Regional and Basin Scale Applications of Ensemble Adjustment Kalman Filter and 4D-Var Ocean Data Assimilation Systems

Andrew Moore1*, Javier Zavala-Garay2, Hernan G. Arango2, Christopher A. Edwards1, Jeffrey Anderson3 and Tim Hoar3

1: Department of Ocean Sciences, University of California, Santa Cruz CA, U.S.A.
2: Department of Marine Sciences, Rutgers University, New Brunswick NJ, U.S.A.
3: National Center for Atmospheric Research, Boulder CO, U.S.A.

* Corresponding author: ammoore@ucsc.edu

Department of Ocean Sciences
University of California Santa Cruz
1156 High Street, Santa Cruz CA 95064
Abstract

The performance of two common approaches to data assimilation, an Ensemble Adjustment Kalman Filter (EAKF) and a 4-dimensional variational (4D-Var) method, is quantified in a popular community ocean model, the Regional Ocean Modeling Systems (ROMS). Two distinct circulation environments are considered: the California Current System (CCS), which is an eastern boundary upwelling regime, and the Indian Ocean (IO) characterized by an equatorial waveguide subject to the energetic seasonal reversals of the Indian and Asian Monsoons. In the case of the CCS, experiments were performed using synthetic observations, so-called Observing System Simulation Experiments (OSSEs). An extensive suite of CCS OSSEs were conducted to explore the performance of both data assimilation approaches to system configuration. For the EAKF, this includes the method for generating the seed ensemble, ensemble size, localization scales, and the length of the assimilation window. In the case of 4D-Var, the influence of assimilation window length, and the formulation of the background error covariance were explored. The performance of the EAKF was found to be influenced most by the size of the ensemble and by the method used to generate the initial seed ensemble where centering of the ensemble was found to yield improvement. For 4D-Var, the assimilation window length is by far the most critical factor, with an increase in system performance as the window length is extended. In general, the EAKF and 4D-Var systems converge to similar solutions over time, which are independent of the starting point. The EAKF employs a First-Guess at Appropriate Time (FGAT) strategy, and some experiments indicate that short FGAT windows can be problematic due to the introduction of frequent initialization shocks. While the EAKF generally out-performs 4D-Var in the OSSEs, analysis of the innovations from the two systems through time indicates that they track each other closely.

Additional Observing System Experiments (OSEs) were performed in the CCS and IO configurations of ROMS using real ocean observations. In this case, the comparison of the EAKF and 4D-Var state estimates with independent observations indicates that the EAKF and 4D-Var state estimates diverge over time, although the 4D-Var estimates are somewhat better by some measures. The relative performance of the EAKF and 4D-Var systems is similar across the wide-range circulation regimes that characterize the CCS and IO, suggesting that the results presented here are a robust indicator of expected performance in other regions of the world ocean.

Key words: Ensemble Kalman filter, 4D-Var, California Current, Indian Ocean
1. Introduction

Data assimilation in ocean models is a mainstream activity at many operational and academic centers, both at regional and global scales (e.g., Edwards et al., 2015; Martin et al., 2015; Carrassi et al., 2018; Moore et al., 2019). The development and maintenance of state-of-the-art data assimilation (DA) systems requires considerable time and resources. Therefore, open access DA tools represent a significant and important resource for the ocean modeling community-at-large. This paper compares the performance of two community tools for DA that are available in support of the Regional Ocean Modeling System (ROMS). One is an Ensemble Kalman Filter (EnKF) that forms part of the Data Assimilation Research Testbed (DART; Anderson et al., 2009), and the other is a 4-dimensional variational (4D-Var) DA system (Moore et al., 2011). The relative performance of each system is assessed for two very different circulation regimes: an eastern boundary current upwelling circulation and a tropical ocean basin. The upwelling regime considered is the California Current System (CCS), which is characterized by a pronounced seasonal cycle strongly controlled by atmospheric conditions (e.g., Hickey, 1979, 1998; Checkley and Barth, 2009). A major characteristic of the CCS is the presence of an energetic mesoscale eddy field associated with eastern boundary current instabilities, which is a challenging environment for most DA systems. The tropical ocean basin considered is the Indian Ocean, which is dominated by both the seasonally varying monsoon forcing and the equatorial waveguide (e.g., Schott and McCreary, 2001). In combination, these yield energetic local and remote influences on the circulation, which also present a significant challenge to DA.

The aim of this paper is threefold. First, it presents a review of two state-of-the-art community ocean DA tools. Second, it serves as a useful reference for users of the ROMS-DART and ROMS-4D-Var systems. And lastly, this work is a prelude to the development of a hybrid data assimilation system in the form of a marriage between the EnKF and 4D-Var, which will be the subject of a future study.

An overview of the data assimilation problem in ROMS is presented in section 2, with a particular emphasis on error covariances. The treatment of the prior error covariance matrix in the EnKF and the background error covariance matrix in 4D-Var is one of the primary factors that sets the two approaches to DA apart. The ROMS configuration for the CCS is introduced in section 3, where the EnKF and 4D-Var frameworks for this case are also described. The results of the CCS DA experiments are presented in sections 4 and 5, which are both divided into several subsections. The results of an extensive suite of Observing System Simulation Experiments (OSSEs) designed to quantify the performance of each DA system are described first using simulated observations from the actual observing network. Specifically, the influence of factors such as ensemble size, correlation and localization length-scales, assimilation window length, and boundary condition errors, on the veracity of ocean circulation estimates is explored. While these experiments provide us with a best-case scenario, they also highlight some of the limitations of the analysis methods as well as various issues that are likely to arise when real ocean observations are assimilated. The results of DA experiments using real in situ and remote sensing data from the CCS are presented in section 6, so-called Observing System Experiments (OSEs). Section 7 focusses on the results of DA experiments in the Indian Ocean. This configuration of ROMS is much more computationally demanding than the CCS, so the simulations, in this case, are limited to using real data. A summary and concluding remarks are
presented in section 8, where the findings and lessons learned from both DA systems applied to
the two circulation environments are discussed.

2. Data assimilation

We begin with a brief overview of linear approaches to the data assimilation. To this end, the
ocean state vector comprising all grid-point values of temperature, salinity, sea surface height,
and two components of horizontal velocity will be denoted as $\mathbf{x}$. If $\mathbf{x}_b$ indicates the background
state (aka the prior, first-guess or forecast), then the best, linear, unbiased circulation estimate
$\mathbf{x}_a$, also referred to as the analysis, arising from data assimilation can be expressed as:

$$\mathbf{x}_a = \mathbf{x}_b + Kd$$  \hspace{1cm} (1)

where $d = \mathbf{y} - H(\mathbf{x}_b)$ is the innovation vector representing the difference between the vector of
observations $\mathbf{y}$ and $\mathbf{x}_b$ evaluated at the observation points via the observation operator $H$, and $K$
is the Kalman gain matrix (Daley, 1991). The Kalman gain can be expressed as:

$$K = BH^T(\mathbf{H} \mathbf{H}^T + \mathbf{R})^{-1}$$  \hspace{1cm} (2)

where $\mathbf{B}$ represents the background (aka prior or forecast) error covariance matrix, $\mathbf{R}$ is the
observation error covariance matrix, and $\mathbf{H}$ is the tangent linearization of the observation
operator\(^1\). Different linear approaches to data assimilation essentially vary according to the
different assumptions made about $\mathbf{B}$ and $\mathbf{R}$. For a linear model $\mathbf{M}$, the background error
covariance matrix is evolved in time in the Kalman Filter (KF) according to:

$$F = \mathbf{MBM}^T + \mathbf{Q}$$  \hspace{1cm} (3)

where $\mathbf{Q}$ is the model error covariance matrix. For nonlinear systems, such as the atmosphere and
ocean, (3) can still be used provided that the evolution of errors is close to linear. This gives rise
to the so-called Extended KF in which the nonlinear forecast model $\mathbf{M}$ is used to evolve the state
and to evaluate $H(\mathbf{x}_b)$, while $\mathbf{M}$ represents a linear approximation of the forecast model (Lorenc,
2003a). However, as is well documented, the large dimension of geophysical problems precludes
direct use of (3), and some approximation must be made. In the Ensemble KF (EnKF) approach,$\mathbf{B}$ (and $\mathbf{F}$) are estimated from an ensemble of model solutions, $\mathbf{x}_n$. Moreover, since $\mathbf{B} \approx
(\mathbf{x}_n - \overline{\mathbf{x}})(\mathbf{x}_n - \overline{\mathbf{x}})^T$ (where an overbar denotes the expected value), the action of $\mathbf{B}$ can be
conveniently represented using the ensemble members directly. More specifically, $\mathbf{BH}^T \approx
(\mathbf{x}_n - \overline{\mathbf{x}})(H(\mathbf{x}_n) - H(\overline{\mathbf{x}}))^T$ and $\mathbf{H} \mathbf{BH}^T \approx (H(\mathbf{x}_n) - H(\overline{\mathbf{x}}))(H(\mathbf{x}_n) - H(\overline{\mathbf{x}}))^T$. A thorough
review of the KF approach can be found in Houtekamer and Zhang (2016).

In 4D-Var, the observation operator $H(\mathbf{x}_b)$ samples the background at the observation locations
in space and time over the time interval that spans the assimilation cycle, thus $H$ includes the
nonlinear forecast model. The background error covariance matrix $\mathbf{B}$ in 4D-Var is typically

\(^1\) Here and throughout, we have adopted the generally accepted compact notation for the operators $\mathbf{H}$ and $\mathbf{R}$ which
represent appropriate block diagonal matrices and each block corresponds to an observation time (e.g. see Gürol et
al., 2014).
assumed to be unchanging from the start of one cycle to the next, so there is no explicit flow
dependence of the background error covariance. However, flow dependence does enter implicitly
via the terms $H B H^T$ and $B H^T$ in (2) since $H$ includes the tangent linearization of the forecast
model. A common approach to modeling the action of $B$ is the solution of a pseudo-heat
diffusion equation (Derber and Rosati, 1989; Weaver and Courtier, 2001; Bennett, 2002).

Lorenc (2003b) provides an excellent review of the ensemble, variational, and hybrid approaches
to data assimilation. Similarly, an excellent review of nonlinear approaches to data assimilation
for geophysical problems can be found in van Leeuwen et al. (2015).

2.1 ROMS-DART

The model used here is the Regional Ocean Modeling System (ROMS), which supports both a
4D-Var (Moore et al., 2011) and an EnKF system. The EnKF is based on the Data Assimilation
Research Testbed (DART) system (Anderson et al., 2009). While there are other reported
applications of ROMS with DART (e.g., E. Curchitser, pers. comm.; Li and Tuomi, 2017), the
system described here is a recent development in which a ROMS-DART interface is now
available in the community versions of DART (https://www.image.ucar.edu/DAReS/DART) and
ROMS (http://www.myroms.org). In the current configuration, the approach used corresponds to
an Ensemble Adjustment KF (EAKF; Anderson, 2001). In addition, a First-Guess at Appropriate
Time (FGAT) approach is used in which all observation forward operators (i.e. $H(x_n)$) are
computed at the nearest model timestep but assimilated as if they were taken at the end time of
the observation window. Since ocean observations typically arrive in a quasi-continuous stream,
the FGAT approach circumvents the need to restart the KF continually, and is, therefore, more
computationally efficient. The efficacy of the FGAT approach will depend on the rate at which
the circulation evolves compared to the assimilation window length and the frequency of the
observations. While the circulation time scale will be strongly influenced by the model
resolution, in cases where the ocean circulation typically varies slowly in time, FGAT is
considered to be a good approximation. In DART, the background error covariance matrix $B$ is
estimated using an ensemble approach, and both covariance localization and covariance inflation
are employed. Both steps are necessary to alleviate the negative impacts of sampling errors
arising from a limited size ensemble. However, localization has the added benefit of rendering a
full rank approximation of $B$.

