FINAL TECHNICAL REPORT

N00014-19-1-2675: Further Developments of the Coupled COAMPS-ROMS Modeling System

Principal Investigator: Andrew M. Moore
Department of Ocean Sciences
University of California Santa Cruz 95062
Tel: 831-459-4632
e-mail: ammoore@ucsc.edu

Accomplishments

What were the major goals and objectives of the project?

This project is an extension of a previous NOPP project (N00014-15-1-2300) aimed at developed a coupled model interface between the Regional Ocean Modeling System (ROMS) and the NRL Coupled Ocean-Atmosphere Mesoscale Prediction System (COAMPS) using the Earth System Modeling Framework (ESMF) National Operational Prediction Capability (NUOPC). An additional component of the original NOPP project was to develop the interface between ROMS and the Data Assimilation Research Testbed (DART) and update existing COAMPS-DART capabilities. With this in mind the following is a summary of the goals of this extension project:

(1) Final testing of the ROMS data ESMF-NUOPC cap files.
(2) Implement the recent release of the NUOPC layer that formally allows nesting between ESM components.
(3) Completion of a manuscript comparing the ROMS-DART and ROMS-4D-Var data assimilation systems.
(4) Preparation of the COAMPS Indian Ocean configuration for coupling with ROMS.
(5) Complete the updated DART-COAMPS interface.
(6) Coupling of COAMPS and ROMS in the Indian Ocean.

What was accomplished towards achieving these goals?

While most of the primary goals of this project were accomplished, some proved to be more technically challenging than originally anticipated and there were insufficient resources available to complete goal (5). As noted above, this is a follow-on project for *N00014-15-1-2300 Toward the Development of a Coupled COAMPS-ROMS Ensemble Kalman Filter and Adjoint with a focus on the Indian Ocean and the Intraseasonal* Oscillation. Therefore, for convenience, continuity, and transparency, the major accomplishments related to the goals of *both* projects combined are described in the three appendices below.

(a) Coupled modeling

While the original focus of the project was the Indian Ocean and intraseasonal variability, it was important to first develop and test the ROMS-COAMPS coupling framework in a region that is very familiar to the ROMS and COAMPS PIs involved in this project. For that reason, we chose to first test the coupled model system configured for the U.S. west coast (see also https://www.myroms.org/wiki/Model_Coupling_WC12). Some of the scientific outcomes of this work are described in Appendix A where the ability of the coupled model to recover the observed relationships at the air-sea interface was explored. The scientific outcomes of the coupled ROMS-COAMPS system configured for the Indian Ocean are described in Appendix B where the influence of air-sea coupling on the path of two tropical storms was explored using the ensemble members generated by an EnKF. While data assimilation in the fully coupled model was ultimately beyond the scope of this project, the work presented in both appendices does address the influence of data assimilation in the ocean component alone on coupled model behavior.

(b) The ROMS-DART EnKF

The ROMS-DART EnKF data assimilation system developed during this project is documented in the published paper by Moore *et al.* (2020) which is included in this report as Appendix C. In this paper we quantified the relative performance of the EnKF and the existing ROMS 4D-Var data assimilation system in two configurations of ROMS, namely, the U.S. west coast and the Indian Ocean. Observing System Experiments (OSEs) using real observations and synthetic observations from Observing System Simulation Experiments (OSSEs) were considered.

What opportunities for training and professional development did the project provide?

Nothing to report.

How were the results disseminated to communities of interest?

A paper describing the relative performance of the ROMS-DART ensemble adjustment Kalman filter and the ROMS-4D-Var data assimilation systems was published in *Progress in Oceanography*:

Moore, A.M., Zavala-Garay, J., Arango, H.J., Edwards, C.A., Anderson, J. and T. Hoar, 2020: Regional and basin scale applications of ensemble adjustment Kalman filter and 4D-Var ocean data assimilation systems. *Progress in Oceanography*, **189**, https://doi.org/10.1016/j.pocean.2020.102450.

What honors or awards were received under this project in this reporting period?

Nothing to report.

Technology Transfer

Nothing to report.

Participants

Name	Role	Person Months
Andrew Moore	PI/PD	4
Hernan Arango	Co-Investigator	7
James Doyle	Co-Investigator	1
Christopher Edwards	Co-Investigator	1
Sasa Gabersek	Co-Investigator	2
Javier Zavala-Garay	Co-Investigator	4

Students

Nothing to report.

Products

Publications-

Moore, A.M., Zavala-Garay, J., Arango, H.J., Edwards, C.A., Anderson, J. and T. Hoar, 2020: Regional and basin scale applications of ensemble adjustment Kalman filter and 4D-Var ocean data assimilation systems. *Progress in Oceanography*, **189**, https://doi.org/10.1016/j.pocean.2020.102450.

Keywords: Ensemble Kalman filter; 4D-Var; California Current; Indian Ocean.
Peer-reviewed.
Federal support is acknowledged.

Other products-

The ESMF-NUOPC layer is now part of the official release of the ROMS community code available from https://www.myroms.org. See also https://www.myroms.org/wiki/Model_Coupling_ESMF for technical details.

The ROMS interface for DART is also available as part of the DART software package from https://dart.ucar.edu.

APPENDICES: ACCOMPLISHMENTS

Appendix A: Air-Sea Interactions at the Ocean Mesoscale in a Regional Coupled Atmosphere-Ocean Model of the California Current System

1. Introduction

The nature of the coupling between the atmosphere and the ocean at the ocean mesoscale has been a subject of relatively intense investigation during the last decade or two. One of the most influential studies that is particularly relevant here is that of Chelton *et al*. (2007; hereafter CSS07) who explored air-sea interactions within the California Current System (CCS). Using satellite-derived observations of sea surface temperature (SST) and surface wind stress, CSS07 demonstrated conclusively how air-sea interaction induces changes in the divergence and curl of the surface wind stress in the vicinity of SST fronts. As shown in several earlier studies (*e.g.* Businger and Shaw, 1984), wind stress increases over warm water and decreases over cold water, as illustrated schematically in Fig. 1a (from CSS07). Despite the fairly complicated nature of the marine boundary layer response to the presence of an SST front (*e.g.* Samelson *et al.*, 2006; Spall, 2007), it is found that a linear relationship emerges between wind stress curl and the crosswind component of the SST gradient, and between the wind stress divergence and the downwind component of the SST gradient. The local control of wind stress curl induced by SST at the mesoscale is particularly noteworthy, since enhancements in local upwelling and downwelling in the CCS due to this mechanism may be significant compared to other factors (*e.g.* Pickett and Paduan, 2003; Huyer *et al*., 2005; Dever *et al*., 2006).

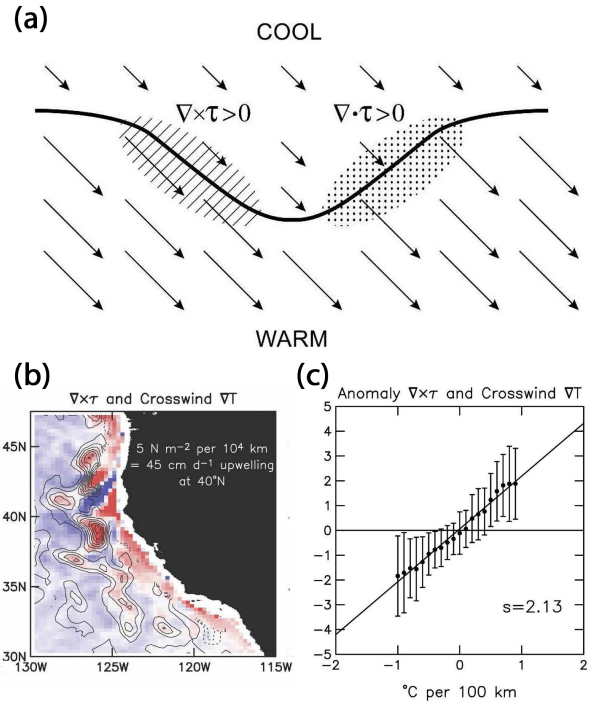

Figure 1: (a) A schematic representation of the influence of an SST front on wind stress. Regions of enhanced wind stress curl ($\nabla \times \boldsymbol{\tau}$) and wind stress divergence ($\nabla \cdot \boldsymbol{\tau}$) are indicated in relation to the orientation of the wind direction and the front. (b) The spatial variations in wind stress curl (colors) and the crosswind component of the SST gradient

(contour lines) derived from spatially high-pass filtered AMSR-E and QuikSCAT observations on 18 August 2002. (c) A binned scatter plot of wind stress curl vs the crosswind component of SST gradient based on summertime (June-Sept) AMSR and QuikSCAT observations for the period 2002-2005. The error bars represent one standard deviation and the best fit straight line through the mean values in each binned interval is also shown. The line slope $s = 2.13$ represents the *coupling coefficient* between the wind stress curl and crosswind SST gradient. Note that in (c) the units of $\nabla \times \boldsymbol{\tau}$ are N m^{-2} per 10^4 km and the units of SST gradient are °C per 100 km. All figures are from CSS07.

Figure 1b from CCS07 shows an example of the relationship between microwave observations of SST from AMSR-E and surface wind stress estimates derived from the QuikSCAT scatterometer. Both data sets have been spatially high-pass filtered using a *loess* smother with a half power cut-off of 75 km. This removes the large-scale features of the atmospheric and oceanic circulation which are generally coupled in the more traditional *equilibrium* sense of Gill (1980) and Lindzen and Nigam (1987) following wave adjustment. At the ocean mesoscale, however, inertial overshoot of the SST front by marine boundary layer winds can modify the downstream air-sea coupling relationships, and spatial high-pass filtering is necessary to identify the important correlations between the atmosphere and oceanic eddies and fronts. Specifically, Fig. 1b shows the wind stress curl and the crosswind component of the SST gradient of the high-pass filtered observations on a single day (18 August 2002) and reveals the remarkable degree of spatial coherency between the two fields. Figure 1c (also from CSS07) shows a binned scatterplot of summertime (June-September) wind stress curl and the crosswind SST gradient for the high-pass filtered data, for the period 2002-05. The linear relation between the two is striking and the slope of the best-fit straight line to the bin means is routinely taken as a measure of the *coupling strength* between the atmosphere and ocean.

Generally, CSS07 found that the strong SST gradients and fairly steady wind conditions that persist over much of the CCS during summer are favorable for strong air-sea interaction. Conversely, the weaker SST gradients and highly variable winds during winter do not favor such pronounced air-sea interactions. In addition, the orographic influences of major coastal promontories on wind stress curl, such as in the lee of Cape Blanco and Cape Mendocino, are also found to weaken its correlation with crosswind SST gradients.

O'Neill *et al.* (2012) have also found that the linear relationship between wind stress curl and crosswind SST gradient, and wind stress divergence and downwind SST gradient are a ubiquitous feature of western boundary currents and in much of the Southern Ocean (see also O'Neill *et al.*, 2005). Although, in these cases the atmosphere-ocean *coupling strength* is strongest during winter when the air-sea temperature difference is largest.

There are also a growing number of modeling studies that confirm the findings from observations. For example, Haack *et al.* (2008) found comparable relationships between SST gradients and wind stress using a high-resolution version of COAMPS (9km) configured for the CCS. Similarly, Small *et al.*, (2008) document the results of several experiments using regional coupled ocean-atmosphere models, including the CCS (Seo *et al.*, 2007), which also agree qualitatively with the observations. However, both horizontal and vertical resolution are important factors that can limit a model's ability to realistically capture important features of the air-sea coupling at the ocean mesoscale.

In addition to the influence of SST gradients on surface wind stress, there are other air-sea interaction processes that can significantly influence the ocean mesoscale circulation at mid-latitudes, particularly in relation to ocean eddies. As discussed by the detailed and inciteful study of Gaube *et al.* (2015), there are two additional air-sea interactions that have an important controlling influence on ocean eddies.

The first involves the impact of the relative motion of the ocean and atmosphere on the surface wind stress. Specifically, from the bulk formula, the surface wind stress can be expressed as $\boldsymbol{\tau} = \rho_a c_d |(\boldsymbol{U}_a - \boldsymbol{u}_o)|(\boldsymbol{U}_a - \boldsymbol{u}_o)$ where $\boldsymbol{U}_a$ and $\boldsymbol{u}_o$ are the surface wind and ocean current velocities, respectively. Thus, when the wind and current are generally in the same (opposite) direction, the wind stress magnitude is weakened (strengthened). In the presence of eddies, the change in wind stress across an eddy leads to a change in the wind stress curl. The vorticity induced by the change in wind stress curl is opposite to that of the eddy, and the associated Ekman pumping velocity acts to attenuate the eddies by generating upwelling in anticyclonic eddies and downwelling in cyclonic eddies.

A second important interaction between the atmosphere and mesoscale ocean eddies involves the torquing action that results from the interaction of the surface wind stress with the vorticity gradient of the ocean surface currents (Dewer and Flierl, 1987). This results in dipoles of Ekman pumping within the interiors of mesoscale eddies which act to displace the eddies depending on the direction of the prevailing wind.

Based on observations, Gaube *et al.* (2015) have explored the relative size of the Ekman pumping velocities induced in middle-latitude eddy cores by the three air-sea interaction mechanisms summarized above. In general, the influence of the surface torque and the relative motion between the surface winds and ocean currents typically dominate, while the SST-induced changes in surface wind stress curl are generally the least important. The combined influence of all three processes is to attenuate ocean eddies on timescales ~100-1000 days.

This appendix summarizes an investigation of the degree of air-sea interaction at the ocean mesoscale in a regional coupled model of the CCS. The coupled model is described in section 2 and the experimental set-up in section 3. An analysis of the sensitivity of air-sea *coupling strength* on atmospheric horizontal resolution, the bulk formulation of surface wind stress, and weakly coupled data assimilation is presented in section 4. A summary and conclusions follows in section 5.

2. Coupled Model Description

The coupled model used in this study was developed as part of this ONR-funded project and comprises the Regional Ocean Modeling System (ROMS) coupled to the atmospheric component of the NRL Coupled Ocean Atmosphere Mesoscale Prediction System (COAMPS). Communication between ROMS and COAMPS was performed using the Earth System Modeling Framework (ESMF) National Operational Prediction Capability (NUOPC). A full technical description can be found at https://www.myroms.org/wiki/Model_Coupling_ESMF. The incorporation of ESMF-NUOPC within ROMS alone represents a significant accomplishment and considerably enhances the utility of the community ROMS system allowing

ROMS to be coupled to a wide range of other Earth System Model components in addition to COAMPS.

The ROMS and COAMPS grids used in this project are shown in Fig. 2. ROMS is configured with a horizontal resolution of 1/10th degree (~10 km) and with 42 terrain-following levels in the vertical. The ROMS grid is centered on the U.S. west coast and captures the circulation of the CCS. COAMPS as employed here comprises two nested grids. The outer-most grid (COAMPS-1 in Fig. 2) has a horizontal resolution ~45 km and spans North America and much of the NE Pacific. The inner-most grid (COAMPS-2 in Fig. 2) has a resolution ~15 km and is centered on the NE Pacific and overlaps the ROMS grid. There are 60 sigma-layers in the vertical in both COAMPS grids.

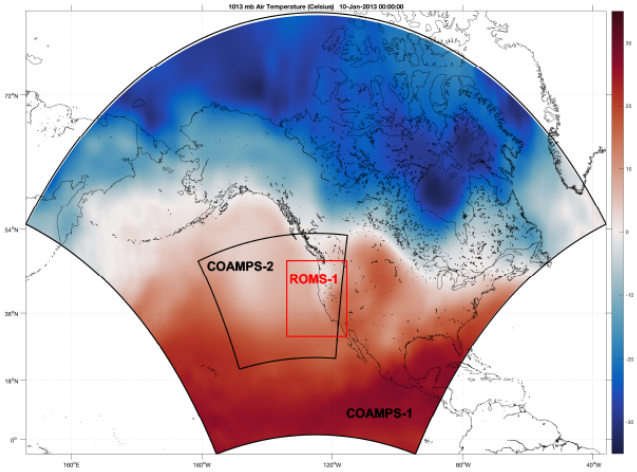

Figure 2: The configuration of the COAMPS-ROMS coupled model. The ROMS grid is centered on the U.S. west coast, and is labelled *ROMS-1*. Two layers of nested COAMPS grids are employed here and labelled *COAMPS-1* and *COAMPS-2*. COAMPS-1 has a horizontal resolution of ~45 km and COAMPS-2 a resolution of ~15 km. The horizontal resolution of ROMS is ~10 km.

3. Experimental Set-up

The period considered here is 1 Jan – 31 Dec 2017. The open boundary conditions for ROMS are taken from a global HYCOM data assimilation product. COAMPS was initialized every 5-days using an atmospheric 4D-Var analysis, and the open boundary conditions are provided by a NOGAPS analysis. ROMS and COAMPS are coupled every 10 minutes using the ESMF-NUOPC interface. In the regions where the ROMS and COAMPS grids do not overlap, SST is taken from a global HYCOM assimilation product. The coupled COAMPS-ROMS model was run for 365 days sequentially in 5-day intervals, starting on 1 Jan 2017. At the end of each 5-day integration, COAMPS was reinitialized using the atmospheric 4D-Var analysis for that day. Two different initialization strategies were adopted in ROMS. In one case, ROMS was initialized using a 4D-Var ocean analysis on 1 Jan 2017 only, after which the ROMS solution at the end of each 5-day interval was used as the ocean initial condition for the next 5-day coupled integration. In a second set of experiments, ROMS was initialized at the start of every 5-day interval using a 4D-Var ocean analysis. Since at least one component of the coupled model was *reinitialized* every 5-days in *all* experiments, the 73 coupled 5-day integrations spanning 2017 in each case should be considered as an ensemble of quasi-independent ocean-atmosphere states, rather than

as one continuous uninterrupted 365-day sequence. Some of the experiments used only the COAMPS-1 grid, others used both COAMPS nested grids, and in some cases the bulk formulation of the surface wind stress was varied. Table 1 summarizes the experiments discussed in section 4.

Experiment name	COAMPS grid	ROMS i.c.	Wind stress
Exp1	45 km	Continuous	$\boldsymbol{\tau} \propto \lvert(\boldsymbol{U}_a - \boldsymbol{u}_o)\rvert(\boldsymbol{U}_a - \boldsymbol{u}_o)$
Exp2	45 km	Continuous	$\boldsymbol{\tau} \propto \lvert\boldsymbol{U}_a\rvert\boldsymbol{U}_a$
Exp3	15 km	Continuous	$\boldsymbol{\tau} \propto \lvert(\boldsymbol{U}_a - \boldsymbol{u}_o)\rvert(\boldsymbol{U}_a - \boldsymbol{u}_o)$
Exp4	15 km	4D-Var	$\boldsymbol{\tau} \propto \lvert(\boldsymbol{U}_a - \boldsymbol{u}_o)\rvert(\boldsymbol{U}_a - \boldsymbol{u}_o)$

Table 1: A summary of coupled COAMPS-ROMS experiments discussed here. The 2nd column indicates the resolution of the COAMPS grid that is *directly* coupled to ROMS. The 3rd column indicates whether ROMS was initialized every 5-days from the ocean state of the preceding 5-day run ("Continuous") or from a 4D-Var analysis for the appropriate day ("4D-Var"). The 4th column indicates the bulk formula used to compute the surface wind stress using only the atmospheric wind ($\boldsymbol{\tau} \propto \lvert\boldsymbol{U}_a\rvert\boldsymbol{U}_a$) or accounting for the relative motion of the atmosphere and ocean ($\boldsymbol{\tau} \propto \lvert(\boldsymbol{U}_a - \boldsymbol{u}_o)\rvert(\boldsymbol{U}_a - \boldsymbol{u}_o)$).

4. Analysis

4.1 SST influences on wind stress

Motivated by the findings of previous studies discussed in section 1, the relationship between SST and wind stress was explored in each of the coupled mode configurations summarized in Table 1. As in section 1, we are primarily interested in the extent to which the atmosphere and ocean are coupled at the ocean mesoscale. The coupled model wind stress and SST were therefore spatially high-pass filtered using a *lowess* smoother in Matlab with a span of 0.12 which closely matches the characteristics of the *loess* filter used by CSS07 (*cf* Figs. 1b and 1c).

Figure 3a shows the August averaged high-pass SST and wind stress magnitude, $|\boldsymbol{\tau}|$, for Exp3 (see Table 1). The degree of coherency between the two fields is very obvious, and agrees qualitatively with similar comparisons based on observations (*e.g.* O'Neill *et al.*, 2012). In general, warm surface conditions coincide with increased wind stress while cooler conditions coincide with reduced wind stress magnitude. The relationship between the wind stress divergence, $\nabla \cdot \boldsymbol{\tau}$, and the downwind component of the SST gradient, $(\nabla T)_d$, computed from the August averaged high-pass fields is shown in Fig. 3b. Again, the degree of coherency between $\nabla \cdot \boldsymbol{\tau}$ and $(\nabla T)_d$ is remarkable. Generally, $\nabla \cdot \boldsymbol{\tau}$ is elevated where $(\nabla T)_d > 0$ and *vice versa*. The exception, however, is the narrow band adjacent to the coast where $(\nabla T)_d$ and $\nabla \cdot \boldsymbol{\tau}$ are negatively correlated. Whether this feature represents a fundamental change in the dominant air-sea interaction processes close to the coast, due perhaps to the influences of the coastline shape, bathymetry or land-sea contrast, deserves further investigation. It is also possible that the calculations of $(\nabla T)_d$ and $\nabla \cdot \boldsymbol{\tau}$ adjacent to the coast are adversely influenced by inconsistencies in the land-sea masking employed in ROMS and COAMPS which have different horizontal resolution, so this will be explored further also. It should be noted that CSS07 excluded wind stress observations within 30 km of the coast due to land contamination of QuikSCAT measurements, and likewise AMSR-E has a relatively low-resolution footprint, so there is no observational evidence for the nearshore correlations in Fig. 3b.

