

An Ensemble Approach for the Estimation of Observational Error Illustrated for a Nominal 1° Global Ocean Model

ALICIA R. KARSPECK

National Center for Atmospheric Research, Boulder, Colorado*

(Manuscript received 15 October 2014, in final form 17 December 2015)

ABSTRACT

Least squares algorithms for data assimilation require estimates of both background error covariances and observational error covariances. The specification of these errors is an essential part of designing an assimilation system; the relative sizes of these uncertainties determine the extent to which the state variables are drawn toward the observational information. Observational error covariances are typically computed as the sum of measurement/instrumental errors and “representativeness error.” In a coarse-resolution ocean general circulation model the errors of representation are the dominant contribution to observational error covariance over large portions of the globe, and the size of these errors will vary by the type of observation and the geographic region. They may also vary from model to model. A straightforward approach for estimating model-dependent, spatially varying observational error variances that are suitable for least squares ocean data assimilating systems is presented here. The author proposes an ensemble-based estimator of the true observational error variance and outlines the assumptions necessary for the estimator to be unbiased. The author also presents the variance (or uncertainty) associated with the estimator under certain conditions. The analytic expressions for the expected value and variance of the estimator are validated with a simple autoregressive model and illustrated for the nominal 1° resolution POP2 global ocean general circulation model.

1. Introduction

Given observations of a geophysical system and a numerical model for simulating that system, data-assimilation methods are designed to produce (in most cases approximate) estimates of the distribution of the model state variables conditional upon observations of the system. This distribution is called the “analysis” in least squares parlance or the “posterior” in Bayesian frameworks. For least squares data-assimilation methods (e.g., variational methods, optimal interpolation, Kalman filtering) the error covariance associated with the “background” or “prior” distribution of the model state variables and the error variance associated with the observations are both required by the algorithms. Indeed, their accurate specification is a necessary condition for the optimality of least squares methods. Even in

real-world modeling applications, wherein error covariances can never be precisely represented and optimality will never be achieved, reasonable specification of the error covariances still plays a critical role in designing a useful assimilation system (Dee 1995).

In addition to errors associated with the measurement process (including instrumental errors), observational errors also include “representativeness errors.” Representativeness error is commonly understood to account for physical processes and scales that are detectable through observation but are not resolvable by the numerical model (e.g., Cohn 1997; Fukumori et al. 1999; Oke and Sakov 2008). This category of errors can also be extended to include errors arising from the process of mapping from the discrete state space of the model to the observed field (sometimes called forward interpolation errors; e.g., Daley 1993). The statistical characteristics of representativeness errors will be model dependent, variable dependent, and geographically inhomogeneous.

The intent of this work is to present a simple framework for estimating observational error variances that include both measurement error and representativeness error. The approach introduced here was motivated by

*The National Center for Atmospheric Research is sponsored by the National Science Foundation.

Corresponding author address: Alicia R. Karspeck, NCAR, P.O. Box 3000, Boulder, CO 80302.
E-mail: aliciak@ucar.edu

design requirements for the assimilation of in situ ocean hydrographic data into a coarse-resolution global ocean model (i.e., horizontal resolution of order 100 km). Coarse-resolution global ocean models do not resolve sharp density fronts or oceanic mesoscale eddies. Thus, in frontal zones or regions where mesoscale variability is active, representativeness error can be the dominant contributor to the total observational error.¹ Our a priori understanding of this motivated the need for a method that admits spatial inhomogeneity.

In this work we are only concerned with the estimation of observational error variance, not the full covariance. While it is possible to extend this framework to the estimation of the time and space covariances we do not cover this here for two reasons: (i) over most of the ocean in situ data are not sufficient to estimate spatial or temporal covariances without assuming parameterized forms (Menemenlis and Chechelnitsky 2000) and (ii) it is common practice in global ocean assimilation to treat observational errors as if they are spatially and temporally uncorrelated (e.g., Wunsch and Heimbach 2007; Karspeck et al. 2013; Derber and Rosati 1989; Behringer et al. 1998; Giese and Ray 2011; Balmaseda et al. 2013; Zhang et al. 2007). This practice, while difficult to justify when observational information is dense, is not unreasonable for hydrographic observations, wherein observations are only occasionally in close physical proximity. At any rate, whether or not it is reasonable, it is also a choice that is routinely made for algorithmic convenience.

The literature on representativeness error and the variety of ways that it can be estimated is vast, and an exhaustive review is not presented here. It is worth mentioning, however, that a great deal of the seminal work on estimating observational error emerged in the context of atmospheric data assimilation (e.g., Rutherford 1976; Lönnberg and Hollingsworth 1986; Daley 1993; Dee and DaSilva 1999; Desroziers and Ivanov 2001; Desroziers et al. 2005). Newer techniques involve the hierarchical Bayesian estimation of observational error (e.g., Li et al. 2009) and the estimation and treatment of correlated and time-varying observational errors (e.g., Waller et al. 2014a,b; Miyoshi et al. 2013; Janjić and Cohn 2006).

In the context of global-scale ocean data assimilation, the observation-based methods of Oke and Sakov

(2008) and Forget and Wunsch (2007) make the assumption that representativeness error primarily stems from the limited resolution of model grids. From this perspective, the observational error variance can be diagnosed by analyzing the differences between observations and their area (Oke and Sakov 2008) or area and time (Forget and Wunsch 2007) averages. The downside of these methods is that they are tailored to the assimilation models for which they will be used only to the extent that averaging domains can be chosen for consistency with a model grid; no consideration of whether the model actually resolves the true large-scale dynamics is made.

Another approach to estimating observational error characteristics is to use *observation-minus-forecast* residuals from the successive cycling of a data-assimilation update and a short-term forecast. These residuals contain a combination of model-resolvable forecast error and unresolvable observation error. Disentangling the two requires that the time and/or geographic length scales of the error sources (and, hence, their covariance characteristics) can be assumed to be sufficiently different as to be robustly and mutually identifiable and that there is sufficient data to make a useful joint estimation. These challenges are well appreciated and have been highlighted in both the oceanographic and atmospheric data-assimilation literature (e.g., Rutherford 1972; Hollingsworth and Lönnberg 1986; Daley 1993; Dee 1995; Dee and DaSilva 1999). For example, in the context of ocean data assimilation, Richman et al. (2005) take the approach of defining observational errors as the component of the observation-minus-forecast residuals in a forced global ocean assimilation that is globally orthogonal to the model's leading spatial eigenvectors—essentially imposing identifiability.

Instead of using observation-minus-forecast residuals, in this work we will adopt the approach of using *observation-minus-simulation* residuals from an ocean model forced with an observationally based estimate of the atmosphere. Menemenlis and Chechelnitsky (2000) and Fu et al. (1993) both take this approach when estimating observation and forecast error covariances for altimeter data. Model simulations are simply forecasts that have been extended long enough that boundary forcing and the internal variability dominate over initial condition information. This strategy allows for the estimation of observational error variance before any assimilation has been performed. (A discussion of the relative advantages of using observation-minus-simulation and observation-minus-forecast residuals in the context of the estimator that will be described in the body of this paper is included in the final section.) Note that the fundamental problem of needing to assume identifiability of the simulation and observational error

¹ For example, the results of Oke and Sakov (2008), Richman et al. (2005), and Forget and Wunsch (2007) support the ranges of 0.5°–3.0°C for the standard deviation of observed temperature error in the Gulf Stream for nominal 1° grids, while measurement error is estimated at 0.005°C (for Argo profiling floats) to 0.1°C [for expendable bathythermographs (XBTs); Johnson et al. (2009)].

covariance structures is still present when using observation-minus-simulation residuals. Both Menemenlis and Chechelnitsky (2000) and Fu et al. (1993) make some progress by choosing highly parameterized forms for their simulation error covariances, but Menemenlis and Chechelnitsky (2000) point out that for the ocean there is rarely sufficient data to infer the forecast or simulation error covariance.

A notable aspect of the forced-ocean residual approaches cited above is that they are based on a single realization of the model state that is meant to represent the expected value of the forecast or simulation distribution. However, the atmospheric state used to force an ocean model is, in actuality, uncertain. This uncertainty will naturally lead to uncertainty in the ocean simulation (e.g., Leeuwenburgh 2007). Typically, the net impact of forcing uncertainty on the ocean state is treated as “process” error (also referred to as “system” or “model” error) and must be inferred as part of the joint forecast and observational error estimation problem (Menemenlis and Chechelnitsky 2000; Fu et al. 1993). However, with the rise in popularity of ensemble data-assimilation systems, ensembles of possible atmospheric states that can be used for forcing ocean models are now available (e.g., Compo et al. 2011; Raeder et al. 2012; Karspeck et al. 2013; Yang and Giese 2013).

The problem of needing to jointly infer both the observational error characteristics and the error characteristics of the simulation from observation-minus-simulation residuals can be mitigated by treating an ensemble of ocean simulations as samples from the simulation distribution. With this additional information about the simulation distribution, it is not necessary to assume or infer a preferred structure of the simulation error in order to separate the observational error and simulation errors; the contribution of simulation error to the observation-minus-simulation residuals can simply be computed directly from the ensemble. That being said, care must be taken to construct a meaningful simulation ensemble and in section 2 we discuss process error in this context and outline some general guidelines for how to construct a suitable ensemble.

Section 3 presents the proposed estimator for observational error variance that is based on observation-minus-simulation residuals from an ensemble of forced-ocean simulations. In the appendix the theoretical mean and variance of the estimator are derived under an explicit set of assumptions and these are referenced in this section. We present the variance of the estimator (not to be confused with the estimator of the observational error variance) to develop an understanding of the factors that limit the usefulness of directly equating the estimation value with the true

observational error variance. In particular, it is shown that the variance of the estimator depends on the variance and autocorrelation of the model simulations, the number of ensembles used for estimation, the number of observations available for computing the residuals, and the size of the true, underlying observational error variance. In section 4, we use Monte Carlo experiments to validate the analytical equations for the properties of the estimator. In section 5, we apply the technique to a state-of-the-art, non-eddy-resolving global ocean general circulation model, showing maps of the estimator. A discussion of the technique in the context of other approaches in the literature is presented in section 6.

