model parameters. *Journal of Marine Systems*, 128, 123-137.
- Kivman, G. A. (2003). Sequential parameter estimation for stochastic systems. *Nonlinear Processes in Geophysics*, 10(3), 253-259.
- Korres, G., Triantafyllou, 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.
- Kuroda, H., & Kishi, M. J. (2004). A data assimilation technique applied to estimate parameters for the NEMURO marine ecosystem model. *Ecological Modelling*, 172(1), 69-85.
- Kwon, E. Y., & Primeau, F. (2006). Optimization and sensitivity study of a biogeochemistry ocean model using an implicit solver and in situ phosphate data. *Global biogeochemical cycles*, 20(4).
- Lagman, K. B., Fennel, K., Thompson, K. R., & Bianucci, L. (2014). Assessing the utility of frequency dependent nudging for reducing biases in biogeochemical models. *Ocean Modelling*, 81, 25-35.
- Lenartz, F., Raick, C., Soetaert, K., & Grégoire, M. (2007). Application of an Ensemble Kalman filter to a 1-D coupled hydrodynamic-ecosystem model of the Ligurian Sea. *Journal of Marine Systems*, 68(3), 327-348.
- Lermusiaux, P. F., & Robinson, A. R. (1999). Data assimilation via error subspace statistical estimation. Part I: Theory and schemes. *Monthly Weather Review*, 127(7), 1385-1407.

- Lévy, M., Krémeur, A.S., & Mémery, L. (2005). Description of the LOBSTER biogeochemical model implemented in the OPA system. Internal report IPSL/LOCEAN.
- Lorenc, A. C. (1986). Analysis methods for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 112(474), 1177-1194.
- Lorenc, A. C., Bell, R. S., & Macpherson, B. (1991). The Meteorological Office analysis correction data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, 117(497), 59-89.
- Losa, S. N., Kivman, G. A., Schröter, J., & Wenzel, M. (2003). Sequential weak constraint parameter estimation in an ecosystem model. *Journal of Marine Systems*, 43(1), 31-49.
- Losa, S. N., Kivman, G. A., & Ryabchenko, V. A. (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), 1-20.
- Madec, G. (2008). NEMO ocean engine. Note du Pole de modélisation, Institut Pierre-Simon Laplace (IPSL), France, No 27 ISSN No. 1288–1619.
- Magri, S., Brasseur, P., & Lacroix, G. (2005). Data assimilation in a marine ecosystem model of the Ligurian Sea. *Comptes Rendus Geoscience*, 337(12), 1065-1074.
- Martin, M. J., Hines, A., & Bell, M. J. (2007). Data assimilation in the FOAM operational short-range ocean forecasting system: A description of the scheme and its impact. *Quarterly Journal of the Royal Meteorological Society*, 133(625), 981-995.
- Matear, R. J. (1995). Parameter optimization and analysis of ecosystem models using simulated annealing: a case study at Station P. *Journal of Marine Research*, 53(4), 571-607.
- Matear, R. J., & Holloway, G. (1995). Modeling the inorganic phosphorus cycle of the North Pacific using an adjoint data assimilation model to assess the role of dissolved organic phosphorus. *Global Biogeochemical Cycles*, 9(1), 101-119.
- Matear, R. J., & Jones, E. (2011). Marine biogeochemical modelling and data assimilation. In *Operational Oceanography in the 21st Century* (pp. 295-317). Springer Netherlands.
- Mattern, J. P., Dowd, M., & Fennel, K. (2010). Sequential data assimilation applied to a physical–biological model for the Bermuda Atlantic time series station. *Journal of Marine Systems*, 79(1), 144-156.
- Mattern, J. P., Fennel, K., & Dowd, M. (2012). Estimating time-dependent parameters for a biological ocean model using an emulator approach. *Journal of Marine Systems*, 96, 32-47.
- 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.
- Mattern, J. P., Fennel, K., & Dowd, M. (2014). Periodic time-dependent parameters improving forecasting abilities of biological ocean models. *Geophysical Research Letters*, 41(19), 6848-6854.
- McClain, C. R. (2009). A decade of satellite ocean color observations. *Annual Review of Marine Science*, 1, 19-42.