The relationship between the wind stress curl, $\nabla \times \boldsymbol{\tau}$, and the crosswind component of the SST gradient, $(\nabla T)_c$, computed from the August averaged high-pass fields is shown in Fig. 3c. The generally positive correlation between $\nabla \times \boldsymbol{\tau}$ and $(\nabla T)_c$ is very obvious. The exception again is the narrow ribbon of negative correlation between $\nabla \times \boldsymbol{\tau}$ and $(\nabla T)_c$ adjacent to the coast. Binned scatterplots of $|\boldsymbol{\tau}|$ vs SST, $\nabla \cdot \boldsymbol{\tau}$ vs $(\nabla T)_d$, and $\nabla \times \boldsymbol{\tau}$ vs $(\nabla T)_c$ are shown in Figs. 3d, 3e and 3f respectively for the high-pass August averaged fields. In each case, bins containing fewer than 10 pairs of data values were excluded. In addition, based on the generally negative correlations adjacent to the coast in Figs. 3b and 3c, data within 200 km of the coast was excluded, as was data within 150 km of the open boundaries. Figures 3d-f show the mean and standard deviation of the ordinate in each case, and the least-squares straight line fit to the bin mean values is also shown. In each case, the positive correlations away from the coast are very evident and robust.

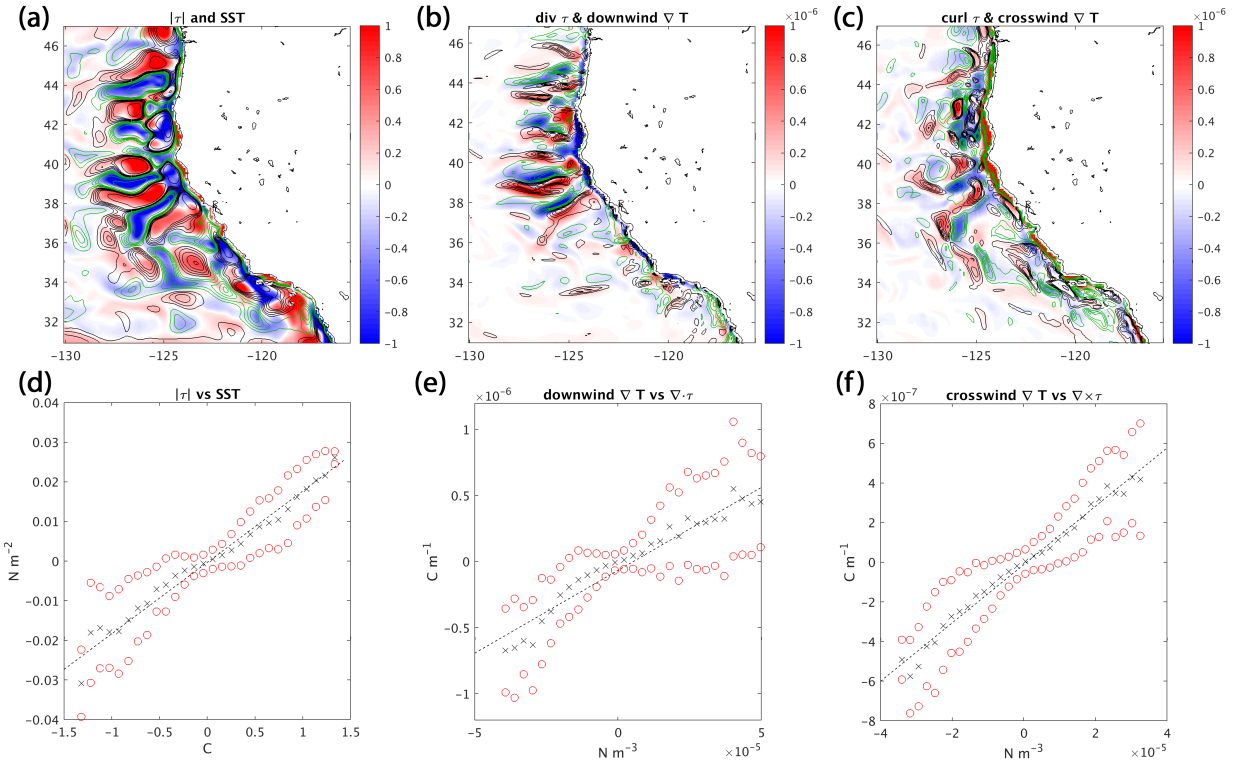

Figure 3: (a) Spatially high-pass filtered August averaged SST (colors) and wind stress magnitude (contours) from Exp3. (b) Same as (a) except for wind stress divergence (colors) and the downwind component of the SST gradient (contours). (c) Same as (a) except for wind stress curl (colors) and the crosswind component of the SST gradient (contours). In all panels red colors and black lines indicate positive values while blue colors and green lines indicate negative values. (d) A binned scatterplot of wind stress magnitude vs SST for the August mean high-pass filtered fields. The means in each bin are denoted by crosses and $\pm$ 1 standard deviation by the circles. The best fit straight line fit to the bin means is indicated by the dotted black line. (e) Same as (d) but for the downwind component of the SST gradient vs wind stress divergence. (f) Same as (d) but for the crosswind component of the SST gradient vs wind stress curl.

4.2 Influence of wind stress formulation on coupling strength

The regression analyses of Figs. 3d-f were performed for all experiments in Table 1 and for each calendar month. As noted in section 1, the gradient of the regression line is a measure of the *coupling strength* between the atmosphere and ocean (*cf* Fig. 1c) and provides a useful way to compare the influence of the different model configurations in Table 1 on air-sea interaction. With this in mind, Fig. 4 shows time series of the *coupling strength* based on $|\boldsymbol{\tau}|$ vs SST, $\nabla \cdot \boldsymbol{\tau}$ vs $(\nabla T)_d$, and $\nabla \times \boldsymbol{\tau}$ vs $(\nabla T)_c$ for each month of 2017 for Exp1 and Exp2. The regression analyses in this case are based on units of N m^{-2} per 10^4 km for $\nabla \cdot \boldsymbol{\tau}$ and $\nabla \times \boldsymbol{\tau}$, and °C per 100 km for the SST gradient which yields estimates of the *coupling strength* that are consistent with those of CSS07 (*cf* "*s*" in Fig. 1c). In both Exp1 and Exp2 ROMS is coupled to the low-resolution (45 km) COAMPS-1 grid. The difference between Exp1 and Exp2 is in the formulation of the surface wind stress. In Exp1, the relative motion between the atmosphere and the ocean is included in the bulk formula for surface stress according to $\boldsymbol{\tau} = \rho_a c_d |(\boldsymbol{U}_a - \boldsymbol{u}_o)|(\boldsymbol{U}_a - \boldsymbol{u}_o)$. Figure 4 shows that Exp1 exhibits a very clear seasonal cycle in *coupling strength* between all fields, with maximum *coupling strength* in the summer in qualitive agreement with the observations. However, the *coupling strengths* between the respective fields are generally lower than observed. For example, Fig. 1c indicates that the *coupling strength*, *s*, between $\nabla \times \boldsymbol{\tau}$ vs $(\nabla T)_c$ in the CCS is 2.13 during summer, while Fig. 4c shows a maximum summertime value of only 1.64. Although it should be noted that Fig. 1c is based on a 4-year average (2002-05) while Fig. 4c represents a single and different year (2017).

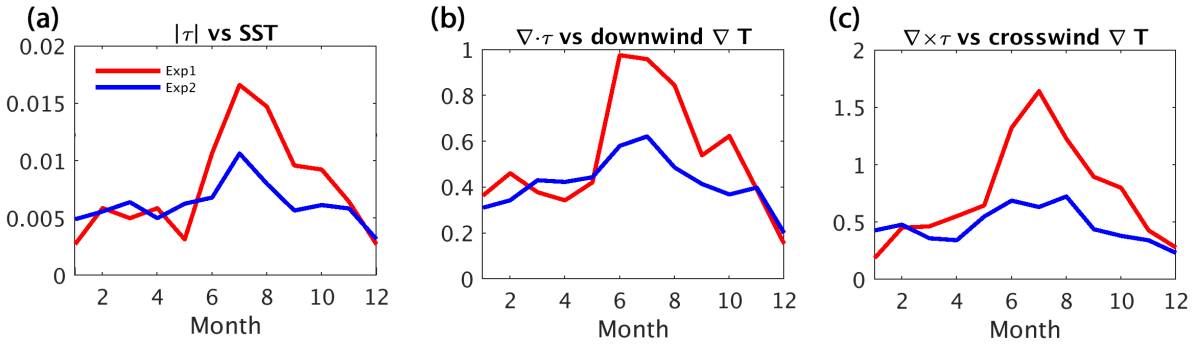

Figure 4: Time series of the regression coefficients (*aka* coupling strength) between (a) SST and wind stress magnitude, (b) the downwind SST gradient and $\nabla \cdot \boldsymbol{\tau}$, and (c) the crosswind SST gradient and $\nabla \times \boldsymbol{\tau}$. Regression coefficients are computed from the spatially high-pass filtered fields from the 45 km COAMPS-1 grid (*cf* Fig.2). Two cases are shown: in Exp1 wind stress is computed by accounting for the relative motion of the ocean and atmosphere (red line), while in Exp2 the relative motion is ignored (blue line). For comparison with Fig. 1c, the units of the regression coefficients are based on SST in °C, SST gradients in °C per 100 km, and wind stress divergence/curl in N m^{-2} per 10^4 km.

In Exp2, the relative motion between the atmosphere and the ocean is ignored in the calculation of the surface stress so that $\boldsymbol{\tau} = \rho_a c_d |\boldsymbol{U}_a| \boldsymbol{U}_a$. In this case, Fig. 4 shows that the seasonal variations in *coupling strength* are reduced significantly compared to Exp1. The winter and spring *coupling strengths* between all fields are similar in both experiments, but the elevated air-sea interaction during summer and fall is suppressed in Exp2. Therefore, the influence of the relative motion between the atmosphere and the ocean appears to be a critical component of atmosphere-ocean coupling during summertime in the CCS.

4.3 Influence of COAMPS resolution on coupling strength

The influence of COAMPS resolution on the *coupling strength* can be evaluated by comparing Exp3 and Exp1. In Exp3 ROMS is coupled directly to the 15 km COAMPS-2 grid while recall that in Exp1 the COAMPS resolution is 45 km. Figure 5 shows time series of *coupling strength* between each field computed from Exp3 and Exp1. Figure 5a indicates that increasing COAMPS resolution generally enhances the *coupling strength* between $|\boldsymbol{\tau}|$ and SST during the period Jan through August. In the case of $\nabla \cdot \boldsymbol{\tau}$ and $(\nabla T)_d$, *coupling strength* is enhanced during most months of the year (Fig, 5b). On the other hand, increasing COAMPS resolution does not significantly enhance the *coupling strength* between $\nabla \times \boldsymbol{\tau}$ and $(\nabla T)_c$ (Fig. 5c).

4.4 The impact of ocean data assimilation on coupling strength

As described in section 3, in Exp3 (Exp1 and Exp2 also) COAMPS was reinitialized at the start of each 5-day coupled integration using a 4D-Var atmospheric analysis while ROMS was initialized from the coupled model solution at the end of the previous 5-day integration. In Exp4, however, ROMS was reinitialized using a 4D-Var ocean analysis. Thus, using the terminology of Zhang *et al.* (2020), Exp4 represents what can be considered as a quasi-weakly coupled data assimilation experiment in that both model components are independently initialized using data assimilation. Figure 5 shows time series of the *coupling strength* between wind stress and SST during Exp4. In Exp4, the *coupling strength* is generally diminished compared to Exp3. In addition, in contrast to Exp3, the *coupling strength* during Exp4 is generally significantly lower during summer and there is no obvious seasonal cycle. The COAMPS initial conditions for each 5-day coupled integration are identical during Exp3 and Exp4. Therefore Fig. 5 indicates that the air-sea interaction processes are quite fragile and sensitive to the periodic disruptions of the ocean state as a result of data assimilation.

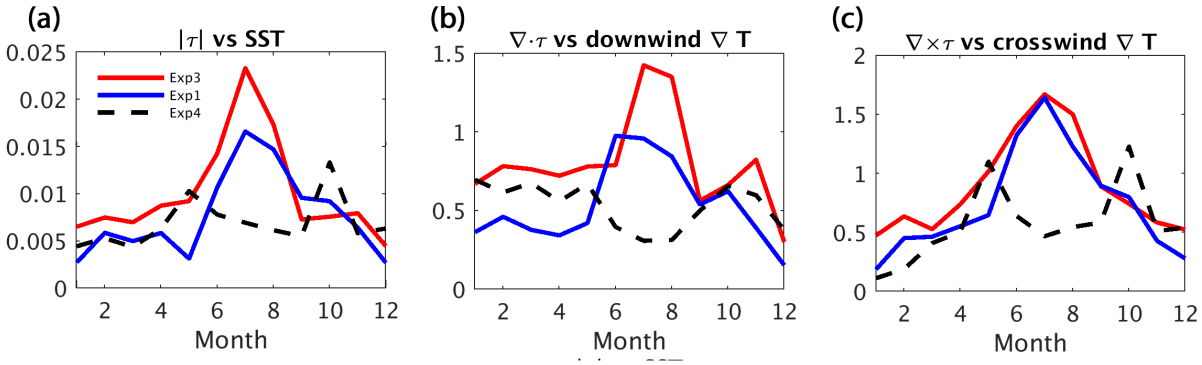

Figure 5: Time series of the regression coefficients (*aka* coupling strength) between (a) SST and wind stress magnitude, (b) the downwind SST gradient and $\nabla \cdot \boldsymbol{\tau}$, and (c) the crosswind SST gradient and $\nabla \times \boldsymbol{\tau}$. Regression coefficients are computed from the spatially high-pass filtered fields, and in all cases wind stress was computed by accounting for the relative motion of the ocean and atmosphere. Results are shown for Exp1 (blue line), Exp3 (red line), and Exp4 (black dashed line). For comparison with Fig. 1c, the units of the regression coefficients are based on SST in °C, SST gradients in °C per 100 km, and wind stress divergence/curl in N m^{-2} per 10^4 km.

5. Summary and Conclusions

Experiments with a prototype regional coupled atmosphere-ocean model of the CCS based on COAMPS and ROMS reveal that, in general, the coupled model is able to qualitatively reproduce the observed relations between SST and surface wind stress at the ocean mesoscale. However, the robustness of these relations is sensitive to the model configuration and model reinitialization procedure. In particular, the influence of the relative motion between the atmosphere and the ocean on surface wind stress appears to be a critical ingredient of atmosphere-ocean coupling for controlling the magnitude of the *coupling strength* and its relative summertime maximum. While atmospheric horizontal resolution significantly influences the *coupling strength* between $|\boldsymbol{\tau}|$ and SST, and $\nabla \cdot \boldsymbol{\tau}$ and $(\nabla T)_d$, the correlation between $\nabla \times \boldsymbol{\tau}$ and $(\nabla T)_c$ is generally much less sensitive to resolution.

Perhaps the most alarming and far-reaching finding here is that the relationship between SST and wind stress at the ocean mesoscale is destroyed when a weakly coupled data assimilation strategy is used. This suggests that the established relationships between SST and wind stress may be quite fragile at the ocean mesoscale, and that strongly coupled data assimilation may be necessary to preserve these important air-sea interactions.

While strongly coupled data assimilation is very challenging and requires specific knowledge of the coupled background error covariances, the analysis of *coupling strength* presented here represent one potential approach for quantifying some of these important covariances, information that could subsequently be used to develop and tune coupled covariance models or parameterizations.

References

Businger, J.A., and Shaw, W.J., 1984: The response of the marine boundary layer to mesoscale variations in sea-surface temperature. *Dyn. Atmos. Oceans*, **8**, 267–281.

Chelton, D.B., Schlax, M.G. and Samelson, R.M., 2007: Summetime coupling between sea surface temperature and wind stress in the California Current system. *J. Phys.Oceanogr.*, **37**, 495-517.

Dever, E.P., Dorman, C.E. and Largier, J.L. 2006: Surface boundary layer variability off northern California, USA during upwelling. *Deep-Sea Res. II*, **53**, 2887–2905.

Dewar, W., and G. Flierl, 1987: Some effects of the wind on rings. *J. Phys. Oceanogr.*, **17**, 1653–1667.

Gaube, P., Chelton, D.B., Samelson, R.M., Schlax, M.G. and O'Neill, L.W., 2015: Satellite observations of mesoscale eddy-induced Ekman pumping. *J. Phys. Oceanogr.*, **45**, 104-132.

Gill, A.E., 1980: Some simple solutions for heat-induced tropical circulation. *Quart. J. R. Meteor. Soc.*, **106**, 447–462.

Haack, T., Chelton, D., Pullen, J., Doyle, J.D. and Schlax, M., 2008: Summertime influence of SST on surface wind stress off the U.S. west coast from the U.S. Navy COAMPS model. *J. Phys. Oceanogr*., 38, 2414–2437.

Huyer, A.E., Fleischbein, J.H., Keister, J., Kosro, P.M., Perlin, N., Smith, R.L. and Wheeler, P.A., 2005: Two coastal upwelling domains in the northern California Current system. *J. Mar. Res.*, **63**, 901–929.

Lindzen, R.S., and Nigam, S., 1987: On the role of sea surface temperature gradients in forcing low level winds and convergence in the tropics. *J. Atmos. Sci*., **44**, 2418–2436.

O'Neill, L.W., Chelton, D.B., Esbensen, S.K., and Wentz, F.J., 2005: High-resolution satellite measurements of the atmospheric boundary layer response to SST variations along the Agulhas Return Current. *J. Climate*, **18**, 2706–2723.

O'Neill, L.W., Chelton, D.B., and Esbensen, S.K., 2012: Covariability of surace wind and stress responses to sea surface temperature fronts. *J. Climate*, **25**, 5916-5942.

Pickett, M.H., and Paduan, J.D., 2003: Ekman transport and pumping in the California Current based on the U.S. Navy's high-resolution atmospheric model (COAMPS). *J. Geophys. Res*., **108**, 3327, doi:10.1029/2003JC001902.

Samelson, R.M., Skyllingstad, E.D., Chelton, D.B., Esbensen, S.K., O'Neill, L.W., and Thum, N., 2006: On the coupling of wind stress and sea surface temperature. *J. Climate*, **19**, 1557-1566.

Seo, H., Miller, A.J., and Roads, J.O., 2007: The Scripps coupled ocean–atmosphere regional (SCOAR) model, with applications in the Eastern Pacific sector. *J. Climate*, **20**, 381–402.

Small, R.J., de Szoeke, S.P., Xie, S.P., O'Neill, L., Soe, H., Song, Q., Cornillon, P., Spall, M. and Minobe, S., 2008: Air-sea interacrion over ocean fronts and eddies. *Dyn. Atmos. Oceans*, **45**, 274-319.

Spall, M.A., 2007: Midlatitude wind stress-sea surface temperature coupling in the vicinity of ocean fronts. *J. Climate*, **20**, 3785-3801.

Zhang, S., Liu, Z., Zhang, X., Wu, X., Han, G., Zhao, Y., Yu, X., Liu, C., Liu, Y., Wu, F., Lu, F., Li, M. and Deng, Z., 2020: Coupled data assimilation and parameter estimation in coupled ocean-atmosphere models: a review. *Climate Dynamics*, **54**, 5127-5144.

Appendix B: Coupled Model Simulations of the Indian Ocean

1. Introduction

Intra-seasonal variability in the tropical Indian Ocean is dominated by the Madden-Julian Oscillation (MJO). Numerical modeling studies and intensive observational field campaigns have shown that air-sea interactions are an important ingredient of the MJO dynamics, although many of the details are still not well understood. Coupled ocean-atmosphere models are an important tool for simulating the MJO, both in terms of understanding the dynamics of individual events and predicting their evolution through time. While many models struggle to generate MJO events spontaneously, data assimilation can be used to initialize subseasonal-to-seasonal (so-called S2S) forecasts using coupled models. However, data assimilation in coupled models is a major challenge, and is a very active area of research.

The work reported on here as part of this ONR funded project represents a stepping-stone toward developing a data assimilation capability for a regional coupled ocean-atmosphere model based on the Regional Ocean Modeling System (ROMS) and the Coupled Ocean-Atmosphere Mesoscale Prediction System (COAMPS). The ROMS-COAMPS system has been configured for the Indian Ocean, and this appendix will focus on two aspects of the coupled model simulations: (i) the ability of the coupled system to maintain MJO events and the impact of weakly coupled data assimilation, and (ii) ensemble prediction of tropical storms spawned by individual MJO events.

2. Coupled Model Description

The model used here comprises the atmospheric component of COAMPS coupled to ROMS, both configured for the Indian Ocean. COAMPS was configured with 15 km horizontal resolution and 60 σ-levels in the vertical and spans the region 35.6ºW-104.4ºW, 20.2ºS-29.2ºN. The ROMS grid has a variable resolution ranging from ~7 km in the equatorial waveguide to ~26 km at the northern and southern boundaries. The ROMS grid does not extend as far east as COAMPS, and SST from global HYCOM was used to constrain COAMPS in the region of the maritime continent. Since ROMS is forced at the boundaries with global HYCOM as well, this avoids spurious discontinuities in the SST field passed to COAMPS.