2. The modeling framework

a. Ensemble of forced-ocean simulations

Consider a deterministic numerical modeling framework for the physics and dynamics of the ocean where the discrete multivariate model state vector \mathbf{z}^t is connected through time t using

$$\mathbf{z}^{t+\Delta t} = M\mathbf{z}^t + \mathbf{f}^t, \quad (1)$$

where \mathbf{f} can be interpreted as the net forcing resulting from time-varying surface fluxes of heat, fresh water, and momentum. Let us assume that we have access to a set of possible realizations of atmospheric surface variables from which k realizations of surface forcing can be derived and used to force a set of k ocean simulations. If M is linear and the net ocean forcing is an independent realization of a normally distributed process,² then each of the k ocean state ensembles can be viewed as an identically distributed random variable $\mathbf{Z} = \mathbf{z}$ with mean μ and covariance Λ ; that is,

$$[\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3, \dots, \mathbf{Z}_k] \sim N(\mu, \Lambda), \quad (2)$$

where each \mathbf{Z}_j , $j = 1, 2, 3, \dots, k$ exists in the discrete grid space and time intervals of the numerical model. Note that in the absence of a time superscript \mathbf{Z} is meant to signify a vector defined over the discrete space and time of the numerical model. The covariance Λ will reflect the time and space autocorrelation and non-stationarity that emerge naturally through the model dynamics and the atmospheric forcing. In section 3a we will also add the assumption that M is an “unbiased” model, but in practice the requirement is just that if

² Neither of these things will be strictly true for most state-of-the-art global ocean models, but it is a useful approximation at coarse resolutions.

there is a stationary bias relative to observations it can be estimated and removed.

In this context, variations in \mathbf{f} are the sole source of process error in the system. In its broadest definition, process error may also include the effects of model parameter uncertainty and numerical imprecision (e.g., Fu et al. 1993; Blanchet et al. 1997). However, it is important to note that there is no unique definition of what constitutes process error; it must be determined subjectively through consideration of how the model is intended to be used. For example, if the intention of the practitioner is to use a version of the model with variable parameters, then the additional process error that emerges as a result of parameter uncertainty should be included in the generation of the ensemble. On the other hand, if static parameters are to be used (as is most common in practice) then errors emerging as a result of parameter misspecification will fall into the category of either bias or unresolvable processes. Because we are focused here on the use of a global ocean model with static parameters, static topography, compiled code with fixed precision, etc., it is appropriate to assume that forcing variability defines the error subspace of the model. To the extent that it is desirable in other applications to consider additional forms of process error [parameter uncertainty likely being the next most important for non-eddy-resolving ocean general circulation models; P. Gent (2015, personal communication)] the ensemble of ocean states can be constructed to contain these uncertainties through manipulation of the numerical model M . As a general rule, if the practitioner generates an ensemble that underestimates (overestimates) the process error, then the estimate of the observational error variance will tend to be too high (too low).

b. The treatment of unresolved processes as observational representativeness error

Before progressing, we consider why errors of representation can be lumped together with measurement errors under the general term observation error. The real ocean, in contrast to the modeled ocean state, exists in the continuous time and space that we associate with reality. It includes scales and physical processes that are unresolvable by the model and/or impossible to accurately map from model state variables. Imagine that the real ocean state can be expressed as the sum of a random variable \mathbf{Z} , which is the model-resolvable ocean state and a random variable $\tilde{\mathbf{Z}}$ that represents all remaining unresolvable processes. Note that $\tilde{\mathbf{Z}}$ is not synonymous with the process error discussed in the previous subsection; on the contrary, the process error (and its manifestation as simulation error through integration of

the model) defines the boundaries of what is resolvable by the model.

Naturally, observations are noisy records of the real ocean and here we make the standard assumption that they are normally distributed conditional on knowledge of the real ocean state; that is,

$$p(\mathbf{y} | \mathbf{z}, \tilde{\mathbf{z}}) = N(\tilde{\mathbf{H}}[\mathbf{z} + \tilde{\mathbf{z}}], \Lambda_m), \quad (3)$$

where the variance/covariance matrix Λ_m represents uncertainty due to the measurement process and $\tilde{\mathbf{H}}$ is a (hypothetical) mapping from the continuous multivariate space of reality to the discrete time–space location of the measurement. Like \mathbf{z} , \mathbf{y} is generally a multivariate vector of observations defined at any time and space location. In practice, $\tilde{\mathbf{H}}$ must be approximated by the discrete operator \mathbf{H} that maps from the vector model space to the time–space location of the observations and $\delta\mathbf{H}$ is the error due to this approximation (i.e., $\delta\mathbf{H} \equiv \tilde{\mathbf{H}} - \mathbf{H}$) and (3) becomes

$$\begin{aligned} p(\mathbf{y} | \mathbf{z}, \tilde{\mathbf{z}}) &= N([\mathbf{H} + \delta\mathbf{H}][\mathbf{z} + \tilde{\mathbf{z}}], \Lambda_m) \\ &= N(\mathbf{H}\mathbf{z} + \delta\mathbf{H}\mathbf{z} + \tilde{\mathbf{H}}\tilde{\mathbf{z}}, \Lambda_m). \end{aligned} \quad (4)$$

The terms $\delta\mathbf{H}\mathbf{z}$ and $\tilde{\mathbf{H}}\tilde{\mathbf{z}}$ are the components of the representativeness error due to the use of approximate forward operators and unresolved processes. While forward operator error might be appreciable in some applications (e.g., assimilation of satellite radiances in the atmosphere) for the problem of assimilation of in situ hydrography, forward operator errors are relatively small (i.e., $\delta\mathbf{H}\mathbf{z} \ll \tilde{\mathbf{H}}\tilde{\mathbf{z}}$). Moving forward, we assume that the forward operator error is zero.

The goal here is to use observations to constrain the model-resolvable vector random variable \mathbf{Z} , not $\mathbf{Z} + \tilde{\mathbf{Z}}$. This is an important distinction; in sequential methods that involve successive cycles of assimilation and forecasting (such as Kalman filtering) requiring the numerical model to evolve states that include unresolved processes can degrade the performance of the system (e.g., Lorenc 1986; Daley 1993; Fukumori et al. 1999). We must also, necessarily, admit that only the portion of the observations that we can map using \mathbf{H} can be used as a constraint.

Thus, the target posterior probability distribution is $p(\mathbf{z} | \mathbf{y})$ and Bayes's rule can be used to express this target posterior distribution as

$$p(\mathbf{z} | \mathbf{y}) \propto p(\mathbf{z})p(\mathbf{y} | \mathbf{z}). \quad (5)$$

The first factor on the right-hand side is the prior probability distribution for \mathbf{Z} and the second factor is the data likelihood. Using (3), we can express the likelihood

function $p(\mathbf{y} | \mathbf{z})$ as a marginalization over the unresolved processes:

$$p(\mathbf{y} | \mathbf{z}) = \int_{\tilde{\mathbf{z}}} p(\mathbf{y} | \mathbf{z}, \tilde{\mathbf{z}}) p(\tilde{\mathbf{z}}) d\tilde{\mathbf{z}}. \quad (6)$$

Importantly, the modeling framework described by (1) implicitly assumes that the evolution of the model state $\mathbf{Z} = \mathbf{z}$ is independent of the realization of $\tilde{\mathbf{Z}} = \tilde{\mathbf{z}}$. In other words, there are assumed to be no rectified effects of the realization the unresolved processes on the realization of the model state—and conversely, the realizations of the unresolved process are not model state dependent. We also make the standard assumption that the unresolved processes are normally distributed with mean zero (unbiased) (i.e., $\tilde{\mathbf{Z}} \sim N[0, \tilde{\mathbf{\Lambda}}]$). To be clear, the assumption that there is no covariance between \mathbf{Z} and $\tilde{\mathbf{Z}}$ does not necessarily mean that these two fields do not impact one another. For example, the effects of the statistical properties of the unresolved processes on the resolved state may be accounted for via model parameterization, in which case they are implicitly included in M . This is consistent with the above assumptions unless the parameterizations are state dependent. If the model parameterizations are constructed with state dependency in the mean of $\tilde{\mathbf{Z}}$ then clearly the assumption of zero covariance will be violated. Or, if the covariance of the unresolved processes is a function of the model state, then the marginal likelihood is no longer normally distributed in \mathbf{z} —violating the standard least squares assumption of a Gaussian likelihood function.³ We do not explore these possibilities in this work, but we mention it here for completeness and because it represents a substantive limitation in conceptually separating resolved and unresolved processes.

Noting the above-mentioned assumptions and their caveats, performing the integral in (6), the observational likelihood remains normally distributed,

$$p(\mathbf{y} | \mathbf{z}) = N(\mathbf{H}\mathbf{z}, \mathbf{\Lambda}_o) \mathbf{\Lambda}_o \equiv \tilde{\mathbf{H}} \tilde{\mathbf{\Lambda}} \tilde{\mathbf{H}}^T + \mathbf{\Lambda}_m, \quad (7)$$

and the variance associated with the unresolved process $\tilde{\mathbf{H}} \tilde{\mathbf{\Lambda}} \tilde{\mathbf{H}}^T$ is the representativeness error. Thus, for the practical purpose of doing Bayesian assimilation using (5), assuming Gaussian distributions, and measurement and measurement errors that are uncorrelated with each other and with the resolvable model process, the measurement and representativeness errors can be summed. Readers interested in a more nuanced discussion of representation error and how it can be treated in the context of

data assimilation are referred to [Hodyss and Nichols \(2015\)](#). The development of an estimator for $\mathbf{\Lambda}_o$ under certain conditions is the central goal of this paper. For clarity and consistency with the literature [see [Cohn \(1997\)](#)], it should also be noted that in the special case of (5)–(7) applied to a sequential filtering update algorithm (like the Kalman filter), there is an implied restriction of \mathbf{y} and \mathbf{z} to time t and an implied conditioning on observations at previous times. In this special case, it is necessary to assume that the measurement and representativeness error are uncorrelated in time in order to support the time-serial processing the Kalman filter employs.

