Improving Variational Data Assimilation through Background and Observation Error Adjustments

JANN PAUL MATTERN, CHRISTOPHER A. EDWARDS, AND ANDREW M. MOORE

Department of Ocean Sciences, University of California, Santa Cruz, Santa Cruz, California

(Manuscript received 21 September 2017, in final form 20 November 2017)

ABSTRACT

A procedure to objectively adjust the error covariance matrices of a variational data assimilation system is presented. It is based on popular diagnostics that utilize differences between observations and prior and posterior model solutions at the observation locations. In the application to a data assimilation system that combines a three-dimensional, physical–biogeochemical ocean model with large datasets of physical and chlorophyll *a* observations, the tuning procedure leads to a decrease in the posterior model-observation misfit and small improvements in short-term forecasting skill. It also increases the consistency of the data assimilation system with respect to diagnostics, based on linear estimation theory, and reduces signs of overfitting. The tuning procedure is easy to implement and only relies on information that is either prescribed to the data assimilation system or can be obtained from a series of short data assimilation experiments. The implementation includes a lognormal representation for biogeochemical variables and associated modifications to the diagnostics. Furthermore, the effect of the length of the observation window (number and distribution of observations) used to compute the diagnostics and the effect of neglecting model dynamics in the tuning procedure are examined.

1. Introduction

The availability of large observational datasets and the increased importance of numerical models as scientific tools is leading to a growing use of data assimilation (DA). Application of DA to high-dimensional models in the geosciences in general, and to biogeochemical (BGC) ocean models in particular, have created significant improvements in model skill based on a variety of metrics (Edwards et al. 2015; Stammer et al. 2016). Yet, many opportunities for improvement remain for DA methods. In particular, some of the most widely used DA techniques require the specification of observation error and background error values, which have a strong influence on DA results but are difficult to determine objectively. In applications, these uncertainties are often based on heuristic approaches that can lead to inconsistencies (with respect to linear estimation theory) between specified error values and the response of the DA system. Recently, the diagnostics introduced in Desroziers et al. (2005) set the foundation for a variety of approaches to more objectively estimate observation and background error values for variational and ensemble DA systems.

Observation and background error values are typically specified in the forms of $\mathbf{R}$ and $\mathbf{B}$, the observation and background error covariance matrices, respectively. The dimensions of $\mathbf{R}$ are determined by the number of observations used in the DA application. Observation error covariances, the entries of $\mathbf{R}$, are determined by instrument noise, the error of representativeness, and other error sources, such as errors in the preprocessing of the data. These observation error contributions and their structures are typically not well understood and are difficult to estimate objectively.

The dimensions of $\mathbf{B}$ are determined by the length of the model state vector (model grid size $\times$ number of model variables), which is upward of 10^6 for typical three-dimensional ocean models. Because of its size alone, the entries of $\mathbf{B}$ that represent the error of the prior estimate are difficult to determine. In practical applications, background error values are often based on an ensemble of model simulations or a single (long) model simulation without DA, thereby assuming that background errors are proportional to variability in the nonassimilative model output. Inflation coefficients are used in some applications to rescale observation or background error values, thereby modifying the weight that observations or model solutions have in the assimilation. Some operational DA

Corresponding author: Jann Paul Mattern, jmattern@ucsc.edu

DOI: 10.1175/MWR-D-17-0263.1

systems employ complex covariance models that can perform updates to the error estimates over the course of a DA simulation (see, e.g., Dee et al. 2011). Recently, the error covariance diagnostics presented in Desroziers et al. (2005) have gained in popularity and have found use in a variety of applications to estimate entries or subsets of entries of **R**, **B**, or inflation coefficients.

The diagnostics in Desroziers et al. (2005) (hereafter referred to as error covariance diagnostics) are based on linear estimation theory, which forms the theoretical basis of variational and many ensemble-based DA techniques [see, e.g., Talagrand (1999) for a summary]. They can be easily computed based on differences between the observation values and the background (prior) and analysis (posterior) model solutions at the observation locations and are, therefore, also often referred to as observation-minus-background, observation-minus-analysis, and analysis-minus-background statistics. Their computation requires an analysis solution and, thus, output from one or multiple DA experiments (a more detailed description of the error covariance diagnostics and their computation is presented in section 2c).

The error covariance diagnostics were originally introduced as consistency diagnostics, allowing for a relatively simple way to check the error specifications of DA systems. They have found a wide variety of applications in various DA systems. Early use of the diagnostics include the estimation of observational error values and inflation coefficients for an ensemble Kalman filter–based DA system (Li et al. 2009), the estimation of observational errors including error correlations (off-diagonal elements of **R**) in a variational DA application (Bormann et al. 2010), and the estimation of the sensitivity of DA-based model forecasts on entries in **R** and **B** (Daescu and Todling 2010). More recently, the use of error covariance diagnostics has gained in popularity, especially applied to atmospheric models; several studies, including Stewart et al. (2014), Waller et al. (2016b), Bormann et al. (2016), and Campbell et al. (2017), use them to estimate observation error values for different observation types. The estimation includes error correlations, and each study highlights the importance of including these correlations in DA systems, especially for spatially and temporally dense observations such as satellite data. Other uses of the error covariance diagnostics include the estimation of model error in weak constraint DA systems [using an extension of the error covariance diagnostics; Bowler (2017)] and accounting for the lack of model error in strong constraint DA systems through adjustments of **R** (Howes et al. 2017). Applied to ocean models, the error covariance diagnostics have been used in the adjustment of observation errors and in the estimation of a time-dependent inflation coefficient accounting for changes in the global observing network in a historical ocean reanalysis (Yang et al. 2017). Similar statistics, based on the difference between observations and nonassimilative model solutions at the observation locations and obtained from an ensemble of simulations, are used in Karspeck (2016) to estimate observational error values for a 1°-resolution global ocean model.

In practical applications, estimates of **R** and **B** obtained from the error covariance diagnostics will not be exact and are dependent on the presumed error structure. As noted in Desroziers et al. (2005), the simultaneous estimation of **R** and **B** requires that the spectra of the matrices differ sufficiently, and Ménard (2016) highlights additional limitations of the error covariance diagnostics in the joint estimation of observation and background error values. Several studies further show that if observation error correlations exist in the data but are omitted in the formulation of **R**, observation error estimates obtained from the error covariance diagnostics tend to be inaccurate and can degrade assimilation results (Bormann et al. 2016; Ménard 2016; Waller et al. 2016b). Despite their known limitations, useful results can still be obtained from the error covariance diagnostics, even if the errors are not perfectly specified (Bormann et al. 2016; Waller et al. 2016b), and many studies report improvements in assimilation results after modifications of the DA systems based on the error covariance diagnostics (Bormann et al. 2016; Campbell et al. 2017; Cordoba et al. 2017).

In this study, we apply error-covariance-diagnostics-based covariance adjustments to a four-dimensional variational (4D-Var) DA system consisting of a three-dimensional coupled physical–BGC ocean model with physical and satellite chlorophyll *a* data. This DA setup has previously been presented in Mattern et al. (2017), where manual modifications to **B** were used to reduce the magnitude of increments to the BGC model variables and increase their magnitude for the physical variables. The approach presented here is based on a fixed-point iteration (FPI) that iteratively determines multiplicative factors that adjust parts of **R** and **B** associated with different subsets of observations, a technique similar to that presented in Desroziers and Ivanov (2001) and Bölöni and Horvath (2010). We investigate if the adjustments have a positive effect on the error covariance diagnostics and, beyond that, on other metrics, such as the model forecasting skill. As this represents the first application of error covariance diagnostics to a coupled physical–BGC DA system, our examination highlights the balance between increments to the physical and the BGC variables when observations of both types are assimilated jointly. We thus examine if this procedure avoids the need for manual adjustments, which are cumbersome and frequently based

TABLE 1. The data used for assimilation.

Observed variables	Data source	URL	Data type
Chlorophyll *a*	MODIS (on board *Aqua*)	https://oceandata.sci.gsfc.nasa.gov/MODIS-Aqua	Satellite
SSH	AVISO	http://www.aviso.altimetry.fr	Satellite
Salinity (*S*)	Aquarius	http://podaac.jpl.nasa.gov/aquarius	Satellite
Temperature (*T*)	OSTIA	http://ghrsst-pp.metoffice.com/ostia/	Satellite
T + *S*	Met Office	http://hadobs.metoffice.com/en4/	In situ (profiles)
T + *S*	UCSD glider	http://spray.ucsd.edu	In situ (glider)
T + *S*	Argo	http://www-argo.ucsd.edu/	In situ (float)

on subjective criteria. By using two configurations of our DA system with different error structures (one strongly simplified), we further examine the effect of the error structure on the convergence characteristics of the FPI and the DA results. Furthermore, we assess the sensitivity of the FPI technique on the number of assimilation cycles (length of the observation window and number of observations) used to generate the error covariance diagnostics (section 2c), the effect of model dynamics on the covariance estimates (section 2d), and the attractiveness of the fixed point we obtained with our procedure (section 2e).

2. Methods

a. Model and observations

The coupled physical–BGC model is based on the Regional Ocean Modeling System [ROMS; version 3.7, revision 737; Haidvogel et al. (2008)]. The model domain covers the California Current System (CCS; 30° to 48°N, coastline to 134°W) at a horizontal resolution of 0.1° × 0.1°; it is divided into 42 terrain-following layers vertically. Boundary conditions and physical forcing (wind, solar radiation, air temperature, pressure, and humidity) are based on output from COAMPS (Doyle et al. 2009). More details about the physical model are provided in Veneziani et al. (2009) and Raghukumar et al. (2015), which use a setup identical to our present application.

The BGC model is the North Pacific Ecosystem Model for Understanding Regional Oceanography (NEMURO; Kishi et al. 2007), which contains 11 BGC variables, including two phytoplankton that represent different size classes: large phytoplankton (LP) simulate diatoms, dominant in the coastal waters of the CCS, while small phytoplankton (SP) represent smaller species more prevalent offshore. LP and SP are assumed to have fixed but different nitrogen-to-chlorophyll *a* ratios, affecting the observation operator H for chlorophyll *a* and the setup of our FPI below. NEMURO parameter values are taken from and listed in Mattern and Edwards (2017), a parameter estimation study using the same model domain. The setup of the coupled physical–BGC model is identical to that in Mattern et al. (2017), where more information can be found.

In our experiments, we assimilate satellite-derived surface chlorophyll *a* data as the only BGC data jointly with physical data for temperature, salinity, and sea surface height (SSH) anomaly. The physical data include in situ, satellite-based, and reanalysis-based data. All data sources are listed in Table 1.

b. 4D-Var-based assimilation system and log transformation

The DA system uses the incremental, strong-constraint 4D-Var implemented in ROMS; our implementation computes increments to the initial conditions for the model's state variables over 4-day assimilation cycles, using an iterative conjugate-gradient algorithm in order to minimize the cost function:

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x}_b - \mathbf{x})^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}_b - \mathbf{x}) + \frac{1}{2}[\mathbf{y} - H(\mathbf{x})]^{\mathrm{T}}\mathbf{R}^{-1}[\mathbf{y} - H(\mathbf{x})]. \quad (1)$$

Here, $\mathbf{x}$ is the model state vector, and $\mathbf{x}_b$ is its initial estimate, also referred to as the background model state. The vector $\mathbf{y}$ contains the observations, typically at different points in time and space, and H is an operator mapping $\mathbf{x}$ to $\mathbf{y}$, accounting for temporal progression based on the model dynamics and spatial interpolation. The $\mathbf{B}$ and $\mathbf{R}$ are the background and observation error covariance matrices, respectively, that are the focus of this study and that are adjusted in the procedure detailed in the sections below. The 4-day cycle length was chosen as a duration over which the nonlinear model is typically well approximated by its tangent linear representation.