2.2 ROMS-4D-Var

ROMS also supports a 4D-Var data assimilation system, both the primal and dual formulations
(Moore et al., 2011a). Specifically, the incremental form of 4D-Var is used (Courtier et al.,
1994) in which the increments $\delta x = x - x_b$ represent the departure of state-vector from the
background. The analysis $x_a = x_b + \delta x$ is identified as that increment that minimizes the
linearized cost function:

$$J = \delta x^T B^{-1} \delta x + (H \delta x - d)^T R^{-1} (H \delta x - d).$$

(4)

It is assumed that the ocean state variables can be decomposed into a balanced and unbalanced
component in which the unbalanced components are mutually uncorrelated. Following Weaver
et al. (2005), the background error covariance $\mathbf{B}$ can then be factorized as $\mathbf{B} = \mathbf{K}_b \mathbf{C} \mathbf{C}^T \mathbf{K}_b^T$, where $\mathbf{C}$ is the univariate correlation matrix of the unbalanced variables, $\mathbf{\Sigma}$ is a diagonal matrix of standard deviations of the errors in the unbalanced variables, and $\mathbf{K}_b$ is a multivariate balance operator (Derber and Rosati, 1989). The correlation matrix $\mathbf{C}$ is modeled using the diffusion operator approach of Weaver and Courtier (2001), which requires assumptions to be made about the spatiotemporal variations of the error correlations. Since this approach is computationally costly, $\mathbf{C}$ is typically assumed to be time-invariant.

Two approaches are adopted in ROMS for specifying $\mathbf{\Sigma}$. One approach uses error statistics that reflect those associated with asymptotic climatological errors. For this reason, $\mathbf{B}$ is often referred to as a climatological background error covariance. In the second approach, an error model based on the method of Mogensen et al. (2012) is used in which background errors are assumed to be proportional to vertical derivatives of the state. This choice is based on the underlying assumption that the difference between the background value of a state variable $\phi_b$ and the true value $\phi_t$ is due to a vertical displacement of the profile, in which case, to first-order, $\phi_t(z) \approx \phi_b(z + \delta z) + (\partial \phi_b / \partial z) \delta z$, where $\delta z$ is the displacement. In other words, it is assumed that the background and true profiles have a similar shape and that the true value of $\phi$ lies somewhere in the background water column. The error in $\phi_b$ is, therefore, approximately given by $(\partial \phi_b / \partial z) \delta z$. Following Mogensen et al. (2012), we assume that the standard deviation of the error is given by $\sigma_\phi \sim |(\partial \phi_b / \partial z) \delta z|$, with the following constraints:

$$\sigma_\phi = \begin{cases} \max(\bar{\sigma}_\phi, \sigma_{\phi}^{ml}) & \text{if } z \geq -D_{ml} \\ \max(\bar{\sigma}_\phi, \sigma_{\phi}^{do}) & \text{if } z < -D_{ml} \end{cases}$$

where:

$$\bar{\sigma}_\phi = \min(|(\partial \phi_b / \partial z) \delta z|, \sigma_{\phi}^{max}).$$

The prescribed parameter $\sigma_{\phi}^{max}$ is the maximum value of $\sigma_\phi$, while $\sigma_{\phi}^{ml}$ and $\sigma_{\phi}^{do}$ are the prescribed minimum values allowed in the mixed layer and deep ocean, respectively. In ROMS, the mixed layer depth $D_{ml}$ is computed using the method described by Kara et al. (2000). Mogensen et al. (2012) apply (5) and (6) only to temperature, and a different formulation is used for other state-variables in conjunction with a balance operator. As described below, we do not use a balance operator in the experiments reported here, and (5) and (6) were used for all state variables.

3. The California Current System (CCS)

3.1 Model and observations

The ROMS-CCS system is well documented (Veneziani et al., 2009a) and has a history of 4D-Var data assimilation (Broquet et al., 2009a,b, 2011; Moore et al., 2011b; Moore et al., 2013; Gürol et al., 2014; Neveu et al., 2016; Song et al., 2016; Mattern et al., 2018). The model domain spans the region 30°N-48°N, 134°W-115.5°W, and is shown in Fig. 1a. The horizontal resolution employed in the present study is 1/10° and with 42 terrain-following $\sigma$-levels in the
vertical, which yield a vertical resolution of 0.3-8 m over the continental shelf, and 7-100 m in the deep ocean. The model was forced by surface fluxes of momentum, heat, and freshwater that were computed using atmospheric fields from the Coupled Ocean-Atmosphere Mesoscale Prediction System (COAMPS; Doyle et al., 2009) and the bulk flux formulation of Fairall et al. (2003). The model was constrained at the open boundaries by fields from the global Simple Ocean Data Assimilation (SODA) product (Carton and Giese, 2008).

Figure 1: The ROMS-CCS configuration used in the EAKF and 4D-Var experiments. The SST and SSH on a date close to the start of each experiment are shown in (a) and (b), respectively. The locations of all in situ hydrographic observations collected during the Jan-April 2003 period are shown in (c), and time series of log10 of the number of observations of each type available during each 24-hour period are shown in (d). Here and throughout the paper, units are as follows: temperature (deg C), SSH (m), and salinity (nondimensional).

The observations assimilated into the model take the form of along-track satellite observations of sea surface temperature from the AVHRR, MODIS and AMSR platforms, estimates of sea surface height (SSH) based on a daily gridded product from the Archiving, Validation, and Interpretation of Satellite Oceanographic data (AVISO; Dibarboure et al., 2011), and in situ observations taken from the quality-controlled EN3 archive (v2a) maintained by the Met Office (Ingleby and Hudson, 2007). The in situ observations are in the form of temperature and salinity measurements from various platforms, including Argo profiling floats, XBTs, CTDs, and tagged marine mammals. While we are ultimately interested in the relative performance of each assimilation system in the case of real observations, it is also very illuminating to conduct experiments using simulated observations from the same observing system. Results from both approaches are presented in sections 4, 5, and 6.

3.2 EAKF configuration

Observations were assimilated into ROMS using the EAKF-FGAT. The baseline configuration comprises a 50-member ensemble using a 1-day FGAT window with Gaspari-Cohn localization.
(Gaspari and Cohn, 1999). Spatially- and temporally-varying adaptive inflation (Anderson, 2009) using a constant inflation standard deviation of 0.6 and an inflation damping with parameter 0.9 that retains 90% of the inflation value each assimilation step was used. The sampling error correction algorithm of Anderson (2012) was applied to minimize the need for the fixed localization. Observation quality control was performed based on a comparison of the observations with the ensemble mean, and observations were rejected that depart from the ensemble mean by more than three standard deviations of the expected total error. An extensive set of sensitivity experiments was employed as detailed in Table 1 and compared to the baseline case. Also, different methods were explored for generating the initial seed population of ensemble members that are required to initialize the EAKF. The parameter choices and ensemble methods used will be discussed in more detail in the following sections.

<table>
<thead>
<tr>
<th>Experiment name</th>
<th>FGAT (days)</th>
<th>Ensemble size N</th>
<th>H localization (km)</th>
<th>V localization (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>1</td>
<td>50(51)</td>
<td>130</td>
<td>200</td>
</tr>
<tr>
<td>EAKF1a</td>
<td>2</td>
<td>50(51)</td>
<td>130</td>
<td>200</td>
</tr>
<tr>
<td>EAKF1b</td>
<td>4</td>
<td>50(51)</td>
<td>130</td>
<td>200</td>
</tr>
<tr>
<td>EAKF2a</td>
<td>1</td>
<td>10(11)</td>
<td>130</td>
<td>200</td>
</tr>
<tr>
<td>EAKF2b</td>
<td>1</td>
<td>30(31)</td>
<td>130</td>
<td>200</td>
</tr>
<tr>
<td>EAKF3a</td>
<td>1</td>
<td>50(51)</td>
<td>260</td>
<td>200</td>
</tr>
<tr>
<td>EAKF3b</td>
<td>1</td>
<td>50(51)</td>
<td>65</td>
<td>200</td>
</tr>
<tr>
<td>EAKF4a</td>
<td>1</td>
<td>50(51)</td>
<td>130</td>
<td>-</td>
</tr>
<tr>
<td>EAKF4b</td>
<td>1</td>
<td>50(51)</td>
<td>130</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 1: The table shows the parameters used in each EAKF experiment, and the experiment designation. As described in section 4.1.1, two different approaches were employed to generate the initial ensemble perturbations, which are referred to as methods A and B. In addition, each set of perturbations were used in two different ways, either once only or in ± pairs to center the resulting ensemble. In the text, the letters A or B are appended to the experiment name shown in column 1 to indicate the ensemble perturbations used, and a prefix C is used to denote a centered ensemble. For example, A_EAKF1a denotes configuration EAKF1a initialized using the uncentered ensemble generation method A, while CB_EAKF1a denotes the same experiment configuration using a centered ensemble generation method B. The numbers in parentheses in column 3 are the ensemble sizes used when centering is employed.

3.3 4D-Var configuration

The 4D-Var configuration is well documented elsewhere (e.g., Neveu et al., 2016; hereafter, N16), so only a brief description will be given here. The baseline 4D-Var configuration is based on that of N16 and comprises 8-day assimilation windows, using only the model initial conditions as the control variable. In all experiments, the dual 4D-Var formulation was used in which the cost function (4) is minimized by conjugate gradient searches in observation space (e.g. Courtier, 1997). The background error standard deviations, $\Sigma$, were based on either climatological standard deviations (denoted $\Sigma_c$) of a long run of the model without data assimilation, varying from month-to-month, or on (4) and (5) (denoted $\Sigma_m$). The background error correlations were assumed to be isotropic, homogeneous, and time-invariant, and a range of values was considered in the experiments reported below. In all cases, two outer-loops and seven inner-loops were used, which requires approximately 90% of the computational effort of the
baseline EAKF experiment when running in serial over the same time interval. Increasing the
number of outer-loops was found to have minimal influence on the results presented here. Our
experience to date with the balance operator in ROMS has shown that its performance is
somewhat mixed, so as in N16, the balance operator was not used, and we rely instead on natural
dynamic adjustments via the term $HBH^T$ to provide cross-covariance information. Table 2
summarizes the 4D-Var experiments performed as part of this study.

<table>
<thead>
<tr>
<th>Experiment name</th>
<th>Window (days)</th>
<th>Scaling ($\alpha$)</th>
<th>H decorrelation (km)</th>
<th>V decorrelation (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M4DVAR1</td>
<td>1</td>
<td>1</td>
<td>25</td>
<td>90</td>
</tr>
<tr>
<td>M4DVAR2</td>
<td>2</td>
<td>1</td>
<td>25</td>
<td>90</td>
</tr>
<tr>
<td>M4DVAR4</td>
<td>4</td>
<td>1</td>
<td>25</td>
<td>90</td>
</tr>
<tr>
<td>M4DVAR8</td>
<td>8</td>
<td>1</td>
<td>25</td>
<td>90</td>
</tr>
<tr>
<td>C4DVAR8h25</td>
<td>8</td>
<td>-</td>
<td>25</td>
<td>90</td>
</tr>
<tr>
<td>C4DVAR8h50</td>
<td>8</td>
<td>-</td>
<td>50</td>
<td>90</td>
</tr>
<tr>
<td>C4DVAR8h75</td>
<td>8</td>
<td>-</td>
<td>75</td>
<td>90</td>
</tr>
<tr>
<td>C4DVAR8v30</td>
<td>8</td>
<td>-</td>
<td>25</td>
<td>30</td>
</tr>
<tr>
<td>C4DVAR8v60</td>
<td>8</td>
<td>-</td>
<td>25</td>
<td>60</td>
</tr>
<tr>
<td>C4DVAR8v90</td>
<td>8</td>
<td>-</td>
<td>25</td>
<td>90</td>
</tr>
<tr>
<td>M4DVAR8a</td>
<td>8</td>
<td>0.25</td>
<td>25</td>
<td>90</td>
</tr>
<tr>
<td>M4DVAR8b</td>
<td>8</td>
<td>0.5</td>
<td>25</td>
<td>90</td>
</tr>
<tr>
<td>M4DVAR8c</td>
<td>8</td>
<td>1.5</td>
<td>25</td>
<td>90</td>
</tr>
</tbody>
</table>

Table 2: The table shows the parameters used in each 4D-Var experiment, and the experiment designation. As
described in section 4.2.3, two different approaches were employed to model the background error standard
deviations. The prefix C denotes experiments that use the climatological standard deviations $\Sigma_c$, while the prefix M
denotes experiments that use the error model $\Sigma_m$ described by equation (4) and (5). The data assimilation window
length is indicated in column 2, and the scaling factor $\alpha$ that was applied to the standard parameter values in (4) is
indicated in column 3, where applicable. Columns 4 and 5 show the horizontal and vertical decorrelation lengths
assumed for the background errors in all state variables.