3. An estimator for the observational error variance

Moving forward, we focus on the special (but common) case where the observational error variance for a given data type is assumed to be time stationary over a fixed, but limited, geographic region. The choice of these regional boundaries is, naturally, application specific and in many cases may be a heuristic choice based on expert insight into the system behavior. Henceforth, we will refer to this more specific scalar parameter as σ_o^2 , and the estimator or this parameter will be defined for each geographic region independently.⁴

Within each region, a set of n observations available through time and indexed by i , will be used for estimation. Each scalar observation is a realization of the random variable Y_i , distributed as in (7). Each ensemble member j of the model simulations can be mapped to the time and geographic location of any observation using $X_{ij} = \mathbf{H}_i \mathbf{Z}_j$, for $j = 1, 2, 3, \dots, k$, such that Y_i and X_{ij} are now collocated. For each region, we can then form Φ , which will serve as an estimator for σ_o^2 :

$$\Phi = \frac{1}{n} \sum_{i=1}^n (Y_i - \langle X_i \rangle)^2 - \frac{1}{n} \sum_{i=1}^n \frac{k+1}{k} S_i^2. \quad (8)$$

The average of the k model ensembles at time and space location i is denoted $\langle X_i \rangle$ and S_i^2 is the $(k-1)$ normalized) ensemble variance; that is, $S_i^2 = 1/(k-1) \sum_{j=1}^k (X_{ij} - \langle X_i \rangle)^2$. The estimator presented in (8) is the central result of this paper. The first term contains the root-mean-square of deviations of the observational value from the ensemble mean value mapped to the

³Note that the estimator presented in the next section will still be unbiased in the case of a non-Gaussian likelihood.

⁴Note that it is certainly possible to construct a more complex hierarchical spatial model of the observational error variance. And there may be some benefit to doing so for very data-sparse applications. But since one of the goals of this method is ease of implementation, this simpler framework of regionally independent estimation is presented here.

location of the observation. In essence, this term contains a contribution from the observational error variance and the model simulation error variance [see (A4) in the appendix]. This can be nonintuitive for readers who are accustomed to interpreting an ensemble mean as the true state. In an ensemble framework, the correct interpretation is that the true state is statistically indistinguishable from the model simulation ensemble (see item *i* below). As such, it contains one realization of the model simulation error. The second term then contains the sample variance of the ensemble, which is basically a sample estimate of the model simulation variance. The difference of these terms is an estimate of the observational error variance. There is a natural balance in these terms; as the spread of the ensemble (term two) becomes smaller, the ensemble mean will be more reflective of the true state, and the simulation error component of term one will also decrease. The remainder of this section is concerned with the conditions under which Φ can be considered an unbiased estimate of σ_o^2 and the theoretical variance of the estimator. The properties of the estimator presented below are all derived in the appendix.

Properties of the estimator

The estimator Φ is unbiased only if its expected value is equal to σ_o^2 . Here we list the series of conditions for which this holds:

- (i) The model-resolvable true state (denoted \mathbf{Z}_*) along with the simulation ensembles are independently and identically distributed. This is known as a “reliable” ensemble and allows (2) to be extended to include \mathbf{Z}_* ; that is, $[\mathbf{Z}_*, \mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3, \dots, \mathbf{Z}_k] \sim N(\mu, \Lambda)$.
- (ii) There is no correlation between observation errors and simulation errors (i.e., deviations from the mean).
- (iii) The model solution is unbiased relative to the observations (or the bias can be estimated and removed).

Regarding condition i, keep in mind that \mathbf{Z}_* is defined by the error subspace of the simulation ensemble. Thus, the ensemble is reliable by construction. Condition i will be violated by the use of an ensemble that does not represent the range of model-resolvable variability that could emerge in the actual anticipated use of the model.

If the above conditions are met, then, as shown in the appendix,

$$\mathbb{E}(\Phi) = \sigma_o^2, \tag{9}$$

where \mathbb{E} is the expectation operator. If we make the additional assumptions that

- (iv) observation errors are uncorrelated in space and time,

then (from the derivation in the appendix) we can write the variance of Φ as

$$\text{Var}(\Phi) = \frac{2}{n} \left[\sigma_o^4 + \frac{2(k+1)}{k} \sigma_o^2 \bar{\sigma}^2 + \frac{(k+1)^2}{k(k-1)} \frac{n}{\nu_{\text{eff}}} \bar{\sigma}^4 \right], \tag{10}$$

where $\bar{\sigma}^2 \equiv 1/n \sum_{i=1}^n \sigma_i^2$, $\bar{\sigma}^4 \equiv 1/n \sum_{i=1}^n \sigma_i^4$, and σ_i^2 is the simulation error variance at each time–space location *i*; that is, $\sigma_i^2 = \mathbf{H}_i \Lambda \mathbf{H}_i^T$. An important note here is that there is no need to assume that the simulation variance is stationary or that deviations from the mean are time uncorrelated. The term ν_{eff} serves as an effective degrees of freedom that is associated with the autocorrelation and possibly nonstationary variance of the simulation:

$$\nu_{\text{eff}} = \frac{n}{1 + \beta}, \quad \beta = \frac{2}{n\sigma^4} \sum_{i=1}^n \sum_{i'=i+1}^n \sigma_i^2 \sigma_{i'}^2 \rho_{i,i'}^2. \tag{11}$$

Here $\rho_{i,i'}$ is the autocorrelation in the model simulation between locations *i* and *i'*. Although the assumption of uncorrelated observational errors is standard in data assimilation, there is evidence that both measurement errors and representativeness errors may be correlated in geoscience applications (e.g., Stewart et al. 2008; Bormann et al. 2010; Waller et al. 2014a). Readers interested in deriving the properties of the estimator assuming correlated observational errors can do so from (A7) in the appendix. We do not explicitly consider that case here, but suffice to say that the estimator will still be unbiased, but the variance of Φ will increase. We now consider three special cases that lead to significant simplification of (11).

In the special case of no autocorrelation in the model simulations, ν_{eff} is trivially equal to *n*.

In the case that the model ensemble variance is stationary (i.e., $\sigma_i^2 = \sigma_{i'}^2$), observations are available at regular time intervals Δt , and the model simulation has a regional autocorrelation function ρ_τ defined at lags $\tau \Delta t$, then $\beta = 2/n \sum_{\tau=1}^{n-1} (n-\tau) \rho_\tau^2$. Further assuming that the range of significant autocorrelation is much smaller than $n \Delta t$, the factor $(n-\tau)/n$ will be nearly unity for nonzero ρ_τ , and we can approximate ν_{eff} as

$$\nu_{\text{eff}} \approx \frac{n}{1 + 2 \sum_{\tau=1}^{n-1} \rho_\tau^2}. \tag{12}$$

In practice, the assumption that data are available at regular intervals may be too restrictive. Imagine instead

that something is known about the average simulation autocorrelation between all unique locations of the n observations, which we will just call $\overline{\rho^2}$. As before, we make the simplifying assumption that the simulation variance is stationary. In this case, we have $\beta = (n-1)\overline{\rho^2}$. And if $n \gg 1$, then we can approximate ν_{eff} as

$$\nu_{\text{eff}} \approx \frac{n}{1 + n\overline{\rho^2}}. \quad (13)$$

The value of having an analytic form for the variance of Φ is that it illuminates the strategies that could be used to increase the precision of Φ . A discussion along these lines is in the final section of this paper.

4. A Monte Carlo verification of the properties of the estimator

In this section we use Monte Carlo sampling to demonstrate that (9) and (10) accurately reflect the first two moments of the estimator Φ . For the purpose of this demonstration, consider (1) in a univariate context with f a time-uncorrelated, zero-mean random variable with prescribed variance $\sigma_f^2 = 3$ and $M = 0.5$. This model is then simply a first-order autoregressive process, with $\sigma^2 = \sigma_f^2/(1 - M^2)$ and $\rho_\tau = \rho_1^T = M^T$. Pseudo observations can be constructed by integrating (1) through time, at each time $t = i\Delta t$ ($i = 1, 2, 3, \dots, n$) adding random draws of the observational error, which is normally distributed with zero mean and variance $\sigma_o^2 = 5$. We generate a total of $n = 100$ pseudo observations. To apply the estimator Φ to the problem of trying to infer the observational error variance, an ensemble of k integrations of (1) must be run. These are analogous to the ensemble of forced-ocean states that would be available in the real problem. This set of observations and the corresponding collection of ensembles would be the only information available to infer σ_o^2 .

In any real application there is only one set of observations and collection of ensembles. Thus, although the estimator Φ is a random variable, only one sample of Φ can be computed. However, in this demonstration, the goal is to show the distribution of Φ , so that the Monte Carlo estimates of the mean and variance can be compared to the theoretical values. Many samples of Φ are generated by repeatedly creating a sequence of n pseudo observations, regenerating the k ensembles from new integrations of (1) and computing a sample of Φ . For this demonstration, 200 000 samples are used.

Figure 1a shows the Monte Carlo generated discrete probability density function of Φ , computed for the aforementioned values of k , n , variance, and autocorrelation

properties of the autoregressive model (i.e., σ^2 and ρ_τ) and σ_o^2 . The Monte Carlo and theoretical moments are indicated in the panel and are in excellent agreement. Following the parameters outlined in Table 1, Figs. 1b–f show the theoretical (solid lines) and Monte Carlo generated values (dots) of the expected value and variance of Φ as these system and observational parameters vary. The agreement is nearly exact, providing a check on (9), (10), and (11). Equations (10) and (12) and Figs. 1b–f indicate that the variance of Φ increases with increasing σ_o^2 , σ^2 , and ρ_τ and decreases with larger n and k . For the parameter set here explored here, the variance of Φ is only a very weak function of the number of ensembles beyond a threshold value of about 10. Similarly, it is not until the lag-one autocorrelation reaches 0.6 that there is a strong functional dependence on the autocorrelation. On the other hand, for this parameter set the variance of Φ is a strong function of the observation and simulation variance and the number of observations.

5. Illustration using the POP2 ocean model and in situ hydrographic data

Here we apply the observational error variance estimator to in situ temperature data that can be assimilated into the Parallel Ocean Program version 2 (POP2; Smith et al. 2010) global ocean general circulation model. POP2 is a level-coordinate model, with 60 vertical levels, configured with a nominal horizontal resolution of 1° , increasing to $1/4^\circ$ meridionally near the equator. POP2 is the ocean component of the Community Earth System Model (CESM; Gent et al. 2011). Further details on the model can be found in Danabasoglu et al. (2012). Note that the goal of this illustration is not to make the most precise estimation of the observational error but to simply illustrate the methodology.

