JOURNAL OF GEOPHYSICAL RESEARCH: OCEANS, VOL. 118, 2746–2760, doi:10.1002/jgrc.20213, 2013

# Particle filter-based data assimilation for a three-dimensional biological ocean model and satellite observations

Jann Paul Mattern,[1,2] Michael Dowd,[1] and Katja Fennel[2]

Received 15 November 2012; revised 11 March 2013; accepted 22 April 2013; published 30 May 2013.

[1] We assimilate satellite observations of surface chlorophyll into a three-dimensional biological ocean model in order to improve its state estimates using a particle filter referred to as sequential importance resampling (SIR). Particle Filters represent an alternative to other, more commonly used ensemble-based state estimation techniques like the ensemble Kalman filter (EnKF). Unlike the EnKF, Particle Filters do not require normality assumptions about the model error structure and are thus suitable for highly nonlinear applications. However, their application in oceanographic contexts is typically hampered by the high dimensionality of the model's state space. We apply SIR to a high-dimensional model with a small ensemble size (20) and modify the standard SIR procedure to avoid complications posed by the high dimensionality of the model state. Two extensions to the SIR include a simple smoother to deal with outliers in the observations, and state-augmentation which provides the SIR with parameter memory. Our goal is to test the feasibility of biological state estimation with SIR for realistic models. For this purpose we compare the SIR results to a model simulation with optimal parameters with respect to the same set of observations. By running replicates of our main experiments, we assess the robustness of our SIR implementation. We show that SIR is suitable for satellite data assimilation into biological models and that both extensions, the smoother and state-augmentation, are required for robust results and improved fit to the observations.

**Citation:** Mattern, J. P., M. Dowd, and K. Fennel (2013), Particle filter-based data assimilation for a 3-dimensional biological ocean model and satellite observations, *J. Geophys. Res. Oceans*, *118*, 2746–2760, doi:10.1002/jgrc.20213.

## 1. Introduction

[2] Numerical ocean models are growing in their capabilities and their significance for ocean research and prediction. At the same time, an ever larger number of observing platforms, ranging from underwater observatories to satellites, provide an unprecedented wealth of ocean information. Data assimilation procedures, which combine models and observations, represent the principal means by which a large number of observations can be used to improve model estimates and forecast abilities, and quantify uncertainty. Two major categories of data assimilation methods are: (1) variational methods, such as 4DVAR [*Bennett*, 2002] and (2) ensemble methods, such as the ensemble Kalman filter (EnKF) [*Evensen*, 2009]. Ensemble, or sample-based, approaches have become popular in recent years due to their relative ease of implementation, and their straightforward treatment of model error. They treat the data assimilation problem from a probabilistic perspective and use Bayesian principles to blend model predictions with observations [*Wikle and Berliner*, 2007].

[3] Particle filters represent a very general class of ensemble-based statistical data assimilation techniques that offer complete solutions to nonlinear and non-Gaussian data assimilation problems [*Ristic et al*., 2004]. For particle filters, there are no restrictive assumptions about the probability distributions, unlike the EnKF which assumes that the prediction and filtering distributions are normal, and that the likelihood function is linear and normal. When these assumptions are violated the EnKF solutions are suboptimal [*van Leeuwen*, 2010], which has been demonstrated for nonlinear systems [*Dowd*, 2007]. Hence, the generality of the particle filter makes it attractive for application to the complex, highly nonlinear models encountered in oceanography and meteorology. Yet, particle filters have been applied infrequently due to practical problems in the ensemble representation in high dimensional applications [*van Leeuwen*, 2009] (also see discussion in section 5). Here, we present a particle filter that focuses on parameter and chlorophyll state estimation for a 3-D biological ocean model, a complex system with a high-dimensional state space. In this application a number of modifications are introduced that improve the particle filter's robustness and allow for effective data assimilation.

[1]Department of Mathematics and Statistics, Dalhousie University, Halifax, Nova Scotia, Canada.
[2]Department of Oceanography, Dalhousie University, Halifax, Nova Scotia, Canada.

Corresponding author: J. P. Mattern, Department of Mathematics and Statistics, Dalhousie University, Halifax, NS B3H 4R2, Canada. (paul.mattern@dal.ca)


2169-9275/13/10.1002/jgrc.20213

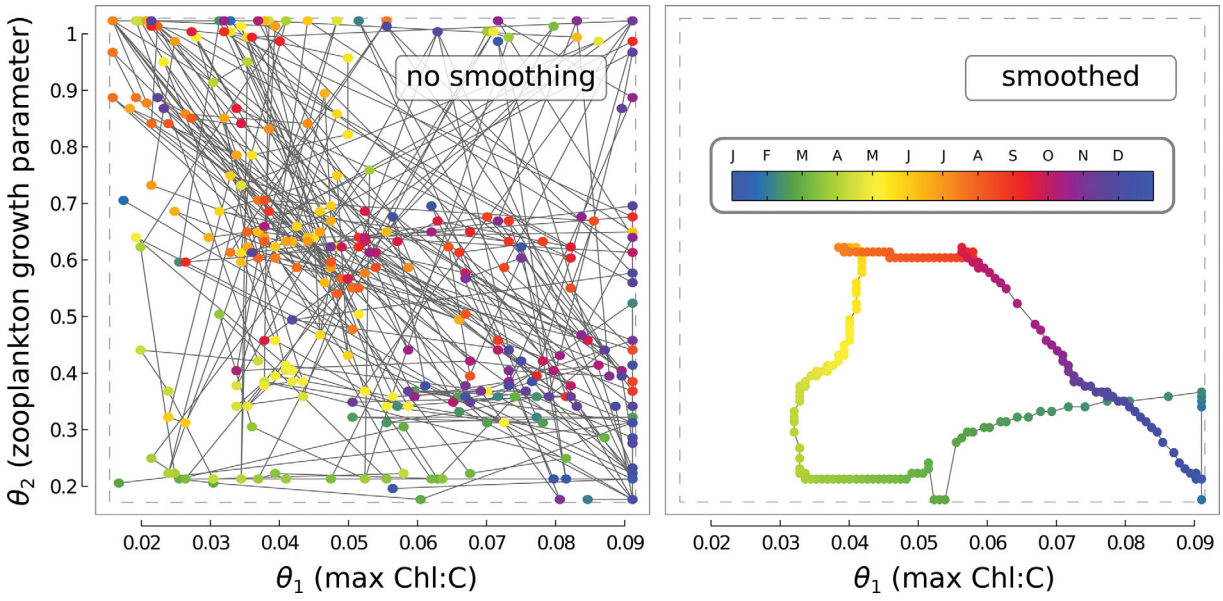


**Figure 1.** Time evolution of the optimal parameter values for individual images within our satellite dataset. Each dot represents the pair of optimal parameter values for one satellite image. (a) Without any smoothing the parameter development is dominated by high frequency variation; (b) by smoothing the underlying distance function in time, the low frequency signal is more apparent. Plots are taken from *Mattern et al*. [2012].

[4] As a basis, we use sequential importance resampling (SIR) [*Gordon et al*., 1993; *Kitagawa*, 1996], the most standard particle filter algorithm. Like all particle filters, SIR is a sequential or recursive technique which assimilates observations in sequence, a setup that is suitable for online estimation, i.e., forward operation and prediction. For the purpose of state estimation, particle filters use an ensemble of model simulations, run in parallel, that allow for the approximation and propagation of model uncertainty, and blending with measurement information. While applications of other ensemble techniques, especially the EnKF, are numerous in oceanography [*Ciavatta et al*., 2011; *Hu et al*., 2012], most SIR particle filters have only been applied in the context of relatively simple models [*Dowd*, 2007; *Mattern et al*., 2010a] or using synthetic (typically model-generated) observations [*Annan and Hargreaves*, 2010] which do not contain outliers or observation errors. Here, we apply SIR to a complex 3-D biological ocean model of the Middle Atlantic Bight (MAB) in the North Atlantic [*Hu et al*., 2012; *Mattern et al*., 2012], and assimilate daily satellite observations of chlorophyll derived from ocean color.

[5] In our approach, we evade problems posed by the high dimensionality of the model by allowing errors in a few select biological parameters only. By ignoring model errors arising from other sources, such as physical model inputs, the procedure is effectively operating in a much lower dimensional error subspace. While only two biological parameters are selected to contain errors in this application, the error subspace approach can be easily extended to more parameters and other stochastic inputs.

[6] Our approach for particle filter-based data assimilation is motivated by previous research. In *Mattern et al*. [2012], we used the same ocean model and the same set of observations to optimize biological parameters using polynomial chaos, a statistical emulator technique. We obtained chlorophyll state estimates and biological parameters estimates from the emulator. We also emphasized the use of time dependent (seasonally) evolving parameter values, which explained the observations considerably better that the optimal fixed parameters (see, e.g., Figure 1). The time-dependence of the parameters is in part attributed to seasonal changes in the plankton populations (species composition), which cannot be adequately represented by just a few bulk plankton variables and fixed parameters [*Mattern et al*., 2012]. Here we use the same model setup as a suitable test bed for combined state and parameter estimation with a particle filter.

[7] The goal in this study is to explore the potential of this particle filter configuration and identify effective approaches that allow us to obtain model chlorophyll and biological parameter estimates through data assimilation.

## 2. Methods

### 2.1. Particle Filtering Overview

[8] Before presenting our particle filter implementation, we provide a short overview of the basic particle filtering procedure. More detailed descriptions of particle filters and some of their extensions can be found in *Ristic et al*. [2004], *Dowd* [2007], and *van Leeuwen* [2009].

[9] For particle filters and ensemble-based data assimilation techniques in general, models are considered to be stochastic, i.e., the model state is represented by a multivariate probability distribution. This probability distribution is approximated with an ensemble of particles. Each particle, or ensemble member, contains a particular model state that represents a sample from its probability distribution.

Starting with an initial ensemble of particles (which we generate by running the model multiple times with different parameter values), sequential data assimilation techniques perform a sequence of assimilation steps to propagate the ensemble forward in time and update it with information from observations as they become available. Each assimilation step consists of two distinct substeps: the forecast step, and the observation update step.