A quality control check was performed on the observations based on the background following
the approach of Andersson and Järvinen (1999), with a rejection criterion similar to that
employed by the EAKF.

3.4 A single profile illustration

One of the advantages of the EAKF is that the ensemble provides flow-dependent covariance
information about the so-called “errors of the day,” which is propagated from one cycle to the
next. While on the other hand, 4D-Var relies on climatological covariance information, it too
employs implicit flow-dependent covariances via the term $HBH^T$. However, this information is
replaced at the beginning of each cycle. Information about both the errors of the day and
climatological error is useful, and recent work indicates that a marriage of the two is beneficial
(Lorenc et al., 2015). Before proceeding, it is illuminating to reflect on the similarity and
differences in covariance information resulting from the two approaches by considering the
increments that arise from the assimilation of a single temperature profile since this will provide
a useful reference point for the experiments in later sections.
Figure 2: The background SSH in the vicinity of central California on 8 Jan 2003. The triangle indicates the position of the single vertical temperature profile assimilated on the same day. 4D-Var analysis increments on 8 Jan 2003 of (b) SST and (c) sea surface salinity (SSS). (d) A zonal, vertical section of the 4D-Var temperature increment at the same latitude as the observation point. EAKF increments on 8 Jan 2003 of (e) SST and (f) SSS. (g) A zonal, vertical section of the EAKF temperature increment at the same latitude as the observation point.

The location chosen for the simulated vertical temperature profile is close to San Francisco Bay, as indicated in Fig. 2a. Figure 2a also shows the SSH in the vicinity of the region and reveals the nature of the mesoscale circulation at the observation time. A full water column profile of temperature was assimilated into the model on 8 Jan 2003 using the EAKF and 4D-Var. In the 4D-Var example, the assimilation window spans the 4-day interval 4-8 Jan for this case, and the end of the window corresponds to the observation time. The background state used in 4D-Var and the ensemble mean prior to the EAKF update are identical, and as such, we are considering the first assimilation cycle of both systems. The parameters used in the EAKF and 4D-Var experiments correspond to those of CB_EAKF1b (Table 1) and M4DVAR4 (Table 2), respectively. Figure 2 shows the increments in SST and sea surface salinity (SSS) on 8 Jan from each experiment. In the case of the EAKF, these are the increments that are applied at the end of the FGAT window on 8 Jan, while in the case of 4D-Var the increment is computed on 4 Jan and propagated forward in time to 8 Jan. The EAKF increments (Figs. 2e and 2f) are primarily negative, and coherent over a distance ~100 km. The horizontal extent of the increment reflects the localization radius used, which in this experiment was 130 km, and should not be confused with any imposed correlation length. The 4D-Var increments, on the other hand (Figs. 2b and 2c), have a more complicated structure and comprise both positive and negative values. The horizontal scale of the increments is commensurate with the horizontal length scale that is used to model $B$, which is 25 km in this case. The complex structure is due to the time evolution of $x_b$ via the covariance $BB^T$ and the term $BH^T$, and we expect similar structures to develop in the EAKF over time as the filter spins-up. The vertical structure of the temperature increments on 8 Jan along a zonal section passing through the observation point is shown in Figs 2d and 2g for
the EAKF and 4D-Var, respectively. Overall, the vertical structure and vertical extent are similar in both cases, although the horizontal length-scale of the EAKF increment is generally larger than that of 4D-Var at all depths.

4. CCS Twin Experiments

Three time-intervals were considered for the CCS experiments corresponding to (a) 4 Jan - 14 April 2003, (b) 4 Jan - 31 Dec 2003 and (c) 4 Jan 2003 - 31 Dec 2004. The time interval (a) was used to perform an extensive suite of experiments to quantify the performance of the EAKF and 4D-Var system to variations in the system configuration. Time intervals (b) and (c) were used for longer experiments using selected system configurations that will be discussed in later sections. Experiments spanning time interval (a) will be considered first of all, and a time series of the number of the different types of observations is shown in Fig. 1d, while the distribution of all of the in situ observations collected during this period is shown in Fig. 1c. Two series of experiments were performed with the EAKF and 4D-Var in the CCS: in the first case, simulated observations taken from a nature run were assimilated. In contrast, in the second series, real ocean observations were assimilated.

The first series of experiments that will be described are those in which simulated observations sampled from the actual observing array were assimilated into the model. This type of analysis represents a traditional Observing System Simulation Experiment (OSSE), and the simulated observations were generated by sampling a reference solution, often referred to as a nature run. In this case, the nature run was a 13-year run of the model without data assimilation spanning the period 1999-2012 using the COAMPS fields to compute surface forcing and the SODA open boundary conditions. Simulated observations were taken from the nature run, and unbiased random observation errors were added, sampled from a Gaussian distribution with covariance $R$. Since the same model, and surface and boundary inputs were used for the data assimilation experiments, the model can be considered as perfect, in which case $Q = 0$ in (3). The 4 Jan 2003 initial condition used for the data assimilation experiments will be denoted as $x_0$, and two cases were considered: (i) $x_0^1$ taken from the 1980-2010 4D-Var reanalysis product of N16 (referred to as WCRA31) in which all available real observations were assimilated, and (ii) $x_0^1$ taken from a run of the model without data assimilation spanning the period 1980-2003. The ocean surface forcing fields in both cases were derived from a combination of ECMWF products and the Cross-Calibrated Multi-Platform winds (CCMP). The ocean state $x_0$ on 4 Jan 2003 differs significantly in both cases from the nature run solution on the same day. Since $x_0^1$ also represents our best estimate of the true ocean state on this date, it is also the nominal initial state used for the real observation experiments described in section 6. Representative fields of SST and SSH on a date close to the start of each experiment are shown in Fig. 1a and Fig. 1b, respectively.

4.1 EAKF experiments

4.1.1 Ensemble generation strategies

Several different strategies were investigated for generating perturbations for the initial ensemble on 4 Jan 2003. In the first approach, referred to as Method A, analysis increments $\delta \mathbf{x}_j$ from January were randomly selected from the WCRA31 reanalysis of N16. An ensemble of initial
conditions was then created according to $x_0 + \delta x_j$. The rationale for this approach is that the $\delta x_j$ should be representative of uncertainties in the initial conditions, and perturbations chosen this way also have the advantage that they are dynamically balanced (although this does not guarantee that $x_0 + \delta x_j$ will be in balance). The second strategy used, Method B, was to sample the nature run at random times during January of different years, excluding of course Jan 2003, the starting point for the data assimilation experiments. The difference, $\Delta x_j$, between randomly chosen pairs of these January solutions were then used as initial perturbations for the ensemble members, according to $x_0 + \Delta x_j$. Two additional variants of Methods A and B were also considered, in which each perturbation was used twice according to $x_0 \pm \delta x_j$ or $x_0 \pm \Delta x_j$. By including the unperturbed ensemble member, this approach yields an ensemble mean initial condition equal to the unperturbed ensemble member, $x_0$, and the ensemble is centered about $x_0$. Centering of the ensemble members in this way can be beneficial because it ensures that no bias is introduced as a result of the limited size ensemble, and it provides for a fair comparison with the 4D-Var experiments (see section 5) which use $x_0$ as the prior for the first assimilation cycle. In all of the centered EAKF experiments, the ensemble size has one additional member (from the unperturbed initial condition $x_0$) compared to the uncentered counterpart. It should be noted, however, that centering the ensemble in this way leads to linear dependence between corresponding pairs of ensemble perturbations. Nonetheless, as we shall see, centering has a positive impact on the performance of the EAKF in the case where a long spin-up of the filter is not possible due to the limited duration of the observation record.

It should also be noted that all ensemble members are subjected to the same atmospheric conditions and open boundary conditions. There will be variations in the ocean surface fluxes across the ensemble members due to differences that arise in SST, but the diversity in ocean surface forcing fields will still be relatively low. Even though we are assuming that the state of the atmosphere and the ocean at the open boundary conditions is known perfectly in the OSSEs, it is generally accepted that even in this case, perturbing the forcing and boundary conditions of each ensemble member can be an effective strategy to account for sampling errors due to the limited ensemble size. A lack of such diversity in the surface fluxes can significantly influence the performance of the EAKF, as we shall see in later sections. Previous studies using the same ROMS configuration by Veneziani et al. (2009b) have demonstrated that the circulation is relatively insensitive to variations in the open boundary conditions, so the lack of diversity in the conditions imposed at the open boundaries across the ensemble may be less of an issue.

As described in section 2.1, the FGAT approach is employed in ROMS-DART, and adjustments made to each ensemble member at the end of the assimilation window. In the case of 4D-Var, however, the observations were assimilated at the actual observation times. Table 1 summarizes the various parameter combinations used in the EAKF for each of the ensemble generation methods. As noted earlier, the baseline experiment is the one to which all others will be compared.
Figure 3 shows a time series of the root mean square (RMS) error in the ensemble mean temperature ($T$), salinity ($S$), the meridional component of velocity ($v$) and SSH ($\zeta$) for all model grid-points, where the nature run is taken as the truth, for a series of experiments using the baseline configuration. Also shown is the case where no data were assimilated, where the model was initialized from $x_0$ and run forward in time subject only to the COAMPS-derived surface forcing and the SODA open boundary conditions. Figure 3 indicates that the baseline experiment, regardless of the ensemble generation method, can slowly recover the nature run solution over time, with immediate improvements in $T$ (Fig. 3a) and $\zeta$ (Fig. 3d) compared to the no assimilation case. The estimates for $S$ (Fig. 3b) and $v$ (Fig. 3c; and for zonal velocity also, not shown) are initially only marginally better than the no assimilation case, although method B without centering leads to an initial degradation of $S$ (Fig. 3b). The small increases in RMS error in $\zeta$ at the end of each calendar month (Fig. 3d) are because no SSH observations were available on those days. However, the system immediately recovers the following day when new observations become available.

Method A typically leads to a slower reduction in error over time than method B, although this is due to the difference in the amplitude of the ensemble perturbations. The perturbations used in method A drawn from the WCRA31 reanalysis are typically ~5 times smaller than those generated by randomly sampling the model state using method B. The results from a repeat of the method B baseline experiment with the ensemble perturbations scaled by a factor of 0.2 is also shown in Fig. 3 and has a convergence rate similar to that of method A, indicating that a more aggressive choice of ensemble perturbation amplitudes can significantly accelerate...
recovery of the truth. The best baseline case of all results from a centered version of method B, which converges somewhat faster than the uncentered case for $T$ and $S$.