An ensemble of 30 POP2 ocean simulations is produced by forcing with 30 unique samples of the atmospheric state from an ensemble reanalysis of the atmosphere (Raeder et al. 2012) produced with the nominal 2° resolution Community Atmosphere Model, version 4 (CAM4; Neale et al. 2013). This ensemble atmospheric reanalysis assimilates temperature and winds from radiosondes and aircraft- and satellite-derived drift winds. The ocean ensemble was initialized on 1 January 2004 from a climatological ensemble of POP2 ocean states. The ocean simulations were integrated through 31 December 2006, but 2004 was treated as a spinup period and only years 2005 and 2006 were used for estimating the observation error variance. In situ temperature observations from the 2009 World Ocean Database (WOD09; Johnson et al. 2009) were

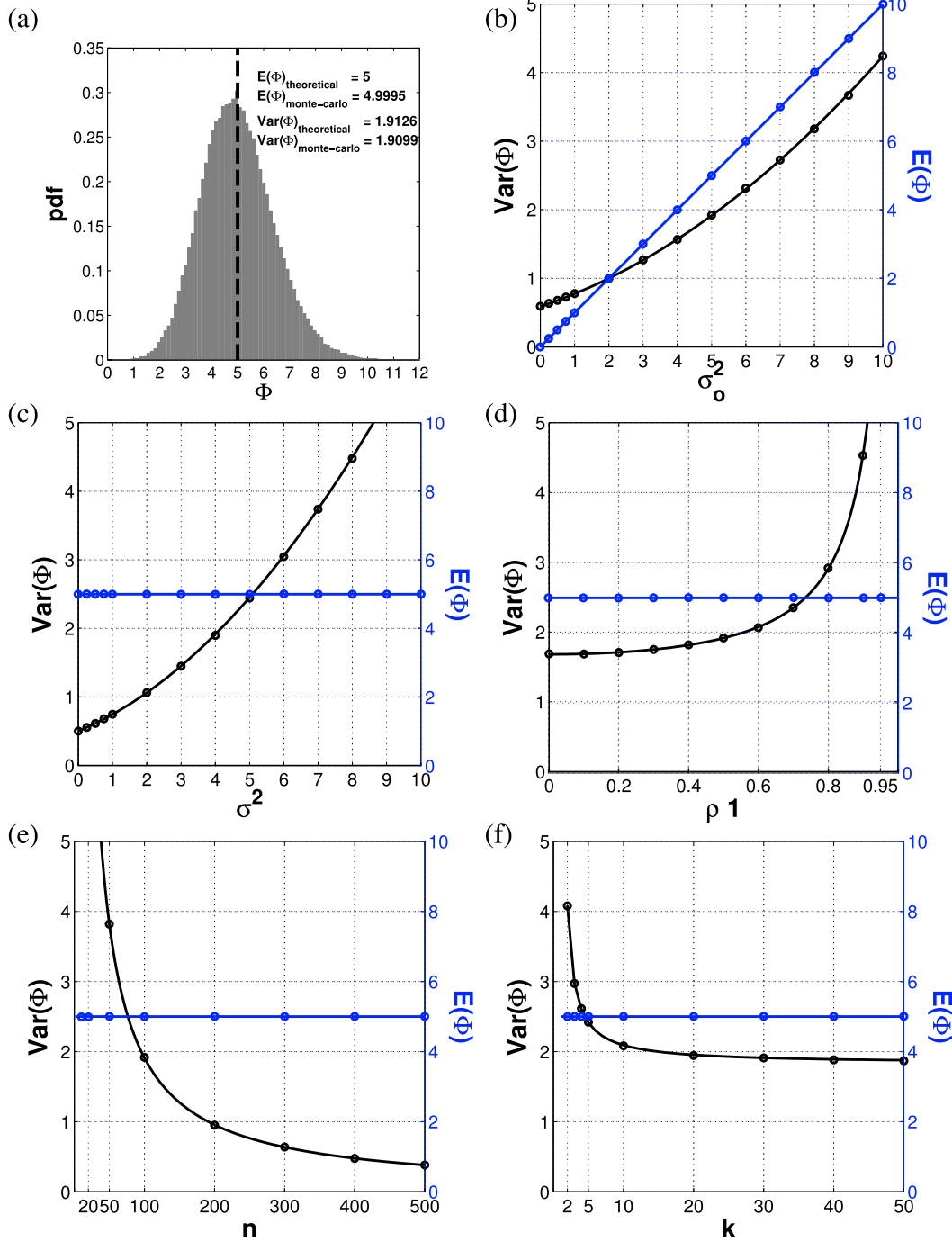


FIG. 1. Monte Carlo estimates of the mean and variance of the estimator for comparison to the theoretical values. (a) The probability density function Φ for the control parameters specified in Table 1. (b) Varying σ_o^2 , solid lines represent the theoretical moments, and the dots represent the Monte Carlo computed values. (c)–(f) As in (b), but for variations in the ensemble variance of the simulation, the lag-one autocorrelation of the system, the number of observations, and the number of ensemble members.

used to compute observation-minus-simulation residuals for use in (8). During 2005 and 2006, most of the in situ temperature data in the *WOD09* comes from autonomous drifting profiling floats (Argo),

bathythermographic (XBT) instruments, moored thermistors, and surface drifting buoys.

To compute residuals and ensemble sample statistics as called for by (8) the temperature field from each

TABLE 1. List of Monte Carlo experiments using a first-order autoregressive model for demonstrating the validity of (9), (10), and (11).

Demonstration name	σ_o^2	σ^2	ρ_1	n	k
Control	5	4	0.5	100	30
Vary σ_o^2	0.1–10	4	0.1–0.9	100	30
Vary σ^2	5	0.1–10	0.5	100	30
Vary ρ_1	5	4	0.1–0.9	100	30
Vary n	5	4	0.5	10–500	30
Vary k	5	4	0.5	100	2–50

ocean ensemble was interpolated to the time and space location of the observations. (This interpolation is simply the application of the forward observation operator **H**.) Because there are systematic biases between the observations and the ensemble forced-ocean simulation that should not be included in the estimation of observational error, the residuals from the year 2005 are used to estimate the systematic bias within 2° grid boxes at standard levels. A two-dimensional Gaussian filter was then applied to this field to smooth the solution. This bias was then removed from the simulation ensembles in the year 2006, prior to applying (8) for 5° boxes.⁵ Note that given a longer simulation, it would be reasonable to make a climatologically varying estimate of the bias. The number of observations used within each grid box is illustrated in Fig. 2 at 100- and 1000-m-depth levels.

For this illustration, we assume time stationarity of the observational error variance, although for some applications there may be benefit to computing a seasonally dependent estimator. With enough data, this would amount to simply applying the estimator to seasonally differentiated subsets of the data.

The Φ estimate of the standard deviation of the observational error at 100- and 1000-m depths is shown in Fig. 3. Zonal versus depth cross sections at the equator and at 35°N are shown in Fig. 4. At the equator the standard deviation of the observational error is highest along the main thermocline, where unresolvable circulation or wind errors will naturally be amplified by the stratification. At 35°N , the Gulf Stream and Kuroshio western boundary currents stand out as regions of high observational error variance. This is due in part to the high eddy activity in these currents—activity that is unresolvable in coarse-resolution ocean models. Our estimates can be visually compared to the solutions of Forget and Wunsch (2007), who also estimate

⁵ Five-degree boxes were used because we found that there were not enough observations in 2° boxes over 1-yr time frames to make meaningful estimates of the observational error variance.

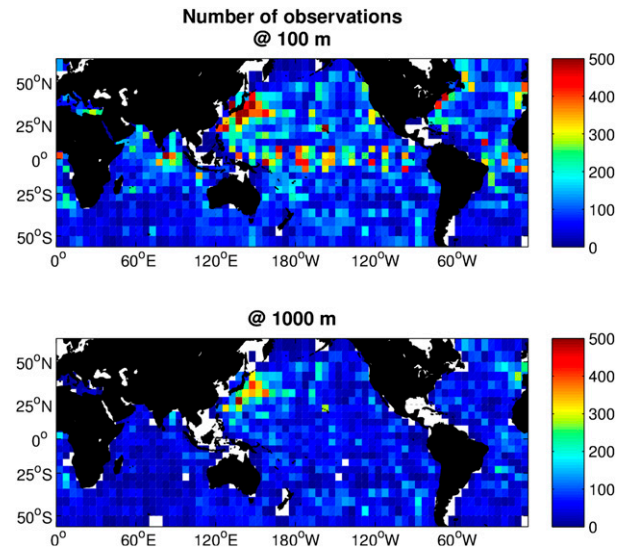


FIG. 2. Number of in situ temperature observations in each 5° box used to compute the estimator. White boxes indicate regions where there were no observations available.

observational error statistics for in situ ocean hydrographic data. These two independent estimates are in excellent agreement. As mentioned in the introduction, theirs is a purely observationally based estimate and is comparable to ours because they use a horizontal space-averaging baseline for their estimates of 1° , which is approximately the same resolution as the POP2 model. The similarity in the estimates suggests that, roughly speaking, a lack of resolution is the primary source of observational error in this model. Note that they also apply a spatial smoothing to their estimates.