[10] The forecast step simulates the time evolution of the ensemble. Each particle is moved forward in time with the numerical model. For this purpose, the model state associated with each particle is used as an initial condition for the model and a model simulation is started for each particle. The model simulations are run until the next point in time with available observations and a new ensemble is formed from the most recent model states. In the subsequent observation update step, the new ensemble is updated with information from the newly available observations. This update typically accounts for observation errors as well as model uncertainty which is derived from the ensemble. While the forecast step is conceptually the same in all sequential techniques, including the EnKF, the filters differ in their implementation of the observation update step.

[11] The SIR observation update step consists of a weighting of each ensemble member dependent on the current observations, and a subsequent weight-based resampling. The weights assigned to each particle represent their distance from the current observations. In particle filter theory, the weight of the $i^{\text{th}}$ ensemble member $x_t^{(i)}$ given the current observation $y_t$ is equal to the likelihood $p(y_t|x_t^{(i)})$ (the probability of the observations given the ensemble member). In practice, this likelihood is often unknown (and in case of high-dimensional model states, typically infeasible to compute) and thus the weights need to be approximated.

[12] After the weights have been determined, a weight-based resampling of the ensemble is performed. An ensemble member with high fidelity to the observations, and therefore a large weight, has a higher probability of being drawn during the resampling than a lower-weighted particle. The sampling is performed with replacement so that particles can be drawn more than once. This means that at the end of the resampling the new ensemble will typically include multiple replicates of high-weighted particles, while some low-weighted particles do not get resampled (in other words, the resampling is a weighted bootstrap). Through this procedure, the current observations have now been assimilated into the updated ensemble, and the ensemble-approximated model distribution has been made more consistent with the observations. The model-generated states that enter the observation update are left intact and remain true to the model dynamics, as the resampling introduces no potentially undesirable shifts in the model state of ensemble members.

### 2.2. Model Overview

[13] In this study we apply our particle filter to a 3-dimensional physical-biological ocean model. This model has previously been employed in the data assimilation studies of *Mattern et al*. [2012] and *Hu et al*. [2012] and a thorough model description is given in the latter. Here, we provide a short summary. The domain of the model is the MAB, a region in the northwestern North Atlantic that stretches from Cape Cod in the north to Cape Hatteras in the south (Figure 2). The model is based on the regional ocean modeling system [*Haidvogel et al*., 2008] and consists of a 3-dimensional physical model coupled with a biological component. The model domain is divided into $130 \times 82$ horizontal grid cells with a resolution varying from $\approx 5.5$ km in the estuaries to $\approx 8.0$ km in the open ocean. In the vertical, the model water column is divided into 36 layers with a higher resolution near the surface. The physical model includes runoff of major rivers, tides, and uses open boundary conditions for all state variables, as described in detail in *Hu et al*. [2012]. The biological component, which we focus on in this study, is described in *Fennel et al*. [2006] and has been used successfully in various modelling studies [*Fennel et al*., 2008; *Fennel and Wilkin*, 2009; *Previdi et al*., 2009; *Fennel*, 2010; *Bianucci et al*., 2011]. It contains 7 state variables (phytoplankton, zooplankton, chlorophyll, nitrate, ammonium and small and large detrital nitrogen) that simulate a simplified nitrogen cycle. Chlorophyll is modelled separately from phytoplankton to account for the effects of photoacclimation, which allows phytoplankton species to regulate their chlorophyll content based on the availability of light and nutrients [*Geider et al*., 1998]. A separate chlorophyll variable can be compared directly to observations, thereby facilitating assimilation. There is no feedback from the biological component to the physical model, so that the physical state variables remain unaffected by changes in the biological state variables.

[14] The initial and boundary conditions for the biological variables are taken from an implementation of the same model in a larger scale domain of the northeast North American shelf and adjacent deep ocean [*Fennel et al*., 2008]. For the physical model, open boundary conditions for temperature, salinity, subtidal frequency velocity and sea level are taken from the larger scale MAB and Gulf of Maine regional model described in *Chen and He* [2010].

### 2.3. Chlorophyll Observations

[15] The assimilated observations are daily images of surface chlorophyll concentrations derived from the SeaWiFS satellite for 2006. Each of the available 350 images represents one or an average of multiple satellite passes. The observations are obtained at a resolution of 1 km; to facilitate model-data comparison, they are interpolated onto the model grid. A more detailed description of the observations is given in *Hu et al*. [2012].

[16] By assimilating daily observations, instead of weekly or monthly composites, we aim to retain short-term features in the images and allow for better prediction of the timing of events such as phytoplankton blooms. Assimilation of daily images also tests the particle filter's ability to deal with outliers and missing values: Large portions of the images can be missing due to clouds and other effects, so that some images contain only localized information about parts of the domain. In addition, significant noise is present in the satellite dataset, especially close to the coast. As noted in *Mattern et al*. [2012], high noise levels and the abundance of colored dissolved organic material impede

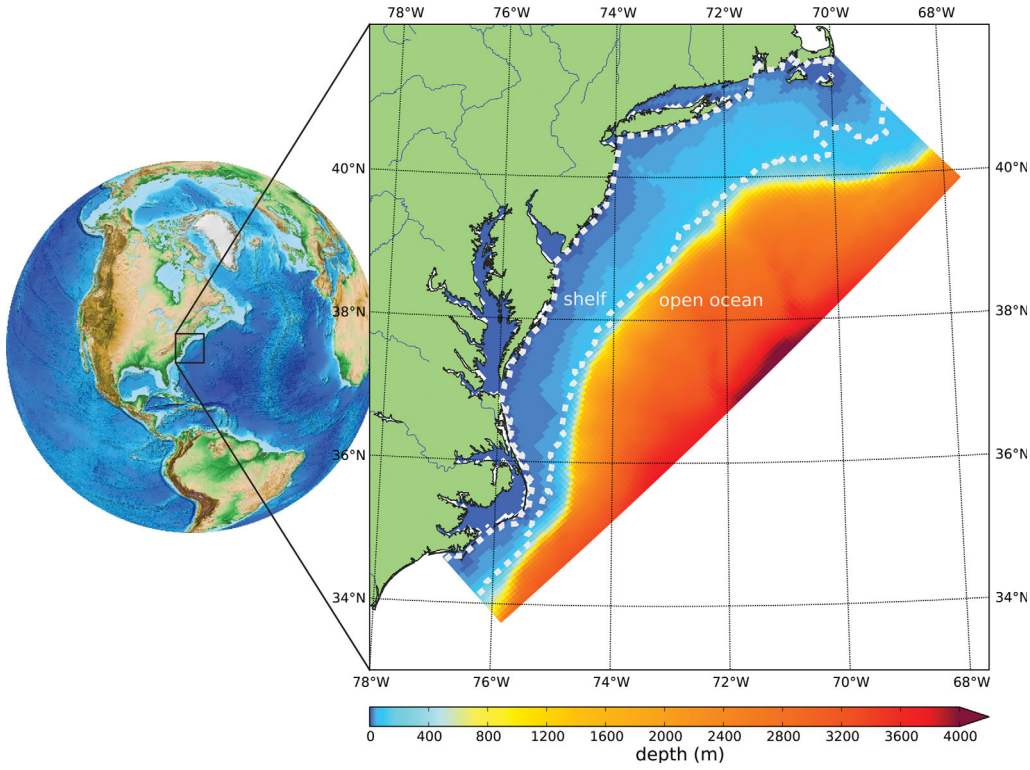


**Figure 2.** The model domain and its bathymetry. The dashed white lines mark the boundaries of the shelf and open ocean regions used in our analysis.

data assimilation and parameter estimation in the estuaries. For this reason, we exclude the estuaries from the assimilation experiments presented here.

### 2.4. Biological Parameters of Interest

[17] The ensemble generation is an important aspect of our particle filter implementation and relies on treating two selected parameters of the biological model as stochastic. In this section we will briefly describe the role of the two parameters in the model and motivate why we selected these parameters in particular.

[18] The first parameter of interest, $\theta_1$, is the phytoplankton maximum chlorophyll-to-carbon ratio, it regulates the mechanism by which phytoplankton can adapt their chlorophyll content. The second parameter $\theta_2$ is the maximum grazing rate of zooplankton, a multiplicative factor controlling the rate at which zooplankton consume phytoplankton. While $\theta_1$ affects the chlorophyll variable in the model directly, $\theta_2$ has a more indirect effect on chlorophyll, through altering the grazing pressure on phytoplankton and hence their abundance. Our motivation for selecting these two parameters is twofold: Firstly, the model exhibits a strong sensitivity to relative changes in both parameters, especially $\theta_1$. Consequently, even small parameter adjustments can have a large effect on the chlorophyll output of the model. Secondly, both parameters are physiological plankton parameters known to vary greatly among species and can thus be considered stochastic with a high uncertainty. Varying these parameters presents a suitable way to produce variable chlorophyll output in the model. While it would be easy to incorporate more parameters into the particle filter procedure, we limit ourselves to these two parameters because we have examined their effects and identified their optimal values in a previous study [*Mattern et al.*, 2012]. This allows us to compare the particle filter-based parameter estimates to our previous results.

## 3. Particle Filter Implementation

[19] In this section, we describe the aspects of the particle filter implementation that are important to this application and deserve special attention. While the SIR resampling is straightforward to implement, the computation of the weights can be implemented in various ways and needs careful consideration.

[20] In this application, each particle in the ensemble is a realization of the biological state of the system consisting of the concentrations of the seven biological state variables in each of the $82 \times 130(\text{horizontal}) \times 36(\text{vertical}) = 383760$ grid cells of the model. As the state of the physical ocean model is unaffected by the biological parameters, it remains deterministic and does not need to be considered by the SIR procedure, or even included in the state vector. With only the biological variables included, a state vector is $7 \times 383760 = 2686320$ dimensional. For some of our SIR experiments, we additionally append two model parameters to the state, and this is described in more detail in section 3.2 below.