![Figure 4: Rank histograms based on SST observations on three different days for (a-c) the method B baseline case with ensemble perturbations scaled by 0.2, and (d-f) the centered method B baseline case. Note the expanded scale of the ordinate in (a).](image)

### 4.1.2 Ensemble spread and covariance inflation

Rank histograms (aka Talagrand diagrams) provide a quantitative assessment of the behavior of the baseline case. In a rank histogram, the observation values are sorted into bins that are based on the distribution of the values of the ensemble members sampled at the observation points. Rank histograms, therefore, are one measure of the reliability of the ensemble. In other words, if the true state of the system, as revealed by the observations, is a likely member of the ensemble, then there is an equal likelihood of finding any observation in each of the bins. Therefore, an indicator of a reliable ensemble is a flat rank histogram. When producing rank histograms for observed quantities, the ensemble members were perturbed with independent random draws from the observation error distribution; since this distribution is Gaussian here, this is equivalent to random draws from the likelihood (Anderson 1996). Figure 4 shows rank histograms based on the SST observations on three different dates for the method B baseline case with ensemble perturbations scaled by 0.2 (Figs. 4a-c) and the centered method B baseline case (Figs. 4d-f). The dates chosen correspond to assimilation cycles near the beginning (4 Jan, Figs. 4a,d), middle (25 Feb, Figs. 4b,e), and end (17 April, Figs. 4c,f) of the experiment period. In the case of the reduced amplitude method B ensemble, Fig. 4a shows that in January, there is insufficient spread in the ensemble, as indicated by the U-shape of the histogram. However, as the experiment proceeds, the adaptive covariance inflation scheme that is employed can adjust the ensemble spread so that by April (Fig. 4c), the rank histogram is essentially flat, indicating that the ensemble is reliable. Conversely, Fig. 4d shows that the centered method B baseline ensemble
enjoys too much spread, as indicated by the inverted U-shape of the rank histogram. In this case, also, the adaptive inflation can effectively adjust the ensemble leading to a reliable ensemble by as early as February (Fig. 4e).

Figure 5: SST inflation factors on 4 Jan and 17 April for (a,b) the method B baseline with ensemble perturbations scaled by 0.2, and (c,d) the centered method B baseline case. Note the different color bar scales used in each case.

Figure 6: Time series of the root mean square inflation factors for (a) SST and (b) SSH for the method B baseline with ensemble perturbations scaled by 0.2 (black line), and the centered method B baseline case (red line).

The inflation factors for SST on 4 Jan and 17 April for the same two baseline experiments are shown in Fig. 5. In the case of the small-amplitude ensemble, Fig. 5a indicates that the inflation factors are quite large in Jan over much of the domain. However, as the ensemble becomes more reliable, the inflation factors decrease significantly (Fig. 5b). A rapid adjustment in the inflation
factors and ensemble spread occurs over just a few days, as illustrated in Fig. 6a black line, which shows a time series of the mean inflation factor for SST for the same experiment. After a rapid fall in early Jan, the inflation factor continues a slow, gradual decline toward unity. The same behavior is present in other fields, such as SSH, as shown in Fig. 6b. In the centered method B case, the presence of too much spread in the ensemble results in inflation factors close to unity (red line) over most of the domain (Figs. 5c and d). In this case, there is little change in the inflation factors over time, as illustrated in Fig. 6, which reveals a slight decrease for those of SST and SSH from a value just above one toward unity.

The ensemble of the method A baseline exhibits similar behavior to the small amplitude method B case, while the full amplitude method B baseline is like the centered method B case (not shown).

4.1.3 FGAT window length

The influence of increasing the length of the FGAT time interval on the quality of the circulation estimates was explored in experiments EAKF1a and EAKF1b, where the window length was increased first to 2-days then to 4-days. The time series of the RMS errors in temperature for experiments CB_EAKF1a and CB_EAKF1b are shown in Fig. 7a and indicate that increasing the length of the FGAT window leads to no appreciable change in the quality of the circulation estimates compared to the baseline case. Other fields exhibit similar behavior (not shown).

Figure 7: Time series of RMS errors in ensemble mean temperature for EAKF cases involving variations in (a) FGAT window length (2-day, red; 4-day black), (b) ensemble size (11 members, red; 31 members, black), (c) horizontal localization radius (260 km, red; 65 km, black), and (d) vertical localization radius (no vertical localization, red; 100 m, black). In each case, the RMS error for the centered 51-member method B baseline (blue dashed line) and no assimilation case (black dashed line) are also shown.
4.1.4 Ensemble size

The impact of the ensemble size on the veracity of the circulation estimates was quantified by experiments EAKF2b and EAKF2a, where the ensemble size was reduced from 51 to 31 and 11 members, respectively. The time series of the RMS errors in $T$ for experiments CB_EAKF2a and CB_EAKF2b are shown in Fig. 7b and indicate that reducing the ensemble size to 11 members degrades the circulation estimates significantly compared to the baseline, although 31 members is only marginally worse than the baseline case. Similar results were found for other fields (not shown).

4.1.5 Localization radius

The influence of the horizontal and vertical localization scale on the circulation estimates was explored in experiments EAKF3a, EAKF3b, EAKF4a, and EAKF4b. As noted in Section 3.2, a Gaspari-Cohn localization function was used in all experiments, which in the baseline simulation has a half-width of 130 km in the horizontal and 200 m in the vertical. In experiments EAKF3a and EAKF3b, the horizontal localization half-width was increased by a factor of 2 and decreased by a factor of 2 respectively, while keeping the vertical localization half-width scale fixed. Figure 7c shows time series of RMS errors in temperature from experiments CB_EAKF3a and CB_EAKF3b and reveals that, in general, increasing the horizontal localization scale degrades considerably the quality of the circulation estimates compared to the baseline case by admitting spurious correlation information in $M$. While reducing the horizontal localization scales marginally improves the temperature estimates during Jan, the estimates are degraded during later months suggesting the baseline choice of 130 km is more appropriate.

In experiment CB_EAKF4a, the vertical localization was deactivated altogether, and Fig. 7d indicates that this significantly degrades the circulation estimates by allowing spurious correlations to influence the entire water column. Stratification dictates that some level of vertical localization is necessary since errors in $T$ and $S$ above the thermocline are unlikely to be significantly correlated with errors in the deep ocean. Reducing the localization length scale in the vertical, as in CB_EAKF4b, leads to a marginal degradation of the temperature field compared to the baseline case, suggesting that the baseline value of 200 m is a reasonable choice. Other fields exhibit similar behavior (not shown).

4.2 4D-Var experiments

The simulated observations were also assimilated using 4D-Var and a variety of system configurations that are summarized in Table 2. In the experiments using background error standard deviations computed using (5), the default parameters are listed in Table 3. The default parameters lead to peak values in the standard deviations of ~0.6 Cº, ~0.1, and ~0.1 ms$^{-1}$ for temperature, salinity and velocity respectively which are about a factor two larger than the typical peak prior ensemble spreads in the EAKF experiments.
Table 3: The default parameters used in the background standard deviation equations (5) and (6).

<table>
<thead>
<tr>
<th></th>
<th>(\sigma_{max})</th>
<th>(\sigma_{ml})</th>
<th>(\sigma_{do})</th>
<th>(\delta z)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature</td>
<td>0.66 C</td>
<td>0.1 C</td>
<td>0.04 C</td>
<td>40 m</td>
</tr>
<tr>
<td>Salinity</td>
<td>0.05</td>
<td>0.1</td>
<td>0.056</td>
<td>40 m</td>
</tr>
<tr>
<td>Velocity</td>
<td>0.12 ms(^{-1})</td>
<td>0.10 ms(^{-1})</td>
<td>0.04 ms(^{-1})</td>
<td>500 m</td>
</tr>
<tr>
<td>SSH</td>
<td>0.05 m</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

4.2.1 Assimilation window length

The influence of the length of the 4D-Var data assimilation window was explored in experiments M4DVAR1, M4DVAR2, M4DVAR4, and M4DVAR8 (see Table 2) which all use \(\Sigma_m\) based on the default parameters for (5) and (6) (Table 3). In each experiment, the ocean analysis \(x_a\) at the end of each 4D-Var cycle was used as the background initial condition \(x_b\) for the next assimilation cycle. The time series of the RMS circulation errors for each of the 4D-Var experiments are shown in Fig. 8 and indicate that, in general, the performance of the system improves as the length of the assimilation window increases. Except for M4DVAR1, the RMS errors are generally lower than the case where no data are assimilated. The poor performance of M4DVAR1 is associated mainly with large errors immediately below the thermocline (not shown here). For SSH, the RMS errors all converge to approximately the same level after ~2 months, although initially shorter assimilation windows are better at recovering the true SSH.

![Figure 8: Time series of the RMS error in (a) temperature, (b) salinity, (c) meridional velocity and (d) SSH for experiments using 1-day (blue line), 2-day (red line), 4-day (green line) and 8-day (black line) data assimilation windows. Also shown is the case where no data were assimilated (black dashed line).](image)

The steady improvement in the performance of the 4D-Var system with increasing window length has also been reported elsewhere (e.g., Pires et al., 1996; Swanson et al., 1998). One important factor is associated with the implicit flow-dependence of the background error covariance in observation space described by \(HBH^T\) and \(BH^T\) in (2). Recall that in 4D-Var \(H\) represents the tangent linear ROMS sampled at the observation times and locations while \(H^T\) is
the adjoint of ROMS forced at the observation points. Both the tangent linear and adjoint models are linearized about the time-evolving background circulation $x_b(t)$. Thus, extending the length of the 4D-Var assimilation window allows the time-evolution of background circulation to exert an ever-increasing influence on the specified background error covariance matrix $B$. The benefit of this appears to outweigh the potentially negative influence of violating the tangent linear assumption as the window length increases (Pires et al., 1996). In addition, increasing the window length increases the number of observations that will be available to control the ocean state, which will also be an important contributing factor.

4.2.2 **Horizontal and vertical correlation lengths**

As discussed in section 2.2, the background error correlation matrix is modeled as a diffusion operator. This approach requires that typical horizontal and vertical correlation lengths for the background errors must be chosen. While in reality these length scales will vary spatially and temporally, at present, ROMS supports only user-defined homogeneous and isotropic correlation functions, so a single length scale must be chosen for each state variable that is representative of conditions over the entire domain. This is clearly an over-simplification of the complex error covariances that will exist in reality. However, an alternative interpretation of $B$ is that it acts as a smoothing penalty function in the 4D-Var cost function. The cost function in (4) is an example of a penalty function used in optimal control theory where smoothing constraints, referred to as Tikhonov regularization (Bertsekas, 1982), are often included to penalize the smoothness of the solutions. Thus, $B$ can be thought of as a means of controlling the scale and smoothness of the resulting 4D-Var increments. Appropriate background error decorrelation length scales can be estimated using semi-variogram analyses (Banerjee et al., 2004; Matthews et al., 2011).

With this in mind, Fig. 9a shows the impact on the efficacy of the circulation estimates for the temperature to variations in the horizontal correlation length while keeping the vertical correlation length fixed at 90 m. In these experiments, the climatological standard deviations $\Sigma_c$ were used in conjunction with an 8-day assimilation window. Figure 9a indicates that increasing the assumed horizontal length scale of the background error (i.e., increasing the degree of smoothing of the increments) leads to a significant degradation of the 4D-Var circulation estimates. Other state variables exhibit similar behavior (not shown).

The influence of the vertical correlation on the quality of the 4D-Var temperature estimates is shown in Fig. 9b for the case using $\Sigma_c$, a fixed horizontal correlation length of 25 km, and an 8-day assimilation window. In this case, increasing the vertical correlation length (i.e., the degree of smoothing in the vertical) leads to significant improvement in the quality of the estimates. Similar results were found for the other state variables (not shown).

