

Evaluating the Impact of the Number of Satellite Altimeters Used in an Assimilative Ocean Prediction System

ROBERT W. HELBER, JAY F. SHRIVER, AND CHARLIE N. BARRON

Naval Research Laboratory, Stennis Space Center, Mississippi

OLE MARTIN SMEDSTAD

QinetiQ North America, Stennis Space Center, Mississippi

(Manuscript received 9 December 2008, in final form 20 October 2009)

ABSTRACT

The impact of the number of satellite altimeters providing sea surface height anomaly (SSHA) information for a data assimilation system is evaluated using two comparison frameworks and two statistical methodologies. The Naval Research Laboratory (NRL) Layered Ocean Model (NLOM) dynamically interpolates satellite SSHA track data measured from space to produce high-resolution (eddy resolving) fields. The Modular Ocean Data Assimilation System (MODAS) uses the NLOM SSHA to produce synthetic three-dimensional fields of temperature and salinity over the global ocean. A series of case studies is defined where NLOM assimilates different combinations of data streams from zero to three altimeters. The resulting NLOM SSHA fields and the MODAS synthetic profiles are evaluated relative to independently observed ocean temperature and salinity profiles for the years 2001–03. The NLOM SSHA values are compared with the difference of the observed dynamic height from the climatological dynamic height. The synthetics are compared with observations using a measure of thermocline depth. Comparisons are done point for point and for 1° radius regions that are linearly fit over 2-month periods. To evaluate the impact of data outliers, statistical evaluations are done with traditional Gaussian statistics and also with robust nonparametric statistics. Significant error reduction is obtained, particularly in high SSHA variability regions, by including at least one altimeter. Given the limitation of these methods, the overall differences between one and three altimeters are significant only in bias. Data outliers increase Gaussian statistical error and error uncertainty compared to the same computations using nonparametric statistical methods.

1. Introduction

Sea surface height anomaly (SSHA) measurements from space provide a global and nearly real-time representation of ocean dynamical features and are the most important observation component of the U.S. Navy's operational global ocean prediction system. Observations of SSHA are particularly useful for mapping eddies, meandering currents, and fronts that are associated with, for example, planetary waves and geostrophic currents. This information is utilized by assimilation into the $1/32^\circ$ Naval Research Laboratory (NRL) Layered Ocean Model (NLOM; Shriver et al. 2007). NLOM has demonstrated skill as a dynamical interpolator of satellite

altimeter observations transforming the along-track data into high-horizontal-resolution (eddy resolving) SSHA fields that accurately represent the mesoscale variability.

Because NLOM is a layered model with high horizontal but low vertical resolution, it does not independently provide a realistic vertical structure. A more complete depiction is made by using the SSHA fields from NLOM within the Modular Ocean Data Assimilation System (MODAS) to construct subsurface three-dimensional temperature and salinity fields. These global fields, which we call synthetics, are generated through the use of gridded statistical relationships between SSHA, sea surface temperature (SST), and historical in situ profile observations of temperature and salinity (Fox et al. 2002a,b). MODAS synthetics are generated for several experimental cases that differ according to the selection of zero to three satellite altimeter data streams to be assimilated by NLOM. These synthetic fields of temperature

Corresponding author address: Robert W. Helber, Naval Research Laboratory, Stennis Space Center, MS 39529-5004.
E-mail: robert.helber@nrlssc.navy.mil

and salinity provide the basis for the present study and are compared with independent in situ ocean profile observations to evaluate their relative accuracy.

A main goal of this analysis is to evaluate the relative prediction accuracy resulting from different numbers and configurations of altimeters providing data for NLOM and MODAS. Another equally important goal is to develop validation methodologies for error assessment of ocean prediction systems in general. To accomplish these objectives, we apply two comparison frameworks and two statistical methodologies. Comparisons are made point for point and in binned overlapping 1° radius regions 60 days long. Error analyses are then computed using both traditional Gaussian and also robust nonparametric statistical methods. The metric used for the evaluation is thermocline depth (TD) because the SSHA is most highly correlated with the thermocline (Hurlburt 1986).

During the analysis time period, there were three satellite altimeter platforms used in daily operations. Maintaining that level of satellite coverage is expensive, and thus a clear understanding regarding the prediction impact of multiple satellites is of interest. Recent scientific literature shows the importance of satellite altimeters and their influence on ocean prediction. Use of the Navy's Coastal Ocean Model (NCOM) to predict SSHA, in assimilative and nonassimilative modes, was evaluated relative to tide gauges by Barron et al. (2004). Other studies assimilate altimeter SSHA and surface drifter data together (e.g., Lin et al. 2007; Fan et al. 2004; Ishikawa et al. 1996). In a study by Smedstad et al. (2003), identical twin numerical experiments simulated the SSHA error versus the number of assimilated satellite altimeters in a $1/6^\circ$ version of NLOM. The present analysis provides a similar error comparison but with assimilation of real altimeter measurements in $1/32^\circ$ NLOM and MODAS and in situ observations as the validation reference.

Numerical models can capture the statistical characteristics of the spatial and temporal variability relatively well. Climate studies often require that the underlying mechanisms and general character of the prediction is correct (e.g., Doney et al. 2007), while the specific skill of predicting temperature and salinity at a point is less important. Other validation studies evaluate the long-term temporal variability of models (Kara and Hurlburt 2006). More detailed predictions are often required by the fishing industry, the navy, and others for prediction of the synoptic variability. For example, the positions of fronts are often used by the maritime community to evaluate model skill (e.g., Oey et al. 2005).

An even more detailed method for comparing model results is to compare the model at the exact location of the observation point for point. This type of comparison

is a very stringent measure for model evaluation, because the observations are influenced by all the physical processes present in the ocean, some of which may not be adequately represented in the prediction system. For example, turbulent diffusion is parameterized, and internal waves and tides are not included in the prediction systems we evaluate here. Point-for-point comparisons can exaggerate the influence of small displacement errors of mesoscale features. If the predictions represent the mesoscale features relatively well but with slight shifts in the horizontal, relatively large point-for-point errors may occur.

A second type of comparisons begins with overlapping data bins with a 1° radius and 60-day time window. The data in these bins are then linearly fit in a least squares sense, providing a statistical representation (superposition data) of thermocline depth within the bin. This methodology is applied to produce both superobservation and superprediction data. The main advantages are an evening of the data weights between high- and low-density observation regions and estimates of time rate of change. Results using this procedure are compared with results from point-for-point comparisons to evaluate the impact of the comparison framework.

Both traditional Gaussian and robust nonparametric statistical methods are used in the present analysis. Ocean observations rarely exhibit Gaussian data distributions that are assumed when using Gaussian statistical methods. To evaluate the impact of non-Gaussian data distributions, nonparametric and Gaussian statistical methods are applied to all comparisons in this analysis.

In section 2, the observations are described, and special attention is paid to the vertical resolution because it has a large influence on the TD computations. The dynamically interpolated SSHA fields from NLOM are described in section 3a, and sections 3b and 3c describe MODAS and the details of the synthetic variants, respectively. The TD metric and superdata and statistical methods are described in section 4. The results are presented in section 5, with a summary of conclusions in section 6.

2. Profile observations

Observed profiles of temperature and profile pairs of temperature and salinity from the World Ocean Database 2005 (WOD05; Boyer et al. 2006), the U.S. Navy's Master Oceanographic Observation Dataset (MOODS; Teague et al. 1990), and Argo (Roemmich et al. 2004) are combined. Because the data sources for MOODS and WOD05 are mostly the same, much of these data are duplicated, but each set contains unique profiles. The WOD05 has less stringent quality-control (QC)

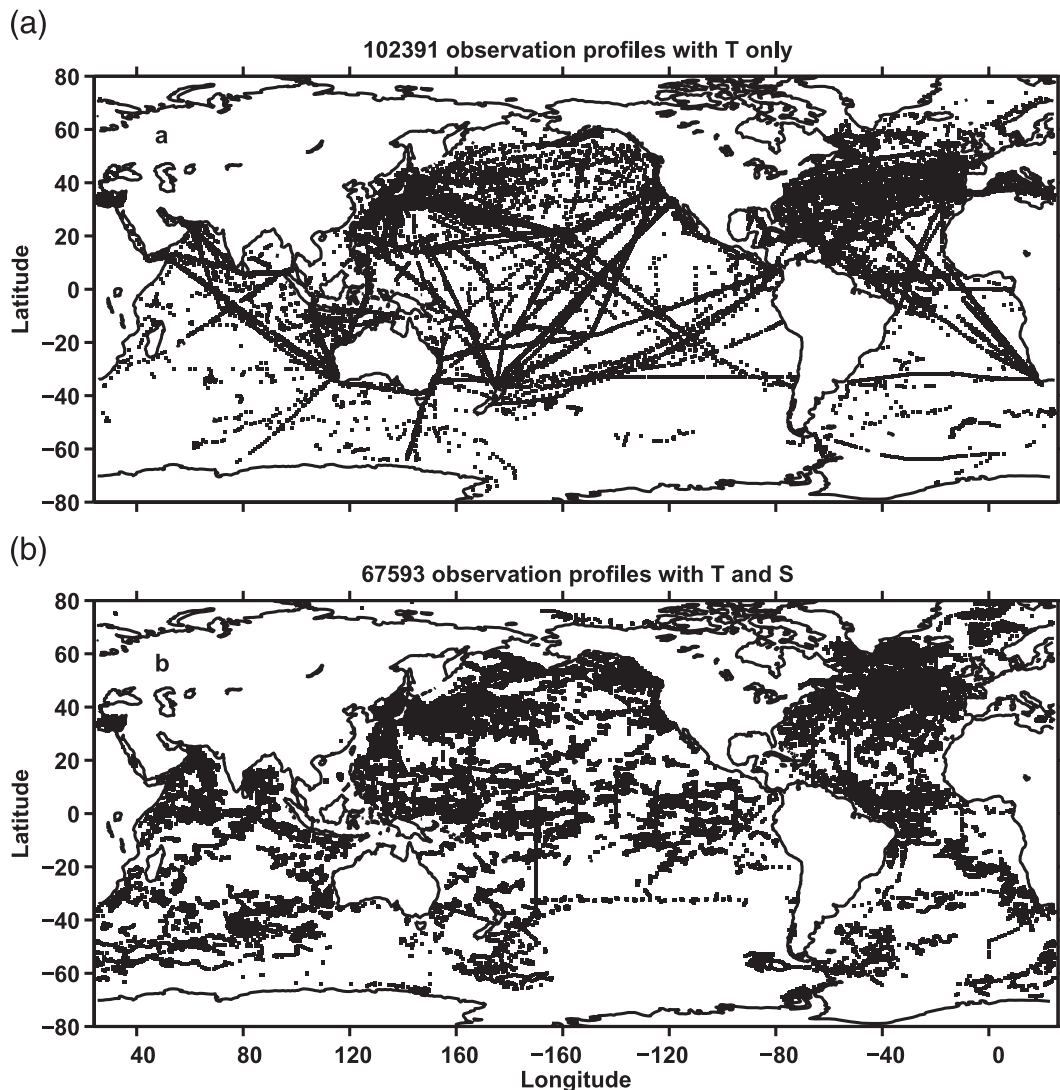


FIG. 1. The locations of the publicly available profile observations that have (a) temperature and (b) paired values of temperature and salinity.

procedures than MOODS. Argo data are contained in the MOODS and WOD05 datasets but not up to the present time. In addition, the Argo profiles in MOODS are from the real-time system and thus lack some of the QC procedures specifically designed for floats performed at the Argo data acquisition centers (DAC). Also, further QC recommendations are periodically available, such as the pressure sensor error that was discovered for certain float types (available online at <http://www.argo.ucsd.edu>). For these reasons, the latest Argo data from the DACs have been used in place of the same profiles present in WOD05 and MOODS.

