1. Introduction/background

To understand the ocean’s role in climate variability, its physical state must be accurately known, including temperature and salinity structures, circulation and property transports, and air–sea exchanges of momentum, heat, and freshwater. Hydrographic measurements are important, but even with global projects, such as the World Ocean Circulation Experiment, efforts to determine time-mean transports by combining one-time transects from different seasons and years are complicated by inconsistencies because of the time variability of the ocean (Ganachaud and Wunsch 2003; Wunsch 1978). More recent subsurface datasets, such as the Argo array of profiling floats, sample the spatial and temporal variability of temperature and salinity but have limited horizontal resolution and depth range. Satellite observations, on the other hand, provide good spatial and temporal coverage of surface parameters, such as sea surface height (SSH) and sea surface temperature (SST), but are not adequate to fully describe subsurface structures and their evolution in time. Existing estimates of air–sea exchanges are subject to substantial systematic errors and limited resolution. Currently, it is not practical to use data alone to fully characterize seasonal-to-decadal variability in the oceans.

Ocean state estimation synthesizes sparse ocean observations by using the dynamics of a numerical ocean circulation model. This process provides estimates of the time-varying ocean circulation with temporal and spatial coverage unavailable from observations alone. However, ocean processes take place on spatial scales ranging from centimeters to hundreds of kilometers and temporal scales ranging from seconds to years. The spatial resolution of syntheses is limited by computing power, with the consequence that small-scale processes must be parameterized. Before using such a synthesis for analysis, it is important to know its uncertainties and test its sensitivity with respect to the data used as constraints and the prior errors of those data.
The Estimating the Circulation and Climate of the Ocean (ECCO) consortium (see Stammer et al. 2002a) provides such dynamically consistent estimates of the time-varying ocean circulation by constraining the Massachusetts Institute of Technology (MIT) general ocean circulation model with most available ocean datasets. This model, which uses the adjoint approach to ocean state estimation, has been used to provide global syntheses of ocean data as long as 50 yr on a 1° grid (Stammer et al. 2002b, 2003, 2004; Wunsch and Heimbach 2007; Köhl et al. 2007). These results have been used to study various oceanic problems and to initialize coupled ocean–atmosphere circulation models (Dommenget and Stammer 2004; Pierce et al. 2004; Ponte et al. 2007).

The present work is an analysis of the sensitivity of the ECCO data assimilating system in a regional setting to changes in the type and weighting of assimilated datasets. This model, which uses the adjoint approach to ocean state estimation, has been used to provide global syntheses of ocean data as long as 50 yr on a 1° grid (Stammer et al. 2002b, 2003, 2004; Wunsch and Heimbach 2007; Köhl et al. 2007). These results have been used to study various oceanic problems and to initialize coupled ocean–atmosphere circulation models (Dommenget and Stammer 2004; Pierce et al. 2004; Ponte et al. 2007).

The present work is an analysis of the sensitivity of the ECCO data assimilating system in a regional setting to changes in the type and weighting of assimilated datasets. The assimilation procedure used by this system attempts to minimize a weighted model–data misfit. In this case, the model has 1° horizontal resolution, so mesoscale processes are unresolved. The subsurface data, however, are sparse pointwise measurements, which subsample these processes. The mismatch between processes observed in the data and those present in the model leads to so-called representation errors, which are accounted for, along with measurement errors, in the weights applied to the data. Ideally, full error covariances would be used for this purpose, but incomplete knowledge of such covariances and computational expenses are limiting factors. If the weights used for assimilation of these data are too high, the projection of aliased mesoscale signals onto the large-scale synthesis can lead to inaccurate solutions. Conversely, weights that are too low decrease the influence of the subsurface data and prevent the assimilation process from taking full advantage of this extensive dataset. An analysis of the weighting scheme is needed to determine if the assimilation’s use of available subsurface information is adequate. To this end, a model estimate assimilating only surface data is compared with one assimilating only subsurface data to determine the relative influence of each type of data. In addition, an analysis is performed with higher weights on the subsurface data to demonstrate its sensitivity to these weights and to determine the costs and benefits associated with changing them.

The structure of the paper is as follows; section 2 describes the model and the four experiments used in the analysis. Section 3 compares the results from each experiment with two of the assimilated datasets to demonstrate how consistent the solutions are with the observations. Section 4 assesses the results through a comparison with independent data. Section 5 compares our estimates with an existing global ECCO estimate, in terms of both the model–data misfits and the adjustments to control parameters. Finally, section 6 has discussion and conclusions.

2. Model description

The present analysis employs the ECCO system in a regional setting. The model and the adjoint assimilation approach are explained in detail by Stammer et al. (2002b) and Köhl et al. (2007); here, we present only a conceptual description. First, the model runs forward to simulate the time-varying ocean state. From this state, a weighted model–data misfit, called the cost function, is calculated. The adjoint model is then used to provide the gradient of the cost function with respect to the control variables. A standard descent algorithm uses these gradients to provide adjustments to these control parameters, which are defined here as surface forcing and initial conditions, to reduce the misfit while maintaining dynamical consistency. A new forward run is then performed by using the adjusted forcing and initial conditions, and a new cost function is calculated. The process is repeated iteratively until the cost function is minimized.

The model cost function provides a quantitative measure of the overall misfit of the model relative to the observations. It is calculated as follows:

$$ J = \sum (\text{model} - \text{data})^2 W_{\text{data}} + \sum (\text{control adjustments})^2 W_{\text{control}}, $$

where the first term is composed of the misfit between the simulated ocean state and the constraining data and the second term quantifies the changes to the control parameters. Each type of data and each control variable have a corresponding weight matrix determined from the inverse of the prior error. Minimizing the cost function is equivalent to finding the best dynamically consistent fit of the model to the data.

For the present work, the region of study is the Pacific Ocean north of 26°S (Fig. 1). The model has horizontal resolution of 1° in latitude and longitude and 50 vertical levels. The level thicknesses increase from 5 m at the surface to 500 m at depth. The model output spans the time period 1992–2004 and consists of daily SSH fields; monthly means of temperature, salinity, and zonal, meridional, and vertical velocity; and adjustments to the initial conditions and surface forcing as determined by the assimilation. Initial temperature and salinity fields are taken from the Levitus 1994 climatology (Levitus and Boyer 1994; Levitus et al. 1994), and the National...
Centers for Environmental Prediction (NCEP) reanalyses of heat flux, freshwater flux, and zonal and meridional wind stress are used for initial surface forcing (Kalnay et al. 1996).