4.2.3 **Background error standard deviations**

The influence of the choice of parameters in the error model (5) and (6) for $\Sigma_m$ on the RMS errors of the temperature is summarized in Fig. 9c. In each case, the default parameters in Table 3 were all rescaled by a factor $\alpha$ except for $\delta z$, where $\alpha$ was varied between 0.5 and 1.5 (see Table 2). In each case, a horizontal correlation length of 25 km, a vertical correlation length of 90 m, and an 8-day assimilation window were used. In these experiments, the data assimilation
methodology was modified, so that during the first 4D-Var cycle starting on 4 Jan, $\Sigma_c$ was used. This procedure is based on the assumption that at the time of the first cycle, the errors in the background initial conditions are probably best described by climatology since no data have yet been assimilated. This inference was confirmed by computing the innovation statistics described by Desroziers et al. (2005). Specifically, $\sigma_b^2 = (y_a - y_b)^T(y - y_b)/N$ was evaluated where the overbar denotes the expected value, $y$ is the vector of observations, and $y_a$ and $y_b$ are respectively the analysis and background evaluated at the observation locations and observation times, and $N$ is the number of observations. Desroziers et al. (2005) show that if $B$ and $R$ are chosen consistently, then $HBH^T \approx (y_a - y_b)(y - y_b)^T$. Therefore, for a given subset of observations (e.g., temperature), $\sigma_b^2$ is the expected value for the background error variance. The statistic $\sigma_b^2$ is, therefore, a useful measure of the consistency between the expected background error based on the a priori choice of $B$ and $R$, and the specified values of error variance in $\Sigma_c$ or $\Sigma_m$. Figure 9d shows the time series of $\log_{10}(\sigma_b^2)$ and $\log_{10}(\text{Tr}(\Sigma_c^2)/N)$ for all SST observations during each 4D-Var cycle. During the first 4D-Var cycle, the expected background errors in SST are consistent with those specified from climatology. However, during subsequent cycles, the expected error is an order of magnitude lower than that given by $\Sigma_c$, showing that the choice of $\Sigma_c$ is inappropiate. This adjustment occurs because, after the first assimilation cycle, the model SST is much closer to the truth than a randomly chosen ocean state. Figure 9d shows the situation in experiment M4DVAR8 where $\Sigma_c$ is used during the first 4D-Var cycle, and $\Sigma_m$ with the default parameters is used thereafter. In this case, the agreement between $\sigma_b^2$ and $\text{Tr}(\Sigma_m^2)$ is greatly improved, indicating that $\Sigma_m$ is a more appropriate choice of background errors after the first cycle.

![Figure 9](image-url)
also shown for comparison (black dashed). (d) Time series of \( \log_{10}\sigma^2 \) (solid black line) and \( \log_{10}(\text{Tr}(\Sigma^2)/N) \) (black dashed line) for SST observations for experiment C4DVAR8h25, and of \( \log_{10}\sigma^2 \) (solid red line) and \( \log_{10}(\text{Tr}(\Sigma^2)/N) \) (red dashed line) for experiment M4DVAR8.

Figure 9c indicates that the efficacy of the 4D-Var circulation estimates of temperature is relatively insensitive to the scaling that is applied to the parameters in (5) and (6).

5. A Comparison of CCS EAKF and 4D-Var

In this section, we compare the performance of representative best-case EAKF and 4D-Var configurations during the extended period that spans all of 2003. The RMS errors of the CB baseline EAKF and M4DVAR8 experiments for the period Jan-Dec 2003 are shown in Figs. 10a-c for cases initialized using \( x_0 \) and \( x_{ii}^{\text{opt}} \) on 4 Jan 2003. Centering the ensemble in the case of the EAKF ensures that the ensemble mean at the beginning of the first assimilation cycle is identical to the 4D-Var background. Figures 10a-c show that, overall, the level of error in the circulation estimates is similar for both the EAKF and 4D-Var. The errors in the EAKF estimates are generally less sensitive to the starting point \( x_0 \) than those in the 4D-Var cases. 4D-Var errors are typically larger for the case starting from \( x_{ii}^{\text{opt}} \), although, by the end of Dec 2003, they are similar to those starting from \( x_0 \). Figures 10a-c therefore indicate that after about 12 months, the EAKF and 4D-Var circulation estimates display a similar degree of accuracy regardless of the first-guess estimate \( x_0 \). The temperature errors in all cases increase during the spring and summer, mirroring the case where no data are assimilated.
The fact that the EAKF and 4D-Var analysis sequences converge to similar circulations when starting from different first-guess fields for the first assimilation cycle, indicates that both methods are performing as expected and that their respective implementations are robust. In the case of 4D-Var, it also suggests that there are apparently no multiple minima of the cost function associated with different solution trajectories through time.

Figures 10a-c also show the envelope bounded by the maximum and minimum RMS error of each of the 51 members of the EAKF ensembles. While it is gratifying to note that the ensemble mean is more accurate than any single ensemble member, the broadening of the envelope throughout the year in both EAKF experiments is troubling and indicative of an over-dispersive ensemble. This is explored further in Fig. 11, which shows the RMS errors in $T$ and $S$ averaged over different depth ranges for the CB EAKF baseline and M4DVAR8 starting from $x_0^{10}$. Over the upper 100 m (Figs. 11a and 11b), which includes the thermocline, the errors in $T$ and $S$ are similar for both cases, and mirror the errors computed over the full depth. However, the temperature errors are a little higher in the case of 4D-Var during summertime. The spatial structure of the errors in $T$ averaged over the same depth range are shown in Figs. 12a-c on 7 July 2003, and reveal that errors are primarily associated with mesoscale features. This suggests that the increase in errors during the spring and summer may be associated with elevated eddy activity downstream of Cape Mendocino (cf. Fig. 1) that accompanies a strengthening of the California Current during upwelling season, a feature that is also observed (Kelly et al., 1998).

Immediately below the thermocline, 4D-Var generally performs better than the EAKF in the case of $T$ (Fig. 11c), while the errors in $S$ are similar (Fig. 11d). Figures 12e-g show the spatial structure of the errors in $T$ for this depth range on 7 July 2003 and indicate that the EAKF generally performs worst equatorward of 40°N. The divergence in the ensemble of the EAKF evident in Figs. 10a-c is associated with a significant increase in errors of $T$ and $S$ at mid-water column depths and below, as revealed by Figs. 11e and 11f. Starting around spring, the errors in both $T$ and $S$ exhibit an upward trajectory that continues for the rest of the year. The spatial structure of the errors in $S$ on 14 Dec 2003 averaged over 1000-2000 m are shown in Figs. 12i-k, and clearly illustrate the development of significant salinity errors in several locations during the EAKF case (Fig. 12k).
Figure 11: Time series of RMS errors in the ensemble mean temperature and salinity for the CB baseline EAKF case (blue line), the M4DVAR8 case (red line) and CB_EAKF1b (solid black line) for the period Jan-Dec 2003 computed over different depth ranges: (a,b) 0-100 m, (c,d) 100-200 m and (e,f) 1000-2000 m. Also shown for comparison is the case where no data were assimilated (black dashed line). The starting initial condition on 4 Jan 2003 in all cases was $x_0$.

The error growth in the CB baseline EAKF experiment continues beyond the end of 2003, and eventually, the computation becomes unstable, as illustrated in Figs. 10d-f, which shows RMS errors for selected experiments for the extended period Jan 2003-Dec 2004. By August 2004, the errors in all field estimates of the CB_EAKF baseline have increased significantly before the calculation terminates. On the other hand, the 4D-Var errors continue to decline slowly in the case of $S$ and $v$, while in $T$, they exhibit a pronounced seasonal cycle with maximum errors during the summer each year.

We set forth two hypotheses to explain the divergence of the EAKF ensemble and eventual failure of the baseline experiment. The first is associated with initialization shocks that are introduced into the model at the start of each 1-day FGAT assimilation cycle. No attempt was made to dynamically balance the EAKF field estimates at the start of each cycle, and it is known that inertia-gravity waves that are generated in this system, as a result, take $\sim$24-48 hours to dissipate (Raghukumar et al., 2015). Consequently, with a 1-day FGAT interval, insufficient time elapses to significantly dissipate the inertia-gravity waves generated during the update of the previous assimilation cycle before the circulation fields are updated again. The accumulation of gravity wave energy over time could, therefore, be responsible for the divergence of the ensemble seen in Figs. 10a-c. Our second hypothesis relates to the initialization of the vertical mixing coefficients of temperature and salinity ($A_{T,S}$) and velocity ($A_v$) after each EAKF update. In all of the cases presented here, the generic length scale formulation of Umlauf and Burchard
(2003) was used to compute time-dependent vertical mixing coefficients. This approach necessitates the solution of additional equations for turbulent kinetic energy (TKE) and the mixing length scale (MLS). Since TKE and MLS are prognostic variables, they should be considered as part of the state-vector that is updated by both the EAKF and 4D-Var. However, this is not done in the current formulation of ROMS, and as such, no attempt is made to update TKE and MLS during each assimilation cycle. During the CB_EAKF baseline experiment, $A_{T,S}$ and $A_p$ were found to increase with time, reaching very large values at the time that the calculation failed. It is possible that this behavior is caused by the lack of updates of TKE and MLS. On the other hand, it could also be a symptom of the development of an ever-increasingly unstable water column as the deep ocean salinities and temperatures are exposed to large anomalies (cf. Fig. 12k).

Figure 12: Maps of the RMS errors averaged over different depth ranges on various dates for cases where no data are assimilated, M4DVAR8, the CB_EAKF baseline, and CB_EAKF1b. (a-d) temperature errors, 0-100 m, 7 July, (e-h) temperature errors, 100-200 m, 7 July, and (i-l) salinity errors, 1000-2000 m, 4 Dec. Note that the range of the color bar is not the same in all like panels.

Based on these two hypotheses, the EAKF experiment using a 4-day FGAT window, CB_EAKF1b (see Table 1), was extended to the end of 2004. Increasing the time interval between EAKF updates in this way will allow sufficient time for the inertia-gravity waves.
excited as part of the initialization shock to dissipate. The resulting RMS errors for CB_EAKF1b are shown in Figs. 10d-f for the case using $x_0^i$. In this case the EAKF is well behaved during the entire 2-year period in that the RMS errors generally decline over time, and there is no sign of divergence of the ensemble. As discussed by Houtekamer et al. (2005), an additional measure of the reliability of the ensemble is if $d^T d = Tr\{(y_b - \bar{y}_b)(y_b - \bar{y}_b)^T + R\}$ where $d$ is the innovation vector, $y_b$ is a vector for each ensemble member evaluated at the observation locations and observation times, and an overbar denotes the ensemble mean. Figures 10d and 10e shows time series of $d^T d/N$ and $Tr\{(y_b - \bar{y}_b)(y_b - \bar{y}_b)^T + R\}/N$ for CB_EAKF1b calculated for all temperature and salinity observations where $N$ is the number of observations of each type assimilated. After a few months of filter spin-up, the agreement between the mean squared innovation and the total expected error variance per observation is very good, indicating that the ensemble is indeed reliable during much of the experiment.

Figure 11 shows how the RMS errors in $T$ and $S$ vary with depth during CB_EAKF1b and indicate that this case is typically superior to M4DVAR8 at all depths. In agreement with 4D-Var, Fig. 10d shows that temperature errors during CB_EAKF1b undergo a seasonal cycle, being largest during the summer. The spatial structure of the errors on selected dates are also shown in Figs. 12d, 12h, and 12l and indicate that they are generally lower everywhere, and the problems in the deep ocean identified earlier have been eliminated.

### 5.1 Innovation statistics

Another useful way to compare the EAKF and 4D-Var systems is to examine the statistics of the innovations of each approach, which for a linear system, are expected to be normally distributed with zero mean and covariance $(H BH^T + R)$. With this in mind, Figs. 13a-d show the probability density functions (pdfs) of the innovations for SST and in situ salinity computed for the entire Jan 2003 - Dec 2004 period from both systems. Also shown for comparison are normal distributions with the same mean and standard deviation as the innovations. The EAKF and 4D-Var innovations are very similar to each other, and in both cases, are approximated very well by a normal distribution. The same is true for innovations of SSH and in situ temperature observations (not shown).
Figure 13: Probability density functions (pdfs) of the Jan 2003 - Dec 2004 innovations in (a) 4D-Var SST, (b) 4D-Var \textit{in situ} S, (c) EAKF SST and (d) EAKF \textit{in situ} S. The red line shows the pdf for a normal distribution with the same mean and standard variation. Time series of the first four moments of the innovation pdfs computed over a 15-day moving window for the EAKF (thick lines) and 4D-Var (thin lines) for (e) SST, (f) SSH, (g) \textit{in situ} T and (h) \textit{in situ} S. The pdf mean is shown in blue, the standard deviation in red, the skewness in black and the kurtosis scaled by a factor of 0.1 in cyan.

