An Ensemble Adjustment Kalman Filter for the CCSM4 Ocean Component

ALICIA R. KARSPECK, STEVE YEAGER, GOKHAN DANABASOGLU, TIM HOAR, NANCY COLLINS, KEVIN RAEDER, JEFFREY ANDERSON, AND JOSEPH TRIBBIA
National Center for Atmospheric Research,* Boulder, Colorado

(Manuscript received 2 July 2012, in final form 10 April 2013)

ABSTRACT

The authors report on the implementation and evaluation of a 48-member ensemble adjustment Kalman filter (EAKF) for the ocean component of the Community Climate System Model, version 4 (CCSM4). The ocean assimilation system described was developed to support the eventual generation of historical ocean-state estimates and ocean-initialized climate predictions with the CCSM4 and its next generation, the Community Earth System Model (CESM). In this initial configuration of the system, daily subsurface temperature and salinity data from the 2009 World Ocean Database are assimilated into the ocean model from 1 January 1998 to 31 December 2005. Each ensemble member of the ocean is forced by a member of an independently generated CCSM4 atmospheric EAKF analysis, making this a loosely coupled framework. Over most of the globe, the time-mean temperature and salinity fields are improved relative to an identically forced ocean model simulation without assimilation. This improvement is especially notable in strong frontal regions such as the western and eastern boundary currents. The assimilation system is most effective in the upper 1000 m of the ocean, where the vast majority of in situ observations are located. Because of the shortness of this experiment, ocean variability is not discussed. Challenges that arise from using an ocean model with strong regional biases, coarse resolution, and low internal variability to assimilate real observations are discussed, and areas of ongoing improvement for the assimilation system are outlined.

1. Introduction

In this paper, we document the development of an ensemble adjustment Kalman filter (EAKF) (Anderson et al. 2009) data assimilation system for the ocean component of the Community Climate System Model, version 4 (CCSM4). The ocean assimilation system described here was developed to support the eventual generation of historical ocean-state estimates and ocean-initialized climate predictions with the CCSM4 and its next generation, the Community Earth System Model (CESM).1

This development is part of a broader initiative at the National Center for Atmospheric Research (NCAR) to build assimilation capabilities for the atmosphere, land, sea ice, and ocean components of the community model.

There is currently an array of global ocean assimilation products available to the climate-science community that employ various ocean general circulation models and assimilation algorithms. The assimilation methods used to construct these products are all least squares methods that attempt to minimize the difference between an ocean model solution and a set of observations. Broadly speaking, they are distinguishable from one another by the constraints levied upon the solution, the way that prior knowledge of the state of the system is formulated, and whether observations can influence state estimates in the past. For example, adjoint [four-dimensional variational data assimilation (4DVAR)] methods, for example, the Estimating the Circulation and Climate of the Ocean (ECCO)2 products described

1 At the time of writing, the ensemble ocean states from this system have already been used as ocean initial conditions for a set of CCSM4 decadal predictions contained in phase 5 of the Coupled Model Intercomparison Project (CMIP5; Taylor et al. 2012).

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

Corresponding author address: Alicia R. Karspeck, NCAR, P.O. Box 3000, Boulder, CO 80307.
E-mail: aliciak@ucar.edu

DOI: 10.1175/JCLI-D-12-00402.1

© 2013 American Meteorological Society
by Wunsch and Heimbach (2007) and Köhl et al. (2006), strongly enforce the dynamical constraints of the model, allowing observations from the future to exert influence on the past ocean state. In contrast, methods such as 3DVAR, optimal interpolation, and Kalman filters only use past and current observations to estimate the ocean state. Within this latter category, the way in which background information is specified is one important distinguishing characteristic. The 3DVAR and optimal interpolation methods typically specify parameterized background covariance estimates, for example, the National Centers for Environmental Prediction (NCEP) Ocean Reanalysis system, version 4 (ORAS4) (Mogensen et al. 2012) and Global Ocean Data Assimilation System (GODAS) (Behringer and Xue 2004). Some implementations of these can include parametric covariance forms that are calendar month and/or model state dependent, for example, the Simple Ocean Data Assimilation (SODA) (Carton and Giese 2008). Notable exceptions to parametric methods are a newer class of methods that use a fixed set of ensembles to compute sample background statistics, for example, the Australian Bureau of Meteorology Bluelink Ocean Data Assimilation System (BODAS) (Oke et al. 2008) and Global Ocean Reanalysis and Simulations (GLORYS2V1) (Ferry et al. 2012). Instead of specified covariance forms, Kalman filters use time-varying background covariance estimates that are determined via model dynamics. This naturally enables the multivariate physical relationships to be captured by the evolving background covariance. Ensemble Kalman filters, such as the one we use in the application presented here, share this general property of the Kalman filter. In the context of global ocean circulation models, variations on the ensemble Kalman filter are currently being employed with the Max Planck Institute Ocean Model (Leeuwenburgh 2007), the Poseidon ocean general circulation model (Keppenne and Rienecker 2001, 2002), and the Modular Ocean Model (MOM) (Zhang et al. 2007).

Ensemble Kalman filters are attractive for global ocean-state estimation for a number of reasons. From an operations perspective they are relatively easy to implement, with minimal interaction with preexisting code. This is especially important for use in a community model that undergoes frequent changes. Unlike the Kalman filter, which would be computationally intractable for use with a global ocean general circulation model, ensemble Kalman filters are scalable to large state spaces and can be made parallel for computation on multiple computer processors (Anderson and Collins 2007). Methodologically, they fall into the class of advanced data assimilation techniques that allow for time-varying model-determined background states. This enables the use of prior knowledge that is potentially more complex and informative than would be possible with either stationary background statistics or parameterized covariance forms. It also allows for covariability between ocean variables. For prediction purposes, they also have the natural benefit of delivering an ensemble of states that can potentially be used as initial conditions for probabilistic forecasts. And, in contrast to 4DVAR global ocean-state estimation systems, filters assimilate only past observations, making their historical state estimates appropriate for testing and calibrating ocean-initialized retrospective climate forecasts.

In this initial effort, the ocean assimilation system has been run over an 8-yr period from 1 January 1998 to 31 December 2005 with boundary forcing provided by an ensemble atmospheric analysis based on the atmospheric component of CCSM4. Because the period of assimilation is too short to evaluate interannual variability, we focus on the time-mean state. Insights from this short-term assimilation experiment will form the backbone of our understanding of how the assimilation system can be improved in future iterations.

The CCSM4 ocean model used for the data assimilation experiment is based on the Parallel Ocean Program, version 2 (POP2) (Smith et al. 2010). POP2 is a primitive equation level-coordinate global ocean model. The model has 384 (latitude) \times 320 (longitude) grid points in the horizontal, corresponding to a nominal 1° grid with increased resolution in the tropics. There are 60 vertical levels, with 10-m resolution in the upper 200 m, gradually expanding to 250-m resolution below 3000-m depth. Details of the ocean model are given in Danabasoglu et al. (2012). To highlight the role that in situ data can play in constraining the ocean, we compare the analysis to an ocean state driven by identical atmospheric forcing without the assimilation of data.