The model domain is bounded by land to the north and to the east, but the southern and western boundaries are open to exchange with the South Pacific and with the Indian Ocean through the Indonesian Throughflow. These open boundaries are indicated in Fig. 1. Temperature, salinity, and zonal and meridional velocity obtained from a global ECCO state estimate [ECCO Global Ocean Data Assimilation Experiment (ECCO-GODAE); see Wunsch and Heimbach 2007] are prescribed on the boundaries. The ECCO-GODAE estimate available after 177 iterations was used. Although the horizontal resolution of the ECCO-GODAE estimate is the same as in our experiment, a significant difference is the increase in vertical resolution from 23 levels to 50. Linear interpolation was used to obtain 50-level boundary conditions from the global 23-level solution. Additionally, the weights used to obtain the ECCO-GODAE estimate differ from the weights used in any of the experiments described later. This estimate is used to provide boundary conditions because it spans the full time period from 1992 to 2004, and no synthesis for the full time period using any of the weighting schemes described later was available.

The weighting scheme used as a starting point for this analysis is the same as that used for a global ECCO run previously performed by Köhl et al. (2007). The weighting matrices are diagonal, with the inverse of the prior variances on the diagonal. Weights for the control variables of initial temperature and salinity are derived from a global error profile from the Levitus climatology. Heat, freshwater, and momentum fluxes, which are also control variables, have spatially varying weights defined as 1/3 the local standard deviation of NCEP values. Spatially varying weights for assimilation of sea surface height anomaly are derived from variability of altimetric height measurements. The same one-dimensional profile used to weight the initial temperature and salinity is applied to assimilation of subsurface measurements of temperature and salinity, including conductivity–temperature–depth sensors (CTDs), expendable bathythermographs.

FIG. 1. The region of study. The mean transport streamfunction in the top 800 m from the WEIGHTED experiment is shown. Stars indicate the positions of Station Aloha north of Hawaii and OSP in the northeast Pacific. Open circles indicate the location of open boundaries.
(XBTs), and Argo profiling floats. Separately, weights are assigned to the average drift of temperature, salinity, and vertical velocity between the first and final years of the state estimate and to the difference between a climatology calculated from the 13-year model estimate and the Levitus 1994 climatology. See Köhl et al. (2007) for further details on prior error fields.

Four assimilation experiments were performed. The first experiment, the STANDARD analysis, is constrained by all available data, including satellite measurements such as SSH from altimetry and surface winds from scatterometry; other surface measurements, such as Reynolds SST (Reynolds and Smith 1994) and sea surface salinity (SSS); and in situ measurements from XBTs, CTDs, drifters, and Argo profiling floats. Bin averaging of all observations onto the model grid removes some of the eddy noise in these data.

The second and third experiments explore the relative influence of the different datasets. One analysis (NOINSITU) assimilates only surface data of SSH, SST, SSS, and wind, whereas the other (NOSAT) assimilates only subsurface observations from XBTs, CTDs, and Argo profiling floats. Differences between STANDARD and NOINSITU show the influence of subsurface data on the solution, whereas differences between STANDARD and NOSAT show the influence of satellite observations. All weighting matrices are identical to the STANDARD experiment, and differences arise from the withholding of information.

The fourth analysis (WEIGHTED) is constrained by all available data but uses increased weights on subsurface data. There are several reasons for making this adjustment. First, recent research has provided a more accurate estimate of representation error of in situ measurements. Roemmich and Gilson (2009) used the Argo float array to estimate signal and noise variance (Fig. 2). The floats were divided into two randomly selected groups. For 2004–08, 60 monthly anomaly fields of temperature and salinity were estimated separately for each group by objective interpolation. These fields had 1° resolution, with mean and annual cycle removed. Steric height (0–2000 dbar) was calculated for each group, and the noise was estimated from the difference between either group’s estimate and the sample mean. Estimates of zonally averaged signal and noise variance of steric height are shown in Fig. 2, as well as a map of the noise-to-signal ratio. Zonally averaged noise peaks sharply in the Kuroshio region, where the signal peaks as well. The map of noise-to-signal ratio is relatively uniform, about 0.3 outside of the tropics. For a longer dataset, as in the present 12-year study, the signal and noise variances would increase. Although the historical dataset used in the present 12-year study is sparser than Argo, values for assimilation are only provided in those grid boxes that contain data. For those grid boxes containing data, the Argo calculation has some relevance. The STANDARD error bounds for temperature and salinity allow for variations as large as 17.2 cm in steric height through heating and freshening, corresponding to variance of nearly 300 cm², which is much larger than in Fig. 2, albeit for a longer time period. Additionally, the prescribed uncertainty used for hydrographic measurements is large relative to the SSH constraints, which range from 15.1 cm in the Kuroshio region to as low as 2.36 cm in the eastern Pacific at 125°W. Also, the use of purely diagonal covariances must be considered. Measurements of SSH anomaly, because they include corrections for tides and atmospheric conditions, are subject to large-scale coherent errors. As a result, the assumption that error covariances can be scaled with
signal variance is false (Cherniawsky and Sutherland 2008). Taken together, these results indicate that the weights on the subsurface data should be increased. Therefore, in the WEIGHTED experiment, the weights on the subsurface data are increased by a factor of 16 relative to the STANDARD case, which is equivalent to decreasing the prior error by a factor of 4. At 4.3 cm, the uncertainty in steric height from heating and freshening is significantly smaller. The weights applied to surface observations and to the surface fluxes remain the same as in the other three experiments.

After 50 iterations were completed for each of the four experiments, two corrections to the model input fields became available. The first was a small change to the time-invariant runoff field. After an initial adjustment, this correction has minimal impact on the solution. The second was a corrected XBT dataset, which was necessitated by the recent discovery of a time-varying warm bias in the XBT archive (Gouretski and Koltermann 2007). This error is attributable to time-varying fall rate in XBTs, and the fall-rate correction has been estimated by Wijffels et al. (2008), based on consistency of XBT profiles with altimetric height data. The corrected XBT dataset was provided by J. K. Willis (2008, personal communication). This change had substantial effects on the solution, as described later. It was impractical to restart the calculations from iteration 1 because of the computational burden; moreover, the replacement of the subsurface dataset with one that is more consistent with altimetric height provides an instructive example. It is crucial to ensure consistency between different satellite and in situ data types, because the assimilation makes no adjustments other than the forcing fields and initial conditions. The NOINSITU analysis, which was only adjusting to the changes in runoff, needed only 5 further iterations to reach a point where the cost function was no longer decreasing significantly. The STANDARD and NOSAT estimates needed 10 further iterations, and the WEIGHTED solution, which adjusts most strongly to the subsurface data, needed 20 further iterations.