3. Experimental Set-up

The time interval considered here is October 2011 to March 2012, which includes the DYNAMO observational field campaign (from Sept 2011 to January 2012). In all experiments described here, COAMPS was initialized every 5 days from a global NOGAPS 4D-Var analysis interpolated to the COAMPS grid. Several strategies for initializing ROMS were explored as indicated in Table 1. In each case, the coupled model was run for 5-days before reinitializing both components of the model. During *Exp1*, ROMS was initialized at the start of the first 5-day cycle using the ocean state from a multi-year spin-up integration of the ROMS alone forced with surface fluxes derived from ERA-interim. Thereafter, during *Exp1* ROMS was initialized using the ocean state at the end of the previous 5-day coupled model integration. In *Exp2*, ROMS was reinitialized every 5-days using a 4D-Var ocean state estimate in which all available remote

sensing and *in situ* observations from the previous 5-day period were assimilated. *Exp3* was similar to *Exp2* except that the ROMS initial condition was computed using an Ensemble Adjustment Kalman Filter (EAKF) from the DART system. In this case, a First Guess at Appropriate Time (FGAT) approach was used in which all observations over a 2.5-day interval were assimilated into ROMS using a 31-member ensemble. Experiments *Exp2* and *Exp3* represent examples of weakly coupled data assimilation in which data were assimilated independently into both component models. In addition, a *Control* experiment was also performed in which COAMPS was initialized every 5-days and run uncoupled using SST from a HYCOM global analysis as the ocean surface boundary conditions.

Experiment designation	Ocean/ROMS initial conditions
Control	observed SST
Exp1	ocean state at end of previous cycle
Exp2	4D-Var ocean analysis
Exp3	EAKF ocean analysis

Table 1: A summary of the experiments performed, the experiment designation, and the approach used to initialize the ROMS component of the coupled model.

4. Analysis

4.1 MJO hindcast skill

The period of study was chosen because it overlaps with an intensive observational field campaign (DYNAMO) targeting the MJO. During the period Oct-Dec 2011, three MJO events were observed and simulated by COAMPS. Figure 1 quantifies the extent to which the *Control* experiment (Table 1) is able to capture the equatorial zonal wind stress (the Indian Ocean region between 5ºS-5ºN). Specifically, Fig. 1 shows the results of a comparison of τ_x from the *Control* simulation with τ_x from ERA-interim, where the latter is taken as an estimate of the true state. Each 5-day simulation of the *Control* can be viewed as a hindcast, and in Fig. 1a, the correlation between τ_x from the *Control* and ERA-interim is shown as a function of hindcast lead-time and start date. Similarly, Fig. 1b shows, in the same format, the root mean square difference (rmsd) between the equatorial τ_x from the *Control* and ERA-interim. In general, the *Control* struggles to capture the evolution of τ_x during the period Oct-Dec. As noted, this period was characterized by three MJO events which are illustrated in Figs. 1c and 1d, which show identical Hovmoller diagrams of the equatorial τ_x versus time from ERA-interim. The MJO events are indicated by the colored vertical lines in Figs. 1c and 1d (and also in Figs. 1a and 1b). To more clearly illustrate the skill of COAMPS in hindcasting these three events in the presence of the observed SSTs, Figs. 1e and 1f show time series of the correlation and rms errors for τ_x for each 5-day hindcast cycle. The skill in hindcasting the Oct and Dec MJO events is consistent with the range associated with non-MJO periods, although the skill associated with the Nov MJO episode is clearly anomalously low. Animations of the surface wind fields (not shown) indicate that while the NOGAPS analyses (from which COAMPS is initialized) capture the observed MJO events quite well, they subsequently dissipate quite quickly within COAMPS, perhaps because of geostrophic adjustment associated with dynamic imbalances in the initial state. Conversely,

during the period Jan-Mar 2012 when the MJO is not active, Fig. 1 indicates that COAMPS can hindcast quite well the ERA-interim zonal wind stress for many of the 5-day intervals.

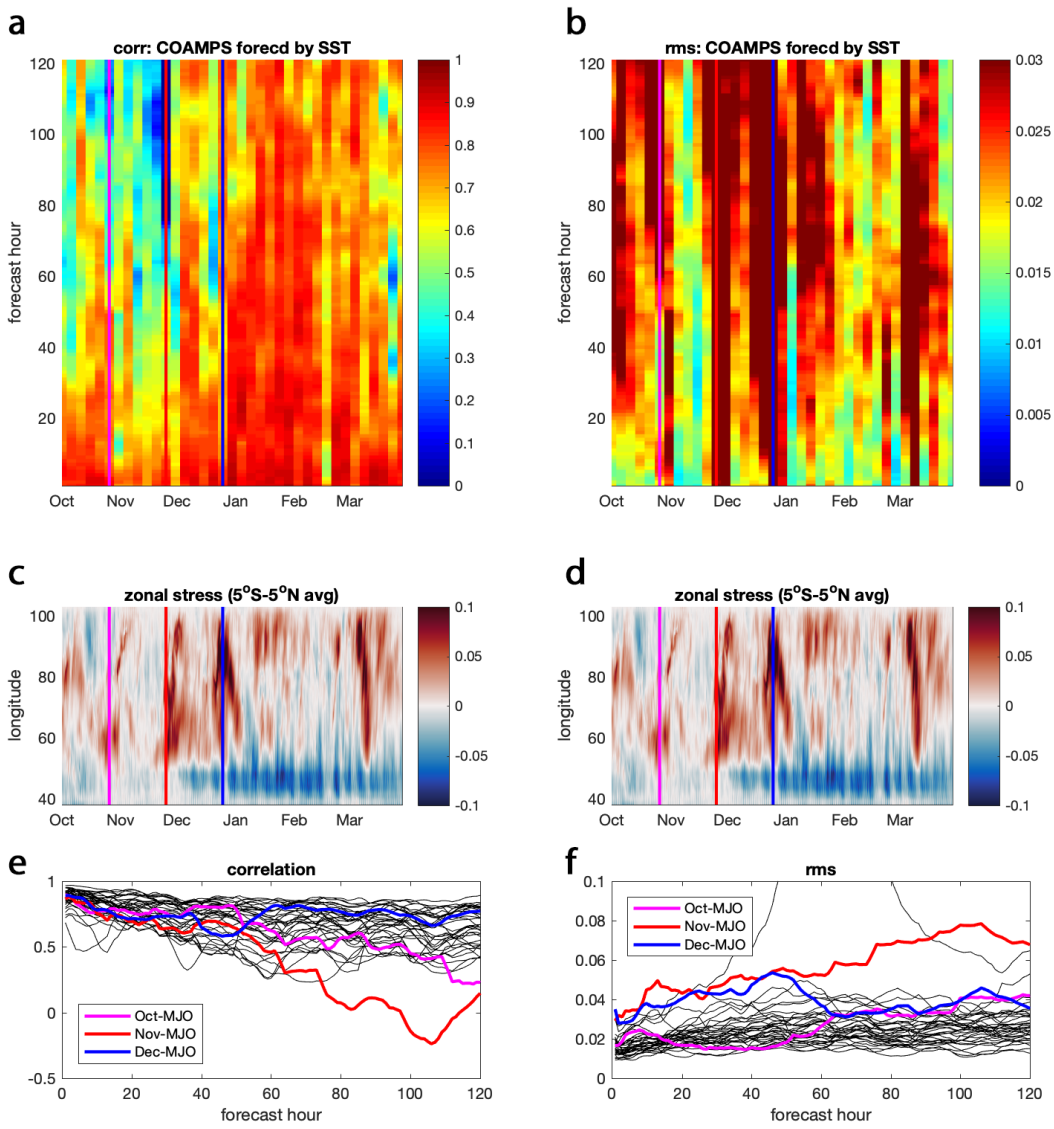

Figure 1: (a) The correlation between the equatorial (*5ºS-5ºN*) zonal wind stress (τ_x) from the *Control* experiment and ERA-interim versus hindcast lead time for each 5-day coupled model integration. (b) The root mean square difference (rmsd) between τ_x from the *Control* experiment and ERA-interim versus hindcast lead time for each 5-day coupled model integration. Hovmoller diagrams of the ERA-interim τ_x (*5ºS-5ºN* mean) are shown in (c) and (d) which are identical. The colored vertical lines (also shown in (a) and (b)) indicate the timing of three MJO events during the Oct 2011 to Mar 2012 period. (e) Times series of the correlation between the equatorial τ_x from the *Control* experiment and ERA-interim versus hindcast lead time for each 5-day coupled model integration. The periods spanned by the three MJO events highlighted in (c) and (d) are indicated by the colored lines. (f) Same as (e) except for the rms difference.

The hindcast skill for τ_x during *Exp1*, *Exp2* and *Exp3* is shown in Fig. 2. In this case the difference between the correlation and rmsd of each experiment and that of the *Control* is

illustrated to more clearly highlight the influence of the different ROMS initialization strategies. In general, Figs. 2b and 2f suggest that *Exp1* yields generally lower hindcast skill for τ_x than the *Control* during the MJO active period Oct-Dec, while during the non-active interval Jan-Mar, the skill is generally a little higher.

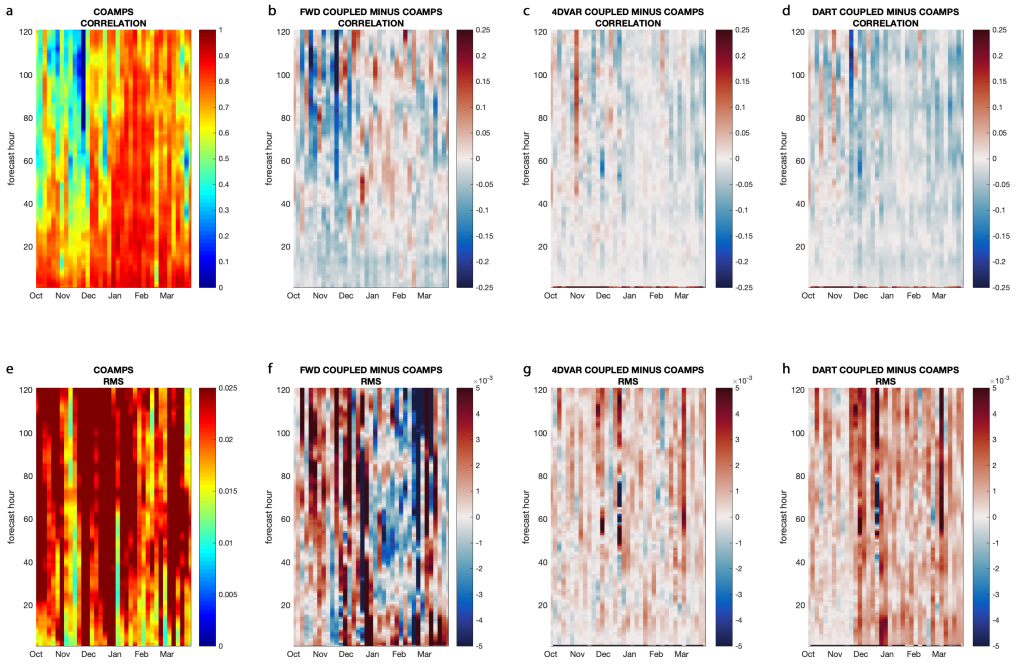

Figure 2: (a) The correlation between the equatorial τ_x from the *Control* experiment and ERA-interim versus hindcast lead time for each 5-day coupled model integration. (b) The difference between the τ_x correlations of *Exp1* with ERA-interim and those of the *Control* in (a). (c) Same as (b) but for *Exp2*. (d) Same as (b) but for *Exp3*. (e) The rmsd between τ_x from the *Control* experiment and ERA-interim versus hindcast lead time for each 5-day coupled model integration. (f) The difference between the rmsd of *Exp1* and ERA-interim and the rmsd of the *Control*. (g) Same as (e) except for *Exp2*. (h) Same as (e) except for *Exp3*.

During *Exp2*, where ROMS was initialized every 5-days using a 4D-Var analysis, Fig. 2c indicates that the hindcast skill in τ_x based on correlation is marginally higher than the *Control* during the active MJO period and worse during the non-active period. However, in the case of *Exp3*, where ROMS was initialized using analyses from the DART EAKF, Figs. 2d and 2h indicate that the hindcast skill is generally always lower than the *Control*.

4.2 Tropical storm ensembles

During the active MJO period Oct-Dec 2011, several tropical depressions and cyclones were spawned in the Indian Ocean. Because *Exp3* is based on an ensemble Kalman filter, we also have the opportunity to explore the sensitivity of hindcasts for the storm track and intensity to the ocean state using the 31 individual ensemble members. To this end, two case studies were considered: (i) tropical depression ARB 04 (Nov 26-Dec 1, 2011), with an observed minimum central pressure of 998mb (peak winds 55 km hr^{-1}), and (ii) tropical cyclone Thane (Dec 25-31, 2011), with an observed minimum central pressure of 969mb (peak winds 140 km hr^{-1}), which

developed into a category 2 storm. For both cases, the 31 individual ROMS ensemble members from the EAKF on Nov 25 (TD ARB 4) and Dec 25, 2011 (TC Thane) were used to initialize an ensemble of coupled model integrations of 5-days duration. In each case, COAMPS was initialized with the same NOGAPS analysis for the appropriate day.

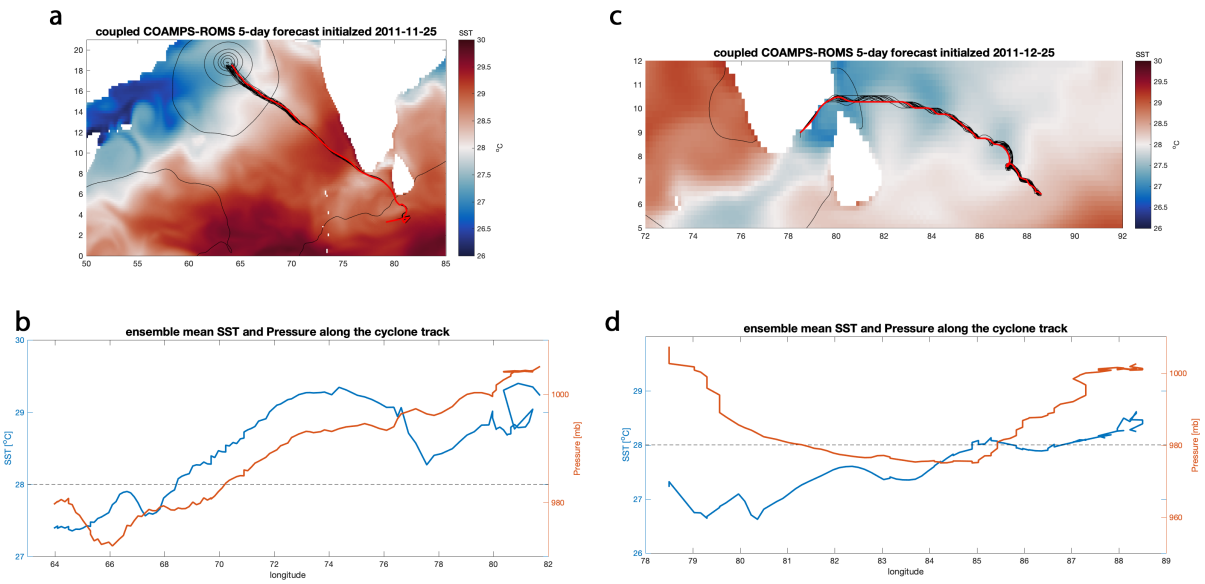

Figure 3: An ensemble of storm tracks from the coupled model for (a) tropical depression ARB 4, and (c) tropical cyclone Thane. Individual ensemble members are indicated by the thin black lines and the red line shows the ensemble mean. In each case, the color map shows the ensemble mean SST, and the contours show the ensemble mean sea level pressure. (b) The SST (blue line) and minimum surface pressure (red line) versus longitude along the path of tropical depression ARB 4. (d) Same as (b) but for tropical cyclone Thane. The horizontal dashed line in (b) ad (d) indicates 28ºC, the generally accepted threshold required to maintain a tropical storm via convective instability.

Figures 3a and 3c show the ensemble of tracks associated with each storm, along with the path of the ensemble mean. For reference, the ensemble mean SST and sea level pressure are also shown for the Nov 25 and Dec 25 start dates. In both cases, the storm track is relatively insensitive to the ocean state, probably because both were relatively weak storm systems. Nonetheless, Fig. 3 illustrates that the coupled model and DART-EAKF have tremendous potential utility in the Indian Ocean for ensemble coupled prediction of tropical cyclones. Figures 3a and 3c indicate that both storms move over regions of cooler SST as they progress northwestward along their respective paths. Hence, they may be expected to weaken. Figures 3b and 3d illustrate how the central pressure of each storm along its track varies with the underlying SST. As anticipated, the central pressure falls as the storm moves over warmer water and rises again as it passes over water cooler than ~28ºC, the generally accepted threshold required to maintain convective instability.

The relationship between tropical cyclone Thane and ocean temperature is further revealed in Fig. 4, showing shows vertical sections of the upper ocean temperature along the cyclone track on selected days. Figure 4 clearly shows that the upper ocean temperature behind the storm cools as the cyclone progresses northwestward. The formation of a cold wake by hurricanes and tropical cyclones is well documented in observations and some models. The cooling is associated

with a combination of increased evaporative cooling and entrainment of cold water at the base of the ocean mixed layer as it deepens in response to an increase in wind-induced turbulent mixing.

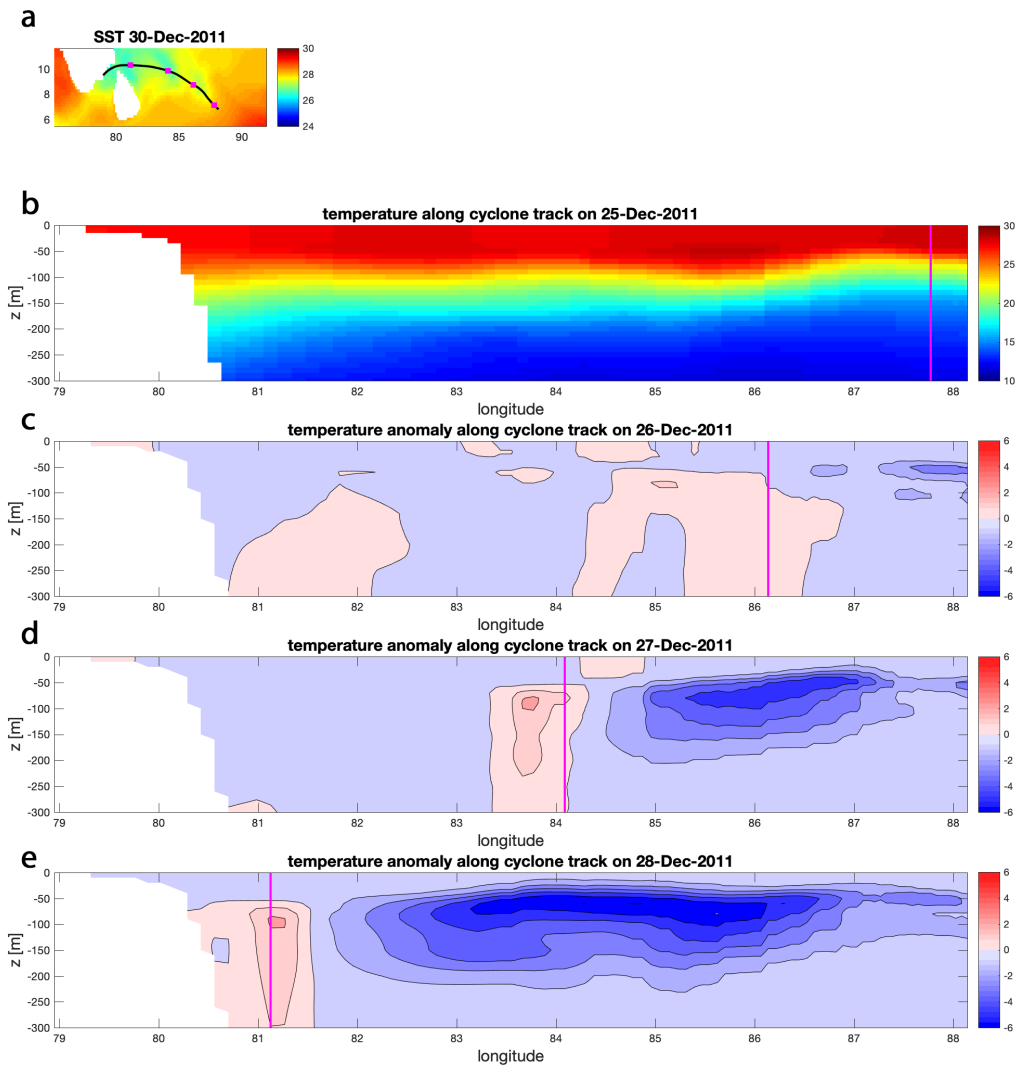

Figure 4: (a) The path of tropical cyclone Thane superimposed on the SST. The magenta squares show the location of the storm on 25, 26, 27 and 28 December. (b) A vertical section of the upper ocean temperature (0-300 m) on Dec 25 along the full length of the cyclone track. The location of tropical cyclone Thane on this date is indicated by the magenta line. (c) An upper-ocean vertical section (0-300 m) of the difference in temperature between 26 and 25 Dec along the cyclone track. The location of tropical cyclone Thane on this date is indicated by the magenta line. Blue (red) indicates that the ocean is cooler (warmer) on 26 Dec than it was on the 25 Dec. (d) Same as (c) except for 27 Dec. (e) Same as (c) except for 28 Dec.