The data for this study are temperature profiles from 2001 through 2003 that have the shallowest depth level above 12 m, the deepest depth level below 150 m, and

at least temperature values. Most profiles are expendable bathythermographs (XBTs), but many are from conductivity–temperature–depth (CTD) recorders. The total number of profiles for this time period is 225 160, with 73 549 having profiles pairs of temperature T and salinity S (Fig. 1a) and 151 611 having T only (Fig. 1b). An additional restriction is that, in the upper 150 m, no profile may have any gaps in depth levels exceeding 25 m. This restriction reduces the numbers to 222 772 total, 72 586 for T and S , and 150 186 for T only.

Although the data are quality controlled, errors including profile location, XBT drop rate, and other potential errors may still exist. The effect of location errors and how the QC procedures can miss this type of error are discussed by Kara et al. (2009).

3. Prediction system

a. NLOM

NLOM is a global numerical primitive equation layered formulation model with six dynamic layers and a bulk mixed layer. The operational eddy-resolving $1/32^\circ$ global ocean prediction system has $1/32^\circ$ by $45/1024^\circ$ (latitude by longitude) resolution and an embedded mixed layer depth (MLD) model independent from the six dynamic layers (Shriver et al. 2007). The mixed layer can extend across more than one dynamical layer (Wallcraft et al. 2003), and the system assimilates both satellite IR SST (Barron and Kara 2006) and SSHA observations (Barron et al. 2009).

The SSHA assimilation scheme was developed for the efficiency requirements of operational systems and uses an incremental updating technique (Smedstad et al. 2003). First, an optimum interpolation deviation analysis of SSHA is performed using a model forecast as a first guess and the mesoscale covariance functions of Jacobs et al. (2001). The deviation analysis is used to apply weights to all observations according to their space and time locations, in a 3-day window, during which the observations are incrementally updated in the model. The satellite altimeter products used for assimilation are from 1) *Jason-1*, 2) *European Remote Sensing Satellite-2 (ERS-2)/Envisat*, and 3) the *Geosat Follow-On (GFO)*. SST is assimilated through a heat flux relaxation scheme that depends on a 3-h e -folding time scale and the mixed layer depth. The surface information is transferred through all dynamic layers via the statistical inference technique of Hurlburt et al. (1990). This technique updates the pressure field in all layers below the surface.

An important aspect of NLOM is its estimate of the bottom pressure anomaly, which is used to split the SSH signal into steric and nonsteric components. The SSHA used in this study is NLOM's total SSH minus the SSH proportional to the bottom pressure anomaly minus NLOM's long-term mean steric SSH (Barron et al. 2007). As a result, this SSHA is the steric SSH anomaly that is required for MODAS (see section 3b).

A modified version of the $1/12^\circ$ 5-min gridded elevations/bathymetry for the world (ETOPO5) bottom topography (National Oceanic and Atmospheric Administration 1986) is used in the NLOM simulations. The topography is first interpolated to the model grid and then smoothed twice with a nine-point smoother to reduce energy generation at smaller scales that are poorly resolved by the model. The maximum depth of the model is set at 6500 m. The minimum depth, set at 200 m, is used as the model boundary, with a few exceptions where shallower depths are needed to connect semienclosed seas. Shelf regions are excluded, because the NLOM formulation

requires all layers to exist with positive thickness at all points.

Although assimilation of SSHA has an effect on NLOM's embedded MLD through the divergence of near-surface layer currents, its influence is more strongly seen in the TD. The influence of SSHA on MLD is indirect through the prognostic equation for MLD (Wallcraft et al. 2003),

$$\frac{\partial(h_m)}{\partial t} + \nabla \cdot (h_m \mathbf{v}_1) = \omega_m, \quad (1)$$

where h_m is the thickness of the embedded mixed layer, ω_m is the entrainment rate at the base of the mixed layer, and \mathbf{v}_1 is the velocity of NLOM's top layer. Assimilated SSHA is used to update layer thicknesses (which are related to the pressure gradient) and a geostrophic correction is applied to currents outside the near-equatorial band (approximately 5°N – 5°S) to rebalance the currents with respect to the new model SSHA. Because divergence is a second-order computation, we do not use the NLOM embedded MLD in the analysis metrics. Instead, we compare the dynamic height deviations of in situ profile observations from climatology with NLOM's steric SSHA. We also use NLOM's SSHA within MODAS to generate three-dimensional synthetic temperature and salinity fields for direct comparison with in situ profile observations.

A set of simulations using NLOM were run for years 2001–03, differing only in the number of altimeter data streams assimilated, from 0 to 3. Each altimeter data stream provides, in one day, approximately 35 000 observations globally. An additional three-altimeter simulation was started from a different initial condition (same day, different climatological year). Both three-altimeter cases produce similar results, except in areas where the variability is nondeterministic.

b. MODAS

MODAS estimates profiles of temperature and salinity from inputs of SST and/or SSHA using gridded polynomials at standard depths with coefficients determined by least squares fit to historical observations (Fox et al. 2002a,b). The SSHA input for the greatest MODAS accuracy is the offset from the steric height anomaly relative to 1000 m in the background climatology. If SST and/or SSHA are not available, the profile estimates revert to the MODAS bimonthly climatology. Both the MODAS background climatology and regression coefficients are defined in a database centered on 15 January, 15 March, 15 May, 15 July, 15 September, and 15 November. Synthetics for a given day are based on coefficients linearly interpolated in time and space to the desired location.

TABLE 1. The SSHA estimates and other details used in the generation of MODAS subsurface synthetics. All cases but CLIM use SST from the observation profile. The labels are used in the figures and text to denote each case.

| No. | Label | Description |
|-----|--------|---|
| 1 | CLIM | The MODAS background climatology without SSHA or SST input (the only case without the observed SST). |
| 2 | TONLY | The MODAS synthetic without SSHA input. |
| 3 | MODAS | The two-dimensional optimally interpolated SSHA fields produced using all three available altimeters by MODAS independently of NLOM. |
| 4 | BEST | The difference of the dynamic height computed from the observation profile and the dynamic height of the MODAS background climatology. This is the SSHA value that corresponds to MODAS's best possible representation of the observation. This is unavailable in practical applications and represents the upper limit on prediction accuracy. |
| 5 | A0 | NLOM SSHA produced without SSHA assimilation. Surface fluxes and SST assimilation is still used. |
| 6 | A1J | The NLOM SSHA assimilation is from <i>Jason-1</i> only. |
| 7 | A1G | The NLOM SSHA assimilation is from <i>GFO</i> only. |
| 8 | A2EG | The NLOM SSHA assimilation is from <i>ERS-2/Envisat</i> and <i>GFO</i> . |
| 9 | A3EGT | The NLOM SSHA assimilation from all available altimeters: <i>Jason-1</i> , <i>ERS-2/Envisat</i> , and <i>GFO</i> . |
| 10 | A3EGT2 | As in A3EGT, but the initial conditions are perturbed slightly to help identify the nondeterministic response to the initial conditions. |

Climatological profiles are linearly interpolated in space and time as well.

For this analysis, we provide SSHA and SST to MODAS to produce synthetic profiles of temperature and salinity from which we compute the thermocline depth parameters described in section 4a. To test the impact of assimilating altimeter data, we pair SST from the observed ocean profiles with SSHA from a series of NLOM experiments. In addition, climatology profiles are defined by a zero SSHA and climatological SST.

c. Synthetics

A set of synthetic profiles is derived using MODAS at each of the observed profile locations and times described in section 2. All of the synthetics use the shallowest temperature value from the observed profile as the SST input for MODAS; the climatology profiles use climatological SST and zero SSHA. Eight different SSHA estimates are used to create different sets of synthetics. The descriptions and labels for each are listed in Table 1. For this analysis, we focus on the relative impact that the number of satellite altimeters have on the combined NLOM and MODAS predictions of SSHA and thermocline depth.

4. Methods

a. Metrics

Two TD parameters computed from temperature profile observations and synthetics made during 2001–03 are listed in Table 2. The parameters are based on two isothermal layer depth (ILD) algorithms computed using freely available software described in Lorbacher et al. (2006) and Kara et al. (2000). These methodologies

can also be applied to density profiles to compute MLD, but this is not considered here because not all profiles have both temperature and salinity and we are primarily considering thermocline depth.

The ILD algorithms have been well tested, and their characteristics are known. For example, the Lorbacher et al. (2006) ILD selects the first curvature peak of the temperature profile and tends to be relatively shallow. Because the Lorbacher et al. (2006) ILD method is curvature based, the label we use is ILD_{∇} , where the subscript ∇ represents the gradient associated with curvature. The Kara et al. (2000) ILD method selects the depth where a change in temperature relative to the near surface exceeds a threshold. Because the Kara et al. (2000) ILD is based on a change in temperature, the label we use is ILD_{Δ} , where the subscript Δ indicates that it is a threshold methodology. Hereafter, the unadorned ILD will refer to the isothermal layer depth in general. The ILD_{Δ} values tend to be deeper than ILD_{∇} and are most closely associated with the seasonal ILD (e.g., Helber et al. 2008).

TABLE 2. The upper thermocline metric parameters that are computed from observations and synthetics.

| No. | Label | Description |
|-----|----------------|--|
| 1 | ILD_{Δ} | Temperature threshold ILD Kara et al. (2000) |
| 2 | ILD_{∇} | Temperature profile curvature ILD Lorbacher et al. (2006) |
| 3 | TD_{Δ} | TD below ILD_{Δ} |
| 4 | TD_{∇} | TD below ILD_{∇} |
| 5 | TS_{Δ} | $TS, (\Delta T/2)/(TD_{\Delta} - ILD_{\Delta})$; ΔT is from the TD_{Δ} computation |
| 6 | TS_{∇} | $TS, (\Delta T/2)/(TD_{\nabla} - ILD_{\nabla})$; ΔT is from the TD_{∇} computation |

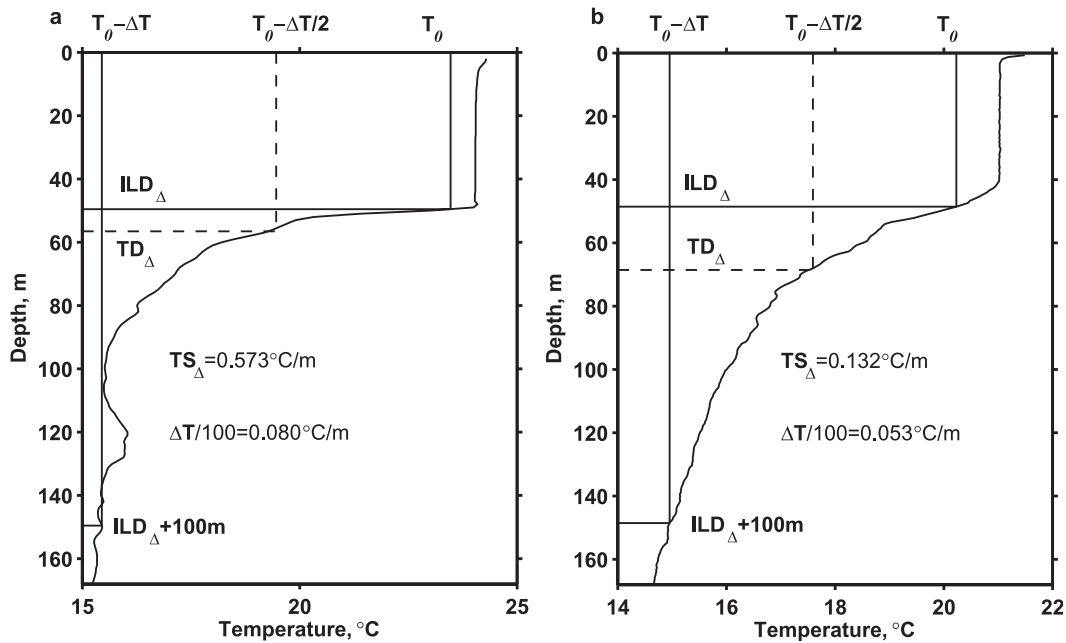


FIG. 2. The computation of TD for a typical temperature profile with a relatively (a) strong and (b) weak thermocline. The temperatures at the ILD, TD, and ILD + 100 m are T_0 , $T_0 - \Delta T/2$, and $T_0 - \Delta T$, respectively. The thermocline gradient is computed two ways, as labeled. The ILD is computed using the threshold method (ILD_{Δ}) and has a value of (a) 49.5 and (b) 48.6 m and the corresponding TD_{Δ} is (a) 56.5 and (b) 68.6 m.