In practice, it is impossible to be certain that the cost has reached a global minimum in a nonlinear problem such as this one. Computational constraints play a role in determining how many iterations are completed. In this analysis, one iteration required 3.5 days of computer time. After 50 iterations, the total cost functions of each of the four experiments were decreasing by less than 0.1% of their initial value with each subsequent iteration. For these reasons, the optimization was stopped at this point, although more iterations were performed as previously described once the corrections were made. This solution could be a local minimum rather than the global minimum, which is the ultimate goal of the optimization. The existence of temperature drifts and other issues suggest that the solutions are not fully converged. However, the distinctions between the solutions discussed later became evident early in the optimization process and appear robust. This was considered to be adequate for the sensitivity study performed here.

To demonstrate the trade-off in cost between decreasing model-data misfit and increasing adjustments to the control variables and to indicate some of the differences between the model results, the evolution of several components of the cost function is shown in Fig. 3. Discontinuities seen at iteration 50 are a result of the aforementioned corrections. Total cost (Fig. 3a) includes the sum of the squares of the misfits between each of the datasets and the model estimate multiplied by the weights. All misfits are included in this cost calculation, regardless of whether the data are assimilated into the specific estimate (e.g., the NOSAT cost includes misfits relative to satellite data). The total cost decreases by the largest percentage in the WEIGHTED estimate, to around one-tenth of its initial value. Total costs of the STANDARD and NOINSITU estimates are nearly identical, indicating that the subsurface data, at the weights used in the STANDARD estimate, do not provide much improvement to the model-data misfit. The NOSAT estimate, however, is considerably degraded, indicating the importance of surface data over subsurface data in the estimates calculated here. Figures 3b,c,d break down the total cost, showing contributions from XBT misfits, SSH misfits, and heat flux adjustments, respectively, divided by the number of data points in the dataset. Because the weights are calculated from the inverse of the prior variance, the average cost is the average ratio of misfit variance to prior variance and should converge to unity.

In these experiments, the squared difference between model estimates of temperature and XBT observations decreased from more than 12 to between 2 and 6 times the prior variance of XBT data (Fig. 3b). The decrease in XBT cost is largest in the WEIGHTED experiment, which has the highest penalty associated with this misfit, and smallest in the NOINSITU experiment, which does not assimilate these data. The XBT cost contributions are nearly identical in the NOSAT and STANDARD experiments, indicating that the assimilation of surface data has little effect on the estimation of subsurface temperature fields. One might expect the NOSAT experiment to match the XBT data more closely than it does, because there are no competing misfits from surface data to balance it. However, in NOSAT, the relative cost of adjusting the control variables is too high to justify the changes necessary for a reduced misfit. In the WEIGHTED experiment, higher weights on the XBTs
result in higher costs from the model–data misfit, such that the relative cost of adjustments is small, so large adjustments are made and a smaller model–data misfit is attained.

A similar effect is noted in the SSH cost component (Fig. 3c). The largest SSH misfit occurs in the NOSAT estimate, which does not assimilate those data. The cost of SSH misfit in the NOINSITU experiment is almost identical to the STANDARD cost, indicating that the estimation of surface fields is not affected by the additional assimilation of subsurface data in the STANDARD experiment. As in the case of the NOSAT experiment with XBT data, the cost of adjusting the control variables to match the SSH data more closely is higher than the reduction in misfit that would result from such changes. The misfit between the WEIGHTED model estimate and the measured SSH is slightly larger than that in the STANDARD run, indicating that the improved match to the XBT data in the WEIGHTED experiment comes at the expense of the SSH model–data misfit.

The costs of the adjustments made to the heat flux are shown in Fig. 3d. Although misfits to XBTs and SSH decrease as iterations progress, adjustments to control variables and their associated costs increase. For the first 50 iterations, adjustments relative to NCEP surface fluxes determine the cost. At iteration 50, as mentioned previously, the experiments were restarted using the adjusted control variables determined by the model. The adjustments after this point are calculated relative to this new forcing rather than the initial NCEP forcing. The largest adjustments occur in the WEIGHTED
experiment, indicating the changes necessary for increased accuracy in the fit to the subsurface data. Although much larger than the adjustments of the other three model estimates, the average cost of the WEIGHTED adjustment is still below 0.2, indicating that the magnitude of the changes is, on average, less than 20% of the prior variance. Therefore, these adjustments are well within the expected limits. Although only the heat flux adjustments are shown in Fig. 3d, these results are consistent for adjustments to the other forcing fields. The adjustments to the initial conditions are much smaller and are not discussed further in this analysis.

3. Comparison with assimilated data

Comparisons between the state estimates and the constraining data are informative in several ways. First of all, differences indicate the extent to which the estimates are consistent with observations. The term “consistent” here implies agreement between the model estimate and the data, within the prior assumptions about uncertainties associated with both the observations and the model estimate. Second, because the experiments assimilate different subsets of the data, the differences between the estimates demonstrate the influence of each type of data. Therefore, we compare the estimates to the XBT temperature profiles, which compose the largest subsurface dataset, and to the satellite altimetric height dataset, the dynamically important surface dataset.

a. Heat content/SST

For this time period, the XBT temperature data are the subsurface dataset with the best spatial and temporal coverage. Even so, its coverage is inhomogeneous, as seen in Fig. 4, which shows the depth- and time-integrated number of measurements for each location. To account for the sparse nature of the data, at each time step, the heat content was computed from the model output only at the positions where XBT data are available. Maps of root-mean-square difference (RMSD) between data- and model-estimated heat content (mean removed) are shown in Fig. 5. The RMSD between the STANDARD model estimate and the data (Fig. 5a) shows a region of high RMSD off the coast of Japan, where the Kuroshio separates from the coast and meanders eastward toward the Kuroshio Extension region. This is a highly variable region with intense mesoscale activity, which is absent in the model and is aliased by sparse observations, so significant discrepancies between the data and model estimates are expected. Figures 5b,c, which show the difference between the STANDARD RMSD (Fig. 5a) and the NOINSITU and NOSAT RMSDs, respectively, indicate that both the NOINSITU and the NOSAT experiments have higher
RMSDs than the STANDARD experiment in this region. The difference between WEIGHTED and STANDARD RMSDs (Fig. 5d) indicates that the WEIGHTED estimate is more consistent with the observed heat content than the STANDARD experiment. Overall, the NOINSITU RMSDs are higher than any of the other model estimates, particularly in the Kuroshio region and the northeastern Pacific. This is expected because NOINSITU is not constrained by these data. The NOSAT estimate matches the data more closely than the STANDARD estimate in some regions but has higher RMSDs in most of the North Pacific, indicating that the inclusion of surface data in the analysis generally increases the consistency of the estimate with observations of subsurface quantities.