Section 2 gives an overview of the subsurface hydrographic observations that are assimilated, and the ocean EAKF system and its forcing are described in section 3. Section 4a provides some diagnostic checks on the performance of the assimilation system. In section 4b, we demonstrate that in the global average the assimilation of subsurface hydrographic data brings the ocean simulation closer to observations than an atmospheric-forced simulation without assimilation. However, below 1000 m there are identifiable regions of degradation due to the assimilation. Evaluation of the model simulations against satellite-derived sea surface temperature (SST) and sea surface height (SSH) data products

---

3 Associated with the Mercator Ocean project.
is presented in section 4c. Since one potential application of this system is ocean-initialized climate prediction, we focus our regional evaluations on the Northern Hemisphere Atlantic and the tropical Pacific, areas that have been identified as potentially relevant for that application. These are presented in sections 4d and 4e. In section 5, we discuss the challenges that emerged in this work and outline areas of potential improvement for the system.

2. Subsurface temperature and salinity observations

Subsurface temperature and salinity observations assimilated into the ocean model for the 1 January 1998 to 31 December 2005 period are taken from the World Ocean Database 2009 (hereafter WOD09) (Johnson et al. 2009). Figure 1 shows the number of temperature and salinity observations in the WOD09 during this period partitioned by the observational platform. Over this 8-yr period the data come primarily from autonomous drifting profiling floats (Argo floats; http://www.argo.ucsd.edu), ship-deployed conductivity–temperature–depth (CTD), expendable bathythermographic (XBT) instruments, moored thermistors, and surface drifting buoys including those from the Global Temperature and Salinity Profile Program (GTSPP). The WOD09 performs quality checks on all data in the archive, including (but not limited to) checks for duplicate observations, unrealistic values, and excessive vertical gradients. We used the standard depth level WOD09 product wherein raw data have been interpolated onto 40 standard depth levels ranging from the surface to 6000 m. The WOD09 corrects the XBT standard level data for the known errors in the drop rate equation (Hanawa et al. 1995) and the time-dependent temperature biases described in Levitus et al. (2009). Some of the profiling floats in the WOD09 were adjusted by the Argo project for drifts in their pressure sensors (Barker et al. 2011). Only those data that pass all WOD09 quality control standards are included in the assimilation. See Johnson et al. (2009) for a complete description of the WOD09 quality control procedures and interpolation scheme.

Over the time period of our assimilation, the expansion of the Argo network of autonomous drifting floats resulted in a significant increase in the observational coverage of temperature and salinity in the upper 2000 m of the global ocean. Modest deployment began in the year 2000 and by 2006 Argo floats were providing broadly distributed global coverage. Prior to the Argo era, the upper-ocean temperatures were observed primarily by XBT instruments deployed by ships of opportunity. These tended to be located along shipping routes (leaving large regions of the ocean unobserved) and are restricted to the upper 1000 m of the ocean. Salinity coverage, which came primarily from CTDs prior to Argo, was even more sparse. While only about 15% of the Argo observations are located in the 1000–2000-m depth range, this represents roughly an 18-fold increase in the number of observations taken between 1998 and 2005.

Other notable features in the observing network are the deployment of animal-mounted temperature sensors in the eastern midlatitude Pacific in 1998 (and to a lesser degree in 1999) and the GTSPP drifter deployment in the eastern midlatitude Atlantic in 2001. These observational platforms only provide data in the upper 250 m of the water column and in very limited geographic domains.

During the 8-yr period from 1998 to 2005, less than 1% of the observations in the WOD09 database are located below 2000 m. Those that are available come primarily from CTD casts taken at targeted locations with no repeat measurements. There are so few observations relative to the upper ocean that the deep ocean is essentially unconstrained by direct observations; changes to the ocean state below 2000 m are nearly all the result of data assimilation adjustments that originate from shallow observations and from the dynamical system.

3. An ensemble adjustment filter for the POP2 ocean model

The assimilation of observed data into a numerical model can be framed in a Bayesian context (Jazwinski 1970), wherein the prior multivariate state distribution
(the "forecast") is updated by observations to form a posterior distribution (the "analysis"). Ensemble methods of assimilation circumvent the computationally expensive problem of forming and manipulating the full multivariate distribution by assuming that the relevant features of the distribution can be estimated from the statistics of the ensemble members. This sample-based formulation is particularly important when dealing with large state spaces, as we have when performing a global ocean assimilation.

The EAKF is a deterministic variant of the ensemble Kalman filter, which was first introduced in the geosciences literature by Evensen (1994). Like all Kalman filters, it uses a nonstationary prior distribution that is evolved by the model dynamics. The EAKF update adjusts the daily ocean model forecast ensemble members such that the sample mean and covariance are consistent with the traditional Kalman filter update. (Higher order moments are not explicitly adjusted.) This new ensemble (the analysis) is then evolved by the model to the next time step, where it becomes the new forecast. This process advances sequentially through time. We implement the EAKF using the Data Assimilation Research Test-bed (DART) software (Anderson et al. 2009), which has been developed and is maintained at NCAR. We refer to Anderson and Collins (2007) for the technical details of the DART assimilation algorithm.

The optimality of the Kalman filter (and by extension the EAKF) is guaranteed only when the following assumptions are satisfied: (i) the forecast and observation errors are normally distributed, uncorrelated in time, and uncorrelated with one another; (ii) the model dynamics and mapping from the state space of the model to the state space of the observations are both linear, ensuring that the distributions remain Gaussian; and (iii) the forecast and observational error covariances can be represented exactly. The EAKF used here, which processes the observations sequentially at each assimilation update cycle, additionally assumes that the observational errors are uncorrelated with one another. The often noted assumption that models must be "unbiased" or capable of perfectly representing the system is related to the aforementioned assumptions of time-uncorrelated forecast errors and perfectly specified error covariances.

Like all global ocean models, POP2 does not satisfy these assumptions perfectly. Ocean general circulation models have systematic errors, are not perfectly linear in their dynamics, and have state spaces that are too large to allow for exact specification of their error covariances. Nevertheless, least squares data assimilation methods (of which the EAKF is an example) remain popular because they are tractable and useful, even if not optimal.

Here, a 48-member EAKF is used to assimilate observations of subsurface temperature and salinity from the WOD09 (described in section 2) into the POP2 global ocean model at a daily frequency from 1 January 1998 to 31 December 2005. Each day of the assimilation cycle, all standard level observations within a ±12-h window are assimilated at midnight model time. Both temperature and salinity observations are allowed to impact the multivariate prognostic state of the ocean model (consisting of the potential temperature, salinity, velocity, and sea level height on the model grid). Allowing multivariate covariance helps maintain the dynamical temperature–salinity and geostrophic balances in the model.

Each ensemble member of the ocean model is forced at the air–sea interface by a unique sample of the atmospheric state from an EAKF analysis of the atmosphere (Raeder et al. 2012). The atmospheric analysis uses the nominal 2°-resolution Community Atmosphere Model, version 4 (CAM4) (Neale et al. 2013) and assimilates temperature and winds from radiosondes, aircraft, and satellite-derived drift winds. In the atmospheric assimilation system, all members of the ensemble are forced by the Hurrell et al. (2008) SST and sea ice boundary dataset, which, over the period of the assimilation, is essentially the NCEP optimally interpolated SST (Olv2) (Reynolds et al. 2002). Sea ice concentrations in this dataset are derived from the Hadley Centre Global Ice and Sea Surface Temperature (HADISST) dataset (Rayner et al. 2003) and modified to be consistent with the NCEP OIv2 climatology.