### 3.1. SIR Weighting

[21] In the resampling step of the SIR procedure, a weight is assigned to each ensemble member proportional to its likelihood (see section 2.1). In practice, the likelihood

is typically unknown and can only be approximated. In this implementation, we base the computation of the associated weights on the assumption that the likelihood is inversely related to a suitable distance measure for model state and observations.

[22] Here, each observation is a satellite chlorophyll image obtained at a higher resolution than the model grid. To facilitate the comparison with model chlorophyll, we first interpolate each satellite chlorophyll observation onto the model grid and then transform the interpolated observation into a real-valued image. For our model grid, the transformation into an image is straightforward: Because the horizontal model grid is approximately a rotated regular grid, each surface grid cell is simply mapped onto a single image pixel. The observations contain missing values, which are transformed into pixels containing missing values in the image (on average, the proportion of missing values is 75%). Using the same approach, we also transform the corresponding model surface chlorophyll field into an image. The model-data comparison is now reduced to the problem of comparing two real-valued images, both of which typically contains missing values (land grid cells within the model domain are turned into missing values, see *Mattern et al*. [2010b]).

[23] For image comparison, we use the adapted grey block (AGB) distance measure introduced by *Mattern et al*. [2010b], which was also used in the emulator study by *Mattern et al*. [2012]. To compute an AGB distance value, two images are compared at different resolution levels by dividing them into successively smaller blocks and computing the mean value of the pixels within each block. The average absolute difference is then computed at each resolution level, from the highest resolution, where each block consists of only 1 pixel, to the lowest, where a single block encompasses the entire image. Finally, the average absolute difference values for every resolution level are summed to obtain the AGB distance value. For the purpose of comparing satellite images, this multiresolution approach has proven to be advantageous over standard approaches such as the regular root mean square error because it is less sensitive to noise and adapted to work with missing values in the images [*Mattern et al*., 2010b]. The AGB is also suitable for our purposes, because it factors in features at various different spatial scales that may be present in the model output or the satellite observations. However, our SIR weighting implementation is flexible and any suitable model-data distance measure could be used.

[24] To transform the AGB distance values into weights, we use the following procedure:

[25] 1. Compute $d_i = \mathrm{d}_{\mathrm{AGB}}(x_t^{(i)}, y_t)$, the AGB distance between the current satellite observation and ensemble member $i$ for each $i = 1, \ldots, n_{\mathrm{ens}}$.

[26] 2. Calculate raw weights as the inverse of the AGB values: $\hat{w}^{(i)} = \frac{1}{d_i}$.

[27] 3. Normalize the raw weights and transform them according to: $\tilde{w}^{(i)} = \left(\frac{\hat{w}^{(i)}}{\max_j \hat{w}^{(j)}}\right)^{a_{\mathrm{weight}}}$, where $a_{\mathrm{weight}}$ is an adjustable parameter that spreads (or contracts) the distribution of weights (discussed in more detail below).

[28] 4. Determine final weights as $w_t^{(i)} = \frac{\tilde{w}_t^{(i)}}{\sum_{j=1}^{n_{\mathrm{ens}}} \tilde{w}_t^{(j)}}$, so that their sum is one.

[29] The parameter $a_{\mathrm{weight}}$ in step (3) adjusts the weights before they enter the resampling process. It does not change the relative ranking of the weights but affects their variance. The parameter $a_{\mathrm{weight}}$ brings the weights closer together (for $a_{\mathrm{weight}} < 1$) or spreads them apart (for $a_{\mathrm{weight}} > 1$). In practice, such an adjustment is needed if the values of the weights are too similar or differ by orders of magnitude. The former is the case in our application: typical values of $d_i$ (step (1)) are strongly dependent on the satellite observation but do not vary much across the ensemble of model simulations. Consequently, the normalized raw weights obtained in step (2) are tightly clustered (more than 70% of the raw weights are between 0.9 and 1.0 during a typical assimilation run). If no adjustment is made to the weights before resampling takes place (beside the normalization in step (4)), the ensemble members are resampled at almost equal probability. As a result, the effect of the resampling on the ensemble is very small.

[30] Figure 3 shows the effect of varying $a_{\mathrm{weight}}$. For $a_{\mathrm{weight}} = 1$, which corresponds to no weight adjustment, the SIR resampling will have almost no effect even in cases where parameter and weights are strongly correlated, i.e an increase in $\theta_1$ is accompanied by an increase in weights (Figure 3a). For higher values of $a_{\mathrm{weight}}$, ensemble members with higher weights become more likely to be picked during resampling (Figure 3b) up to a point where one ensemble member dominates the ensemble (Figure 3c). In the latter case the ensemble is likely to collapse to just a few unique particles. The adjustment of $a_{\mathrm{weight}}$ has a similar effect as changing the observation error distribution, which is done, for example, to decrease the effects of outliers [*van Leeuwen*, 2003]. In this application, both the true likelihood and the value of $a_{\mathrm{weight}}$ that would best approximate the likelihood are unknown. Furthermore, we are limited to a small number of particles, so that it is essential to select $a_{\mathrm{weight}}$ to extract a maximum amount of information from the observations while avoiding ensemble collapse. We set $a_{\mathrm{weight}} = 16$ for this application. This was based on a series of experiments in which we simulated thousands of sequences of 5 to 10 consecutive assimilation steps, using our satellite observations. In each experiment, we measured the average raw weight (the weight prior to normalization, see step (2)) of the ensemble after having performed several resampling steps, and estimated the probability of ensemble collapse. The parameter $a_{\mathrm{weight}}$ was chosen to maximize the average raw weight, while keeping the probability of ensemble collapse low.

[31] The effects of this weighting implementation on the ensemble structure and its development from one assimilation step to the next are visualized in Figure 4 for a typical SIR simulation with 10 ensemble members. It gives a detailed view of the time history of the particles and the effect of the weights on the resampling of the ensemble. By tracing each particle back to its "parent" particle (the particle it was resampled from), we determined that 7.5 is the average number of assimilation steps it takes to trace the ensemble at a given assimilation step back to its last

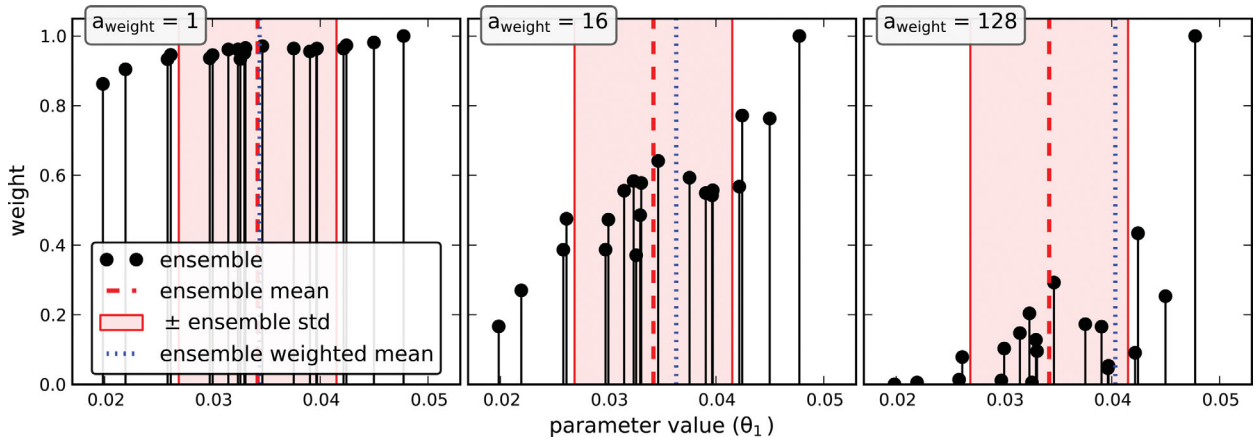


**Figure 3.** Black dots show the weight of ensemble members during resampling over the corresponding parameter value $\theta_1$ for a weighting parameter $a_{\text{weight}}$ of (a) 1, (b) 16, and (c) 128. Weights are obtained according to the procedure described in section 3.1. Also shown are the ensemble mean and standard deviation of $\theta_1$ (dashed and solid red lines) and the weighted ensemble mean (dotted blue line). The distance between weighted and unweighted ensemble mean acts as an indicator for the ensemble shift that can be expected from the SIR resampling. For $a_{\text{weight}} = 1$ the ensemble is not likely to shift significantly after resampling, due to the small differences in weights. By increasing $a_{\text{weight}}$, differences in weight become more pronounced, increasing the expected ensemble shift. For $a_{\text{weight}} \geq 128$, one or just a few ensemble members carry high weights, increasing the chance of ensemble collapse.

common "ancestor" particle (see green line in Figure 4). This number indicates that the expected number of ensemble collapse events is relatively low and that particles with high likelihoods do not get too much weight. At the same time it is apparent from Figure 4 that the weighting is effective at steering the ensemble.