5. Summary and Conclusions

Preliminary investigations have been performed of coupled ocean-atmosphere simulations of the Indian Ocean using ROMS and COAMPS. One focus of this work was the simulation of the

MJO. The period considered corresponds to active and break periods of the MJO and overlaps with the DYNAMO intensive observational field campaign. Uncoupled integrations of the atmospheric component of COAMPS alone subject to observed SST, and initialized with a global NOGAPS analysis, show that COAMPS is unable to maintain MJOs, despite these events being captured quite well by NOGAPS. Coupled model simulations are perhaps a little better during MJO active periods, but further experiments show that attempts to improve the ocean state using a weakly coupled data assimilation approach do not significantly improve the coupled model's ability to hindcast the MJO, and can, in fact, degrade the model simulations in some cases.

The second focus of this work was the utility of coupled model ensembles which arise, quite naturally, from the ensemble Kalman filter employed in one of the experiments described here. Specifically, we have explored the influence of the ocean state on tropical storm hindcasts using an ensemble of ocean state estimates. While we did not find much sensitivity in the path and intensity for the relatively weak storms considered here, the coupled model in tandem with the EAKF shows great promise as a tool for ensemble prediction of tropical cyclones. It was also gratifying to see that in the coupled model, a cold wake forms behind a developing tropical cyclone, confirming what is commonly seen in observations.

Appendix C: ROMS-DART

Progress in Oceanography 189 (2020) 102450

Contents lists available at ScienceDirect

Progress in Oceanography

journal homepage: www.elsevier.com/locate/pocean

Regional and basin scale applications of ensemble adjustment Kalman filter and 4D-Var ocean data assimilation systems

Andrew Moore [a,*], Javier Zavala-Garay [b], Hernan G. Arango [b], Christopher A. Edwards [a], Jeffrey Anderson [c], Tim Hoar [c]

[a] Department of Ocean Sciences, University of California, Santa Cruz, CA, USA
[b] Department of Marine Sciences, Rutgers University, New Brunswick, NJ, USA
[c] National Center for Atmospheric Research, Boulder, CO, USA

ARTICLE INFO

Keywords:
Ensemble Kalman filter
4D-Var
California Current
Indian Ocean

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 of 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.

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

* Corresponding author at: Department of Ocean Sciences, University of California Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA.
E-mail address: ammoore@ucsc.edu (A. Moore).

https://doi.org/10.1016/j.pocean.2020.102450
Received 13 May 2020; Received in revised form 16 August 2020; Accepted 29 September 2020
Available online 3 October 2020

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 $\boldsymbol{x}$. If $\boldsymbol{x}_b$ indicates the background state (aka the *prior*, *first-guess* or *forecast*), then the best, linear, unbiased circulation estimate $\boldsymbol{x}_a$, also referred to as the analysis, arising from data assimilation can be expressed as:

$$\boldsymbol{x}_a = \boldsymbol{x}_b + \boldsymbol{K}\boldsymbol{d} \tag{1}$$

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

$$\boldsymbol{K} = \boldsymbol{B}\boldsymbol{H}^T\left(\boldsymbol{H}\boldsymbol{B}\boldsymbol{H}^T + \boldsymbol{R}\right)^{-1} \tag{2}$$

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

$$\boldsymbol{F} = \boldsymbol{M}\boldsymbol{B}\boldsymbol{M}^T + \boldsymbol{Q} \tag{3}$$

where $\boldsymbol{Q}$ 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 $\mathscr{M}$ is used to evolve the state and to evaluate $H(\boldsymbol{x}_b)$, while $\boldsymbol{M}$ 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, $\boldsymbol{B}$ (and $\boldsymbol{F}$) are estimated from an ensemble of model solutions, $\boldsymbol{x}_n$. Moreover, since $\boldsymbol{B} \approx \overline{(\boldsymbol{x}_n - \overline{\boldsymbol{x}})(\boldsymbol{x}_n - \overline{\boldsymbol{x}})^T}$ (where an overbar denotes the expected value), the action of $\boldsymbol{B}$ can be conveniently represented using the ensemble members directly. More specifically, $\boldsymbol{B}\boldsymbol{H}^T \approx \overline{(\boldsymbol{x}_n - \overline{\boldsymbol{x}})\left(H(\boldsymbol{x}_n) - \overline{H(\boldsymbol{x}_n)}\right)^T}$ and $\boldsymbol{H}\boldsymbol{B}\boldsymbol{H}^T \approx \overline{\left(H(\boldsymbol{x}_n) - \overline{H(\boldsymbol{x}_n)}\right)\left(H(\boldsymbol{x}_n) - \overline{H(\boldsymbol{x}_n)}\right)^T}$. A thorough review of the KF approach can be found in Houtekamer and Zhang (2016).

In 4D-Var, the observation operator $H(\boldsymbol{x}_b)$ 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 $\boldsymbol{B}$ in 4D-Var is typically 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 $\boldsymbol{H}\boldsymbol{B}\boldsymbol{H}^T$ and $\boldsymbol{B}\boldsymbol{H}^T$ in (2) since $\boldsymbol{H}$ includes the tangent linearization of the forecast model. A common approach to modeling the action of $\boldsymbol{B}$ 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.* $H(\boldsymbol{x}_n)$) 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,

[1] Here and throughout, we have adopted the generally accepted compact notation for the operators $\boldsymbol{H}$ and $\boldsymbol{R}$ which represent appropriate block diagonal matrices and each block corresponds to an observation time (*e.g.* see Gürol et al., 2014).

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 $\boldsymbol{B}$ 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 $\boldsymbol{B}$.

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 $\delta\boldsymbol{x} = \boldsymbol{x} - \boldsymbol{x}_b$ represent the departure of the state-vector from the background. The analysis $\boldsymbol{x}_a = \boldsymbol{x}_b + \delta\boldsymbol{x}$ is identified as that increment that minimizes the linearized cost function:

$$J = \delta\boldsymbol{x}^T\boldsymbol{B}^{-1}\delta\boldsymbol{x} + (\boldsymbol{H}\delta\boldsymbol{x} - \boldsymbol{d})^T\boldsymbol{R}^{-1}(\boldsymbol{H}\delta\boldsymbol{x} - \boldsymbol{d}) \tag{4}$$

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 $\boldsymbol{B}$ can then be factorized as $\boldsymbol{B} = \boldsymbol{K}_b\Sigma\boldsymbol{C}\Sigma^T\boldsymbol{K}_b^T$, where $\boldsymbol{C}$ is the univariate correlation matrix of the unbalanced variables, Σ is a diagonal matrix of standard deviations of the errors in the unbalanced variables, and $\boldsymbol{K}_b$ is a multivariate balance operator (Derber and Rosati, 1989). The correlation matrix $\boldsymbol{C}$ 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, $\boldsymbol{C}$ is typically assumed to be time-invariant.

Two approaches are adopted in ROMS for specifying Σ. One approach uses error statistics that reflect those associated with asymptotic climatological errors. For this reason, $\boldsymbol{B}$ 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 φ_b and the true value φ_t is due to a vertical displacement of the profile, in which case, to first-order, $\varphi_t(z) \approx \varphi_b(z+\delta z) + (\partial\varphi_b/\partial z)\delta z$, where δz 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 φ lies somewhere in the background water column. The error in φ_b is, therefore, approximately given by $(\partial\varphi_b/\partial z)\delta z$. Following Mogensen et al. (2012), we assume that the standard deviation of the error is given by $\sigma_\varphi \sim |(\partial\varphi_b/\partial z)\delta z|$, with the following constraints:

$$\sigma_\varphi = \begin{cases} max\left(\widehat{\sigma_\varphi}, \sigma_\varphi^{ml}\right) if\ z \geq -D_{ml} \\ max\left(\widehat{\sigma_\varphi}, \sigma_\varphi^{do}\right) if\ z < -D_{ml} \end{cases} \tag{5}$$

where:

$$\widehat{\sigma_\varphi} = \min\left(|(\partial\varphi_b/\partial z)\delta z|, \sigma_\varphi^{max}\right) \tag{6}$$

The prescribed parameter σ_φ^{max} is the maximum value of σ_φ, while σ_φ^{ml} and σ_φ^{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., 2017). 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 σ-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).

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 Huddleston, 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

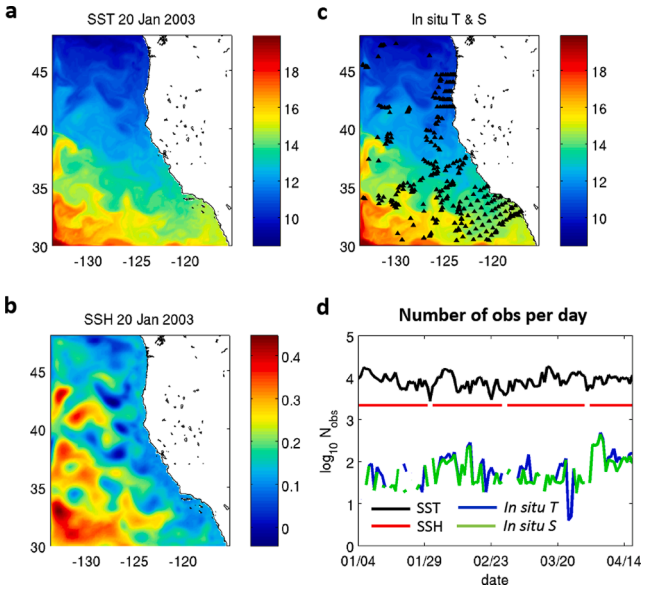

Fig. 1. The ROMS-CCS configuration used in the EAKF and 4D-Var experiments. The SST and SSH on a date close to the start of each experiment are shown in (a) and (b), respectively. The locations of all *in situ* hydrographic observations collected during the Jan-April 2003 period are shown in (c), and time series of $\log_{10}$ 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).

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.

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, Σ, were based on either climatological standard deviations (denoted Σ_c) of a long run of the model without data assimilation, varying from month-to-month, or on (5) and (6) (denoted Σ_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 $\boldsymbol{HBH}^T$ to provide cross-covariance information. Table 2 summarizes the 4D-Var experiments performed as part of this study.

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 $\boldsymbol{HBH}^T$. 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.

The location chosen for the simulated vertical temperature profile is close to San Francisco Bay, as indicated in Fig. 2a. Fig. 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 $\boldsymbol{x}_b$ 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. Fig. 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 (Fig. 2e and f) are

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.

Experiment name	FGAT (days)	Ensemble size N	H localization (km)	V localization (m)
Baseline	1	50(51)	130	200
EAKF1a	2	50(51)	130	200
EAKF1b	4	50(51)	130	200
EAKF2a	1	10(11)	130	200
EAKF2b	1	30(31)	130	200
EAKF3a	1	50(51)	260	200
EAKF3b	1	50(51)	65	200
EAKF4a	1	50(51)	130	–
EAKF4b	1	50(51)	130	100

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 Σ_c while the prefix M denotes experiments that use the error model Σ_m described by Eqs. (5) and (6). The data assimilation window length is indicated in column 2, and the scaling factor α that was applied to the standard parameter values in (5) 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.

Experiment name	Window (days)	Scaling (α)	H decorrelation (km)	V decorrelation (m)
M4DVAR1	1	1	25	90
M4DVAR2	2	1	25	90
M4DVAR4	4	1	25	90
M4DVAR8	8	1	25	90
C4DVAR8h25	8	–	25	90
C4DVAR8h50	8	–	50	90
C4DVAR8h75	8	–	75	90
C4DVAR8v30	8	–	25	30
C4DVAR8v60	8	–	25	60
C4DVAR8v90	8	–	25	90
M4DVAR8a	8	0.25	25	90
M4DVAR8b	8	0.5	25	90
M4DVAR8c	8	1.5	25	90

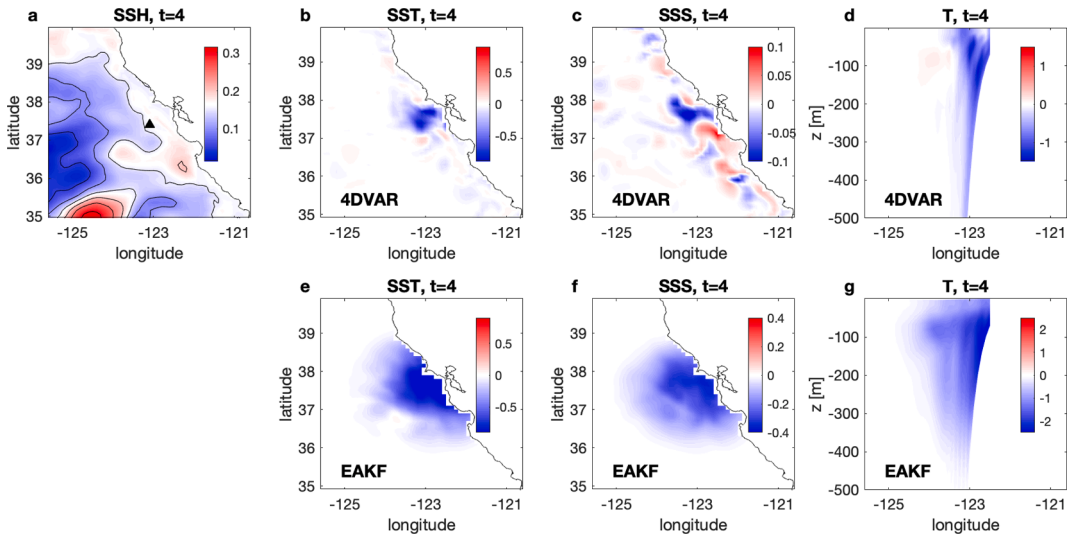

Fig. 2. The background SSH in the vicinity of central California on 8 Jan 2003. The triangle indicates the position of the single vertical temperature profile assimilated on the same day. 4D-Var analysis increments on 8 Jan 2003 of (b) SST and (c) sea surface salinity (SSS). (d) A zonal, vertical section 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.

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 to be confused with any imposed correlation length. The 4D-Var increments, on the other hand (Fig. 2b and c), 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 $\boldsymbol{B}$, which is 25 km in this case. The complex structure is due to the time evolution of $\boldsymbol{x}_b$ via the covariance $\boldsymbol{HBH}^T$ and the term $\boldsymbol{BH}^T$, 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 Fig. 2d and g 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 $\boldsymbol{R}$. 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 $\boldsymbol{Q} = 0$ in (3). The 4 Jan 2003 initial condition used for the data assimilation experiments will be denoted as $\boldsymbol{x}_0$, and two cases were considered: (i) $\boldsymbol{x}_0^{\mathrm{i}}$ 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) $\boldsymbol{x}_0^{\mathrm{ii}}$ 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 $\boldsymbol{x}_0$ on 4 Jan 2003 differs significantly in both cases from the nature run solution on the same day. Since $\boldsymbol{x}_0^{\mathrm{i}}$ 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 $\delta\boldsymbol{x}_j$ from January were randomly selected from the WCRA31 reanalysis of N16. An ensemble of initial conditions was then created according to $\boldsymbol{x}_0 + \delta\boldsymbol{x}_j$. The rationale for this approach is that the $\delta\boldsymbol{x}_j$ 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 $\boldsymbol{x}_0 + \delta\boldsymbol{x}_j$ 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, $\Delta\boldsymbol{x}_j$, between randomly chosen pairs of these January solutions were then used as initial perturbations for the ensemble members, according to $\boldsymbol{x}_0 + \Delta\boldsymbol{x}_j$. Two additional variants of Methods A and B were also considered, in which each perturbation was used twice according to $\boldsymbol{x}_0 \pm \delta\boldsymbol{x}_j$ or

$\boldsymbol{x}_0 \pm \Delta\boldsymbol{x}_j$. By including the unperturbed ensemble member, this approach yields an ensemble mean initial condition equal to the unperturbed ensemble member, $\boldsymbol{x}_0$, and the ensemble is centered about $\boldsymbol{x}_0$. 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 $\boldsymbol{x}_0$ 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 $\boldsymbol{x}_0$) 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.

Fig. 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 (ζ) 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 $\boldsymbol{x}_0^{\mathrm{i}}$ and run forward in time subject only to the COAMPS-derived surface forcing and the SODA open boundary conditions. Fig. 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 ζ (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 ζ 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

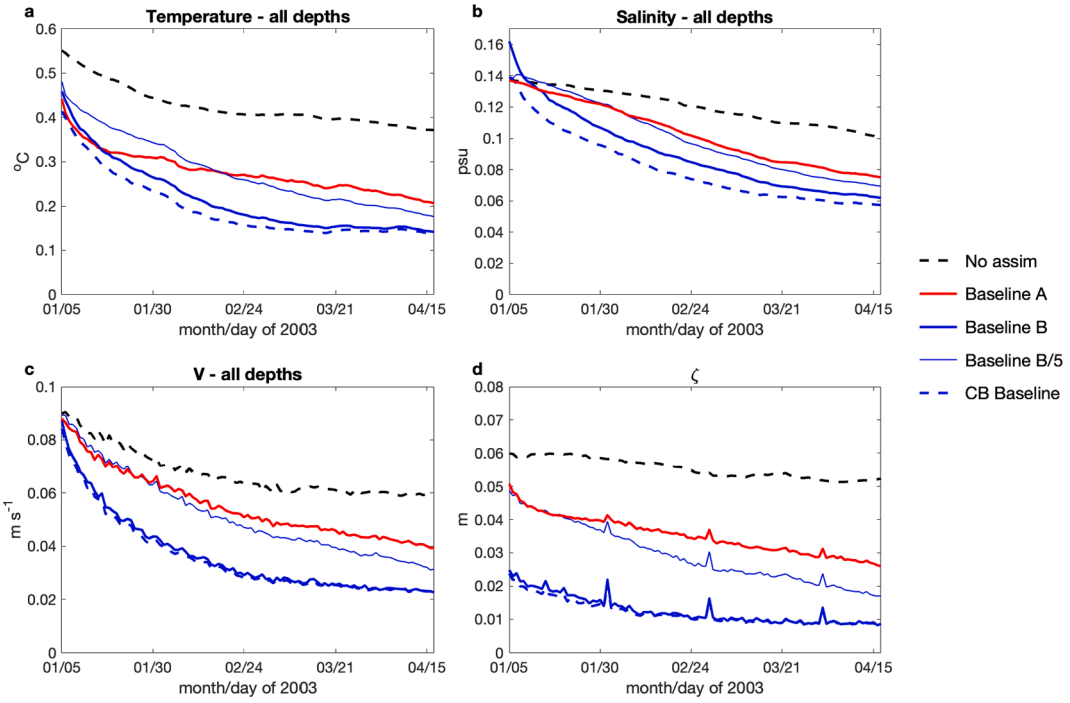

Fig. 3. Time series of the root mean square error in ensemble mean (a) temperature, (b) salinity, (c) meridional velocity, and (d) SSH for various EAKF experiments using different methods to generate the initial ensemble. The baseline cases using methods A (red line) and B (bold blue line) are shown, along with a centered baseline case using method B (blue dashed line). A separate case using method B, where all ensemble perturbations were scaled by a factor of 0.2, is also shown (thin blue line). For reference, the case where no data were assimilated is included (black dashed line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

somewhat faster than the uncentered case for *T* and *S*.

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). Fig. 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 (Fig. 4a–c) and the centered method B baseline case (Fig. 4d–f). The dates chosen correspond to assimilation cycles near the beginning (4 Jan, Fig. 4a, d), middle (25 Feb, Fig. 4b, e), and end (17 April, Fig. 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).

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 (Fig. 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

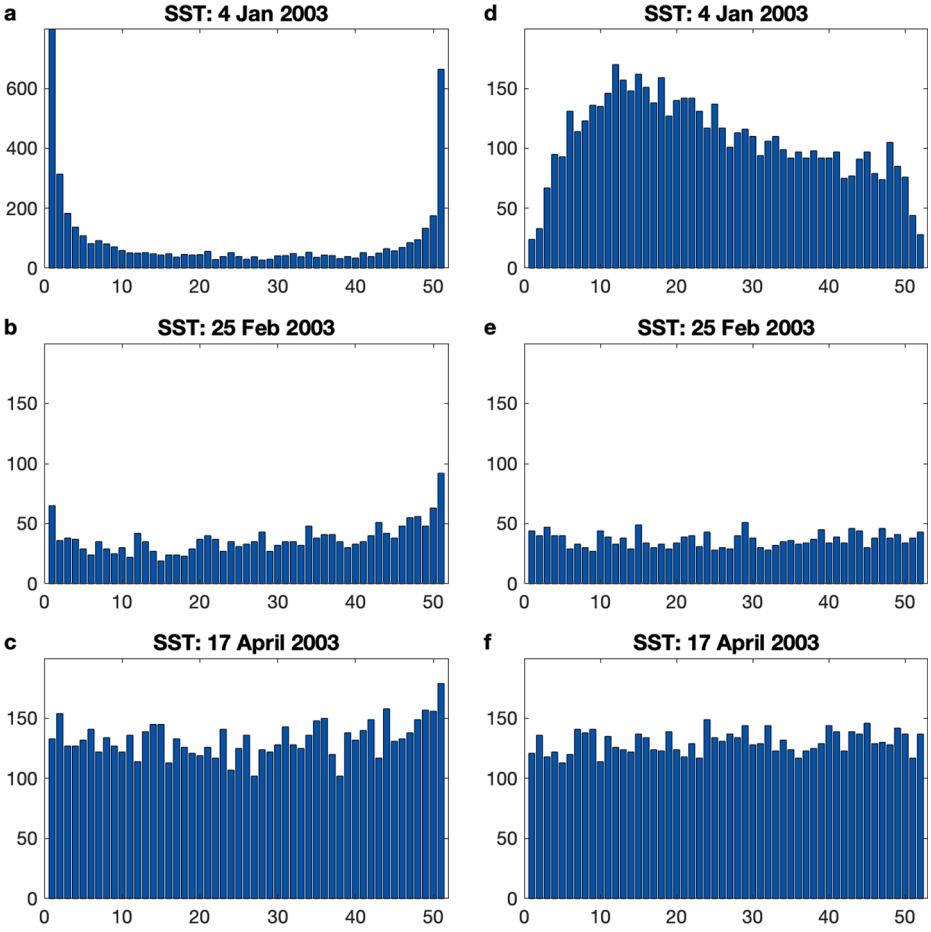

Fig. 4. Rank histograms based on SST observations on three different days for (a-c) the method B baseline case with ensemble perturbations scaled by 0.2, and (d-f) the centered method B baseline case. Note the expanded scale of the ordinate in (a).