In this configuration, the fluxes to the ocean at the air–sea interface are a function of both the prescribed atmospheric state and the evolving SST and surface velocities. The fluxes are calculated based on the bulk formulas described in Large and Yeager (2009). A weak salinity restoring to the Polar Science Center Hydrographic Climatology (Levitus et al. 1998; Steele et al. 2001) with a 4-yr time scale is applied over the upper 50 m of the ocean surface. The global mean salinity is adjusted at every model time step such that the restoring does not contribute to the global salt budget. Fluxes of freshwater into the ocean from river runoff are prescribed as the seasonal mean climatological values estimated by Dai and Trenberth (2002). We do not use an active sea ice model but, instead, prescribe daily sea ice fractions based on satellite measurements from the Special Sensor Microwave Imager (SSM/I) (Cosimo 1999). Note that these same sea ice data form the backbone of the HADISST sea ice concentrations used to force the atmospheric analysis. We refer the reader to Large et al. (1997) and Danabasoglu (2004) for the details of under sea ice fluxes of heat and freshwater into the ocean.
This loosely coupled setup differs from a fully coupled assimilation procedure in two important ways. First, because the ocean and atmosphere models are integrated independently, the air–sea fluxes forcing each component will not be identical. Second, from the data assimilation perspective, we do not exploit the possible covariances between the oceanic and atmospheric variables.

When the degrees of freedom in a system vastly exceed the number of ensemble members, as in the present case, spurious correlations can degrade the fidelity of the assimilation (Houtekamer and Mitchell 2001). This is an issue common to most ensemble-based methods; it reflects the fact that background error covariance matrices computed with a small number of samples will be rank deficient. In the EAKF the standard remedy for this problem is to use a localization function to restrict the impact of observations on geographically distant state variables (Anderson 2007). Here we use a compactly supported fifth-order piecewise polynomial localization function (Gaspari and Cohn 1999) with an isotropic radius of ~11° in the horizontal. No localization is being applied in the vertical, allowing observations at any depth to impact the entire water column. Issues related to this choice are discussed in sections 4b and 4d.

In the EAKF framework, the presence of sufficient spread in the prior ensemble is necessary for observations to impact the ocean state. In theory, ensemble spread can be maintained by the intrinsic internal variability of the model dynamics. In practice, however, very few geophysical models produce sufficient spread from internal variability alone. In a nominal 1°-resolution ocean model, as we use here, the contribution of oceanic internal variability to the ensemble spread is minor, and nearly all of the ensemble variability is imparted by the use of an ensemble of atmospheric boundary forcing. This is essentially a form of variance inflation that can be understood as ocean model error that stems from our uncertainty in the atmospheric conditions. Most ensemble filtering methods underestimate variance because of statistical sampling error stemming from the use of a finite number of ensemble members (Furrer and Bengtsson 2007). Artificial variance inflation can be used to counter both this variance loss and to simulate other model errors that are not captured by the variance in atmospheric forcing. That being said, it was not clear in the design of this system whether artificial inflation would be necessary. So, while the option to artificially inflate the ensemble spread can potentially be exercised, in the current configuration we do not use any explicit variance inflation beyond the boundary forcing.

Observations that differ from the daily model prior mean value by more than three times the square root of the sum of the observational and prior predicted error variance are excluded from the assimilation. In unbiased models, this outlier rejection criterion acts as an additional quality control measure on the observations. In models with significant biases, this can lead to the selective culling of observations that are not in agreement with the model. Examples of this are discussed in sections 4c and 4e. Observations located within model grid cells with a land boundary are also excluded from the assimilation because of the ambiguity this poses for interpolation from the state space of the model to the observational state space.

The EAKF algorithm requires estimates of the error variance associated with the observations that are being assimilated. The observational error includes the instrumental error and the error due to unresolved time and space scales in the model (called the “representativeness error”). In coarse-resolution models, the representativeness error is the dominant observational error source (e.g., Oke and Sakov 2008; Forget and Wunsch 2007, and references therein). The true magnitude of this error will depend on the dynamics and resolution of the model as well as the geographic location of the observations. For simplicity in the current configuration, however, a single error estimate is used globally. For temperature, the standard deviation of the error is set at 0.5°C across all platforms. For salinity, the standard deviation of the error is set at 0.5 practical salinity units. While reasonable in some regions of the ocean, these choices are likely too small in regions with high mesoscale variability, such as the western boundary currents, and too high in more quiescent areas, such as the abyssal ocean.

The ensemble of 1 January model states for initializing the filter was chosen randomly from a set of multi-decadal integrations of the POP2 ocean model forced with a dataset of historical atmospheric states (Large and Yeager 2009). These integrations included simulations with both active sea ice models and prescribed ice coverage, and a range of surface salinity restoring time scales. This was our best estimate of a climatological distribution of POP2 ocean states and results in significant initial variance at all levels in the ocean.

4. Results

a. Daily assimilation statistics

In the optimal case, when all the Kalman filter assumptions are exactly satisfied, the EAKF update will draw the ensemble mean of the analysis closer to the
true state of the ocean than the forecast in a least squares sense. In practice, of course, there is no truth against which to evaluate this proposition. Only observations can be used for evaluation, and, as mentioned in section 3, these contain a combination of instrumental and representativeness errors. Because of these errors and because of the least squares nature of the Kalman filter solution, there is no guarantee that the magnitude of the analysis-minus-observation statistic \( e^2_a \), hereafter the “analysis squared misfit”) will always be less than the magnitude of the forecast-minus-observation statistic \( e^2_f \) (“forecast squared misfit”) at the location of an individual observation. Here \( a_i \) (\( f_i \)) is the analysis (forecast) ensemble mean at the time and location of an individual observation \( o \) indexed by \( i \). Further, and for the reasons outlined in the preceding section, the system is not optimal.

Heuristically, however, we can expect that, if the observational error is not so high that data are ignored and if the background statistics used by the filter are not wildly incorrect, then when averaged over a sufficiently large number of observations the assimilation update will draw the model solution closer to the observations on the whole. The reduction of the average squared misfit due to the assimilation update (i.e., \( N^{-1} \sum e^2_f, N^{-1} \sum e^2_a \)) is shown in the upper panel of Fig. 2 for temperature observations. For this measure, the summation is over all \( N \) observations in 30-day windows at each model level.\(^4\) (Only levels above 1000 m are shown because below this level there are very few observations in each time bin in the early record.) That all values are positive provides a qualitative check that the assimilation update is drawing the model solution closer to the observations.

In contrast to the heuristic argument made above, the EAKF update equation for the covariance guarantees that the ensemble variance at all points in the state will be reduced by each daily assimilation cycle. As before, averaged over all assimilated temperature observations in 30-day windows at depth levels from 100 to 1000 m, the lower panel of Fig. 2 shows the daily reduction in ensemble variance at the location of temperature observations. That all values are positive provides a diagnostic check on this constraint.