- Metref, S., Cosme, E., Snyder, C., & Brasseur, P. (2013). A non-Gaussian analysis scheme using rank histograms for ensemble data assimilation. *Nonlinear Processes in Geophysics*, 21, 869-885.
- Mikaloff Fletcher, S. E., Gruber, N., Jacobson, A. R., Doney, S. C., Dutkiewicz, S., Gerber, M., Follows, M., Joos, F., Lindsay, K., Menemenlis, D., Mouchet, A., Müller, S. A., & Sarmiento, J. L. (2006). Inverse estimates of anthropogenic CO₂ uptake, transport, and storage by the ocean. *Global Biogeochemical Cycles*, 20(2).
- Moisan, J. R., Hofmann, E. E., & Haidvogel, D. B. (1996). Modeling nutrient and plankton processes in the California coastal transition zone: 2. A three-dimensional

- physical-bio-optical model. *Journal of Geophysical Research: Oceans* (1978–2012), 101(C10), 22677-22691.
- Morel, A., & Prieur, L. (1977). Analysis of variations in ocean color. *Limnol. Oceanogr.*, 22(4), 709-722.
- Najjar, R. G., Sarmiento, J. L., & Toggweiler, J. R. (1992). Downward transport and fate of organic matter in the ocean: Simulations with a general circulation model. *Global Biogeochemical Cycles*, 6(1), 45-76.
- Natvik, L. J., & Evensen, G. (2003a). 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.
- Natvik, L. J., & Evensen, G. (2003b). Assimilation of ocean colour data into a biochemical model of the North Atlantic: Part 2. Statistical analysis. *Journal of Marine Systems*, 40, 155-169.
- Nerger, L., Hiller, W., & Schröter, J. (2005). A comparison of error subspace Kalman filters. *Tellus A*, 57(5), 715-735.
- Nerger, L., & Gregg, W. W. (2007). Assimilation of SeaWiFS data into a global ocean-biogeochemical model using a local SEIK filter. *Journal of Marine Systems*, 68(1), 237-254.
- 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), 87-102.
- O'Dea, E. J., Arnold, A. K., Edwards, K. P., Furner, R., Hyder, P., Martin, M. J., Siddorn, J. R., Storkey, D., While, J., Holt, J. T., & Liu, H. (2012). An operational ocean forecast system incorporating NEMO and SST data assimilation for the tidally driven European North-West shelf. *J. Oper. Oceanogr.*, 5, 3–17.
- Oke, P. R., Allen, J. S., Miller, R. N., Egbert, G. D., & Kosro, P. M. (2002). Assimilation of surface velocity data into a primitive equation coastal ocean model. *Journal of Geophysical Research: Oceans* (1978–2012), 107(C9), 5-1.
- Oort, A. H. (1983). Global atmospheric circulation statistics, 1958-1973 (No. 14). US Department of Commerce, National Oceanic and Atmospheric Administration.
- Oschlies, A., & Garçon, V. (1998). Eddy-induced enhancement of primary production in a model of the North Atlantic Ocean. *Nature*, 394(6690), 266-269.
- Oschlies, A., & Schartau, M. (2005). Basin-scale performance of a locally optimized marine ecosystem model. *Journal of Marine Research*, 63(2), 335-358.
- Ourmières, Y., Brasseur, P., Lévy, M., Brankart, J. M., & Verron, J. (2009). On the key role of nutrient data to constrain a coupled physical–biogeochemical assimilative model of the North Atlantic Ocean. *Journal of Marine Systems*, 75(1), 100-115.
- Palmer, J. R., & Totterdell, I. J. (2001). Production and export in a global ocean ecosystem model. *Deep Sea Research Part I: Oceanographic Research Papers*, 48(5), 1169-1198.
- Pastres, R., Ciavatta, S., & Solidoro, C. (2003). The Extended Kalman Filter (EKF) as a tool for the assimilation of high frequency water quality data. *Ecological modelling*, 170(2), 227-235.
- Pelc, J. S., Simon, E., Bertino, L., El Serafy, G., & Heemink, A. W. (2012). Application of model reduced 4D-Var to a 1D ecosystem model. *Ocean Modelling*, 57, 43-58.
- Pham, D. T., Verron, J., & Christine Roubaud, M. (1998a). A singular evolutive extended Kalman filter for data assimilation in oceanography. *Journal of Marine systems*, 16(3), 323-340.
