Naval Research Laboratory

Stennis Space Center, MS 39529-5004

NRL/MR/7320--13-9364

Validation Test Report for the Improved Synthetic Ocean Profile (ISOP) System, Part I: Synthetic Profile Methods and Algorithm

Robert W. Helber
Tamara L. Townsend
Charlie N. Barron
Jan M. Dastugue

Ocean Dynamics and Prediction Branch
Oceanography Division

Michael R. Carnes

QinetiQ North America
Stennis Space Center, Mississippi

March 15, 2013

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
15-03-2013	Memorandum Report	

4. TITLE AND SUBTITLE

Validation Test Report for the Improved Synthetic Ocean Profile (ISOP) System, Part I: Synthetic Profile Methods and Algorithm

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
0601153N

6. AUTHOR(S)

Robert W. Helber, Tamara L. Townsend, Charlie N. Barron, Jan M. Dastugue, and Michael R. Carnes[1]

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
73-4490-03-5

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

Naval Research Laboratory
Oceanography Division
Stennis Space Center, MS 39529-5004

8. PERFORMING ORGANIZATION REPORT NUMBER

NRL/MR/7320--13-9364

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

Office of Naval Research
One Liberty Center
875 North Randolph Street, Suite 1425
Arlington, VA 22203

10. SPONSOR / MONITOR'S ACRONYM(S)

ONR

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

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

[1]QinetiQ North America, Stennis Space Center, MS 39529

14. ABSTRACT

A new method is developed to utilize the relatively abundant satellite observations of sea surface temperature (SST) and sea surface height (SSH) and estimates of mixed layer depth (MLD) to create synthetic profiles of temperature and salinity for assimilation in the Navy's Global Ocean Forecasting System (GOFS). The traditional methods for creating synthetics have limitations particularly with regard to accuracy of the vertical ocean gradients. To address this issue, the Improved Synthetic Ocean Profile (ISOP) system has been developed to more correctly represent the ocean's vertical gradients using a three-layer approach accounting for the ocean's (1) surface mixed layer, (2) vertical gradients below the mixed layer through the thermocline, and (3) relatively quiescent depths. The present document covers the details of the background historical in situ observation database, the three-layer approach, the algorithms for each layer, and the global ISOP stand-alone synthetic validation.

15. SUBJECT TERMS

Ocean modeling
Data assimilation
Probabilistic prediction
Ocean climatology
Ocean forecasting

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Robert Helber
Unclassified Unlimited	Unclassified Unlimited	Unclassified Unlimited	Unclassified Unlimited	125	19b. TELEPHONE NUMBER *(include area code)* (228) 688-5430

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Table of Contents

List of Tables

1. Introduction

Operational ocean data assimilation aims to constrain numerical ocean simulations for accurate forecasts of the ocean for up to 96 hours into the future. Most of the data used for assimilation comes from spaced based observations of sea surface height (SSH) and sea surface temperature (SST). While subsurface temperature and salinity observations are also used, there are far fewer observations with far less geographical coverage below the surface.

For more than a decade and a half the Navy has implemented a method for utilizing the abundant surface observations to create three-dimensional synthetic fields of temperature and salinity for data assimilation. Synthetic vertical profiles are constructed by projecting remotely observed SSH and SST downward from the surface using a global database of statistical relationships. The original synthetic temperature and salinity generation system, which is still in use at the time of writing, is the Modular Ocean Data Assimilation System (MODAS) (Fox, et al. 2002).

Since Navy operations rely heavily on the analysis of ocean acoustics, the vertical structure of the ocean is of particular interest because many ocean acoustic phenomena are governed by the vertical gradients of sound speed. Since sound speed depends upon temperature and salinity, there is a need for accurate prediction of the vertical structure of the ocean temperature and salinity. The MODAS system is designed only to estimate the values of temperature and salinity without constraint on the vertical gradients. Inaccurate performance of the MODAS temperature and salinity vertical gradient structure has motivated efforts to improve the Navy's methods for synthetic profile creation. To correct the quality of the synthetic vertical ocean gradients, a new system was developed as a result of the NRL 6.2 project, Improved Synthetic Ocean Profile (ISOP).

The new ISOP system employs a three layer approach that accounts for the (1) ocean's surface mixed layer, (2) the vertical gradients found below the mixed layer and through the thermocline, and (3) the relative quiescence of the deep ocean. The entire approach is centered on providing correct representation of the ocean's vertical gradients as observed by historical in situ ocean observations. This approach has been described in a patent application that has been submitted (Helber, Carnes and Barron 2011). The present document is the first of two parts that describe the ISOP system and its capabilities.

The ISOP VTR has two parts because there are two modes of operation for generating synthetic ocean profiles. The first mode provides a stand-alone capability for generating ISOP synthetics anywhere in the global ocean. The second mode is where ISOP is used with the Navy's Coupled Ocean Data Assimilation (NCODA) system (Cummings 2005, Smith, et al. 2012), providing synthetics for assimilation in the numerical models the Hybrid Coordinate Ocean Model (HYCOM) (Metzger, et al. 2010) and Navy Coastal Ocean Model (NCOM) (Barron, et al. 2006). While the core methods of the ISOP system for generating ocean synthetic profiles is identical for both modes of operation, there is a substantial amount of extra details associated with running ISOP within a data assimilation system to warrant two separate documents. The present document covers the details of the background historical in situ observation database, the three layer approach, the algorithms for each layer, and the global ISOP stand-alone synthetic validation. The second part "Validation Test Report for the Improved Synthetic Ocean Profile (ISOP) System, Part II: Synthetic Profile Assimilation in Navy Ocean Forecast Systems" covers the details of how ISOP is run in the NCODA system including validation of the full data assimilation of ISOP with HYCOM and NCOM.

Manuscript approved January 31, 2013.

2. Background

Many different approaches to the construction of synthetic temperature and salinity profiles have been published. Multivariate regression between subsurface variables, temperature or salinity anomaly at each depth, and surface variables, steric height anomaly and possibly surface temperature anomaly, were used in (Gilson, et al. 1998) and (Fox, et al. 2002). The primary difficulty with their method is that the steric height anomaly at the surface relative to some deep reference level must be computed for each profile before the regression analysis can be performed. Either a significantly reduced database consisting of only sufficiently deep profiles that have both temperature and salinity must be used or a method must be used to extend profiles and compute salinity when only temperature is available. This problem was removed in (Mellor and Ezer 1991), (Ezer and Mellor 1994), (Ezer and Mellor 1997), (Masina, Pinardi and Navarra 2001), (Fischer 2000), and (Guinehut, et al. 2004) since the required statistics were computed from previous ocean model simulations rather than from observed profiles. The drawback of these methods is that the solution is dependent on the quality of the original simulations that may differ from the true ocean.

Multivariate regression between surface parameters and the amplitudes of empirical orthogonal functions (EOFs) computed from historical temperature and salinity profiles were studied as the basis for predicting synthetic profiles in (Carnes, Mitchell and DeWitt 1990) and (Carnes, Teague and Mitchell 1994). These regression methods are currently being used in the US Navy's MODAS system, which lacks vertical gradient constraint that we wish to include.

The elimination or simplification of the regressions can be accomplished by reducing the number of EOFs to the number of independent variable (surface variables) for cases where the relationship between the dependent variables can be established. In a study by (Pascual and Gomis 2003), EOFs of the vertical structure of dynamic height profiles was computed. Since the vertical gradient of dynamic height versus pressure is the specific volume anomaly, then the amplitudes of dynamic height EOFs also apply to the specific volume anomaly profiles computed from the dynamic height EOFs. Therefore, given the surface dynamic height and the surface specific volume anomaly, the amplitudes of the first two EOFs can be determined.

The most promising approach to creating synthetics from surface observations is the variational method proposed by (Fujii and Kamachi 2003a). The method eliminates regressions and the need to compute steric height anomalies from historical profiles while properly incorporating the expected error statistics. A key feature of the method is the construction of the first guess error covariance matrix from a reduced number of dominant EOF modes computed from historical in situ ocean temperature and salinity profile observations. It is the vertical structure of the coupled temperature and salinity profile EOF modes that are utilized for constructing a vertical covariance matrix useful for projecting the surface ocean information into the subsurface synthetic profiles. (Fujii and Kamachi 2003a) used the vertical covariance matrix in a one-dimensional variational analysis for a 10° by 10° region of the Pacific Ocean east of Japan. It was also applied in the equatorial Pacific in a three-dimensional variational analysis system in (Fujii and Kamachi 2003b).

The Fujii and Kamachi method described above makes use of the historical vertical structure from observations but does not constrain the solution to the vertical gradient of temperature and salinity. The use of truncated EOFs, for example, inherently reduces the ability to represent sharp gradients. To account for sharp vertical gradients at the bottom of the ocean mixed layer, the ISOP system is designed with a three layer format. The first layer is the surface mixed layer that is relatively uniform from the surface down to mixed layer depth (MLD). The MLD is an input to the ISOP system. The second layer is the dynamic layer that represents the ocean from the bottom of the MLD through the thermocline and down to 1000 m. The synthetic solution for this layer is derived using a technique based on the (Fujii

and Kamachi 2003a) approach applied globally and with additional constraints for the vertical gradients of temperature and salinity. The third layer is from 1000 m to the bottom of the ocean is blended from the temperature and salinity values of layer 2 at 1000 m to climatology in the deep ocean. Together, the three ISOP layers provide a complete water column synthetic profile useful for assimilation into numerical models that is highly constrained to accurately represent vertical gradients in the ocean.

3. Data preparation

The goals for the ISOP data aggregation are to obtain the largest collection of in situ observations that will characterize the temperature and salinity vertical structure and vertical gradients of the global ocean. With this data the ISOP temperature and salinity vertical covariance and climatology databases are constructed. The horizontal covariance structure is not required for projecting surface observations information into the subsurface ocean to create synthetics.