Figure 3 shows the time evolution of the daily forecast misfit \( N^{-1} \sum e^2_f \) and ensemble variance. As in Fig. 2,

\[^4\) Observations associated with high-volume regional campaigns, such as the 2001 GTSPP drifter deployment in the Atlantic and the animal-mounted sensor program in the California Current, are excluded from this measure because they bias the global metric.
the summation is over all \( N \) observations in 30-day windows at each model level. Here we see that, even though data assimilation is consistently reducing the misfit (Fig. 2), there is not a simple monotonic decrease in the global average forecast misfit as the assimilation progresses. Especially in the years prior to 2003, before the Argo program resulted in more uniform global observational coverage, global forecast misfits are larger and more variable, with a clear seasonal cycle in the upper ocean. In the global average during this time, the misfit reduction owing to data assimilation is not sufficient to reduce error growth associated with the seasonal cycle. After 2003, however, the misfit is lower and more stable. As expected, the ensemble variance decrease is very pronounced in the first few years of the assimilation, with the ensemble contraction primarily coming from the data assimilation (see lower panel of Fig. 2). The variance in the upper 200 m is relatively stable beginning in 1999 owing to variance imparted by the ensemble atmospheric forcing. Below this level, however, the variance contracts throughout the life of the assimilation.

As mentioned previously, the EAKF update equations assume that the error in both observations and forecast is random, implying that there should be no systematic component to their difference. Unfortunately, this strict assumption is rarely satisfied in practical ocean modeling, and the presence of systematic errors in the forecast relative to the observations has implications for the optimality of the ensemble Kalman filter. Time-mean bias is one type of systematic error, which is relatively simple to diagnose. Examination of the forecast misfits through time within geographical subsections of the model domain can be used to evaluate what fraction of the ocean suffers from a loss of optimality due to a time-mean bias. We bin the forecast-minus-observation statistics at each of the WOD09 standard depth levels into \( 5^\circ \times 5^\circ \) boxes. The total forecast mean squared misfit within each box can be decomposed as

\[
\frac{1}{N} \sum_{i=1}^{N} (f_i - o_i)^2 = (\bar{f} - \bar{o})^2 + \frac{1}{N} \sum_{i=1}^{N} [(f_i - \bar{f}) - (o_i - \bar{o})]^2.
\]

Here, subscript \( i \) indexes all \( N \) observations at all times located within a geographical box. The overbars indicate averaging over all \( N \) observations within each geographical box from 1 January 2000 to 31 December 2005. (The 2-yr period from 1 January 1998 to 31 December 1999 is assumed to be an adjustment period and is not included in this evaluation.) The first term on the right contains the contribution from the squared time-mean bias and the second term is the contribution from the time-varying component of the squared forecast misfit. Throughout the remainder of this paper, the term bias should be understood as defined in (1): \( \bar{f} - \bar{o} \). It should be noted that these statistics are computed only for the period of the assimilation experiment. As a result, decadal-scale and longer variability in the forecast misfit...
is implicitly included in the bias term. While it would be preferable to assess time-mean bias over a longer time interval, this limited analysis gives us a sense of regions that are potentially problematic. The magnitude of the bias term relative to the total square misfit gives an indication of whether the bias may be important. Figure 4 shows that the fraction of $5^\circ \times 5^\circ$ regional boxes where the bias component of the squared misfit is less than 20\% of the total forecast squared misfit. Circles (squares) are for temperature (salinity).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig4}
\caption{At each depth level, fraction of $5^\circ \times 5^\circ$ regional boxes where the bias component of the squared misfit is less than 20\% of the total forecast squared misfit. Circles (squares) are for temperature (salinity).}
\end{figure}

Here, the monthly-mean fields from the assimilation (hereafter Assim) will be compared against an ocean simulation in which no temperature and salinity observations have been ingested. To generate this control simulation (hereafter NoAssim), we perform a set of 48 POP2 integrations in which the initial conditions (on 1 January 1998) and atmospheric forcing are identical to those used in Assim.

Figure 5 shows the time evolution of the horizontal mean spread (ensemble standard deviation) in temperature for Assim normalized by the horizontal mean spread in NoAssim as a function of ocean depth. For both temperature and salinity (not shown), within the first month of assimilation there is a contraction in the fractional spread at all depth levels to about 40\% percent of the spread without assimilation. This rapid reduction in spread is typical of the spinup period of an ensemble data assimilation system. In the upper 50 m this equilibrates to a near-constant level of relative ensemble spread within the first two years of the assimilation. Deeper levels are slower to reach a steady state, and below 250 m the spread in temperature and salinity continue to decrease throughout the eight years of the assimilation.

The rapid adjustment of the shallow layers of the ocean to a steady ensemble spread is expected because the flux-driven ensemble variance is sufficient to balance the contraction by observational constraints. However, below this surface layer there is not sufficient variability imparted by the atmosphere to compensate for the ensemble contraction. Because no vertical localization is used in this system, the spread at depth is contracting both in response to observations at the local depth level and as a result of vertical covariance with upper-ocean observations.

Figure 6 shows the total mean-squared misfit over all observations from 1 January 2000 to 31 December 2005 for Assim and NoAssim as a function of depth. (The 2-yr period from 1 January 1998 to 31 December 1999 is assumed to be an adjustment period and is not included in this evaluation.) Here the squared misfits are defined as in (1) except that the forecast value is time and space interpolated from the monthly average fields from Assim and NoAssim. We compare in this way for consistency between the Assim and NoAssim experiments; while daily statistics are available for the Assim experiment, due to computational limitations only monthly averaged ocean fields for the NoAssim experiment were retained.$^5$ While this is not a comparison to independent data, which are used in the sections

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig5}
\caption{At each depth level, fraction of $5^\circ \times 5^\circ$ regional boxes where the bias component of the squared misfit is less than 20\% of the total forecast squared misfit. Circles (squares) are for temperature (salinity).}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig6}
\caption{At each depth level, fraction of $5^\circ \times 5^\circ$ regional boxes where the bias component of the squared misfit is less than 20\% of the total forecast squared misfit. Circles (squares) are for temperature (salinity).}
\end{figure}

$^5$ In the Assim experiment, where both daily and monthly squared misfits can be computed, the two methods show very close agreement.
that follow, the monthly averaged fields are computed from the daily forecasts, prior to the assimilation of any observations.

At all depths to 2000 m the horizontal average mean-squared misfit in temperature and salinity are lower with assimilation. The greatest improvement attributed to assimilation tends to be below the thermocline from 250-m to 1000-m depth. One reason why we do not see more improvement below 1000 m is that the Argo profiling floats below 1000 m have very little influence when they finally become plentiful after 2003 because the ensemble spread at depth is so depleted by that time.

The difference between the mean square misfit in Assim and NoAssim stems from a reduction of the bias component [see (1)] of the squared misfit in Assim (shown by the gray lines in Fig. 6). This is commonplace for many ocean assimilation systems (e.g., Balmaseda et al. 2007; Carton and Giese 2008). In part, this results because a large fraction of the time-varying component of the forecast misfit is actually observational error, which is irreducible. The portion of the time-varying component that is resolvable by the model and potentially reducible by the data assimilation system cannot be robustly assessed with only six years of simulation.

This is because most time scales of variability in a coarse-resolution model are seasonal or longer (giving us a maximum of six instances for verification) and because POP2 simulations forced by observationally based fluxes without assimilation are known to reasonably simulate variability of upper-ocean variables even without assimilation (Danabasoglu et al. 2012; Doney et al. 2007; Yeager et al. 2012).