A time series of heat content anomaly in the upper 750 m of the ocean, which was calculated from optimally interpolated XBT data (Willis et al. 2004, hereafter WRC), is also available for comparison with the model estimate. This product has been corrected for the time-varying XBT bias previously mentioned. There are several issues to be considered when comparing this product with the model results. First, the WRC product is a result of statistical interpolation of a sparse dataset, such that differences between the ECCO analyses and the product could be a result of inconsistencies in data-poor regions. Another concern is a large secular increase in heat content found in the ECCO estimates. Although heat content is thought to increase over the period 1992–2004 (Levitus et al. 2005), the drift in the model estimates implies a heat uptake that is larger than expected. Additionally, the increase noted in Levitus et al. (2005) could be exaggerated because of the previously noted bias in XBT measurements. In their estimate, Köhl et al. (2007) found (after 69 iterations) an increase in heat content corresponding to a global net ocean heat uptake of 1.1 W m$^{-2}$, which is slightly larger than indicated by Levitus et al. (2005). In the region of interest, the Pacific Ocean north of 26°S, their average heat uptake was 1.5 W m$^{-2}$. The average heat uptakes for the WRC product and for each of the four estimates are found in Table 1. The NOSAT estimate’s uptake of only 1.49 W m$^{-2}$ is the smallest among the experiments. The largest uptake of 4.45 W m$^{-2}$ is found in the NOINSITU estimate. The WEIGHTED estimate has a heat uptake of 1.96 W m$^{-2}$, whereas the STANDARD estimate is slightly higher at 2.52 W m$^{-2}$. The WRC product has an average heat uptake of only 0.77 W m$^{-2}$.

As mentioned, these solutions are not fully converged and the drift in temperature is, in part, an artifact of that
lack of convergence. Despite this, the difference in heat uptakes illustrates the trade-offs involved in assimilating different and sometimes conflicting data. The observed SSH data show an increase of 0.20 cm yr$^{-1}$ in this region. Without assimilation, the estimate of SSH shows an increase of 0.10 cm yr$^{-1}$. In response to this misfit, the NOINSITU estimate matches the SSH increase (0.19 cm yr$^{-1}$) with increased heat content and thus dynamic height through increased surface heat flux. The resulting temperature misfit does not increase the cost in the NOINSITU estimate. The cost of the NOSAT estimate, conversely, is not affected by SSH misfit, so its solution has a low SSH trend of 0.10 cm yr$^{-1}$, but its heat uptake is closer to observations than the other three estimates. The STANDARD and WEIGHTED estimates represent the middle ground between these extremes, and both have an increase of SSH of 0.11 cm yr$^{-1}$.

To determine if estimated variability is correlated with that in the WRC product, a linear trend is removed from each estimate. Figure 6 shows detrended estimated heat content, which is smoothed over one year. Correlation coefficients between model estimates and the WRC product are listed in Table 1. The WEIGHTED experiment has the highest correlation coefficient at 0.75. The STANDARD estimate is lower at 0.51, indicating that higher weights on subsurface data result in a representation of subsurface variability more consistent with the data. Neither the NOSAT nor the NOINSITU estimate has a correlation that is statistically significant, indicating the importance of both surface and subsurface information when estimating an integrated quantity such as heat content.

b. Sea surface height

SSH has been measured continuously by satellites since 1992. Gridded along-track SSH anomalies from the Ocean Topography Experiment (TOPEX)/Poseidon and Jason satellites are assimilated by the STANDARD, WEIGHTED, and NOINSITU experiments. Using a method similar to that used for heat content to account for inhomogeneously distributed data, maps of the RMSD between observed and model-estimated SSH anomalies are created (Fig. 7). Figure 7a shows the RMSD between STANDARD SSH anomaly and the altimetric height data, and Figs. 7b–d show the difference between the STANDARD RMSD and the NOINSITU, NOSAT, and WEIGHTED RMSDs, respectively. As with heat content, RMSD is high in the Kuroshio Extension region, an area with intense mesoscale activity. The differences between NOINSITU and STANDARD (Fig. 7b) are small in magnitude and evenly split between positive and negative. This indicates that the skill of the two experiments in estimating SSH is about equal, implying that the effect of assimilating subsurface data on the estimation of SSH with the weights used in the STANDARD experiment is minimal. The NOSAT experiment is much less consistent with data. This is logical, because the SSH observations are not assimilated by the NOSAT analysis. However, the differences between the NOSAT estimate and the satellite data are more significant and more pervasive than the differences between the NOINSITU estimate and the heat content (Fig. 5b). In other words, the NOINSITU estimate of heat content is only slightly worse than the NOSAT and WEIGHTED estimates of heat content, whereas the NOSAT estimate of sea surface height is significantly worse than the other estimates of sea surface height. This implies that, for overall consistency, assimilating the satellite data is more important than assimilating the subsurface data. The WEIGHTED RMSD is higher than the STANDARD RMSD in most regions, implying an inconsistency between subsurface and SSH observations, because the strongly increased weights are somewhat detrimental to the objective of minimizing SSH misfit. Because

| Table 1. Basin-averaged heat uptake, correlation coefficients between detrended model estimates of heat content and detrended WRC, and 95% confidence intervals for the correlations. |
|---------------------------------|-------------------|-----------------|-----------------|
| Net heat uptake (W m$^{-2}$)    | Correlation with WRC 95% | Confidence intervals |
| WRC product                    | 0.77              | —               | —               |
| STANDARD                       | 2.52              | 0.51            | 0.38–0.62       |
| NOSAT                          | 1.49              | 0.31            | 0.15–0.45       |
| NOINSITU                       | 4.45              | −0.10           | −0.26 to 0.07   |
| WEIGHTED                       | 1.96              | 0.75            | 0.66–0.81       |

Fig. 6. Time series of heat content anomaly as a function of time from an optimally interpolated XBT dataset and from each of the four estimates. All time series were detrended.
both the satellite and subsurface observations subsample mesoscale features, this could indicate that the WEIGHTED analysis is “matching” subsurface noise, whereas the STANDARD estimate is matching surface noise. Even so, the WEIGHTED estimate of SSH anomalies is more consistent with the observations than the NOSAT experiment (Fig. 7c), reinforcing the importance of using all available datasets.

Empirical orthogonal functions (EOFs) provide a metric for comparison of large-scale variability. EOFs represent the decomposition of a time series into orthogonal modes representing the most energetic patterns of variability (Lorenz 1956). Each EOF has two parts: the spatial pattern of variability and the associated amplitude time series. Monthly estimates of SSH anomalies were smoothed over 12 months to remove the annual cycle as well as other higher-frequency variability. The first EOFs of SSH for the data and each model estimate are shown in Fig. 8. The dominant feature of the EOF of the data is a tropical signal, with a high region east of 160°W and a low area in the west, between the equator and 20°N. There is also a smaller-amplitude feature in the North Pacific with low values in the western and central North Pacific and high values in the northern and eastern coastal regions. This is similar to the pattern known as the Pacific decadal oscillation (Mantua et al. 1997). The STANDARD, NOINSITU, and WEIGHTED estimates (Figs. 8a,b,d, respectively) are very similar, showing the tropical signal with approximately the right magnitude and a slightly weak subtropical signal. The NOSAT estimate (Fig. 8c) differs most from the data, with a less intense equatorial signal that extends farther south than in the other estimates. This indicates that not assimilating SSH can lead to a result that is inconsistent with a known pattern of variability.