For a small percentage of profiles, each ILD method has certain shortcomings. For example, for profiles where the water column is mixed all the way to the bottom of the profile, ILD_V returns a value of zero. This is because the algorithm cannot find a curvature peak (Kara et al. 2009). For similar cases, particularly at high latitudes, the threshold value for ILD_{Δ} is too large and returns an ILD that is unrealistically deep or at the bottom of the profile. To eliminate this problem, profiles are discarded if the ILD estimates are at the bottom of the profile and more than 50 m from the ocean bottom (based on the navy’s 2-min bathymetry) because in these cases the observations do not allow for an accurate estimate of the ILD.

The TD parameter is computed by first determining two temperatures: 1) the temperature of the profile at the ILD T_0 and 2) the change in temperature between T_0 and the temperature 100 m below the ILD, ΔT . The depth below the ILD where the temperature is equal to $T_0 - \Delta T/2$ defines TD. Because we have two estimates of the ILD, there are two corresponding estimates of TD, TD_{Δ} and TD_V , which correspond to ILD_{Δ} and ILD_V , respectively.

The TD_{Δ} is shown graphically with labels in Figs. 2a,b. For profiles with sharp thermoclines just below the ILD_{Δ} , the TD_{Δ} is relatively close to the ILD_{Δ} (Fig. 2a). For weak thermoclines, the TD_{Δ} is further from the ILD_{Δ} (Fig. 2b). In this way, the TD_{Δ} characterizes the

location of the highest gradients of the upper thermocline from a simple, widely applicable calculation. Because the computation is a bulk estimate of the high temperature gradient depth, it is insensitive to random noise in the profile. Finite difference computation of vertical gradients in observation profiles is inherently noisy and thus avoided in this analysis. The strength of the thermocline (TS) is the gradient of the upper thermocline between the ILD and TD (Table 2; Fig. 2).

b. Superposition

As mentioned in the introduction, observation time and locations are grouped into analysis windows to produce superposition data (“superdata”). For this analysis, the windows are 1° radius circular regions 60 days long computed each month for the entire analysis time period from January 2001 through December 2003. The 1° radius circles are centered on even-numbered latitude and longitude locations. For consistent comparison, the predictions are sampled or interpolated to the time and three-dimensional locations of the observations and are binned in the same manner.

In the analysis windows, superdata are defined as linear least squares fits to the model or observation data. The superdata values at the location and time of the analysis window (x, y, τ_0) are the mean a_0 and temporal

trend a_1 . The linear fit minimizes the squared residuals ε over all the data within the analysis window, where the residual ε_j for the j th data point, data_j , within the analysis window is defined as

$$\varepsilon_j = a_0(x, y, \tau_0) + a_1(x, y, \tau_0)(\tau_j - \tau_0) - \text{data}_j, \quad (2)$$

where τ_j is the time vector in units of days and τ_0 is the value at the start of the 60-day period such that a_0 is the mean and a_1 is the rate of change per day. The residual, ε , is minimized in a linear least squares sense and the mean fit error is computed as

$$e = \sqrt{\frac{1}{N} \sum_{j=1}^N [\varepsilon_j]^2}, \quad (3)$$

where N is the number of data points in the analysis window. These fit parameters, a_0 , a_1 , and e , are computed to produce the superobservations or superpredictions only for bins where the data span at least 30 of 60 days.

The superdata methods reduce the weight of individual data points in regions of the ocean with relatively high observation density. This is because the subsequent statistical analysis treats every superdata point the same. As a result, regions with a relatively large number of points have the same weight as regions with a small number of points. Both superdata and point-for-point results are discussed in section 5.

c. Statistical error analysis methods

Standard data quality-control procedures are not perfect and allow some erroneous data through. In an attempt to reduce the impact of this noise, we employ robust nonparametric statistical methods, because these methods are less sensitive to the influence of outlier data values (e.g., Rousseeuw and Leroy 1987). Because the nonparametric methods have analogous counterparts in traditional Gaussian statistics, we list those first and explain the robust methods second.

For assessment of prediction accuracy, we compute the mean bias (MB) as

$$\text{MB} = \frac{1}{N} \sum_{j=1}^N (P_j - O_j), \quad (4)$$

the root-mean-square error (RMSE) as

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{j=1}^N (P_j - O_j)^2}, \quad (5)$$

and correlation coefficient R , as

$$R = \frac{\frac{1}{N-1} \sum_{j=1}^N (P_j - \bar{P})(O_j - \bar{O})}{\left[\frac{1}{N-1} \sum_{j=1}^N (P_j - \bar{P})^2 \right]^{1/2} \left[\frac{1}{N-1} \sum_{j=1}^N (O_j - \bar{O})^2 \right]^{1/2}}, \quad (6)$$

where N is the number of samples. For the superdata comparison framework, N would be the number of analysis windows; for the point-for-point comparison framework, N would be the number of actual observation points. The corresponding P_j and O_j are the NLOM SSHA or MODAS synthetic predictions and the observations, respectively, in either point-for-point or superdata form. The sign convention of bias is $P_j - O_j$, such that positive bias corresponds to synthetics larger than the observations. Mean values are denoted with a bar, \bar{P} and \bar{O} , and are computed as

$$\bar{P} = \frac{1}{N} \sum_{j=1}^N P_j \quad \text{and} \quad \bar{O} = \frac{1}{N} \sum_{j=1}^N O_j.$$

In addition, the unbiased RMSE is also computed such that

$$\text{RMSE}_{\text{unbiased}} = \sqrt{\frac{1}{N} \sum_{j=1}^N [(P_j - \bar{P}) - (O_j - \bar{O})]^2} \quad (7)$$

and

$$\text{RMSE} = \sqrt{\text{MB}^2 + \text{RMSE}_{\text{unbiased}}^2}. \quad (8)$$

Equations (4)–(8) are the standard formulation for computing statistics on variables with Gaussian distributions. In practice, oceanographic data are rarely Gaussian and using the standard computations is often misleading. The robust or nonparametric methods used in this analysis are the biweight mean and standard deviation described by Lanzante (1996) and used by others (e.g., Zou and Zeng 2006). With these methods, the analogous quantity for mean bias is

$$\text{MB}^{\text{bw}} = \langle P_j - O_j \rangle^{\text{bw}}, \quad \text{for } j = 1, 2, 3, \dots, N, \quad (9)$$

and for root-mean-square error is

$$\text{RMSE}_{\text{unbiased}}^{\text{bw}} = \langle \langle P_j - O_j \rangle \rangle^{\text{bw}}, \quad \text{for } j = 1, 2, 3, \dots, N, \quad (10)$$

where $\langle \rangle^{bw}$ represents the biweight mean operator and $\langle \langle \rangle \rangle^{bw}$ represents the biweight standard deviation operator (see the appendix for details). Because the biweighted standard deviation is an unbiased estimate of the variability scale, Eq. (8) is used to provide the biweighted estimate of RMSE,

$$RMSE^{bw} = \sqrt{(MB^{bw})^2 + (RMSE_{unbiased}^{bw})^2}. \quad (11)$$

The biweight methods are robust in that they are nonparametric, meaning that the data distribution needs not resemble an assumed form. The biweight methodology applies less weight to data points that are statistically considered outliers of the natural distribution of the data. The details of the biweight computations are described in the appendix. To evaluate the impact of the robust error statistics, the results from both Gaussian and nonparametric statistical methods are presented in section 5.

For consistent comparisons of data with different error variance levels, we use “summary” diagrams (Jolliff et al. 2009). The advantage of these diagrams is that many experimental results can be presented on one figure for quick evaluation of the complete error statistics. The quantities are scaled by the standard deviation of the observations so that the magnitudes of values on a summary diagram tend to be less than one. A value of one indicates that the errors are as large as the observation standard deviation. The y axis indicates the scaled MB,

$$MB' = \frac{MB}{\left[\frac{1}{N} \sum_{j=1}^N (O_j - \bar{O})^2 \right]^{1/2}}, \quad (12)$$

or the biweight version,

$$MB^{bw'} = \frac{MB^{bw}}{\langle \langle O_j \rangle \rangle^{bw}}, \quad (13)$$

and the x axis denotes the scaled unbiased RMSE,

$$RMSE'_{unbiased} = \frac{RMSE_{unbiased}}{\left[\frac{1}{N} \sum_{j=1}^N (O_j - \bar{O})^2 \right]^{1/2}}, \quad (14)$$

or the biweight version,

$$RMSE_{unbiased}^{bw'} = \frac{RMSE_{unbiased}^{bw}}{\langle \langle O_j \rangle \rangle^{bw}}. \quad (15)$$

To investigate the impact of outliers, results from both the Gaussian statistics [Eqs. (12) and (14)] and the nonparametric statistics [Eqs. (13) and (15)] are presented in summary diagrams and discussed in section 5.

The scaling is important because it facilitates comparison of different datasets with varying variance levels. This is the case in this experiment because TD_{Δ} and TD_{∇} have different variance levels by virtue of differences in the ILD methodology. The summary diagrams allow consistent comparison of these side by side.

Statistical significance is computed with the use of bootstrap standard error estimates (Efron and Tibishirani 1986). The error bars presented in the figures of section 5 represent a 100 independent draw bootstrap standard error estimate.