The data consist of temperature and salinity profiles combined from the Navy's MOODS (Master Oceanographic Observation Data Set) profile archive (Bauer 1982, Jugan and Beresford 1991) with classified profiles removed, the WOD 2005 (World Ocean Database) (Boyer, et al. 2006, NODC 2009), and delayed-mode Argo (Roemmich and Gilson 2009) profiles downloaded from the USGODAE server (www.usgodae.org/argo/argo.html) on October 11, 2007. Only profiles having both temperature and salinity were used. Argo profiles were first removed from MOODS and WOD 2005. Then, a duplicate check was performed that removed any profiles from MOODS that appeared in the WOD 2005 data set. The final combined data set consisted of 3,506,496 profiles. The number of profiles measured by each of 22 different categories of instruments is listed in the second column of Table 1. The first column of Table 1 is an instrument code number used in MOODS.

The full data set of profiles was examined using a GUI-based editor. The editor displayed all profiles in geographical blocks up to 10º of latitude by 10º of longitude in size. Smaller subsets of profiles, grouped by position and time of year within the largest block, could be examined in an automated sequence of profile overlays of temperature versus depth, salinity versus depth, potential density versus depth, potential density versus salinity, temperature versus potential density, and salinity versus potential density. Outliers on each plot could then be easily identified and then marked for deletion by the operator. The editing GUI also allowed profiles to be individually edited, but this capability was not used often due to the large amount of time required if applied to the full data set. The manual editing identified and removed 668,416 profiles. The number of profiles of each measurement type in this first set of deletions is listed in Table 2. Altogether, 668,416 (19.1 % of the total) profiles were deleted from the database up to this point. Next, all profiles where the deepest observation depth was less than 200 m, but the bottom bathymetry depth was greater than 400 m were removed. This step removed another 172,105 (4.9 % of the original total) profiles. The final data set contained 1,969,081 profiles containing both temperature and salinity from the various instrument types as shown in columns 6 of Table 2.

The geographical distributions of temperature and salinity profiles are displayed in Figure 1, as the $\log_{10}$ of the number of profiles within each 5º x 5º geographic block. The distribution of profiles in terms of their year of measurement is shown in Figure 2. Because the MBTs were eliminated from the final data set, the number of temperature and salinity profiles is about equal in the years before 1960, but after XBTs began to be used in the 1960s, the number of temperature-only profiles far outnumbered profiles with both temperature and salinity. The relative number of profiles with both temperature and salinity has been steadily rising since the late 1990s due to the use of Argo floats in deep water.

The final step in the preparation of the profile data set interpolated each profile to a set of 78 fixed depths between 0 m and 6600 m using the piece-wise cubic interpolating polynomial interpolator

e.g. (Fritsch and Carlson 1980), similar to the Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) interpolating method in MATLAB and GNU OCTAVE. The advantage of the pchip technique is that over intervals of the original data that are monotonically increasing or decreasing, the interpolated curve is also monotonic in the same direction. Also, the interpolated curve contains local extremes only at points where the original data do. As a result, overshoots that might occur with splines near sudden gradient changes in the original data are eliminated. The 78 standard depths are listed by spacing depth regimes in Table 3.

4. Three-layer synthetic profile model

The purpose of the ISOP system is to use surface ocean observations to construct synthetic profiles of temperature and salinity that most accurately represent the unobserved subsurface ocean conditions. The inputs ISOP is designed to accept include surface observations or estimates of temperature and height and their uncertainties, mixed layer depth, and a first-guess subsurface temperature and salinity profile and their uncertainties. All inputs, their description, and likely sources are listed in Table 4 ISOP Inputs. Given appropriate inputs, ISOP is capable of creating a synthetic temperature (T) and salinity (S) profile for any date at any ocean latitude and longitude. At any one time or place any or all observation may not be available when a synthetic is required. For this reason, the ISOP system has the flexibility that allows for any or all of these inputs to be absent. In the case of no inputs, the output is the ISOP climatology T and S profile for the month and location of the requested synthetic. As shown in Table 4, each input value also has a specified value of uncertainty except for the mixed layer depth.

Synthetic profiles are constructed by combining separate profile segments from three adjacent layers to output a complete temperature and salinity profile with its associated uncertainty (Table 5). The top layer (Layer 1) extends from the surface to the input MLD, and is computed by linear regressions dependent on the MLD and the profile gradient at the MLD. The second or middle layer (Layer 2) extends from the MLD to either the ocean bottom or to 1000 m, whichever is smaller. The synthetic temperature and salinity profiles in the second layer are computed using a one-dimensional variational technique using inputs of SST, SSHA, MLD, and first-guess estimates of temperature (T) and salinity (S) profile pairs. The third layer (Layer 3) extends from 1000 m to the ocean bottom, using linear regression coefficients for blending Layer 2 synthetics towards climatology with depth.

All three layers of the ISOP system utilize a static database to construct synthetics anywhere for any month in the global ocean. The top and bottom layer databases contain only regression coefficients stored on a ½° global grid. The middle, dynamic layer, database has covariance information, constructed from an m-mode EOF representation of the ocean vertical structure, also on a ½° global grid that is used to minimize a cost-function. The databases for the three layers are combined with the input data to create the synthetics. In section 4.1, the surface mixed layer methods and database construction is described. In section 4.2, the Layer 2 cost function is defined and derived. Cost function minimization is performed by first setting the differentiated cost function equal to zero analytically. Section 4.3 describes the synthetic profile below 1000 m.

4.1. Layer 1, Surface mixed layer density model

The top ISOP layer, Layer 1, extends from the ocean surface down to the input value of MLD. ISOP does not associate the MLD with an uncertainty and therefore defines Layer 1 depth to be exactly equal to the input MLD. The structure of the Layer 1 synthetic profile is computed in two ways depending on the application of ISOP. One option is to shift the input first guess profile to match the estimated surface potential density at a depth of 4 m, thereby retaining the mixed layer structure of the

first guess profile. The shifted profile is then used as the synthetic Layer 1 solution. The second option is to construct the Layer 1 temperature and salinity using a model derived from historical observations.

4.1.1. Option One

In cases where it is desirable for ISOP Layer 1 to match the structure of the first-guess profile, an option exist where ISOP will shift and interpolate the first-guess profile to match the input SST value and the Layer 2 synthetic solution at the MLD. This option is selected using the ISOP software flag, model_ml set equal to true.

4.1.2. Option Two

For this option, the Layer 1 synthetic profile is required to match the estimated surface potential density at a depth of 4 m. The Layer 1 potential density profile, $\sigma_\theta(z)$, is constructed in such a way that it matches the Layer 2 potential density and vertical gradient of potential density, $G_{MLD} = \partial\sigma_\theta/\partial z\big|_{z=MLD}$, at the MLD. For this reason, the Layer 2 synthetic profile described in the next section is computed first.

The method used in ISOP for determining the MLD from a profile of temperature and salinity is the potential-density-threshold method. The threshold method defines the MLD as the depth below a reference depth where the potential density is greater by a specified amount than at the reference depth. Use of a 4-m subsurface reference depth, rather than using the surface, is intended to isolate the determination of MLD from the diurnal variations of near-surface temperature. For climatological profiles, (Kara, Rochford and Hurlburt 2000) used a 10-m reference depth. We use a reference depth of 4 m so as not to eliminate the possibility of very shallow mixed layers that exist in model and observational data. A threshold density difference, $\Delta\sigma_{\theta,Thresh}$ = 0.15 kg/m^3, is used unless the computed MLD is greater than 400 m. Then, the threshold value is reduced to 0.05 kg/m^3. If the MLD computed using this new threshold is still greater than 400 m, then it is reduced to 0.025 kg/m^3, and then to 0.01 kg/m^3 and finally to 0.001 kg/m^3, until the MLD is less than 400 m. The Layer 1 potential density profile is then constructed with the same change in density between 4 m and the MLD as the final threshold value.

The statistical model for the ISOP mixed layer (Layer 1) was developed by first examining the relationship between the potential density anomaly and the density gradient at the MLD. The potential density anomaly is scaled by the potential density threshold value from the MLD calculation, $\sigma_{\theta,Thresh}$, and is given by

$$\Delta\sigma_\theta(z) = \left\{\sigma_\theta(z) - \sigma_\theta(MLD)\right\}\Big/\sigma_{\theta,Thresh}, \tag{1}$$

which is a function of depth. The density gradient at the mixed layer is also scaled by the potential density threshold value

$$G_{MLD} = \partial\sigma_\theta/\partial z\big|_{z=MLD}\Big/\sigma_{\theta,Thresh}. \tag{2}$$

The quantities given by equation (1) and (2) are investigated at several locations around the world.

Comparisons indicated that the scaled potential density anomalies, $\Delta\sigma_\theta(z)$, grouped at constant scaled depths, *z'=z/MLD*, are approximately proportional to the product, $G_{MLD}MLD$. For example, Figure 3 shows a scatter plot of points colored by the value, $\Delta\sigma_\theta(z')$, at a scaled depth of z' = 0.812 on a graph of G_{MLD} versus *MLD* determined from 398 profiles surrounding the position at 30° N and 140° E. In this plot, the scaled density anomaly is about -1.0 (dark blue) near the surface and about zero (red) near the MLD.

There is an approximate inverse relationship between G_{MLD} and MLD for fixed potential density anomaly and depth. This can be seen in Figure 3 by looking at the pattern dots of a similar color follow. For example, yellow dots of equal potential density anomaly have small (large) potential density gradient at large (small) MLD. The relationship is looser at small values of potential density anomaly or as the scaled depth, z' (not shown here), approaches unity.