The degree to which the pdfs change through time is illustrated in Figs. 13e-h, which show time series of the mean, standard deviation, skewness, and kurtosis of the innovation pdfs computed over a 15-day moving window for period Jan 2003 - Dec 2004. The time evolution of the EAKF and 4D-Var mean and standard deviations are generally very similar to each other for the innovations associated with each observation type. The mean is close to zero throughout the 2-year window indicating no significant bias. After an initial period of a week or so, the standard deviations settle down to near-constant values in both data assimilation systems. The skewness of a normal distribution is zero, and except for the first month or so, the skewness is generally low and centered around zero, indicating that by-and-large the innovation pdfs are well approximated through time by a normal distribution. There are more significant variations in the skewness of the \textit{in situ} observations. Some of these variations are due to the relatively small sample size (i.e., there are relatively few \textit{in situ} observations available during most 15-day moving windows). However, in the case of \textit{in situ} salinity observations, both the EAKF and 4D-Var innovations are significantly positively skewed during the early months of 2003 (i.e., the distribution leans more toward negative innovations). The kurtosis of a normal distribution is 3 and has been scaled by a factor of 0.1 in Figs. 13e-h for convenience of visualization. Both the EAKF and 4D-Var innovation pdfs have a kurtosis that is generally close to a value of 3 through time for all observation types. The exception is the first few weeks of each experiment in which elevated values of kurtosis indicate that at these times, the tails of the distribution are lower than those of the expected normal distribution (i.e., there are fewer extreme values than expected).

Figure 14: Time series of the correlation between the EAKF and 4D-Var innovations over a 15-day moving window associated with observations in SST (black line), SSH (red line), \textit{in situ} T (cyan line) and \textit{in situ} S (dark blue line).

While the statistics of the pdfs in Fig. 13 indicate that the innovations of both EAKF and 4D-Var are both well approximated by a normal distribution, it is of interest to know the extent to which the innovation values for each observation correspond to each other in both assimilation systems. This relationship is quantified in Fig. 14, which shows time series of the correlation between
innovations associated with each observation type computed over a 15-day moving window for the period Jan 2003 - Dec 2004. Except for the first week or so, the correlation between the EAKF and 4D-Var innovations for all observation types is between 0.8 and 0.9 (i.e., $r^2 \approx 0.6-0.8$), indicating that both assimilation systems are producing very similar values for the innovations associated with individual observations through time.

5.2 Imperfect forcing and boundary conditions

In all of the OSSEs considered so far, the state of the atmosphere and open boundary conditions were assumed to be known and error-free. In this section, these assumptions are relaxed, and the influence of errors in these critical boundary conditions on the ability of the EAKF and 4D-Var systems to recover the nature run will be considered. As noted in section 4, the nature run was generated by computing surface fluxes of heat, freshwater, and momentum from atmospheric fields derived from COAMPS. Open boundary conditions were derived from SODA. The experiments CB_EAKF1b and M4DVAR8 were repeated a second time to explore the influence of errors and uncertainties in the surface fluxes and wind stress. However, in this case, the surface forcing was computed using the combination of ECMWF and CCMP fields described earlier. The open boundary conditions used were the same as those of the nature run.

Figure 15: Time series of the RMS errors in (a) $T$ and (b) $S$ for repeat experiments CB_EAKF1b (black line) and M4DVAR8 (red line) using an incorrect realization of the atmospheric conditions. The gray shaded region represents the envelope bounded by the maximum and minimum RMS errors of each of the 51 members of the CB_EAKF1b ensemble. RMS errors for a 4D-Var experiment where the surface fluxes and wind stresses are also adjusted are also shown (solid blue line). For reference, the errors from a model where no data were assimilated are also shown (black dashed line). Panels (c) and (d) show the RMS errors in $T$ and $S$ for experiments using incorrect open boundary conditions: black line - CB_EAKF1b, red line - M4DVAR8. Also shown are the RMS errors of a 4D-Var experiment in which all open boundary conditions were adjusted (solid blue line). Each panel also shows time series of the mean squared innovation $d^T d / N$ (cyan line) and the expected total error variance of the ensemble $Tr[(y_b - \bar{y}_b)(y_b - \bar{y}_b)^T + R] / N$ (magenta line) for all temperature and salinity observations, where $N$ is the number of observations assimilated in each case. The scale for $d^T d / N$ and $Tr[(y_b - \bar{y}_b)(y_b - \bar{y}_b)^T + R] / N$ is shown on the right.
Figures 15a and 15b show time series of the RMS errors in $T$ and $S$ from these new experiments. The errors initially decrease in both cases during the first 2-3 months. However, the spring transition is followed by an increase in errors in $T$ (Fig. 15a), to the extent that 4D-Var is inferior to the case where no data were assimilated. During the fall and winter, the errors decrease again. In the case of $S$ (Fig. 15b), the errors are similar in both cases, although during summer the EAKF errors are higher. Figure 15a indicates that the temperature errors in the case without data assimilation also increase during the summer. Further investigation reveals that this is associated with large differences in the surface heat flux derived from ECMWF and COAMPS fields that are mostly attributable to differences in solar heating due to differing distributions of marine stratus, a common feature during the summer. During this time, Figs. 15a and 15b also indicate that the spread of the EAKF ensemble is too small compared to the mean squared innovation and is not reliable. As noted in section 4.1.1, the under-dispersive nature of the ensemble is probably associated with the fact that each ensemble is subject to the same atmospheric conditions and oceanic open boundary conditions. The influence of perturbing these fields also will be the subject of a future study.

The ROMS 4D-Var system also includes an option whereby the surface fluxes and wind stress can be included as part of the control vector and adjusted to better fit the model to the observations. Figures 15a and 15b show the results of an additional 4D-Var experiment based on M4DVAR8 in which the surface heat flux ($Q_H$), freshwater flux ($Q_w$), and both components of surface wind stress ($\mathbf{\tau}$) were adjusted during each 8-day 4D-Var cycle. Following Broquet et al. (2011), the background error covariance matrix for the additional control vector components was modeled using a diffusion operator with assumed horizontal decorrelation length scales of 100 km for $Q_H$ and $Q_w$ and 300 km for $\mathbf{\tau}$. The standard deviations assumed for the background errors were chosen to be uniform over the model domain and specified according to $2\sigma_T$ for $\mathbf{\tau}$, $2\sigma_w$ for $Q_w$, and $(2 + 8\sin^2(j\pi/12))\sigma_H$ for $Q_H$, where $\sigma_T$, $\sigma_w$, and $\sigma_H$ are the annual, domain-averaged climatological standard deviations of $\mathbf{\tau}$, $Q_w$, and $Q_H$, respectively, and $j$ denotes the calendar month such that $j=1$ is Jan, $j=2$ is Feb, etc. The time dependence imposed on the standard deviation for $Q_H$ mimics the seasonal cycle in the uncertainty associated with the net solar forcing with the largest uncertainty during summertime. Figure 15b shows that while this configuration of 4D-Var has little influence on errors in $S$, Fig. 15a reveals that the impact on $T$ is considerable.

It should be noted that the surface forcing corrections computed by 4D-Var are predicated on the assumption that errors in the background surface flux estimates are unbiased and normally distributed. However, more often than not, surface forcing errors take the form of a bias, which is the case here, so it is encouraging that the 4D-Var system can improve the state estimate in this situation also. A more appropriate approach though would be to include a variational bias correction term in the cost function (4) as described by Dee (2005) and Balmaseda et al. (2007).

To explore the influence of errors and uncertainties in the open boundary conditions, the experiments CB_EAKF1b and M4DVAR8 were repeated using open boundary conditions that were offset by one year. Thus, for the 2003 experiments, the open boundary conditions were derived from the SODA product for 2004. The atmospheric conditions were assumed to be error-free. Time series of RMS errors for these experiments are shown in Figs. 15c and 15d. In this case, the errors are similar for both the EAKF and 4D-Var, and although the EAKF performs
better, the ensemble spread is generally a little too low compared to the mean squared innovation. Also shown in Figs. 15c and 15d are results of a 4D-Var experiment in which all open boundary variables were included as part of the control vector and adjusted to better fit the model to the observations. Given the lower resolution of the SODA boundary data compared to ROMS (0.25° - 0.4° vs. 0.1°), the background error decorrelation lengths were chosen to be 100 km in the horizontal and 30 m in the vertical for all boundary variables following N16, and the background error standard deviations were taken to be the monthly climatological standard deviations of each SODA field along the ROMS open boundaries. Figures 15c and 15d show that this approach leads to only marginal improvement in RMS errors. As in the case of the surface flux adjustments, the background boundary condition errors are typically systematic, while the underlying assumption in 4D-Var is that they are unbiased and normally distributed. Therefore, a variational bias approach would be more appropriate in the case also.

5.3 Errors relative to unassimilated observations

In practice, of course, the true circulation will never be known, so we must validate the data assimilation analyses against independent observations or against model hindcasts initialized from the analyses with new observations before they are assimilated into the model. In the case of hindcasts initialized from 4D-Var analyses, the difference between the hindcast and the new observations corresponds to the innovations for the portion of the hindcast interval that overlaps the next assimilation cycle. For the hindcasts initialized from the EAKF analyses, the hindcast is sampled at the actual space-time locations of the new observations, and so in general the differences will not correspond to the FGAT innovations of the next cycle(s). Here we perform these comparisons using the simulated observations since this will provide a useful benchmark for the experiments that use real observations in section 6. The circulation estimates from the CB baseline EAKF, CB_EAKF1b, and M4DVAR8 cases were all used as initial conditions for 8-day hindcasts, and the hindcasts compared to the new observations collected during the hindcast interval (i.e., before they were assimilated into the model during the next assimilation cycles). An 8-day interval was used for convenience because this corresponds to the 4D-Var window length.

Figure 16 shows time series of the RMS difference between model hindcasts and simulated observations during the hindcast interval. Specifically, Fig. 16 shows the envelope bounded by the maximum and minimum of the error evaluated on each hindcast day. During most hindcast cycles, the smallest errors are associated with the hindcasts on day 1 and the largest errors with hindcast day 8. For comparison, the standard deviation of the unbiased random error added to the simulated observations is also shown. Several noteworthy features emerge in Fig. 16. First, the errors of hindcasts initialized from both the EAKF and 4D-Var analyses are always typically close to the observation error, although the EAKF takes ~3-4 months to reach this limit. Second, despite the divergence of the ensemble in the 1-day FGAT CB EAKF baseline case, there is no hint of any problem with the analyses based on the available observations because there are no deep temperature and salinity profiles. Finally, while CB_EAKF1b using a 4-day FGAT window is superior in terms of the true error (cf. Fig. 10d-f), Fig. 16 indicates that this is not reflected in the error of hindcasts initialized from the 4-day FGAT EAKF analyses, particularly in the case of SST (Fig. 16a).
Figure 16: Time series of RMS errors of hindcasts initialized from the CB EAKF baseline, M4DVAR8, and CB_EAKF1b circulation estimates. The hindcast errors were computed relative to independent observations of (a) SST, (b) SSH, (c) in situ T, and (d) in situ S before these data were assimilated during the next assimilation cycle. The hindcast duration is eight days in each case, and hindcast errors were computed for each hindcast day. The shaded regions are bounded by the maximum and minimum daily hindcast errors over the 8-day window. The errors for the CB EAKF baseline are shown as blue shading, the errors for CB_EAKF1b as green shading, and those for the M4DVAR8 forecasts as red shading. The 8-day hindcast errors for the model run without data assimilation is also shown (black dashed line). For reference, the standard deviation of the random errors added to the simulated observations is also shown (thin black line). Also shown are time series of the true hindcast errors as compared to the nature run for CB_EAKF1b (green line) and M4DVAR8 (red line). Time series of RMS errors of hindcasts initialized from CB_EAKF1b (green) and M4DVAR8 (red) using imperfect atmospheric fields during the analysis and hindcast cycles are shown in (e)-(h).