In our setup, we assume normal distributions for all physical variables, which is the standard 4D-Var approach. For the BGC variables and chlorophyll *a* we assume lognormal distributions. The lognormal assumption for chlorophyll *a* better represents its characteristics in nature

(Campbell 1995); for the BGC variables the lognormal assumption has shown advantages in DA scenarios (Song et al. 2012). These assumptions imply that the distribution of the errors in the physical variables is normal, whereas the distribution of the errors in the log-scaled BGC variables is normal [for details, see Fletcher and Zupanski (2006)]. This change requires modifications to the standard cost function in Eq. (1), and these changes are described in detail in Song et al. (2012, 2016a). The structure of J remains essentially the same as in the standard, purely Gaussian DA, but with $\mathbf{B}$ and $\mathbf{R}$ describing the statistics of variables with either normal or lognormal distributions. Applications of this 4D-Var setup with lognormal distributions are described in Song et al. (2016a) and Mattern et al. (2017). In the following, we implicitly presume that all variables are appropriately transformed to fulfill the assumptions of the log-transformed 4D-Var in Song et al. (2016a). This implies that the entries in $\mathbf{B}$ and $\mathbf{R}$ associated with BGC variables are covariances of the log-transformed variables; covariance entries between different variables (log transformed or not) are zero. The lognormal assumption for chlorophyll *a* and the contribution of LP and SP requires modifications to the FPI, which are detailed in appendix B. The basic technique, however, is general and does not rely on the log transformation.

c. Error covariance diagnostics

Following Desroziers et al. (2005), we define the following properties that are dependent on the observations $\mathbf{y}$, the prior (background) model solution at the observation locations $H(\mathbf{x}_b)$, and the posterior (analysis) model solution at the observation locations $H(\mathbf{x}_a)$:

$$\mathbf{d}_a^o = \mathbf{y} - H(\mathbf{x}_a), \tag{2}$$

$$\mathbf{d}_b^o = \mathbf{y} - H(\mathbf{x}_b), \tag{3}$$

$$\mathbf{d}_b^a = H(\mathbf{x}_a) - H(\mathbf{x}_b) \tag{4}$$

Thus, $\mathbf{d}_a^o$ and $\mathbf{d}_b^o$ represent model-data misfit for the posterior and prior model solutions, respectively, and $\mathbf{d}_b^a$ quantifies the DA increments at the observation locations. For a linearized observation operator $\mathbf{H}$, Desroziers et al. (2005) show that the following relationships should approximately hold for correctly specified variational DA systems:

$$\mathbb{E}(\mathbf{d}_b^a \mathbf{d}_b^{o\mathrm{T}}) \approx \mathbf{HBH}^{\mathrm{T}}, \tag{5}$$

$$\mathbb{E}(\mathbf{d}_a^o \mathbf{d}_b^{o\mathrm{T}}) \approx \mathbf{R}, \tag{6}$$

where $\mathbb{E}(\cdot)$ is the expected value. Hence, Eqs. (5) and (6) provide diagnostics that can be used to assess the consistency of the specified background and observation errors with respect to the error covariance diagnostics. In practice, it is simpler to evaluate the following (weaker) relationships that are only based on the diagonal elements of the matrices in Eqs. (5) and (6) and are easy to compute in DA applications:

$$\tilde{\sigma}_b^{(i)} = \sqrt{\frac{1}{|O_i|}\sum_{j\in O_i} \mathbf{d}_{bj}^a \mathbf{d}_{bj}^o} \approx \sqrt{\frac{1}{|O_i|}\sum_{j\in O_i} (\mathbf{HBH}^{\mathrm{T}})_{jj}} = \overline{\sigma}_b^{(i)},$$

$$\tilde{\sigma}_o^{(i)} = \sqrt{\frac{1}{|O_i|}\sum_{j\in O_i} \mathbf{d}_{aj}^o \mathbf{d}_{bj}^o} \approx \sqrt{\frac{1}{|O_i|}\sum_{j\in O_i} \mathbf{R}_{jj}} = \overline{\sigma}_o^{(i)}, \tag{7}$$

where the O_i, for $i = 1, \ldots, n_i$, denote subsets of observation indices, allowing for the distinction of observation subsets, such as different observation types. The error covariance diagnostics $\overline{\sigma}_b^{(i)}$ and $\overline{\sigma}_o^{(i)}$ are based on $\mathbf{B}$ and $\mathbf{R}$, properties that must be defined prior to performing DA, while $\tilde{\sigma}_b^{(i)}$ and $\tilde{\sigma}_o^{(i)}$ can only be determined after the assimilation is completed. In our FPI implementation below, the computation of $\overline{\sigma}_b^{(i)}$ does not include the tangent linear model dynamics contained in $\mathbf{H}$ (rather, $\overline{\sigma}_b^{(i)}$ is based on the entries of $\mathbf{B}$ at the observation locations), a potential issue noted in Neveu et al. (2016) and Bowler (2017) that is further investigated in section 2d. Our implementation also does not include a bias removal term, as used in Waller et al. (2016a) and Cordoba et al. (2017).

d. Fixed-point iteration

Building on ideas in Desroziers and Ivanov (2001) and Desroziers et al. (2005), we construct an FPI to adjust background and observation errors in an attempt to improve the consistency of the error statistics used in our DA system with respect to the error covariance diagnostics. Based on Eq. (7), we set up the following iterative procedure:

$$\overline{\sigma}_{b(k+1)}^{(i)} = \tilde{\sigma}_{b(k)}^{(i)} \quad \text{and} \quad \overline{\sigma}_{o(k+1)}^{(i)} = \tilde{\sigma}_{o(k)}^{(i)} \quad \text{for} \quad i = 1, \ldots, n_i \quad \text{and} \quad k = 0, \ldots, n_k. \tag{8}$$

The iteration starts at $k = 0$ with default values for $\mathbf{B}$ and $\mathbf{R}$ that determine $\overline{\sigma}_{b(k)}^{(i)}$ and $\overline{\sigma}_{o(k)}^{(i)}$. After performing a DA simulation, typically consisting of multiple cycles, $\tilde{\sigma}_{b(k)}^{(i)}$ and $\tilde{\sigma}_{o(k)}^{(i)}$ can be determined. Here, we compute the error covariance diagnostics for each DA cycle individually and then average across cycles to obtain $\tilde{\sigma}_{b(k)}^{(i)}$ and $\tilde{\sigma}_{o(k)}^{(i)}$. In the next step, $\mathbf{R}$ and $\mathbf{B}$ are adjusted to satisfy the equalities in Eq. (8). The rows and columns in $\mathbf{B}$ associated with observation type i are multiplied by $\lambda_b^{(i)} = \tilde{\sigma}_{b(k)}^{(i)}/\overline{\sigma}_{b(k)}^{(i)}$. The $\mathbf{R}$ is treated analogously, using the multiplier $\lambda_o^{(i)} = \tilde{\sigma}_{o(k)}^{(i)}/\overline{\sigma}_{o(k)}^{(i)}$; because $\mathbf{R}$ is diagonal, this amounts to multiplying all diagonal elements $\mathbf{R}_{jj}$ associated with observation type i by $[\lambda_o^{(i)}]^2$. This step

TABLE 2. The initial values for the diagonal entries of **B** and **R** used in DA configurations DAC1 and DAC2. Background error values for DAC1 are based on a long model simulation and are not shown here.

Variable	Units	$(\mathbf{B}_{jj})^{1/2}$ DAC2 surface	$(\mathbf{R}_{jj})^{1/2}$ DAC2	$(\mathbf{R}_{jj})^{1/2}$ DAC1
SSH	m	0.03	0.03	0.03
Temperature (T)	°C	1.0	0.3	In situ: 0.1, other: 0.3
Salinity (S)	—	0.1	0.3	In situ: 0.01, other: 0.3
u velocity	$\mathrm{m\,s^{-1}}$	0.03	—	—
v velocity	$\mathrm{m\,s^{-1}}$	0.03	—	—
LP	$\mathrm{mmol\,N\,m^{-3}}$	0.3	—	—
SP	$\mathrm{mmol\,N\,m^{-3}}$	0.3	—	—
Log-chlorophyll	$\log(\mathrm{mmol\,chl}a\,\mathrm{m^{-3}})$	—	0.3	0.3
Unobserved BGC	Various	0.3	—	—

determines the new values for $\overline{\sigma}^{(i)}_{b(k+1)}$ and $\overline{\sigma}^{(i)}_{o(k+1)}$, k is incremented, and the above procedure is repeated in the next iteration. In the following, we drop the FPI index k from our notation; when error covariance diagnostics appear in the same equation, they have the same FPI index.

In our FPI procedure, we use different observation types for adjusting **B** and **R**. For **B**, we distinguish between temperature, salinity, SSH, and log-chlorophyll, the latter yielding two multipliers for LP and SP (which are not independent; see appendix B). For **R**, we may additionally split up temperature and salinity observations into satellite and in situ categories (see below for details). Consequently, each iteration determines the values for 9 to 11 multipliers, and background errors for unobserved variables remain unchanged.

e. The data assimilation configurations

We test the FPI procedure for two DA configurations that only differ in their initial values for **B** and **R** and that showcase the range of complexity possible in different DA systems. In the more complex DA configuration 1 (DAC1), the entries in **B** are based on variations around the climatological monthly mean of a long (>10 yr) model simulation without DA. That is, daily snapshots of the model state are aggregated for each climatological month, and their variances determine the diagonal elements of **B**, resulting in 12 covariance matrices that are selected and used in the DA based on the start of the DA cycle (note that our FPI algorithm does not distinguish between months and that the same $\lambda^{(i)}_b$ multipliers are applied to all **B**). This configuration is nearly identical to that used in Mattern et al. (2017), where a multiplicative factor of 0.1 is used to reduce the BGC background error values in order to prevent undesired changes to the biological model state in unobserved variables. For the same reason, DAC1 also uses a multiplicative reduction of 0.1 for entries in **B**, but here, it is only applied to entries for unobserved variables that are not adjusted by the FPI.

DA configuration 2 (DAC2) has a much simpler setup for **B**: diagonal entries associated with surface values are constant for each variable (values for the variables are listed in Table 2), and a logistic decline with depth is then applied to the surface values to determine subsurface entries. This depth reduction is meant to lower the impact of DA at depth where oceanic variability is weaker, and the prior estimate is presumably more accurate; below 200 m, variance values are approximately 10% of their magnitude at the surface. In ROMS, decorrelation length scales determine the off-diagonal elements for **B**; as in Song et al. (2016b), we use a horizontal decorrelation length scale of 50 km for all variables and a vertical one of 30 and 7 m for physical and BGC variables, respectively. In all cases **B** is univariate, block diagonal.

Our two DA configurations also differ in their entries for **R**. DAC2 uses only four observation types, whereas DAC1 distinguishes between in situ and satellite temperature and salinity observations (values are listed in Table 2). All initial observation error values are roughly based on those used in a previous DA application (Mattern et al. 2017) but are simplified to one value for each observation type.