Since the inverse relationship is only approximate, a more accurate estimate was made by fitting observations of scaled potential density anomaly, $\Delta\tilde{\sigma}$, to a polynomial equation at each global grid location given by

$$\Delta\sigma_\theta\left(z'_k, G_{MLD}, MLD\right) = a_{1,k} + a_{2,k}G_{MLD} + a_{3,k}G_{MLD}^2 + a_{4,k}MLD + a_{5,k}MLD^2 + a_{6,k}G_{MLD}MLD + a_{7,k}G_{MLD}^2MLD + a_{8,k}G_{MLD}MLD^2 \quad (3)$$

The solution is obtained by minimizing the sum,

$$\sum_{i=1}^{N}\left(\Delta\sigma_\theta\left(z'_k, G_{MLD,i}, MLD_i\right) - \Delta\tilde{\sigma}_{\theta,i}\left(z'_k\right)\right)^2, \quad (4)$$

with respect to each of the coefficients, a_1 through a_8, where *k* is the depth index and *i* is the profile index. The *i*-th observed profile of scaled potential density anomalies are given by

$$\Delta\tilde{\sigma}_{\theta,i}\left(z'_k\right) = \left\{\tilde{\sigma}_{\theta,i}\left(z'_k\right) - \tilde{\sigma}_{\theta,i}\left(1\right)\right\}\Big/\sigma_{\theta,Thresh}, \quad (5)$$

where the threshold value, $\sigma_{\theta,Thresh}$, may be different for each measured profile.

The set of coefficients was determined at each of the K scaled depths determined by the formula

$$z'_k = \left(1 + \log_{10}\left[0.1 + 0.05(k-1)\right]\right)\Big/\left(1 + \log_{10}\left(1.1\right)\right), \text{ k = 1, K.} \quad (6)$$

The profile defined by equation (6) provides wider spacing near the surface and finer spacing near the MLD ($z' = 1$). The fit of this equation to the observations is represented by the contours from -0.1 to 1.0 at intervals of 0.1 in Figure 3.

The Layer 1 potential temperature and salinity profiles are computed from the Layer 1 synthetic potential density profile in the mixed layer using simple relationship computed by linear regression from historical profiles. The temperature coefficient is computed at each point of the global grid by minimizing the sum of differences,

$$\sum_{i=1}^{N}\left[\left(\theta_i\left(z'\right) - \theta_i\left(MLD\right)\right) - \left(\hat{\theta}_i\left(z'\right) - \hat{\theta}_i\left(MLD\right)\right)\right]^2 \quad (7)$$

with respect to the coefficient, a_T, where,

$$\theta_i\left(z'\right) - \theta_i\left(MLD\right) = a_T\left(z'\right)\left(\hat{\sigma}_{\theta,i}(z') - \hat{\sigma}_{\theta,i}(MLD)\right). \quad (8)$$

The salinity coefficient is computed by minimizing

$$\sum_{i=1}^{N}\left[\left(S_i\left(z'\right) - S_i\left(MLD\right)\right) - \left(\hat{S}_i\left(z'\right) - \hat{S}_i\left(MLD\right)\right)\right]^2 \quad (9)$$

with respect to the coefficient, a_S, where,

$$S_i\left(z'\right) - S_i\left(MLD\right) = a_S\left(z'\right)\left(\hat{\sigma}_{\theta,i}(z') - \hat{\sigma}_{\theta,i}(MLD)\right). \quad (10)$$

Once the ISOP Layer 1 databases have been constructed, the potential density profile can first be computed from the MLD, the threshold potential temperature difference, the vertical gradient at the MLD, the surface potential density, and the coefficients in Equation 3. Next, the potential temperature, θ, and salinity, S, are computed at each scaled depth, z', from the synthetic potential density anomaly according to,

$$\theta(z') = \theta(MLD) + a_T(z')\left(\sigma_\theta\left(z'\right) - \sigma_\theta(MLD)\right) \quad (11)$$

$$S(z') = S(MLD) + a_S(z')\left(\sigma_\theta\left(z'\right) - \sigma_\theta(MLD)\right). \quad (12)$$

The last step is then to compute the temperature from the potential temperature, salinity, and pressure, and to then interpolate the profiles from the scaled depths to depth using the MLD.

Typical examples of the regression between the potential temperature anomalies and potential density anomalies at two different scaled depths, at the same geographic location as used in Figure 3, are shown in Figure 4. Figure 4a shows the data and regression results at a scaled depth of about 72% of the MLD and Figure 4b shows the results at a depth of about 94% of the MLD. Generally, the correlation between the temperature and potential density anomalies increases as the depth approaches the MLD. At this location, changes in density are due mainly to temperature changes so that the correlation between salinity anomalies and potential density anomalies are generally low as shown in Figure 5.

4.2. Layer 2 - Dynamic climatology approach

Synthetic temperature and salinity profiles in the second layer (Layer 2, from the mixed layer to 1000 m) are computed using a one-dimensional variational technique using inputs of SST, SSHA, MLD, and first-guess estimates of temperature (T) and salinity (S) profile pairs. The solution is obtained directly by performing the following steps:

1) Define the cost function (described in this section)
2) Differentiate the cost function with respect to the unknown solution of T and S and their EOF amplitudes
3) Set the differentiated cost function equal to zero
4) Put the resulting set of equations, including the input data values, in the form $\mathbf{D} \cdot \mathbf{g} = \mathbf{e}$ (equation 48, below)
5) Invert the $\mathbf{D}$ matrix using upper and lower matrix decomposition

Steps 1) through 3) are performed analytically once and described in detail in this section. Steps 4 and 5 are performed (by ISOP software) every time a synthetic is required. Using ISOP software, the matrix equation $\mathbf{D} \cdot \mathbf{g} = \mathbf{e}$ is constructed (using the analytical solution derived in steps 1) through 3) when the inputs (SST, SSHA, MLD, first-guess T and S) for a given location and time are provided. The result is a synthetic ocean profile for that time and location.

All inputs, except for MLD, must also have corresponding uncertainty estimates (see Table 4). The cost function constraints, in addition to the input data, included climatological T and S values and an m-mode EOF representation of the ocean vertical structure as derived from historical observations. The climatology and EOF representation of the ocean are computed from the historical observations described in Section 3 above.

The ISOP climatology is constructed using similar methods as those used for constructing the Generalized Digital Environmental Model (GDEM) version 4 (Carnes, Helber, et al. 2010). The purpose of the ISOP climatology is to provide a mean and standard deviation of the ocean. These quantities are used to construct the correlation matrix employed in the EOF analysis, which is described in Sections 4.2.1 and 5.2.3. The primary difference between the ISOP climatology and GDEM is that ISOP only uses paired profiles of T and S whereas GDEM uses all T profiles available. Paired profiles of T and S are required for ISOP because the T and S co-variability is important for accurate representation of the dynamic structure of the ocean.

4.2.1. The Cost Function Constraints

The approach used to construct synthetic temperature and salinity profiles within the middle layer (below the mixed layer; Layer 2) is derived by minimizing a cost function. The cost function J is minimized with respect to the synthetic temperature (T) and salinity (S) profile pairs and two sets of EOF amplitudes. The cost function with ten terms labeled is given by:

$$
\begin{aligned}
J = & \underbrace{(x - x_{cl})^T B^{-1} (x - x_{cl})}_{J_1} + \underbrace{(d - d_{cl})^T B_g^{-1} (d - d_{cl})}_{J_2} \\
& + \underbrace{(x_{fg} - x)^T R^{-1} (x_{fg} - x)}_{J_3} + \underbrace{(d_{fg} - d)^T R_g^{-1} (d_{fg} - d)}_{J_4} \\
& + \underbrace{\sum_{i=1}^{N} \left(\frac{T_i' - \hat{T}_i'}{u_i} \right)^2}_{J_5} + \underbrace{\sum_{i=1}^{N-1} \left(\frac{\Delta T_i' - \left(\hat{T}_{i+1}' - \hat{T}_i' \right)}{w_i} \right)^2}_{J_6} \\
& + \underbrace{\sum_{i=1}^{N} \left(\frac{S_i' - \hat{S}_i'}{u_{i+N}} \right)^2}_{J_7} + \underbrace{\sum_{i=1}^{N-1} \left(\frac{\Delta S_i' - \left(\hat{S}_{i+1}' - \hat{S}_i' \right)}{w_{i+N-1}} \right)^2}_{J_8} \\
& + \underbrace{\frac{\left(\tilde{T}_{MLD}' - \hat{T}_{MLD}' \right)^2}{\varepsilon_{SST}^2}}_{J_9} + \underbrace{\frac{\left(\tilde{h}_{MLD} - \hat{h}_{MLD} \right)^2}{\varepsilon_h^2}}_{J_{10}}
\end{aligned}
\tag{13}
$$

Since equation 13 is long with many variables, Table 6 contains a brief explanation of each symbol. Before the details of each term of equation 13 are described, however, the following paragraph explains the general purpose of each term.

Term J_1 constrains the synthetics to the ISOP climatological T and S field while term J_2 constrains the synthetics to the vertical gradient of the ISOP climatological T and S field. Terms J_3 and J_4 constrain the synthetics to the first-guess T and S field and the vertical gradient of the first-guess T and S field. Terms J_5, J_6, J_7 , and J_8 constrain the synthetics to the m-mode EOF representation of the T and S field and the m-mode EOF representation of the vertical gradient T and S field (for the prototype ISOP system, $m=6$). Term J_9 constrains the synthetics to the observed temperature at the mixed layer depth while term J_{10} constrains dynamic height of the synthetics to the observed sea surface height anomaly. The following paragraphs describe in detail all the terms in equation 13.

Terms J_1, J_2, J_3, and J_4 are related and each contain either the vectors x or d. The analysis $\hat{T}$ and $\hat{S}$ variables (or the synthetic solution variables being sought) are represented by the vector x, given by

$$
x = \left[\hat{T}_1, \hat{T}_2, \hat{T}_3, \cdots, \hat{T}_N, \hat{S}_1, \hat{S}_2, \hat{S}_3, \cdots, \hat{S}_N \right]. \tag{14}
$$

The top half of x are temperature from 1 to N and the bottom half from $N+1$ to $2N$ are the salinity deviations from climatology, where N is the number of depth levels in the profile. The vector d contains the analysis $\hat{T}$ and $\hat{S}$ variable differences between consecutive depth levels

$$d = \left[\hat{T}_2 - \hat{T}_1, \hat{T}_3 - \hat{T}_2, \cdots, \hat{T}_N - \hat{T}_{N-1}, \hat{S}_2 - \hat{S}_1, \hat{S}_3 - \hat{S}_2, \cdots, \hat{S}_N - \hat{S}_{N-1} \right] \tag{15}$$

The top values from *1* to *N-1* are the temperature differences and the bottom values from *N* to *2N-2* are the salinity differences. The vectors x_{cl}, d_{cl}, x_{fg}, and d_{fg} are of the same form as x and d but represent the climatology and first-guess fields. The climatology field (T_{cl} and S_{cl}) is the observed climatological mean computed from ISOP historical temperature and salinity paired profiles. The first-guess field (T_{fg} and S_{fg}) is intended to be input from a reasonably accurate model forecast.