### 3.2. State-Augmentation

[32] State-Augmentation is an extension that permits parameter estimation within the particle filter framework and was introduced by *Kitagawa* [1998]. In this extension, the model state $x_t$ is augmented by a vector of model parameters, $\theta_t \in \mathbb{R}^{n_\theta}$ to form

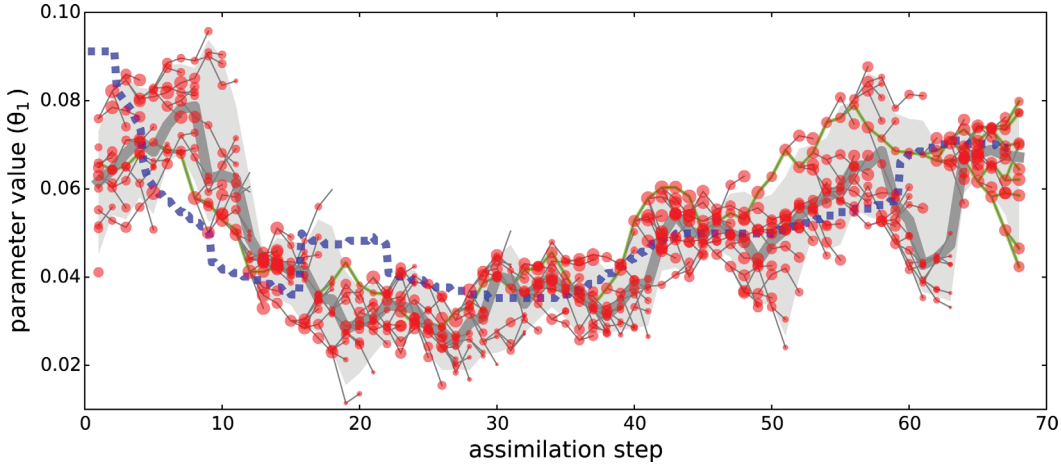


**Figure 4.** A detailed view of the development of $\theta_1$ during a typical particle filter experiment with an ensemble size of $n_{\text{ens}} = 10$ over 68 assimilation steps. Each red dot marks one particle, the area of the dot is proportional to its weight during the resampling (overlapping dots cause an increase in intensity). The thin lines connect each particle to its "parent" particle from which it was resampled and its "children" which are resampled from it. The lines highlighted in green trace the last ensemble at assimilation step 68 back through their ancestors. The (vertical) parameter value offsets between a parent particle and its children are due to the parameter noise that is applied during resampling. The ensemble median and the region between the 0.1 and 0.9 ensemble quantiles are displayed in the background (dark gray line and light gray area, respectively), for the plots in Figure 6 we use this, less detailed view of the ensemble. For reference, the blue dashed line shows the smoothed optimal parameter values (see Figure 6). The data for this plot is taken from one of our particle filter experiments in section 4.5, it uses our standard particle filter configuration.

**Table 1.** Optimal Parameter Values and Other Properties for the Particle Filter Experiments

| Name | Description | Optimal Value | Initial Distribution | Truncated Distribution[a] | Parameter Perturbation sd[b] |
|---|---|---|---|---|---|
| $\theta_1$ | Maximum Chl:C ratio in phytoplankton | 0.047 | U(0.013, 0.094) | U(0.040, 0.094) | 0.004 |
| $\theta_2$ | Maximum grazing rate of zooplankton | 0.650 | U(0.150, 1.050) | U(0.450, 1.050) | 0.045 |

[a]Used in the state-augmentation experiment in section 4.4.
[b]Applied to $\theta_1$ and $\theta_2$ after the particle filter resampling; equal to $\frac{1}{20}$ of the range of its initial distributions.

$$x_t^\star = \begin{pmatrix} x_t \\ \theta_t \end{pmatrix}.$$

[33] The new, augmented state vector replaces the regular state vector in the particle filter procedure (section 2.1) so that the ensemble of state vectors now carries information about parameter values, which are weighted and resampled along with all other information contained in the vector. The process of resampling parameter values yields a time-dependent parameter distribution which can adapt to improve the fit to the observations. Suitable parameter values produce good state estimates which are characterized by being similar to the observations. This, in turn, results in high likelihood values for state vectors with suitable parameter values, ensuring that they are resampled with a higher probability than less suitable ones. As a result, after a number of resampling steps, the ensemble should be populated by the parameter values that best fit the observations [*Dowd*, 2011].

[34] In this application we augment the state vector by $n_\theta = 2$ parameters. As described in section 2.4, these parameters are two biological parameters, the maximum carbon-to-chlorophyll ratio ($\theta_1$) and the maximum zooplankton growth rate ($\theta_2$). Initial values for the parameters are drawn from parameter specific uniform distributions, which are listed in Table 1 and based on parameter ranges used in *Mattern et al*. [2012] (parameter ranges are also shown in Figure 1). In our implementation of the particle filter procedure, the parameters in the state vector are treated the same way as all other entries in $x_t^\star$, with the exception that perturbations of the model state only affect the parameters.

### 3.3. Ensemble Generation and the Error

[35] In our final experiments, we examine the influence of the ensemble size on our particle filter implementation. Perturbations to the state vectors are necessary in our particle filter setup to prevent the ensemble from degenerating. If no perturbations are introduced after resampling, two replicates of the same ensemble member would result in two identical model simulations, since both model state and parameter values are resampled together. Over the course of multiple assimilation steps, replicates of ensemble members with high weights could replicate further and the ensemble degenerate into a collection of one or just a few unique ensemble members. This is referred to as ensemble degeneracy or ensemble collapse and presents a well-known issue with particle filters [*Dowd*, 2006]. Many schemes have been developed to address this problem [*van Leeuwen*, 2009]. Here, we simply perturb the parameters that are part of the augmented state by adding normally distributed random noise to the parameter values after they have been resampled. This approach avoids modifications to the rest of the state vector, thus keeping the resampled model states true to the model. While the model contains other error sources, introduction of noise through parameters and other model inputs is an effective way to add uncertainty into the model results.

[36] In our implementation, the standard deviation of the parameter noise is specific to each parameter and listed in Table 1. In the rare cases where the noise causes a parameter value to become negative, the sign of that value is switched, ensuring positive parameter values. No measures are taken to prevent parameter values from exceeding any upper limit. From an implementation point of view, it is important that enough noise is added to each parameter to effectively counteract ensemble degeneracy, but not so much that state-augmentation becomes meaningless. That is, if the addition of noise causes the parameter values to deviate far from their original values, the positive effect of resampling the parameters will be lost.

[37] The evolution of the ensemble and its error sources can be summarized in a state-space equation. With state-augmentation, the state-space equation of this application is

$$\begin{pmatrix} x_{t+1} \\ \theta_{t+1} \end{pmatrix} = \begin{pmatrix} F_{\text{model}}(x_t, \theta_t) \\ \theta_t \end{pmatrix} + \begin{pmatrix} 0 \\ \varepsilon_\theta \end{pmatrix},$$

[38] where $F_{\text{model}}$ represents the model which propagates the biological state $x_t$ from one assimilation time step to the next, using the parameters in $\theta_t$. The parameter vector $\theta_t$ is only modified by the additive error term $\epsilon_\theta$ (as described above, no error is introduced to the biological state directly), which allows it to adapt to the observations through the filter-based estimation procedure.

[39] We know already that in this application time varying values of $\theta_1$ and $\theta_2$ can describe the data significantly better than the optimal fixed parameter set [*Mattern et al*., 2012]. To allow the particle filter to capture the time dependence of suitable parameter values, we choose a relatively high standard deviation for the parameter noise, permitting the parameter values to easily readjust over the course of assimilation (Figure 4).

[40] The variation of $\theta_1$ and $\theta_2$ gives rise to the idea of an error subspace: all of the error introduced into the ensemble originates in perturbations of these two parameters. Therefore, the particle filter ensemble needs to approximate the two-dimensional parameter distribution, rather than the much higher dimensional distribution of the state space. As a consequence, we may expect good representation of the distribution even for a relatively small ensemble.

[41] One shortcoming of this approach is that it is based on the assumption that the model error can be explained by a few parameters alone: While the simplified

**Table 2.** Configuration of the Particle Filter Experiments

| | Name | Section | $n_{ens}$ | S-A[a] | ADA | $n_{ADA}$ | Trunc. dist.[b] | $n_{rep}$[c] |
|---|---|---|---|---|---|---|---|---|
| | Reference simulation | 4.1 | | | The reference simulation does not use SIR | | | |
| 1 | Baseline | 4.2 | 20 | ✓ | ✓ | 4 | | 13 |
| 2.1 | No ADA | 4.3 | 20 | ✓ | | | | 3 |
| 2.2 | Strong ADA | 4.3 | 20 | ✓ | ✓ | 9 | | 5 |
| 3.1 | No state-augmentation | 4.4 | 20 | | ✓ | 4 | | 1 |
| 3.2 | Truncated baseline | 4.4 | 20 | ✓ | ✓ | 4 | ✓ | 1 |
| 3.3 | Truncated, no S-A[a] | 4.4 | 20 | | ✓ | 4 | ✓ | 1 |
| 4.1 | $n_{ens} = 40$ | 4.5 | 40 | ✓ | ✓ | 4 | | 11 |
| 4.1 | $n_{ens} = 10$ | 4.5 | 10 | ✓ | ✓ | 4 | | 11 |
| 4.2 | $n_{ens} = 5$ | 4.5 | 5 | ✓ | ✓ | 4 | | 11 |

[a]State-augmentation.
[b]Sing the truncated distribution.
[c]Umber of replicate simulations.

representation of phytoplankton and zooplankton in the model indicates that $\theta_1$ and $\theta_2$ are major error sources (which is also evidenced by the their temporal variation, Figure 1), there are numerous more error sources (e.g., other parameters, boundary and initial conditions, or the model's spatial and temporal discretization). We can expect good results from the error subspace approach if the parameters that are varied are the major sources of model error that can induce enough variability into the ensemble to avoid its degeneration.

### 3.4. Asynchronous Data Assimilation

[42] Particle filters that assimilate observations sequentially, and one-by-one, are typically very sensitive to outlying observations, which can induce strong, undesired shifts of the ensemble. To increase the robustness of the particle filter in the presence of outliers, data can be assimilated asynchronously, i.e., not individually as each observation becomes available, but as a set of observations at a later point in time. Many implementations of asynchronous data assimilation (ADA) have been developed for sequential techniques, e.g., for particle filters [*Godsill et al*., 2004; *van Leeuwen*, 2009; *Dowd*, 2011] and the EnKF [*Evensen and van Leeuwen*, 2000; *Sakov et al*., 2010] and are often referred to as fixed lag smoothers as they are meant to smooth the evolution of the ensemble and remove undesired shifts. The general idea is to run the model and collect several observations before assimilating them into the model together.

[43] Here, we implemented a simple ADA scheme mainly because the noise level in our satellite observations is high and they contain a large number of missing values. The latter may cause a single image to contain only localized information about part of the domain. Both factors contribute strongly to the rapid changes of optimal parameter values [*Mattern et al*., 2012], which are shown in Figure 1.