In the amplitude time series associated with these EOFs (Fig. 8f), the most significant feature, replicated in all estimates, is a peak just before the end of 1997, which is associated with a strong El Niño event that occurred in 1997–98. This, together with the tropical location of the largest spatial feature (Fig. 8e), indicates that the most energetic mode of interannual variability in SSH is associated with the El Niño oscillation. All time series in Fig. 8f are highly correlated (correlation coefficients >0.9).

4. Comparison with independent data

This section compares model estimates with two time series not used as constraints. These data are independent
of the experiments and thus provide a more stringent assessment of the model estimate’s skill.

a. Station Aloha

The Hawaii Ocean Time series (HOT) consists of hydrographic measurements at Ocean Station Aloha, which is situated north of Oahu, Hawaii, at 22°45′N, 158°00′W (Karl and Lukas 1996). Temperature and salinity have been measured at this location approximately once per month since October 1988. Vertical resolution of the data is 2 m, but data are interpolated onto the model’s 50-level vertical grid for these comparisons. Differences resulting from aliasing effects are expected when comparing data taken at one point in time and space to monthly means averaged over model grid boxes, but Stammer et al. (2008) have shown that aliasing resulting from mesoscale structure, although present, does not dominate the comparison between an ECCO estimate and the observations at this location. The similarities and differences will give some perspective on the consistency of the present ECCO analyses with the data.

The time evolution of estimated minus observed temperature at Station Aloha is shown in Figs. 9a–d for each of the four experiments. The temperature data from Station Aloha are shown in Fig. 9e for reference. In all cases, the estimated surface temperature is higher than observed. This is related to problems with the model physics in simulating the wind-induced mixed layer deepening in the summer and the associated cooling of SST (Köhl et al. 2007). The overestimation is weakest in the WEIGHTED time series (Fig. 9d), whereas the NOSAT estimate (Fig. 9c) has the highest surface temperatures. This disparity in surface temperatures is the
strongest during summer, when estimated stratification is stronger than observed. The strong summer stratification is confined to a shallower depth in the model estimates than in the data. As a result, temperatures at depths greater than 100 m are too low. This difference is magnified during the fall, when summer stratification breaks down earlier in the estimates than in the data, with the result that, in all analyses, the estimated subsurface temperatures are as much as 3°C cooler than observed. The difference in thermocline structure results in cooler temperatures in the model estimates to a depth of almost 500 m. From about 1992 to 1997, the largest differences are centered at 200–250 m, whereas from 1998 to 2004 they are at a shallower depth of about 150 m. The subsurface bias has similar magnitude in all experiments. The basic structure of overestimated surface temperatures, a too-shallow thermocline, and underestimated subsurface temperatures persists throughout the time series and is consistent in all model estimates. The similarity of the structures of the differences points to an underlying problem in model physics, which cannot and should not be overcome through alteration of the control variables. Despite this, there is still merit in looking at the consistency of the variability of the estimates with the data.

The time evolution of corresponding salinity differences, model estimate minus data, is shown in Fig. 10. Figure 10e shows that subsurface salinity at Station Aloha has a more complicated structure than temperature, increasing to a subsurface maximum and then decreasing with depth. However, the structure of the differences is similar to temperature. Near the surface, to about 100 m, estimated salinity is higher than observed in all except the WEIGHTED analysis. The subsurface salinity maximum is shallower in all model estimates than in the data. Below this point, estimated salinity is consistently lower than measured salinity, particularly from about 200 to 300 m. The magnitude of this structure changes with time. During 1998, the difference between estimated and measured salinity in the top 100 m increases markedly. Observed salinity increases at this time, but salinity simulated by all four experiments increases even more. After this shift, in the STANDARD, NOINSITU, and NOSAT estimates, the
positive surface bias dominates the difference figures and the bias in the halocline weakens. In the subsurface region, the bias gets weaker in these three estimates. In the WEIGHTED estimate, surface salinity is close to observations, but the subsurface bias strengthens. This indicates a decrease in subsurface salinity in the WEIGHTED estimate relative to the observations, whereas the other three estimates increase relative to observations. Overall, the structure of the differences is similar in all four estimates, reinforcing the concept of a shortcoming in model physics common to all four analyses.

The mean and RMS differences between estimated and measured temperature as a function of depth are shown in Figs. 11a,c, respectively. Figure 11a shows that, although all solutions overestimate surface temperature, the WEIGHTED estimate is, on average, the closest. However, within 50 m of the surface, all four model estimates become cooler than the data. The WEIGHTED estimate has the smallest bias to about 150 m; below that, the NOINSITU estimate is the least biased, but all estimates are similar. With the mean removed, the RMS differences between the data and the WEIGHTED model estimate are smaller than those of the other estimates for the full depth range shown (Fig. 11c). At the surface, the NOSAT estimate has the highest RMS difference; however, below 200 m, the NOINSITU estimate is the least consistent with the observations. This indicates that, when the bias is removed, the time evolution of temperature structure in the WEIGHTED experiment is most consistent with observations. The prior errors from which the weights are derived are also shown in Fig. 11c. Near the surface, the RMS difference between the WEIGHTED estimate and the observations is smaller than the prior errors; however, below about 50-m depth, the prior error is significantly lower than any of the RMS differences. Additionally, the structure of the RMS differences, with a maximum at 300 m, indicates that the vertical structure of the prior error is not entirely appropriate for this region. This is a result of using a one-dimensional error profile globally. The subsurface maximum, found in the region of the thermocline, is located at different depths in different locations and disappears when global smoothing is applied. This is another argument in favor of a spatially varying error field.

Mean and RMS differences between estimated and observed salinity as a function of depth are shown in
Figs. 11b,d, respectively. The mean difference is the smallest in the WEIGHTED model estimate at the surface (Fig. 11b), where the other three solutions overestimate salinity by more than 0.2 psu. All four experiments underestimate the salinity at around 200-m depth by about 0.2 psu and overestimate salinity below about 300 m. Figure 11d shows that, in the top 200 m, the STANDARD has slightly smaller RMS difference than the other three, but the WEIGHTED estimate has the smallest RMS difference from about 200 to 500 m. This confirms that, in addition to being the least biased, the time evolution of the WEIGHTED experiment is as consistent with observations as the STANDARD experiment and more consistent than the other two experiments. As for temperature, the prior error is shown. None of the four estimates has an RMS differences at or below the value of the prior error in the top 600 m. For much of the depth range, the prior error is significantly smaller than the RMS differences.