3. Results

a. Convergence

We first examine the FPI's convergence characteristics using an FPI setup where each iteration of the FPI is based on 10 DA simulations, each consisting of two DA cycles. We thus refer to this DA setup as the 10×2 setup. The 10 simulations are evenly spaced across a 3-yr period of interest (2013–15; the start date of the first simulation is 5 January 2013, and that of the last is 6 September 2015) in order to capture and better

represent interannual and intrannual differences in data availability, model misfit, (monthly) background error values, and more. In our application, it is especially important to include intrannual differences in the DA cycles, largely because of seasonal differences in the BGC model dynamics, while the length of each simulation (here, two cycles) is less crucial. Initial conditions for each of the 10 simulations are provided by a DA simulation without covariance adjustments; more than 3 million observations are assimilated in the 20 nonconsecutive DA cycles of the 10 × 2 setup (Fig. 1, center column).

In our tests, the FPIs converge rapidly for both DAC1 and DAC2 (Fig. 2). In both DA configurations, relatively large changes to $\overline{\sigma}_b$ and $\overline{\sigma}_o$ are made during the first three iterations, after which only smaller adjustments occur for all observation types. We stop each FPI after five iterations, at which stage most multipliers have values close to 1 (the largest deviation from 1 is the DAC1 log-chlorophyll *a* background error multiplier with a value of 0.880; see Fig. 2d), and we do not expect significant improvements from additional iterations. The largest overall relative changes occur for the in situ temperature and in situ salinity observation error, which increase by factors of 7 and 10, respectively (with a noticeable effect in the results below). Apart from these large changes, $\overline{\sigma}_b$ and $\overline{\sigma}_o$ remain between 25% and 250% of their original values for all other observation types.

Despite differences in configurations that cannot be eliminated by the FPI, such as the spatial structure of background errors in DAC1 or differences in entries for unobserved variables, $\overline{\sigma}_b$ and $\overline{\sigma}_o$ converge to very similar values for DAC1 and DAC2. For DAC2, where no distinction is made between in situ and other observation types, the $\overline{\sigma}_o$ for temperature and salinity converge toward the same values as DAC1's $\overline{\sigma}_o$ for the satellite observations, which far outnumber in situ observations and, thus, largely determine the error covariance diagnostics. These similarities between the final entries in **B** and **R** for the two DA configurations indicate that the procedure converges toward the same fixed point. Further tests in section 3e highlight that the FPI converges to nearly identical values if the initial entries in **B** and **R** are modified by random multipliers. Following the FPI procedure, can we now expect similar DA results for the tuned DAC1 and DAC2?

b. Consistency and model-observation misfit

In our assessment, we evaluate the effects of the covariance adjustments by performing a yearlong DA simulation for each of the two DA configurations before and after the covariance adjustments. Each simulation starts on 1 January 2013 and spans all of 2013, during

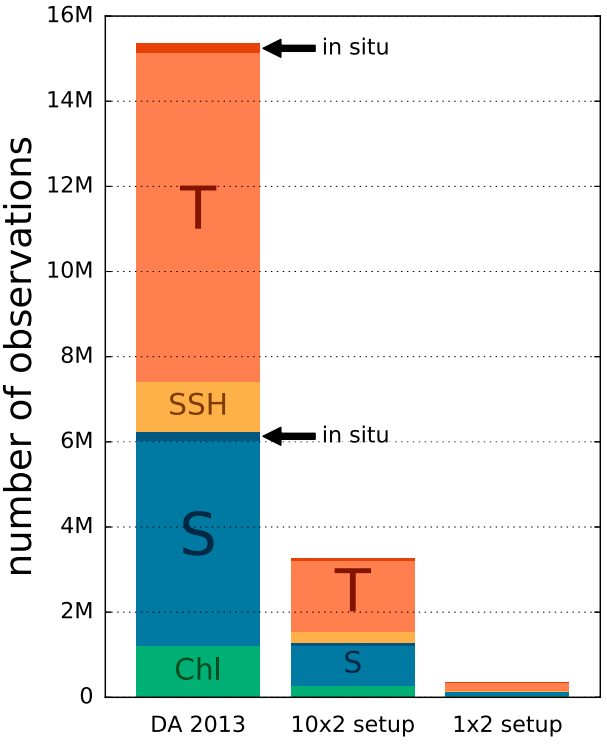

FIG. 1. The number of temperature (*T*; orange), SSH (yellow), salinity (*S*; blue), and chlorophyll *a* (Chl; green) observations used in our experiments; darker colors mark in situ observations (see arrows). Our FPI is based on the (center) 10 × 2 setup. The prior and posterior model-observation misfits in section 3b were determined based on (left) a longer DA simulation spanning all of 2013, while the FPI sensitivity experiments in section 3c are based on (right) the 1 × 2 setup.

which more than 15 million observations (Fig. 1, left column) are assimilated in 92 cycles. To compare the DA configurations, we used several metrics that distinguish between observation types and DA cycles. That is, the metrics below were computed for each cycle and observation type individually (and are, thus, dependent on $O_{i,c}$, the observations associated with observation type i and DA cycle c).

We are primarily interested in improving the consistency of the DA system, with respect to the error covariance diagnostics, and assessing its effect on the DA result. To quantify this consistency, we define the metric e_σ as

$$e_\sigma^{(i,c)} = \frac{1}{2}\left|\frac{\tilde{\sigma}_o^{(i,c)} - \overline{\sigma}_o^{(i,c)}}{\overline{\sigma}_o^{(i,c)}}\right| + \frac{1}{2}\left|\frac{\tilde{\sigma}_b^{(i,c)} - \overline{\sigma}_b^{(i,c)}}{\overline{\sigma}_b^{(i,c)}}\right|, \qquad (9)$$

representing an evenly weighted average of fractional discrepancies between expected statistics. A lower value of $e_\sigma^{(i,c)}$ indicates higher consistency with respect to the observations included in $O_{i,c}$. The FPI leads to an overall

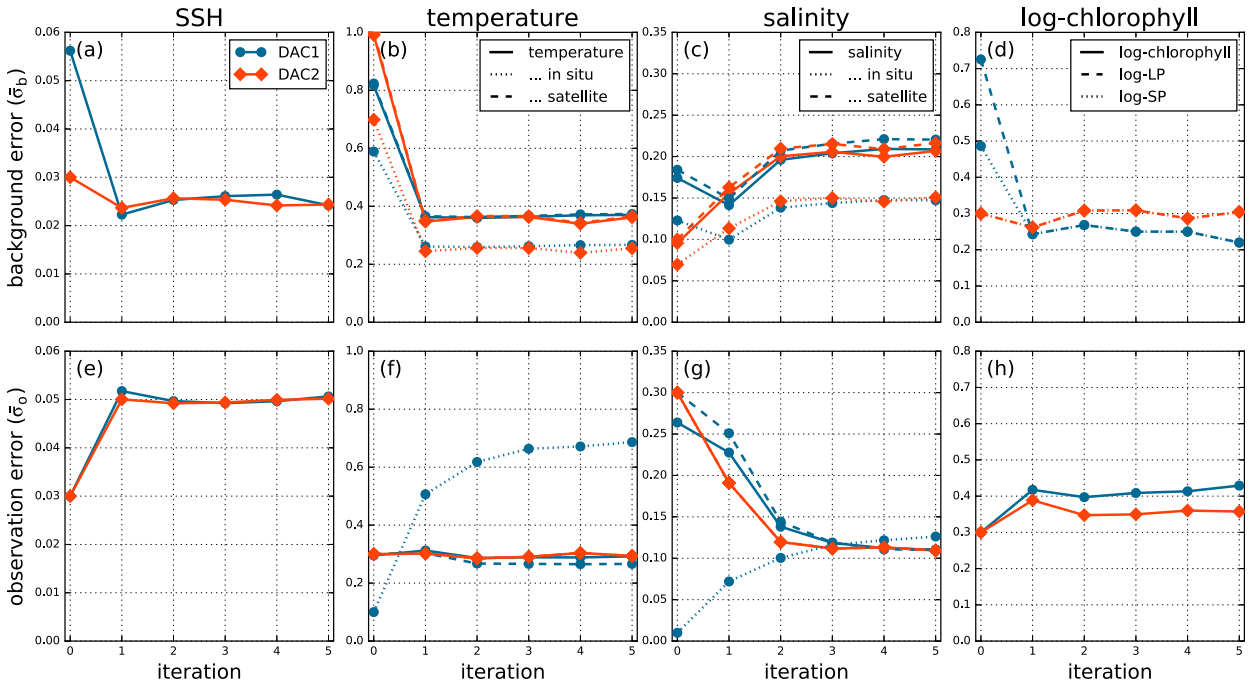

FIG. 2. Convergence of (a)–(d) $\overline{\sigma}_b$ and (e)–(h) $\overline{\sigma}_o$ for DAC1 (blue) and DAC2 (red). Each column corresponds to an observation variable, and in situ (dotted lines) and satellite (dashed line) observations are distinguished by line style (for reference, in situ and satellite observation results for DAC2 are included, as well as aggregate temperature and salinity results for DAC1). (d) For log-chlorophyll, we distinguish between $\overline{\sigma}_b$ for LP and SP, which are identical for DAC2.

improvement in consistency for all observation types for both DAC1 and DAC2 (Figs. 3a,b). As expected, the FPI leads to generally improved consistency values. Only a small number of cycles show an increase in e_σ following the tuning procedure; this occurs mostly for two observation types, log-LP and log-SP in DAC1. Notably, the largest improvement in consistency occurs for in situ temperature and salinity in DAC1, which have especially high e_σ values in the untuned system. The FPI also leads to an improvement in an alternative consistency metric that is based on the posterior (minimum) value of the cost function: in a consistent DA system, $J(\mathbf{x}_a)$ is expected to be χ^2-distributed with n_{obs} degrees of freedom, where n_{obs} is the number of observations. Hence, $\mathbb{E}[J(\mathbf{x}_a)/n_{\text{obs}}] = 1$ and $J(\mathbf{x}_a)/n_{\text{obs}} \approx 1$ can be used as a consistency diagnostic [often referred to as "J_{min}" diagnostic; see Talagrand (1999) and Ménard (2016)]. In both configurations, the value of $J(\mathbf{x}_a)/n_{\text{obs}}$ is closer to 1 in the tuned systems, with a bigger improvement for DAC1; while there is still large variability between cycles, the cycle average $J(\mathbf{x}_a)/n_{\text{obs}}$ is moved from 1.97 to 0.99 by the tuning (Fig. 4a). For DAC2, the untuned system performs slightly better, and the FPI leads to an improvement from 0.88 to 0.92 (Fig. 4b).

Besides consistency, an important measure of performance is the decrease in model misfit due to DA. We quantify the posterior model-observation misfit as the RMSE between the posterior model solution and observations as

$$e_a^{(i,c)} = \sqrt{\frac{1}{|O_{i,c}|} \sum_{j \in O_{i,c}} \mathbf{d}_{aj}^{o^2}}. \tag{10}$$

The FPI procedure results in improved posterior RMSEs, which are strongly dependent on the observation type (Figs. 3c,d). For DAC1, the FPI creates a large reduction in the posterior salinity RMSE (>40% on a cycle average) in combination with small reduction for satellite temperature. A small increase in the posterior RMSE occurs for chlorophyll *a*, and a larger increase occurs for the in situ observations, consistent with the increase in observation error for these observation types produced by the FPI (cf. Fig. 2). For DAC2, there is a notable decrease in salinity and log-chlorophyll RMSE (>20%), while the posterior temperature error remains, on average, unchanged. In both configurations, the FPI has negligible impact on the SSH misfit of the posterior solution.