[44] Our ADA procedure is as follows: weights are computed at every analysis step but resampling occurs only every $n_{ADA} + 1$ steps. Resampling takes into account the current and the $n_{ADA}$ previous observations by averaging the last $n_{ADA} + 1$ weights for each ensemble member. Note that the model does not need to be stopped and restarted when no resampling occurs. It is sufficient to store model output at the time steps corresponding to the available observations, so that the weights can be computed when a full assimilation step is performed. The above procedure represents a simple way to combine $n_{ADA} + 1$ observational time steps into one assimilation step. Outlying observations still enter the SIR resampling but their effect on the ensemble is greatly reduced by the averaging process. We assess the effect of this ADA scheme in section 4.3.

## 4. Experiments and Results

### 4.1. Particle Filter and Model Setup

[45] We use the following basic setup for all experiments. Unless noted otherwise, our experiments are particle filter simulations from 1 January to 31 December, 2006. After an adjustment phase of 10 days, in which the initial values of $\theta_1$ and $\theta_2$ take effect, the 350 daily satellite observations are assimilated into the model. The initial conditions for the physical and biological variables were spun up for 1 year; initial parameters are taken from the distributions listed in Table 1 (see Table 2 for an overview of all experiments).

[46] To assess our particle filter results quantitatively, we compare them to a model simulation with optimized, fixed parameters. This reference simulation uses the identical model setup and optimized values for $\theta_1$ and $\theta_2$ (taken from *Mattern et al*. [2012]) that minimize the average AGB distance based on all observations that we use for our particle filter assimilation (including the removal of the estuaries data, see section 2.3). This means that our reference simulation produces the best chlorophyll output we can expect from a single deterministic model simulation. Therefore we consider particle filter simulations with similar performance to the reference simulation a success.

[47] In the following, we introduce our particle filter baseline experiment (section 4.2) and further experiments to assess the individual effects of ADA (section 4.3), state-augmentation (section 4.4) and the ensemble size (section 4.5). Because the particle filter results are based on parameters that are drawn from a uniform distribution initially and assigned random noise, we perform several replicate simulations of our experiments, in order to assess their robustness (see Table 2). The replicate simulations are exact copies of our particle filter experiments, except for the random terms. For experiments with replicate simulations, Table 3 lists the mean absolute residual and its standard

**Table 3.** Absolute Residuals (Mean ± Standard Deviation) for the Particle Filter Experiments

| | Average Absolute Residuals (mg chlorophyll $m^{-3}$) | | | |
|---|---|---|---|---|
| Name | Shelf Region[a] | Open Ocean Region[a] | April, Open Ocean | December, Open Ocean |
| Reference simulation | 0.628 | 0.241 | 0.579 | 0.400 |
| Particle filter baseline | 0.586 ± 0.021 | 0.251 ± 0.015 | 0.585 ± 0.115 | 0.356 ± 0.005 |
| ↪Worst[b] | 0.616 | 0.275 | 0.775 | 0.368 |
| ↪Best[b] | 0.551 | 0.228 | 0.417 | 0.349 |
| No ADA | 0.671 ± 0.101 | 0.239 ± 0.021 | 0.389 ± 0.016 | 0.361 ± 0.006 |
| Strong ADA | 0.676 ± 0.063 | 0.299 ± 0.021 | 0.919 ± 0.211 | 0.373 ± 0.023 |
| No state-augmentation | 0.661 | 0.260 | 0.725 | 0.381 |
| $n_{ens} = 40$ | 0.567 ± 0.010 | 0.243 ± 0.009 | 0.511 ± 0.041 | 0.359 ± 0.003 |
| $n_{ens} = 10$ | 0.598 ± 0.028 | 0.247 ± 0.011 | 0.525 ± 0.086 | 0.366 ± 0.010 |
| $n_{ens} = 5$ | 0.637 ± 0.069 | 0.263 ± 0.025 | 0.638 ± 0.218 | 0.365 ± 0.016 |
| Truncated baseline | 0.583 | 0.243 | 0.484 | 0.357 |
| Truncated, no S-A[c] | 0.803 | 0.344 | 1.012 | 0.345 |

[a]See Figure 5 for a map of the regions.
[b]The residuals of the worst and best results of the baseline experiment in each column are taken from different simulations.
[c]State-augmentation

deviation. In the table, we also included the best and the worst results of the baseline experiment in order to provide a full account of the performance encountered in the replicate simulations.

### 4.2. Experiment 1: Baseline Experiment

[48] Our baseline experiment uses ADA (with $n_{ADA} = 4$), state-augmentation, and an ensemble size of $n_{ens} = 20$. In order to compare the performance of the baseline experiment to our reference simulation, we focus on the average surface chlorophyll content in two subregions of the model domain (Figure 2 and Table 3): The division into estuaries (not considered here due to high noise in the observations), shelf and open ocean regions is mainly motivated by the different levels of surface chlorophyll present in these regions throughout the year and the associated differences in the plankton dynamics. The reference simulation fits the observations best in the months following the chlorophyll spring bloom (May, June and July) when chlorophyll levels stay relatively low and do not fluctuate much. Large discrepancies appear during the bloom in April in the open ocean region where the model overestimates the surface chlorophyll content, and late in the year in December when chlorophyll is underestimated. In order to examine if the particle filter can improve this fit, we include the mean absolute chlorophyll residual for April and December in Table 3, along with the mean absolute residuals for the entire year.

[49] In comparison to the reference simulation, our baseline particle filter configuration performs, on average, slightly better in the shelf region and slightly worse in the open ocean region (Table 3). For a better overview, we have included the simulations that achieved the worst, the best and the median results among replicate simulations (based on absolute residuals in the open ocean region) in Figures 5 and 6. There are some obvious differences between individual simulations that use the same configuration: Our median simulation (Figure 5a), for example, shows a small improvement of fit to the observations in April. In contrast, the best simulation does a much better job at predicting the decline of the chlorophyll bloom in early April, whereas the worst simulation overestimates chlorophyll in April beyond the high level of the reference simulation (Figure 5b).

[50] As indicated by the results in Figure 5b, the highest variation among reference simulations can be found in the April chlorophyll residuals (Table 3), where most but not all particle filter simulations can improve upon the reference simulation and where the best results show a very strong improvement. All particle filter simulations improve upon the reference simulation for the December residuals and the standard deviation among simulations is considerably smaller. In order to examine if the variation of the particle filter residuals is affected by ADA, we performed a follow-up experiment in section 4.3.

[51] The parameter development in our baseline experiment (Figures 6a and 6b; for $\theta_2$ see online appendix) reveals that the values of the augmented parameters do not converge to a fixed value but roughly retrace the smoothed optimal time-dependent values for $\theta_1$. This means that the assimilation can capture the time-dependence of $\theta_1$, the parameter with the strongest effect on model chlorophyll. For $\theta_2$, which has a smaller impact on surface chlorophyll [*Mattern et al.*, 2012], optimal parameters are retraced much less closely. The extent to which the chlorophyll estimates benefit from state-augmentation is investigated in section 4.4.

### 4.3. Experiment 2: Effect of ADA

[52] In our second experiment, we assessed the effects of ADA on the particle filter results. For this purpose, we performed several replicate simulations where ADA was deactivated (each observation was assimilated individually) and simulations where ADA was increased over a larger smoothing window of size $n_{ADA} = 9$.

[53] The average, year-long absolute residuals and their variability among the experiments have notably increased without ADA in the shelf region (Table 3), the particle filter becomes less robust and the general performance over the time span of the entire experiment drops. Yet, the average results in the open ocean region are slightly improved compared to the baseline experiment, and a stronger improvement is visible in the April residuals, when the associated variability is lower too. One possible

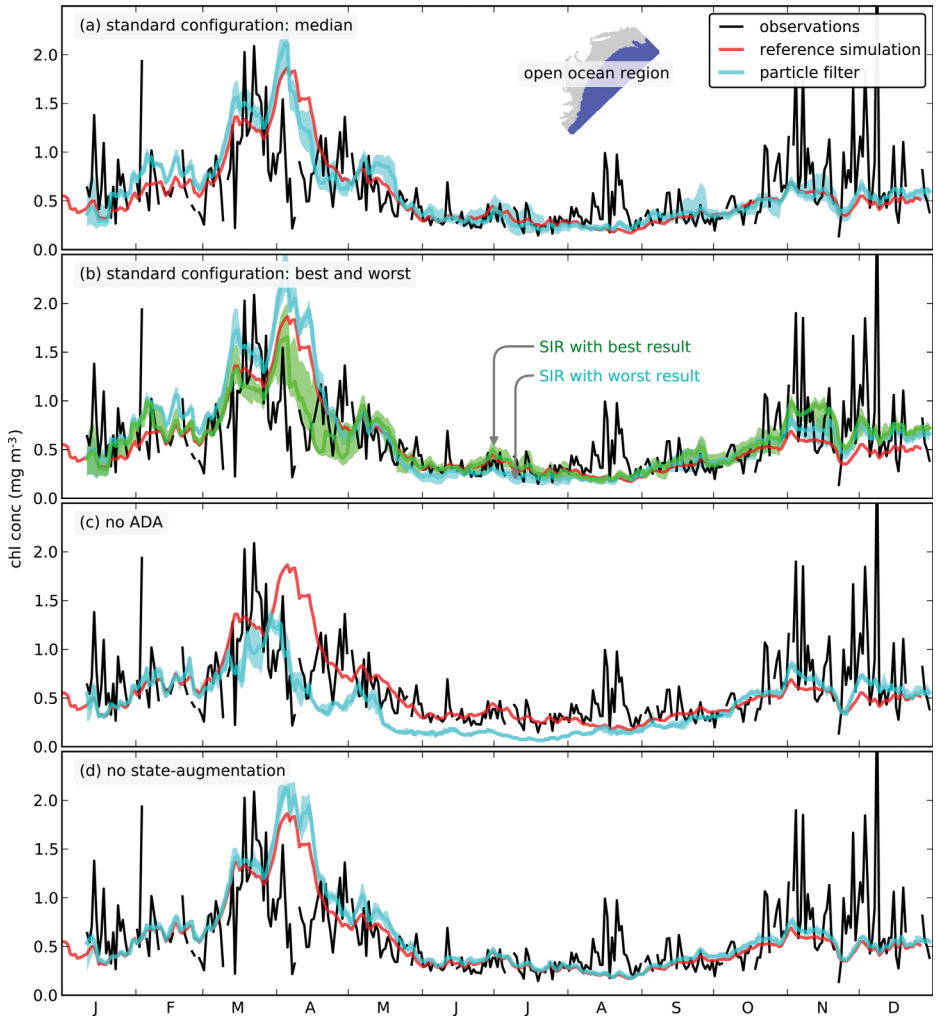


**Figure 5.** The surface chlorophyll development in the open ocean region (see inset; highlighted in blue) for selected particle filter experiments in comparison to the observations and the reference simulation with optimized parameters. For the particle filter, both ensemble median and the region between the 0.1 and the 0.9 quantile are shown. The panels correspond to different particle filter experiments. (a) The particle filter simulation that achieved the median fit to the observations among the replicates of the baseline experiment (section 4.2) and (b) those with the worst and best fit are displayed. (c) An experiment without ADA (section 4.3) and (d) an experiment without state-augmentation (section 4.4) are depicted.