Term J_1 constrains the synthetics to the climatology while preserving the covariance structure given by B, which is the covariance of the vertical structure of the ocean derived from in situ observations. The covariance matrix B is constructed from the correlations of in situ observations in the form of a correlation matrix such that

$$B = UCU \tag{16}$$

where C is the correlation matrix and U is a diagonal standard deviation matrix with elements equal to the square root of the diagonal elements of B. The correlation matrix, C, is computed from historical observations of temperature and salinity profiles for the global ocean and is described in Section 5.2 below.

Our approach for computing B from in situ observations employs correlations rather than covariances because the large difference in the variances of temperature and salinity, and the large change in variances versus depth. Without the normalization of the correlation matrix, the large temperature variance and the larger near surface variance would dominate the variability represented by the first several EOFs, thus losing much of the salinity variability and deep variability in the discarded EOFs. Substituting correlation for covariance in EOF analyses is a common approach (e.g. Wilks, 2006). The final form of our equations uses a subset of m EOFs, those with the largest eigenvalues. The prototype version of ISOP has m=6.

To remove the expensive inverse calculation of C, the Jordan decomposition, a special case of singular value decomposition when applied to symmetric matrices (Hardle and Simar, 2003) is applied to C in equation (16), resulting in,

$$B = U\Gamma\Lambda\Gamma^T U\,, \tag{17}$$

where Λ is a diagonal matrix with elements, λ_i ,equal to the singular values of C. Since C is a correlation matrix, the singular values are also the eigenvalues (or EOFs) of C. The columns of the orthogonal matrix, Γ, are the combined *T-S* eigenvectors, γ_i , of C. The inverse of B is then just,

$$B^{-1} = U^{-1}\Gamma\Lambda^{-1}\Gamma^T U^{-1}\,, \tag{18}$$

where the inverses of the diagonal matrices are just diagonal matrixes with elements equal to the inverses of the original elements. Defining a new vector of EOF amplitudes,

$$a = \Gamma^T U^{-1}(x - x_{cl}) \tag{19a}$$

or written in summation form

$$a_i = \sum_{k=1}^{2N} \gamma_{ik} u_k^{-1} \left(x_k - x_{cl,k} \right) \tag{19b}$$

where 2N is the number of depths from 0 to 1000 m (N=47) for both temperature and salinity, we obtain for the first term J_1 of equation (13),

$$J_1 = \left(x - x_{cl}\right)^T B^{-1} \left(x - x_{cl}\right) = a^T \Lambda^{-1} a = \sum_{i=1}^{m} a_i^2 / \lambda_i\,, \tag{20}$$

where a_i are the elements of a, u_k are the elements of U, and m is the number of EOFs retained for the analysis. For the first ISOP prototype system $m=6$. Since the a amplitudes are part of the synthetic solution, the a_i elements are unknown and are the solution obtained by the minimization of the cost function.

Term J_3 constrains the synthetics to the first guess while preserving the covariance structure given by R. If these were observed profiles rather than model forecast profiles, the observation error covariance matrix would typically contain non-zero terms only along the diagonal. Model profile errors, however, are expected to be correlated among depths, and we simply assume that R is proportional to B, i.e., $R=B/F_{fg}$ and $R^{-1}=F_{fg}B^{-1}$. Then, term J_3 in equation (13) can be written as

$$J_3 = \left(x_{fg} - x\right)^T R^{-1} \left(x_{fg} - x\right) = F_{fg} \sum_{i=1}^{m} \left(a_i^2 + a_{fg,i}^2 - 2a_i a_{fg,i}\right) / \lambda_i \,, \tag{21}$$

using $x_{fg} - x = (x_{fg} - x_{cl}) - (x - x_{cl})$, and where $a_{fg,i}$ is the ith EOF amplitude of the vector

$$a_{fg} = \Gamma^T U^{-1} (x_{fg} - x_{cl}) \,, \tag{22a}$$

which is similar to equation 19, but computed from the input first-guess temperature and salinity profiles. Since a_{fg} is computed from the input first-guess profiles using

$$a_{fg,i} = \sum_{k=1}^{2N} \gamma_{ik} u_k \left(x_{fg,k} - x_{cl,k}\right) \tag{22b}$$

the first-guess amplitudes are an input to the minimization. They are computed before the minimization of the cost function.

By the same procedure, the inverse covariance matrix of the vertical difference errors, B_g^{-1}, in term J_2, is transformed into

$$B_g^{-1} = W^{-1} \Phi M^{-1} \Phi^T W^{-1} \,, \tag{23}$$

where M^{-1} is the diagonal matrix of inverses of the eigenvalues (with elements μ_i), Φ is the matrix of eigenvectors (with elements $\phi_{i,j}$, for the i-th depth of the j-th EOF), and W is the diagonal matrix of the standard deviations of vertical differences (with elements w_k w_i down the diagonal). Similarly to equation 22, the amplitudes are defined as

$$b = \Phi^T W^{-1} \left(d - d_{cl}\right) \tag{24a}$$

Or

$$b_i = \sum_{k=1}^{2N} \phi_{ik} w_k \left(d_k - d_{cl,k}\right). \tag{24b}$$

As with the U matrix, the upper half of W contains the temperature vertical difference standard deviations and the lower half contains the salinity vertical difference standard deviations. In addition, we assume as before with B and R, that B_g is proportional to R_g, i.e., $R_g=B_g/G_{fg}$ and $R_g^{-1}=G_{fg}B_g$, where G_{fg} is a constant. Terms J_2 and J_4 can them be written as:

$$J_2 = \left(d - d_{cl}\right)^T B_g^{-1} \left(d - d_{cl}\right) = \sum_{i=1}^{m} b_i^2 \,/\, \mu_i \tag{25}$$

$$J_4 = \left(d_{fg} - d\right)^T R_g^{-1} \left(d_{fg} - d\right) = G_{fg} \sum_{i=1}^{m} \left(b_i^2 + b_{fg,i}^2 - 2b_i b_{fg,i}\right) / \mu_i \tag{26}$$

Since the vertical difference amplitudes b_i are amplitudes of the synthetic solution, the elements are unknown and are the solution obtained by the minimization of the cost function. Since b_{fg} is computed

from the input first-guess profiles, they are an input to the minimization. They are computed before the minimization of the cost function.

For the terms J_5, J_6, J_7, and J_8 in equation (13), the m-mode EOF representation of temperature, T and salinity, S, is use to constrain the synthetics. The temperature anomaly components, $T_i' = T_i - T_{cl,i}$, and salinity anomaly components, $S_i' = S_i - S_{cl,i}$, are used rather than using their combined form as components of x. Similarly, we will use $\Delta T_i' = \Delta T_i - \Delta T_{cl,i}$ and $\Delta S_i' = \Delta S_i - \Delta S_{cl,i}$, where the vertical difference is given by $\Delta(\ )_i = (\ )_{i+1} - (\ )_i$, rather than their combined form as components of d, above. Then, the temperature and salinity anomalies and their vertical differences anomalies computed from the EOFs are

$$T_k' = \sum_{i=1}^{m} u_k a_i \gamma_{k,i} \tag{27}$$

$$S_k' = \sum_{i=1}^{m} u_{k+N} a_i \gamma_{k+N,i} \tag{28}$$

$$\Delta T_k' = \sum_{i=1}^{m} w_k b_i \varphi_{k,i} \tag{29}$$

$$\Delta S_k' = \sum_{i=1}^{m} w_{k+N-1} b_i \varphi_{k+N-1,i} \ . \tag{30}$$

The analysis or the solution being sought for the synthetic temperature and salinity profiles is given in equation (13) by $\hat{T}_i$ and $\hat{S}_i$. The terms J_5, J_6, J_7, and J_8 of the cost function (13) attempt to keep the final synthetic temperature and salinity as well as their vertical differences close to the values computed from the EOFs.

By inserting equations (27), (28), (29), and (30) into terms J_5, J_6, J_7, and J_8 of (13), we obtain

$$J_5 = \sum_{i=1}^{N} \left(\frac{T_i' - \hat{T}_i'}{u_i} \right)^2 = \sum_{i=1}^{N} \left(\frac{\sum_{k=1}^{m} \left(u_i a_k \gamma_{i,k} \right) - \hat{T}_i'}{u_i} \right)^2 \tag{31}$$

$$J_6 = \sum_{i=1}^{N-1} \left(\frac{\Delta T_i' - \left(\hat{T}_{i+1}' - \hat{T}_i' \right)}{w_i} \right)^2 = \sum_{i=1}^{N-1} \left(\frac{\sum_{k=1}^{m} \left(w_i b_k \varphi_{i,k} \right) - \left(\hat{T}_{i+1}' - \hat{T}_i' \right)}{w_i} \right)^2 \tag{32}$$

$$j_7 = \sum_{i=1}^{N} \left(\frac{S_i' - \hat{S}_i'}{u_{i+N}} \right)^2 = \sum_{i=1}^{N} \left(\frac{\sum_{k=1}^{m} \left(u_{i+N} a_k \gamma_{i+n,k} \right) - \hat{S}_i'}{u_{i+N}} \right)^2 \tag{33}$$

$$J_8 = \sum_{i=1}^{N-1} \left(\frac{\Delta S_i' - \left(\hat{S}_{i+1}' - \hat{S}_i' \right)}{w_{i+N-1}} \right)^2 = \sum_{i=1}^{N-1} \left(\frac{\sum_{k=1}^{m} \left(w_{i+N-1} b_k \varphi_{i+N-1,k} \right) - \left(\hat{S}_{i+1}' - \hat{S}_i' \right)}{w_{i+N-1}} \right)^2 \tag{34}$$

Term J_9 of equation (13) will be solved in terms of the analysis temperature and salinity values, $\hat{T}$ and $\hat{S}$. The synthetic profile temperature at the mixed layer depth can be written as a weighted sum of the temperature profile components, so that term J_9 can be written as

$$J_9 = \frac{\left(\tilde{T}'_{MLD} - \hat{T}'_{MLD}\right)^2}{\varepsilon^2_{SST}} = \frac{\left(\tilde{T}'_{MLD} - \sum_{i=1}^{N} Q_i \hat{T}'_i\right)^2}{\varepsilon^2_{SST}} \tag{35}$$

where Q is a linear interpolation vector which give the interpolated $\hat{T}$ value at the MLD. Q is zero except at the two depths bounding the MLD. This term constrains the synthetic temperature toward the specified temperature at the base of the mixed layer.