5. Results

a. Global consistency

Inspection of the model results relative to observations reveals a remarkable similarity that bolsters confidence in the U.S. Navy’s ocean prediction system. For example, Figs. 3a,b show the mean SSHA values (superdata fit parameter a_0) for the global observations and the A3EGT SSHA (the SSHA used in the A3EGT synthetics described in Table 1), respectively. Each dot represents regions that have enough data to span 30 days of the 60-day bins. The parameter a_0 represents the mean of 1° diameter circles centered on even latitude and longitudes for a 60-day period starting on 1 January for the years 2001–03. In this format, comparing the two fields, the spatial variability appears similar and the magnitude is close. Similarly, TD_{Δ} from observations and the A3EGT synthetics are shown in Figs. 4a,b. Comparisons with ILD climatologies (available online at <http://www7320.nrlssc.navy.mil/nmld/nmld.html>; Kara et al. 2003) are consistent, though the TD_{Δ} is expected to be deeper.

b. SSHA accuracy

To gain an understanding of the relative accuracy of the NLOM SSHA estimates before using them to construct synthetics with MODAS, we compare them to the BEST SSHA estimate (the SSHA used for the BEST synthetics described in Table 1). This estimate is the closest we have to direct in situ observations of SSHA. As mentioned in section 4c, consistent comparisons can be made with summary diagrams and, consequently, all SSHA errors (relative to the BEST SSHA) using both nonparametric and Gaussian statistical methods are presented in Fig. 5. Using the superdata comparison

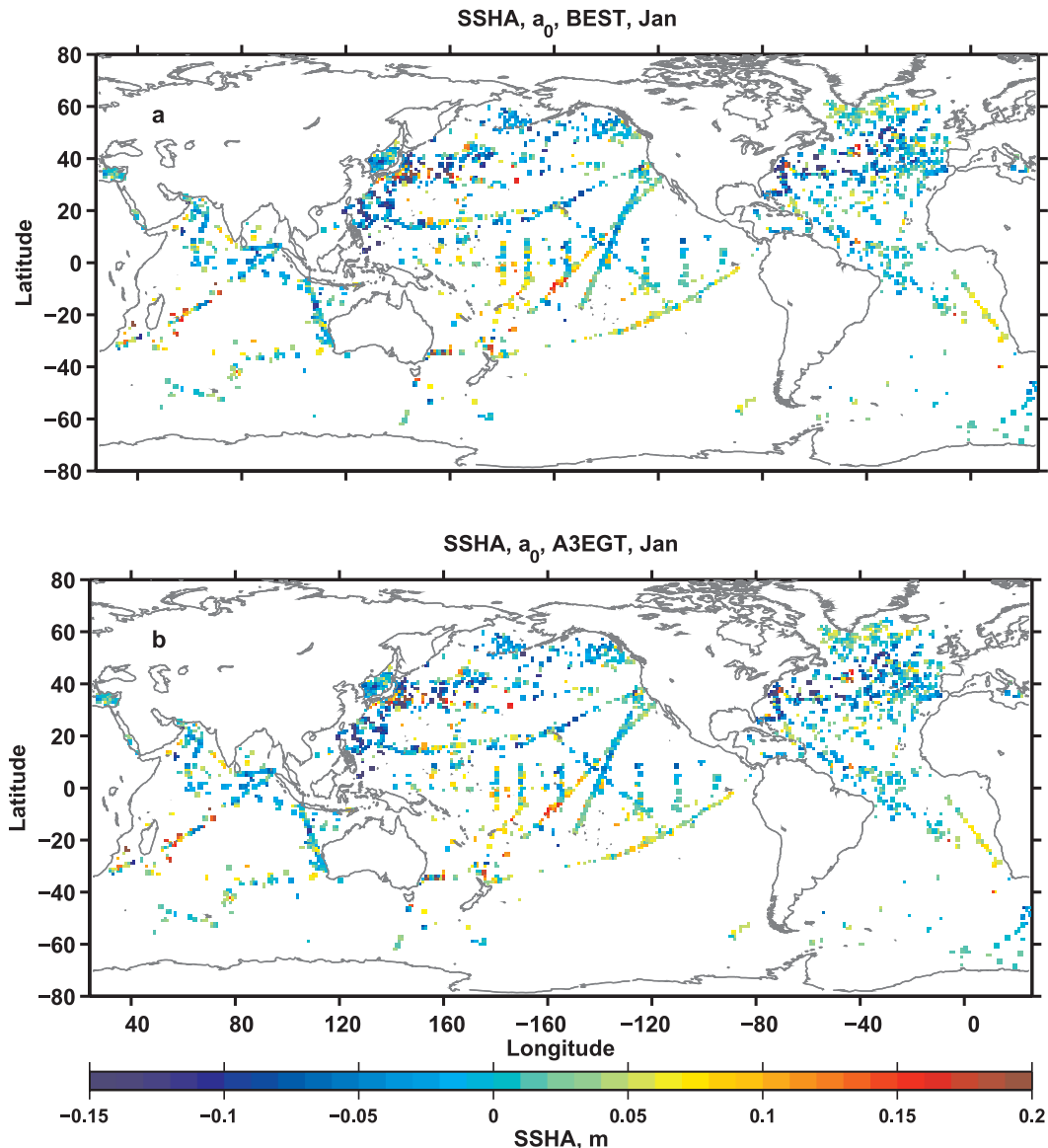


FIG. 3. The fit parameter a_0 (superdata mean; in m) for the (a) BEST and (b) A3EGT SSHA estimates for the global ocean for a 60-day analysis window starting 1 Jan for the years 2001–03. Values exist only where data span at least 30 of the 60 days.

framework (Fig. 5a), we find that the SSHA estimates using two or three altimeters have the smallest bias, whereas the MODAS SSHA estimate has the smallest RMSE. In Fig. 5 and the rest of the summary diagrams, the greater accuracy is indicated by values closer to the top-left corner of the figures. Farther toward the left represents smaller RMSE, and farther toward the top represents smaller bias. Bias is decreasing upward, because in the cases considered here it is always negative. MODAS tends to have a shallow bias in MLD and TD because of the unrealistically smooth vertical gradients of the synthetics.

The Gaussian statistical methods produce similar results to those for the nonparametric methods, with the exception of the A0 case (Fig. 5a), which has larger RMSE. For the traditional Gaussian statistical results, the bootstrap standard error bars are unstable and sensitive to the number of independent draws. For this reason, some of these error bars extend outside the axis limits in Fig. 5. The point-for-point comparison framework has a reduction in bias error with an increase in RMSE error only for Gaussian statistics (Fig. 5b).

Because standard error bars for the one- and two-altimeter cases do not overlap, two altimeters are

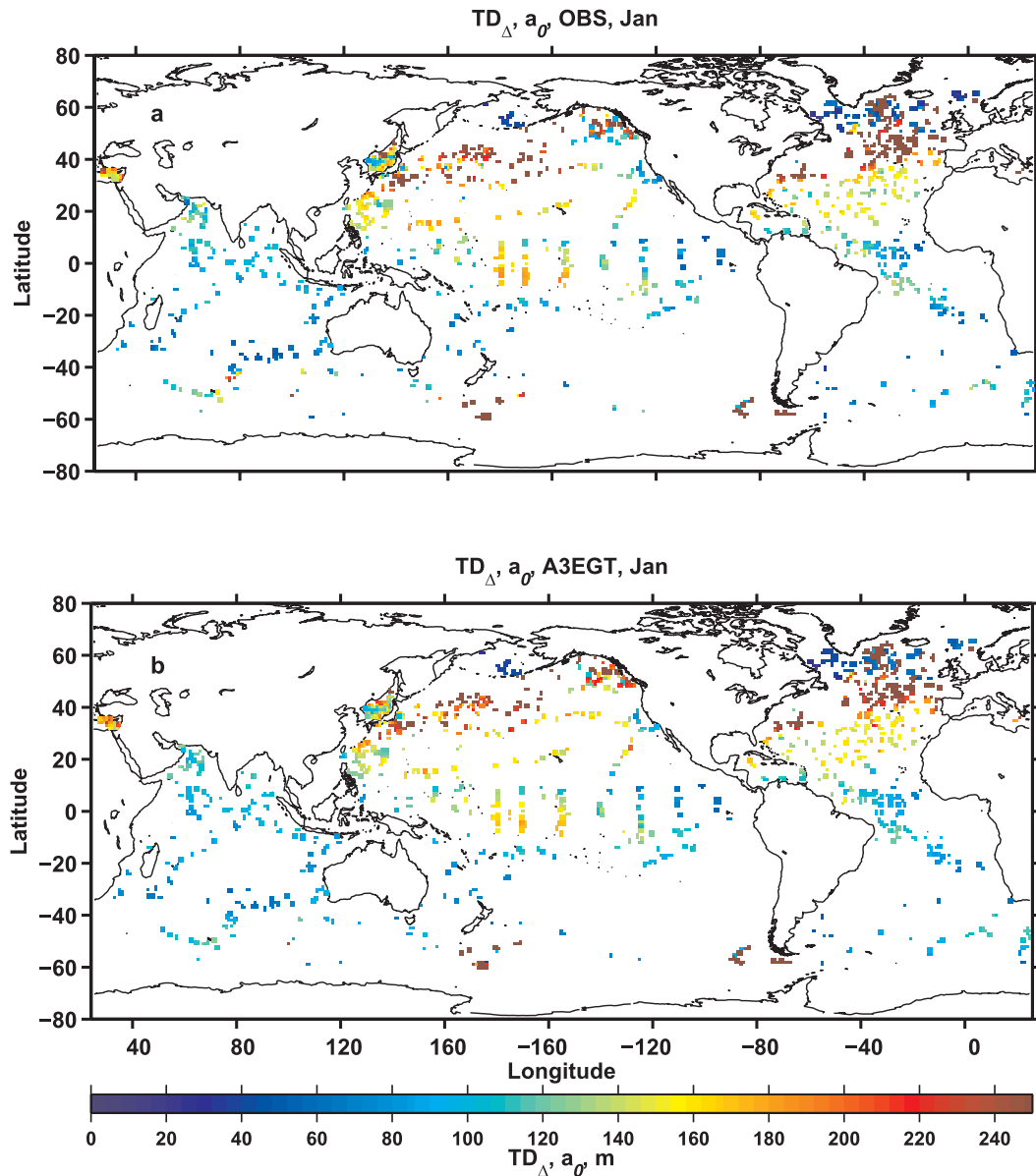


FIG. 4. The fit parameter a_0 (superdata mean; in m) for the (a) observed and (b) A3EGT synthetic TD_{Δ} estimate for the global ocean for a 60-day analysis window starting 1 Jan for the years 2001–03.

significantly more accurate than just one. One altimeter is also clearly better than the nonassimilative NLOM case. It is interesting to note that the A1G has smaller errors than the A1J, suggesting the higher horizontal resolution of the GFO orbit may reduce the overall bias. This is consistent with the expected error estimates of Jacobs et al. (2002).

c. TD accuracy

The primary metric for this analysis is the TD, which is computed from the MODAS synthetics. An advantage of this metric is that we are able to compute TD from

both the observation and the synthetic profiles, whereas for SSHA there is no direct in situ observation. In this section, summary diagrams for all synthetics for both TD metrics are presented in Figs. 6a–d. In these figures, the lowest error is again closest to the top-left corner of the plot, because there is a shallow TD bias ($MB' < 0$) for all synthetics and $RMSE' > 0$. The error bars are from a 100 independent draw bootstrap standard error estimate. Figure 6a uses both the superdata and non-parametric methods and therefore should be the most reliable global error estimate. Figure 6b also uses the superdata methods, but with Gaussian statistics. Figures 6c,d