The variance of our estimator will be a function of the simulation variance and autocorrelation at the time–space locations of the observations, the number of observations, the number of ensembles, and the underlying true observational error. Of course, excepting the number of ensembles and the number of observations, we do not know these parameters exactly. However within each 5° box approximate values can be assigned using sample estimates from the 2005 period. This is a very rough estimate based on the understanding that the years 2005 and 2006 will share simulation and observational network characteristics and that the sample statistics will be correct within an order of magnitude. Based on these estimated parameter values, Fig. 5 shows spatial maps of the approximate variance of the estimator at 100- and 1000-m depths. Note that the color scale is logarithmic; by these approximations there is enormous spatial and depth inhomogeneity in the variance of the estimator. What dominates the uncertainty in the estimator? In Fig. 6, the horizontal average of the

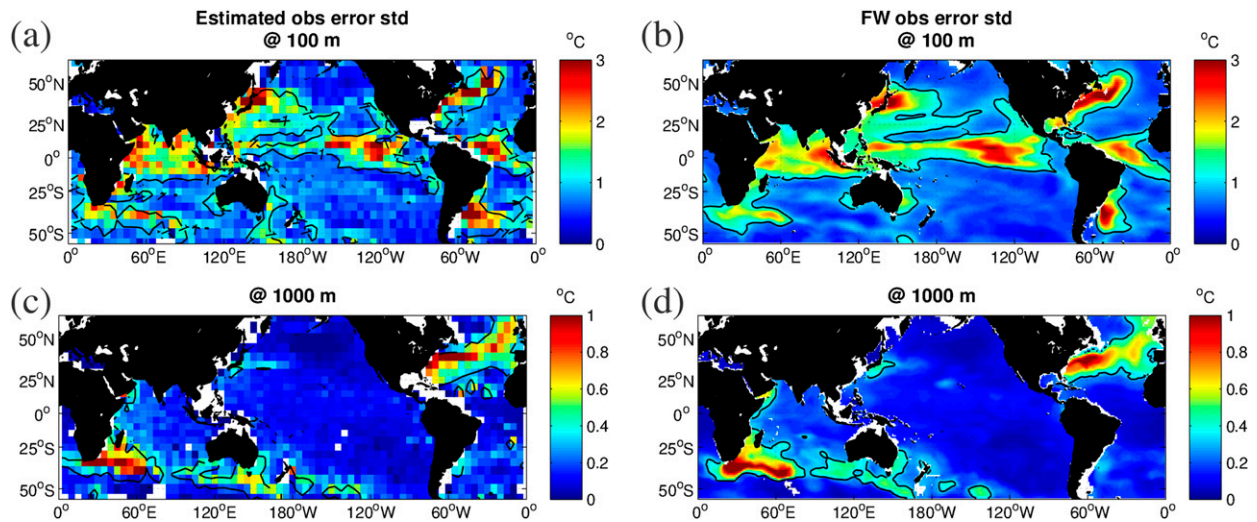


FIG. 3. (left) Estimate of the standard deviation of observational error computed as the square root of the expected value of the estimator Φ . (right) Estimates as described in Forget and Wunsch (2007). At 100-m depth, the line indicates the 1°C contour. At 1000-m depth, the line indicates the $\frac{1}{3}$ °C contour.

percent contribution of each of the three terms in (10) to the total variance of Φ is shown as a function of depth. This suggests that, in this system, the vast majority of the uncertainty in the estimation of observational error variance is a function of the underlying true observational error and is, thus, reducible only through increasing the number of observations used for estimation.

6. Discussion

This paper presents a simple ensemble-based technique for estimating observational error variance for ocean data-assimilation systems. Given an ensemble of atmospheres that reflect the uncertainty in the forcing to the ocean, *observation-minus-simulation* residuals from the corresponding forced-ocean ensembles can be used to form an unbiased estimator for the observational error variance. We also derive an uncertainty associated with the estimator in the case where the standard assumption of uncorrelated observational error is made.

We show that the uncertainty in the estimator Φ is a function of the true value of the observational error variance, the model simulation variance and covariance, the number of observations used in the estimate, and the number of simulation ensembles. It can be understood from (10) that as the total number of observations and ensembles available for estimation increases and the simulation ensemble variance and covariance decreases, the precision of the estimator will improve. While the size of the underlying observational error variance is an irreducible factor in the uncertainty in Φ , the remaining factors can be altered. If possible, more data points and

more ensembles can be used, increasing n and k . And, of course, a different simulation distribution, one with reduced variance and covariance, but that is still reliable, could be constructed.

If one only considers the expected value of the estimator, it can appear that there is no benefit to choosing an ensemble that is both reliable and minimizes the variance of the simulation ensemble. On the other hand, if estimator precision is a goal, there is some benefit to decreasing the simulation ensemble spread and autocorrelation. The extent to which that benefit is significant will depend on the relative contributions of the three terms in (10).

This reduction of uncertainty in the simulation ensemble can result from the assimilation of data. If a data-assimilation update is performed optimally, with observational and background error covariances perfectly known and perfectly represented and the model unbiased, then the forecasts used in sequences of *observation-minus-forecast* residuals will remain reliable, and the estimator shown here will remain unbiased, but with increased precision. On the other hand, it is impossible to do optimal assimilation since we have no a priori knowledge of the observational error variance—and forecasts following a suboptimal update will always violate the reliability criteria. Desroziers et al. (2005), who developed a set of diagnostic statistics that could be applied after successive cycles of assimilation to assess whether the observation and forecast errors had been correctly specified, use analogous reasoning. They show that in the case of perfectly specified errors, the observation error variance is equal to the expected value of the sample covariance of observation-minus-forecasts

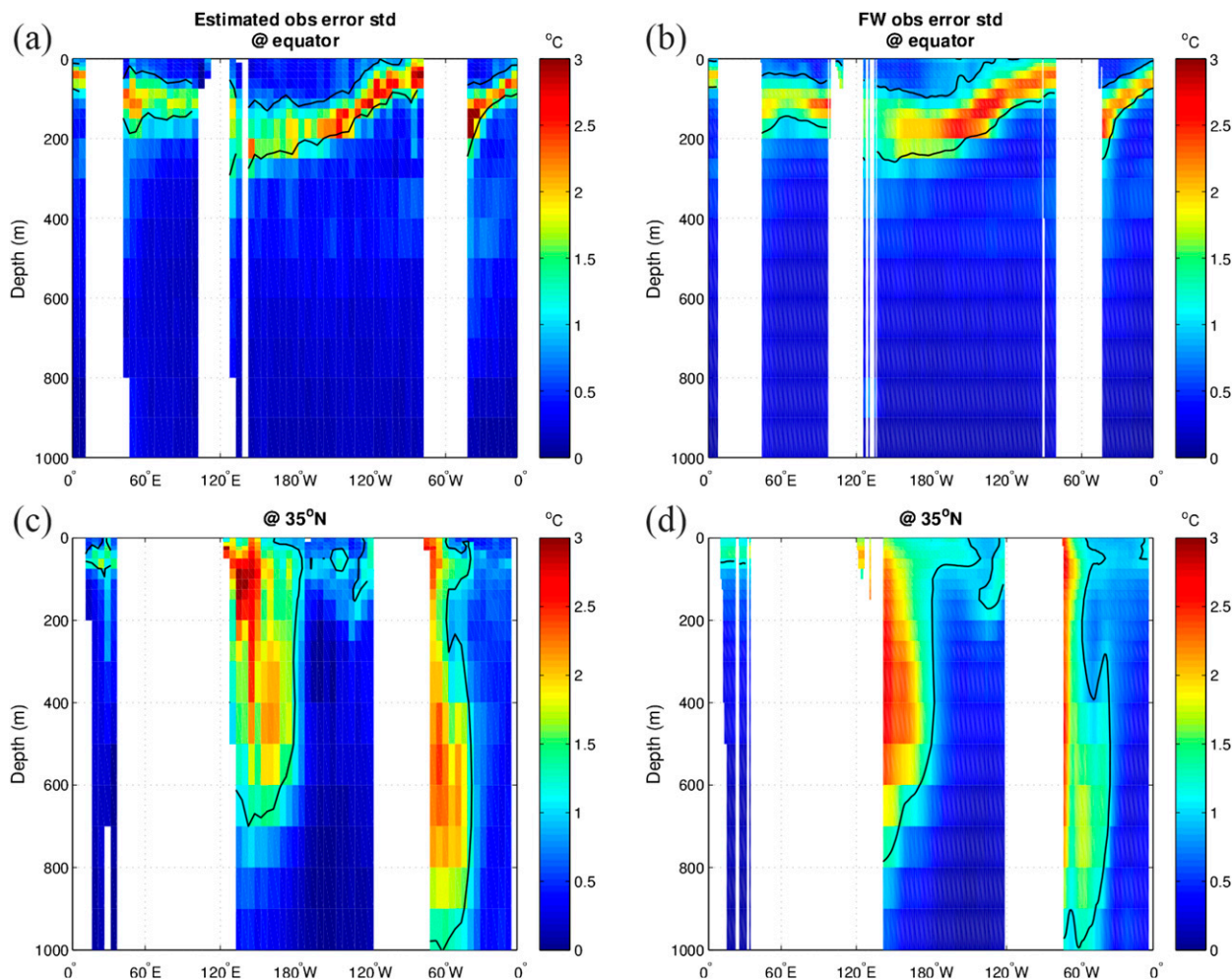


FIG. 4. As in Fig. 3, but for longitude–depth transects at the equator and 35°N. In all panels the line indicates the 1°C contour.

residuals and observation-minus-analysis residuals. The advantage of using simulation residuals instead of forecast residuals from a sequential data-assimilation cycling is that the assumption of correct error statistics need only apply to the model for the estimator to remain unbiased. However, the cost of this less stringent assumption is that precision in the estimator will decrease.

We have focused on an ensemble of forced-ocean simulations for the estimation. But, in fact, the estimator Φ is equally valid for an ensemble of ocean states from a freely evolving coupled ocean–atmosphere model. In this case, the covariance and autocorrelation of the atmosphere ensemble will be larger and will lead to larger variances and autocorrelations in the ocean ensemble as well. It is then a matter of expert opinion whether or not the resulting variance of Φ is too high for useful estimation.

We have also focused on the estimation of observational error variance and have not attempted to estimate the possible time or space covariance. As mentioned in the

manuscript, it is straightforward to extend the estimator to multivariate space so that observational covariances could be computed. However, the lack of time and space proximate hydrographic observations makes the actual estimation practically impossible. Estimation of observational covariances is more sensible for densely sampled observations such as those derived from satellite remote sensors.