The cost function constrains the steric height anomaly of the synthetic towards the input surface value of SSHA. Term J_{10} of equation (13) applies this constraint in terms of steric height at the MLD referenced to 1000 m depth. The input surface SSHA is projected to the MLD by subtracting an estimate of the steric height anomaly at the surface relative to the MLD

$$\hat{h}_{MLD} = \hat{h} - h_{0 \to MLD}\,. \tag{36}$$

The input SSHA is denoted by $\hat{h}$ and the steric height anomaly at the surface relative to the MLD given by $h_{0 \to MLD}$ is computed using either the first-guess profile or constant T and S values in the mixed layer, depending on the Layer 1 option (see Section 4.1).

The analysis steric height at the MLD is given by $\hat{h}_{MLD}$. In term J_{10} in equation (13) is computed as

$$\hat{h}_{MLD} = \int_{ref}^{MLD} \delta(\hat{T}, \hat{S}, p)\, dz\,, \tag{37}$$

by integrating the normalized specific volume anomaly,

$$\delta = \left(\nu(\hat{T}, \hat{S}, p) - \nu(0,35,p)\right) \Big/ \nu(0,35,p)\,, \tag{38}$$

computed from the temperature and salinity of the synthetic profile (from the state vector, x), where ν is the specific volume. Since we want the cost function (13) to be written as a linear function of the amplitudes, a_i, of the EOFs, the calculation of δ is performed using a locally linearized approximation,

$$\delta(\hat{T}, \hat{S}, p) = \delta(T_{cl}, S_{cl}, p) + \alpha(T_{cl}, S_{cl}, p)\left(\hat{T} - T_{cl}\right) + \beta(T_{cl}, S_{cl}, p)(\hat{S} - S_{cl}) \tag{39}$$

where $T_{\rm cl}$ and $S_{\rm cl}$ are the temperature and salinity of the climatology field, and

$$\alpha = \left.\frac{\partial \delta}{\partial T}\right|_{S,p} \quad \text{and} \quad \beta = \left.\frac{\partial \delta}{\partial S}\right|_{T,p} \tag{40}$$

are the thermal and haline expansion coefficients estimated near each climatology temperature, salinity, and pressure. To reduce errors caused by the linearization, once the synthetic profile has been computed, the values of α and β are recomputed (at the average of synthetic and background T and S) and then the synthetic profile is computed again.

The steric height anomaly is then computed from the integral of the normalized specific volume anomaly from the reference level to the MLD,

$$\hat{h}_{MLD} = \int_{ref}^{MLD} \delta\left(\hat{T}, \hat{S}, p\right) dz = \sum_{i=1}^{N} X_i \left(\delta_{cl,i} + \alpha_i \hat{T}'_i + \beta_i \hat{S}'_i\right) \tag{41}$$

where X is a vector of depth intervals required to perform the discrete trapezoidal vertical integration. The last term J_{10} of equation (13) is then,

$$J_{10} = \frac{\left(\tilde{h}_{MLD} - \hat{h}_{MLD}\right)^2}{\varepsilon_h^2} = \frac{\left(\tilde{h}_{MLD} - \sum_{i=1}^{N}\left\{X_i\left(\delta_{cl,i} + \alpha_i \hat{T}_i' + \beta_i \hat{S}_i'\right)\right\}\right)^2}{\varepsilon_h^2} \tag{42}$$

In the next section, methods for minimizing the cost function are described.

4.2.2. Cost Function Minimization

The synthetic profile values, $\hat{T}_i$ and $\hat{S}_i$, for $i=1,N$ are computed by solving the system of equations created by minimizing the cost function, (13). Written compactly with the use of the ten J_i terms describe in the last section, the cost function is given by

$$J = \sum_{k=1}^{10} J_k \,. \tag{43}$$

The solution is the minimization of (43) with respect to each of the $2m$ component values of the unknown EOF amplitudes and $2N$ analysis values of $\hat{T}_i$ and $\hat{S}_i$.

The prototype system minimizes (43) directly by setting its derivative equal to zero. The derivatives are partial derivatives with respect to the m amplitudes for the T and S constraints,

$$\frac{\partial J}{\partial a_j} = \sum_{k=1}^{10} \frac{\partial J_k}{\partial a_j} = 0 \tag{44}$$

the m amplitudes for the vertical difference T and S constraints,

$$\frac{\partial J}{\partial b_j} = \sum_{k=1}^{10} \frac{\partial J_k}{\partial b_j} = 0 \tag{45}$$

the $N\hat{T}'$ and $N\hat{S}'$ analysis solutions,

$$\frac{\partial J}{\partial \hat{T}_j'} = \sum_{k=1}^{10} \frac{\partial J_k}{\partial \hat{T}_j'} = 0 \tag{46}$$

$$\frac{\partial J}{\partial \hat{S}_j'} = \sum_{k=1}^{10} \frac{\partial J_k}{\partial \hat{S}_j'} = 0 \tag{47}$$

making a total of $2m+2N$ unknowns and therefore $2m+2N$ partial derivative equations that are set to zero. The minimized solution occurs when each of the $2m+2N$ partial derivative equations are each equal to zero. When (44), (45), (46), and (47) are satisfied, the solution is obtained. All non-trivial partial derivatives are derived in detail in Appendix A.

The system of equations given by (44), (45), (46), and (47) can be represented by

$$\mathbf{D} \cdot \mathbf{g} = \mathbf{e} \,, \tag{48}$$

where $\mathbf{g}$ is the vector of $2m+2N$ unknowns,

$$\mathbf{g}^T = \left[a_1, \ldots, a_m, b_1, \ldots, b_m, \hat{T}_1', \ldots, \hat{T}_N', \hat{S}_1', \ldots, \hat{S}_N'\right]. \tag{49}$$

In the Appendix A, the set of $2m+2N$ equations are derived and written in a form where terms that do not depend on the unknowns in (48) are on the right hand side of the equal sign and are in $\boldsymbol{e}$. The $\boldsymbol{D}$, in (48) contain the terms on the left side of the equal sign. Also described in Appendix A are the methods for solving the system which include upper and lower matrix decomposition.

4.2.3. Layer 2 Error Estimates

Data assimilation systems are built around observation and model error estimates. For smooth integration of ISOP into the Navy's forecasting systems, error estimates of the synthetics are essential. The fundamental idea that facilitates our error estimate is the fact that the Layer 2 synthetics are derived from a linear combination of estimated or observed inputs. The inputs are ocean observations of surface temperature ($\tilde{T}_{MLD}$), sea surface height anomaly ($\tilde{h}_{MLD}$), and EOF amplitudes of the first-guess (a_{fg}, b_{fg}). With the use of eigenvectors (γ, ϕ), standard deviations (u, w) estimated from historical ocean profiles, and expected error variances $\left(\varepsilon_{SST}^2, \varepsilon_h^2, \frac{\lambda}{F_{fg}}, \frac{\mu}{G_{fg}}\right)$, synthetic T and S are constructed using the linear equation (48). Given this framework, it follows that we can estimate the synthetic temperature and salinity as a linear equation

$$x_k = K_k + \sum_{i=1}^{2m+2}\left(I_i \frac{\partial x_k}{\partial I_i}\right) \tag{50}$$

where I_i is the i-th input value and K_k is a constant for the k-th depth. There are $2m+2$ input values, $2m$ EOF amplitudes $\left(a_{fg}, b_{fg}\right)$ and 2 observations $\left(\tilde{T}_{MLD}, \tilde{h}_{MLD}\right)$. Given (50), the expected error variance, $\varepsilon_{x,k}^2$, of the T and S synthetics, x_k, can be computed as function of the expected error variances, $\varepsilon_{I,j}^2$. Expected error variance is given by

$$\varepsilon_x^2 = \left(x - x_{cl}\right)^T \left(x - x_{cl}\right) \tag{51}$$

or

$$\varepsilon_{x,k}^2 = \left(x_k - x_{cl,k}\right)^2 = \left[K_k + \sum_{i=1}^{2m+2}\left(I_i \frac{\partial x_k}{\partial I_i}\right) - \overline{K_k + \sum_{i=1}^{2m+2}\left(I_i \frac{\partial x_k}{\partial I_i}\right)}\right]^2 \tag{52}$$

since the climatology is the average ocean state. This reduces to

$$\varepsilon_{x,k}^2 = \left(x_k - x_{cl,k}\right)^2 = \left[\sum_{i=1}^{2m+2}\left(I_i \frac{\partial x_k}{\partial I_i} - \overline{I_i \frac{\partial x_k}{\partial I_i}}\right)\right]^2. \tag{53}$$

This is to a good approximation equal to

$$\varepsilon_{x,k}^2 = \left(x_k - x_{cl,k}\right)^2 = \left[\sum_{i=1}^{2m+2}\left(\left(I_i - \overline{I_i}\right)\frac{\partial x_k}{\partial I_i}\right)\right]^2 \tag{54}$$

or

$$\varepsilon_{x,k}^2 = \sum_{i=1}^{2m+2}\left(\varepsilon_{I,i}^2 \left(\frac{\partial x_k}{\partial I_i}\right)^2\right) \tag{55}$$

assuming each input is independent.