In the context of Fig. 6, it is worth keeping in mind that some of the adjustments made by the data assimilation scheme persist without continual readjustment. Because of this, the differences between the bias components of Assim and NoAssim are not, in and of themselves, a measure of whether the Kalman filter assumption of unbiased forecast errors is violated. From the perspective of the data assimilation scheme, it is only those errors that are continually present or emergent, as measured by (1), that indicate a suboptimality with the system. While Fig. 6 gives a global measure of the bias difference between Assim and NoAssim, Fig. 4 provides a sense of the fractional horizontal ocean area that is potentially affected by model bias in the context of the assimilation.

Spatial maps of the bias component of misfit relative to observations for temperature and salinity from
1 January 2000 to 31 December 2005 are shown in Figs. 7 and 8 at 100-m-, 700-m, and 1400-m depth. As before, these are comparisons of the WOD09 daily observations to the associated monthly-mean model output in $5^\circ \times 5^\circ$ boxes. At 100 m major temperature and salinity biases associated with the positions of the western and eastern boundary currents and the frontal zones of the Antarctic Circumpolar Current (ACC) are reduced by assimilation (albeit not completely eliminated). These signals tend to be barotropic down to ~1000 m and so show up at both 100 and 700 m. At 700-m depth, the assimilation system is effective at reducing the large-scale biases in temperature and salinity poleward of about 25°.

At 1400-m depth there are significant differences in the solutions due to assimilation, but assimilation does not consistently lead to better agreement with the observations. Most of the adjustment occurs in the first few months of assimilation (when the deep ocean ensemble spread is still large) before the deep Argo observations become plentiful. This suggests that vertical covariances in the model are responsible for most of the differences at these deep levels. Some regions do show improvement with assimilation, however. In the eastern North Atlantic the warm, salty plume of water from the Mediterranean outflow is better simulated. Assimilation also cools and freshens the waters of the Atlantic deep western boundary current south of the Grand Banks at 35°N, bringing them into better agreement with observations (not shown). At least in part, the deep Northern Hemisphere Atlantic is better simulated than other basins because there was a relatively high level of observational coverage from CTDs in the 1000–2000-m range in 1998 when the assimilation was initiated and the ensemble spread was still high.

c. Comparison to satellite data

We use satellite SST and SSH to further assess whether the time-mean ocean state is improved through data assimilation. As described in section 3, the Hurrell SST is a monthly-mean dataset based on the NCEP OIv2 optimally interpolated SST. During the 1 January 2000–31 December 2005 period of comparison the Hurrell SST dataset largely comprises satellite data from the Advanced Very High Resolution Radiometer, with in situ data from ships and buoys only used for correcting satellite bias. There is no significant overlap between the
data used in the Hurrell SST and the WOD09 data that we assimilate. However, it should be noted that this is the same SST dataset used as the lower boundary condition for the atmospheric assimilation. We also compare to satellite SSH, which reflects both density changes throughout the water column and dynamical mass redistribution. The SSH dataset used here is based on distributions of monthly-mean global $1/8^\circ$ gridded dynamic topography from satellite altimetry (Ssalto/Duacs multi-mission altimeter products; http://www.aviso.oceanobs.com/duacs/; hereafter AVISO). Global area-weighted averages of SSH for the model and the AVISO data have been subtracted to compensate for the unknown geoid. No altimetry data were assimilated into the ocean model, making this a completely independent data source. Observed SST and SSH fields have been bilinearly interpolated onto the POP2 grid for comparison to model simulations.

For both SSH and SST, the bias component of the total mean-squared misfit at each grid point is computed as $\sigma^2 = (x - \bar{x})^2$, where $x$ and $\bar{x}$ are the time-mean values of model and data (respectively) computed over all 72 months from January 2000 to December 2005. Figure 9 shows the geographic regions where the squared

![Figure 7](image-url)
misfits are altered through assimilation. Blue indicates model grid boxes where assimilation reduces the square error between model and data and gray indicates regions where the assimilation degrades the solution. Regions that are not statistically significant or where any fraction of the grid cell is covered in ice are left unmarked. The appendix describes the parametric bootstrap technique used to assess the statistical significance of differences in the mean squared misfits.

Overall, the time-mean signal for both SST and SSH is improved in most regions of the globe where statistical significance can be established. For SST (SSH) 25% (41%) of the global sea surface area is improved with assimilation, 4% (12%) is degraded, and 71% (47%) is not statistically significant. There are notable improvements in the areas associated with western boundary currents and coastal upwelling regions. However, there are also smaller-scale regions that show degradation due to assimilation. Along sharp frontal zones, such as the western boundary currents and the ACC, large-scale adjustments that improve the position of the mean gradients can erroneously alter neighboring regions if the model does not capture the correct placement of eddies and meanders.

In the midlatitudes, adjustments to the depth of the eastern boundary thermocline at the beginning of the assimilation improve the representation of coastal upwelling. Over multiple years, however, these initial adjustments propagate westward across the basin as planetary Rossby waves, which can degrade both the
mean and variability (not shown). These features become increasingly difficult for the assimilation system to correct because the thermocline deepens to the west where there are fewer observational constraints and lower ensemble spread.

The formation of new water masses at the beginning of the assimilation when the spread is large can also degrade the solution. If the density changes are large enough to destabilize the water column, the changes can drive spurious vertical motions. This behavior appears to be responsible for the degradation of SSH in the Southern Hemisphere subtropical gyre of the Pacific and may be a factor in the SSH degradation in the Labrador Sea. Because of vertical covariance with the surface fields, dense water masses also form in the deep ocean at the start of the assimilation. If they are formed at bathymetric slopes, the water can spread along the ocean floor, filling deep basins. This can lead to spurious time mean and variability in the SSH. We observe this to be the case in various places in the Southern Ocean and in the equatorial Atlantic (discussed further in section 4d).

In Assim, the largest SST improvement in the subtropical and subpolar regions of the Northern Hemisphere Pacific and Atlantic Oceans stems from more realistic pathways of the Kuroshio and the Gulf Stream/North Atlantic Current systems. The locations of these warm western boundary currents are associated with large gradients of SST and SSH, so small errors in the position and sharpness of the current fronts can lead to large discrepancies in the surface fields.

In both basins assimilation sharpens the front and moves the latitude of coastal current separation southward. In the Atlantic, assimilation also steers the path of the North Atlantic Current (the northern extension of the Gulf Stream) north at around 45°W. Figure 10 illustrates the SST bias reduction due to assimilation in both basins. We present the results using NoAssim as the baseline of comparison because this highlights that the spatial pattern of change owing to assimilation is broadly consistent with the biases in NoAssim. Assimilation acts to reduce the bias, in this case improving the pathway of the western boundary currents. Off the coast

![SSH time mean](image)

![SST time mean](image)

*Fig. 9.* Blue indicates model grid boxes where assimilation reduces the bias component of the squared misfit between model SSH (or SST) and satellite-derived AVISO SSH (or Hurrell SST) relative to the NoAssim experiment. Model grid boxes in gray indicate regions where assimilation increases the model data misfit (degrading the solution). Only those grid points that are statistically significant are marked (see appendix).
of Japan the southward shift in the separation latitude of the Kuroshio with Assim leads to a localized cooling of 6°C on the northern side of the current and a broader warming of ~1°C on the equatorward side. Assimilation essentially eliminates the bias we see in the NoAssim run. In the Gulf Stream region the major SST biases in NoAssim are related both to the latitude of separation and to the sharp northward turn of the current into the subpolar North Atlantic. These biases are reduced by 35%–50% in Assim (see caption in Fig. 10). The SSH fields exhibit similar patterns of improvement (not shown).