explanation for these results can be found in the evolution of the parameter values (Figure 6c). In instances where there is a large deviation between the model and the chlorophyll observations, the daily assimilation can quickly adjust model state and parameter values. This can be advantageous in the short term (April) where it can lead to a more consistent improvement compared to the standard experiment. Yet strong shifts can occur even when the particle filter estimates are close to the observations, these have a negative impact on assimilation performance and frequently force the parameter estimates away from the optimal parameter values. For example, a strong shift in $\theta_1$ in May (see Figure 6c) causes a long-lasting underestimation of chlorophyll (Figure 5c). Overall, the negative shifts outweigh the positive ones, evidenced by the year-long results in Table 3.

[54] One additional problem that affects the simulations without ADA is the low ensemble spread (compare Figure 5c to 5a or 5b) which hinders a quick recovery from shifts in the ensemble. The low spread is caused by the faster succession of assimilation steps (daily compared to every 5 days in the baseline experiment) which contract the ensemble more frequently, an effect that can be counteracted by increasing the parameter noise. Furthermore, the use of ADA decreases the number of independent observations that are assimilated. More independent observations, in the simulations without ADA, can have a negative effect on particle filter performance, and have been shown to increase the probability of ensemble collapse [*Snyder et al.*, 2008; *van Leeuwen*, 2009].

[55] For $n_{\mathrm{ADA}} = 9$, the mean performance degrades compared to the baseline experiment (Table 3). Here it appears that the wider gap between assimilation steps prohibits a quick adjustment of the ensemble, causing changes in the state due to assimilation to become sluggish.

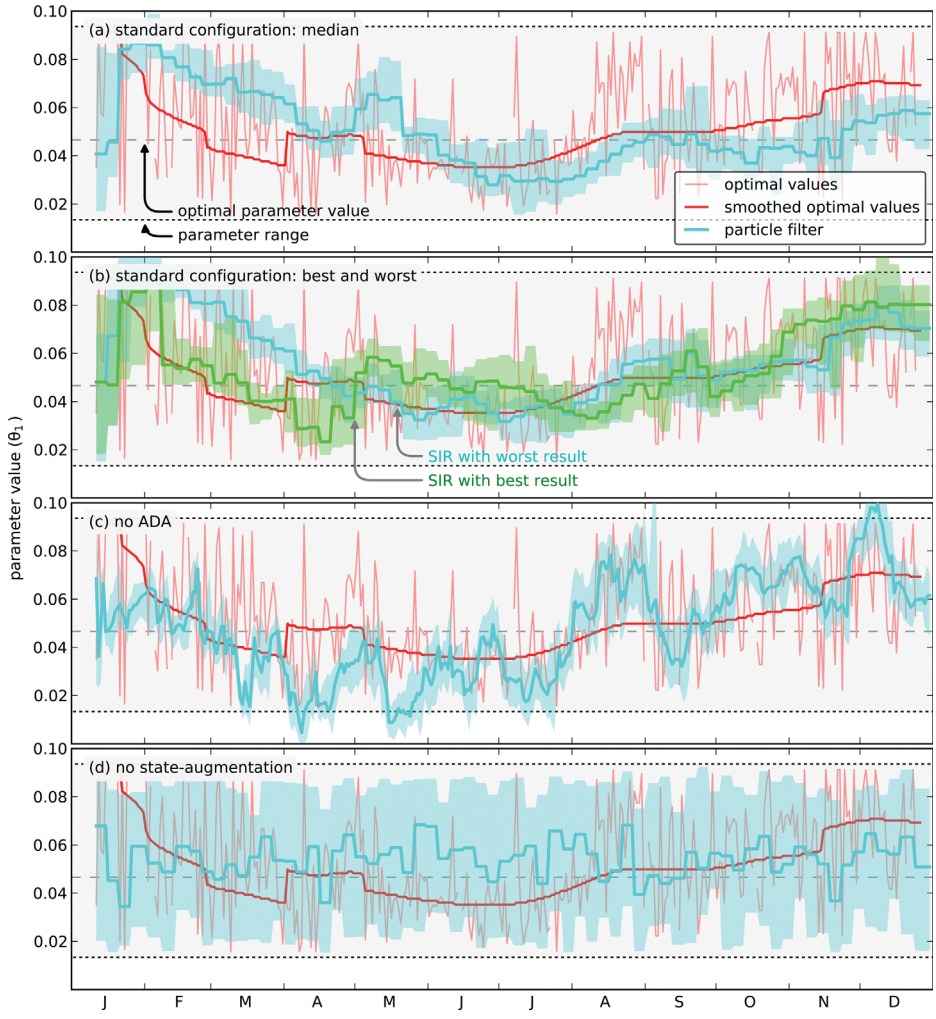


**Figure 6.** The development of $\theta_1$ for the same experiments depicted in Figure 5. The faint red line and the thick red line correspond to the optimal parameter values obtained in *Mattern et al.* [2012] (compare Figure 1). For the particle filter, both ensemble median and the region between the 0.1 and the 0.9 quantile are shown. The light gray area between the dotted horizontal line marks the parameter range, the optimal fixed parameter value is denoted by the dashed horizontal line (see Table 1).

[56] In summary, the ADA has a stabilizing effect on our particle filter simulations. Without ADA, resampled parameter values appear less smooth in time and the particle filter state estimates degrade due to outliers in the data. In contrast, too much smoothing results in a dampened response. By averaging the weights over five time steps with our ADA procedure, we were able to reduce the problem posed by outliers in the observations while still permitting parameter values to change in time.

### 4.4. Experiment 3: Effect of State-Augmentation

[57] In this experiment we examine the effects of state-augmentation which serves as a way to jointly estimate parameter and state values with particle filters and provides the ensemble with a parameter memory: suitable parameter values are "remembered" by carrying them over from one assimilation step to the next. Our baseline simulation with state-augmentation exhibits a clear pattern in the ensemble of parameter values, retracing the optimal time-dependent parameter values for $\theta_1$ (Figure 6a). This indicates that a memory of suitable parameter values is advantageous and may be required to obtain agreement with the chlorophyll observations.

[58] To assess the influence of state-augmentation, we performed a particle filter simulation with the same configuration as the baseline experiment, but deactivated state-augmentation. Parameter values were no longer resampled but instead redrawn from a static distribution at the end of every assimilation step, rendering the parameter distribution essentially memoryless (illustrated in Figure 6d). In a first experiment, we used the same uniform distributions that was used as the initial parameter distribution in the baseline experiment. To assess the influence of the parameter distribution in particular, we performed an additional experiment replacing the original distribution with a truncated one (see Table 1).

[59] The ensemble median of our particle filter simulation without state-augmentation remains close to the reference simulation (Figure 5d) and consequently the performance of the two simulations are very similar (Table

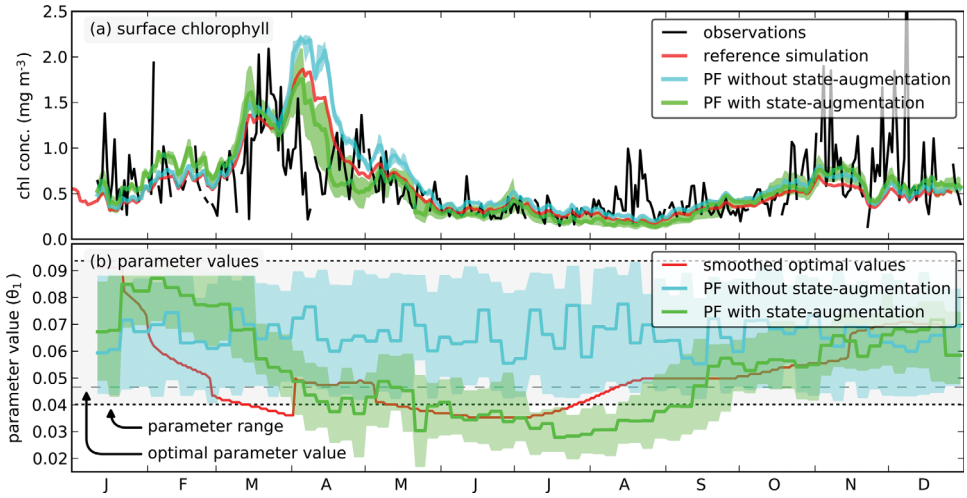


**Figure 7.** The development of the (a) surface chlorophyll concentration and (b) $\theta_1$ for the particle filter with truncated parameter distributions (compare section 4.4) with and without state-augmentation. In both panels, the ensemble median and the region between the 0.1 and the 0.9 quantile are shown for both particle filter experiments. Figure7a also contains the observations and the reference simulation, Figure7b shows the smoothed optimal values (obtained from *Mattern et al*. [2012]). The light gray area between the dotted horizontal line in Figure7b marks the truncated parameter range, the optimal fixed parameter value is denoted by the dashed horizontal line (see Table 1).