A lower posterior error does not necessarily constitute a superior posterior model solution, as a closer model fit to any particular dataset may be due to overfitting (fitting the solution to the random observation

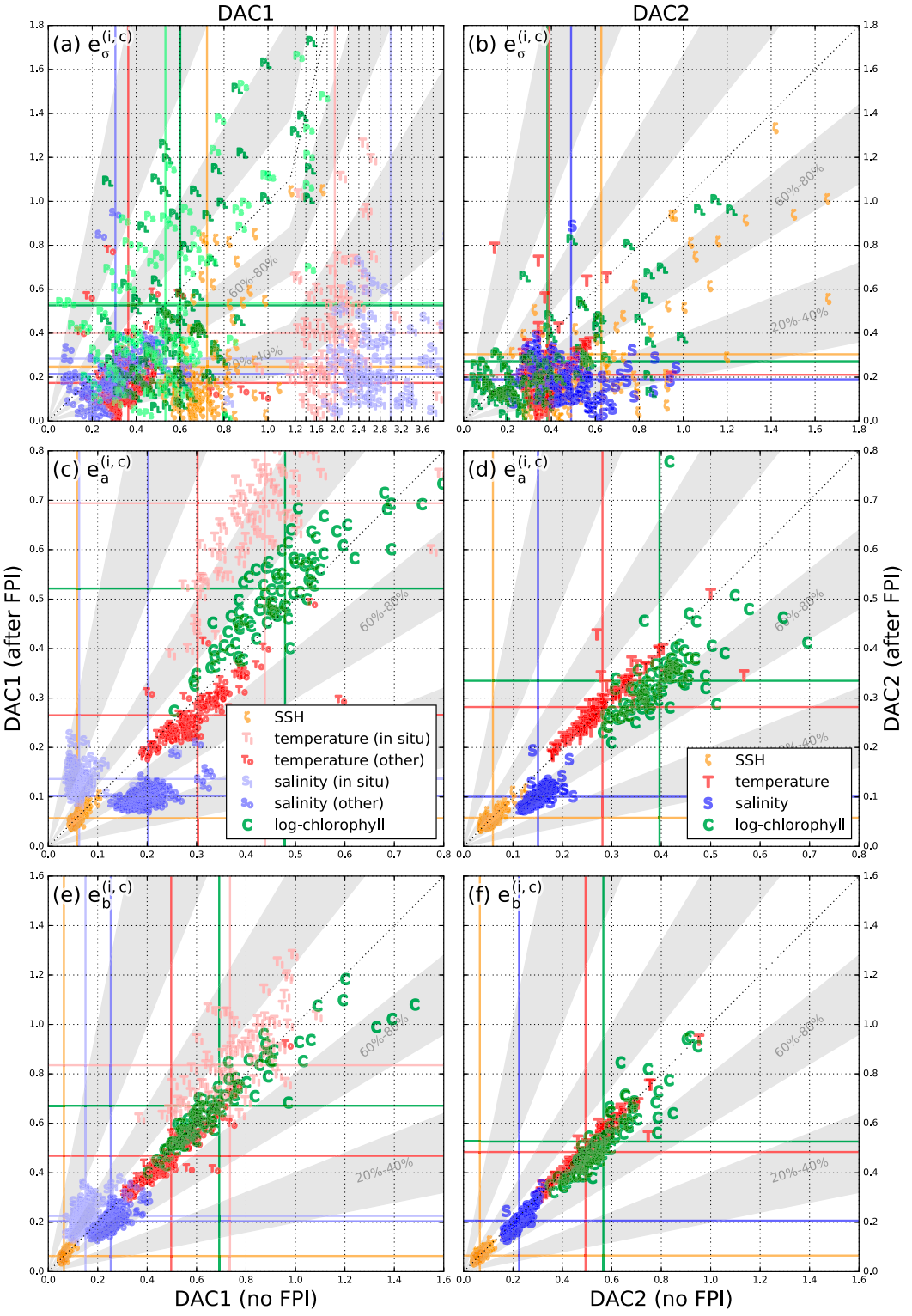

FIG. 3. Scatterplots for the metrics introduced in section 3b for (a),(c),(e) DAC1 and (b),(d),(f) DAC2; each panel shows the results of the DA before (x axis) and after the FPI (y axis) was used to adjust **B** and **R**. Average values for each observation type are marked by solid lines. For DAC1, the results are split into in situ and satellite observations for temperature and salinity. Note that the x axis in (a) is compressed beyond 1.1 to accommodate high e_σ values and that log-chlorophyll is split into LP (dark green P_L) and SP (light green P_S).

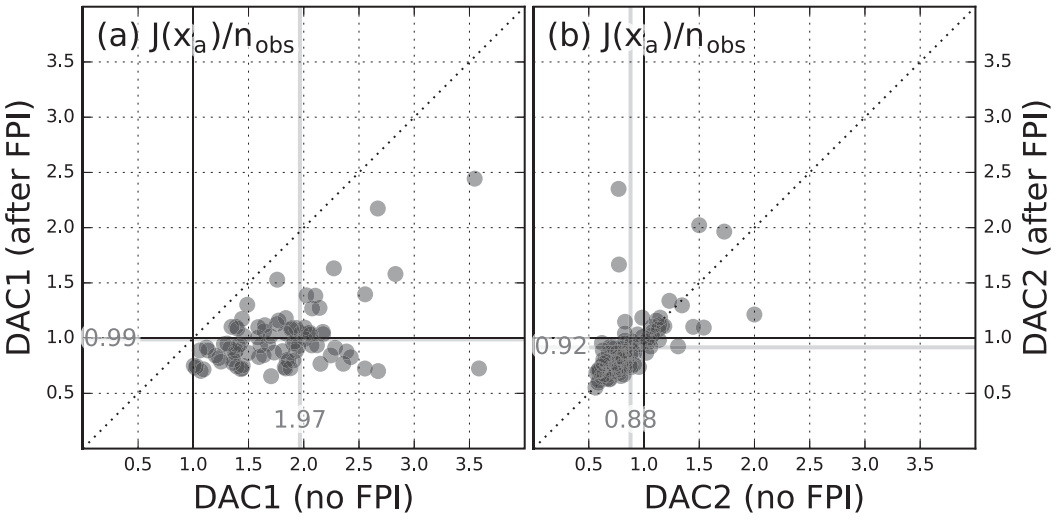

FIG. 4. Scatterplots of $J(\mathbf{x}_a)/n_{\text{obs}}$ for each DA cycle of (a) DAC1 and (b) DAC2; both panels show the results of the DA before (x axis) and after the FPI (y axis) was used to adjust **B** and **R**. Average values are denoted by numbers and marked by solid lines.

error). Thus, e_a by itself is not a useful measure of DA performance, and to complement it, we include the prior RMSE,

$$e_b^{(i,c)} = \sqrt{\frac{1}{|O_{i,c}|}\sum_{j\in O_{i,c}} \mathbf{d}_{bj}^{o^2}}, \tag{11}$$

which is based on unassimilated observations. Because observations are assimilated in consecutive cycles, e_b is a measure of DA performance and can be thought of as short-term forecasting skill. Furthermore, e_b is not subject to overfitting (presuming uncorrelated observation errors) and can, in conjunction with e_a, be used as an overfitting indicator: if a DA system achieves a much lower posterior RMSE without a reduction of the prior RMSE in the next DA cycle, this indicates that the posterior error reduction may be due to overfitting. It should be noted that e_b can also be interpreted as a consistency diagnostic on innovations (Desroziers et al. 2005). In general, our FPI implementation leads to smaller changes in the prior RMSE in comparison to the posterior, yet, overall, the FPI improves the forecasting skill for all observation types except in situ salinity, in situ temperature, and SSH (Figs. 3e,f; again, SSH remains nearly unaffected by the FPI). Interestingly, the prior RMSE for chlorophyll *a* is slightly better with tuning. Thus, the lower posterior RMSE of the untuned system in DAC1 did not translate into an improved forecast, an indication of overfitting that was effectively removed by the FPI. A similar conclusion can be drawn from the results for the in situ observations, which are fit more tightly in the untuned system, resulting in a larger misfit to the satellite observations, which vastly outnumber in situ observations for both temperature and salinity (see Fig. 1). The covariance tuning increases the observation error for in situ temperature and salinity, thereby relaxing the fit to in situ observations but improving it for the satellite data. Another possible factor explaining the discrepancy between in situ and satellite model misfit for temperature and salinity may be that the observation values from the different data sources do not agree well. [The agreement is not easily quantifiable, as in situ observations are mostly subsurface, while satellite observations are from the ocean surface; a possible disagreement between data could perhaps be rectified by bias correction, examined in Waller et al. (2016a), but we did not implement a bias correction in this application.]

c. *FPI sensitivity*

In our FPI experiments above, we used a setup where each iteration requires 10 assimilative simulations, each consisting of two cycles. Even though these simulations can be run in parallel, a considerable computational expense is associated with each iteration. To assess the sensitivity of the FPI to the DA setup and gauge the potential for drastically reducing the computational cost of the FPI, we created a second FPI setup where each iteration consists of 1, rather than 10, simulations, and in which just over 300 000 observations are assimilated (approximately 10% of the 10 × 2 setup; Fig. 1, right column). Because the single simulation still consists of two DA cycles, we refer to this setup as 1 × 2. In the following, we test the 1 × 2 setup in both DA

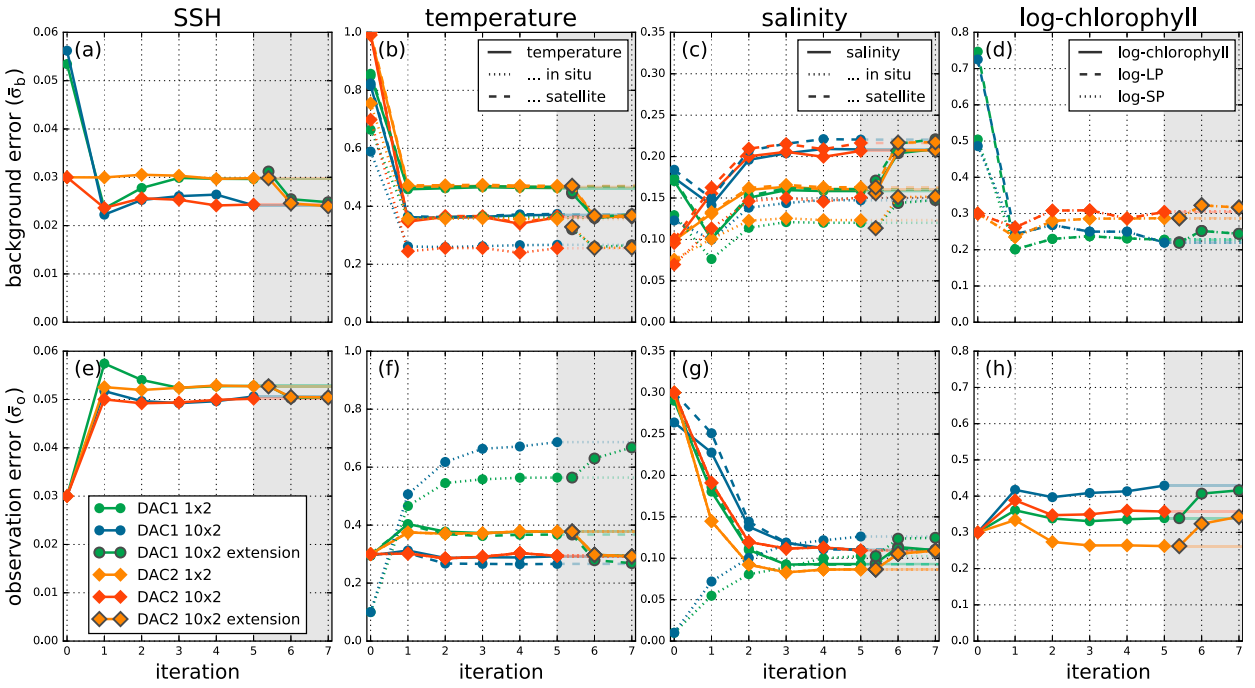

FIG. 5. Convergence of (a)–(d) $\overline{\sigma}_b$ and (e)–(h) $\overline{\sigma}_o$ for DAC1 and DAC2 using both the 1×2 and 10×2 FPI setups (the 10×2 results are copied from Fig. 2). Each column corresponds to an observation type, and in situ and satellite observations are distinguished by line style (for reference, in situ and satellite observation results for DAC2 are included, as well as aggregate temperature and salinity results for DAC1). (d) For log-chlorophyll, we distinguish between $\overline{\sigma}_b$ for LP and SP, which are identical for DAC2. A set of extension experiments (gray-shaded background) uses the final entries of 1×2 FPI as the initial values for a new 10×2 FPI.

configurations, DAC1 and DAC2, and compare the results to the 10×2 setup.