While it is not altogether clear why assimilation is more effective in adjusting the path of the Kuroshio, it likely stems from the special complexity of the Gulf Stream region; variations in the pathway have been linked to, among others, bottom topography (Warren 1963), interaction with the deep western boundary current (e.g., Yeager and Jochum 2009), changes in the Labrador Current (Hameed and Piontkovski 2004), and remotely forced westward propagating Rossby waves impinging on the east coast of the United States (Gangopadhyay et al. 1992). These are all indirect influences that can compete against the assimilation of local in situ observations. Because the Gulf Stream region has such severe biases without assimilation, it presents a special challenge for the data assimilation system. The outlier rejection criteria (described in section 3), which is designed to filter true observational outliers, can act to cull a disproportionate number of observations in severely biased regions. In the upper 1000 m in the 5° box centered at 46°N, 39°W (the “+” in Fig. 10), for example, the percentage of observations ignored increases from 30% in 1998 to 45% in 2005. The increase in rejected observations over the life of the assimilation system: decreasing the local ensemble spread by assimilating observations increases the number of observations rejected in the following assimilation cycle, and as more observations are ignored the solution drifts back toward its naturally biased state. This is analogous to the feedback that leads to the well-known problem of “filter divergence.”

The problem can be remedied to some extent by an artificial inflation of the ensemble spread. However, in this model, artificially inflating the ensemble can have an even more problematic outcome: in the frontal regions, increasing the variance in the forecast ensemble leads to the assimilation of more observations because fewer are deemed outliers and a tighter alignment with each one because the forecast is considered more uncertain. This can force the model into numerically unstable configurations. We observed in offline experiments with inflation that numerical instabilities arose when the coarsely resolved ocean cleaved tightly to the observed sharp and meandering fronts of the Gulf Stream. Whether an assimilation system rejects observations or adheres to them closely is intrinsically related to the appropriate specification of the representativeness error,
which helps to determine the middle ground between the true state of the ocean and a state that is numerically stable for a coarsely resolved model. Work on specification of regionally dependent representativeness error is ongoing.

d. Meridional overturning circulation and heat transport in the Atlantic

In this section, we use the Atlantic meridional overturning circulation (AMOC) and its associated northward heat transport (NHT) as basin-scale diagnostic quantities to frame our understanding of the changes to the Atlantic Ocean state that arise from assimilating subsurface data. We focus on this region because it has been posited that the decadal variability of the AMOC may be a predictable component of North Atlantic climate (Griffies and Bryan 1997; Msadek et al. 2010; Teng et al. 2011; Yeager et al. 2012).

The time-mean AMOC for the Assim and NoAssim runs are shown in Fig. 11. The latitude of maximum overturning in both runs is close to 35°N, with Assim transporting about 5.5 additional sverdrups (Sv = 10⁶ m³ s⁻¹) in the upper 1000 m. In Assim there is greater equatorward transport at deeper levels, reflecting an intensification of the southward flowing deep western boundary current (DWBC). Assimilation also leads to a deep equatorial countercirculation of about 7.5 Sv that is absent in NoAssim. In Assim, this flow is driven by a surge of cold dense water over the equatorial sill west of the mid-Atlantic ridge and through the narrow bathymetric channels that connect the two hemispheres. This gravity current is a transient flow resulting from the large temperature and salinity assimilation increments that occur within the first couple of months of assimilation. (Similar flows are observed in the deep closed basins of the Southern Ocean.) From 1998 to 2005, this circulation diminishes from 15 Sv to 5 Sv, suggesting that it is a spurious feature of the filter initialization. Muñoz et al. (2011) show that there are other commonly used ocean reanalysis products that contain

![Fig. 11. (a),(b) Atlantic meridional overturning circulation time averaged from 2000–05 for Assim and NoAssim. The AMOC is defined as the streamfunction of the zonally integrated northward volume transport in the Atlantic Ocean basin. Each level indicates the depth integrated transport from the surface. Zero contour in bold, contour intervals are 4 Sv; shaded areas are positive contours. The maximum value of AMOC is 24 Sv in Assim and 18.5 Sv in NoAssim. (c) AMOC profiles at 26.5°N from Assim (solid line), NoAssim (dashed line), and estimates from RAPID (gray line) for the 21-month overlapping period April 2004 to December 2005.](image-url)
a deep equatorial countercirculation in the time-mean AMOC as well; however, the origin of these flows and their realism is unclear.

While there are no direct observations of the AMOC, estimates based on a transatlantic section of hydrographic moorings and current monitors, called the Rapid Climate Change–Meridional Overturning Circulation and Heatflux Array (RAPID–MOCHA), deployed along 26.5°N (Cunningham et al. 2007) are available back to 2004, providing a baseline against which to evaluate the Assim and NoAssim profiles. We compare only the 21-month period of overlap from April 2004 to December 2005. At this latitude the volume transport in the upper 1000 m is approximately 3 Sv greater in Assim than in NoAssim, bringing the transport into closer agreement with the estimates from RAPID (Fig. 11c). This increase in transport appears to come from an enhancement of the northward Gulf Stream flow at deeper levels (500–1000 m). Over the same 21-month period, assimilation increases the mean NHT at 26.5°N from 0.8 to 0.95 PW (Fig. 12). This brings the NHT into better agreement with the observationally based estimate of 1.4 PW reported in Johns et al. (2011). The NHT increase of 0.15 PW is consistent with the relationship between changes in the AMOC and NHT diagnosed in Msadek et al. (2013). In their study, this relationship is set primarily by the overturning circulation acting on the zonal mean vertical temperature gradient. Relative to NoAssim, Assim at this latitude shows a coherent large-scale pattern of warmer upper 1000-m water and cooler waters in the 1000–2000-m depth range (consistent with data from the WOD09) that acts to increase NHT.

Between the equator and 30°N the impact of the spurious deep countercirculation in the Assim AMOC can be observed on the NHT. This circulation is transporting large amounts of cool water into the abyssal planes of the Atlantic, driving a reduction in the net NHT in Assim. The set of ocean reanalyses that contain deep equatorial countercirculations presented in Muñoz et al. (2011) also show a marked reduction in NHT between the equator and 30°N. This highlights the danger of using metrics that rely on correctly simulating the currents and temperatures of the abyssal ocean, where there are almost no observations, as a means of evaluating model or assimilation performance.

The DWBC, which carries cold, recently ventilated waters from the far North Atlantic down the coast of North America, is thought to be the primary pathway of the deep southward flow of the overturning circulation. Modeling studies have demonstrated that the passage of the DWBC beneath the Gulf Stream can impact the latitude at which the Gulf Stream separates from the coast of North America, suggesting that it is an important feature of the deep circulation to simulate in models (e.g., Yeager and Jochum 2009; and references therein).