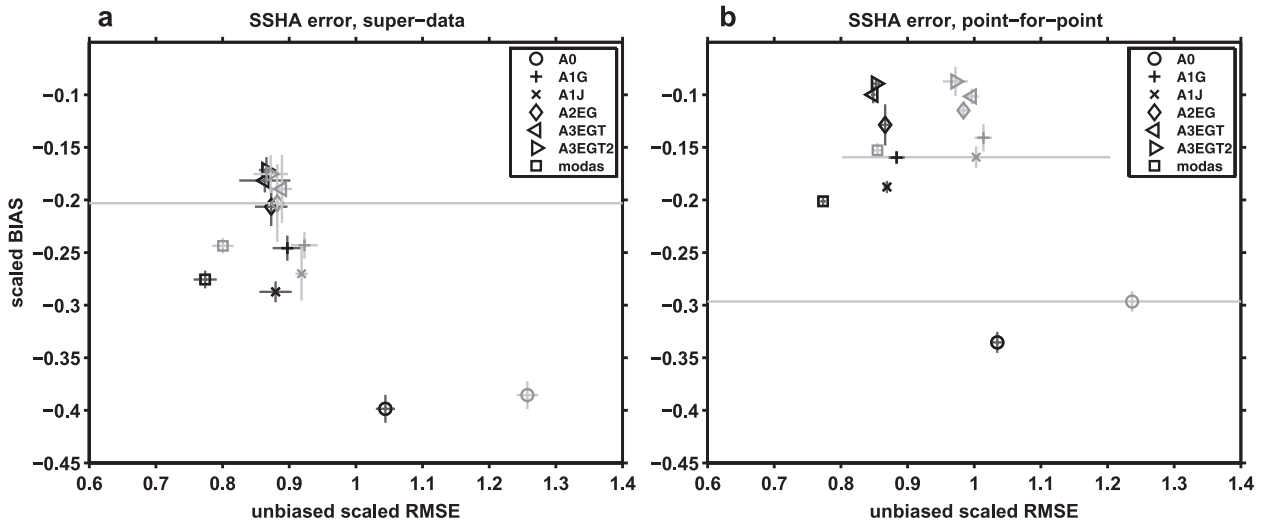


FIG. 5. The (a) superdata and (b) point-for-point summary diagram of biweight statistics $RMSE'_{unbiased}$ (x axis) and $MB^{bw'}$ (y axis, black; 25% gray error bars) and Gaussian statistics $RMSE'_{unbiased}$ (x axis) and MB' (y axis, 50% gray; 75% gray error bars) for SSHA for the six NLOM and MODAS SSHA estimates denoted by the symbols in the legend. The SSHA values are those associated with the synthetics described in Table 1. The error bars represent a 100 independent draw bootstrap standard error estimate.

are point-for-point comparisons with nonparametric and Gaussian statistics, respectively.

The statistical methods have the largest impact on the results (cf. Fig. 6a with 6b and Fig. 6c with 6d). Note that the axis limits are larger in Figs. 6b,d. Outliers increase the errors when using Gaussian statistical techniques because the nonparametric statistics reduce the weight of points that exist far from the median absolute deviation (see appendix), whereas the impact of the superdata versus point-for-point comparisons is relatively small (cf. Fig. 6a with 6c; cf. Fig. 6b with 6d). There is, however, less separation between the TD_{Δ} and TD_{∇} metrics for the point-for-point framework. There is also a slight increase in error levels for the point-for-point framework, suggesting that there are larger errors in regions of high data density.

Notice that the synthetics are more skillful for the threshold method TD_{Δ} , and the gradient method TD_{∇} has the larger error for all cases. This occurs because MODAS synthetic profiles tend to have overly smooth vertical gradients that result in a shallow bias for ILD_{Δ} and therefore TD_{Δ} . The shallow bias is exacerbated by the curvature-based ILD_{∇} algorithm when operated on synthetics because of the reduced curvature relative to the observed profiles.

The synthetic with the smallest overall error for both metrics using nonparametric statistics is BEST (Figs. 6a,c), which shows the optimal performance of the MODAS approached by using as input the actual dynamic height deviation from the observed profiles (see Table 1). The smallest bias is for CLIM and TONLY, which have nearly

identical errors. The next smallest error is with the three-altimeter synthetics A3EGT and A3EGT2. The largest error, in MB and RMSE, tends to occur with the zero-altimeter synthetic case A0. In terms of RMSE, all synthetics, except BEST and A0, tend to have nearly the same RMSE. The inclusion of altimeter assimilation tends to have the largest impact on the MB.

The results using Gaussian statistics are less clear because the standard error bars are larger and the estimates are clustered closer together (Figs. 6b,d). The general trend is maintained, where fewer altimeters have larger errors, but the separations between values are generally not significant.

The authors have not found a suitable nonparametric estimate for the correlation coefficient. The traditional estimate computed using Eq. (6) for both comparison frameworks is shown in Fig. 7. Although all correlations are high, the differences between the cases are not significant, with the exception of the difference between the A0 and the other altimeter assimilative cases. A major difference in correlation relative to the results shown in Fig. 6 is the low correlation values of CLIM and TONLY. This is consistent because climatologies do not vary on subseasonal time scales; therefore, we would expect lower correlations. Climatologies are designed to represent long space and time scale averages such as those computed for Fig. 6.

The overall results show that a significant error reduction is obtained by including at least one altimeter. The difference between one and three altimeters is significant only in bias. Given the limitations of the time

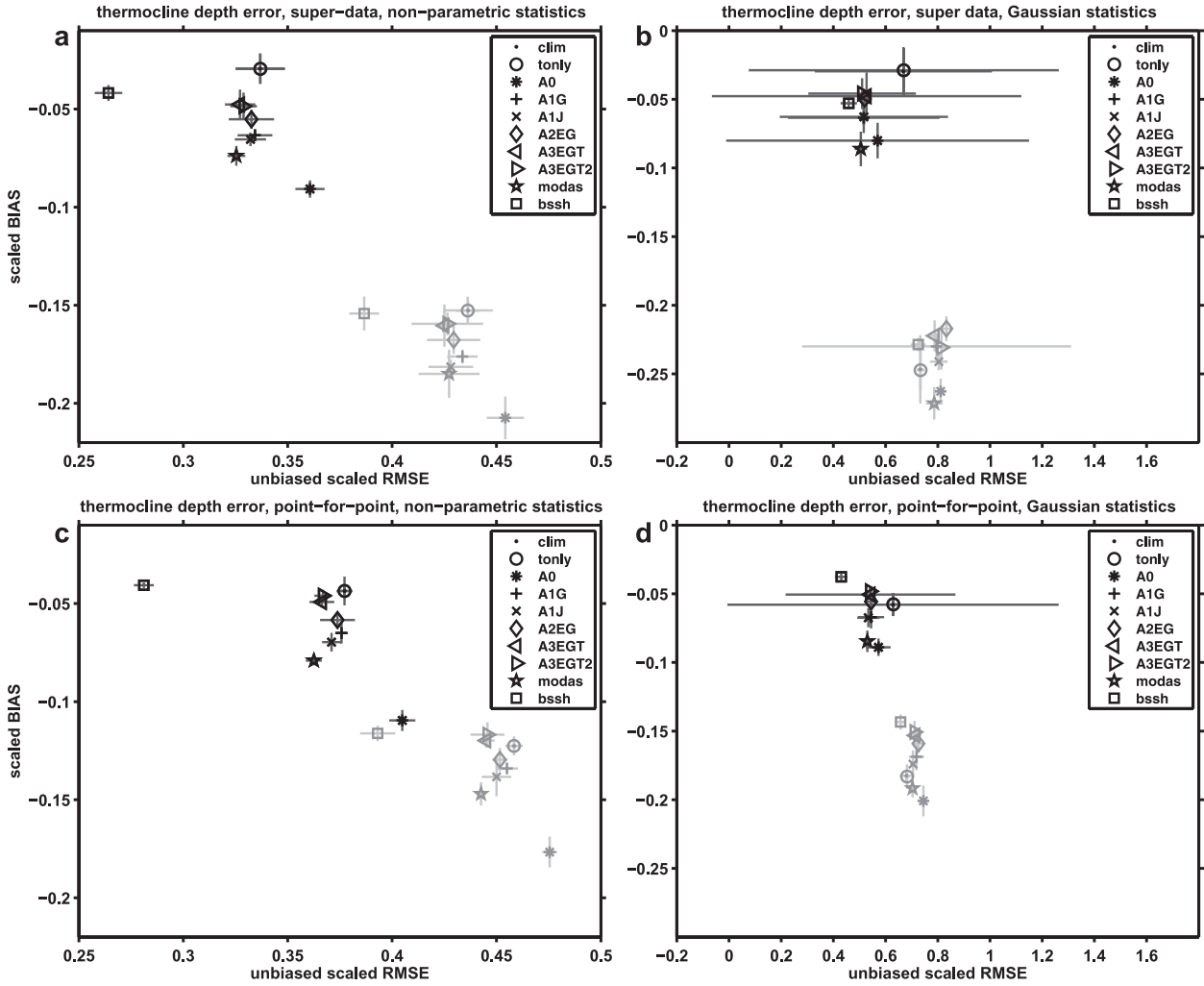


FIG. 6. (a) The superdata and nonparametric statistics summary diagram of $RMSE_{unbiased}^{bw}$ (x axis) and MB^{bw} (y axis) for TD_{Δ} (black and 75% gray error bars) and TD_{∇} (50% and 25% gray error bars) for the 10 synthetics denoted by the symbols in the legend as described in Table 1. (b) The superdata and Gaussian statistics summary diagram of $RMSE'_{unbiased}$ (x axis) and MB' (y axis) for TD_{Δ} (black and 75% gray error bars) and TD_{∇} (50% and 25% gray error bars). (c) The point-for-point data and nonparametric statistics summary diagram of $RMSE_{unbiased}^{bw}$ (x axis) and MB^{bw} (y axis) for TD_{Δ} (black and 75% gray error bars) and TD_{∇} (50% and 25% gray error bars). (d) The point-for-point data and Gaussian statistics summary diagram of $RMSE'_{unbiased}$ (x axis) and MB' (y axis) for TD_{Δ} (black and 75% gray error bars) and TD_{∇} (50% and 25% gray error bars). The error bars represent a 100 independent draw bootstrap standard error estimate.

and space scales inherent in the NLOM/MODAS assimilation, we are unable to discern a significant difference in the errors using two or three altimeters. The use of nonparametric statistics tends to be more effective in discriminating differences among the MODAS SSHA-derived predictions.

d. Error variability

To explore the seasonality of the error and fit parameters over the annual cycle, we choose two regions in the northwest Pacific Ocean. The high variability (HV) region bounded by 20° – 50° N and 120° E– 160° W has high

SSHA variability because of the meandering of the Kuroshio Current. The low variability (LV) region bounded by 0° – 20° N and 130° E– 150° W has relatively low SSHA variability.

In the HV region, the fit error e is driven by the depth of TD_{Δ} such that e is larger when TD_{Δ} is deeper in the months of December, January, and March (cf. solid lines of Figs. 8a,c). The percent error, which is the percent of the error relative to the observed TD_{Δ} given by

$$\% \text{ error} = 100 \left[\frac{a_0^{(A3EGT)} - a_0^{(obs)}}{a_0^{(obs)}} \right], \quad (16)$$

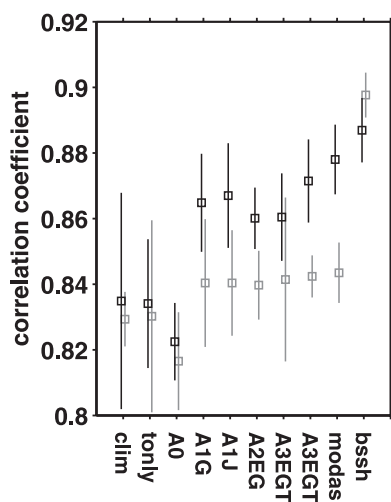


FIG. 7. The correlation coefficient of observed vs synthetic TD_{Δ} for each of the cases in Table 1 as listed on the x axis. The correlation is computed on superdata (black) and point-for-point (gray) values, and the corresponding vertical lines represent the 100 draw bootstrap standard error estimate.

is related to the rate of change fit parameter a_1 . When the magnitude of a_1 is large the magnitude of the percent error is also large (cf. solid black lines of Figs. 8b,d). The percent error tends to be negative because the synthetics have a shallow bias.