3). The resampling of states seems to have little effect on the evolution of the ensemble when it is combined with a random assignment of parameter values. In fact, the ensemble median of the average surface chlorophyll concentration remains consistently above the output of the reference simulation throughout the simulation (Figure 5d). This bias is present even when it implies a severe overestimation of the chlorophyll concentration (see for example the April result in Figure 5d). In this particle filter configuration, the resampling cannot effectively correct chlorophyll estimates. Instead, the chlorophyll estimates appear to be strongly dependent on the static parameter distributions chosen for the experiment. This parameter distribution is centered on parameter values close to those used in reference simulation, explaining the close match between the simulation without state-augmentation and the reference simulation.

[60] To further assess the influence of the parameter distributions, we performed an additional experiment with modified distributions. Again, we used uniform distributions but truncated their range by one third on the lower side (see Table 1). As a result, the optimal parameter values are no longer close to the mean of their respective distributions. The use of these new distributions adds a bias to the chlorophyll estimates and causes a clear decline in the fit to the observations for the particle filter without state-augmentation (Figure 7). With state-augmentation turned on, however, the effects of the ill-chosen initial parameter distribution are barely noticeable, as the resampling quickly determines the time evolution of parameter values (Figure 7b).

[61] A closer investigation reveals that the particle filter implementation without state-augmentation produces, on average, considerably lower weights during resampling when compared to particle filter with state-augmentation. Due to the random assignment of parameter values and the small ensemble size, none of the particles experiences a lasting benefit from suitable parameter values. This negative effect cannot be effectively counteracted by the SIR resampling.

[62] In summary, it is apparent that the parameter memory provided by state-augmentation is an important asset in this application. It allows the particle filter to move the parameter distributions to suitable regions in parameter space which are allowed to vary in time. In addition, good results can be obtained without prior knowledge of suitable parameters values, as state-augmentation provides the particle filter with a means to adjust parameters quickly and break out of ill-posed initial distributions.

### 4.5. Experiment 4: Ensemble Size

[63] In our final experiments, we examine the influence of the ensemble size on our particle filter implementation: In a typical particle filter application, the ensemble represents a sample from the probability distribution of the state variables. In our implementation, where the ensemble operates in a lower dimensional error subspace, the augmented part of the ensemble approximates the probability distribution of the two parameters $\theta_1$ and $\theta_2$. Regardless of the exact dimension of the approximated probability distribution, it is expected that a larger ensemble and therefore a higher sample size leads to a better representation and more accurate results. However, large ensembles are not practical because they result in a higher computational cost due to additional model simulations.

[64] In addition to the standard ensemble size of 20, we tested sizes of 5, 10, and 40. Computational constraints also force us to keep the number of replicate experiments at 11, a relatively low number which might not eliminate the influence of outliers on the results. Gradually increasing the ensemble size from 5 to 40 leads to the expected improvement in average performance (Table 3). It is notable that, in

comparison to the baseline experiment, a decrease in the ensemble size to 10 yields no strong decline in performance across the results listed in Table 3 (a further reduction of the ensemble size to 5, however, causes a larger drop in the average performance). It may seem surprising that 10 ensemble members are sufficient to approximate the high-dimensional distribution of the model state. This is a result of the two-dimensional error subspace that the particle filter operates in. Since state-perturbations are only caused by two parameters, it is sufficient that the ensemble can approximate the two-dimensional parameter distribution and its time-evolution to obtain adequate results. An expanded state-augmentation that incorporates more than two parameters will likely profit more from larger ensemble sizes.

[65] Besides average performance, the increase of the ensemble size also improves the precision of the particle filter, as indicated by the standard deviations listed in Table 3. The reduction of the standard deviation associated with increasing the ensemble size from 20 (baseline experiment) to 40 highlights the importance of large ensembles. With 40 ensemble members, particularly bad particle filter results (such as the worst replicate of the baseline simulation in Figure 5b) are less likely to occur.

[66] In summary, increasing the ensemble size leads, as expected, to an increased accuracy and precision of our particle filter results. While 10 ensemble members are sufficient to achieve good results, larger ensembles markedly increase the robustness of our particle filter.

## 5. Discussion and Conclusions

[67] We have implemented a particle filter algorithm for data assimilation of real daily satellite observations of chlorophyll into a time dependent 3-D biological ocean model for the purpose of joint state and parameter estimation. Particle filters offer advantages over other ensemble data assimilation technique in that it can produce the correct target distributions for the state and parameters for general nonlinear models and non-Gaussian errors, a feature which is not supported by the widely used EnKF [*van Leeuwen*, 2009].

[68] One of the main problems in using particle filters for data assimilation and complex models is ensemble collapse and the accompanying low ensemble spread that prevents effective data assimilation. Ensemble collapse is often caused by an undersampling of the state space [*Snyder et al*., 2008]. We have shown in this study that ensemble collapse can also be caused by outlying observations which are ubiquitous in biological applications. In high-dimensional applications, the state space is typically sparsely populated by a small ensemble relative to the number of dimensions. As a consequence, one or just a few ensemble members tend to lay close to the observation (in the sense of obtaining a high likelihood) and receive high weights in the particle filter-based assimilation procedure. This leads to the resampling of the few highly weighted ensemble members and the collapse of the ensemble. A similar process can occur due to the presence of outlying observations. Outliers deviate notably from the rest of the observations and tend therefore not to lay close to those regions of the state space where previous observations have consolidated the ensemble. Again, only the few close ensemble members receive high weights. As the experiment in section 4.3 shows, outlying observations can bring an ensemble close to collapse and can therefore effectively impede data assimilation.

[69] In this application, we avoid the problem posed by the high dimensionality of our model by effectively operating the data assimilation in a lower dimensional error subspace. This is achieved by introducing stochasticity into the ensemble solely by varying a few model parameters that the model is very sensitive to. Furthermore, the stochastic parameters are incorporated into the augmented state. The variation of parameters as a means by which stochastic variation is added to the ensemble, is especially suitable for biological models [*Annan*, 2001]. These models contain parameters with a large range of possible values that are in practice not very well-known and whose effect on the model output is great. For this study we chose parameters that the model chlorophyll is very sensitive to, ensuring adequate spread in the predictive ensemble with respect to chlorophyll for most of the simulation. At some points however, the predictive ensemble does not cover the observations well and as a consequence the improvement obtained by the particle filter assimilation remains relatively low (e.g., the surface chlorophyll development in November and December in the open ocean region in Figure 5).

[70] It should be noted, however, that other model errors, especially from physical sources, are not represented and would have to be considered for a more comprehensive assessment [*Palmer et al*., 2005]. We expect that the addition of more error sources, physical or biological (e.g., in the form of additional parameters via state-augmentation), will improve the particle filter state estimates at the cost of a requirement for larger ensembles. Parameter estimates could likely worsen, however, as dependencies between the biological parameters and the additional error sources become an issue. More work is required to assess the exact benefits and drawbacks of a higher dimensional error subspace and the effect of adding physical error sources on biological state and parameter estimation. Recently, some other particle filtering approaches suitable for state estimation for large scale systems have been proposed: *van Leeuwen* [2010] suggests modifying the initial distribution of the error sources (in this application: $\theta_1$ and $\theta_2$) with the aim of preventing ensemble collapse by using future observational information, and *Chorin and Tu* [2009] suggest a nonlinear particle filter based on implicit sampling that behaves well for low numbers of particles. *Briggs et al*. [2013] suggest using a particle smoother in space along with the estimation in time. Our use of error subspace provides an alternative approach, that is useful for both state and parameter estimation.

[71] To address the detrimental effect of observational noise and outliers, we made use of ADA in our particle filter implementation. Without ADA, single anomalous observations can introduce strong shifts into the ensemble during assimilation (Figure 6d), an effect which can persist over multiple assimilation steps and decrease the fit to the observations significantly. In fact, the particle filter configuration without ADA produced some of the worst results among our particle filter experiments (Table 3). The simple ADA scheme we introduced in section 3.4 is

straightforward to implement and essentially averages over multiple observations before assimilating them into the ensemble, increasing the robustness of the particle filter. The second implementation aspect that increases the particle filter's robustness is the weighting during resampling. In this study, we base the particle filter's weights on transformed inverse distance values, and introduce a parameter that allows for suitable spreading of the weights. This approach is flexible, as it allows the use of various types of observations along with suitable model-data distance measures.

[72] An effect of using complex numerical models is that their computational cost precludes the use of large ensemble sizes (even in very high dimensional problems, such as numerical weather prediction, ensemble sizes are between 5 and 100) [*Gneiting and Raftery*, 2005]. Ensemble-based data-assimilation techniques (e.g., particle filters) that are built on these relatively small samples can show strong variability in their results from one experiment to the next. To quantify this variability we ran multiple replicates of our main experiments (see Table 2 for the number of replicates for each experiment). Most particle filter experiments stayed close to the reference simulation, yet we observed some variation, both beneficial and detrimental. The use of multiple realizations to quantify sampling variability is a standard tool in statistics [*Gelman and Rubin*, 1992], and its use is important in assessing the effectiveness of all data assimilation schemes relying on small ensembles.

[73] State-augmentation allows for joint state and parameter estimation within the particle filter procedure [*Dowd*, 2011]. In our experiments, the particle filter benefits from the resampling of parameters, improving chlorophyll estimates and allowing parameters to diverge from ill-posed initial distributions. For the two biological parameters that are part of the state-augmentation it is realistic to assume time-dependence [*Mattern et al*., 2012]. Parameter estimates obtained via state-augmentation could identify the low frequency time evolution of $\theta_1$, the chlorophyll-to-carbon ratio in all of our replicate experiments (Figures 6a and 6b). For $\theta_2$, the less sensitive maximum grazing rate of zooplankton, the estimates are less robust and vary greatly amongst different particle filter simulations. In our application, state-augmentation is more useful in improving state estimates than in obtaining robust parameter estimates.