Dynamic height provides an integrated, quantitative comparison between the four ECCO analyses and the data. For this comparison, the estimated temperature and salinity are subsampled to times when Aloha data are available. These time series are then interpolated to all months and smoothed with a 12-month running mean to remove high-frequency variability. Dynamic height is calculated relative to 235 m to emphasize the surface properties and then relative to 935 m to include the full thermocline. In Fig. 12a, the effects of the increased salinity near the surface in the STANDARD, NOINSITU, and NOSAT estimates are manifested in decreased dynamic height in the second half of the time series.
leaving those three estimates biased low, whereas WEIGHTED still matches relatively well. The correlations are 0.40, 0.49, 0.58 and 0.65, for the NOINSITU, NOSAT, STANDARD, and WEIGHTED estimates, respectively. Relative to 935 m, all four estimates are clearly biased low, as shown in Fig. 12b. This results from the low temperatures shown in Fig. 9a–d. Some of the interannual variability is still captured, and the correlations are 0.43, 0.41, 0.48 and 0.41 for the NOINSITU, NOSAT, STANDARD, and WEIGHTED estimates, respectively. The WEIGHTED correlation is no better than the others for this depth, but the lower bias, particularly in the first half of the time series, still indicates that this estimate’s subsurface structure is the most consistent with observations.

b. Ocean Station Papa

Another independent time series has been recorded in the northeastern Pacific Ocean at 50°N, 145°W (Whitney and Freeland 1999). This station, known as Ocean Station Papa (OSP), was originally established in 1949. During the time period spanned by the ECCO analyses, measurements were taken approximately 3–6 times per year. Although this is less frequent than Station Aloha, there are still a significant number of profiles available for comparison with the assimilation output. As with Station Aloha, the OSP station data are interpolated onto the lower-resolution model vertical grid for comparison.

The time evolution of estimated minus observed temperatures at OSP is shown in Figs. 13a–d for each of the model estimates. All four solutions underestimate the surface temperature throughout the time series, sometimes by as much as 3°C. Starting in about 1996, estimated temperatures between 100- and 300-m depths are significantly warmer than observations. Figure 13e shows that observed temperatures did not change significantly, indicating that these signals come from a warming in the model estimates. The warming deepens to almost 500 m by the end of the time series, reaching magnitudes of up to 2°C. The structure of the differences is the same in all four estimates, with the main difference being that both the underestimation at the surface and the subsurface warming have distinctly smaller magnitudes in the WEIGHTED estimate. This temperature
drift is another indication that the analyses are not fully converged.

The time evolution of estimated minus observed salinity at OSP is shown in Figs. 14a–d. Figure 14e shows that observed salinity structure at OSP is simpler than at Station Aloha, with minimum values at the surface, a sharp halocline near 150 m, and increasing salinity below that point. The main feature of the difference plots is a band of much higher estimated than measured salinity at about 100-m depth. This indicates that the halocline is shallower in the model estimates than in the data. Resulting differences exceed 0.6 psu at times. Differences at both shallower and deeper levels are of a much smaller magnitude. This systematic bias in all four estimates suggests that a problem with mixing is leading to a misplaced halocline. As noted at Station Aloha, changes to the control variables will not overcome deficiencies in the physics of the analysis. However, the variability of the solutions and the magnitude of the discrepancies are still instructive.

Comparisons between estimated and measured temperature and salinity as a function of depth at OSP are shown in Fig. 15. In temperature (Fig. 15a), in the top 100 m, the WEIGHTED solution deviates from the mean observations by less than the other three estimates. Below 100 m, the NOINSITU estimate is closest to the data, whereas the other three estimates have similar structures. Figure 15c shows that the RMS difference between estimated and observed temperature is smallest for the WEIGHTED analysis throughout the depth range. In the top 100 m, the NOSAT estimate is the furthest from the data; deeper, the NOINSITU estimate has the highest RMS difference. Additionally, the prior error is shown for comparison, and the WEIGHTED RMS is the only one smaller than the prior error at any depth shallower than 600 m. As with Station Aloha, there is a subsurface peak in RMS, but at this higher latitude that peak is shallower, at 200 m. In salinity, Fig. 15b demonstrates again that, in the mean, the most significant difference is the incorrect halocline structure. The effect is slightly smaller in the WEIGHTED estimate, a result balanced by a slight overestimation of salinity just below the halocline, when all three of the other estimates match the mean salinity closely. Figure 15d
shows that, in the top 150 m, the WEIGHTED estimate has distinctly smaller RMS differences in salinity than the other three. At the surface, the WEIGHTED estimate is slightly smaller than the prior error. Below 150 m, the four estimates are nearly identical; below 300 m, all are slightly smaller than the prior error.

Dynamic height provides a quantitative comparison between the data and the ECCO analyses. For dynamic height relative to 135 m (Fig. 12c), all four analyses have statistically significant correlations of 0.81, 0.70, 0.65, and 0.36 for WEIGHTED, NOINSITU, NOSAT, and STANDARD, respectively. The NOSAT estimate shows a distinct low bias, probably because of distinctly cooler water in the surface layer. Relative to 935 m (Fig. 12d), however, the estimates of dynamic height from NOINSITU, STANDARD, and NOSAT all increase consistently, whereas observed dynamic height decreases throughout the time series. The increase in the estimates can be attributed to the deep warming trend evident in Fig. 13. This is a reflection of model drift. To limit the influence of the drift on the comparison of estimated to observed variability, all time series are detrended. Detrended, NOINSITU, STANDARD, and NOSAT do not have a correlation with the observed dynamic height that is statistically significant at the 95% level. The detrended WEIGHTED estimate has a correlation of 0.38 with the detrended OSP dynamic height, demonstrating the best consistency with observed interannual variability at this location.

5. Comparison with global estimate

Of the four experiments described in this work, the WEIGHTED estimate is most consistent with assimilated data as well as independent data. However, other global estimates exist, such as the ECCO-GODAE solution used for boundary conditions. It is important to determine how the present analysis compares with these other solutions.

a. Model–data misfits

One metric for comparison with the ECCO-GODAE analysis is the cost. Table 2 lists average cost contributions of each type of data for the STANDARD, WEIGHTED, and ECCO-GODAE analyses. As in Fig. 3, in a “perfect” solution, the model–data misfit
would be the same as the prior error of the variable and the average cost would be equal to unity. The calculation is made by using the STANDARD weights for all three estimates.