For the low variability region, which is represented by the dashed–dotted lines in Fig. 8, the effects of seasonality are relatively weak for a_0 , a_1 , and e , which change little throughout the year. In Figs. 8d,e, percent error and correlation coefficient for both HV and LV regions are considered for both the A3EGT synthetics (in black) and CLIM (in gray) cases. In Figs. 8d,e, the gray and black dashed–dotted lines are relatively close together compared to the solid lines. This suggests that the A3EGT and CLIM cases have similar errors in the LV region. In the HV regions, however, the CLIM case has larger magnitude in Fig. 8d and lower correlation on average in Fig. 8e. The average correlation coefficient in the HV region for a_0 determined from the A3EGT synthetics is 0.93, higher than 0.83 for climatology. Average correlation is also relatively low in the LV region, 0.84 for both A3EGT and CLIM. In the LV region, the accuracy of the A3EGT synthetics is not significantly different from climatology. The value added by altimetry in the MODAS synthetics is evident only in the region of high SSH variability, where the signal stands out from the background noise.

The global distribution of observed a_0 values can be seen in Fig. 9. For the 60-day period starting 1 February, the northwest Pacific Ocean Kuroshio Extension region

(the HV region in the solid box of Fig. 9) has relatively deep TD_{Δ} in the annual cycle. In contrast, the 60-day period starting in August has relatively shallow TD_{Δ} in the Kuroshio Extension region (see also Fig. 8). The seasonal cycle is much weaker in the LV region (dashed–dotted box in Fig. 9).

The rate of change of TD_{Δ} is given by the fit parameter a_1 , as shown in Fig. 10 globally for the 60-day periods starting in March and November. In the Kuroshio Extension (HV) region, a_1 is shoaling most rapidly in March and April and deepening in November and December (Fig. 8). This pattern for the annual cycle of TD_{Δ} rate of change applies generally to most regions in the mid-latitude Northern Hemisphere.

The annual cycle of the percent error in A3EGT synthetics is linked to the TD_{Δ} rate of change. Wherever the magnitude of a_1 is large, the percent error magnitude is large. For the Kuroshio Extension region, the 60-day periods starting in April and December have relatively large percent error (Fig. 8); the global distributions of errors for these months are shown in Fig. 11. The largest errors tend to be where SSHA variability is large: for example, in the eddy shedding regions of the Kuroshio and the Gulf Stream. In regions of the ocean where the SSHA variability is weak, such as the LV region, the errors are relatively small.

6. Summary and conclusions

The impact of the number of satellite altimeters on upper-ocean predictions of thermocline depth (TD) are evaluated using global, data assimilating, layered model SSHA analyses. By varying the number of altimeters from zero to three, the prediction accuracy is determined relative to a global set of in situ profile observations. In addition, methods for evaluating prediction accuracy are presented using two comparison frameworks and two statistical methodologies. Comparisons are made point for point and for binned and fitted superdata regions (1° radius, 60 days). Both traditional Gaussian and non-parametric statistical methodologies are applied.

The general results show that accuracy provided by the satellite altimeter datasets is greater in regions of high SSHA variability, and significant error reduction is achieved with the addition of at least one satellite altimeter dataset. Under the limitations of the analysis methods, additional error reduction when assimilating data from three altimeters versus one is significant only in bias. These conclusions are drawn from both the point-for-point and superdata frameworks when using nonparametric statistical methods. The lack of significant skill improvement between two and three satellite altimeter datasets may be a consequence of the length

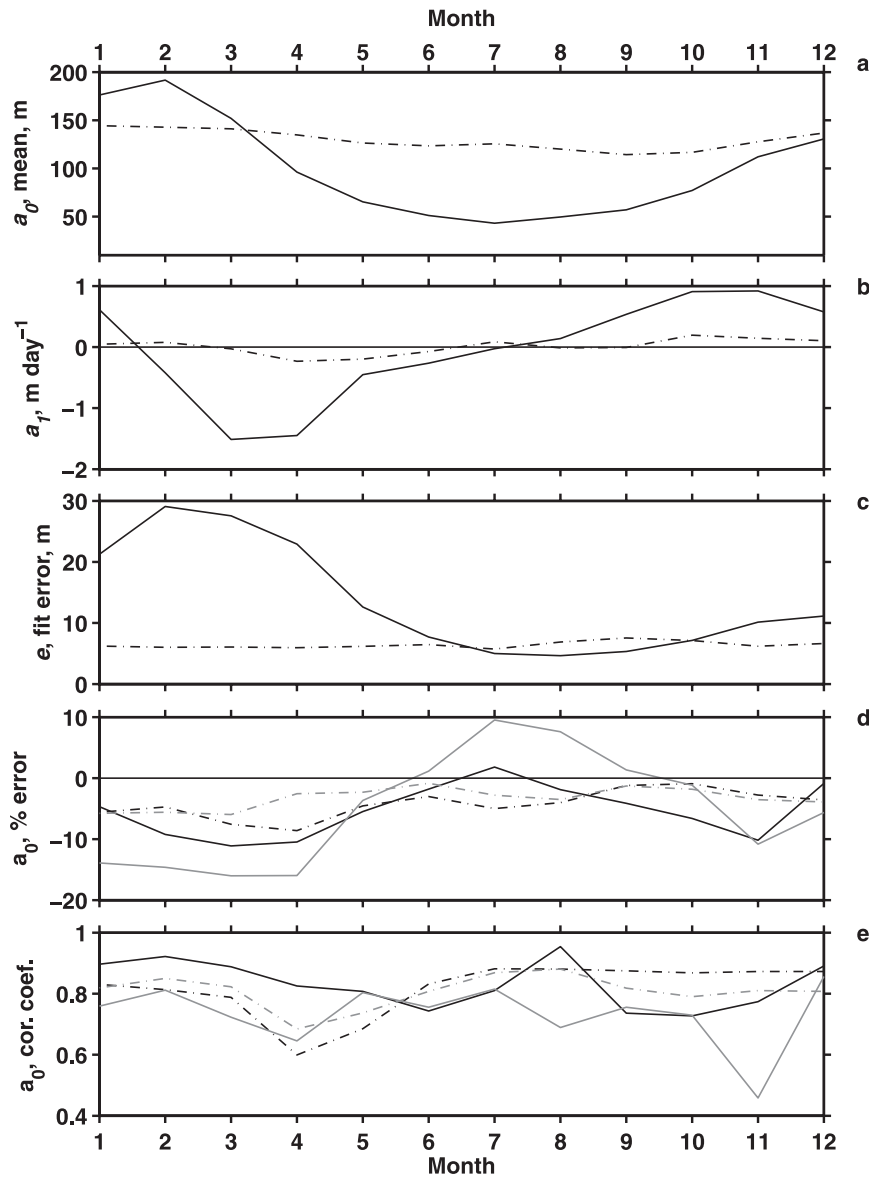


FIG. 8. The median fit parameters (a) a_0 , (b) a_1 , and (c) e and the (d) percent error and (e) correlation of a_0 from A3EGT synthetics (in black) and CLIM (in gray) by month, relative to the fit parameters determined directly from observations. The solid lines are for the region bounded by 20°–50°N and 120°E–160°W and the dashed–dotted lines are for the region bounded by 0°–20°N and 130°E–150°W in the northwest Pacific Ocean.

and time scales associated with the NLOM assimilation and the smoothing associated with gridded climatological coefficients of MODAS. As a result, this system is unable to take full advantage of the added spatial detail provided by multiple altimeters, even though each additional data stream adds approximately 35 000 data points globally per day. The simulated error experiment of Smedstad et al. (2003) also indicates that the marginal reduction in SSHA error decreases for each additional satellite data stream.

As expected, the BEST synthetics have the greatest accuracy but are unavailable in practical applications because the observed profiles are used for their derivation. The BEST SSHA estimate is fully consistent with the observed profile within the context of MODAS synthetics. This upper accuracy limits of MB' and RMSE' (bias and rms error scaled by the observations standard deviation) are approximately -0.05 and 0.25 , respectively, for nonparametric statistics and the superdata methodology. The next greatest MB' accuracy is found

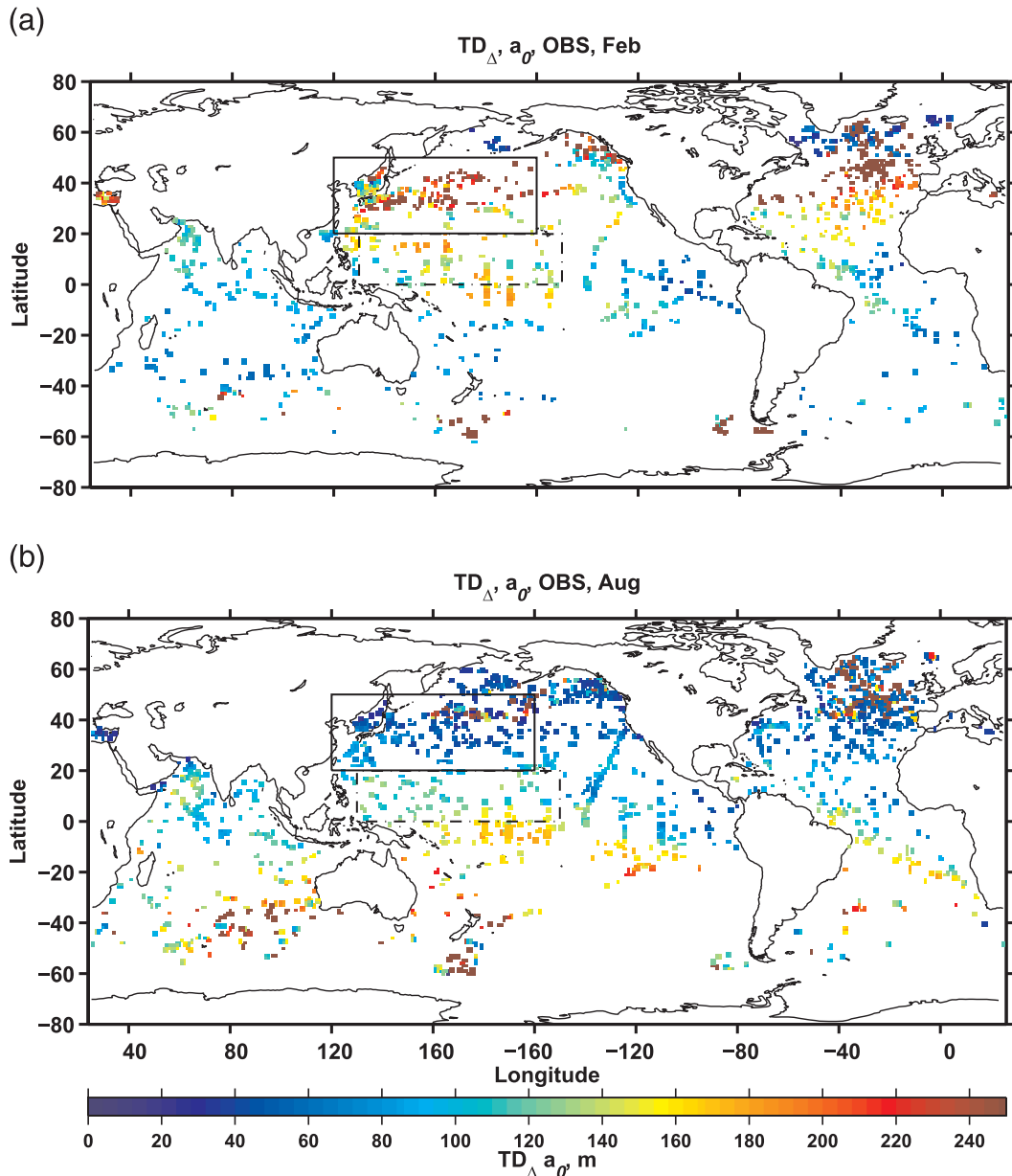


FIG. 9. The fit parameter a_0 (superdata mean; in m) for the observed TD_{Δ} for the global ocean for a 60-day analysis window starting (a) 1 Feb and (b) 1 Aug for the years 2001–03. The rectangles represent the regions for the computations in Fig. 8.

in the results from CLIM and TONLY. This somewhat surprising result is due to climatologies being specifically designed to globally minimize MB' and $RMSE'$. The correlation levels of CLIM and TONLY, however, are lower than the altimeter assimilative MODAS cases. The TONLY estimates are nearly the same as CLIM, suggesting that SST alone has little influence on TD.