- Pham, D. T., Verron, J., & Gourdeau, L. (1998b). Filtres de Kaiman singuliers évolutifs pour l'assimilation de données en océanographie. *Comptes Rendus de l'Académie des Sciences-Series IIA-Earth and Planetary Science*, 326(4), 255-260.
- Popova, E. E., Lozano, C. J., Srokosz, M. A., Fasham, M. J. R., Haley, P. J., & Robinson, A. R. (2002a). Coupled 3D physical and biological modelling of the mesoscale variability observed in North-East Atlantic in spring 1997: biological

- processes. *Deep Sea Research Part I: Oceanographic Research Papers*, 49(10), 1741-1768.
- Popova, E. E., Srokosz, M. A., & Smeed, D. A. (2002b). Real-time forecasting of biological and physical dynamics at the Iceland-Faeroes Front in June 2001. *Geophysical Research Letters*, 29(4), 14-1.
- Prieß, M., Piwonski, J., Koziel, S., Oschlies, A., & Slawig, T. (2013). Accelerated parameter identification in a 3D marine biogeochemical model using surrogate-based optimization. *Ocean Modelling*, 68, 22-36.
- Prunet, P., Minster, J. F., Ruiz-Pino, D., & Dadou, I. (1996a). Assimilation of surface data in a one-dimensional physical-biogeochemical model of the surface ocean: 1. Method and preliminary results. *Global Biogeochemical Cycles*, 10(1), 111-138.
- Prunet, P., Minster, J. F., Echevin, V., & Dadou, I. (1996b). Assimilation of surface data in a one-dimensional physical-biogeochemical model of the surface ocean: 2. Adjusting a simple trophic model to chlorophyll, temperature, nitrate, and pCO₂ data. *Global Biogeochemical Cycles*, 10(1), 139-158.
- Raick, C., Alvera-Azcarate, A., Barth, A., Brankart, J. M., Soetaert, K., & Grégoire, M. (2007). Application of a SEEK filter to a 1D biogeochemical model of the Ligurian Sea: twin experiments and real in-situ data assimilation. *Journal of Marine Systems*, 65(1), 561-583.
- Reynolds, R. W. (1988). A real-time global sea surface temperature analysis. *Journal of climate*, 1(1), 75-87.
- Ristic, B., Arulampalam, S., & Gordon, N. (2004). *Beyond the Kalman filter: Particle filters for tracking applications* (Vol. 685). Boston: Artech house.
- Robinson, A. R., & Lermusiaux, P. F. (2002). Data assimilation for modeling and predicting coupled physical-biological interactions in the sea. *The Sea*, 12, 475-536.
- Rose, K. A., Megrey, B. A., Werner, F. E., & Ware, D. M. (2007). Calibration of the NEMURO nutrient-phytoplankton-zooplankton food web model to a coastal ecosystem: Evaluation of an automated calibration approach. *Ecological modelling*, 202(1), 38-51.
- Rousseaux, C. S., & Gregg, W. W. (2012). Climate variability and phytoplankton composition in the Pacific Ocean. *Journal of Geophysical Research: Oceans* (1978–2012), 117(C10).
- Rubin, D. B. (1988) Using the SIR algorithm to simulate posterior distributions. In: *Bayesian Statistics 3*, Oxford University Press, 395–402.
- Samuelson, A., Bertino, L., & Hansen, C. (2009). Impact of data assimilation of physical variables on the spring bloom from TOPAZ operational runs in the North Atlantic. *Ocean Science*, 5(4), 635-647.
- Sasaki, Y. (1958). An Objective Analysis Based on the Variational Method. *Journal of the Meteorological Society of Japan*, 36, 77-88.
- Schartau, M., Oschlies, A., & Willebrand, J. (2001). Parameter estimates of a zero-dimensional ecosystem model applying the adjoint method. *Deep Sea Research Part II: Topical Studies in Oceanography*, 48(8), 1769-1800.
- Schartau, M., & Oschlies, A. (2003). Simultaneous data-based optimization of a 1D-ecosystem model at three locations in the North Atlantic: Part I—Method and parameter estimates. *Journal of Marine Research*, 61(6), 765-793.