For almost all data, both STANDARD and WEIGHTED average costs are lower than those from the ECCO-GODAE analysis. There are several reasons for this. First, these estimates benefit from increased vertical resolution. Additionally, for regional estimates such as these, the assimilation will focus on locally important features. In a global estimate, inaccuracies in the Antarctic Circumpolar Current might dominate the assimilation as a result of the magnitude of the signals in that region, whereas misfits in the Kuroshio would be small by comparison. In the North Pacific region, on the other hand, the Kuroshio itself is a major feature, and its accurate representation is an important component of an optimal solution. However, the most significant difference between the two estimates, as demonstrated by the difference between STANDARD and WEIGHTED average costs, is the increased weight on the subsurface data in the WEIGHTED experiment. It is clear that this change has decreased model–data misfits for most data. The cost of SSH is higher in the WEIGHTED estimate.

### Table 2. Average costs of the assimilated datasets for the STANDARD, WEIGHTED, and ECCO-GODAE estimate.

<table>
<thead>
<tr>
<th></th>
<th>STANDARD (iteration 60)</th>
<th>WEIGHTED (iteration 70)</th>
<th>ECCO-GODAE (iteration 177)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTD $T$</td>
<td>2.52</td>
<td>1.95</td>
<td>3.25</td>
</tr>
<tr>
<td>CTD $S$</td>
<td>1.45</td>
<td>1.03</td>
<td>1.65</td>
</tr>
<tr>
<td>Argo $T$</td>
<td>3.76</td>
<td>2.48</td>
<td>4.26</td>
</tr>
<tr>
<td>Argo $S$</td>
<td>2.52</td>
<td>1.54</td>
<td>1.99</td>
</tr>
<tr>
<td>XBT</td>
<td>3.60</td>
<td>3.25</td>
<td>4.26</td>
</tr>
<tr>
<td>SSH (altimetric height)</td>
<td>1.29</td>
<td>2.35</td>
<td>1.32</td>
</tr>
<tr>
<td>SST (Reynolds)</td>
<td>2.66</td>
<td>1.01</td>
<td>2.01</td>
</tr>
<tr>
<td>SSS</td>
<td>1.78</td>
<td>1.75</td>
<td>2.40</td>
</tr>
<tr>
<td>SST (TMI)*</td>
<td>2.49</td>
<td>0.96</td>
<td>1.89</td>
</tr>
</tbody>
</table>

* TMI = Tropical Rainfall Measuring Mission (TRMM) Microwave Imager.
indicating a larger misfit, but the increase is minor relative to the decreases in all other cost components.

Maps of RMSD of heat content anomalies and SSH anomalies provide another comparison between the WEIGHTED and ECCO-GODAE estimates (Fig. 16). The RMSDs between XBT-derived heat content anomaly and model-estimated heat content anomaly for the ECCO-GODAE and WEIGHTED estimates are shown in Figs. 16a,c, respectively. The difference, ECCO-GODAE minus WEIGHTED (Fig. 16e), reveals that, over much of the North Pacific, the WEIGHTED RMSD is lower than that of the ECCO-GODAE estimate. This suggests that the WEIGHTED experiment is more consistent with the observed heat content. The RMSDs between observed and estimated SSH anomaly for ECCO-GODAE and WEIGHTED are shown in Figs. 16b,d, respectively. In contrast with heat content, the negative values dominating the difference map (ECCO-GODAE minus WEIGHTED; Fig. 16f) indicate that WEIGHTED estimates of SSH anomaly are less consistent with the observations than ECCO-GODAE estimates almost everywhere. This confirms what was suggested in Table 2 by the higher average cost of SSH.

b. **Surface forcing adjustments**

The WEIGHTED state estimate has a smaller model-data misfit than the ECCO-GODAE state estimate. This has been achieved in a dynamically consistent way by adjusting the fluxes of heat, freshwater, and momentum imparted by the surface forcing. Because all analyses were initially forced by NCEP reanalysis fields, their adjustments can be compared directly. Although
the adjustments are made at two-day intervals, only monthly means of forcing are available for the ECCO-GODAE solution, so the comparisons are made between monthly adjustments.

The average costs of each of the four forcing terms are listed in Table 3 for STANDARD, WEIGHTED, and ECCO-GODAE estimates. As in Table 2, the calculation is made using the STANDARD weights for all three estimates. The calculation for the ECCO-GODAE estimate is restricted to the North Pacific region. In this case, all results are significantly less than one, indicating that the adjustments made to the forcing terms are, on average, much smaller than the prior errors.

Average costs of the adjustments made to zonal and meridional wind stress by the WEIGHTED estimate are larger in magnitude than those for the ECCO-GODAE estimates. Adjustments in the STANDARD estimate are smaller than either of the other two. The spatial structures of the mean adjustments to the wind stress are shown in Fig. 17. The adjustments to zonal wind stress show similar patterns of strengthening of the major wind systems: the trades near the equator and the westerlies in the northern Pacific. This result is supported by prior research indicating that the wind fields of the NCEP reanalysis are biased low relative to shipboard wind measurements (Smith et al. 2001). The adjustments in the WEIGHTED estimate have a larger magnitude. In meridional wind stress, the solutions are distinctly different. The WEIGHTED estimate has an increase in northward wind stress near the equator on the east side of the basin, whereas the ECCO-GODAE estimate has a much weaker increase at that location, countered by a negative signal farther east. In the ECCO-GODAE estimate, the dominant feature is a decrease in meridional wind stress in the Bering Sea, whereas in the WEIGHTED estimate there is a strong positive adjustment just east of the Okhotsk Sea. Adjustments in the eastern part of the basin near the equator are also of opposite signs. The magnitude of the adjustments is generally larger in the WEIGHTED estimate; however, as previously discussed, these adjustments are still generally within the uncertainty of the measurements. The STANDARD estimate adjustments have the same structure as the WEIGHTED estimates, with much smaller magnitudes.

Adjustments to the heat flux have similar costs in the ECCO-GODAE and WEIGHTED estimates; STANDARD costs are significantly smaller. All estimates require a net decrease in heat flux that is equivalent to an increase in heat absorption by the North Pacific. In Fig. 18c, the WEIGHTED estimate has some of the same features seen in the ECCO-GODAE estimate (Fig. 18a), particularly a decrease in surface flux (more heat into the ocean) in the western North Pacific just south of the Kuroshio and an increase in surface flux (more heat into the atmosphere) north of this region. This dipole feature mirrors a similar structure in zonal wind stress (Fig. 17a,c,e). This could indicate a problem with the model physics in the location or intensity of the Kuroshio. The relatively low 1° resolution cannot accurately represent the high-intensity jet structure of the western boundary current, and this may well be reflected in the adjustment to the surface forcing in this region. These features are much stronger for the WEIGHTED estimate than in the ECCO-GODAE estimate, which could be a result of the regional model’s focus on the Kuroshio, or could be another indication of the increased subsurface weights or vertical resolution. The strongest feature in the ECCO-GODAE mean heat flux adjustment is the region of strong increase in heat flux (less heat into the ocean) just south of the equator in the eastern Pacific. This feature is completely absent from the WEIGHTED estimate. Because thermocline variations are important in this region, this difference could be related to the difference in vertical resolution between the two estimates. Further analysis is necessary to understand this difference. The spatial structures of the adjustments from the STANDARD analysis are again very similar to those in the WEIGHTED analysis.