In terms of $RMSE'$, all of the synthetics, with the exception of BEST and A0, have nearly the same value. The differences in $RMSE'$ of these synthetics are not

statistically significant. The lowest accuracy in both MB' and $RMSE'$ comes from the zero-altimeter synthetic A0. This is because, without SSHA assimilation, NL0M is not constrained at all by SSHA. These lower accuracy limits of MB' and $RMSE'$ are approximately -0.2 and 0.45 , respectively, for nonparametric statistics and the superdata methodology.

In this paper, the statistical methodologies are found to have a relatively large impact on the results. Evaluation based on traditional statistics that assume Gaussian

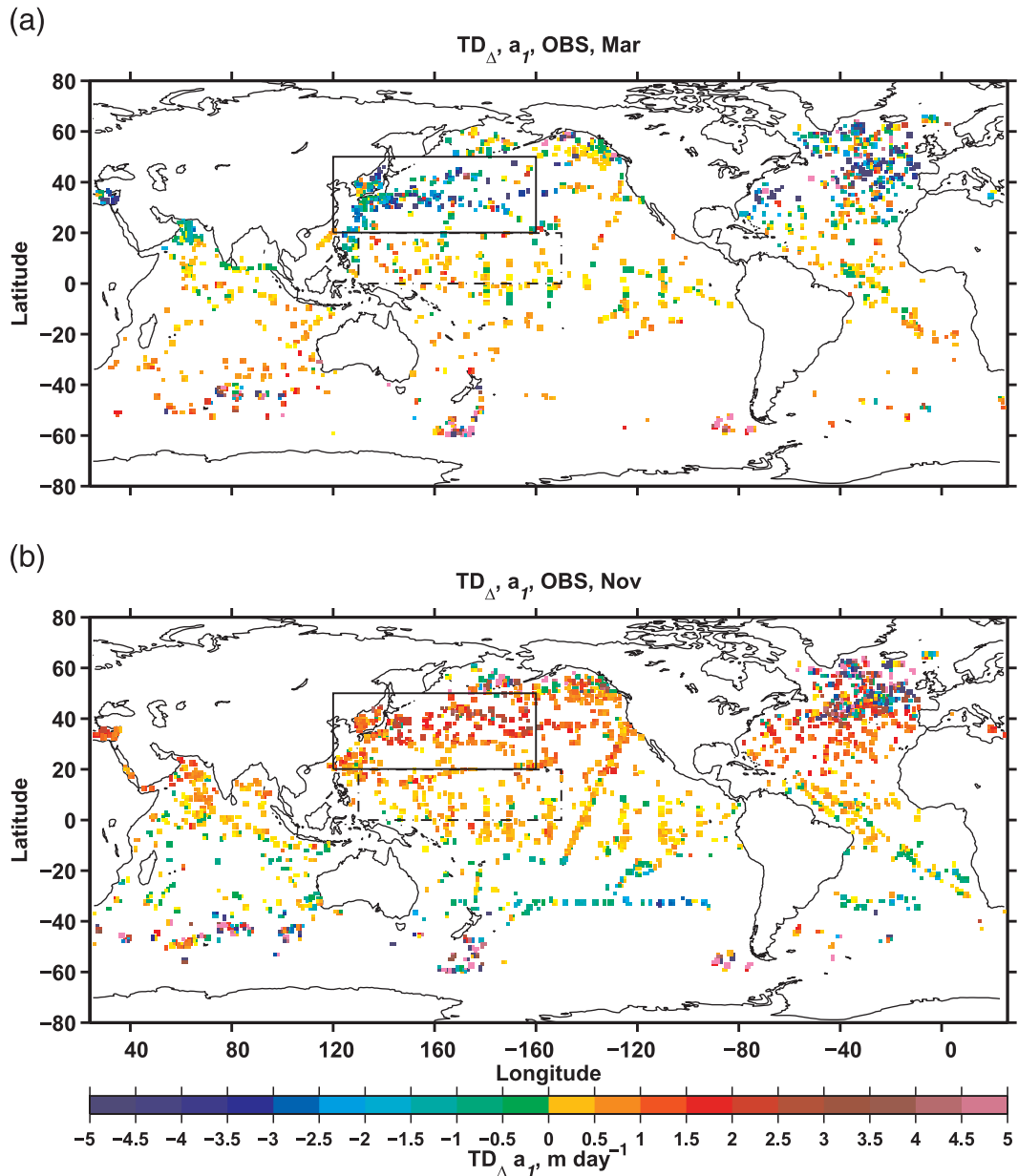


FIG. 10. The fit parameter a_1 (superdata rate of change; in $m\ day^{-1}$) for the observed TD_{Δ} for the global ocean for a 60-day analysis window starting (a) 1 Mar and (b) 1 Nov for the years 2001–03.

data distributions result in inflated error and error uncertainty estimates. Using nonparametric statistics circumvents the need for strict and often unfounded assumptions about the data distribution. The two comparison frameworks used in this analysis have a lesser impact. The effect of the superdata methods is to even the influence of data between low- and high-density observation regions. Because the global error results for the point-for-point framework are slightly larger than for the superdata framework, high-density observation regions tend to occur in areas with larger errors.

The errors associated with SSHA (Fig. 5) are generally larger than the errors found for TD (Fig. 6). This is due to representation errors associated with the SSHA comparison, because there are no direct in situ estimates of SSHA. Instead, we use the difference in observed dynamic height from climatology. Climatology, however, is not identical to the mean sea surface as measured from satellites or associated with NLOM. These misrepresentations add to the errors found in this analysis. Because MODAS is designed to provide realistic three-dimensional temperature and salinity structure, the TD

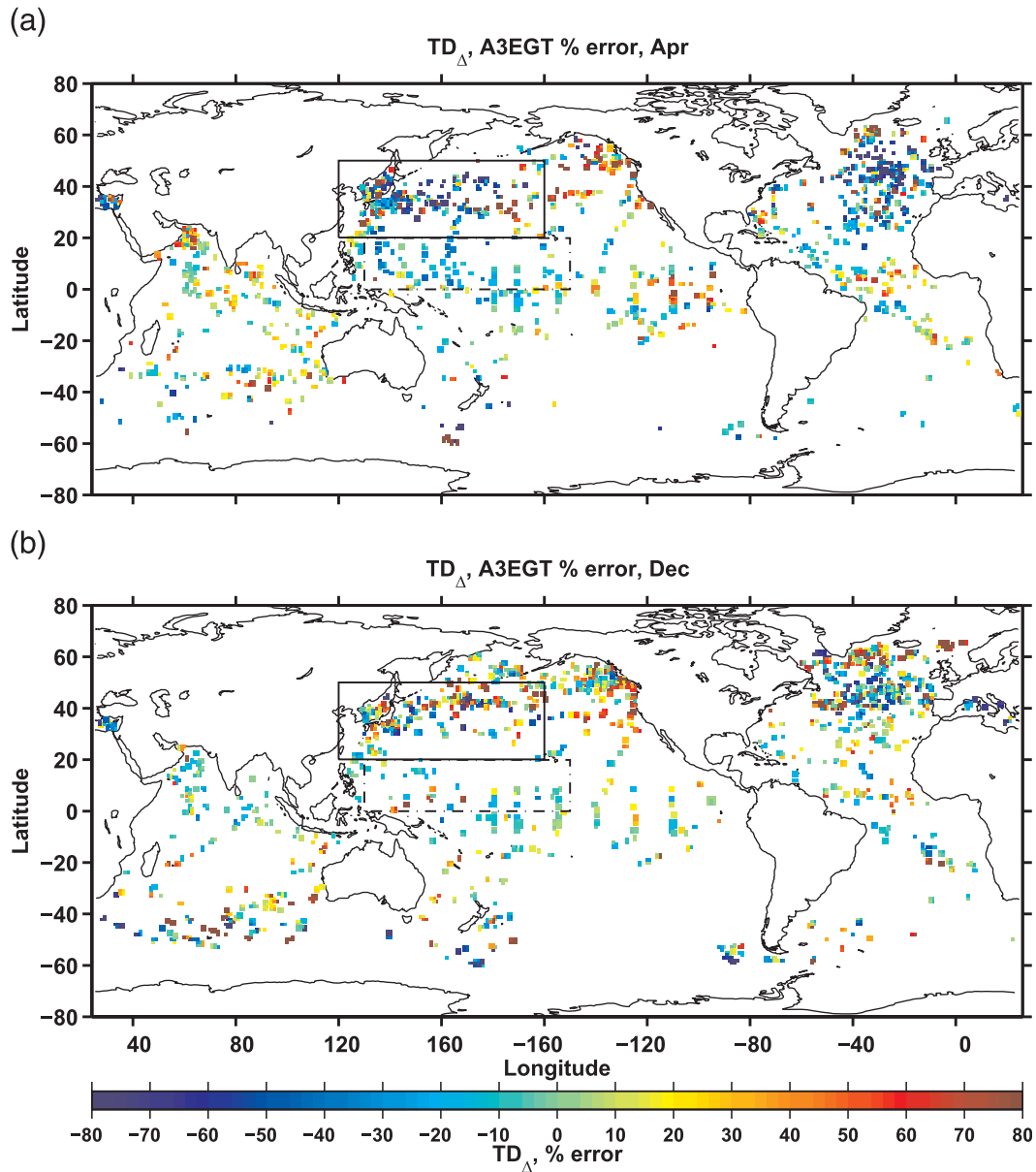


FIG. 11. The percent error for TD_{Δ} A3EGT fit parameter a_0 (superdata mean) for the global ocean for a 60-day analysis window starting (a) 1 Apr and (b) 1 Dec for the years 2001–03.

estimates have greater skill relative to direct in situ profiles observations.

The metric TD_{∇} has poorer accuracy than TD_{Δ} because ILD_{∇} introduces a shallow bias in the ILD estimates. The ILD_{∇} methodology is based on the first curvature peak found in the profile and therefore returns the depth of the purely isothermal layer near the surface. This exacerbates the already shallow bias of MODAS, further degrading the synthetic accuracy. Synthetics and other predictions tend to smooth the sharp gradients found in observations, further reducing the accuracy of both TD_{∇}

and TD_{Δ} . The more unbiased threshold ILD -based estimate TD_{Δ} does not reinforce the shallow bias of MODAS and leads to more accurate TD_{Δ} estimates.