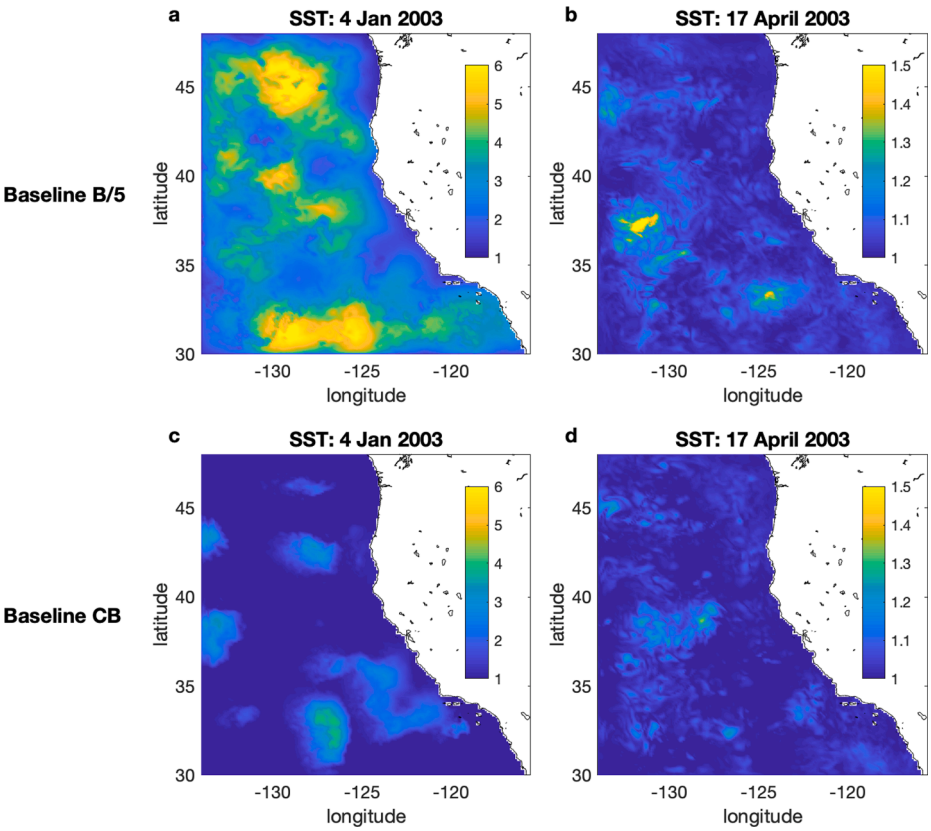

Fig. 5. SST inflation factors on 4 Jan and 17 April for (a, b) the method B baseline with ensemble perturbations scaled by 0.2, and (c, d) the centered method B baseline case. Note the different color bar scales used in each case.

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).

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. Fig. 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 $\boldsymbol{B}$. 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 °C, ~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.

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 Σ_m based on the default parameters for (5) and (6) (Table 3). In each experiment, the ocean analysis $\boldsymbol{x}_a$ at the end of each 4D-Var cycle was used as the background initial condition $\boldsymbol{x}_b$ 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,

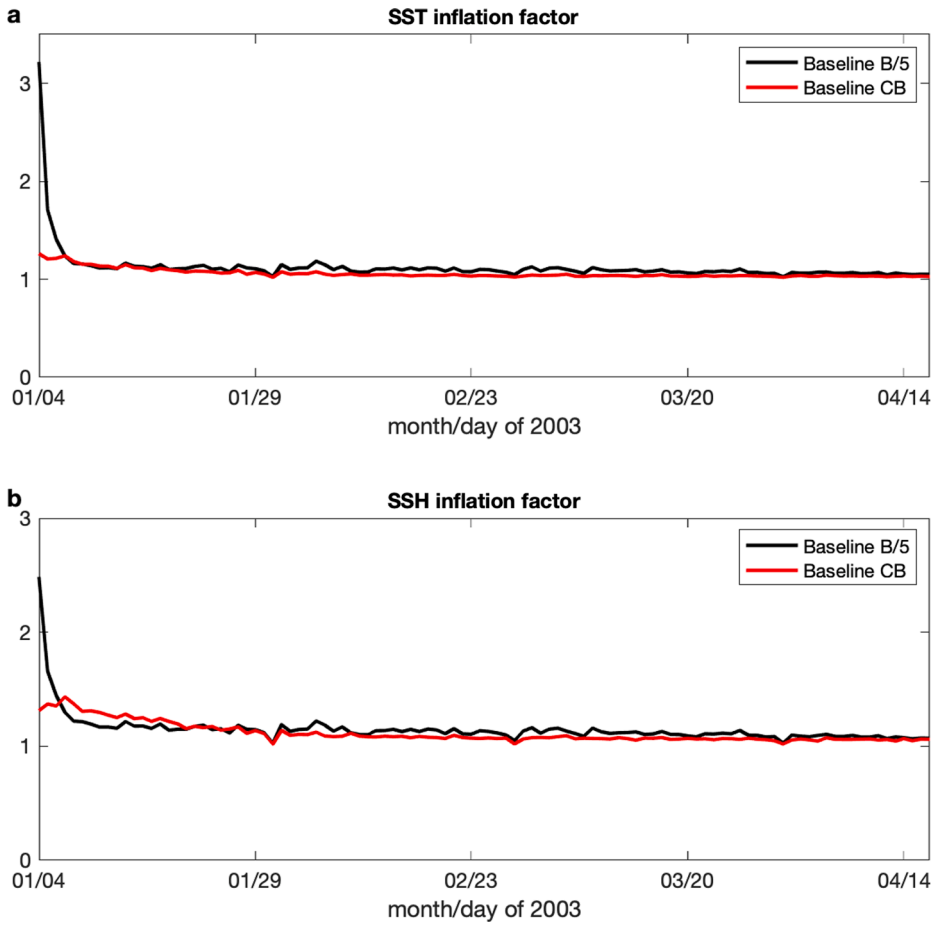

Fig. 6. Time series of the root mean square inflation factors for (a) SST and (b) SSH for the method B baseline with ensemble perturbations scaled by 0.2 (black line), and the centered method B baseline case (red line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

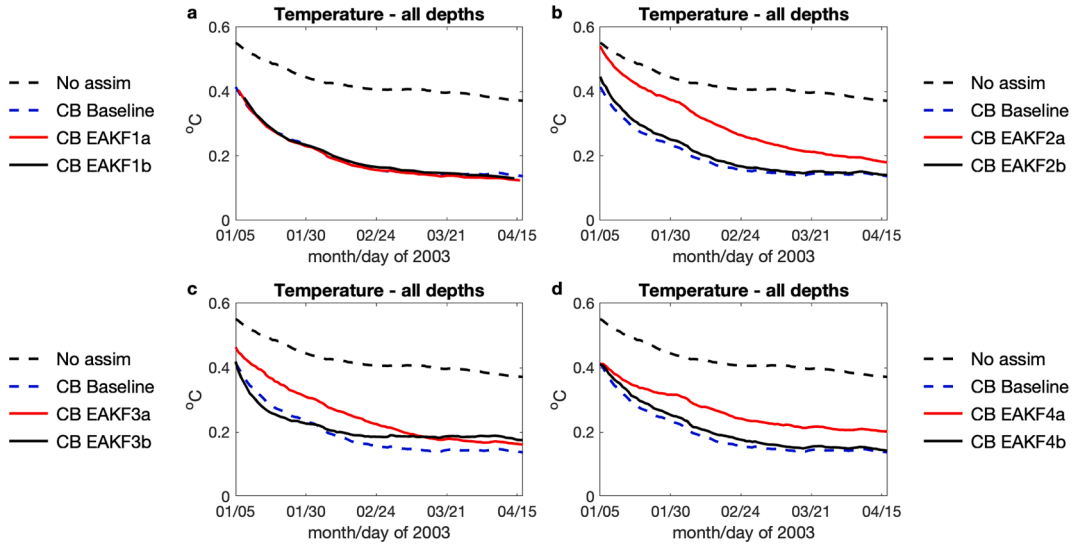

Fig. 7. Time series of RMS errors in ensemble mean temperature for EAKF cases involving variations in (a) FGAT window length (2-day, red; 4-day black), (b) ensemble size (11 members, red; 31 members, black), (c) horizontal localization radius (260 km, red; 65 km, black), and (d) vertical localization 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. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
The default parameters used in the background standard deviation Eqs. (5) and (6).

	σ^{max}	σ^{ml}	σ^{do}	δz
Temperature	0.66 °C	0.1 °C	0.04 °C	40 m
Salinity	0.05	0.1	0.056	40 m
Velocity	0.12 ms^{-1}	0.10 ms^{-1}	0.04 ms^{-1}	500 m
SSH	0.05 m	–	–	–

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.

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 $\boldsymbol{HBH}^T$ and $\boldsymbol{BH}^T$ in (2). Recall that in 4D-Var $\boldsymbol{H}$ represents the tangent linear ROMS sampled at the observation times and locations while $\boldsymbol{H}^T$ 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 $\boldsymbol{x}_b(t)$. 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 $\boldsymbol{B}$. The benefit of this appears to out-weigh 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 $\boldsymbol{B}$ 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, $\boldsymbol{B}$ 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 Σ_c were used in conjunction with an 8-day assimilation window. Fig. 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 Σ_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 Σ_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 α except for δz, where α 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

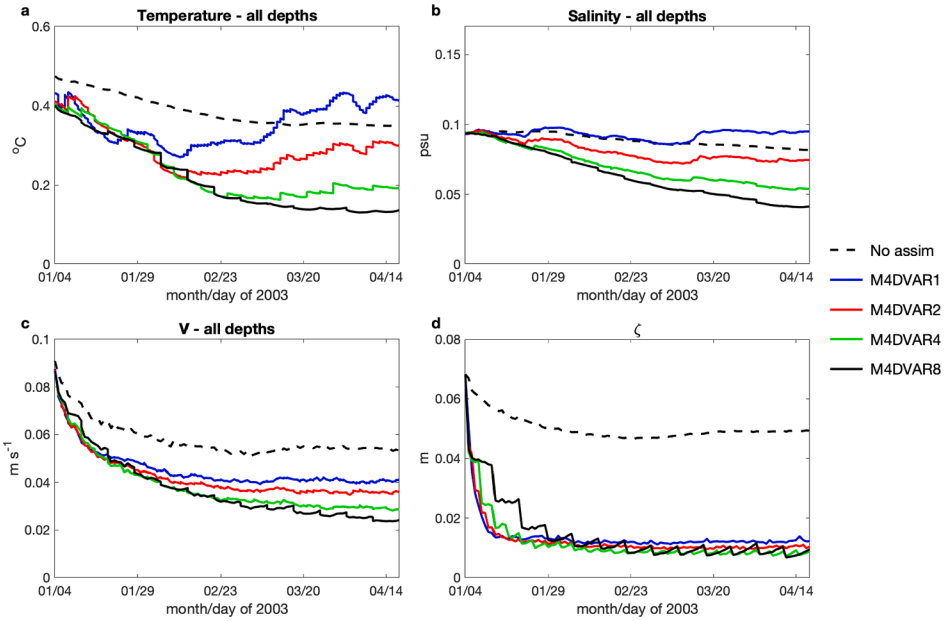

Fig. 8. Time series of the RMS error in (a) temperature, (b) salinity, (c) meridional velocity and (d) SSH for experiments using 1-day (blue line), 2-day (red line), 4-day (green line) and 8-day (black line) data assimilation windows. Also shown is the case where no data were assimilated (black dashed line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

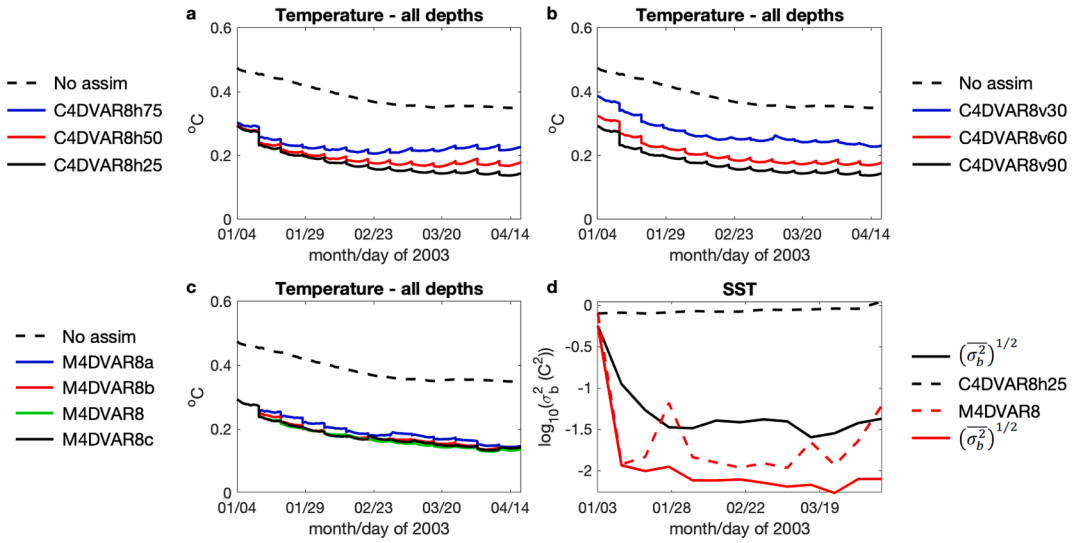

Fig. 9. Time series of RMS errors in temperature for 4DVAR cases involving variations in (a) horizontal correlation length (25 km, black; 50 km, red; 75 km, blue) with vertical correlation of 90 m and using Σ_c, (b) vertical correlation length (90 m, black; 60 m, red; 30 m, blue) with horizontal correlation of 25 km and using Σ_c, and (c) the scaling factor α for the parameters in the standard deviation Eqs. (5) and (6) for Σ_m and using a 25 km horizontal correlation length and a 90 m vertical correlation length ($\alpha = 0.25$, blue; $\alpha = 0.5$, red; $\alpha = 1$, green; $\alpha = 1.5$, black). In all cases, an 8-day assimilation window was used. The case with no data assimilation is also shown for comparison (black dashed). (d) Time series of $\log_{10}\overline{\sigma_b^2}$ (solid black line) and $\log_{10}\left(\mathrm{Tr}\left(\Sigma_c^2\right)/N\right)$ (black dashed line) for SST observations for experiment C4DVAR8h25, and of $\log_{10}\overline{\sigma_b^2}$ (solid red line) and $\log_{10}\left(\mathrm{Tr}\left(\Sigma_m^2\right)/N\right)$ (red dashed line) for experiment M4DVAR8. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

used. In these experiments, the data assimilation methodology was modified, so that during the first 4D-Var cycle starting on 4 Jan, Σ_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, $\overline{\sigma_b^2} = \left(\boldsymbol{y}_a - \boldsymbol{y}_b\right)^T\left(\boldsymbol{y} - \boldsymbol{y}_b\right)/N$ was evaluated where the overbar denotes the expected value, $\boldsymbol{y}$ is the vector of observations, and $\boldsymbol{y}_a$ and $\boldsymbol{y}_b$ 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 $\boldsymbol{B}$ and $\boldsymbol{R}$ are chosen consistently, then $\boldsymbol{HBH}^T \approx \overline{\left(\boldsymbol{y}_a - \boldsymbol{y}_b\right)\left(\boldsymbol{y} - \boldsymbol{y}_b\right)^T}$. Therefore, for a given subset of observations (*e.g.*, temperature), $\overline{\sigma_b^2}$ is the expected value for the background error variance. The statistic $\overline{\sigma_b^2}$ is, therefore, a useful measure of the consistency between the expected background error based on the *a priori* choice of$\boldsymbol{B}$ and $\boldsymbol{R}$, and the specified values of error variance in Σ_c or Σ_m. Fig. 9d shows the time series of $log_{10}\overline{\sigma_b^2}$ and $log_{10}\left(\mathrm{Tr}\left(\Sigma_c^2\right)/N\right)$ 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 Σ_c, showing that the choice of Σ_c is inappropriate. This adjustment occurs because, after the first assimilation cycle, the model SST is much closer to the truth than a randomly chosen ocean state. Fig. 9d shows the situation in experiment M4DVAR8 where Σ_c is used during the first 4D-Var cycle, and Σ_m with the default parameters is used thereafter. In this case, the agreement between $\overline{\sigma_b^2}$ and $\mathrm{Tr}\left(\Sigma_m^2\right)$ is greatly improved, indicating that Σ_m is a more appropriate choice of background errors after the first cycle.

Fig. 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 Fig. 10a–c for cases initialized using $\boldsymbol{x}_0^{\mathrm{i}}$ and $\boldsymbol{x}_0^{\mathrm{ii}}$ 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. Fig. 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 $\boldsymbol{x}_0$ than those in the 4D-Var cases. 4D-Var errors are typically larger for the case starting from $\boldsymbol{x}_0^{\mathrm{ii}}$, although, by the end of Dec 2003, they are similar to those starting from $\boldsymbol{x}_0^{\mathrm{i}}$. Fig. 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 $\boldsymbol{x}_0$. 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.

Fig. 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 $\boldsymbol{x}_0^{\mathrm{ii}}$. Over the upper 100 m (Fig. 11a and b), 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

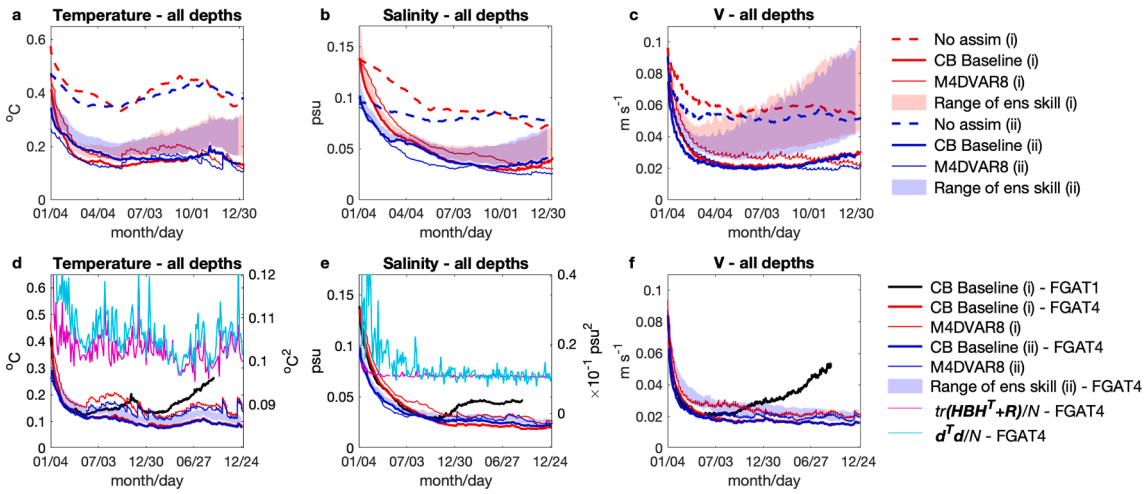

Fig. 10. Time series of RMS errors in ensemble mean (a) temperature, (b) salinity, and (c) meridional velocity for the period Jan-Dec 2003 for the CB baseline EAKF case using x_0^{i} (thick red line) and x_0^{ii} (thick blue line). Also shown are the corresponding M4DVAR8 cases (thin red line and thin blue line, respectively). The shading represents the envelopes bounded by the maximum and minimum RMS error of each of the 51 members of the CB EAKF baseline ensemble for the two cases. Also shown for comparison are the cases initialized from x_0^{i} (dashed red line) and x_0^{ii} (dashed blue line) where no data were assimilated. Time series of RMS errors for the extended period Jan 2003–Dec 2004 are shown in (d), (e) and (f) for the ensemble mean of the CB EAKF baseline case using x_0^{i} (thick black line) and CB_EAKF1b starting from x_0^{i} (thick red line) and x_0^{ii} (thick blue line). The blue shaded region is the envelope bounded by the maximum and minimum RMS error of each ensemble member of the latter case. The corresponding M4DVAR8 cases (thin red line for the x_0^{i} case and thin blue line for x_0^{ii}) are also shown. Also shown in (d) and (e) are time series of mean squared innovation $\boldsymbol{d}^T\boldsymbol{d}/N$ (cyan line) and the expected total error variance of the ensemble per observation $Tr\left\{\overline{(\boldsymbol{y}_b-\overline{\boldsymbol{y}_b})(\boldsymbol{y}_b-\overline{\boldsymbol{y}_b})^T}+\boldsymbol{R}\right\}\Big/N$ (magenta line) for all temperature and salinity observations, where N is the number of observations assimilated in each case. The scale for $\boldsymbol{d}^T\boldsymbol{d}/N$ and $Tr\left\{\overline{(\boldsymbol{y}_b-\overline{\boldsymbol{y}_b})(\boldsymbol{y}_b-\overline{\boldsymbol{y}_b})^T}+\boldsymbol{R}\right\}\Big/N$ is shown on the right. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

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 Fig. 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). Fig. 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 Fig. 10a–c is associated with a significant increase in errors of T and S at mid-water column depths and below, as revealed by Fig. 11e and f. 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 Fig. 12i–k, and clearly illustrate the development of significant salinity errors in several locations during the EAKF case (Fig. 12k).