The disparities between the adjustments in freshwater flux are significant. The cost of freshwater adjustments in the ECCO-GODAE estimate are an order of magnitude larger than those in the WEIGHTED estimate, which are themselves an order of magnitude larger than the STANDARD estimate costs. All estimates require a net flux of freshwater into the North Pacific (precipitation and/or runoff; Figs. 18b,d,f). Spatially, some features are similar, such as an increase in evaporation (negative) south of the equator and an increase in precipitation (positive) north of the equator. There are, however, distinct differences. The ECCO-GODAE estimate indicates a strong decrease in precipitation in the North Pacific near 45°N. A similar feature is much weaker in the WEIGHTED estimate. Increased precipitation in the tropics in the west is seen in the ECCO-GODAE adjustments but absent in the WEIGHTED

<table>
<thead>
<tr>
<th></th>
<th>STANDARD</th>
<th>WEIGHTED</th>
<th>ECCO-GODAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zonal wind stress</td>
<td>0.0410</td>
<td>0.144</td>
<td>0.0606</td>
</tr>
<tr>
<td>Meridional wind stress</td>
<td>0.0154</td>
<td>0.0764</td>
<td>0.0278</td>
</tr>
<tr>
<td>Heat flux</td>
<td>0.0510</td>
<td>0.213</td>
<td>0.211</td>
</tr>
<tr>
<td>Freshwater flux</td>
<td>0.00277</td>
<td>0.0164</td>
<td>0.107</td>
</tr>
</tbody>
</table>
adjustments. Overall, the ECCO-GODAE solution has much more structure on small spatial scales than the WEIGHTED estimate. It is not clear whether differences result from increased subsurface weights in the WEIGHTED estimate indicate a lack of convergence or stem from differences in vertical resolution or other sources. It is likely partly due to the fact that these adjustments are smaller, and their contributions to the cost function are lower. The smoothness of the WEIGHTED solution makes it seem unlikely that noise is being fit too closely. The STANDARD estimate is very similar to the WEIGHTED estimate, with lower magnitudes.

6. Discussion and conclusions

The results of state estimation depend on many factors, including the model resolution, the amount and type of data being assimilated, and the prior variances used to constrain both the data and the control variables. One purpose of this study was to compare the influence of surface observations as constraints on ocean state estimates to that of subsurface observations. Comparisons between assimilation results excluding satellite and in situ data separately demonstrated that assimilating surface observations without in situ data provides a
better estimate than assimilating in situ observations without surface data; overall, however, the importance of using all available data is evident. Having established that, the effect of increasing the weighting on the subsurface observations was examined. Although the WEIGHTED solution is less consistent than the STANDARD solution with observed SSH anomaly, in comparisons with both assimilated and independent subsurface data, it is more consistent with most observations than the STANDARD estimate.

The changes that occurred when the improved XBT dataset was introduced into the assimilation are also important to this analysis. Figure 19 uses probability distribution functions (PDFs) of model-estimated minus observed temperature to show the change in temperature bias as the optimization progresses. This assessment, although somewhat crude in that it ignores possible regional differences, nonetheless gives a perspective on the overall distribution of temperature differences. Initially, model estimates exceed CTD and Argo observations, but XBT data are biased high relative to model output (Fig. 19a). After 50 iterations, the bias of XBT data is still evident (Fig. 19b). Once the corrected XBT dataset is introduced, the PDFs of the three assimilated datasets are almost indistinguishable (Fig. 19c). The unadjusted XBT data are shown for comparison, but these
data are no longer being assimilated. These results lead to several conclusions. First of all, the new dataset is more consistent with the other available data, which supports the validity of the corrections that were applied. Second, the model is responding appropriately to this new dataset through the assimilation process. Finally, the resulting estimate shows little to no global bias with respect to any of the assimilated datasets.

A comparison of the WEIGHTED estimate with a global, data assimilating model solution (ECCO-GODAE) reinforced the conclusion that the increased weights on the subsurface data provided a state estimate that is more consistent with observations but is still within expected limits in its adjustments to the control variables. Neither the ECCO-GODAE nor the WEIGHTED analyses have the ability to resolve mesoscale processes, so the increased consistency of ECCO-GODAE with SSH could be a result of that analysis fitting more closely to unresolved noise in SSH observations, whereas the WEIGHTED analysis fits unresolved noise in subsurface data. Neither can ever be expected to be perfect.

The comparison with the global estimate also demonstrated that adjustments to control parameters from the WEIGHTED experiment have similar patterns to those from the ECCO-GODAE estimate. The relative smoothness of the WEIGHTED results indicate that it is unlikely that noise is being fitted. Overall, these results indicate that further exploration of modified weights on subsurface data is warranted. Comparisons with independent time series at two locations indicate that, as a result of global smoothing, the one-dimensional weighting system lacks the subsurface peak in variability found in the thermocline, which is another aspect of the weighting system that could be improved.

There are still issues to be considered with respect to data assimilation systems. This analysis found evidence of temperature drift and errors in model physics in the mixed layer and in the Kuroshio region, all of which should be considered, but these do not diminish the result of the experiments. The goal was to determine if the subsurface observations could be matched more consistently without a loss of realism in the overall solution, which might be expected if aliased mesoscale features were projected onto the final estimate. However, within the ECCO framework, the combined goal of consistency with observations and consistency with model dynamics in ocean state estimation has been achieved by the WEIGHTED estimate to a greater degree than any of the other estimates in this analysis. This could result in an improved ocean state estimate for analysis of the dynamics of the North Pacific between 1992 and 2004.

Acknowledgments. We thank Ibrahim Hoteit, Armin Kohl, and Bruce Cornuelle for assistance with the setup of the regional ECCO model and valuable discussion during the experiments. Reanalysis surface forcing fields from the National Center for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) are obtained through a computational grant at NCAR. Computational support from the Scripps Observations and Modeling Center (COMPAS) is gratefully acknowledged. Computing resources used in running the optimization were provided by the National Center for Supercomputing Applications. The global state estimates were provided by the Estimating the Circulation and Climate of the Ocean (ECCO) consortium, which is funded by the National Oceanographic Partnership Program (NOPP). This work is supported in part through ONR (NOPP) ECCO Grants N00014-99-1-1049, NOAA Grant NA17RJ1231 (GODAE, Argo) and by the NASA Ocean Surface Topography Science Working Team through JPL Contract 961424. This is a contribution of the consortium for Estimating the Circulation and Climate of the Ocean (ECCO), which is funded by the National Oceanographic Partnership Program.
REFERENCES