- Schlitzer, R. (2000). Applying the adjoint method for biogeochemical modeling: Export of particulate organic matter in the world ocean. *Inverse methods in global biogeochemical cycles*, AGU Geophysical Monograph Series, 114, 107-124.
- Schlitzer, R. (2002). Carbon export fluxes in the Southern Ocean: results from inverse modeling and comparison with satellite-based estimates. *Deep Sea Research Part II: Topical Studies in Oceanography*, 49(9), 1623-1644.
- Schlitzer, R., Usbeck, R., & Fischer, G. (2004). Inverse modeling of particulate organic carbon fluxes in the South Atlantic. In *The South Atlantic in the Late Quaternary* (pp. 1-19). Springer Berlin Heidelberg.

- Semovski, S., & Woźniak, B. (1995). Model of the annual phytoplankton cycle in the marine ecosystem-assimilation of monthly satellite chlorophyll data for the North Atlantic and Baltic. *Oceanologia*, 37(1), 3-31.
- Sheng, J., Greatbatch, R. J., & Wright, D. G. (2001). Improving the utility of ocean circulation models through adjustment of the momentum balance. *Journal of Geophysical Research: Oceans* (1978–2012), 106(C8), 16711-16728.
- 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.
- 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., & Bertino, L. (2012). Gaussian anamorphosis extension of the DEnKF for combined state parameter estimation: application to a 1D ocean ecosystem model. *Journal of Marine Systems*, 89(1), 1-18.
- Smith, K. W., & McGillicuddy Jr, D. J. (2011). Dynamical interpolation of surface ocean chlorophyll fields via data assimilation with an iterative ensemble smoother. *Journal of Marine Systems*, 85(3), 96-105.
- Song, H., Edwards, C. A., Moore, A. M., & Fiechter, J. (2012). Incremental four-dimensional variational data assimilation of positive-definite oceanic variables using a logarithm transformation. *Ocean Modelling*, 54, 1-17.
- Spitz, Y. H., Moisan, J. R., Abbott, M. R., & Richman, J. G. (1998). Data assimilation and a pelagic ecosystem model: parameterization using time series observations. *Journal of Marine Systems*, 16(1), 51-68.
- Spitz, Y. H., Moisan, J. R., & Abbott, M. R. (2001). Configuring an ecosystem model using data from the Bermuda Atlantic Time Series (BATS). *Deep Sea Research Part II: Topical Studies in Oceanography*, 48(8), 1733-1768.
- Talagrand, O., & Courtier, P. (1987). Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory. *Quarterly Journal of the Royal Meteorological Society*, 113(478), 1311-1328.
- Talagrand, O. (1997). Assimilation of observations, an introduction. *Journal of the Meteorological Society of Japan*, 75, 191-209.
- 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(1), 200-217.
- Thomasell, A., & Welsh, J. G. (1963). Studies of techniques for the analysis and prediction of temperature in the Ocean. Part I: The objective analysis of sea-surface temperature, Interim Report 7046-70, The Travelers Research Center, Inc.
- Tjiputra, J. F., Polzin, D., & Winguth, A. M. (2007). Assimilation of seasonal chlorophyll and nutrient data into an adjoint three-dimensional ocean carbon cycle model: Sensitivity analysis and ecosystem parameter optimization. *Global Biogeochemical Cycles*, 21(1).
- Torres, R., Allen, J. I., & Figueiras, F. G. (2006). Sequential data assimilation in an upwelling influenced estuary. *Journal of Marine Systems*, 60(3), 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., Korres, G., Hoteit, I., Petihakis, G., & Banks, A. C. (2007). Assimilation of ocean colour data into a Biogeochemical Flux Model of the Eastern Mediterranean Sea. *Ocean Science*, 3(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.
- Vallino, J. J. (2000). Improving marine ecosystem models: use of data assimilation and mesocosm experiments. *Journal of Marine Research*, 58(1), 117-164.
- Valsala, V., & Maksyutov, S. (2010). Simulation and assimilation of global ocean pCO₂ and air-sea CO₂ fluxes using ship observations of surface ocean pCO₂ in a simplified biogeochemical offline model. *Tellus B*, 62(5), 821-840.