In terms of their convergence characteristics, the two FPI setups produce similar values, yet noticeable differences remain for some variables, especially for salinity background errors and log-chlorophyll observation errors (Fig. 5). For the metrics examined in section 2b, the 10×2 setup results in an improvement in consistency (lower e_σ; not shown) over the 1×2 setup. With respect to the prior and posterior model error (not shown), the two setups exhibit systematic differences for the different observation types but no substantial improvement for either 10×2 or 1×2.

While the 1×2 setup comes at a lower computational cost, it is based on a lower number of observations. The results above indicate that the 1×2 setup provides less representative error covariance diagnostics and may be potentially less suitable for general use in our DA application. More evidence for this conclusion is given when examining the background and observation error multipliers derived from the 92-cycle, yearlong simulations that were used to generate the results in section 2b. That is, we use the **B** and **R** produced by the 1×2 FPI to perform a new 92-cycle DA simulation for 2013 and then compute the multipliers λ_b and λ_o from the input and output of this long simulation. The 1×2 multipliers differ from those based on the 92-cycle simulation for the 10×2 setup (dashed lines in Fig. 6). Specifically, the 1×2 multipliers deviate further from 1.0, strongly indicating that the 1×2 setup is less representative of the 92-cycle simulation than the 10×2 setup. The reason for this is that as the FPI converges, background and observation error multipliers should approach 1.0. Presuming now that we have an FPI setup that is fully representative of the 92-cycle DA simulation, and a new 92-cycle simulation is started using the **B** and **R** obtained from the FPI, the multipliers obtained from this simulation should be nearly identical to 1.0 as well. Deviations from 1.0, thus, indicate nonrepresentativeness.

In light of these results, we created a new 10×2 FPI that uses the final values obtained from the 1×2 FPI in its first iteration. This new FPI setup converges toward the previous 10×2 values in just two iterations (for both DAC1 and DAC2; see extension experiments in Fig. 5). This result suggests that in practical applications, the first iterations of an FPI can be based on a small number of DA cycles, which come at a lower computational cost. After a few iterations in which partial convergence is achieved, the FPI switches to a more representative, yet more expensive, setup that uses a higher number of DA

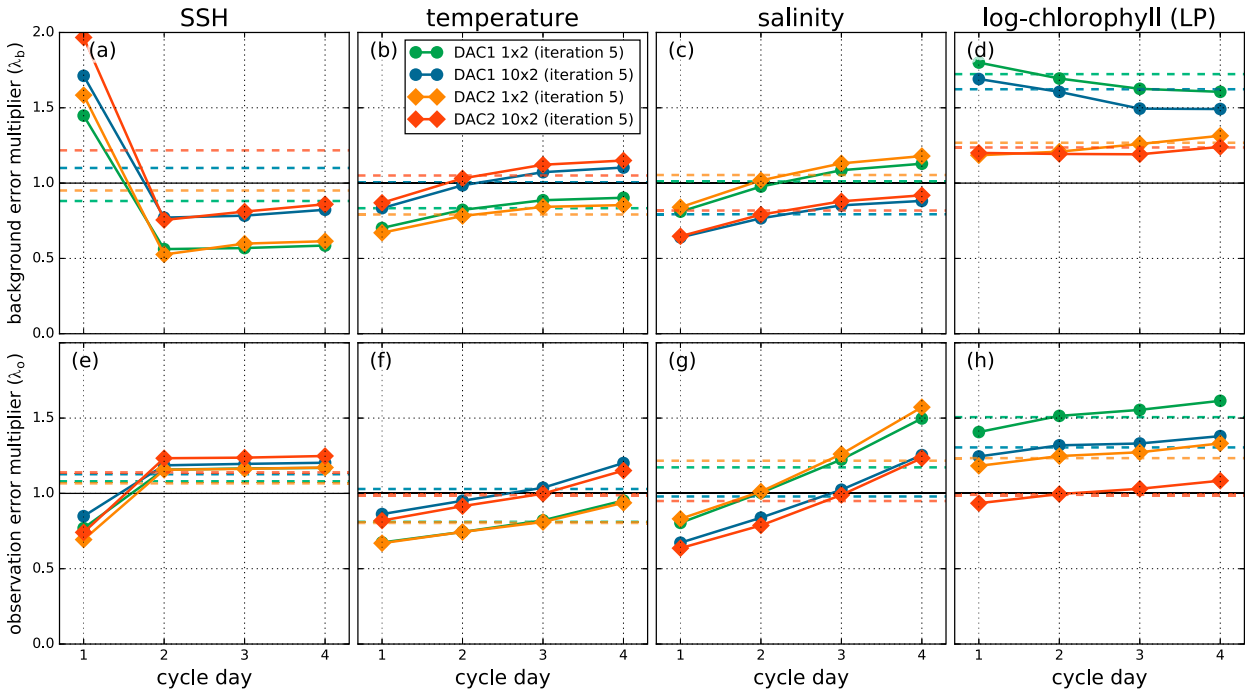

FIG. 6. The effect of basing the FPI multipliers on subsets of observations that correspond to the first, second, third, and fourth days of each assimilation cycle (cycle day, see section 3d). The (a)–(d) background error and (e)–(h) observation error multipliers are shown for (left to right) each observation variable. The average multiplier for each observation type is marked by a dashed horizontal line in each panel. The multipliers are derived from the 92-cycle, yearlong simulations for DAC1 and DAC2 and using both the 10×2 setup (the basis for the results presented in section 3) and the computationally less expensive 1×2 setup, introduced in section 3c.

cycles and more observations for full convergence (see section 4 for further discussion).

d. *Effect of model dynamics on covariance estimates*

While our FPI procedure adjusts the values in **B** directly, the error covariance diagnostics in Eq. (5) are based on $\mathbf{HBH}^{\mathrm{T}}$. In 4D-Var applications such as this, the operator **H** contains the tangent linear model dynamics; hence, it is uncertain if a multiplicative adjustment of **B** translates to an adjustment of similar magnitude to $\mathbf{HBH}^{\mathrm{T}}$ after passing through the model dynamics. To quantify the effect of the tangent linear model dynamics described by **H** on the error covariance diagnostics estimates, we split the observations into subsets based on the time since the start of the assimilation cycle. Specifically, we split all observations used in our 92-cycle DA simulation for 2013 into four distinct observation subsets, one for each day of the assimilation cycle (the first subset contains all observations from the first day of each DA cycle, the second contains observations from each second day, etc.). At the start of each assimilation cycle, the model is initialized, and there is no effect of dynamics; as the cycle progresses and the model runs forward, the effect of model dynamics increases. This effect propagates into $\tilde{\sigma}_b$, which incorporates H and, therefore, accounts for model dynamics. To measure the effect of model dynamics, we compute $\tilde{\sigma}_b$ for each of the four observational subsets in which the impact of model dynamics becomes more pronounced from day 1 to day 4. In our application, the effect of model dynamics on estimates of $\tilde{\sigma}_b$ and, subsequently, the corrective multipliers, remains relatively small (see Fig. 6). The largest change in estimates occurs between days 1 and 2 for SSH for background and also observation error multipliers; yet, as observed in the previous results, the FPI adjustments have little effect on model SSH. When excluding the SSH results from the analysis, the maximum standard deviation between the four background error multiplier estimates is 0.13 (for salinity) and 0.32 for observation error multipliers (again, for salinity). Mean values are, as expected, close to 1.0 for most variables (exceptions like the background error multipliers for log-chlorophyll highlight the benefits of the 10×2 over the 1×2 when estimating the error covariance diagnostics for all of 2013 and also show better error covariance diagnostics estimates for DAC2, compared to DAC1).

e. *Attractiveness of the fixed-point solutions*

In a final set of experiments, we examine the attractiveness of the fixed points in order to investigate if there is a broader domain in which the iteration converges toward the

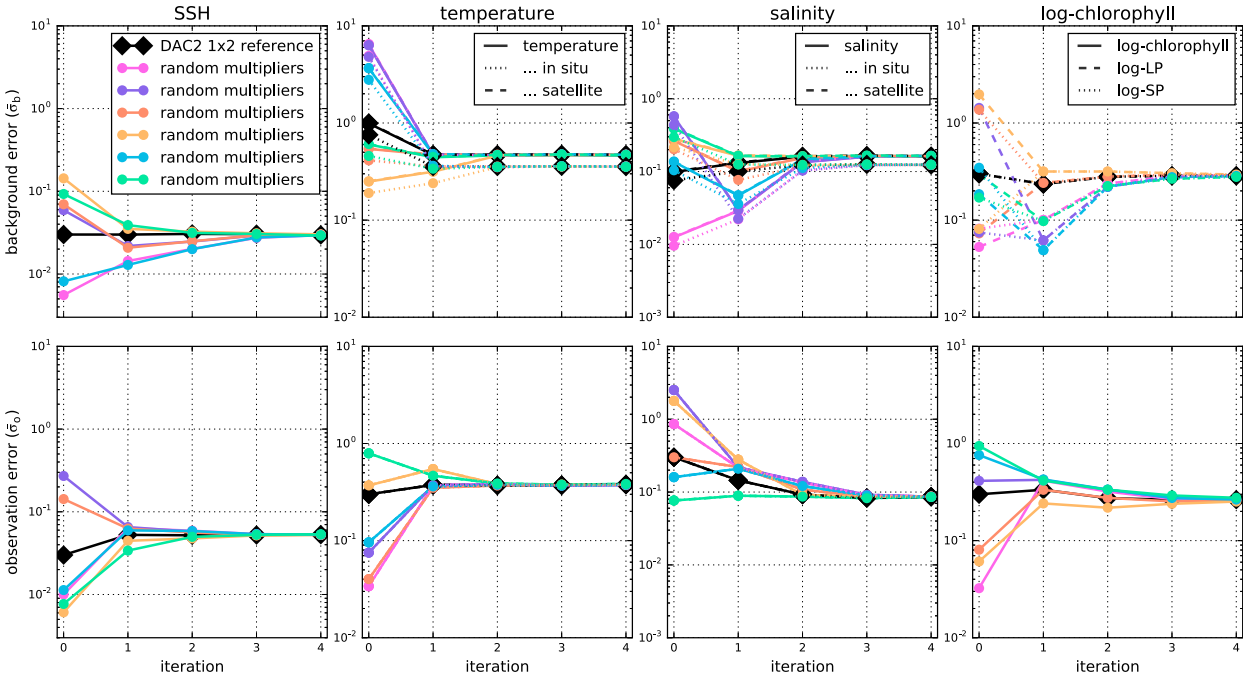

FIG. 7. Convergence of (top) $\overline{\sigma}_b$ and (bottom) $\overline{\sigma}_o$ for six FPIs that were initialized with randomized multipliers, in comparison to the reference DAC2 1 × 2 FPI (black).

same set of values. The complexity of the functions on which our FPIs are based [see Eqs. (5) and (6)], incorporating prior and posterior model solutions, prohibits an analytical examination. Instead, we start several FPIs in which **B** and **R** are modified by random multipliers and investigate their convergence characteristics. Because of the computational expense of the FPIs, we restrict our analysis to performing only six FPIs for DAC2 and the 1 × 2 setup.