The framework presented here relies on relatively strict assumptions regarding the independence of simulation and observational errors and normally distributed processes. While these assumptions are rarely realized precisely, they tend to be useful approximations. For much of the global ocean, the large-scale oceanic response to atmospheric perturbations is well approximated with linear dynamics (e.g., Anderson and Gill 1975; Anderson and Killworth 1977; Malanotte-Rizzoli 1996); however, in some regions (e.g., high-latitude convective regimes) the extent to which nonlinear dynamics lead to nonnormal distributions

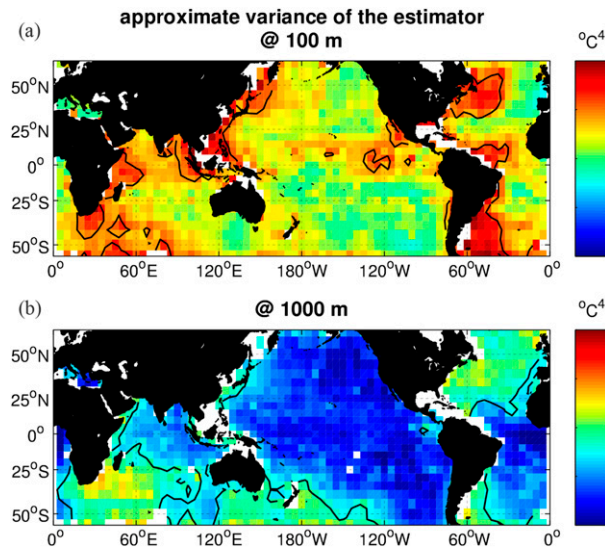


FIG. 5. The approximate variance of the estimator Φ . Note that the color scale is logarithmic. For 100-m (1000 m) depth the black line indicates the 1°C^4 (0.001°C^4).

could potentially be an issue. This is also an issue for any nonlinear forward operator. However, as mentioned in the appendix, even when errors are non-Gaussian, the estimator presented here is still unbiased but the variance of the estimator is no longer given by the formula presented in this work. The assumption of independence between the resolved and unresolved process poses another interesting challenge. This assumption goes far beyond this work; it is fundamental to the formulation of most data-assimilation methods. The development of state-dependent observational error statistics would be a fruitful line of research, as would the development of techniques for the joint estimation of resolvable and unresolvable processes.

Finally, it is worth commenting on one of the basic assumptions in this work: that of a “reliable” ensemble with which to generate our estimate of observational error. How can one verify that a model ensemble has been suitably formed? Unfortunately, when real data are used for benchmarking, there is no objective answer to the question of whether or not the ensemble is reliable because measures of reliability must account for observational errors [see Hamill (2001) and Anderson (1996)]. Thus, there is no way of knowing with certainty whether any emerging inconsistencies result from the misspecification of background errors or observational errors. It is a frustratingly circular state of affairs. A possible route forward is to attempt to establish reliability based on independent data sources that can be used to aggregate over time–space scales that can be assumed to be well resolved by the model. It might then be possible to assume that the error of representation (and measurement) is much much less than the background error—such that

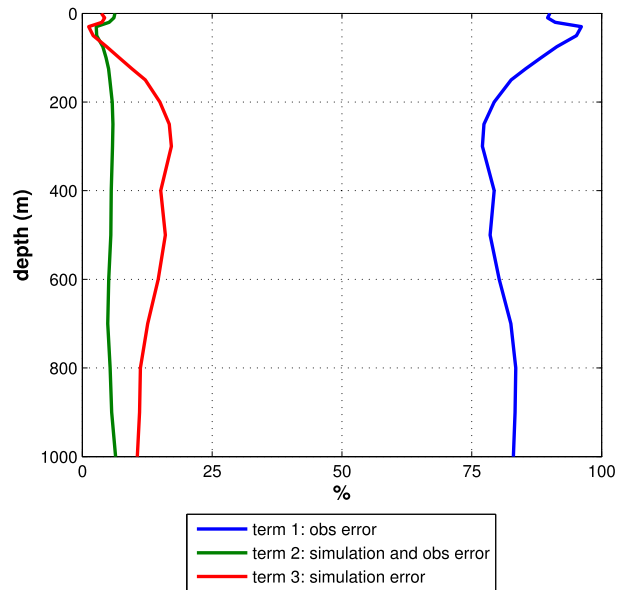


FIG. 6. For each depth, the horizontal average of the percent contribution of each of the three terms to the total variance of Φ . Blue line: term 1 from (10), representing the estimated contribution to the variance of Φ that emerges from the magnitude of the observational error; green line: contribution from term 2, which emerges from a combination of the simulation and observational error variance; red line: contribution from term 3, which represents the simulation error variance and the reduced degrees of freedom associated with auto correlation in the simulation ensemble.

traditional measures of reliability such as the “uniformity of the rank histogram” (Hamill 2001) could be interpreted with less ambiguity.

Acknowledgments. Special thanks to Alexey Kaplan, Jeff Anderson, Robert Miller, and Doug Nychka for helpful discussions on this topic. Thanks also to the three anonymous reviewers for their constructive reading of the manuscript. This work was funded in part by the NOAA Climate Program Office under the Climate Variability and Predictability Program Grants NA09OAR4310163 and NA13OAR4310138 and by the NSF Collaborative Research EaSM2 Grant OCE-1243015.

APPENDIX

Deriving the First Two Moments of the Estimator

This appendix derives the expected value and variance of the observational error estimator Φ . All notation is consistent with the main body of the paper. We begin by rewriting the estimator (8) from the main text in vector form,

$$\Phi = \frac{1}{n} [\mathbf{Y} - \mathbf{H}(\mathbf{Z})]^T [\mathbf{Y} - \mathbf{H}(\mathbf{Z})] - \frac{1}{n} \frac{k+1}{k-1} \frac{1}{k} \sum_{j=1}^k [\mathbf{z}_j - \langle \mathbf{Z} \rangle]^T \mathbf{H}^T \mathbf{H} [\mathbf{z}_j - \langle \mathbf{Z} \rangle], \quad (\text{A1})$$

where observed values Y_i are in a column vector \mathbf{Y} and the linear matrix operator $\mathbf{H} = [\mathbf{H}_1, \mathbf{H}_2, \mathbf{H}_3, \dots, \mathbf{H}_n]$ can be used to map each state-space vector \mathbf{Z}_j into an observation space column vector. (As in the main text, $\langle \cdot \rangle$ denotes an ensemble average.)

We will derive the moments of Φ using the algebra of random variables (e.g., Springer 1979). The algebra is made considerably more straightforward by treating the random variables \mathbf{Z} and \mathbf{Y} as the sum of a mean value and a deviation from the mean, where the deviations are zero-mean random variables. From the main text (2), (7), and the reliability criteria, we can express the model ensembles, the model-resolvable true state of the ocean, and the observations as

$$\mathbf{HZ}_j = \mathbf{H}\boldsymbol{\mu} + \boldsymbol{\varepsilon}_j, \quad \boldsymbol{\varepsilon}_j \sim N(0, \boldsymbol{\Sigma}), \quad j = 1, 2, 3, \dots, k$$

$$\mathbf{HZ}_* = \mathbf{H}\boldsymbol{\mu} + \boldsymbol{\varepsilon}_*, \quad \boldsymbol{\varepsilon}_* \sim N(0, \boldsymbol{\Sigma})$$

$$\mathbf{Y} = \mathbf{HZ}_* + \boldsymbol{\varepsilon}_o, \quad \boldsymbol{\varepsilon}_o \sim N(0, \boldsymbol{\Sigma}_o). \tag{A2}$$

The zero-mean random variables $\boldsymbol{\varepsilon}_j$, $\boldsymbol{\varepsilon}_*$, and $\boldsymbol{\varepsilon}_o$ are all vectors of length n and $\boldsymbol{\Sigma} = \mathbf{H}\boldsymbol{\Lambda}\mathbf{H}^T$ and $\boldsymbol{\Sigma}_o$ are $[n \times n]$ covariance matrices. The body of the paper is only concerned with the case in which the off-diagonal elements of $\boldsymbol{\Sigma}_o$ are zero and the diagonal elements are a constant σ_o^2 (regionally stationary, uncorrelated observational errors). But we derive the complete matrix form of the moments of the estimator and do not apply this simplification until later in the appendix.

We use (A2) to express (A1) in terms of the vector deviations associated with \mathbf{Y} and \mathbf{Z} :

$$\begin{aligned} \Phi &= \frac{1}{n} [\boldsymbol{\varepsilon}_* + \boldsymbol{\varepsilon}_o - \langle \boldsymbol{\varepsilon} \rangle]^T [\boldsymbol{\varepsilon}_* + \boldsymbol{\varepsilon}_o - \langle \boldsymbol{\varepsilon} \rangle] - \frac{1}{n} \frac{k+1}{k-1} \frac{1}{k} \sum_{j=1}^k [\boldsymbol{\varepsilon}_j - \langle \boldsymbol{\varepsilon} \rangle]^T [\boldsymbol{\varepsilon}_j - \langle \boldsymbol{\varepsilon} \rangle] \\ &= \frac{1}{n} [\boldsymbol{\varepsilon}_*^T \boldsymbol{\varepsilon}_* + \boldsymbol{\varepsilon}_o^T \boldsymbol{\varepsilon}_o + 2\boldsymbol{\varepsilon}_*^T \boldsymbol{\varepsilon}_o - 2(\boldsymbol{\varepsilon}_* + \boldsymbol{\varepsilon}_o)^T \langle \boldsymbol{\varepsilon} \rangle + \langle \boldsymbol{\varepsilon} \rangle^T \langle \boldsymbol{\varepsilon} \rangle] - \frac{1}{n} \frac{k+1}{k-1} \frac{1}{k} \sum_{j=1}^k [\boldsymbol{\varepsilon}_j^T \boldsymbol{\varepsilon}_j + \langle \boldsymbol{\varepsilon} \rangle^T \langle \boldsymbol{\varepsilon} \rangle - 2\boldsymbol{\varepsilon}_j^T \langle \boldsymbol{\varepsilon} \rangle]. \end{aligned} \tag{A3}$$

Using the identity^{A1} $\langle \boldsymbol{\varepsilon} \rangle^T \langle \boldsymbol{\varepsilon} \rangle = 1/k^2 \sum_{j=1}^k \boldsymbol{\varepsilon}_j^T \boldsymbol{\varepsilon}_j + 2/k^2 \sum_{j=1}^k \sum_{m=j+1}^k \boldsymbol{\varepsilon}_j^T \boldsymbol{\varepsilon}_m$, we can rewrite Φ (terms are labeled for later convenience):

$$\begin{aligned} \Phi &= \frac{1}{n} \left[\underbrace{\boldsymbol{\varepsilon}_*^T \boldsymbol{\varepsilon}_*}_A + \underbrace{\boldsymbol{\varepsilon}_o^T \boldsymbol{\varepsilon}_o}_B + \underbrace{2\boldsymbol{\varepsilon}_*^T \boldsymbol{\varepsilon}_o}_C - \underbrace{\frac{2}{k} (\boldsymbol{\varepsilon}_* + \boldsymbol{\varepsilon}_o)^T \sum_{j=1}^k \boldsymbol{\varepsilon}_j}_D \right. \\ &\quad \left. - \underbrace{\frac{1}{k} \sum_{j=1}^k \boldsymbol{\varepsilon}_j^T \boldsymbol{\varepsilon}_j}_E + \underbrace{\frac{4}{(k-1)(k)} \sum_{j=1}^k \sum_{m=j+1}^k \boldsymbol{\varepsilon}_j^T \boldsymbol{\varepsilon}_m}_F \right]. \end{aligned} \tag{A4}$$