The core of the DWBC drops below 2500 m as it follows the topographic slope around the Grand Banks and on toward the equator. This has been documented observationally by Joyce et al. (2005), who estimate that the majority of the 14–19 Sv of transport in the DWBC near 37°N is below 2500 m. This transect is marked in Fig. 13. For comparison to the Joyce et al. data, we define two layers, one from 1000–2500 m (the upper layer) and one from 2500 m to the sea floor (lower layer). In the two layers combined below 1000 m, Assim has a stronger DWBC (Fig. 13) with 12.9 Sv of total transport compared to 6.3 Sv of transport in NoAssim. In NoAssim, these transports are nearly evenly distributed between the layers (2.8 Sv and 3.5 Sv in the upper layer and one from 2500 m to the sea floor (lower layer). In the two layers combined below 1000 m, Assim has a stronger DWBC (Fig. 13) with 12.9 Sv of total transport compared to 6.3 Sv of transport in NoAssim. In NoAssim, these transports are nearly evenly distributed between the layers (2.8 Sv and 3.5 Sv in the upper and lower layers, respectively) while transport in the Assim case is primarily in the lower layer (4.7 Sv vs 8.2 Sv), making it more consistent with the magnitude and depth distribution reported in the literature.

It is interesting to note that data assimilation changes the properties of the Nordic Seas overflow source water, making it colder, fresher, and denser at the level of the overflow sills (not shown). This brings the temperature closer to the observational estimates reported by Legg et al. (2009) and to the raw observational values in the WOD09. The freshening, however, overshoots the Legg et al. values. The net increase in density in the overflow region is consistent with a deeper DWBC. However,
changes to the DWBC south of the Grand Banks emerge within the first few months of assimilation, long before changes in the overflow region properties could advect to the midlatitudes. Instead, it appears that assimilation is making local changes to the density structure along the entire path of the DWBC that support a deeper flow.

e. Tropical Pacific

The equatorial Pacific is characterized by a deep thermocline in the western part of the basin relative to the east. Although SST in the tropical Pacific is controlled by a collection of processes, the broad-scale feature of cool surface water in the east and warm water in the west is related to the proximity of the thermocline to the surface. This is captured well by both Assim and NoAssim compared to the Hurrell SST observations. The model SST shows only small differences due to assimilation but of the correct sign (Fig. 14).

Off the coast of Central America, the NoAssim SST is too cool to the north of the equator and too warm to the south, such that the northern SST front is too weak. In their diagnosis of systematic errors in the CCSM version 3 coupled model, Large and Danabasoglu (2006) attribute SST bias in the eastern equatorial Pacific mainly to problems with the incoming solar radiation and with weak coastal wind stresses in the atmospheric model. Because the atmospheric forcing fields are identical in Assim and NoAssim, biases remain in these regions. This is part of a broader pattern of systematic errors along the upwelling regions of the eastern ocean.

![Figure 13](https://example.com/figure13.png)

**FIG. 13.** Vectors of volume transport in the North Atlantic deep western boundary current: blue (green) lines represent southward (northward) transport. Transport with data assimilation between (a) 1000-m and 2500-m depth and (c) 2500 m to the ocean floor; (b),(d) the transport without data assimilation. For clarity, only grid cell transports greater than 0.5 Sv are plotted. Circled vector near 30°N, 15°W in (a) is 4 Sv; the dashed line marks the transect used for comparison to transport estimates from Joyce et al. (2005).

![Figure 14](https://example.com/figure14.png)

**FIG. 14.** Difference between (top) the Hurrell SST and the NoAssim SST in the tropical Pacific and (bottom) the Assim SST and the NoAssim SST. Zero contour indicated in bold, with gray shading over positive contours; contour intervals are 0.2°C.
boundaries that the assimilation of ocean data reduces but does not remove (see also the California Current and the Canary Current in Fig. 10).

The complex picture of how assimilation impacts the equatorial region can be seen succinctly by comparing the World Ocean Atlas 2009 (WOA09) subsurface temperature climatology (Locarnini et al. 2010). Assimilation warms the upper thermocline along the entire equator, nearly eliminating the NoAssim temperature bias in the western half of the basin (Fig. 15). In the east, however, between 50-m and 100-m depths, this warming actually degrades the solution relative to the WOA09 climatology. Although there have been suggestions that the WOA09 is too smooth at the equator (e.g., Danabasoglu et al. 2012), we confirm this degradation by comparing directly to measurements of temperature from the Tropical Atmosphere Ocean (TAO; http://www.pmel.noaa.gov/tao) moored array averaged from 1 January 2000 to 31 December 2005.6 (Note that data from these moorings are included in the collection of assimilated observations.) The depth of the 20°C isotherm in Assim is in good agreement with observations in the west but is nearly 30 m too deep in the far east (not shown).

The upper layers of the eastern equatorial Pacific are well observed, so assimilation could potentially be more effective in reducing the bias. However, as noted earlier, the ocean bias in this region appears to be linked to the atmospheric forcing and may also reflect inherent problems with the ocean model. For example, insufficient mixing in the eastern tropical thermocline may prevent the formation of proper covariant relationships with the temperatures in the warm pool. The outlier rejection criteria resulted in the culling of approximately 35% of the subsurface observations located in the east equatorial Pacific (i.e., above 150 m, within 5° of the equator, east of 140°W). To put this fraction in perspective, only about 4% are rejected west of 140°W and less than 1% are rejected below 150 m. This offers an explanation of why Assim does not closely match the observations in this region. As discussed in section 5, we are actively exploring how this might be remedied through changes to the assimilation system.

Daily acoustic Doppler current profile (ADCP) data from the TAO project from 1 January 2000 to 31 December 2005 were time averaged at four locations on the equator for comparison of the equatorial undercurrent. Everywhere along the equator assimilation leads to a deepening of the undercurrent maximum (Fig. 16). In the western and central part of the basin, at 165°E and 170°W, this brings the undercurrent into closer agreement with the TAO ADCP measurements.7 However, consistent with the lack of near-surface temperature improvement in Assim, there is no improvement in the undercurrent position or strength east of 150°W. But, as with temperature, zonal velocities are improved below 100 m.

5. Summary and discussion

We have reported on the development of a new data assimilation system for the POP2 ocean component of CCSM4. In the current configuration, subsurface temperature and salinity data from the WOD09 have been assimilated using a 48-member EAKF, where each ensemble member of the ocean is forced with a sample from the posterior atmospheric analysis from CAM4. On balance, the assimilation of subsurface data improves the time-mean ocean-state estimate relative to an identically forced ocean model simulation. Most of this improvement is in the upper 1000 m of the ocean. Because the simulation is only 8 years in length, we do not present an assessment of the variability. Simulations of greater than 30 years, which are currently planned, will allow us

6 The temperature differences due to assimilation exceed the TAO temperature array instrumental error reported by Freitag et al. (1994) by more than an order of magnitude.

7 The differences we show are roughly an order of magnitude greater than the accuracy of the data as reported in Johnson et al. (2001).
to robustly determine how data assimilation alters the seasonal cycle and variability in the ocean model. Identifiable issues notwithstanding, the results of this first endeavor are encouraging and (more importantly) shed light on ways that the system can be made more accurate and reliable for ocean-state estimation and for use in ocean-initialized climate predictions.