The error growth in the CB baseline EAKF experiment continues beyond the end of 2003, and eventually, the computation becomes unstable, as illustrated in Fig. 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 ~24–48 h 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 Fig. 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_v 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).

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 Fig. 10d–f for the case using x_0^{ii}. 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

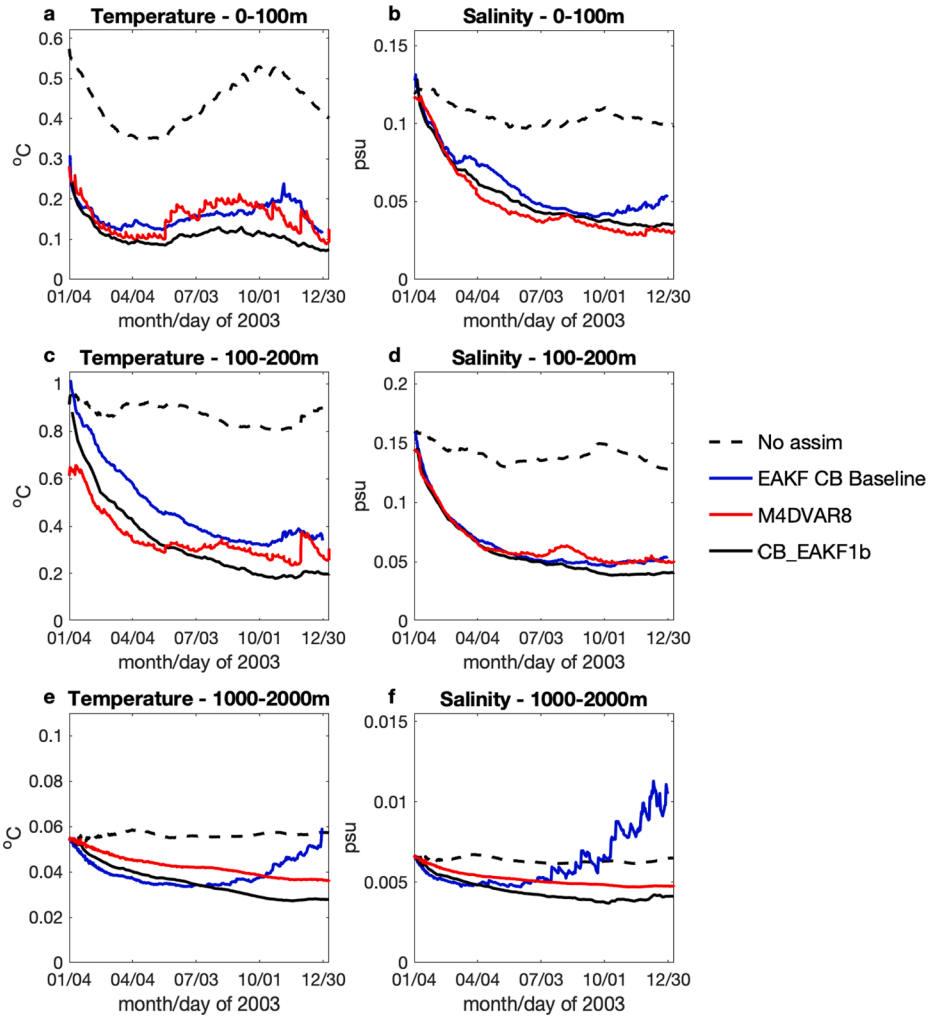

Fig. 11. Time series of RMS errors in the ensemble mean temperature and salinity for the CB baseline EAKF case (blue line), the M4DVAR8 case (red line) and CB_EAKF1b (solid black line) for the period Jan-Dec 2003 computed over different depth ranges: (a, b) 0–100 m, (c,d) 100–200 m and (e,f) 1000–2000 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^{ii}. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the ensemble. As discussed by Houtekamer et al. (2005), an additional measure of the reliability of the ensemble is if $\boldsymbol{d}^T\boldsymbol{d} = Tr\left\{\overline{(\boldsymbol{y}_b - \overline{\boldsymbol{y}_b})(\boldsymbol{y}_b - \overline{\boldsymbol{y}_b})^T} + \boldsymbol{R}\right\}$ where $\boldsymbol{d}$ is the innovation vector, $\boldsymbol{y}_b$ is a vector for each ensemble member evaluated at the observation locations and observation times, and an overbar denotes the ensemble mean. Fig. 10d and 10e shows time series of $\boldsymbol{d}^T\boldsymbol{d}/N$ and $Tr\left\{\overline{(\boldsymbol{y}_b - \overline{\boldsymbol{y}_b})(\boldsymbol{y}_b - \overline{\boldsymbol{y}_b})^T} + \boldsymbol{R}\right\}\Big/N$ 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.

Fig. 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 Fig. 12d, h, and l 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 $(\boldsymbol{HBH}^T + \boldsymbol{R})$. With this in mind, Fig. 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).

The degree to which the pdfs change through time is illustrated in Fig. 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 *in situ* observations. Some of these variations are due to the relatively small sample size (*i.e.*, there are relatively few *in situ* observations available during most 15-day moving windows). However, in the case of *in situ* salinity observations, both the EAKF and 4D-Var innovations are significantly positively

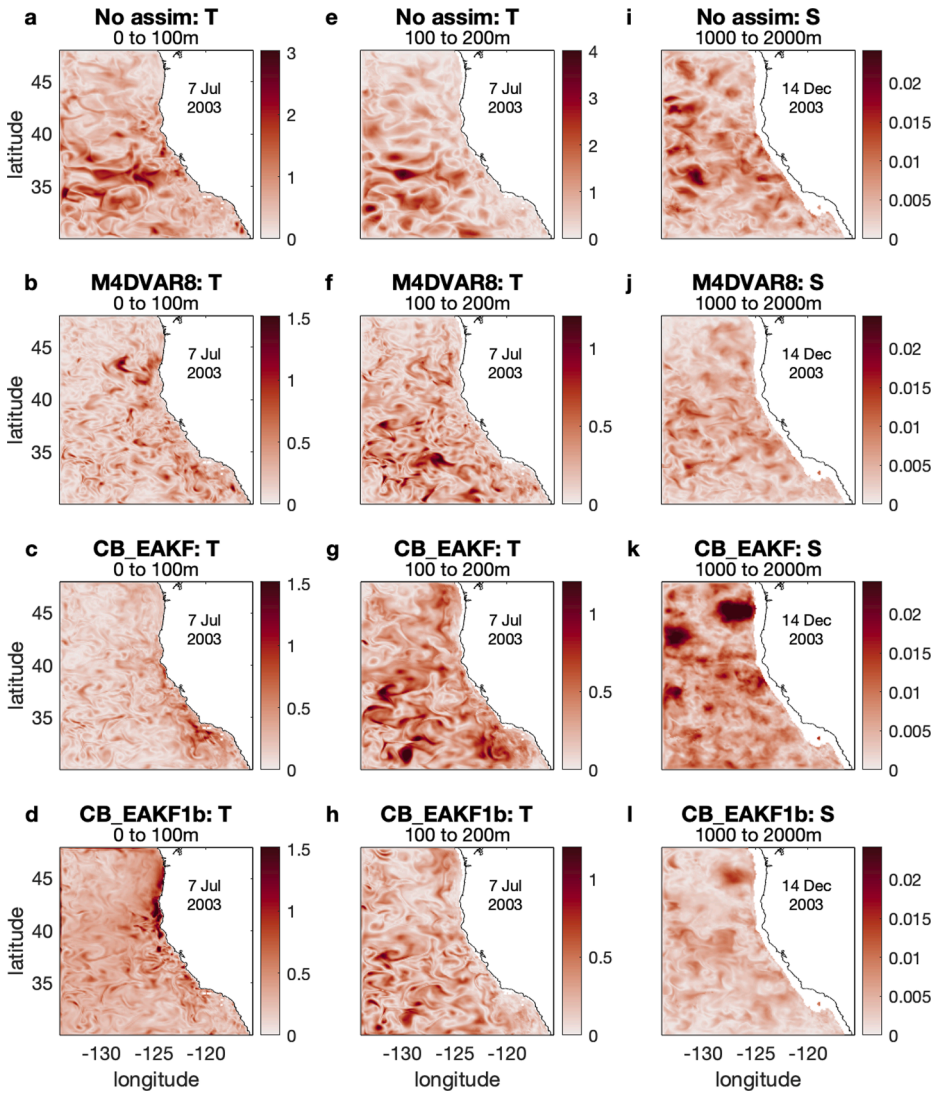

Fig. 12. Maps of the RMS errors averaged over different depth ranges on various dates for cases where no data are assimilated, M4DVAR8, the CB_EAKF baseline, and CB_EAKF1b. (a–d) temperature errors, 0–100 m, 7 July, (e–h) temperature errors, 100–200 m, 7 July, and (i–l) salinity errors, 1000–2000 m,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

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 Fig. 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).

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.*, $r^2 \sim 0.6$–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.

Fig. 15a and b 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

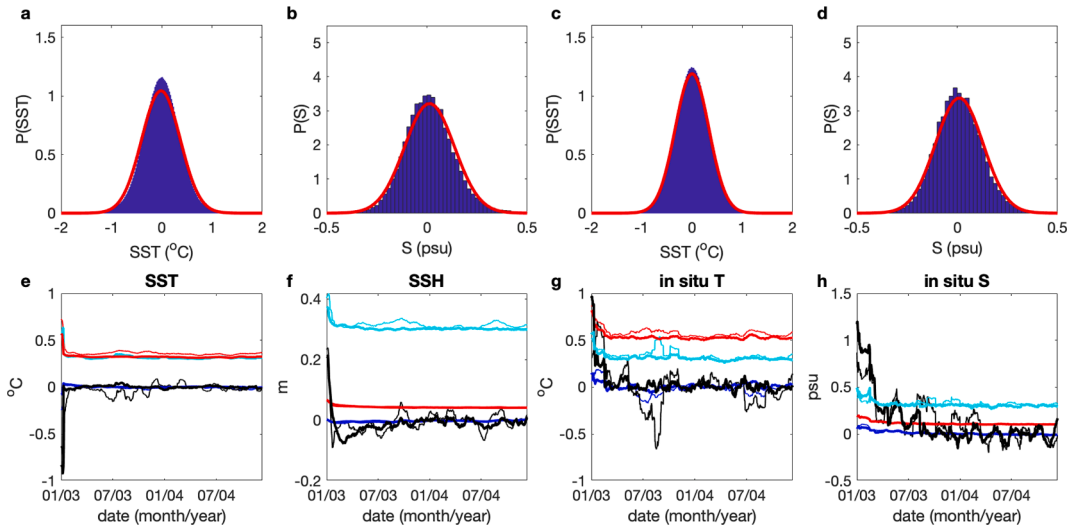

Fig. 13. Probability density functions (pdfs) of the Jan 2003–Dec 2004 innovations in (a) 4D-Var SST, (b) 4D-Var *in situ S*, (c) EAKF SST and (d) EAKF *in situ S*. The red line shows the pdf for a normal distribution with the same mean and standard variation. Time series of the first four moments of the innovation pdfs computed over a 15-day moving window for the EAKF (thick lines) and 4D-Var (thin lines) for (e) 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. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

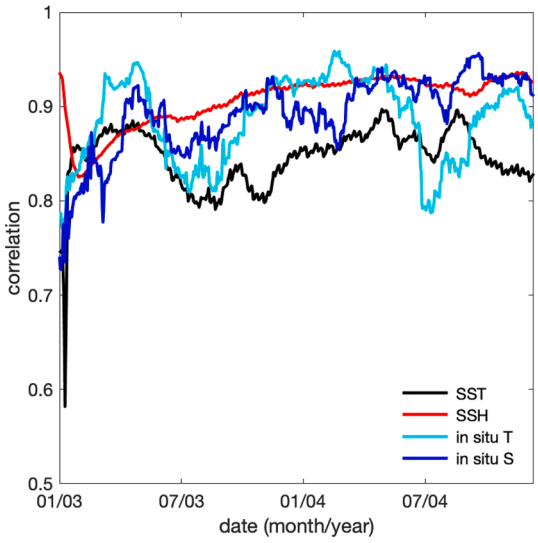

Fig. 14. Time series of the correlation between the EAKF and 4D-Var innovations over a 15-day moving window associated with observations in SST (black line), SSH (red line), *in situ T* (cyan line) and *in situ S* (dark blue line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

both cases, although during summer the EAKF errors are higher. Fig. 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, Fig. 15a and b 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. Fig. 15a and b 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 ($\boldsymbol{\tau}$) 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 $\boldsymbol{\tau}$. The standard deviations assumed for the background errors were chosen to be uniform over the model domain and specified according to $2\sigma_\tau$ for $\boldsymbol{\tau}$, $2\sigma_w$ for Q_w, and $\left(2+8\sin^2(j\pi/12)\right)\sigma_H$ for Q_H, where σ_τ, σ_w and σ_H are the annual, domain-averaged climatological standard deviations of $\boldsymbol{\tau}$, 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. Fig. 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 Fig. 15c and d. 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 Fig. 15c and d are results of a 4D-Var experiment in which *all* open boundary variables were included as part

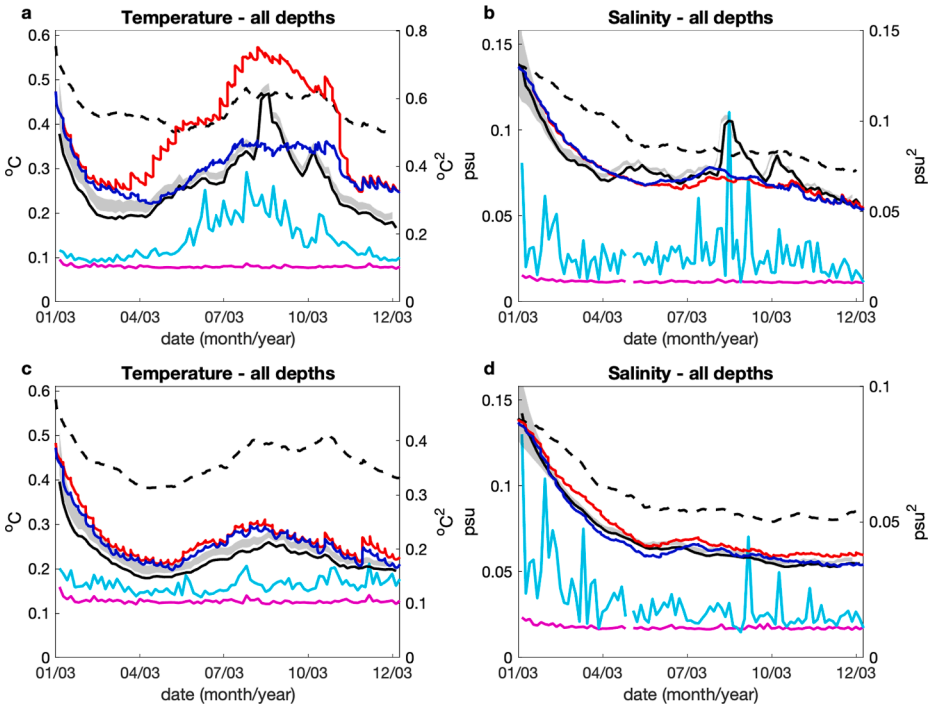

Fig. 15. Time series of the RMS errors in (a) T and (b) S for repeat experiments CB_EAKF1b (black line) and M4DVAR8 (red line) using an incorrect realization of the atmospheric conditions. The gray shaded region represents the envelope bounded by the maximum and minimum RMS errors of each of the 51 members 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 $\boldsymbol{d}^T\boldsymbol{d}/N$ (cyan line) and the expected total error variance of the ensemble $Tr\left\{\overline{(\boldsymbol{y}_b-\overline{\boldsymbol{y}_b})(\boldsymbol{y}_b-\overline{\boldsymbol{y}_b})^T}+\boldsymbol{R}\right\}/N$ (magenta line) for all temperature and salinity observations, where N is the number of observations assimilated in each case. The scale for $\boldsymbol{d}^T\boldsymbol{d}/N$ and $Tr\left\{\overline{(\boldsymbol{y}_b-\overline{\boldsymbol{y}_b})(\boldsymbol{y}_b-\overline{\boldsymbol{y}_b})^T}+\boldsymbol{R}\right\}/N$ is shown on the right. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

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. Fig. 15c and d 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.

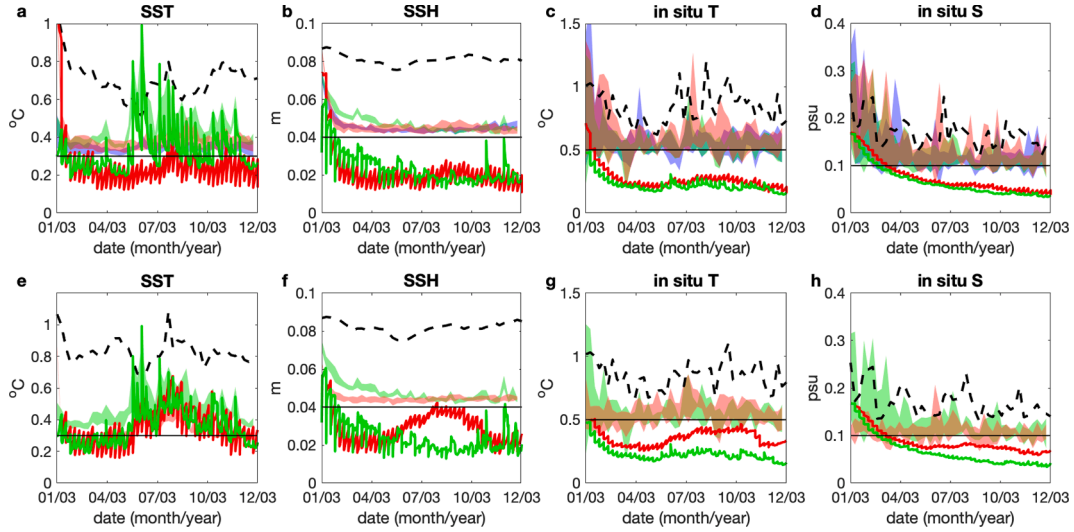

Fig. 16. Time series of RMS errors of hindcasts initialized from the CB EAKF baseline, M4DVAR8, and CB_EAKF1b circulation estimates. The hindcast errors were computed relative to independent observations of (a) SST, (b) SSH, (c) *in situ T,* and (d) *in situ S* before these data were assimilated during 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). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

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.

Fig. 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).

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.

Fig. 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. Fig. 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 Fig. 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 (Fig. 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. Fig. 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

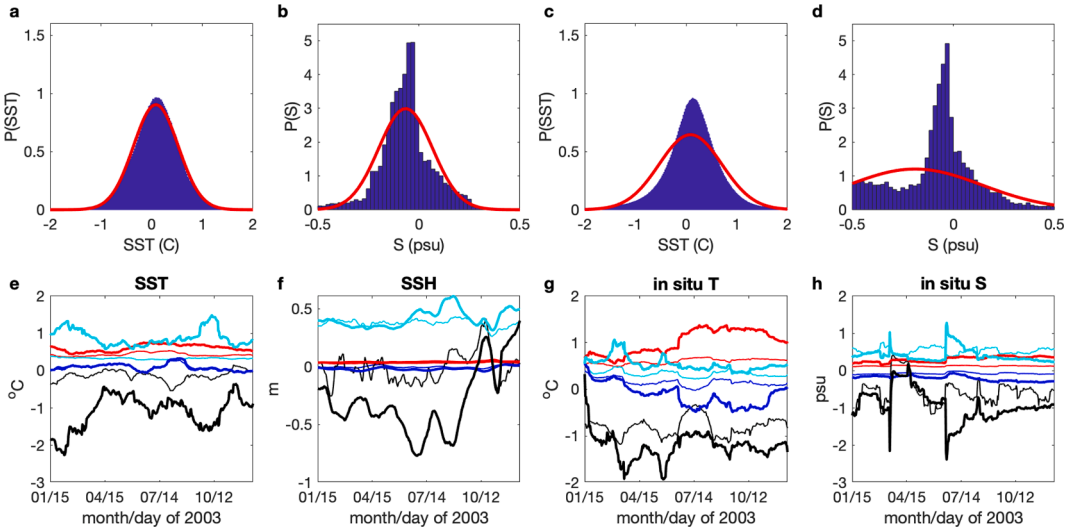

Fig. 17. Probability density functions (pdfs) of the Jan-Dec 2003 innovations when real ocean observations are assimilated: (a) 4D-Var SST, (b) 4D-Var *in situ S*, (c) EAKF SST and (d) EAKF *in situ S*. The red line shows the pdf for a normal distribution with the same mean and standard variation. Time 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) 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. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

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 Fig. 17b and d. 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).