At the start of the experiment, **B** and **R** are modified by random multipliers for each observation type. Each multiplier, randomly chosen from the interval of [0.1, 10.0], is squared when applied to **B** and **R** (just like λ_b and λ_o used in the FPI procedure). Using the modified error covariance matrices as initial conditions, we then start a new FPI. In each of our six FPIs, the entries of **B** and **R** converge toward the same values as the reference FPI (not modified by multipliers) within five iterations (Fig. 7). Based on this evidence, the fixed point appears to be attractive, and we consider it unlikely that there are other fixed points within close proximity that may yield more consistent results with respect to the error covariance diagnostics. Yet, by modifying **B** and **R** using multipliers, we do not change the structure of the specified errors. This experiment, thus, does not provide an answer to the question of if the structures of **B** and **R** are specified correctly or if the FPI would converge to similar solutions if their structures change (although the latter question is, to some extent, addressed in our comparison of the convergence of DAC1 and DAC2 in section 3a).

4. Discussion and conclusions

We presented a simple way to objectively adjust the covariance matrices of variational DA systems based on the error covariance diagnostics, which are easy to compute based on properties prescribed to the DA system or that can be obtained from one or more DA cycles. In the way presented here, the modification of covariance matrices is easy to implement and consists of a mere rescaling of variances and associated off-diagonal elements of **B** (up- and downweighting the contribution of individual observation types to the cost function). The result is a more consistent DA system (based on the error covariance diagnostics) with improved prior and posterior model-observation fit. In this first application of error covariance diagnostics to a coupled physical–BGC DA system, the covariance adjustments eliminate signs of overfitting present in the untuned DA system and balance the weighting of physical and chlorophyll *a* observations, which are assimilated jointly. This can be an important issue in BGC DA where increments to the physical variables are weighted against biogeochemical increments, which, in a previous setup of our DA system without error-covariance-diagnostics-based tuning, required manual modifications of the error covariances in order to achieve meaningful improvement in chlorophyll *a* state estimates (Mattern et al. 2017). The error covariance

diagnostics, thus, provide an objective way to determine the weight of each observation type in the assimilation, eliminating subjective manual decisions.

We tested the procedure on two DA configurations that predominantly differ in their prescribed background error structure. DAC1, the first configuration, bases its background error values on the statistics of a long model simulation without DA, while the second configuration, DAC2, uses constant surface variance values with a depth decline, severely reducing the complexity of the prescribed values. The covariance adjustments led to improvements in both configurations, and, after the adjustments, they exhibit similar characteristics in terms of the statistics we examined. While the improvements for DAC1 can, to some extent, be attributed to the downweighting of in situ observations, which showed indications of overfitting prior to tuning, the FPI does more than eliminate somewhat obvious shortcomings in our configurations. In particular, it also improves DAC2 where no distinction is made between in situ and other observations (either in terms of different observation error values or in the tuning procedure itself). Furthermore, the simpler but tuned DAC2 clearly outperforms the untuned DAC1 with respect to the posterior and prior model-observation fit, and it provides very similar results after tuning. This outcome suggests that covariance estimates with a very simple structure, together with covariance tuning, may offer an alternative way to determine values for $\mathbf{B}$ without the requirement of an ensemble or a long model simulation.

FPIs based on error covariance diagnostics may converge toward incorrect error values if the error structure is not correctly modeled. One issue that several studies note is difficulties in estimating background and observation error values jointly (Desroziers et al. 2005; Ménard 2016; Bowler 2017). Under certain conditions, the contributions of the two error types cannot be separated, resulting in incorrect adjustments to $\mathbf{R}$ and $\mathbf{B}$. In some of our initial experiments with the FPI (not shown), we observed only relatively small differences in the background error values when these quantities are estimated by themselves or jointly with observation errors (only conducted for DAC2 where values for $\overline{\sigma}_b$ remained within $\pm 8\%$ of the joint estimates). Another issue arises when correlated observation errors are modeled as uncorrelated errors (Waller et al. 2016c; Ménard 2016; Campbell et al. 2017). In this scenario, the observation error is typically underestimated by the error covariance diagnostics, effectively upweighting observations instead of attributing their impact to spatial correlations. This effect is likely influencing our application, where we assume uncorrelated errors for all observations, including satellite data for temperature, salinity, and chlorophyll *a*. Our FPI implementation does not make adjustments to the error structure and, thus, cannot mitigate this likely shortcoming of our error specification. Despite these limitations of the error covariance diagnostics, studies like Waller et al. (2016b), Campbell et al. (2017), and Cordoba et al. (2017) show that even with incorrectly modeled errors, it is possible to obtain useful results from the error covariance diagnostics unless the errors are severely misspecified (Bormann et al. 2016). Our results confirm these findings: despite the assumption of uncorrelated observation errors, and no matter which of the two background error specifications is used (DAC1 or DAC2), the error-covariance-diagnostics-based FPI yields the improvements in the assimilation results that we noted above.

A limitation of the specific approach presented here is that it makes no adjustments to the background error values for unobserved variables. This is especially an issue in complex BGC ocean models, where typically no observations exist for most of the biological variables. In our application, we set the background error values for unobserved variables to relatively low values, a pragmatic approach to limit DA adjustments to variables, which we cannot objectively assess with the error covariance diagnostics presented here. A possible alternative that we did not explore would be the use of correlations in the state vector to establish correlations for background error values and perform a spatially dependent adjustment of the background error values for unobserved variables [localized, in order to minimize the effect of spurious correlations, similar to the treatment of inflation factors in Anderson (2009)]. Relatedly, a second limitation is that the approach presented here scales the full background error field for each variable, thereby assuming that its underlying spatial structure is correct. Given enough observations to provide reliable statistics, it is easy to allow for structural changes by dividing the observations into more types: for example, by distinguishing between coastal and offshore observations or observations at different times of year, resulting in spatial or temporal structure in the background and observation error values. It is further possible to adjust off-diagonal elements by estimating length scales (see, e.g., Ménard 2016). A third limitation of our approach is that the influence of the tangent linear model dynamics is ignored in the computation of $\mathbf{HBH}^{\mathrm{T}}$. While Bowler (2017) suggests basing the estimation of background errors on observations from the start of a DA cycle, we find that model dynamics had only a minor effect on our results (see section 3d). They may, however, be more important in other applications, especially when long DA cycles are used.

The covariance tuning is based on an FPI, which converges quickly in about five iterations. In our experiments, we observed that the FPI converges to similar, but

not identical, values if it is based on 1 instead of 10 simulations with two cycles each (see section 3c). Our results show that the 20 (nonconsecutive) cycle statistics, which are based on approximately 10 times the number of observations, are more representative and, thus, preferable to the two-cycle statistics. This suggests an FPI setup for practical applications that bases its first iterations on computationally cheap DA simulations, followed by iterations using more expensive simulations. The first iterations are used to achieve an initial convergence toward rough estimates of the error covariance values; to achieve the final convergence, the following iterations use more representative DA simulations consisting of more cycles and incorporating more observations.

Acknowledgments. We thank two anonymous reviewers for their constructive comments. This research was, in part, supported by Grant OCE-1566623 from the National Science Foundation Division of Ocean Sciences. Any opinions, findings, and conclusions or recommendations expressed here are those of the authors and do not necessarily reflect the views of the National Science Foundation. We also gratefully acknowledge the support through Grant NA16NOS0120021 of the National Oceanographic and Atmospheric Administration and the Central and Northern California Ocean Observing System.

APPENDIX A

Diagnostics for Log-Transformed Variables

While the diagnostics for the fixed-point iteration are derived in Desroziers et al. (2005) assuming Gaussian error statistics, it is not obvious that the theory applies more generally to variables with non-Gaussian distributions and error statistics. Here, we reexamine the assumptions of the quadratic, incremental form of lognormal 4D-Var (Song et al. 2016a) used in the present data assimilation system and show that within the constraints of the linearized theory, the diagnostics remain appropriate.

As discussed more extensively in Fletcher and Zupanski (2006) and Song et al. (2016a), the cost function for lognormal variables can be expressed as

$$J_L(\mathbf{x}) = \frac{1}{2}[\log(\mathbf{x}) - \log(\mathbf{x}_b)]^{\mathrm{T}}\mathbf{B}_L^{-1}[\log(\mathbf{x}) - \log(\mathbf{x}_b)] + \frac{1}{2}[\log(\mathbf{y}) - \log(\tilde{\mathbf{x}})]^{\mathrm{T}}\mathbf{R}_L^{-1}[\log(\mathbf{y}) - \log(\tilde{\mathbf{x}})]. \quad \text{(A1)}$$

Background and observation error covariances in log-transformed space are represented by $\mathbf{B}_L$ and $\mathbf{R}_L$, respectively. The nonlinear observation operator H interpolates the state vector both in time and space; thus, model values at the observations can be written as $\tilde{\mathbf{x}} = H(\mathbf{x})$. The desired analysis initial condition $\mathbf{x}_a$ is obtained by finding $\mathbf{x}$ that minimizes Eq. (A1), but nonlinearities introduce minimization challenges.

Nonlinearities in H can be eliminated by developing an incremental form and assuming small deviations to the background. Specifically, $\delta\mathbf{x} = \mathbf{x}_a - \mathbf{x}_b$ represents the nontransformed increment, and the values of the analysis in observation space can be approximated in linearized form

$$\tilde{\mathbf{x}}_a = H(\mathbf{x}_b + \delta\mathbf{x}) \approx H(\mathbf{x}_b) + \mathbf{H}\delta\mathbf{x},$$

where $\mathbf{H}$ represents the tangent linear approximation to H.

Introduction of small increments to Eq. (A1) is not sufficient to render the cost function quadratic, and, therefore, its minimization still requires an iterative procedure. A quadratic form for J_L is obtained through linearization of the natural logarithm function and can be written

$$\log[H(\mathbf{x}_a)] \approx \log[H(\mathbf{x}_b + \mathbf{H}\delta\mathbf{x})] \approx \log[H(\mathbf{x}_b)] + \mathbf{L}\mathbf{H}\delta\mathbf{x},$$

where

$$\mathbf{L} \equiv \left.\frac{\partial\log(\tilde{\mathbf{x}})}{\partial\tilde{\mathbf{x}}}\right|_{\tilde{\mathbf{x}}=H(\mathbf{x}_b)} = \begin{bmatrix} H(\mathbf{x}_b)_1 & 0 & \cdots & 0 \\ 0 & H(\mathbf{x}_b)_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & H(\mathbf{x}_b)_{n_{\mathrm{obs}}} \end{bmatrix}^{-1},$$

and the subscript $1, 2, \ldots, n_{\mathrm{obs}}$ represents the observation index up to a total of n_{obs} observations.