To make (55) useful, we need to derive a method for solving the partial derivative. Starting with equation (19) and shifting a_i to the other side of the equal sign we have

$$\left(x - x_{cl}\right) = U\Gamma a \tag{56a}$$

or in summation form we have

$$x_k - x_{cl,k} = \sum_{i=1}^{m} \gamma_{ik} u_k a_i \ . \tag{56b}$$

Taking the derivative of (56b) we obtain

$$\frac{\partial x_k}{\partial l_j} = u_k \sum_{i=1}^{m} \gamma_{ik} \frac{\partial a_i}{\partial l_j} . \tag{57}$$

This can be used in (55) to compute the error variances.

Alternatively, using the solution for the synthetic minimization, equation (48) can be used in the form

$$\mathbf{g} = \mathbf{D}^{-1}\mathbf{e} \ .$$

The components of **g** are such that

$$\hat{T}'_k = g_{k+2m} = \sum_{p=1}^{2m+2N} D^{-1}_{(k+2m)p} e_p \tag{58}$$

and

$$\hat{S}'_k = g_{k+2m+N} = \sum_{p=1}^{2m+2N} D^{-1}_{(k+2m+N)p} e_p \tag{59}$$

Now the expected error of each synthetic temperature and salinity value can be computed at every depth as

$$\varepsilon^2_{T,k} = \sum_{i=1}^{2m+2} \left(\varepsilon^2_{I,i} \left(\partial \hat{T}'_k / \partial l_i \right)^2 \right) \tag{60}$$

$$\varepsilon^2_{S,k} = \sum_{i=1}^{2m+2} \left(\varepsilon^2_{I,i} \left(\partial \hat{S}'_k / \partial l_i \right)^2 \right) \tag{61}$$

Since D is independent of the input values, then inserting Eq. (58) into (60) and (59) into (61), we obtain error variances of the synthetics,

$$\varepsilon^2_{T,k} = \sum_{i=1}^{2m+2+N_T+N_S} \left(\varepsilon^2_{I,i} \left(\sum_{p=1}^{2m+2M} D^{-1}_{k+2m,p} \partial e_p / \partial l_i \right)^2 \right) \tag{62}$$

and

$$\varepsilon^2_{S,k} = \sum_{i=1}^{2m+2+N_T+N_S} \left(\varepsilon^2_{I,i} \left(\sum_{p=1}^{2m+2M} D^{-1}_{k+2m+N,p} \partial e_p / \partial l_i \right)^2 \right) \tag{63}$$

Differentiating the components of the vector **e** from Eq. (48) and substituting into Eqs. (62) and (63) we obtain the error variance at each depth index, k, for temperature and salinity,

$$\varepsilon_{T,k}^2 = F_m \sum_{i=1}^{m} \left\{ \left(D_{k+2m,i}^{-1} \right)^2 \Big/ \lambda_i \right\} + G_m \sum_{i=1}^{m} \left\{ \left(D_{k+2m,i+m}^{-1} \right)^2 \Big/ \mu_i \right\} + \left(\sum_{i=1}^{N} \left(D_{k+2m,i+2m}^{-1} Q_i \right) \right)^2 \Big/ \varepsilon_{SST}^2 - \left(\sum_{i=1}^{N} \left(D_{k+2m,i+2m}^{-1} X_i \alpha_i \right) \right)^2 \Big/ \varepsilon_h^2 - \left(\sum_{i=1}^{N} \left(D_{k+2m,i+2m+N}^{-1} X_i \beta_i \right) \right)^2 \Big/ \varepsilon_h^2 \tag{64}$$

$$\varepsilon_{S,k}^2 = F_m \sum_{i=1}^{m} \left\{ \left(D_{k+2m+N,i}^{-1} \right)^2 \Big/ \lambda_i \right\} + G_m \sum_{i=1}^{m} \left\{ \left(D_{k+2m+N,i+m}^{-1} \right)^2 \Big/ \mu_i \right\} + \left(\sum_{i=1}^{N} \left(D_{k+2m+N,i+2m}^{-1} Q_i \right) \right)^2 \Big/ \varepsilon_{SST}^2 - \left(\sum_{i=1}^{N} \left(D_{k+2m+N,i+2m}^{-1} X_i \alpha_i \right) \right)^2 \Big/ \varepsilon_h^2 - \left(\sum_{i=1}^{N} \left(D_{k+2m+N,i+2m+N}^{-1} X_i \beta_i \right) \right)^2 \Big/ \varepsilon_h^2 \tag{65}$$

4.3. Layer 3 - Synthetic profiles below 1000 m depth

Temperature and salinity profiles in Layer 3, between 1000 m depth and the bottom, are computed by an anomaly decay scheme from profile values at the bottom of Layer 2. Given that $T_S(z)$ and $S_S(z)$ are the synthetic profiles and $T_G(z)$ and $S_G(z)$ are the GDEM 4 profiles of temperature and salinity, then the Layer 3 profiles are computed as

$$T_S(z) = T_G(z) + (T_S(1000) - T_G(1000)) F_T(z), \quad for\ z > 1000\ m\,. \tag{66}$$

$$S_S(z) = S_G(z) + \left(S_S(1000) - S_G(1000) \right) F_S(z), \quad for\ z > 1000\ m\,, \tag{67}$$

where the units for depth (z) in these equations is meters. $T_s(1000)$ and $S_s(1000)$ are the synthetic temperature and salinity at 1000 m depth computed at the bottom of Layer 2 and F_T and F_S are decay functions that have a value of one at 1000 m, and decay toward zero as depth increases below 1000 m. The decay function profiles were computed between 1000 m and 1800 m from historical observations by linear regression, i.e.,

$$F_T(z) = C_T(z) \left(\sigma_T(z) / \sigma_T(1000) \right), \quad for\ z = 1100\ to\ 1800\ m \tag{68}$$

$$F_S(z) = C_S(z) \left(\sigma_S(z) / \sigma_S(1000) \right), \quad for\ z = 1100\ to\ 1800\ m \tag{69}$$

where the temperature and salinity correlations, C_T and C_S, were computed from historical observations as,

$$C_T(z) = \frac{\sum_{i=1}^{N} \left(\Delta T_i(z) - \overline{\Delta T}(z) \right) \left(\Delta T_i(1000) - \overline{\Delta T}(1000) \right)}{\left(\sum_{i=1}^{N} \left(\Delta T_i(z) - \overline{\Delta T}(z) \right)^2 \sum_{i=1}^{N} \left(\Delta T_i(1000) - \overline{\Delta T}(1000) \right)^2 \right)^{1/2}} \tag{70}$$

$$\Delta T_i(z) = \hat{T}_i(z) - T_G(z) \tag{71}$$

$$\overline{\Delta T}(z) = \frac{1}{N} \sum_{i=1}^{N} \Delta T_i(z) \tag{72}$$

and $\hat{T}_i(z)$ is the temperature observation of the i-th profile at depth z. The correlations for salinity are computed in a similar manner. Below 1800 m, where observations are generally too sparse, the decay function was computed as an extension of the exponential trend between 1000 m and 1800 m, i.e.,

$$F_T(z) = sign\left[C_T(1800)\right]\exp\left(\left(1000 - z\right)/L_{Tv}\right)\left(\sigma_T(z)/\sigma_T(1000)\right), \quad for \quad z = 1800\, to \text{ the } bottom \tag{73}$$

$$F_S(z) = sign\left[C_S(1800)\right]\exp\left((1000 - z)/L_{Sv}\right)\left(\sigma_S(z)/\sigma_S(1000)\right), \quad for\ z = 1800\, to \text{ the } bottom \tag{74}$$

where the units of z is meters. The vertical length scales in units of meters are computed as,

$$L_{Tv} = -800/\log\left(\left|C_T(1800)\right|\right) \tag{75}$$

$$L_{Sv} = -800/\log\left(\left|C_S(1800)\right|\right). \tag{76}$$

As with the upper two layers, statistics for calculation of the synthetics in layer 3 were computed at each point of a global grid with 0.5° spacing in latitude and longitude. Note that Eq. 71 reduces to Eq. 66 when z =1800 m.

4.3.1. Layer 3 error estimates

The error variance of the layer 3 temperatures and salinities down to 1800 m is the sum of two independent errors, one due to the error of the regression and the second due to error in the layer two estimate of T and S at 1000 m ($T_S(1000)$ and $S_S(1000)$). The total error variance is then,

$$\varepsilon_T^2(z) = \sigma_T^2(z)(1 - C_T^2(z)) + \varepsilon_{T,1000}^2 F_T(z) \quad 1000\, m \le z \le 1800\, m \tag{77}$$

$$\varepsilon_S^2(z) = \sigma_S^2(z)(1 - C_S^2(z)) + \varepsilon_{S,1000}^2 F_S(z) \quad 1000\, m \le z \le 1800\, m\,. \tag{78}$$

The first term on the right hand side is the regression error variance written in terms of the correlations. The second term is the expected error variance (Eqs. 62 and 63) due to the error of the temperature or salinity at 1000 m computed from the layer 2 equations.

Below 1800 m depth, there are no statistics computed from the observations from which to estimate errors. However, below 1800 m, the correlations were assumed to decay exponentially with the vertical length scale found at 1800 m. Substituting the exponential correlation functions from Eqs. (73) and (74) into (77) and (78), we obtain estimates of the temperature and salinity below 1800 m as,

$$\varepsilon_T^2(z) = \sigma_T^2(z)(1 - \exp(2(1000 - z)/L_{Tv})) + \varepsilon_{T,1000}^2 F_T(z) \quad z \ge 1800\ m \tag{79}$$

$$\varepsilon_S^2(z) = \sigma_S^2(z)(1 - \exp(2(1000 - z)/L_{Tv})) + \varepsilon_{S,1000}^2 F_S(z) \quad z \ge 1800\ m\,. \tag{80}$$

5. Calculation of the statistical databases for each Layer

The databases of statistical quantities for each of the three layers were computed at each point of a global grid having a resolution of 0.5 degree in latitude and longitude. All results are based on calculations made from the edited global data set of temperature and salinity profiles, described in Section 2, which had been interpolated to 78 standard depths from the surface to 6600 m depth, and including 47 depths from the surface to 1000 m depth. Only profiles having both temperature and salinity observations were used, reducing the number of useable profiles to 1.97 million. The various statistics were computed at each grid point for each of the three ISOP layers using only locally surrounding profiles contained within the rectangular region and within the specified time window, centered on the grid point analysis position, with east-west and north-south lengths, in units of kilometers, defined by

$$\Delta X = (k+1) * 100km * \left[1.3 + 1.7 \exp\left(-\left(\lambda_a / 15.0 \right)^2 \right) \right]$$
$$\Delta Y = (k+1) * 100km \qquad (81)$$

respectively, where λ_a is the latitude in units of degrees. At the equator, the rectangle is twice as large in the east-west direction as in the north-south direction. Beginning with k=1, the number of profiles within the box is counted. If the number is fewer than a specified minimum (200 in Layer 1, 750 in Layer 2, and 50 in Layer 3), then the value of k is incremented upward by one, repeatedly, until at least the specified minimum number of profiles (independent of time of year) are obtained. In Layers 1 and 3, all profiles within the box, independent of time of year were used in the calculations. In Layer 2, monthly analyses were performed using profiles within specific time-of-year bounds which varied with depth. In each layer, each profile is weighted in the analysis according to their distance from the central grid point.