Using summary diagrams, the errors scaled by the observation standard deviation are plotted side by side for consistent comparison. These diagrams are repeated for each statistical methodology and each comparison framework. In each rendition, the results are similar, with the exception that Gaussian statistics inflate the error and error uncertainty levels. The nonparametric statistical methods indicate lower error levels and

larger separation between, for example, A1J and A3EGT synthetics, where the standard error bars do not overlap (Fig. 6). The traditional Gaussian statistical methods are unable to differentiate these cases because this separation is smaller and the standard error bars overlap.

The result that the A1G SSHA estimates tend to have better accuracy than the A1J SSHA estimates is consistent with expected error estimates of Jacobs et al. (2002). The distance between ground tracks in the *GFO* orbit is about 125 km at 35°N with a 17.05-day repeat cycle, whereas the distance between ground tracks in the *Jason-1* orbit is about 260 km at 35°N, with a 9.95-day repeat cycle. The characteristics of the mesoscale tend to have small space and long time scales relative to the altimeter orbit spacing and repeat cycle, respectively (Jacobs et al. 2001). This suggests that the shorter orbit spacing should be an advantage for the A1G SSHA estimate, a hypothesis corroborated in the results of this paper.

Comparing predictions to in situ observations is the ultimate test in that the observations are the closest information we have to the true ocean condition. This type of comparison, however, is inherently plagued with representation errors, because the observations generally have more physical processes influencing the measurements on shorter space and time scales than the predictions are able to resolve.

Acknowledgments. This publication is a contribution to the Assessing UUV Sampling Strategies with Observation System Simulation Experiments project supported by the Office of Naval Research under Program Element 602345N. We thank the National Oceanographic Data Center and T. Boyer for assistance with the World Ocean Database 2005 data, K. Grembowicz of the Naval Oceanographic Office for assistance with the U.S. Navy’s Master Oceanographic Observational Data Set, and J. Dastugue for skillful rendering of figures. Finally, we appreciate the comments of two anonymous reviewers and discussion with J. Cummings and G. Jacobs of the Naval Research Laboratory that have led to important improvements to this manuscript.

APPENDIX

Robust Nonparametric Statistics

To compute the biweight mean, first the median M and the median absolute deviation (MAD) are computed (e.g., Hoaglin et al. 1983). Then, the weights u_j for each of N data values are computed such that

$$u_j = \frac{x_j - M}{cMAD}, \text{ for } j = 1, 2, 3, \dots, N. \quad (A1)$$

The recommended value for the parameter c is 7.5, which corresponds to censoring outliers beyond at least five standard deviations (Lanzante 1996). The censoring is accomplished by using the condition that $u_j = 1.0$ for any j where $|u_j| \geq 1$ (Hoaglin et al. 1983). The biweight estimate of the mean is then

$$\langle x \rangle^{bw} = M + \left\{ \frac{\sum_{j=1}^N (x_j - M)(1 - u_j^2)^2}{\sum_{j=1}^N (1 - u_j^2)^2} \right\}. \quad (A2)$$

It is important to note that this weighting factor has a larger magnitude for larger deviations of x from M . The censoring is accomplished by setting $u_j = 1$ and therefore $1 - u_j^2 = 0$ for deviations from the median M defined by the censor values c . Also note that there is a small typographical error in Lanzante (1996) and that the correction can be found online (at http://www.gfdl.noaa.gov/~jrl/jrl_webpages/manuscripts/resistant/IMPORTANT).

The biweight standard deviation is computed as

$$\langle \langle x \rangle \rangle^{bw} = \frac{\left[N \sum_{j=1}^N (x_j - M)^2 (1 - u_j^2)^4 \right]^{1/2}}{\left| \sum_{j=1}^N (1 - u_j^2)(1 - 5u_j^2) \right|}. \quad (A3)$$

Because Eq. (A3) is an unbiased estimate of the scale of x , this corresponds to the unbiased measure of the standard deviation.

REFERENCES

Barron, C. N., and A. B. Kara, 2006: Satellite-based daily SSTs over the global ocean. *Geophys. Res. Lett.*, **33**, L15603, doi:10.1029/2006GL026356.

—, —, H. E. Hurlburt, C. Rowley, and L. F. Smedstad, 2004: Sea surface height prediction from the global Navy Coastal Ocean Model during 1998–2001. *J. Atmos. Oceanic Technol.*, **21**, 1876–1893.

—, L. F. Smedstad, J. M. Dastugue, and O. M. Smedstad, 2007: Evaluation of ocean models using observed and simulated drifter trajectories: Impact of sea surface height on synthetic profiles for data assimilation. *J. Geophys. Res.*, **112**, C07019, doi:10.1029/2006JC003982.

—, A. B. Kara, and G. A. Jacobs, 2009: Objective estimates of westward Rossby wave and eddy propagation from sea surface height analyses. *J. Geophys. Res.*, **114**, C03013, doi:10.1029/2008JC005044.

Boyer, T. P., and Coauthors, 2006: *World Ocean Database 2005*. NOAA Atlas NESDIS 60, 190 pp.

- Doney, S. C., S. Yeager, G. Danabasoglu, W. G. Large, and J. C. McWilliams, 2007: Mechanisms governing interannual variability of upper-ocean temperature in a global ocean hindcast simulation. *J. Phys. Oceanogr.*, **37**, 1918–1938.
- Efron, B., and R. Tibishirani, 1986: Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat. Sci.*, **1**, 54–77.
- Fan, S., L.-Y. Oey, and P. Hamilton, 2004: Assimilation of drifter and satellite data in a model of the northeastern Gulf of Mexico. *Cont. Shelf Res.*, **24**, 1001–1013.
- Fox, D. N., C. N. Barron, M. R. Carnes, M. Booda, G. Peggion, and J. Van Gurley, 2002a: The Modular Ocean Data Assimilation System. *Oceanography*, **15**, 22–28.
- , W. J. Teague, C. N. Barron, M. R. Carnes, and C. M. Lee, 2002b: The Modular Ocean Data Assimilation System (MODAS). *J. Atmos. Oceanic Technol.*, **19**, 240–252.
- Helber, R. W., C. N. Barron, M. R. Carnes, and R. A. Zingarelli, 2008: Evaluating the sonic layer depth relative to the mixed layer depth. *J. Geophys. Res.*, **113**, C07033, doi:10.1029/2007JC004595.
- Hoaglin, D. C., F. Mosteller, and J. W. Tukey, 1983: *Understanding Robust and Exploratory Data Analysis*. John Wiley and Sons, 447 pp.
- Hurlburt, H. E., 1986: Dynamic transfer of simulated altimeter data into subsurface information by a numerical model. *J. Geophys. Res.*, **91**, 2372–2400.
- , D. N. Fox, and E. J. Metzger, 1990: Statistical inference of weakly correlated subthermocline fields from satellite altimeter data. *J. Geophys. Res.*, **95**, 11 375–11 409.
- Ishikawa, Y., T. Awaji, K. Akitomo, and B. Qiu, 1996: Successive correction of the mean sea surface height by the simultaneous assimilation of drifting buoy and altimetric data. *J. Phys. Oceanogr.*, **26**, 2381–2397.
- Jacobs, G. A., C. N. Barron, and R. C. Rhodes, 2001: Mesoscale characteristics. *J. Geophys. Res.*, **106**, 19 581–19 595.
- , —, D. N. Fox, K. R. Whitmer, S. Klingenberg, D. May, and J. P. Blaha, 2002: Operational altimeter sea level products. *Oceanography*, **15**, 13–21.
- Jolliff, J. K., J. C. Kindle, I. Shulman, B. Penta, M. A. M. Friedrichs, R. Helber, and R. A. Arnone, 2009: Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment. *J. Mar. Syst.*, **76**, 64–82, doi:10.1016/j.jmarsys.2008.05.014.
- Kara, A. B., and H. E. Hurlburt, 2006: Daily inter-annual simulations of SST and MLD using atmospherically forced OGCMs: Model evaluation in comparison to buoy time series. *J. Mar. Syst.*, **62**, 95–119.
- , P. A. Rochford, and H. E. Hurlburt, 2000: An optimal definition for ocean mixed layer depth. *J. Geophys. Res.*, **105**, 16 803–16 821.
- , —, and —, 2003: Mixed layer depth variability over the global ocean. *J. Geophys. Res.*, **108**, 3079, doi:10.1029/2000JC000736.
- , R. W. Helber, T. P. Boyer, and J. B. Elsner, 2009: Mixed layer depth in the Aegean, Marmara, Black and Azov Seas: Part I: General Features. *J. Mar. Syst.*, **78** (Suppl.), 169–180, doi:10.1016/j.jmarsys.2009.01.022.
- Lanzante, J. R., 1996: Resistant, robust and non-parametric techniques for the analysis of climate data: Theory and examples, including applications to historical radiosonde station data. *Int. J. Climatol.*, **16**, 1197–1226.
- Lin, X.-H., L.-Y. Oey, and D.-P. Wang, 2007: Altimetry and drifter data assimilation of loop current and eddies. *J. Geophys. Res.*, **112**, C05046, doi:10.1029/2006JC003779.
- Lorbacher, K., D. Dommenget, P. P. Niiler, and A. Köhl, 2006: Ocean mixed layer depth: A subsurface proxy of ocean-atmosphere variability. *J. Geophys. Res.*, **111**, C07010, doi:10.1029/2003JC002157.
- National Oceanic and Atmospheric Administration, 1986: ETOPO5 digital relief of the surface of the earth. National Geophysical Data Center Data Announcement 86-MGG-07, 19 pp.
- Oey, L.-Y., T. Ezer, G. Forristall, C. Cooper, S. DiMarco, and S. Fan, 2005: An exercise in forecasting loop current and eddy frontal positions in the Gulf of Mexico. *Geophys. Res. Lett.*, **32**, L12611, doi:10.1029/2005GL023253.
- Roemmich, D., S. Riser, R. Davis, and Y. Desaubies, 2004: Autonomous profiling floats: Workhorse for broad-scale ocean observations. *Mar. Technol. Soc. J.*, **38**, 21–29.
- Rousseeuw, P. J., and A. M. Leroy, 1987: *Robust Regression and Outlier Detection*. John Wiley and Sons, 329 pp.
- Shriver, J. F., H. E. Hurlburt, O. M. Smedstad, A. J. Wallcraft, and R. C. Rhodes, 2007: 1/32° real-time global ocean prediction and value-added over 1/16° resolution. *J. Mar. Syst.*, **65**, 3–26.
- Smedstad, O. M., H. E. Hurlburt, E. J. Metzger, R. C. Rhodes, J. F. Shriver, A. J. Wallcraft, and A. B. Kara, 2003: An operational eddy resolving 1/16° global ocean nowcast/forecast system. *J. Mar. Syst.*, **40**, 341–361.
- Teague, W. J., M. J. Carron, and P. J. Hogan, 1990: A comparison between the Generalized Digital Environmental Model and Levitus climatologies. *J. Geophys. Res.*, **95**, 7167–7183.
- Wallcraft, A. J., A. B. Kara, H. E. Hurlburt, and P. A. Rochford, 2003: The NRL Layered Global Ocean Model (NLOM) with and embedded mixed layer submodel: Formulation and tuning. *J. Atmos. Oceanic Technol.*, **20**, 1601–1615.
- Zou, X., and Z. Zeng, 2006: A quality control procedure for GPS radio occultation data. *J. Geophys. Res.*, **111**, D02112, doi:10.1029/2005JD005846.