It is convenient now to introduce the log-space increment $\delta\mathbf{g} = \log(\mathbf{x}_a) - \log(\mathbf{x}_b)$, which itself yields

$$\mathbf{x}_a = \mathbf{x}_b \circ \exp(\delta\mathbf{g}),$$

where $\circ$ indicates a Hadamard (element by element) product. For small increments $\delta\mathbf{g}$,

$$\delta\mathbf{x} = \mathbf{x}_b \circ \exp(\delta\mathbf{g}) - \mathbf{x}_b \approx \mathbf{x}_b \circ (1 + \delta\mathbf{g}) - \mathbf{x}_b \approx \mathbf{x}_b \circ \delta\mathbf{g}.$$

Letting $\mathbf{X}_b$ define a diagonal matrix with the elements of $\mathbf{x}_b$ on the diagonal, the logarithm of the analysis in observation space is expressed in terms of log-space increments

$$\log[H(\mathbf{x}_b + \delta\mathbf{x})] \approx \log[H(\mathbf{x}_b)] + \mathbf{L}\mathbf{H}\mathbf{X}_b\delta\mathbf{g} = \log[H(\mathbf{x}_b)] + \mathbf{H}_L\delta\mathbf{g}$$

and it follows that

$$\log[H(\mathbf{x}_t)] \approx \log[H(\mathbf{x}_b)] + \mathbf{H}_L[\log(\mathbf{x}_t) - \log(\mathbf{x}_b)]. \quad \text{(A2)}$$

Thus, $\mathbf{H}_L = \mathbf{LHX}_b$ represents the tangent linear operator that maps log-space increments $\delta\mathbf{g}$ to observation space.

Denoting the log-space deviations between observations and the background at observation points and times as $\mathbf{d}_b^o = \log(\mathbf{y}) - \log[\mathbf{H}(\mathbf{x}_b)]$, Eq. (A1) assumes its quadratic, incremental form:

$$J_L(\delta\mathbf{g}) = \frac{1}{2}\delta\mathbf{g}^{\mathrm{T}}\mathbf{B}_L^{-1} + \frac{1}{2}(\mathbf{d}_b^o - \mathbf{H}_L\delta\mathbf{g})^{\mathrm{T}}\mathbf{R}_L^{-1}(\mathbf{d}_b^o - \mathbf{H}_L\delta\mathbf{g}). \quad \text{(A3)}$$

The similarity between traditional, incremental 4D-Var assuming Gaussian statistics (Courtier et al. 1994) and quadratic, incremental, lognormal 4D-Var is evident. The solution that minimizes $J_L(\delta\mathbf{g})$ is

$$\delta\mathbf{g} = \mathbf{K}_L\mathbf{d}_b^o, \quad \text{where} \quad \mathbf{K}_L = \mathbf{B}_L\mathbf{H}_L^{\mathrm{T}}(\mathbf{H}_L\mathbf{B}_L\mathbf{H}_L^{\mathrm{T}} + \mathbf{R}_L)^{-1}$$

is a gain matrix defined in terms of log-transformed covariance matrices and linearized representations of the log-transformed model.

Following Desroziers et al. (2005), we examine several consistency diagnostics. For example,

$$\begin{aligned}\mathbf{d}_b^o &= \log(\mathbf{y}) - \log[H(\mathbf{x}_b)],\\ &= \log(\mathbf{y}) - \log[H(\mathbf{x}_t)] + \log[H(\mathbf{x}_t)] - \log[H(\mathbf{x}_b)],\\ &= \tilde{\boldsymbol{\varepsilon}}_L + \mathbf{H}_L\boldsymbol{\varepsilon}_L^b.\end{aligned}$$

Here, $\tilde{\boldsymbol{\varepsilon}}_L$ and $\boldsymbol{\varepsilon}_L^b$ represent the observational and background errors of log-transformed variables, respectively, and these errors are assumed to be uncorrelated. Deviations between log-transformed observations and analysis at observation locations and times become

$$\begin{aligned}\mathbf{d}_a^o &= \log(\mathbf{y}) - \log[H(\mathbf{x}_a)] \approx \log(\mathbf{y}) - \{\log[H(\mathbf{x}_b)] + \mathbf{H}_L\delta\mathbf{g}\}\\ &= (\mathbf{I} - \mathbf{H}_L\mathbf{K}_L)\mathbf{d}_b^o.\end{aligned}$$

And differences between the log-transformed analysis and background in observation space are given by

$$\mathbf{d}_b^a = \log[H(\mathbf{x}_a)] - \log[H(\mathbf{x}_b)] \approx \mathbf{H}_L\delta\mathbf{g}.$$

From these definitions, the following expectation values are obtained:

$$\begin{aligned}\mathbb{E}(\mathbf{d}_b^o\mathbf{d}_b^{o\mathrm{T}}) &= \mathbf{R}_L + \mathbf{H}_L\mathbf{B}_L\mathbf{H}_L^{\mathrm{T}},\\ \mathbb{E}(\mathbf{d}_b^a\mathbf{d}_b^{o\mathrm{T}}) &= \mathbf{H}_L\mathbf{B}_L\mathbf{H}_L^{\mathrm{T}},\\ \mathbb{E}(\mathbf{d}_a^o\mathbf{d}_b^{o\mathrm{T}}) &= \mathbf{R}_L,\\ \mathbb{E}(\mathbf{d}_b^a\mathbf{d}_a^{o\mathrm{T}}) &= \mathbf{H}_L\mathbf{A}_L\mathbf{H}_L^{\mathrm{T}},\end{aligned}$$

where $\mathbf{A}_L$ is the analysis error covariance in log space. These expressions are analogous to those derived by Desroziers et al. (2005), with the appropriate changes in covariance matrices and linearizations to the log-transformed model.

APPENDIX B

Treatment of Log-Chlorophyll

In our 4D-Var DA system, we presume that both the large phytoplankton and small phytoplankton variables contribute toward model chlorophyll using fixed but different carbon-to-chlorophyll *a* ratios. That is, we presume

$$\mathbf{x}^{\mathrm{chl}} = \alpha_{\mathrm{LP}}\mathbf{x}^{\mathrm{LP}} + \alpha_{\mathrm{SP}}\mathbf{x}^{\mathrm{SP}}, \quad \text{(B1)}$$

where α_{LP} and α_{SP} are the carbon-to-chlorophyll *a* ratios for LP and SP, respectively. We further assume that chlorophyll *a* is approximately lognormally distributed (see, e.g., Song et al. 2012, 2016a). This assumption leads to an issue, as we need to assign a distribution to both LP and SP, and, based on Eq. (B1), the two variables have neither normal nor lognormal distributions. Yet, in practice, based on an analysis of the distributions of all variables in the model output, none of the variables has an exact normal or lognormal distribution. Hence, as for all other variables, we assign the distribution that better approximates the LP and SP distributions observed in the model: for both variables, this is the lognormal distribution.

The lognormal assumption poses an obstacle when computing the error covariance diagnostics (see section 2c) for chlorophyll *a*. The properties $\mathbf{d}_a^o$, $\mathbf{d}_b^o$, and $\mathbf{d}_b^a$, on which the statistics in Desroziers et al. (2005) are based, now require the computation of differences in log space. For example, for lognormal variables, $\mathbf{d}_b^o$ becomes

$$\mathbf{d}_b^o = \log(\mathbf{y}) - \log[H(\mathbf{x}_b)], \quad \text{(B2)}$$

where here, and in the following, the log transformation is performed elementwise when applied to a vector. The derivation of the relationships among $\tilde{\sigma}_b$, $\overline{\sigma}_b$, $\tilde{\sigma}_o$, and $\overline{\sigma}_o$ [see Eq. (7)] in Desroziers et al. (2005) requires that the tangent linear assumption holds. For lognormal variables, this entails

$$\log[H(\mathbf{x}_b)] - \log[H(\mathbf{x}_t)] \approx \mathbf{H}_L[\log(\mathbf{x}_b) - \log(\mathbf{x}_t)], \quad \text{(B3)}$$

where $\mathbf{x}_t$ is the presumed true model state, and $\mathbf{H}_L$ is the linearized observation operator specific to the 4D-Var implementation with log transformation [see appendix A

and Eq. (A2) therein for a derivation and section 2.3 in Song et al. (2016a) for further details]. While Eq. (B3), thus, holds approximately true for the lognormal variables, there is a problem when including the weighting with the carbon-to-chlorophyll *a* ratios α_{LP} and α_{SP} and subsequent aggregation of LP and SP to chlorophyll *a* as performed in Eq. (B1) into the observation operators. In the following, we consider a simple (low dimensional) example to showcase the log-chlorophyll problem and the solution we use in our FPI implementation. The example can easily be extended to the general, higher-dimensional case with more variables. Given a single chlorophyll *a* observation $\mathbf{y}^{\mathrm{chl}}$ and $\tilde{\mathbf{x}} = (\tilde{x}^{\mathrm{LP}}, \tilde{x}^{\mathrm{SP}})^{\mathrm{T}} = H(\mathbf{x})$ the model solution for LP and SP at the observation location, the weighting with the carbon-to-chlorophyll *a* ratios, and aggregation can be included in the matrix $\mathbf{A} = (\alpha_{\mathrm{LP}}, \alpha_{\mathrm{SP}})$, so that $\tilde{\mathbf{x}}^{\mathrm{chl}} = \mathbf{A}\tilde{\mathbf{x}}$. Because of the log transformation,

$$\log[\mathbf{A}H(\mathbf{x}_b)] - \log[\mathbf{A}H(\mathbf{x}_t)] \neq \mathbf{A}\mathbf{H}_L[\log(\mathbf{x}_b) - \log(\mathbf{x}_t)],$$

and, thus, $\mathbf{A}$ cannot be easily included in the computation of $\mathbf{d}_a^o$, $\mathbf{d}_b^o$, and $\mathbf{d}_b^a$ in this example or the general case when $\mathbf{A}$ contains multiple instances of weighting and aggregation of LP and SP to chlorophyll *a*.

A solution to the log-transformation issue is, thus, to avoid the weighting and aggregation of LP and SP to chlorophyll *a*. Instead, chlorophyll observations are split into LP and SP, and the comparison is performed for both variables without further transformation so that Eq. (B3) holds. Critically, our approach does not require the knowledge of the LP-to-SP ratio (LP:SP) in the observations (it is typically not known) but requires the assumption that LP:SP is identical in $\mathbf{y}$, $H(\mathbf{x}_a)$, and $H(\mathbf{x}_b)$ at each chlorophyll *a* observation location (it does not need to be identical at different locations). In our simple example, $\mathbf{y} = (y^{\mathrm{LP}}, y^{\mathrm{SP}})^{\mathrm{T}}$ and $\tilde{\mathbf{x}} = (\tilde{x}^{\mathrm{LP}}, \tilde{x}^{\mathrm{SP}})^{\mathrm{T}} = H(\mathbf{x})$. To compute $\mathbf{d}_a^o$, $\mathbf{d}_b^o$, and $\mathbf{d}_b^a$, we then use the identical LP:SP assumption to define two constants, β and γ, that convert chlorophyll *a* to LP and SP, respectively, and for $\mathbf{y}$, $H(\mathbf{x}_a)$, and $H(\mathbf{x}_b)$. It follows

$$\begin{aligned}\mathbf{d}_b^o &= \log(\mathbf{y}) - \log(H\mathbf{x}_b) = \log\begin{pmatrix} y^{\mathrm{LP}} \\ y^{\mathrm{SP}} \end{pmatrix} - \log\begin{pmatrix} \tilde{x}_b^{\mathrm{LP}} \\ \tilde{x}_b^{\mathrm{SP}} \end{pmatrix} \\ &= \log\begin{pmatrix} \beta\, y^{\mathrm{chl}} \\ \gamma\, y^{\mathrm{chl}} \end{pmatrix} - \log\begin{pmatrix} \beta\, \tilde{x}_b^{\mathrm{chl}} \\ \gamma\, \tilde{x}_b^{\mathrm{chl}} \end{pmatrix} \\ &= \log\begin{pmatrix} y^{\mathrm{chl}} \\ y^{\mathrm{chl}} \end{pmatrix} - \log\begin{pmatrix} \tilde{x}_b^{\mathrm{chl}} \\ \tilde{x}_b^{\mathrm{chl}} \end{pmatrix}.\end{aligned}$$

Because both $\mathbf{y}^{\mathrm{chl}}$ and $\tilde{\mathbf{x}}_b^{\mathrm{chl}}$ are known, the result is easy to obtain and does not depend on the value for β or γ and, thus, the actual value of LP:SP. The properties $\mathbf{d}_a^o$ and $\mathbf{d}_b^a$ are computed analogously to $\mathbf{d}_b^o$ using the same assumption.