5.1. Layer 1 – Mixed Layer

The ISOP Layer 1 is the near-surface mixed layer of the synthetic profile. It is computed after the Layer 2 temperature and salinity profiles have been computed because the Layer 1 model of the mixed layer potential density profile requires the value of the potential density gradient (from Layer 2) at the depth of the mixed layer before it can be computed. At the same time, calculation of Layer 2 profiles require knowledge of the Layer 1 steric height at the surface relative to the MLD in order to compute the steric height at the MLD relative to 1000 m from the steric height at the surface relative to 1000 m. Both of these calculations could have been performed simultaneously with suitable manipulation of the equation for both layers. However, instead we perform a short iteration where an estimate of the upper mixed layer temperature and salinity is first made. From that, the steric height at the surface relative to the MLD is computed, which is used in the Layer 2 calculations. Once the Layer 2 profiles are computed, the vertical gradient of the potential density is computed at the MLD, which, in turn, is used in the recalculation of the Layer 1 profiles. The procedure typically converges well enough after only two iterations.

When the ISOP synthetic profiles are constructed and used for assimilation into an ocean model, we typically compute the MLD from the model forecast (at the analysis time) using the potential density threshold method. In addition, rather than compute the mixed layer temperature and salinity profiles using the Layer 1 mixed layer model equations and statistics, we use the model profiles above the MLD. The only further requirement is that the model mixed layer profiles must be linearly stretched; the

temperature is stretched to match a specified sea surface temperature and Layer 2 temperature at the MLD and the salinity is stretched to match the Layer 2 salinity at the MLD.

A global database of statistics required to support calculation of profiles in the mixed layer is constructed from local profiles at each point of a global 0.5° x 0.5° grid. The procedure for selecting profiles for each calculation is described at the top of Section 4. The selection process obtains a minimum of 200 pairs of temperature and salinity profiles. The MLD is determined for each pair using the potential density threshold method. In addition, the vertical gradient of potential density at the MLD is computed. Each profile is then interpolated to the 21 scaled depths (z' in Equation 6) between the surface and the MLD. Using this data set, the ISOP global mixed layer model database is constructed, for each grid location, using the unweighted and time-of-year-independent regressions represented by Equations 4, 8, and 9.

At each grid location, the potential density profile in the mixed layer was computed for each profile used in the construction of the ISOP data base at this grid location. The calculation for each profile used the original profile's MLD, threshold value, and potential density gradient at the MLD. The fraction of the variance accounted for the regression at each scaled depth, z'_k is computed by the following ratio,

$$100\left\{1-\left(\frac{\sum_{i=1}^{N}\left[\left(\Delta\sigma_{\theta,i}\left(z'_k\right)-\overline{\Delta\sigma}_{\theta}\left(z'_k\right)\right)-\left(\Delta\hat{\sigma}_{\theta,i}\left(z'_k\right)-\overline{\Delta\hat{\sigma}}_{\theta}\left(z'_k\right)\right)\right]^2}{\sum_{i=1}^{N}\left[\Delta\hat{\sigma}_{\theta,i}\left(z'_k\right)-\overline{\Delta\hat{\sigma}}_{\theta}\left(z'_k\right)\right]^2}\right)\right\} \tag{82}$$

where the sigma values with the over hat are the observed values and the sigma values without the over hat are the values computed from the ISOP mixed layer model equations, and

$$\Delta\hat{\sigma}_{\theta,i}\left(z'_k\right)=\left\{\hat{\sigma}_{\theta,i}\left(z'_k\right)-\hat{\sigma}_{\theta,i}\left(1\right)\right\}\Big/\sigma_{\theta,Thresh} \tag{83}$$

$$\Delta\sigma_{\theta,i}\left(z'_k\right)=\left\{\sigma_{\theta,i}\left(z'_k\right)-\sigma_{\theta,i}\left(1\right)\right\}\Big/\sigma_{\theta,Thresh} \tag{84}$$

$$\overline{\Delta\hat{\sigma}}_{\theta}=\sum_{i=1}^{N}\left(\Delta\hat{\sigma}_{\theta,i}\right)\Big/N \tag{85}$$

$$\overline{\Delta\sigma}_{\theta}=\sum_{i=1}^{N}\left(\Delta\sigma_{\theta,i}\right)\Big/N \tag{86}$$

The percent of the potential density variance accounted for by the regression is displayed at the scaled depth of 8.867 * MLD in Figure 6.

Figure 7 shows a vertical section through the central Pacific Ocean from south to north along 190° E of statistics from the ISOP Layer 1 (mixed layer) potential temperature model. Figure 8 and 9 show the mixed layer model applied to two example profiles from the Pacific Ocean.

5.2. Layer 2

5.2.1. Data set at each grid position

The global database of required statistics for Layer 2 are computed at each point of the global 0.5° by 0.5° grid using all profiles within the local rectangular region obtained in the profile search procedure described at the beginning of Section 5. The value of the observed temperature or salinity at each standard depth for each of the N profiles within the final rectangular region surrounding a grid point is assigned a weight. If the value at the k^{th} depth of the i^{th} profile is missing, then the weight, $w_{i,j}$ = 0. Otherwise, the weight is defined in terms of the distance of that profile from the grid point, the day of the year when the observation was made, and whether there is a bottom obstruction between the profile position and the grid point position at and below a given depth.

For profile observations at each depth, z (units of meters), we define a function $\delta_b(d - d_m)$ that has a value of one for month m (m=1, 12) only if the day (of the year) of the observation, d, is within the time period, $-\Delta t_{\max}(z) \le d - d_m \le \Delta t_{\max}(z)$, where the day at the center of month m is defined by

$$d_m = 15.25 + 30.5\ (m-1), \tag{87}$$

and the half-width, $\Delta t_{\max}(z)$ (in units of days), is defined by,

$$\Delta t_{\max}(z) = \begin{cases} 45 & 0 \le z < 100m \\ 60 & 100 \le z < 200 \\ 90 & 200 \le z < 400 \\ 120 & 400 \le z < 700 \\ 183 & 700 \le z \le 1000 \end{cases} \tag{88}$$

Otherwise, the value is zero, i.e., $\delta_b(d - d_m) = 0$.

Each profile was previously interpolated to the set 47 analysis depths between the surface and 1000 m (see Section 2). The maximum bottom bathymetry depth, b_{max}, within the 0.5°x0.5° (latitude, longitude) rectangle centered at the analysis position, was determined from the NRL DBDB2 v3.0 global 2-minute-resolution bottom topography (Ko 2009). No analyses were performed at a location where b_{max} < 5 m.

The lengths of the sides of the final rectangle also define the decay scales,

$$\begin{aligned} L_\theta &= \Delta X / \left(110.\cos(\lambda_a \pi / 180\right) \\ L_\lambda &= \Delta Y / 110. \end{aligned} \tag{89}$$

in units of geographical degrees, in the function,

$$w_{k,i} = \delta_d(d - d_m)\,\delta_b(z_k - b_m)\exp(-\left(\left(\theta_i - \theta_a\right)/L_\theta\right)^2 - \left(\left(\lambda_k - \lambda_a\right)/L_\lambda\right)^2 \tag{90}$$

used for weighting the i^{th} observation at the k^{th} depth, z_k, when computing means, variances, and covariances. The weights vary as a function of the distance from between the grid point analysis position, at longitude and latitude given by λ_a and θ_a, respectively, and the position of the i-th profile at λ_i and θ_i. Defining b_{min} as the minimum bottom depth on a rhumb line between the analysis position and the observation position,

$$\delta(z_k - b_{\min}) = 0 \quad z_k > b_{\min} \tag{91}$$

$$\delta(z_k - b_{\min}) = 1 \quad z_k \le b_{\min}. \tag{92}$$

The bottom depth values along the rhumb line were extracted using bilinear interpolation from DBDB2 bathymetry.

The deepest analysis depth at each location was the minimum of either 1000 m or the deepest standard depth (of the 47 depths in the upper 1000 m) with at least two observations remaining after performing the subsetting discussed above.