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.

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 Fig. 19e and f 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.

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. Fig. 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 Fig. 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 $\boldsymbol{x}_o^i$ and $\boldsymbol{x}_o^{ii}$ 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.

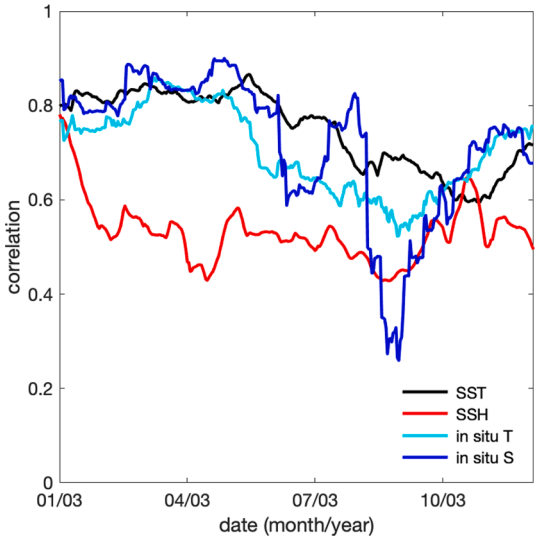

Fig. 18. Time series of the correlation between the EAKF and 4D-Var innovations over a 15-day moving window associated with observations in SST (black line), SSH (red line), *in situ T* (cyan line) and *in situ S* (dark blue line) when real observations are assimilated. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

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 (Dee et al., 2011) at the highest spatial and temporal resolution available (3 h at $^1/_8$ 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 et 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., 2018). The two primary sources of *in situ* temperature and salinity are daily mean profiles from the Research Moored Array for the African-Asian-Australian

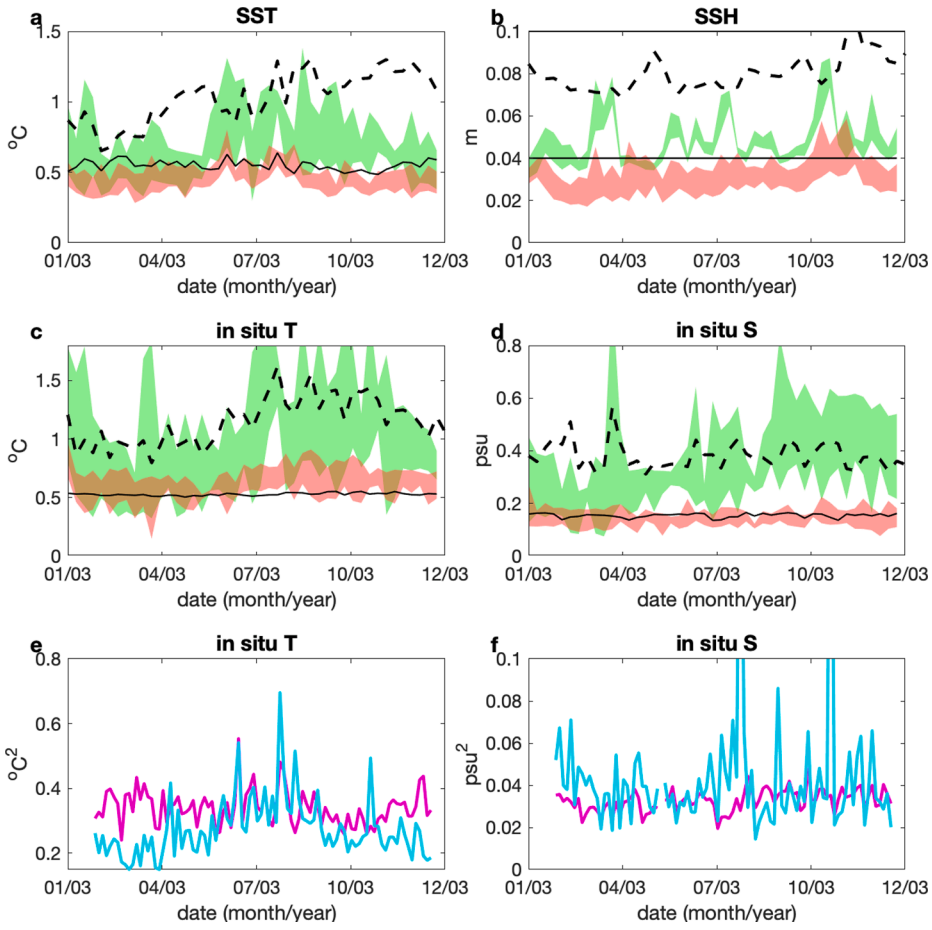

Fig. 19. Time series of RMS errors of hindcasts initialized from the EAKF and 4D-Var circulation estimates derived from real ocean observations in the CCS configuration. The hindcast errors were computed relative to independent observations of (a) SST, (b) SSH, (c) *in situ T*, and (d) *in situ S* before 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 $\boldsymbol{d}^T\boldsymbol{d}/N$ (cyan line) and the expected total error variance of the ensemble per observation $Tr\left\{\overline{(\boldsymbol{y}_b-\overline{\boldsymbol{y}_b})(\boldsymbol{y}_b-\overline{\boldsymbol{y}_b})^T}+\boldsymbol{R}\right\}\Big/N$ (magenta line) for all temperature and salinity observations, where N is the number of observations assimilated in each case. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

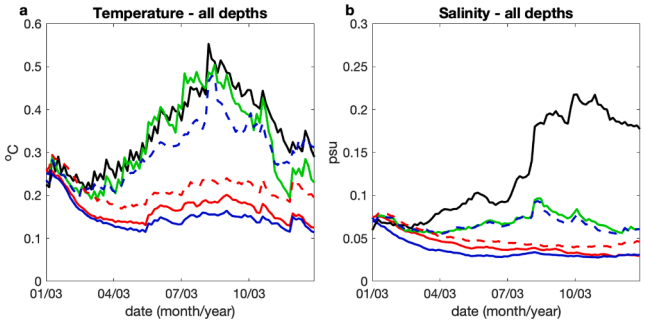

Fig. 20. Time series of the RMS difference between the EAKF and 4D-Var analyses using real observations (black lines): (a) temperature and (b) salinity. Also shown for comparison are the RMS differences between pairs of the EAKF and 4D-Var analyses using simulated observations: CB_EAKF1b ($\boldsymbol{x}_o^i$) and M4DVAR8 ($\boldsymbol{x}_o^i$) – solid blue line; CB_EAKF1b ($\boldsymbol{x}_o^{ii}$) and M4DVAR8 ($\boldsymbol{x}_o^{ii}$) – solid red line; CB_EAKF1b and M4DVAR8 with imperfect forcing – green line; CB_EAKF1b and M4DVAR8 with imperfect forcing adjusted during 4D-Var – dashed blue line; CB_EAKF1b and M4DVAR8 with imperfect open boundary conditions adjusted during 4D-Var – dashed red line. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

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.

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 et 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 $\boldsymbol{x}_0$ was the forward model solution for 1-Sep-2011, and the ensemble perturbations $\pm\delta\boldsymbol{x}$ 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.

Fig. 22b and c 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

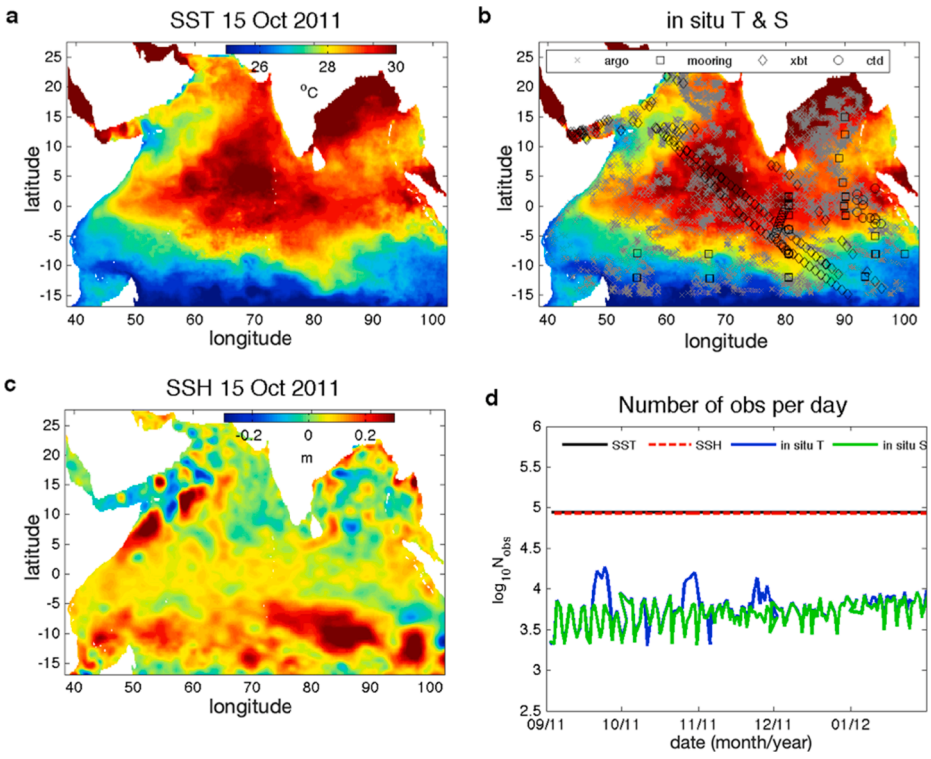

Fig. 21. The ROMS-IO model domain and observations used in the EAKF and 4D-Var experiments. An example of the OSTIA SST and AVISO SSH on 15 October 2011 is shown in (a) and (c) respectively. The locations of all *in situ* hydrographic observations collected between 1-Sep-2011 and 31-Jan-2012 are shown in (b) superimposed on the 15-Oct-2011 SST, and time series of $\log_{10}$ of the number of observations of each type available during each 24-hour window is shown in (d).

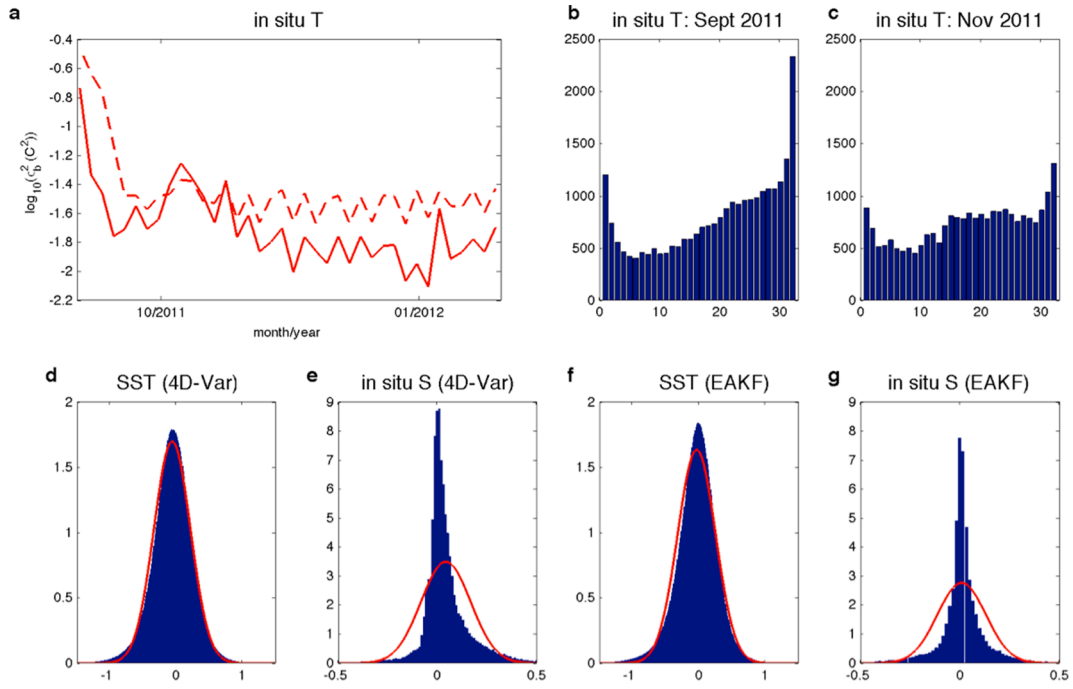

Fig. 22. Diagnostics of 4D-Var and EAKF analyses for ROMS-IO. (a) Time series of $log_{10}\left(\overline{\sigma_b^2}\right)$ (solid red line) and $log_{10}(Tr(\Sigma_c)/N)$ (red dashed line) for *in situ* temperature observations in the 4D-Var analysis. (b) and (c) Rank histograms in the EAKF analysis based on all the *in situ* temperature observations during Sep and Nov (d)-(g) Probability density functions (pdfs) of the Sep 2011–Jan 2012 innovations in (d) 4D-Var SST, (e) 4D-Var *in situ S*, (f) EAKF SST and (g) EAKF *in situ S*. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

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 $\boldsymbol{B}$ is constructed using the seasonal standard deviations of each field (Σ_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 $\boldsymbol{B}$ 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 $\left(\overline{\sigma_b^2}\right)$ and the assumed $(Tr(\Sigma_c)/N)$ 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

Fig. 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).

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

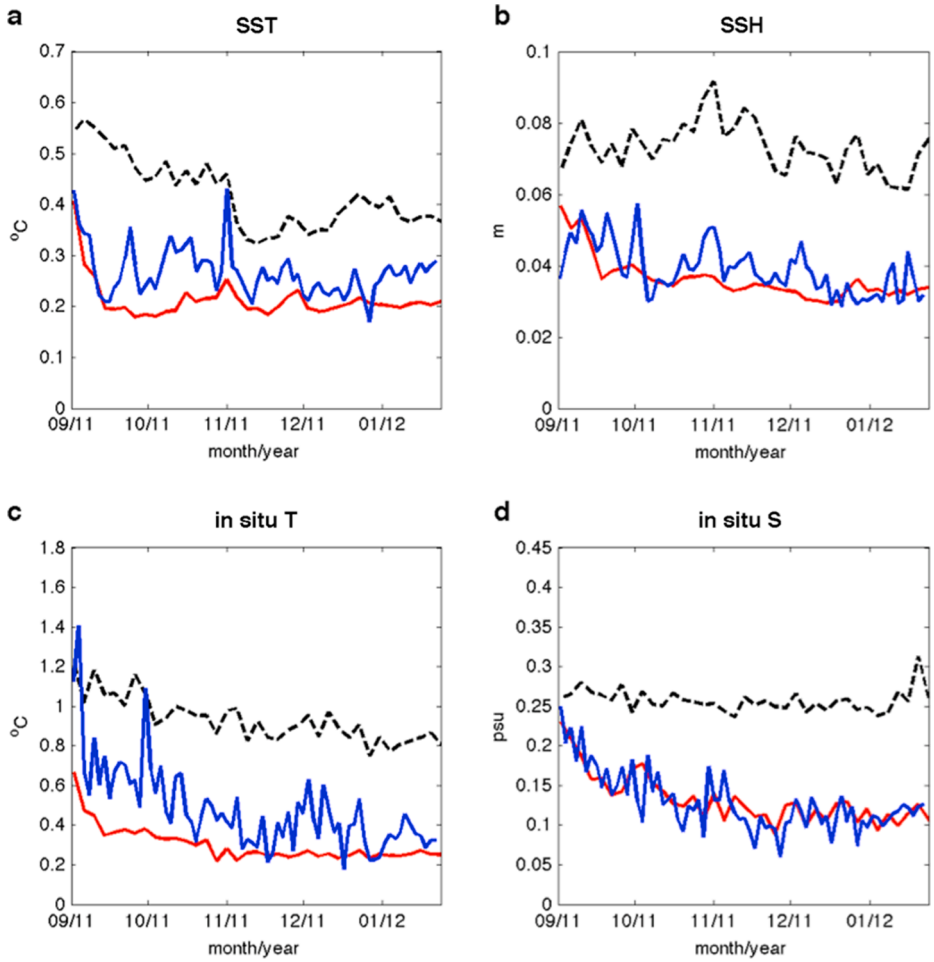

Fig. 23. Time series of RMS errors for the 4D-Var (red lines) and EAKF (blue lines) analyses computed from the assimilated observations. (a) SST, (b) SSH, (c) *in situ* temperature, and (d) *in situ* salinity. Also shown for comparison is the case where no data were assimilated (black dashed line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

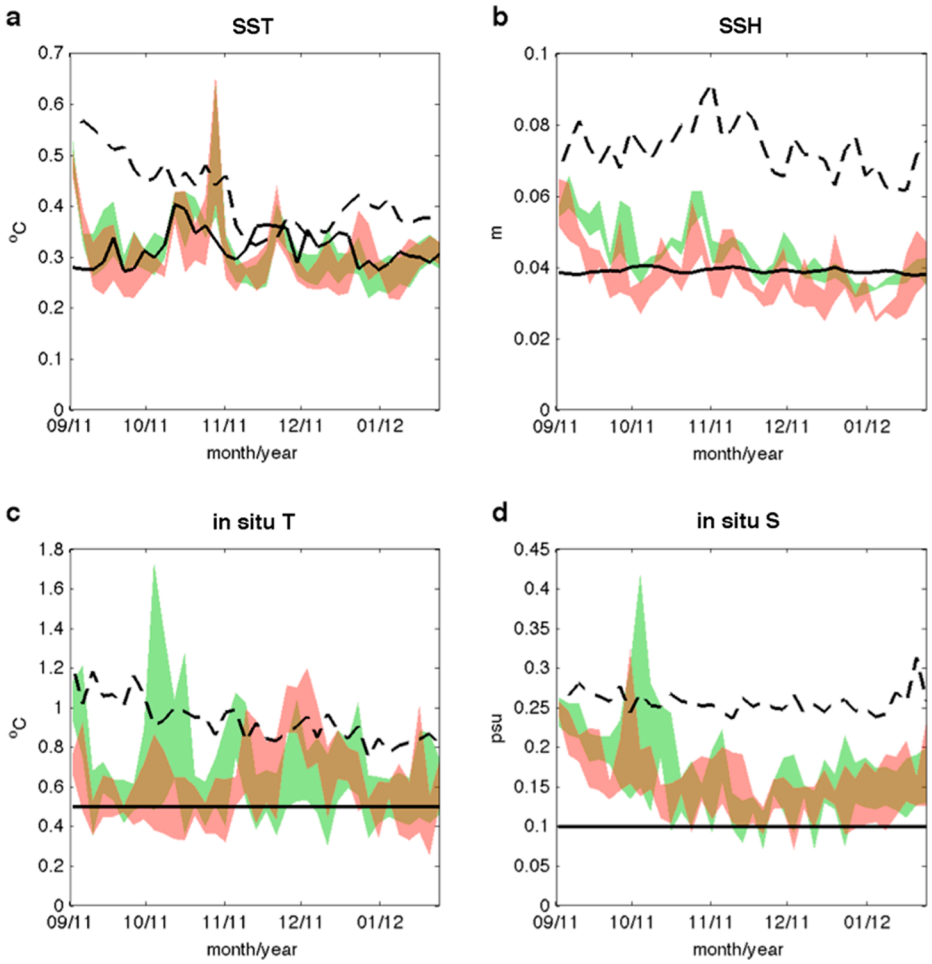

Fig. 24. Time series of RMS errors of hindcasts initialized from the EAKF and 4D-Var circulation estimates in the Indian Ocean configuration. The hindcast errors were computed relative to independent observations of (a) SST, (b) SSH, (c) *in situ T*, and (d) *in situ S* before these data were assimilated 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. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

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.