A major effect assuming identical LP:SP is that the values of $\overline{\sigma}_b$ for LP and SP in the FPI converge toward each other; consequently, the background error entries typically (depending on their spatial structure and observation locations) converge or remain identical, as in the case of DAC2. As a result, the corrective increment applied by the DA system does not significantly alter the LP:SP ratio from that of the prior model state. As no information of LP:SP is contained in the satellite chlorophyll *a* data, this could be considered a positive side effect.

REFERENCES

Anderson, J. L., 2009: Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus*, **61A**, 72–83, https://doi.org/10.1111/j.1600-0870.2008.00361.x.

Bölöni, G., and K. Horvath, 2010: Diagnosis and tuning of background error statistics in a variational data assimilation system. *Quart. J. Hung. Meteor. Serv.*, **114**, 1–19.

Bormann, N., A. Collard, and P. Bauer, 2010: Estimates of spatial and interchannel observation-error characteristics for current sounder radiances for numerical weather prediction. II: Application to AIRS and IASI data. *Quart. J. Roy. Meteor. Soc.*, **136**, 1051–1063, https://doi.org/10.1002/qj.615.

——, M. Bonavita, R. Dragani, R. Eresmaa, M. Matricardi, and A. McNally, 2016: Enhancing the impact of IASI observations through an updated observation-error covariance matrix. *Quart. J. Roy. Meteor. Soc.*, **142**, 1767–1780, https://doi.org/10.1002/qj.2774.

Bowler, N. E., 2017: On the diagnosis of model error statistics using weak-constraint data assimilation. *Quart. J. Roy. Meteor. Soc.*, **143**, 1916–1928, https://doi.org/10.1002/qj.3051.

Campbell, J. W., 1995: The lognormal distribution as a model for bio-optical variability in the sea. *J. Geophys. Res.*, **100**, 13 237–13 254, https://doi.org/10.1029/95JC00458.

Campbell, W. F., E. A. Satterfield, B. Ruston, and N. L. Baker, 2017: Accounting for correlated observation error in a dual-formulation 4D variational data assimilation system. *Mon. Wea. Rev.*, **145**, 1019–1032, https://doi.org/10.1175/MWR-D-16-0240.1.

Cordoba, M., S. L. Dance, G. A. Kelly, N. K. Nichols, and J. A. Waller, 2017: Diagnosing atmospheric motion vector observation errors for an operational high-resolution data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **143**, 333–341, https://doi.org/10.1002/qj.2925.

Courtier, P., J.-N. Thépaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. *Quart. J. Roy. Meteor. Soc.*, **120**, 1367–1387, https://doi.org/10.1002/qj.49712051912.

Daescu, D. N., and R. Todling, 2010: Adjoint sensitivity of the model forecast to data assimilation system error covariance parameters. *Quart. J. Roy. Meteor. Soc.*, **136**, 2000–2012, https://doi.org/10.1002/qj.693.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, https://doi.org/10.1002/qj.828.

Desroziers, G., and S. Ivanov, 2001: Diagnosis and adaptive tuning of observation-error parameters in a variational assimilation. *Quart. J. Roy. Meteor. Soc.*, **127**, 1433–1452, https://doi.org/10.1002/qj.49712757417.

——, L. Berre, B. Chapnik, and P. Poli, 2005: Diagnosis of observation, background and analysis-error statistics in observation space. *Quart. J. Roy. Meteor. Soc.*, **131**, 3385–3396, https://doi.org/10.1256/qj.05.108.

Doyle, J. D., Q. Jiang, Y. Chao, and J. Farrara, 2009: High-resolution real-time modeling of the marine atmospheric boundary layer in support of the AOSN-II field campaign. *Deep-Sea Res. II*, **56**, 87–99, https://doi.org/10.1016/j.dsr2.2008.08.009.

Edwards, C. A., A. M. Moore, I. Hoteit, and B. D. Cornuelle, 2015: Regional ocean data assimilation. *Annu. Rev. Mar. Sci.*, **7**, 21–42, https://doi.org/10.1146/annurev-marine-010814-015821.

Fletcher, S. J., and M. Zupanski, 2006: A data assimilation method for log-normally distributed observational errors. *Quart. J. Roy. Meteor. Soc.*, **132**, 2505–2519, https://doi.org/10.1256/qj.05.222.

Haidvogel, D. B., and Coauthors, 2008: Ocean forecasting in terrain-following coordinates: Formulation and skill assessment of the Regional Ocean Modeling System. *J. Comput. Phys.*, **227**, 3595–3624, https://doi.org/10.1016/j.jcp.2007.06.016.

Howes, K. E., A. M. Fowler, and A. S. Lawless, 2017: Accounting for model error in strong-constraint 4D-Var data assimilation. *Quart. J. Roy. Meteor. Soc.*, **143**, 1227–1240, https://doi.org/10.1002/qj.2996.

Karspeck, A. R., 2016: An ensemble approach for the estimation of observational error illustrated for a nominal 1° global ocean model. *Mon. Wea. Rev.*, **144**, 1713–1728, https://doi.org/10.1175/MWR-D-14-00336.1.

Kishi, M. J., and Coauthors, 2007: NEMURO—A lower trophic level model for the North Pacific marine ecosystem. *Ecol. Modell.*, **202**, 12–25, https://doi.org/10.1016/j.ecolmodel.2006.08.021.

Li, H., E. Kalnay, and T. Miyoshi, 2009: Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter. *Quart. J. Roy. Meteor. Soc.*, **135**, 523–533, https://doi.org/10.1002/qj.371.

Mattern, J. P., and C. A. Edwards, 2017: Simple parameter estimation for complex models—Testing evolutionary techniques on 3-dimensional biogeochemical ocean models. *J. Mar. Syst.*, **165**, 139–152, https://doi.org/10.1016/j.jmarsys.2016.10.012.

——, H. Song, C. A. Edwards, A. M. Moore, and J. Fiechter, 2017: Data assimilation of physical and chlorophyll *a* observations in the California Current System using two biogeochemical models. *Ocean Modell.*, **109**, 55–71, https://doi.org/10.1016/j.ocemod.2016.12.002.

Ménard, R., 2016: Error covariance estimation methods based on analysis residuals: Theoretical foundation and convergence properties derived from simplified observation networks. *Quart. J. Roy. Meteor. Soc.*, **142**, 257–273, https://doi.org/10.1002/qj.2650.

Neveu, E., A. M. Moore, C. A. Edwards, J. Fiechter, P. Drake, W. J. Crawford, M. G. Jacox, and E. Nuss, 2016: An historical analysis of the California Current circulation using ROMS 4D-Var: System configuration and diagnostics. *Ocean Modell.*, **99**, 133–151, https://doi.org/10.1016/j.ocemod.2015.11.012.

Raghukumar, K., C. A. Edwards, N. L. Goebel, G. Broquet, M. Veneziani, A. M. Moore, and J. P. Zehr, 2015: Impact of assimilating physical oceanographic data on modeled ecosystem dynamics in the California Current System. *Prog. Oceanogr.*, **138**, 546–558, https://doi.org/10.1016/j.pocean.2015.01.004.

Song, H., C. A. Edwards, A. M. Moore, and J. Fiechter, 2012: Incremental four-dimensional variational data assimilation of positive-definite oceanic variables using a logarithm transformation. *Ocean Modell.*, **54–55**, 1–17, https://doi.org/10.1016/j.ocemod.2012.06.001.

——, ——, ——, and ——, 2016a: Data assimilation in a coupled physical–biogeochemical model of the California Current System using an incremental lognormal 4-dimensional variational approach: Part 1—Model formulation and biological data assimilation twin experiments. *Ocean Modell.*, **106**, 131–145, https://doi.org/10.1016/j.ocemod.2016.04.001.

——, ——, ——, and ——, 2016b: Data assimilation in a coupled physical–biogeochemical model of the California Current System using an incremental lognormal 4-dimensional variational approach: Part 3—Assimilation in a realistic context using satellite and in situ observations. *Ocean Modell.*, **106**, 159–172, https://doi.org/10.1016/j.ocemod.2016.06.005.

Stammer, D., M. Balmaseda, P. Heimbach, A. Köhl, and A. Weaver, 2016: Ocean data assimilation in support of climate applications: Status and perspectives. *Annu. Rev. Mar. Sci.*, **8**, 491–518, https://doi.org/10.1146/annurev-marine-122414-034113.

Stewart, L. M., S. L. Dance, N. K. Nichols, J. R. Eyre, and J. Cameron, 2014: Estimating interchannel observation-error correlations for IASI radiance data in the Met Office system. *Quart. J. Roy. Meteor. Soc.*, **140**, 1236–1244, https://doi.org/10.1002/qj.2211.

Talagrand, O., 1999: A posteriori evaluation and verification of analysis and assimilation algorithms. *Workshop on Diagnosis of Data Assimilation Systems*, Reading, United Kingdom, ECMWF, 17–28, https://www.ecmwf.int/sites/default/files/elibrary/1999/12547-posteriori-evaluation-and-verification-analysis-and-assimilation-algorithms.pdf.

Veneziani, M., C. A. Edwards, J. D. Doyle, and D. Foley, 2009: A central California coastal ocean modeling study: 1. Forward model and the influence of realistic versus climatological forcing. *J. Geophys. Res.*, **114**, C04015, https://doi.org/10.1029/2008JC004774.

Waller, J. A., S. Ballard, S. Dance, G. Kelly, N. K. Nichols, and D. Simonin, 2016a: Diagnosing horizontal and inter-channel observation error correlations for SEVIRI observations using observation-minus-background and observation-minus-analysis statistics. *Remote Sens.*, **8**, 581, https://doi.org/10.3390/rs8070581.

——, S. L. Dance, and N. K. Nichols, 2016b: Theoretical insight into diagnosing observation error correlations using observation-minus-background and observation-minus-analysis statistics. *Quart. J. Roy. Meteor. Soc.*, **142**, 418–431, https://doi.org/10.1002/qj.2661.

——, D. Simonin, S. L. Dance, N. K. Nichols, and S. P. Ballard, 2016c: Diagnosing observation error correlations for Doppler radar radial winds in the Met Office UKV model using observation-minus-background and observation-minus-analysis statistics. *Mon. Wea. Rev.*, **144**, 3533–3551, https://doi.org/10.1175/MWR-D-15-0340.1.

Yang, C., S. Masina, and A. Storto, 2017: Historical ocean reanalyses (1900–2010) using different data assimilation strategies. *Quart. J. Roy. Meteor. Soc.*, **143**, 479–493, https://doi.org/10.1002/qj.2936.