The expected value of (A4) can be computed for the following assumptions (also outlined in the main text), expressed in terms of the expectation operator:

- (i) The simulation ensemble is reliable ($\boldsymbol{\varepsilon}_*$ and $\boldsymbol{\varepsilon}_j$ are identically distributed); that is, $\mathbb{E}(\boldsymbol{\varepsilon}_j^T \boldsymbol{\varepsilon}_j) = \mathbb{E}(\boldsymbol{\varepsilon}_*^T \boldsymbol{\varepsilon}_*)$ for $j = (1, 2, 3, \dots, k)$.
- (ii) The deviations associated with the simulation ensembles are independent of one another and independent of deviations associated with the model-resolvable true state; that is, $\mathbb{E}(\boldsymbol{\varepsilon}_j^T \boldsymbol{\varepsilon}_*) = \mathbb{E}(\boldsymbol{\varepsilon}_j^T \boldsymbol{\varepsilon}_m) = 0$ for $j, m = (1, 2, 3, \dots, k) j \neq m$.

- (iii) The observation errors are uncorrelated with the simulation- and model-resolvable true state deviations; that is, $\mathbb{E}(\boldsymbol{\varepsilon}_o^T \boldsymbol{\varepsilon}_*) = \mathbb{E}(\boldsymbol{\varepsilon}_o^T \boldsymbol{\varepsilon}_j) = 0$ and

$$\begin{aligned} \mathbb{E}(\Phi) &= \frac{1}{n} \left[\mathbb{E}(\boldsymbol{\varepsilon}_*^T \boldsymbol{\varepsilon}_*) + \mathbb{E}(\boldsymbol{\varepsilon}_o^T \boldsymbol{\varepsilon}_o) - \frac{1}{k} \sum_{j=1}^k \mathbb{E}(\boldsymbol{\varepsilon}_j^T \boldsymbol{\varepsilon}_j) \right] \\ &= \frac{1}{n} \text{Tr}(\boldsymbol{\Sigma}_o), \end{aligned} \tag{A5}$$

where Tr denotes the trace of a matrix. This demonstrates that Φ is an unbiased estimator of the average observational error variance. If the observational error variance is constant, then $\mathbb{E}(\Phi) = \sigma_o^2$. Note that although we have assumed normality of the errors, this is not a necessary condition for the estimator to be unbiased [see identities in Bao and Ullah (2010)]. To compute the variance of Φ , we use the identity $\text{Var}(\Phi) = \mathbb{E}(\Phi^2) - \mathbb{E}(\Phi)^2$. Forming $\mathbb{E}(\Phi^2)$ requires the examination of 21 terms (the product of every unique combination of terms A through F). Drawing on the assumptions outlined above, only nine of those terms are nonzero, and we can write

$$\begin{aligned} \mathbb{E}(\Phi^2) &= \frac{1}{n^2} \mathbb{E}[A^2 + B^2 + C^2 + D^2 + E^2 + F^2 \\ &\quad + 2AB + 2AE + 2BE], \end{aligned} \tag{A6}$$

where the squares should be understood to be a matrix product.

^{A1} This identity is easily derivable with basic algebraic expansion of the ensemble average operator.

We can form the terms in (A6) using a set of derived identities for the expectation of quadratic forms involving two zero-mean, independent, normally distributed multivariate random variables \mathbf{a} and \mathbf{b} (see, e.g., Lemma 2.3; Magnus 1978; Bao and Ullah 2010): $\mathbb{E}([\mathbf{a}^T \mathbf{a}][\mathbf{a}^T \mathbf{b}]) = 0$, $\mathbb{E}([\mathbf{a}^T \mathbf{a}][\mathbf{a}^T \mathbf{a}]) = 2\text{Tr}(\mathbf{C}_a \mathbf{C}_a) + \text{Tr}(\mathbf{C}_a)\text{Tr}(\mathbf{C}_a)$, $\mathbb{E}([\mathbf{a}^T \mathbf{b}][\mathbf{a}^T \mathbf{b}]) = \text{Tr}(\mathbf{C}_a \mathbf{C}_b)$, $\mathbb{E}([\mathbf{a}^T \mathbf{a}][\mathbf{b}^T \mathbf{b}]) = \text{Tr}(\mathbf{C}_a)\text{Tr}(\mathbf{C}_b)$, where \mathbf{C}_a and \mathbf{C}_b are the covariance matrices of \mathbf{a} and \mathbf{b} ;

$$\begin{aligned} \mathbb{E}(A^2) &= 2\text{Tr}(\mathbf{\Sigma}\mathbf{\Sigma}) + \text{Tr}(\mathbf{\Sigma})\text{Tr}(\mathbf{\Sigma}), \\ \mathbb{E}(B^2) &= 2\text{Tr}(\mathbf{\Sigma}_o \mathbf{\Sigma}_o) + \text{Tr}(\mathbf{\Sigma}_o)\text{Tr}(\mathbf{\Sigma}_o), \\ \mathbb{E}(C^2) &= 4\text{Tr}(\mathbf{\Sigma}\mathbf{\Sigma}_o), \\ \mathbb{E}(D^2) &= \frac{4}{k}[\text{Tr}(\mathbf{\Sigma}\mathbf{\Sigma}_o) + \text{Tr}(\mathbf{\Sigma}\mathbf{\Sigma})], \\ \mathbb{E}(E^2) &= \frac{2}{k}\text{Tr}(\mathbf{\Sigma}\mathbf{\Sigma}) + \text{Tr}(\mathbf{\Sigma})\text{Tr}(\mathbf{\Sigma}), \\ \mathbb{E}(F^2) &= \frac{8}{(k)(k-1)}\text{Tr}(\mathbf{\Sigma}\mathbf{\Sigma}), \\ \mathbb{E}(2AB) &= 2\text{Tr}(\mathbf{\Sigma})\text{Tr}(\mathbf{\Sigma}_o), \\ \mathbb{E}(2AE) &= -2\text{Tr}(\mathbf{\Sigma})\text{Tr}(\mathbf{\Sigma}), \\ \mathbb{E}(2BE) &= -2\text{Tr}(\mathbf{\Sigma})\text{Tr}(\mathbf{\Sigma}_o), \end{aligned}$$

Unlike the equalities used in (A5), the above equalities only hold for normally distributed errors. Equation (A6) is then

$$\mathbb{E}(\Phi^2) = \frac{1}{n^2} \left[3\text{Tr}(\mathbf{\Sigma}_o \mathbf{\Sigma}_o) + \frac{4(k+1)}{k}\text{Tr}(\mathbf{\Sigma}_o \mathbf{\Sigma}_m) + \frac{2(k+1)^2}{k(k-1)}\text{Tr}(\mathbf{\Sigma}\mathbf{\Sigma}) \right] \tag{A7}$$

and the exact matrix form of the dispersion of the estimator Φ is

$$\begin{aligned} \text{Var}(\Phi) = \mathbb{E}(\Phi^2) - \mathbb{E}(\Phi)^2 &= \frac{2}{n^2} \left[\text{Tr}(\mathbf{\Sigma}_o \mathbf{\Sigma}_o) \right. \\ &\quad \left. + \frac{2(k+1)}{k}\text{Tr}(\mathbf{\Sigma}_o \mathbf{\Sigma}) + \frac{(k+1)^2}{k(k-1)}\text{Tr}(\mathbf{\Sigma}\mathbf{\Sigma}) \right]. \end{aligned} \tag{A8}$$

If we take $\rho_{i,i'}$ to be the autocorrelation in the model simulation between time-space locations i and i' , and σ_i^2 and $\sigma_{i'}^2$ to be the corresponding variances and the observational errors to be uncorrelated and with a constant variance, then we can write

$$\begin{aligned} \text{Var}(\Phi) &= \frac{2}{n^2} \left[n\sigma_o^4 + \frac{2(k+1)}{k}\sigma_o^2 \sum_{i=1}^n \sigma_i^2 + \frac{(k+1)^2}{k(k-1)} \sum_{i=1}^n \sum_{i'=1}^n (\sigma_i^2 \sigma_{i'}^2 \rho_{i,i'}^2) \right] \\ &= \frac{2}{n} \left[\sigma_o^4 + \frac{2(k+1)}{k}\sigma_o^2 \overline{\sigma^2} + \frac{(k+1)^2}{k(k-1)} \frac{n}{\nu_{\text{eff}}} \overline{\sigma^4} \right], \end{aligned} \tag{A9}$$

where ν_{eff} is an effective degrees of freedom associated with the autocorrelation and possibly nonstationary variance of the model ensemble and is given by

$$\nu_{\text{eff}} = \frac{n}{1 + \beta}, \quad \beta = \frac{2}{n\sigma^4} \sum_{i=1}^n \sum_{i'=i+1}^n \sigma_i^2 \sigma_{i'}^2 \rho_{i,i'}^2. \tag{A10}$$

The exact result for the dispersion of our estimator Φ is given by (A9) and (A10).

REFERENCES

Anderson, D., and A. Gill, 1975: Spin-up of a stratified ocean, with applications to upwelling. *Deep-Sea Res. Oceanogr. Abstr.*, **22**, 583–596, doi:10.1016/0011-7471(75)90046-7.
 —, and P. Killworth, 1977: Spin-up of a stratified ocean, with topography. *Deep-Sea Res.*, **24**, 709–732, doi:10.1016/0146-6291(77)90495-7.
 Anderson, J. L., 1996: A method for producing and evaluating probabilistic forecasts from ensemble model integrations.