Like the ensemble adjustment Kalman filter used here, most data assimilation formulations that are actively used for ocean-state estimation intrinsically assume that models produce forecasts without systematic errors. While the POP2 model simulates the gross behavior of the global ocean, time-mean differences between the real ocean and the simulated ocean were identified here. Computational limitations naturally lead to the use of coarse-resolution models, and this can be a major source of systematic error in ocean simulations. The atmospheric fields used to force an ocean assimilation can have systematic errors as well, and these can force further deviations from the true ocean state. Because a community model is used, we are compelled to use the model in the form released by its developers. So, in practice, systematic errors must be accommodated by the data assimilation system. It may be possible to correct some of the time-mean biases in the ocean assimilation by improving our utilization of available observations and expanding our data sources. But, in general, how to best deal with systematic model errors using a data assimilation system remains an open question. One approach to bias correction in the literature is via augmentation of the model state by an estimate of the bias (e.g., Keppenne et al. 2005; Baek et al. 2006). We refer to Chepurin et al. (2005) for a broader discussion of systematic model error as it relates to ocean data assimilation.

We have identified a number of focus areas for improving the assimilation system. For example, the use of adaptive ensemble variance inflation (Anderson 2009) can potentially increase the information that we can draw out of the observational network. This is especially important at ocean depths greater than 1000 m, where the boundary-forced ensemble spread is small and the Argo

---

**FIG. 16.** Zonal velocity at four locations along the equator from Assim (solid line) and NoAssim (dashed line) and from the 1 Jan 2000 to 31 Dec 2005 average of the TAO ADCP (gray circle line).
network has the potential to provide widespread observational constraints. It may also be important in regions of strong model bias (e.g., western and eastern boundary currents and eastern tropical Pacific). In addition, it is important to refine our observational error estimates to include spatially heterogeneous representativeness error. In coarse-resolution models this can be the dominant observational error source in regions with strong temperature and salinity fronts or strong mesoscale and submesoscale variability. Methods for estimating these errors have been described in a number of papers including Forget and Wunsch (2007), Richman et al. (2005), Oke and Sakov (2008), and others. Developing better estimates of observational error is also necessary for making quantitative assessments of whether the ensemble spread in the system is a reliable indicator of uncertainty. Finally, if and how vertical covariances should be limited by localization is still an open area of investigation. In this work we were able to identify instances where vertical covariances in the model led to spurious water mass formation, but there is also evidence from the literature that short length-scale localization in the vertical can destabilize the water column (Zhang and Rosati 2010).

Information that we have gathered as a result of this initial work with ensemble data assimilation for POP2 will help guide ongoing efforts to extend the analysis back in time and ultimately to develop a fully coupled ocean–atmosphere data assimilation system for CCSM and CESM.

Acknowledgments. We acknowledge the efforts of all the scientists and software engineers who contributed to the development of CCSM4 and the DART data assimilation system. Special thanks are extended to Marianna Vertenstein and Brian Kauffman of the CESM Software Engineering Group, who worked on the CESM–DART infrastructure. We appreciate the initial efforts of Balu Nadiga of the Los Alamos National Laboratory toward interfacing the POP2 ocean model with DART. The CESM project is supported by NSF and the Office of Science (BER) of the U.S. Department of Energy. This work was partially supported by a NOAA Cooperative Agreement (Grant NA06OAR4310119) with the Geophysical Fluid Dynamics Laboratory (GFDL) and by the NOAA Climate Program Office under the Climate Variability and Predictability Program Grant NA09OAR4310163. We thank Anthony Rosati of GFDL for his ongoing support and encouragement of this effort. Computing resources were provided by the Climate Simulation Laboratory at NCAR’s Computational and Information Systems Laboratory (CISL), sponsored by the NSF and other agencies. Support for accessing data from the WOD09 was provided by CISL’s Data Support Section.

APPENDIX

Squared Misfit Test Statistic

We present a method of assessing whether the time-mean (bias) component of the squared misfits \( \bar{\varepsilon} = (\bar{\varepsilon} - \bar{\varepsilon})^2 \), defined in section 4c) in Assim and NoAssim relative to satellite-derived SST and SSH differ from one another in a statistically meaningful way. We use the difference in \( \bar{\varepsilon}^2 \) as a “test statistic.” A distribution of these test statistics can be computed under the hypothetical conditions that the time means of Assim and NoAssim are equal to each other and to the data, and time series of these fields differ only in the realizations of their variability. When differences in the actual squared error exceed 90% of those computed based on this hypothetical condition, we are 90% confident that they would not have occurred by chance alone.

At each point \( j \) on the POP2 grid, the signal can be decomposed as \( x_j = \bar{x}_m + \chi_j \), where \( x_j \) is the time-dependent SST or SSH, \( \bar{x}_m \) signifies a time average over each calendar month \( m \in \{ \text{January, February \ldots December} \} \) in a 6-yr period, and \( \chi_j \) is the remaining anomalous deviation from the mean. The time mean \( \bar{\chi} \) is simply the average of the annual cycle \( \bar{x}_m \). If we assume that the primary source of uncertainty in \( \bar{x}_m \) is sampling errors associated with variations in \( \chi_j \), then we can write a statistical model for \( \bar{x}_m \) as

\[
\bar{x}_m = \mu_m + \varepsilon, \quad \varepsilon \sim N(0, \Sigma),
\]

where \( \mu_m \) is the sample time mean for each calendar month computed from the observed SST and SSH. The random variable \( \varepsilon \) (which represents extraseasonal variations associated with \( \chi_j \)) is normally distributed with a space–time covariance \( \Sigma \). While it adds a level of complexity to model both spatial and temporal covariability associated with \( \chi_j \), neglecting to do so ignores our basic knowledge about the time and space scales associated with extraseasonal variability. Choosing this model creates a stricter test for significance.

To efficiently model the covariance \( \Sigma \), we use \( \chi_j \) at all grid points from the observed SST and SSH to do a singular value decomposition. This separates the space and time variability into orthogonal spatial patterns [empirical orthogonal functions (EOFs)] and orthogonal temporal signals [principal components (PCs)]. Retaining only the leading 20 singular vectors (containing over 80% of the variability), we model each of the PCs independently as a first-order autoregressive process. When
samples are needed, a draw from each of the 20 autoregressive process models associated with the PCs is made, and the EOFs are used to project these time samples into the model grid space.

The strategy for building an empirical distribution of differences in the time mean is to repeat the following sequence 1000 times. (i) Draw three space–time-correlated samples from e, which we call $s_0$, $s_1$, and $s_2$. (ii) Use (A1) to compute samples of $\boldsymbol{\tau}$. (iii) Compute two squared misfits ($e^2$) for the time-mean signal. One will be associated with the difference between $s_0$ and $s_1$ and one with the differences between $s_0$ and $s_2$. This is analogous to computing the bias component of the squared misfits associated with comparing Assim and NoAssim to the observed SST and SSH. (iv) Take the difference between these sampled squared misfits as the test statistic.

Using this procedure, we create 1000 realizations of the test statistic. The critical value at each grid point indicates a difference in time-mean-squared misfit that is exceeded by only 10% of the realizations. This bootstrap is equivalent to a one-tailed 90% significance test and is appropriate in this context because we are assuming that the direction of the difference (whether assimilation is improving or degrading the solution) is correct.

REFERENCES