[74] Predictive skill is often used to assess the capabilities of a model in combination with a data assimilation technique, it quantifies how long a model can retain improvements introduced via assimilation. An assessment of predictive skill in *Hu et al*. [2012] for the same model and the same assimilated observations shows that a localized EnKF can increase the model's surface chlorophyll forecast skill for time spans of a few days to a maximum of 25 days–dependent on the magnitude of the improvement introduced by the EnKF, which differs with the time of year. Because we use the same model, we would expect similar results from our particle filter implementation: Besides a strong dependence of predictive skill on the improvement introduced by the particle filter (which also displays a strong time-dependence, see Figure 5), we would anticipate a positive effect of the state-augmentation, with seasonal parameter values slowing the decline in the model's forecast performance.

[75] In summary, particle filters offer the possibility for sequential ensemble data assimilation to be applied to complex, high-dimensional ocean models and real observations. To work successfully, we provided modifications to the standard particle filter, dealing with problems posed by outlying observations and the high-dimensionality of the model. With these modifications the particle filter achieves a better model-data fit than an optimized model simulation.

[76] **Acknowledgments.** We thank Peter Jan van Leeuwen, Bruce Smith, and Keith Thompson for valuable comments on an earlier version of this manuscript. We also thank Pavel Sakov and an anonymous reviewer whose comments let to improvements. This work was supported by the ONR MURI grant N00014-06-1-0739 to K.F. K.F. is also acknowledging support from ACEnet, NSERC and CFI. M.D. acknowledges support from NSERC.

## References

Annan, J. (2001), Modelling under uncertainty: Monte Carlo methods for temporally varying parameters, *Ecol. Modell.*, *136*(23), 297–302, doi:10.1016/S0304-3800(00)00413-0.

Annan, J., and J. Hargreaves (2010), Efficient identification of ocean thermodynamics in a physical/biogeochemical ocean model with an iterative Importance Sampling method, *Ocean Modell.*, *32*(3-4), 205–215, doi:10.1016/j.ocemod.2010.02.003.

Bennett, A. (2002), *Inverse Modeling of the Ocean and Atmosphere*, Cambridge Univ. Press, Univ. Print. House, Cambridge, U. K.

Bianucci, L., K. L. Denman, and D. Ianson (2011), Low oxygen and high inorganic carbon on the Vancouver Island Shelf, *J. Geophys. Res.*, *116*, C07011, doi:10.1029/2010JC006720.

Briggs, J., M. Dowd, and R. Meyer (2013), Data assimilation for large-scale spatio-temporal systems using a location particle smoother, *Environmetrics*, doi:10.1002/env.2184, in press.

Chen, K., and R. He (2010), Numerical investigation of the middle atlantic bight shelfbreak frontal circulation using a high-resolution ocean hindcast model, *J. Phys. Oceanogr.*, *40*(5), 949–964, doi:10.1175/2009JPO4262.1.

Chorin, A. J., and X. Tu (2009), Implicit sampling for particle filters, *Proc. Natl Acad. Sci*., *106*(41), 17,249–17,254, doi:10.1073/pnas.0909196106.

Ciavatta, S., R. Torres, S. Saux-Picart, and J. I. Allen (2011), Can ocean color assimilation improve biogeochemical hindcasts in shelf seas? *J. Geophys. Res.*, *116*, C12,043, doi:10.1029/2011JC007219.

Dowd, M. (2006), A sequential Monte Carlo approach for marine ecological prediction, *Environmetrics*, *17*(5), 435–455, doi:10.1002/env.780.

Dowd, M. (2007), Bayesian statistical data assimilation for ecosystem models using Markov Chain Monte Carlo, *J. Mar. Syst.*, *68*(3-4), 439–456, doi:10.1016/j.jmarsys.2007.01.007.

Dowd, M. (2011), Estimating parameters for a stochastic dynamic marine ecological system, *Environmetrics*, *22*(4), 501–515, doi:10.1002/env.1083.

Evensen, G. (2009), *Data Assimilation: The Ensemble Kalman Filter*, Springer, Berlin, Germany.

Evensen, G., and P. J. van Leeuwen (2000), An ensemble Kalman smoother for nonlinear dynamics, *Monthly Weather Rev.*, *128*(6), 1852–1867, doi:10.1175/1520-0493(2000)128<1852:AEKSFN>2.0.CO;2.

Fennel, K. (2010), The role of continental shelves in nitrogen and carbon cycling: Northwestern North Atlantic case study, *Ocean Sci*., *6*(2), 539–548, doi:10.5194/os-6–539-2010.

Fennel, K., and J. Wilkin (2009), Quantifying biological carbon export for the northwest North Atlantic continental shelves, *Geophys. Res. Lett.*, *36*, L18605, doi:10.1029/2009GL039818.

Fennel, K., J. Wilkin, J. Levin, J. Moisan, J. O'Reilly, and D. Haidvogel (2006), Nitrogen cycling in the middle Atlantic bight: Results from a three-dimensional model and implications for the North Atlantic nitrogen budget, *Global Biogeochem. Cycles*, *20*(3), doi:10.1029/2005GB002456.

Fennel, K., J. Wilkin, M. Previdi, and R. Najjar (2008), Denitrification effects on air-sea CO(2) flux in the coastal ocean: Simulations for the northwest North Atlantic, *Geophys. Res. Lett.*, *35*, L24608, doi:10.1029/2008GL036147.

Geider, R., H. MacIntyre, and T. Kana (1998), A dynamic regulatory model of phytoplanktonic acclimation to light, nutrients, and temperature, *Limnol. Oceanogr.*, *43*(4), 679–694, doi:10.4319/lo.1998.43.4.0679.

Gelman, A., and D. Rubin (1992), Practical Markov chain Monte Carlo: Rejoinder: Replication without Contrition, *Stat. Sci.*, *7*(4), 503–511, doi:10.1214/ss/1177011148.

Gneiting, T., and A. E. Raftery (2005), Weather forecasting with ensemble methods, *Science*, *310*(5746), 248–249, doi:10.1126/science.1115255.

Godsill, S. J., A. Doucet, and M. West (2004), Monte Carlo smoothing for nonlinear time series, *J. Am. Stat. Assoc.*, *99*(465), 156–168, doi:10.1198/016214504000000151.

Gordon, N., D. Salmond, and A. Smith (1993), Novel approach to nonlinear/non-Gaussian Bayesian state estimation, *Radar and Signal Process., IEE Proc. F*, *140*(2), 107–113, doi:10.1049/ip-f-2.1993.0015.

Haidvogel, D., et al. (2008), Ocean forecasting in terrain-following coordinates: Formulation and skill assessment of the regional Ocean modeling system, *J. Comput. Phys.*, *227*(7), 3595–3624, doi:10.1016/j.jcp.2007.06.016.

Hu, J., K. Fennel, J. P. Mattern, and J. Wilkin (2012), Data assimilation with a local ensemble Kalman filter applied to a three-dimensional biological model of the middle Atlantic bight, *J. Mar. Syst.*, *94*(0), 145–156, doi:10.1016/j.jmarsys.2011.11.016.

Kitagawa, G. (1996), Monte Carlo filter and smoother for non-Gaussian nonlinear state space models, *J. Comput. Graph. Stat.*, *5*(1), 1–25, doi:10.2307/1390750.

Kitagawa, G. (1998), A self-organizing state-space model, *J. Am. Stat. Assoc.*, *93*(443), 1203–1215, doi:10.2307/2669862.

Mattern, J. P., M. Dowd, and K. Fennel (2010a), Sequential data assimilation applied to a physical-biological model for the Bermuda Atlantic time series station, *J. Mar. Syst.*, *79*(1-2), 144–156, doi:10.1016/j.jmarsys.2009.08.004.

Mattern, J. P., K. Fennel, and M. Dowd (2010b), Introduction and assessment of measures for quantitative model-data comparison using satellite images, *Remote Sens.*, *2*(3), 794–818, doi:10.3390/rs2030794.

Mattern, J. P., K. Fennel, and M. Dowd (2012), Estimating time-dependent parameters for a biological ocean model using an emulator approach, *J. Mar. Syst.*, *9697*(0), 32–47, doi:10.1016/j.jmarsys.2012.01.015.

Palmer, T., G. Shutts, R. Hagedorn, F. Doblas-Reyes, T. Jung, and M. Leutbecher (2005), Representing model uncertainty in weather and climate prediction, *Annu. Rev. Earth Planet. Sci.*, *33*(1), 163–193, doi:10.1146/annurev.earth.33.092203.122552.

Previdi, M., K. Fennel, J. Wilkin, and D. Haidvogel (2009), Interannual variability in atmospheric CO2 uptake on the northeast U.S. continental shelf, *J. Geophys. Res.*, *114*, G04,003, doi:10.1029/2008JG000881.

Ristic, B., S. Arulampalam, and N. Gordon (2004), *Beyond the Kalman Filter: Particle Filters for Tracking Applications*, Artech House, London, U. K.

Sakov, P., G. Evensen, and L. Bertino (2010), Asynchronous data assimilation with the EnKF, Dyn. Meteorol. Oceanogr., Tellus Series A, *62*(1), 24–29, doi:10.1111/j.1600-0870.2009.00417.x.

Snyder, C., T. Bengtsson, P. Bickel, and J. Anderson (2008), Obstacles to high-dimensional particle filtering, *Monthly Weather Rev.*, *136*(12), 4629–4640, doi:10.1175/2008MWR2529.1.

van Leeuwen, P. J. (2003), A variance-minimizing filter for large-scale applications, *Monthly Weather Rev.*, *131*(9), 2071–2084, doi:10.1175/1520-0493(2003)131<2071:AVFFLA>2.0.CO;2.

van Leeuwen, P. J. (2009), Particle filtering in geophysical systems, *Monthly Weather Rev.*, *137*(12), 4089–4114, doi:10.1175/2009MWR2835.1.

van Leeuwen, P. J. (2010), Nonlinear data assimilation in geosciences: An extremely efficient particle filter, *Q. J. R. Meteorol. Soc.*, *136*(653), 1991–1999, doi:10.1002/qj.699.

Wikle, C. K., and L. M. Berliner (2007), A Bayesian tutorial for data assimilation, *Physica D*, *230*(12), 1–16, doi:10.1016/j.physd.2006.09.017.