As an additional reference point, Fig. 16 also shows time series of the true hindcast errors computed by comparing the forecasts with the nature run. The errors in the forecasts initialized from the EAKF and 4D-Var analyses are generally comparable, although the 4D-Var-based predictions fare better for SST. While we can only ever expect to fit the model to observations to within the expected combined observation errors and errors of representativeness, it is encouraging to see that the model hindcasts are generally also improving over time when compared to the nature run.

Figures 16e-h show the corresponding errors for hindcasts initialized from EAKF and 4D-Var state estimates in the case where the atmospheric fields are in error. In this case, also, the errors computed relative to unassimilated in situ T, S and SSH are close to the observation errors, despite the very significant errors compared to the nature run (cf. Fig. 15). The only significant departures from the observation errors occur in SST during the summer (Fig. 16e), which is consistent with Fig. 15a.
6. CCS Real Observation Experiments

In the final set of CCS experiments reported here, observations from the real ocean observing system for the period Jan-Dec 2003 were assimilated into the model using the CB_EAKF1b and M4DVAR8 configurations. Figures 17a-d show the pdfs of the SST and in situ salinity innovations for the 4D-Var and EAKF cases for the entire 2003 period. While the 4D-Var pdf for SST is approximated well by a normal distribution (Fig. 17a), the same cannot be said for the EAKF (Fig. 17c) or the in situ observations. The same is generally true for SSH and in situ temperature observations (not shown). Time series of the first four moments of the innovation pdfs computed for a 15-day moving window are shown in Figs. 17e-h. The mean for the SST and SSH innovations is generally close to zero, although there are periods of significant bias in SST during the late spring and early summer. The innovations for the in situ temperature observations (Figs. 17g) in the case of 4D-Var are close to zero throughout the year. However, for the EAKF, they reveal the presence of a negative bias indicating that the model is on average warmer than the observations. The EAKF innovation standard deviation is generally higher than that of 4D-Var for all observation types, particularly in the case of in situ temperature. Figures 17e-h reveal significant variations in skewness for all observation types, and particularly for the EAKF. Similarly, the kurtosis undergoes large fluctuations, particularly in the case of in situ observations of $T$ and $S$, which in conjunction with the skewness, indicate significant departures from a normal distribution at all times, commensurate with Figs. 17b and 17d. This result is in stark contrast to the innovation pdfs of the simulated observation experiments in Fig. 13. The innovation pdfs of the imperfect forcing and boundary condition experiments in section 5.2 are described well by a normal distribution (not shown), so it seems likely that the significant departures from a normal distribution in Fig. 17 are associated with model errors. In addition, the model possesses some significant biases in vertical temperature and salinity structure associated with the presence of different water masses as shown by Veneziani et al. (2009a), and previous studies have shown only part of the bias can be eliminated by data assimilation (Broquet et al., 2009a; Neveu et al., 2016).

![Figure 17: Probability density functions (pdfs) of the Jan-Dec 2003 innovations when real ocean observations are assimilated. (a) 4D-Var SST, (b) 4D-Var in situ $S$, (c) EAKF SST and (d) EAKF in situ $S$. The red line shows the pdf for a normal distribution with the same mean and standard variation. Time series of the first four moments of the innovation pdfs computed over a 15-day moving window for the EAKF (thick lines) and 4D-Var (thin lines) for (e) (f) (g) (h)]
SST, (f) SSH, (g) in situ $T$ and (h) in situ $S$. The pdf mean is shown in blue, the standard deviation in red, the skewness in black and the kurtosis scaled by a factor of 0.1 in cyan.

The degree of correspondence between the EAKF and 4D-Var innovations associated with each type of observation is shown in Fig. 18. During the period Jan-May, the correlation between the SST and in situ observation innovations is ~0.8, indicating that for these observations, the two assimilation systems yield innovations that are, on the whole, reasonably similar. However, after May, the correlations decline quite rapidly, revealing that the two data assimilation systems are diverging.

![Figure 18: Time series of the correlation between the EAKF and 4D-Var innovations over a 15-day moving window associated with observations in SST (black line), SSH (red line), in situ $T$ (cyan line) and in situ $S$ (dark blue line) when real observations are assimilated.](image)

As in the simulated observation experiments, the analyses were used to initialize 8-day hindcasts, which were compared to the new observations before they were assimilated. Time series of RMS difference between the hindcasts and the observations before assimilation are shown in Fig. 19 for the two sequences of hindcasts. In this case, all of the hindcasts initialized from the EAKF analyses are, on average inferior to those initialized using 4D-Var. In the case of in situ $T$ and $S$, they are often worse than the case where no data are assimilated. The reliability of the EAKF ensemble is shown in Figs. 19e and 19f for all temperature and salinity observations respectively and indicates that there are often periods where the ensemble would be considered unreliable. As noted in section 5.2, this occurrence is most likely associated with the fact that each ensemble is subject to identical atmospheric conditions and oceanic open boundary conditions.
Figure 19: Time series of RMS errors of hindcasts initialized from the EAKF and 4D-Var circulation estimates derived from real ocean observations in the CCS configuration. The hindcast errors were computed relative to independent observations of (a) SST, (b) SSH, (c) in situ T, and (d) in situ S before these data were assimilated during the next assimilation cycle. The hindcast duration is eight days in each case, and hindcast errors were computed for each hindcast day. The shaded regions are bounded by the maximum and minimum daily hindcast errors over the 8-day window. The errors for the EAKF-derived forecasts are indicated by green shading, and those resulting from the 4D-Var forecasts as red shading. The 8-day forecast errors for the model run without data assimilation is also shown (black dashed line). For reference, the standard deviation assumed for the combination of instrument error and the error of representativeness is also shown (thin black line). Panels (e) and (f) show time series of mean squared innovation $d^T d/N$ (cyan line) and the expected total error variance of the ensemble per observation $Tr[(y_b - \bar{y}_b)(y_b - \bar{y}_b)^T + R]/N$ (magenta line) for all temperature and salinity observations, where $N$ is the number of observations assimilated in each case.

Since we do not know the true ocean circulation, we will use as an alternative gauge the RMS difference between the EAKF and 4D-Var state estimates compared to that for the different cases considered in section 4 where simulated observations were used. With this in mind, Fig. 20 shows times series of the RMS difference between the temperature and salinity analyses from the EAKF and 4D-Var systems using real observations. Figure 20a indicates that there is a pronounced seasonal cycle in the temperature differences with the two estimates being furthest apart during summer and fall. In contrast, the salinity field estimates diverge continuously throughout the year. It is helpful to view these analysis differences in terms of those derived from simulated observations. Therefore Figure 20 also shows the RMS differences between the CB_EAKF1b and M4DVAR8 with perfect atmospheric and boundary conditions for the two initial first-guess states $x_0^A$ and $x_0^B$ and the RMS differences between the experiments that use imperfect atmospheric and boundary conditions. The RMS EAKF minus 4D-Var analysis temperature differences using real observations are most similar to those that arise from imperfect atmospheric conditions in the OSSEs. However, in the case of salinity, there is no apparent analog.
7. The Indian Ocean

In this section, we document the performance of the ROMS EAKF and 4D-Var data assimilation systems in the Indian Ocean as a representative example of a tropical basin-scale circulation regime.

7.1 Model and observations

The ROMS Indian Ocean application (ROMS-IO) was designed to resolve the main modes of seasonal and mesoscale variability with particular emphasis on the well-observed equatorial region. The model domain spans the area 16°S-27.5°N, 38.5°E-102.3°E and is shown in Fig. 21. The spatial resolution varies from ~29 km at the northern and southern boundaries and increases approaching the equatorial waveguide to ~7 km. In the vertical, 40 terrain-following levels yield a vertical resolution between 6 m and 15 m within the upper equatorial seasonal thermocline (i.e., ~ upper 120 m). The spatial resolution chosen resolves the most critical processes and is also amenable to ensemble and variational data assimilation.

At the open boundaries, the model was constrained by fields from the Global Hybrid Coordinate Ocean Model (HYCOM; Metzger et al., 2014) analyses. The model was gradually nudged to daily means of temperature and salinity in a nudging layer ~200 km wide, where the nudging timescale decreases from 2 days at the lateral boundaries to zero at the interior. Additionally, adaptive open boundary conditions for tracers and currents, as described in Marchesiello et al. (2001), were used. For the barotropic component, Chapman (1985) boundary conditions were used for the vertically averaged velocity, and Flather (1976) boundary conditions suitable for a staggered C-grid (Mason et al., 2010) were used for the free surface elevation. At the surface, the model is forced with air-sea momentum, heat, and freshwater fluxes computed using the bulk formulae parameterizations for the marine boundary layer, as described in Fairall et al. (2003). The atmospheric fields required by these parameterizations were obtained from the ERA-interim analysis at the highest spatial and temporal resolution available (3 hours at ⅛ degree). The river
forcing includes seasonal cycles of river discharge from the five major rivers discharging to the IO basin (the Brahmaputra, Ganges, Irrawaddy, Godavari, and Indus) based on the freshwater discharge database of Dai and Trenberth (2002). A 10-year model integration (2003-2012) produced a reasonable seasonal cycle without significant drifts when compared with satellite estimates of SSH and SST and the temperature and salinity climatology of Chatterjee at al. (2012).

The observations used to evaluate the solutions with and without data assimilation take the form of daily SST composites from the Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA; Donlon et al., 2011), SSH estimates from the gridded products from AVISO, and in-situ observations from different sources compiled in the World Ocean Database 2018 (WOD18; Boyer et al., 2019). The two primary sources of in situ temperature and salinity are daily mean profiles from the Research Moored Array for the African-Asian-Australian Monsoon Analysis and Prediction (RAMA), and profiles from Argo floats. Also included in the in situ database are XBT and CTD profiles, as shown in Fig. 21b.

Figure 21: The ROMS-IO model domain and observations used in the EAKF and 4D-Var experiments. An example of the OSTIA SST and AVISO SSH on 15 October 2011 is shown in (a) and (c) respectively. The locations of all in situ hydrographic observations collected between 1-Sep-2011 and 31-Jan-2012 are shown in (b) superimposed on the 15-Oct-2011 SST, and time series of log₁₀ of the number of observations of each type available during each 24-hour window is shown in (d).

Due to the computational demands of the ROMS-IO application, we concentrate on the 5-month window from 1-Sep-2011 to 31-Jan-2012. This period includes the Dynamics of the Madden-Julian Oscillation DYNAMO; Chen et al., 2015; Jensen at al., 2015) field campaign (Oct-Dec 2011), which provides additional in situ mooring data from CTDs.
7.2 EAKF configuration

Due to the large size of the ROMS-IO grid, it was not possible to perform an extensive series of OSSEs and sensitivity calculations as in the case of the CCS application described in sections 4 and 5. However, the results from the OSSEs in the CCS were a valuable guide for the parameter choices adopted here. The number of ensemble members used was 31, which seems to be a reasonable lower bound in the CCS application and is only marginally worse than the 51-member ensemble case (cf. Fig. 7b). The ensemble of initial conditions needed to initialize the EAKF were created from the 10-year forward run using the centered method B described in section 4.1.1, where \( \mathbf{x}_0 \) was the forward model solution for 1-Sep-2011, and the ensemble perturbations \( \pm \delta \mathbf{x} \) were scaled by a factor of 0.5. As discussed earlier, centering the ensemble ensures that no bias is introduced during the EAKF initialization due to the relatively small number of ensemble members. A 2-day FGAT window was used in conjunction with adaptive inflation and Gaspari-Cohn localization. The required localization scale in the DART input file was chosen based on the spatial decorrelation scales of the forward model anomalies. While it was found that the decorrelation scales vary with latitude and proximity to the coast, we use a value of 250 km which is representative of the equatorial waveguide and open ocean.