- van Leeuwen, P. J. (2009). Particle filtering in geophysical systems. *Monthly Weather Review*, 137(12), 4089-4114.
- Vichi, M., Pinardi, N., & Masina, S. (2007). A generalized model of pelagic biogeochemistry for the global ocean ecosystem. Part I: Theory. *Journal of Marine Systems*, 64(1), 89-109.
- Visinelli, L., Masina, S., Vichi, M., & Storto, A. (2014). Impacts of physical data assimilation on the Global Ocean Carbonate System. *Biogeosciences Discussions*, 11(4), 5399-5441.
- Wang, C. H., Li, X. Y., & Lv, X. Q. (2012). Adjoint Assimilation of SeaWiFS data into a Marine Ecosystem Dynamical Model of the Bohai Sea and the North Yellow Sea. *Procedia Environmental Sciences*, 13, 2045-2061.
- Ward, B. A., Friedrichs, M. A., Anderson, T. R., & Oschlies, A. (2010). Parameter optimisation techniques and the problem of underdetermination in marine biogeochemical models. *Journal of Marine Systems*, 81(1), 34-43.
- Ward, B. A., Schartau, M., Oschlies, A., Martin, A. P., Follows, M. J., & Anderson, T. R. (2013). When is a biogeochemical model too complex? Objective model reduction and selection for North Atlantic time-series sites. *Progress in Oceanography*, 116, 49-65.
- Waters, J., Lea, D. J., Martin, M. J., Mirouze, I., Weaver, A., & While, J. (2015). Implementing a variational data assimilation system in an operational 1/4 degree global ocean model. *Quarterly Journal of the Royal Meteorological Society*, 141: 333-349.
- Weber, L., Völker, C., Schartau, M., & Wolf-Gladrow, D. A. (2005). Modeling the speciation and biogeochemistry of iron at the Bermuda Atlantic Time-series Study site. *Global Biogeochemical Cycles*, 19(1).
- While, J., Haines, K., & Smith, G. (2010). A nutrient increment method for reducing bias in global biogeochemical models. *Journal of Geophysical Research: Oceans* (1978-2012), 115(C10).
- While, J., Totterdell, I., & Martin, M. (2012). Assimilation of pCO₂ data into a global coupled physical-biogeochemical ocean model. *Journal of Geophysical Research: Oceans* (1978-2012), 117(C3).
- Xiao, Y., & Friedrichs, M. A. (2014a). The assimilation of satellite-derived data into a one-dimensional lower trophic level marine ecosystem model. *Journal of Geophysical Research: Oceans*, 119(4), 2691-2712.
- Xiao, Y., & Friedrichs, M. A. M. (2014b). Using biogeochemical data assimilation to assess the relative skill of multiple ecosystem models in the Mid-Atlantic Bight: effects of increasing the complexity of the planktonic food web. *Biogeosciences*, 11(11), 3015-3030.
- Yao, X., & Schlitzer, R. (2013). Assimilating water column and satellite data for marine export production estimation. *Geoscientific Model Development Discussions*, 6(1), 2045-2085.
- Yool, A., Popova, E. E., & Anderson, T. R. (2013). MEDUSA-2.0: an intermediate complexity biogeochemical model of the marine carbon cycle for climate change and ocean acidification studies. *Geoscientific Model Development*, 6, 1767-1811.

- Zhang, P., Wai, O. W., Chen, X., Lu, J., & Tian, L. (2014). Improving Sediment Transport Prediction by Assimilating Satellite Images in a Tidal Bay Model of Hong Kong. *Water*, 6(3), 642-660.
- Zhao, L., Wei, H., Xu, Y., & Feng, S. (2005). An adjoint data assimilation approach for estimating parameters in a three-dimensional ecosystem model. *Ecological modelling*, 186(2), 235-250.
- Zhao, Q., & Lu, X. (2008). Parameter estimation in a three-dimensional marine ecosystem model using the adjoint technique. *Journal of Marine Systems*, 74(1), 443-452.

Met Office
FitzRoy Road, Exeter
Devon EX1 3PB
United Kingdom

Tel (UK): 0870 900 0100 (Int) : +44 1392 885680
Fax (UK): 0870 900 5050 (Int) :+44 1392 885681
enquiries@metoffice.gov.uk
www.metoffice.gov.uk