J. Climate, **9**, 1518–1530, doi:10.1175/1520-0442(1996)009<1518:AMFPAE>2.0.CO;2.
 Balmaseda, M. A., K. Mogensen, and A. T. Weaver, 2013: Evaluation of the ECMWF ocean reanalysis system ORAS4. *Quart. J. Roy. Meteor. Soc.*, **139**, 1132–1161, doi:10.1002/qj.2063.
 Bao, Y., and A. Ullah, 2010: Expectation of quadratic forms in normal and nonnormal variables with applications. *J. Stat. Plann. Inference*, **140**, 1193–1205, doi:10.1016/j.jspi.2009.11.002.
 Behringer, D. W., M. Ji, and A. Leetmaa, 1998: An improved coupled model for ENSO prediction and implications for ocean initialization. Part I: The ocean data assimilation system. *Mon. Wea. Rev.*, **126**, 1013–1021, doi:10.1175/1520-0493(1998)126<1013:AICMFE>2.0.CO;2.
 Blanchet, I., C. Frankignoul, and M. Cane, 1997: A comparison of adaptive Kalman filters for a tropical Pacific Ocean model. *Mon. Wea. Rev.*, **125**, 40–58, doi:10.1175/1520-0493(1997)125<0040:ACOKF>2.0.CO;2.
 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, doi:10.1002/qj.615.

- Cohn, S., 1997: An introduction to estimation theory. *J. Meteor. Soc. Japan*, **75** (1B), 257–288.
- Compo, G., and Coauthors, 2011: The Twentieth Century Re-analysis project. *Quart. J. Roy. Meteor. Soc.*, **137**, 1–28, doi:10.1002/qj.776.
- Daley, R., 1993: Estimating observation error statistics for atmospheric data assimilation. *Ann. Geophys.*, **11**, 634–647.
- Danabasoglu, G., S. Bates, B. Briegleb, S. R. Jayne, M. Jochum, W. Large, S. Peacock, and S. Yeager, 2012: The CCSM4 ocean component. *J. Climate*, **25**, 1361–1389, doi:10.1175/JCLI-D-11-00091.1.
- Dee, D. P., 1995: On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.*, **123**, 1128–1145, doi:10.1175/1520-0493(1995)123<1128:OLEOC>2.0.CO;2.
- , and A. M. DaSilva, 1999: Maximum-likelihood estimation of forecast and observation error covariance parameters. Part I: Methodology. *Mon. Wea. Rev.*, **127**, 1822–1834, doi:10.1175/1520-0493(1999)127<1822:MLEOFA>2.0.CO;2.
- Derber, J., and A. Rosati, 1989: A global oceanic data assimilation system. *J. Phys. Oceanogr.*, **19**, 1333–1347, doi:10.1175/1520-0485(1989)019<1333:AGODAS>2.0.CO;2.
- 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, doi: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, doi:10.1256/qj.05.108.
- Forget, G., and C. Wunsch, 2007: Estimated global hydrographic variability. *J. Phys. Oceanogr.*, **37**, 1997–2008, doi:10.1175/JPO3072.1.
- Fu, L.-L., I. Fukumori, and R. Miller, 1993: Fitting dynamic models to the Geosat sea level observations in the tropical Pacific Ocean. Part II: A linear, wind-driven model. *J. Phys. Oceanogr.*, **23**, 2162–2181, doi:10.1175/1520-0485(1993)023<2162:FDMTTG>2.0.CO;2.
- Fukumori, I., R. Raghunath, L.-L. Fu, and Y. Chao, 1999: Assimilation of TOPEX/Poseidon altimeter data into a global ocean circulation model: How good are the results? *J. Geophys. Res.*, **104**, 25 647–25 665, doi:10.1029/1999JC900193.
- Giese, B. S., and S. Ray, 2011: El Niño variability in simple ocean data assimilation (SODA), 1871–2008. *J. Geophys. Res.*, **116**, C02024, doi:10.1029/2010jc006695.
- Gent, P. R., and Coauthors, 2011: The Community Climate System Model version 4. *J. Climate*, **24**, 4973–4991, doi:10.1175/2011JCLI4083.1.
- Hamill, T. M., 2001: Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev.*, **129**, 550–560, doi:10.1175/1520-0493(2001)129<0550:IORHFV>2.0.CO;2.
- Hodys, D., and N. Nichols, 2015: The error of representation: Basic understanding. *Tellus*, **67A**, 24822, doi:10.3402/tellusa.v67.24822.
- Hollingsworth, A., and R. Lönnberg, 1986: The statistical structure of the short-range forecast errors as determined from radiosonde data. Part I: The wind field. *Tellus*, **38A**, 111–136, doi:10.1111/j.1600-0870.1986.tb00460.x.
- Janjić, T., and S. E. Cohn, 2006: Treatment of observation error due to unresolved scales in atmospheric data assimilation. *Mon. Wea. Rev.*, **134**, 2900–2915, doi:10.1175/MWR3229.1.
- Johnson, D., T. Boyer, H. Garcia, R. Locarnini, O. Baranova, and M. Zweng, 2009: *World Ocean Database 2009 Documentation*. NOAA Printing Office, 175 pp.
- Karspeck, A., S. Yeager, G. Danabasoglu, T. Hoar, N. Collins, K. Raeder, J. Anderson, and J. Tribbia, 2013: An ensemble adjustment Kalman filter for the CCSM4 ocean component. *J. Climate*, **26**, 7392–7413, doi:10.1175/JCLI-D-12-00402.1.
- Leeuwenburgh, O., 2007: Validation of an EnKF system for OGCM initialization assimilating temperature, salinity, and surface height measurements. *Mon. Wea. Rev.*, **135**, 125–139, doi:10.1175/MWR3272.1.
- 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, doi:10.1002/qj.371.
- Lönnberg, R., and A. Hollingsworth, 1986: The statistical structure of the short-range forecast errors as determined from radiosonde data. Part II: The covariance of height and wind errors. *Tellus*, **38A**, 137–161, doi:10.1111/j.1600-0870.1986.tb00461.x.
- Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, **112**, 1177–1194, doi:10.1002/qj.49711247414.
- Magnus, J. R., 1978: The moments of products of quadratic forms in normal variables. *Stat. Neerl.*, **32**, 201–210, doi:10.1111/j.1467-9574.1978.tb01399.x.
- Malanotte-Rizzoli, P., Ed., 1996: *Modern Approaches to Data Assimilation in Ocean Modeling*. Oceanography Series, Vol. 61, Elsevier, 455 pp.
- Menemenlis, D., and M. Chechelnitsky, 2000: Error estimates for an ocean general circulation model from altimeter and acoustic tomography data. *Mon. Wea. Rev.*, **128**, 763–778, doi:10.1175/1520-0493(2000)128<0763:EEFAOG>2.0.CO;2.
- Miyoshi, T., E. Kalnay, and H. Li, 2013: Estimating and including observation-error correlations in data assimilation. *Inverse Probl. Sci. Eng.*, **21**, 387–398, doi:10.1080/17415977.2012.712527.
- Neale, R., J. Richter, S. Park, R. Lauritzen, S. Vavrus, P. Rasch, and M. Zhang, 2013: The mean climate of the Community Atmosphere Model (CAM4) in forced SST and fully coupled experiments. *J. Climate*, **26**, 5150–5168, doi:10.1175/JCLI-D-12-00236.1.
- Oke, P., and P. Sakov, 2008: Representation error of oceanic observations for data assimilation. *J. Atmos. Oceanic Technol.*, **25**, 1004–1017, doi:10.1175/2007JTECHO558.1.
- Raeder, K., J. Anderson, N. Collins, T. Hoar, J. Kay, P. Lauritzen, and R. Pincus, 2012: DART/CAM: An ensemble data assimilation system for CESM atmospheric models. *J. Climate*, **25**, 6304–6317, doi:10.1175/JCLI-D-11-00395.1.
- Richman, J. G., R. N. Miller, and Y. H. Spitz, 2005: Error estimates for assimilation of satellite sea surface temperature data in ocean climate models. *Geophys. Res. Lett.*, **32**, L18608, doi:10.1029/2005GL023591.
- Rutherford, I. D., 1972: Data assimilation by statistical interpolation of forecast error fields. *J. Atmos. Sci.*, **29**, 809–815, doi:10.1175/1520-0469(1972)029<0809:DABSIO>2.0.CO;2.
- , 1976: An operational three-dimensional multivariate statistical objective analysis scheme. *Proc. JOC Study Group Conf. on Four Dimensional Data Assimilation*, Paris, France, WMO, 98–121.
- Smith, R., and Coauthors, 2010: The Parallel Ocean Program (POP) reference manual—Ocean component of the Community Climate System Model (CCSM) and Community Earth System Model (CESM). Los Alamos National Laboratory Tech. Rep LAUR-10-01853, 141 pp. [Available online at <http://www.cesm.ucar.edu/models/ccsm4.0/pop/doc/sci/POPRefManual.pdf>.]
- Springer, M., 1979: *The Algebra of Random Variables*. Wiley, 470 pp.

- Stewart, L., S. Dance, and N. Nichols, 2008: Correlated observation errors in data assimilation. *Int. J. Numer. Methods Fluids*, **56**, 1521–1527, doi:10.1002/fld.1636.
- Waller, J., S. Dance, A. Lawless, and N. Nichols, 2014a: Estimating correlated observation error statistics using an ensemble transform Kalman filter. *Tellus*, **66A**, 23294, doi:10.3402/tellusa.v66.23294.
- , —, —, —, and J. Eyre, 2014b: Representativity error for temperature and humidity using the Met Office high-resolution model. *Quart. J. Roy. Meteor. Soc.*, **140**, 1189–1197, doi:10.1002/qj.2207.
- Wunsch, C., and P. Heimbach, 2007: Practical global oceanic state estimation. *Physica D*, **230**, 197–208, doi:10.1016/j.physd.2006.09.040.
- Yang, C., and B. S. Giese, 2013: El Niño Southern Oscillation in an ensemble ocean reanalysis and coupled climate models. *J. Geophys. Res. Oceans*, **118**, 4052–4071, doi:10.1002/jgrc.20284.
- Zhang, S., M. J. Harrison, A. Rosati, and A. Wittenberg, 2007: System design and evaluation of coupled ensemble data assimilation for global oceanic climate studies. *Mon. Wea. Rev.*, **135**, 3541–3564, doi:10.1175/MWR3466.1.