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 out-performs 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 $\boldsymbol{B}$. 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* Eqs. (5) and (6)). Online documentation is available for DART and ROMS, as well as extensive data assimilation tutorials and hands-on exercises 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 Fig. 10d and e). 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 shorter 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 $\boldsymbol{B}$ (Hamill and Snyder, 2000; Lorenc, 2003a; Buehner, 2005). Hybridization of $\boldsymbol{B}$ in 4D-Var consists of combining the static $\boldsymbol{B}_s$ derived from, say, Σ_c or Σ_m of Section 4.2 with a flow-dependent covariance $\boldsymbol{B}_e$ computed from a localized ensemble. The resulting background error covariance matrix is then given by $\boldsymbol{B} = \alpha\boldsymbol{B}_s + \beta\boldsymbol{B}_e$, where the weights α and β can be computed from statistical considerations (Ménétrier and Auligné, 2015). The system envisaged for ROMS is one in which DART will provide a localized $\boldsymbol{B}_e$. The hybrid $\boldsymbol{B}$ will be used in 4D-Var to compute a new analysis which will then be used to re-center the DART ensemble mean, a approach that has been used successfully in numerical weather prediction (*e.g.* Zhang and Zhang, 2012; Clayton et al., 2013).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

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

Anderson, J.L., 1996. A method for producing and evaluating probabilistic forecasts from ensemble model integrations. J. Climate 9, 1518–1530.
Anderson, J.L., 2001. An ensemble adjustment Kalman filter for data assimilation. Mon. Weather Rev. 129, 2884–2903.
Anderson, J.L., 2009. Spatially and temporally varying adaptive covariance inflation for ensemble filters. Tellus A: Dyn. Meteorol. Oceanography 61 (1), 72–83. https://doi.org/10.1111/j.1600-0870.2007.00361.x.
Anderson, J.L., 2012. Localization and sampling error correction in ensemble Kalman filter data assimilation. Mon. Wea. Rev. 140, 2359–2371.
Anderson, J.L., Hoar, T., Raeder, K., Liu, H., Collins, N., Torn, R., Avellano, A., 2009. The data assimilation research testbed: a community facility. Bull. Amer. Met. Soc. https://doi.org/10.1175/2009BAMS2618.1.
Andersson, E., Järvinen, H., 1999. Variational quality control. Q. J. R. Meteorol. Soc. 125, 679–722.
Balmaseda, M.A., Dee, D.P., Vidard, A., Anderson, D.L.T., 2007. A multivariate treatment of bias for sequential data assimilation: application to the tropical oceans. Q. J. R. Meteorol. Soc. 133, 167–179.
Banerjee, S., Carlin, B., Gelfand, A., 2004. Hierarchical modeling and analysis for spatial data. Monographs on Statistics and Applied Probability 101. Chapman and Hall/CRC Press, Boca Raton, FL, p. 451.
Bennett, A.F., 1985. Array design by inverse methods. Prog. Oceanogr. 15, 129–156.
Bennett, A.F., 2002. Inverse Modeling of the Ocean and Atmosphere. Cambridge University Press, p. 234.
Bertsekas, D.P., 1982. Constrained Optimization and Lagrange Multiplier Methods. Academic Press, Computer Science and Mathematics, p. 395.
Boyer, T.P., Baranova, O.K., Coleman, C., Garcia, H.E., Grodsky, A., Locarnini, R.A., Mishonov, A.V., Paver, C.R., Reagan, J.R., Seidov, D., Smolyar, I.V., Weathers, K., Zweng, M.M., 2019, World Ocean Database 2018. A.V. Mishonov, Technical Editor, NOAA Atlas NESDIS 87.
Broquet, G., Edwards, C.A., Moore, A.M., Powell, B.S., Veneziani, M., Doyle, J.D., 2009a. Application of 4D-variational data assimilation to the California Current System. Dyn. Atmos. Oceans 48, 69–92.
Broquet, G., Moore, A.M., Arango, H.G., Edwards, C.A., Powell, B.S., 2009b. Ocean state and surface forcing correction using the ROMS-IS4DVAR data assimilation system. Mercator Ocean Quarterly Newsletter #34, 5–13.
Broquet, G., Moore, A.M., Arango, H.G., Edwards, C.A., 2011. Corrections to ocean surface forcing in the California Current system using 4D-variational data assimilation. Ocean Model. 36, 116–132.

Buehner, M., 2005. Ensemble-derived stationary and flow-dependent background-error covariances: evaluation in a quasi-operational NWP setting. Q. J. R. Meteorol. Soc. 131, 1013–1043.
Carrassi, A., Bocquet, M., Bertino, L., Evensen, G., 2018. Data assimilation in the geosciences: an overview of methods, issues, and perspectives. Wiley Interdiscip. Rev. Clim. Change 9, e535. https://doi.org/10.1002/wcc.535.
Carton, J.A., Giese, B.S., 2008. A reanalysis of ocean climate using simple ocean data assimilation (SODA). Mon. Wea. Rev. 136, 2999–3017.
Chapman, D.C., 1985. Numerical treatment of cross-shelf open boundaries in a barotropic coastal ocean model. J. Phys. Oceanogr. 15, 1060–1075.
Chatterjee, A., Shankar, D., Shenoi, S.S.C., Reddy, G.V., Michael, G.S., Ravichandran, M., Gopalkrishna, V.V., Rao, E.P.R., Bhaskar, T.V.S.U., Sanjeevan, V.N., 2012. A new atlas of temperature and salinity for the North Indian Ocean. J. Earth Syst. Sci. 121 (3), 559–593.
Checkley, D.M., Barth, J.A., 2009. Patterns and processes in the California Current system. Prog. Oceanogr. 83, 49–64.
Chen, S., Flatau, M., Jensen, T.G., Shinoda, T., Schmidt, J., May, P., Cummings, J., Liu, M., Ciesielski, P.E., Fairall, C.W., Lien, R.-C., Baranowski, D.B., Chi, N.-H., de Szoeke, S., Edson, J., 2015. A study of CINDY/DYNAMO MJO suppressed phase. J. Atmos. Sci. 72, 3755–3779.
Clayton, A.M., Lorenc, A.C., Barker, D.M., 2013. Operational implementation of a hybrid ensemble/4D-Var global data assimilatiom system at the Met Office. Q. J. R. Meteorol. Soc. 139, 1445–1461.
Courtier, P., Thépaut, J.-N., Hollingsworth, A., 1994. A strategy for operational implemenation of 4D-Var using an incremental approach. Q. J. R. Meteorol. Soc. 120, 1367–1388.
Courtier, P., 1997. Dual formulation of four-dimensional variational assimilation. Q. J. R. Meteorol. Soc. 123, 2449–2461.
Dai, A., Trenberth, K.E., 2002. Estimates of freshwater discharge from continents: latitudinal and seasonal variations. J. Hydrometeorol. 3, 660–687.
Daley, R., 1991. Atmospheric Data Analysis. Cambridge University Press, p. 457.
Dee, D.P., 2005. Bias and data assimilation. Q. J. R. Meteorol. Soc. 131, 3323–3343.
Dee, D.P., Uppala, A.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andare, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragnani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Holm, E.V., Isaksen, L., Kallberh, P., Kohler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thepaut, J.-N., Vitart, F., 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q.J.R. Meteorol. Soc. 137, 553–597.
Derber, J., Rosati, A., 1989. A global oceanic data assimilation system. J. Phys. Oceanogr. 19, 1333–1347.
Desroziers, G., Berre, L., Chapnik, B., Poli, P., 2005. Diagnosis of observation, background and analysis-error statistics in observation space. Q. J. R. Meteorol. Soc. 131, 3385–3396.
Dibarboure, G., Pujol, M.-I., Briol, F., Le Traon, P.Y., Larnicol, G., Picot, N., Mertz, F., Ablain, M., 2011. Jason-2 in DUACS: Updated system description, first tandem results and impact on processing and products. Mar. Geod. 34 (3–4), 214–241.
Donlon, C.J., Martin, M., Stark, J.D., Roberts-Jones, J., Fiedler, E., Wimmer, W., 2011. The operational sea surface temperature and sea ice analysis (OSTIA). Remote Sens. Environ. 116, 140–158.
Doyle, J.D., Jiang, Q., Chao, Y., Farrara, J., 2009. High-resolution atmospheric modeling over the Monterey Bay during AOSN II. Deep Sea Res. II 56, 87–99.
Edwards, C.A., Moore, A.M., Hoteit, I., Cornuelle, B.D., 2015. Regional ocean data assimilation. Ann. Rev. Mar. Sci. 7, 6.1–6.22. https://doi.org/10.1146/annurevmarine-010814-015821.
Fairall, C., Bradley, E., Hare, J., Grachev, A., Ebson, J., Young, G., 2003. Bulk parameterization of airsea fluxes: updates and verification for the COARE algorithm. J. Climate 16, 571–591.
Flather, R.A., 1976. A tidal model of the northwest European continental shelf. Mem. Soc. R. Sci. Liege 6 (10), 141–164.
Gaspari, G., Cohn, S.E., 1999. Construction of correlation functions in two and three dimensions. Q. J. Roy. Meteorol. Soc. 125, 723–758.
Gürol, S., Weaver, A.T., Moore, A.M., Piacentini, A., Arango, H.G., Gratton, S., 2014. B-preconditioned minimization algorithms for variational data assimilation with the dual formulation. Q. J. Roy. Meteorol. Soc. 140, 539–556.
Hamill, T.M., Snyder, C., 2000. A hybrid ensemble Kalman filter – 3D variational analysis scheme. Mon. Weather Rev. 128, 2905–2919.
Hickey, B.M., 1979. The California current system: hypotheses and facts. Prog. Oceanogr. 8, 191–279.
Hickey, B.M., 1998. Coastal oceanography of western North America from the tip of Baja, California to Vancouver Island. The Sea 11, 345–393. Eds. A.R. Robinson and K.H. Brink, John Wiley and Sons.
Houtekamer, P.L., Michell, H.L., Pellerin, G., Buehner, M., Charron, M., Spacek, L., Hansen, B., 2005. Atmospheric data assimilation with an ensemble Kalman filter: results with real observations. Mon. Weather Rev. 133, 604–620.
Houtekamer, P.L., Zhang, F., 2016. Review of the ensemble Kalman filter for atmospheric data assimilation. Mon. Weather Rev. 144, 4489–4531.
Ingleby, B., Huddleston, M., 2007. Quality control of ocean temperature and salinity profiles – historical and real-time data. J. Mar. Syst. 65, 158–175.
Jensen, T.G., Shinoda, T., Chen, S., Flatau, M., 2015. Ocean response to CINDY/DYNAMO MJOs in air-sea-coupled COAMPS. J. Meteorol. Soc. Jpn. Ser. II 93A, 157–178.
Kara, A.B., Rochford, P.A., Hurlburt, H.E., 2000. An optimal definition for ocean mixed layer depth. J. Geophys. Res. 105, 16803–16821.
Kelly, K.A., Beardsley, R.C., Limeburner, R., Brink, K.H., Paduan, J.D., Chereskin, T.K., 1998. Variability of the near-surface eddy kinetic energy in California Current based on altimetric, drifter, and moored current data. J. Geophys. Res. 103, 13067–13083.
Li, Y., Tuomi, R., 2017. A balanced Kalman filter ocean data assimilation system with application to the South Australian Sea. Ocean Model. 116, 159–172.
Lorenc, A.C., 1986. Analysis methods for numerical weather prediction. Q. J. R Meteorol. Soc. 112, 1177–1194.
Lorenc, A.C., 2003a. Modelling of error covariances by 4D-Var data assimilation. Q. J. R. Meteorol. Soc. 129, 3167–3182.
Lorenc, A.C., 2003b. The potential of the ensemble Kalman filter for NWP—a comparison with 4D-Var. Q. J. R. Meteorol. Soc. 129, 3183–3203.
Lorenc, A.C., Bowler, N., Clayton, A.M., Pring, S.R., Fairbairn, D., 2015. Comparison of hybrid-4DEnVar and hybrid-4DVar data assimilation methods for global NWP. Mon. Weather Rev. 143, 212–229.
Mason, E., Molemaker, J., Shchepetkin, A.F., Colas, F., McWilliams, J.C., Sangrà, P., 2010. Procedures for offline gerid nesting in regional ocean models. Ocean Model. 35, 1–15.
Marchesiello, P., McWilliams, J.C., Shchepetkin, A., 2001. Open boundary conditions for long-term integration of regional oceanic models. Ocean Model. 3, 1–20.
Martin, M.J., Balmaseda, M., Bertino, L., Brasseur, P., Brassington, G., Cummings, J., et al., 2015. Status and future of data assimilation in operational oceanography. J. Operat. Oceanogr. 8 (Supp. 1), s28–s48. https://doi.org/10.1080/1755876X.2015.1022055.
Mattern, J.P., Song, H., Edwards, C.A., Moore, A.M., Fiechter, J., 2017. Data assimilation of physical and chlorophyll observations in the California Current System using two biogeochemical models. Ocean Model. 109, 55–71.
Matthews, D., Powell, B., Milliff, R., 2011. Dominant spatial variability scales from observations around the Hawaiian islands. Deep Sea Res. Part I 58 (10), 979–987.
Ménétrier, B., Auligné, T., 2015. Optimized localization and hybridization to filter ensemble-based covariances. Mon. Wea. Rev. 143, 3931–3947.
Metzger, E.J., Smedstad, O.M., Thoppil, P.G., Hurlburt, H.E., Cumming, J.A., Wallcraft, A.J., Zamudio, L., Franklin, D.S., Posey, P.G., Phelps, M.W., Hogan, P.J., Bub, F.L., DeHaan, C.J., 2014. US navy operational global ocean and arctic ice prediction systems. Oceanography 27, 32–43.
Mogensen, K., Balmaseda, M.A., Weaver, A.T., 2012. The NEMOVAR ocean data assimilation system as implemented in the ECMWF ocean analysis for system 4. ECMWF Tech. Memorandum 668, 59.
Moore, A.M., Arango, H.G., Broquet, G., Powell, B.S., Zavala-Garay, J., Weaver, A.T., 2011a. The regional ocean modeling system (ROMS) 4-dimensional variational data assimilation systems. Part I: System overview and formulation. Prog. Oceanography 91, 34–49.
Moore, A.M., Arango, H.G., Broquet, G., Edwards, C., Veneziani, M., Powell, B., Foley, D., Doyle, J., Costa, D., Robinson, P., 2011b. The Regional Ocean Modeling System (ROMS) 4-dimensional variational data assimilation systems. Part II: Performance and application to the California Current System. Prog. Oceanogr. 91, 50–73.
Moore, A.M., Edwards, C., Fiechter, J., Drake, P., Arango, H.G., Neveu, E., Gürol, S., Weaver, A.T., 2013. A 4D-Var analysis system for the California current: a prototype for an operational regional ocean data assimilation system. In: Data Assimilation for Atmospheric, Oceanic and Hydrological Applications, Vol. II, Liang Xu and Seon Park, Eds. Springer, Chapter 14, pp. 345–366.
Moore, A.M., Martin, M.J., et al., 2019. Synthesis of ocean observations using data assimilation for operational, real-time and reanalysis systems: a more complete picture of the state of the ocean. Front. Mar. Sci. 6 https://doi.org/10.3389/fmars.2019.00090.
Neveu, E., Moore, A.M., Edwards, C.A., Fiechter, J., Drake, P., Jacox, M.G., Nuss, E., 2016. An historical analysis of the California Current using ROMS 4D-Var. Part I: System configuration and diagnostics. Ocean Model. 99, 131–151.
Pires, C., Vautaud, R., Talagrand, O., 1996. On extending the limits of variational assimilation in non-linear chaotic systems. Tellus 48A, 96–121.
Raghukumar, K.C.A., Edwards, N.L., Goebel, G., Broquet, M., Veneziani, A.M. Moore, Zehr, J.P., 2015. Impact of assimilating physical oceanographic data on modeled ecosystem dynamics in the California Current System. Prog. Oceanoraphy 138, 546–558.
Schott, F.A., McCreary, J.P., 2001. The monsoon circulation of the Indian Ocean. Prog. Oceanogr. 51, 1–123.
Song, H., Edwards, C.A., Moore, A.M., Fiechter, J., 2016. Data assimilation in a coupled physical-biogeochemical model of the California Current System using an incremental lognormal 4-dimensional variational approach: Part 1, Model formulation and biological data assimilation twin experiments. Ocean Model. 106, 131–145.
Swanson, K., Vautard, R., Pires, C., 1998. Four-dimensional variational assimilation and predictability in a quasi-geostrophic model. Tellus 50A, 369–390.
Umlauf, L., Burchard, H., 2003. A generic length-scale equation for geophysical turbulence models. J. Marine Res. 61, 235–265.
van Leeuwen, P.-J., Cheng, Y., Reich, S., 2015. Nonlinear data assimilation. In: Frontiers in Applied Dynamical Systems: Reviews and Tutorials 2. Springer, p. 117.
Veneziani, M., Edwards, C.A., Doyle, J.D., Foley, D., 2009a. A central California coastal ocean modeling study: 1. Forward model and the influence of realistic versus climatological forcing. J. Geophys. Res. 114, C04015. https://doi.org/10.1029/2008JC004774.
Veneziani, M., Edwards, C.A., Moore, A.M., 2009b. A central California coastal ocean modeling study. Part II: Adjoint sensitivities to local and remote forcing mechanisms. J. Geophys. Res. 114, C04020. https://doi.org/10.1029/2008JC004775.

Weaver, A.T., Courtier, P., 2001. Correlation modelling on the sphere using a generalized diffusion equation. Q. J. R. Meteorol. Soc. 127, 1815–1846.
Weaver, A.T., Deltel, C., Machu, E., Ricci, S., Daget, N., 2005. A multivariate balance operator for variational ocean data assimilation. Q. J. R. Meteorol. Soc. 131, 3605–3625.
Zhang, M., Zhang, F., 2012. E4DVar: Coupling an ensemble Kalman filter with four-dimensional variational data assimilation in a limited-area weather prediction model. Mon. Weather Rev. 140, 587–600.

REPORT DOCUMENTATION PAGE

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED	
		START DATE	END DATE
15-04-2022	Final	09-01-2019	01-31-2022

4. TITLE AND SUBTITLE

Further developments of the coupled COAMPS-ROMS modeling system

5a. CONTRACT NUMBER	5b. GRANT NUMBER	5c. PROGRAM ELEMENT NUMBER
	N00014-19-1-2675	
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**

6. AUTHOR(S)

Moore, Andrew M.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of California Santa Cruz
1156 High Street
Santa Cruz CA 95064-1077
U.S.A.

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Research
875 N. Randolph Street Suite 1425
Arlington VA 22203-1995

10. SPONSOR/MONITOR'S ACRONYM(S)

ONR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for Public Release; Distribution is Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The report describes how the project goals were reached. Three key elements are described: (i) the coupled COAMPS-ROMS system that was developed; (ii) the DART-ROMS ensemble Kalman filter interface for ocean data assimilation; (iii) scientific findings of the project. Reference to (i) and (ii) is made primarily via web sites where the codes and documentation are maintained. Three appendices document the scientific findings. Appendix A explores the ability of the COAMPS-ROMS coupled system to capture the observed relationships between atmospheric and oceanic variables at the air-sea interface. This is quantified in a coupled model configured for the U.S. west coast which was the primary initial testbed for the project. Appendix B is concerned with air-sea coupling in the Indian Ocean, and the influence of ocean circulation estimates in the region on tropical storms. The performance of the DART-ROMS ensemble Kalman filter is described in appendix C and compared to the existing 4-dimensional variational data system for ROMS. The results of series of Observing System Experiments (OSEs) and Observing System Simulation Experiments (OSSEs) are presented.

15. SUBJECT TERMS

Data assimilation; coupled modeling; Indian Ocean; California Current system.

16. SECURITY CLASSIFICATION OF:

a. REPORT	b. ABSTRACT	C. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
Unclassified	Unclassified	Unclassified	UU	

19a. NAME OF RESPONSIBLE PERSON	19b. PHONE NUMBER *(Include area code)*
Andrew M. Moore	831-459-4332

INSTRUCTIONS FOR COMPLETING SF 298

1. REPORT DATE.
Full publication date, including day, month, if available. Must cite at least the year and be Year 2000 compliant, e.g. 30-06-1998; xx-06-1998; xx-xx-1998.

2. REPORT TYPE.
State the type of report, such as final, technical, interim, memorandum, master's thesis, progress, quarterly, research, special, group study, etc.

3. DATES COVERED.
Indicate the time during which the work was performed and the report was written.

4. TITLE.
Enter title and subtitle with volume number and part number, if applicable. On classified documents, enter the title classification in parentheses.

5a. CONTRACT NUMBER.
Enter all contract numbers as they appear in the report, e.g. F33615-86-C-5169.

5b. GRANT NUMBER.
Enter all grant numbers as they appear in the report, e.g. AFOSR-82-1234.

5c. PROGRAM ELEMENT NUMBER.
Enter all program element numbers as they appear in the report, e.g. 61101A.

5d. PROJECT NUMBER.
Enter all project numbers as they appear in the report, e.g. 1F665702D1257; ILIR.

5e. TASK NUMBER. Enter all task numbers as they appear in the report, e.g. 05; RF0330201; T4112.

5f. WORK UNIT NUMBER.
Enter all work unit numbers as they appear in the report, e.g. 001; AFAPL30480105.

6. AUTHOR(S). Enter name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. The form of entry is the last name, first name, middle initial, and additional qualifiers separated by commas, e.g. Smith, Richard, J, Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES). Self-explanatory.

8. PERFORMING ORGANIZATION REPORT NUMBER.
Enter all unique alphanumeric report numbers assigned by the performing organization, e.g. BRL-1234; AFWL-TR-85-4017-Vol-21-PT-2.

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES). Enter the name and address of the organization(s) financially responsible for and monitoring the work.

10. SPONSOR/MONITOR'S ACRONYM(S). Enter, if available, e.g. BRL, ARDEC, NADC.

11. SPONSOR/MONITOR'S REPORT NUMBER(S). Enter report number as assigned by the sponsoring/monitoring agency, if available, e.g. BRL-TR-829; -215.

12. DISTRIBUTION/AVAILABILITY STATEMENT. Use agency-mandated availability statements to indicate the public availability or distribution limitations of the report. If additional limitations/ restrictions or special markings are indicated, follow agency authorization procedures, e.g. RD/ FRD, PROPIN, ITAR, etc. Include copyright information.

13. SUPPLEMENTARY NOTES. Enter information not included elsewhere such as: prepared in cooperation with; translation of; report supersedes; old edition number, etc.

14. ABSTRACT. A brief (approximately 200 words) factual summary of the most significant information.

15. SUBJECT TERMS. Key words or phrases identifying major concepts in the report.

16. SECURITY CLASSIFICATION. Enter security classification in accordance with security classification regulations, e.g. U, C, S, etc. If this form contains classified information, stamp classification level on the top and bottom of this page.

17. LIMITATION OF ABSTRACT. This block must be completed to assign a distribution limitation to the abstract. Enter UU (Unclassified Unlimited) or SAR (Same as Report). An entry in this block is necessary if the abstract is to be limited.