Figures 22b and 22c compare the rank histograms during Sep and Nov of 2011 based on all the available subsurface observations (approximately 24,000 per month). While the rank histograms suggest insufficient spread of the initial ensemble (Fig. 22b), the adaptive covariance inflation algorithm yields a more uniform ensemble spread by Nov 2011 (Fig. 22c), as in the CCS case.

7.3 4D-Var configuration

The 4D-Var experiment used a 4-day assimilation window, and only the initial conditions were used as the control variable to compare with the EAKF experiments. The background error covariance \( B \) is constructed using the seasonal standard deviations of each field \( (\Sigma_c) \), which were estimated by fitting the first two harmonics of the annual cycle to the 3-month running mean of the standard deviations of the forward model anomalies. As discussed in section 4.2.3, these standard deviations are representative of the uncertainty in the initial conditions of the first 4D-Var assimilation cycle but are expected to overestimate the true background errors as the 4D-Var cycling procedure continues as demonstrated for the CCS (cf. Fig 9d). With this in mind, the climatological standard deviations were gradually rescaled during the first 4-cycles (i.e., the first 16 days of Sep 2011) down to 25% of the climatological value. This initial damping of the standard deviations was found to approximate the expected background errors for all model variables based on the innovation statistics of Desroziers et al. (2005), as described in section 4.2.3. The background correlation lengths used to model \( B \) were 250 km in the horizontal and 50 m in the vertical, in conjunction with two outer-loops and seven inner-loops. A comparison between the expected \( \left( \hat{\sigma}_b^2 \right) \) and the assumed \( \left( Tr(\Sigma_c)/N \right) \) background error variance during each 4D-Var assimilation cycle for the in situ temperature is shown in Fig. 22a. The error estimates assumed are consistent with those expected.
7.4 Comparison of EAKF and 4D-Var experiments for ROMS-IO

Figures 22d-g compare the pdfs of innovations for SST and in situ observations from the 4D-Var and EAKF experiments for the entire analysis period. The behavior of the innovation pdfs for ROMS-IO are generally consistent with those found in the CCS, with SST pdfs approximately exhibiting normal distributions. However, the in situ salinity pdfs depart significantly from a normal distribution in this case also.

Since in this case we cannot compare the solutions to the true ocean state, we will consider first the fit of the 4D-Var and EAKF analyses to the assimilated observations which is summarized in Fig. 23. After a period of initial adjustment lasting around one month, the RMS errors converge to approximately the same level except for the in situ temperature, which is larger for the EAKF estimate. The RMS differences exhibit more variability in the EAKF, and at times they are comparable to those of the forward solution for some variables (e.g., SST in Nov 2011 and in situ temperature at the end of Sep 2011).
Figure 23: Time series of RMS errors for the 4D-Var (red lines) and EAKF (blue lines) analyses computed from the assimilated observations. (a) SST, (b) SSH, (c) in situ temperature, and (d) in situ salinity. Also shown for comparison is the case where no data were assimilated (black dashed line).

Similar to the CCS experiments with real observations (cf. Fig. 19), Fig. 24 shows validation of the analyses in terms of the RMS difference between hindcasts initialized from the EAKF and 4D-Var analyses and the new observations before they were assimilated. For convenience, the hindcast period is the same as the 4D-Var window length (4 days). Both estimates produce similar hindcast errors, although those derived from the EAKF analyses exhibit some large fluctuations in errors especially when compared to the subsurface in-situ observations. As in the CCS experiments, the a priori observation error is an approximate lower bound on the hindcast error.
Figure 24: Time series of RMS errors of hindcasts initialized from the EAKF and 4D-Var circulation estimates in the Indian Ocean configuration. The hindcast errors were computed relative to independent observations of (a) SST, (b) SSH, (c) in situ T, and (d) in situ S before these data were assimilated during the next assimilation cycle. The hindcast duration is four days in each case, and hindcast errors were computed for each hindcast day. The shaded regions are bounded by the maximum and minimum daily hindcast errors over the 4-day window. The errors for the EAKF-derived forecasts are shown as green shading and those for the 4D-Var forecasts as red shading. The forecast errors for the model run without data assimilation are also shown (black dashed line). The solid black line indicates the standard deviation of the combined instrument and error of representativeness.

8. Summary and Conclusions

This study presents a detailed comparison of the performance of two data assimilation systems that are available as part of the community ROMS. The two systems considered are an ensemble adjustment Kalman filter and a 4-dimensional variational approach. The performance of each system has been evaluated in two diverse dynamical regimes: the California Current system, an eastern boundary current upwelling regime, and the Indian Ocean comprising an energetic equatorial waveguide subject to the large-amplitude seasonal variations of the African-Asian-Australian Monsoon system.

In the CCS, the relative performance of the EAKF and 4D-Var systems was assessed via an extensive suite of OSSEs using simulated observations drawn from the actual observing system. These experiments were designed to quantify the sensitivity of the quality of the ensuing ocean state estimates on the configuration of each DA system. In particular, the OSSEs reveal that the performance of the EAKF is most sensitive to ensemble size, choice of localization scales, and the strategy used to generate the initial ensemble. On the other hand, the performance of 4D-Var is most sensitive to the assimilation window length and a priori assumptions about the decorrelation length scales associated with the background errors. Short FGAT windows were found to create significant issues for the EAKF which are most likely associated with initialization shock or with the reinitialization of the vertical mixing parameterization, or both.
This issue was alleviated by increasing the FGAT window length allowing more time for inertia-gravity waves excited by initialization shocks to dissipate before the next data assimilation cycle, while at the same time enabling the turbulent kinetic energy and mixing length scales of the vertical mixing parameterization to reach an equilibrium. These issues are the subject of an ongoing investigation.

An interesting finding of this study is that the sequence of EAKF and 4D-Var analyses converge to approximately the same ocean state when starting from different first-guess fields for the first assimilation cycle. This result suggests that the 4D-Var cost function does not possess multiple minima which could cause the independent estimates to be attracted to different ocean states.

The innovations in both the EAKF and 4D-Var are well described through time by unbiased normal distributions. Furthermore, the statistics of the innovation pdfs for the EAKF and 4D-Var track each other closely, and the correlation between the innovations associated with individual observations is high, indicating that the two systems remain close and are generating consistent analyses despite the fundamentally different approaches that are employed.

The greatest impediment to the EAKF and 4D-Var, as revealed by the OSSEs, is associated with errors in the surface forcing. In the CCS, where the circulation is dominated by a pronounced seasonal cycle in upwelling and marine stratus, errors in the state of the atmosphere lead to a dramatic reduction in the ability of either DA system to recover the true circulation. The under-dispersive nature of the ensemble during these experiments suggests that the EAKF may, in general, suffer from a lack of diversity of atmospheric (and possibly also open boundary) states in the systems considered here.

While overall, the collective result of all CCS OSSEs indicate that the EAKF generally outperforms 4D-Var, the differences are not large. However, the OSSEs also revealed that evaluating the efficacy of DA analyses based on hindcasts compared to independent observations has its limitations. As anticipated, the model cannot be fit to the observations better than the observation errors allow.

The behavior of the two data assimilation systems applied to the CCS is different in the case when real observations are assimilated. For example, the EAKF and 4D-Var innovation pdfs are generally not well represented by a normal distribution, particularly in the case of in situ observations. Furthermore, unlike the state estimates of the OSSEs, the EAKF and 4D-Var analyses diverge from each other over time. This behavior may be due to the presence of model error and persistent bias, which is absent in the OSSEs. It is also likely, as noted above, that performance of the EAKF here is compromised by subjecting each ensemble member to the same atmospheric state and open boundary conditions. The influence of perturbing the surface forcing and boundary conditions on the performance of the EAKF will be the subject of a future study.

By and large, the relative behavior and performance of the EAKF and 4D-Var systems in the Indian Ocean is consistent with that in the CCS experiments. The only notable difference is that the skill of hindcasts initialized from the two analysis products is more comparable in the Indian Ocean. Thus, both DA systems can function consistently across different circulation
environments, and the relative performance of the EAKF and 4D-Var systems presented here is likely to be representative of that in other regions of the world ocean.

4D-Var has been available in the community release of ROMS for quite some time (Moore et al., 2011a), and the introduction of ROMS within DART increases considerably the data assimilation arsenal available to ROMS users. As already noted, the overall performance of the two data assimilation systems is comparable in the OSSEs and OSEs reported here. Indeed, as shown by Lorenc (1986), with all things being equal both data assimilation approaches would be expected to yield very similar and consistent analyses, which is the case. Thus, which approach a ROMS user might choose is likely predicated on personal choice and the application at hand. The DART EAKF has the significant computational advantage in that multiple ensemble members can be run concurrently, while the current formulations of 4D-Var supported by ROMS are strictly sequential in time. On the other hand, ROMS 4D-Var supports augmentation of the control vector to include the surface forcing, open boundary conditions, and corrections for model error (i.e. weak constraint 4D-Var), an option that is not yet available in the current ROMS-DART release. Both DART and ROMS provide a suite of useful post-processing tools, such as for quantifying the impact of the observations on ocean analyses and subsequent forecasts. The effort required to set up DART and 4D-Var is comparable and, in both cases, requires the user to make some important and informed decisions relating to $B$. In DART, a seed ensemble must be chosen and an appropriate ensemble size, in conjunction with appropriate choices of localization function, localization lengths, and covariance inflation factors, all of which may require some experimentation. In the case of 4D-Var, appropriate parameters for the covariance model must be chosen in the form of horizontal and vertical correlation length scales, as well as parameters for the background error standard deviation thresholds (cf. equations (5) and (6)).

Online documentation is available for DART and ROMS, as well as extensive data assimilation tutorials and hands-on exercise designed for users with all levels of expertise. Perhaps one obvious advantage of the EAKF over 4D-Var is that a forecast ensemble can be readily computed when used in a real-time or operational environment to provide information about forecast uncertainty. However, comparable information can also be computed in the ROMS 4D-Var system with similar computational effort using a practical implementation of the Bennett (1985) array modes of the observing system. Another practical consideration concerns the spin-up of the EAKF in situations when the observation record is limited in time. As shown in Fig. 3, the initial seed ensemble, and ensemble generation method, can significantly influence the time required for the EAKF to spin-up and reach a statistical stable state (see also Figs. 10d and 10e). If the spin-up time is long or a significant fraction of the observation record length, then it may be preferable to use 4D-Var which typically has a quicker spin-up time.

As noted in section 1, this work is a stepping-stone toward the development of a hybrid 4D-Var data assimilation system for ROMS that will be a marriage of the DART system and ROMS 4D-Var. The term hybrid refers to the formulation of the background error covariance matrix $B$ (Hamill and Snyder, 2000; Lorenc, 2003a; Buehner, 2005). Hybridization of $B$ in 4D-Var consists of combining the static $B_s$ derived from, say, $\Sigma_c$ or $\Sigma_m$ of section 4.2 with a flow-dependent covariance $B_e$ computed from a localized ensemble. The resulting background error covariance matrix is then given by $B = \alpha B_s + \beta B_e$, where the weights $\alpha$ and $\beta$ can be computed from statistical considerations (Ménétrier and Auligné, 2015). The system envisaged for ROMS is one in which DART will provide a localized $B_s$. The hybrid $B$ will be used in 4D-
Var to compute a new analysis which will then be used to re-center the DART ensemble mean, an approach that has been used successfully in numerical weather prediction (e.g. Zhang and Zhang, 2012; Clayton et al., 2013).

Acknowledgements

This work was supported by grants from the National Oceanographic Partnership Program (via the Office of Naval Research, N00014-15-2300; N00014-15-1-2545) and the National Oceanic and Atmospheric Administration (NA17NOS0120176).

References