5.2.2. Calculation of means and standard deviation profiles

At each grid location, the profile of the weighted means and standard deviations of temperature and salinity were computed together with similar calculations performed on the differences of vertically adjacent values:

$$\tilde{T}_k = \sum_{i=1}^{N} w_{k,i} T_{k,i} \Big/ \sum_{i=1}^{N} w_{k,i} \tag{93}$$

$$\tilde{S}_k = \sum_{i=1}^{N} w_{k,i} S_{k,i} \Big/ \sum_{i=1}^{N} w_{k,i} \tag{94}$$

$$\sigma^2_{T,k} = \sum_{i=1}^{N} w_{k,i} \left(T_{k,i} - \tilde{T}_k \right)^2 \Big/ \sum_{i=1}^{N} w_{k,i} \tag{95}$$

$$\sigma^2_{S,k} = \sum_{i=1}^{N} w_{k,i} \left(S_{k,i} - \tilde{S}_k \right)^2 \Big/ \sum_{i=1}^{N} w_{k,i} \tag{96}$$

$$\Delta\tilde{T}_k = \sum_{i=1}^{N} w_{k,i} \left(T_{k+1,i} - T_{k,i} \right) \Big/ \sum_{i=1}^{N} w_{k,i} \tag{97}$$

$$\Delta\tilde{S}_k = \sum_{i=1}^{N} \left(S_{k+1,i} - S_{k,i} \right) \Big/ \sum_{i=1}^{N} w_{k,i} \tag{98}$$

$$\sigma^2_{\Delta T,k} = \sum_{i=1}^{N} \left(T_{k+1,i} - T_{k,i} - \Delta\tilde{T}_k \right)^2 \Big/ \sum_{i=1}^{N} w_{k,i} \tag{99}$$

$$\sigma^2_{\Delta S,k} = \sum_{i=1}^{N} w_{k,i} \left(S_{k+1,i} - S_{k,i} - \Delta\tilde{S}_k \right)^2 \Big/ \sum_{i=1}^{N} w_{k,i} \tag{100}$$

where the i subscript indicates the i^{th} profile and the k subscript indicates the value at the k^{th} depth, z_k. The result of each calculation was discarded if the sum of weights ($\sum_{i=1}^{N} w_{k,i}$) was less than 0.1 or if the number of values with non-zero weights was less than 20 in the upper 100 m depth, or less than 10 below 100 m depth.

The mean temperature and salinity profiles were corrected to compensate for biases introduced by the changing number of observations, N_k, at each depth, using the "vertical gradient correction" method developed for the calculation of the GDEM ocean temperature and salinity climatology (Carnes 2009, Carnes, Helber, et al. 2010). This technique computes the corrected temperature and salinity mean, $\overline{T}_k$ and $\overline{S}_k$, by minimizing the cost functions

$$J = \sum_{k=1}^{M} \left(\frac{\overline{T}_k - \tilde{T}_k}{\sigma_{T,k}} \right)^2 + \sum_{k=1}^{M-1} \left(\frac{\overline{T}_{k+1} - \overline{T}_k - \Delta\tilde{T}_k}{\sigma_{\Delta T,k}} \right)^2 \tag{101}$$

$$J = \sum_{k=1}^{M} \left(\frac{\overline{S}_k - \tilde{S}_k}{\sigma_{S,k}} \right)^2 + \sum_{k=1}^{M-1} \left(\frac{\overline{S}_{k+1} - \overline{S}_k - \Delta\tilde{S}_k}{\sigma_{\Delta S,k}} \right)^2 \tag{102}$$

with respect to each of the final corrected mean temperature or salinity values, where M is the index of the deepest depth with useable values. Before performing this calculation, any value of the two sets of standard deviations less than 0.001 were set to 0.001 to remove poorly conditioned equations. The minimization of each equation leads to a tridiagonal system of M equations described in (Carnes, Helber, et al. 2010), where several examples are also shown. This technique is effective in correcting the means because the vertical gradients provide estimates of the between-layer relationships missing from the uncorrected layer-by-layer calculations of the means. Often, the basic shape of different profiles is similar even though the profiles may be shifted relative to each other. In fact, we often find that a running accumulated sum of the mean vertical difference profiles provides a better estimate of the average profile shape than does the mean profile. However, the accumulated vertical profile, starting from the mean value at the first depth, usually slowly drifts from the mean value curve as the vertical accumulation proceeds. However, the accumulation plus a drift correction is essentially provided by the minimization of the cost functions above.

A similar set of calculations was also performed to compute vertical-gradient-corrected means and standard deviations of the consecutive vertical differences of temperature and salinity. The means, standard deviations and covariances of the vertical differences are needed for calculation of synthetic profiles.

5.2.3. Calculation of correlations, eigenvalues, and eigenvectors

The fundamental structure of the Layer 2 synthetic profiles come from the EOF representation of the ocean that is used to construct the cost function (13) described in section 4.2.1. The EOFs are represented in terms of eigenvalues and eigenvectors as part of the covariance matrix decomposition, equation (16). Correlations, rather than covariances, were computed and used in the calculation of eigenvectors because the calculations are combining temperature variability with salinity variability (see section 4.2.1). The correlation matrix, C in Eq. 16, of the combined temperature and salinity profiles were computed for each month using edited profile database and same space and time weighting scheme used for computing the means.

The numerical values of the temperature changes in units of degree Celsius are often much larger than the changes in salinity in units of PSU. Without scaling, the salinity prediction is severely reduced unless a large number of modes (empirical orthogonal modes, or eigenvectors) are used. Also, the scaling helps to improve predictability in the depth domains with lower variability, but at the expense of predictability in the higher variability depths (like in the thermocline). Scaling could have been designed to meet other requirements. For example, temperature and salinity could have been scaled according to their effects on changes in density or on changes in sound speed. Many of the temperature or salinity profiles used in the calculation of the correlations for each month had missing values (or a weight equal to zero), including missing values near the surface if the observation date was not within the day-of-the-year time range of the month being computed, or near the bottom of the profile if it was shorter than surrounding profiles. When computing products of scaled anomalies from two different depths, the product could be computed only if the values were not missing at both depths. As a result, the mean values computed at each depth in the previous section often used a different (and larger) set of observations than was used for a particular correlation calculation. Our approach compensated for the missing data by recomputing the average values for each two-depth pair. For a particular two-depth pair, the mean at each depth was computed from only the subset of values used in the products, and therefore the subset that was not missing at both depths on each profile. For example, if the correlation is being computed between temperature at two different depths, z_i and z_j, then

$$C_{i,j} = \sum_{m=1}^{N} \left(w_{i,m} w_{j,m} \right)^{1/2} \left[\left(\frac{T_{i,m} - \overline{T}_i}{\sigma_{T,i}} \right) - \overline{\theta}_{i(j)} \right] \left[\left(\frac{T_{j,m} - \overline{T}_j}{\sigma_{T,j}} \right) - \overline{\theta}_{j(i)} \right] \Bigg/ \sum_{m=1}^{N} \left(w_{i,m} w_{j,m} \right)^{1/2} \tag{103}$$

where

$$\overline{\theta}_{i(j)} = \sum_{m=1}^{N} \left(w_{i,m} w_{j,m} \right)^{1/2} \left(\frac{T_{i,m} - \overline{T}_i}{\sigma_{T,i}} \right) \Bigg/ \sum_{m=1}^{N} \left(w_{i,m} w_{j,m} \right)^{1/2} \tag{104}$$

and

$$\overline{\theta}_{j(i)} = \sum_{m=1}^{N} \left(w_{i,m} w_{j,m} \right)^{1/2} \left(\frac{T_{j,m} - \overline{T}_j}{\sigma_{T,j}} \right) \Bigg/ \sum_{m=1}^{N} \left(w_{i,m} w_{j,m} \right)^{1/2} \tag{105}$$

Finally, eigenvectors and eigenvalues of the correlation matrix, C, were computed by singular value decomposition. Similarly, the correlation matrix of the combined temperature and salinity vertical (adjacent) differences was computed, and the eigenvalues and eigenvectors also computed. In the end, we computed a separate set of means, standard deviations, eigenvalues, and eigenvectors at each point of the global 0.5°x0.5° grid for each month. We arrange the eigenvector modes in order of their eigenvalues, with the mode having the highest eigenvalue occurring first and with the eigenvalues decreasing as the mode number increases. In the prototype ISOP databases, only the six largest modes were saved. We can determine the fraction of the variance of the entire data set accounted for these modes because the trace of the C correlation matrix is equal to the sum of the eigenvalues of C. The trace of C is the sum of the variance (of the scaled anomalies) at each depth. Figures 10, 11, and 12 show global maps of the percent of total variance accounted for by the first mode, by the sum of the first and second modes, and by the sum of the first three modes, respectively. These figures indicate that, over much of the world's oceans, the running sum of the eigenvalues, starting with mode 1, the mode with the largest eigenvalue, rapidly converges toward the total variance over all depths, i.e., the trace of the correlation matrix. The rapid convergence allows us to efficiently use only a small subset of the eigenvectors with the largest eigenvectors to compute synthetic profiles. The percentage is over 80% by just the first mode over much of the Southern Ocean and in the areas of the Kuroshio Extension and the Gulf Stream Extension. The lowest accumulated variance is found in the equatorial regions of the Indian, Pacific, and Atlantic Oceans. There may be is a certain background noise floor everywhere where fluctuations at one depth are not correlated with those at another or where fluctuations of temperature are not correlated with fluctuations of salinity. In regions where the variability is low and only slightly larger than the noise background, then only a small part of the variability is vertically correlated. In this case, the first few modes will account for only a small fraction of the total variance. Another factor in the interpretation of the eigenvalues is that the temperature and salinity anomalies were scaled by the local standard deviations before calculation of the correlations and eigenvectors, so each depth for both temperature and salinity contributes an equal share to the total (scaled) variance. Scaling makes interpretation of both the eigenvalues and the eigenvectors difficult.

The eigenvectors are computed separately at each grid point location, so they don't necessarily vary smoothly between locations. The shape of eigenvector vertical profiles at adjacent grid points might be nearly the same, but their signs can be different, or the structure found in one mode at one grid point location can suddenly shift into the structure of other modes at the adjacent grid location. However, as shown next, the general structure of eigenvectors, for at least the first mode, generally changes smoothly. The next set of plots shows horizontal (global map) and vertical section views of the first half (temperature) and last half (salinity) of the combined July mode 1 temperature and salinity eigenvectors for two different months. Figure 13, displays the absolute value of the first mode temperature eigenvector at a depth of 160 m, and Figure 14 shows the comparable salinity eigenvector

map. The 160 m depth was chosen because it is typically below the mixed layer and within the main or seasonal thermocline. These two maps delineate the positions of many of the major fronts, particularly in the tropical and subtropical regions. In plots of higher modes, not shown, the spatial scales are generally smaller and the major oceanic features are more difficult to recognize.

Vertical sections of the first mode eigenvectors of the scaled temperature and salinity are shown in Figure 15. In each plot, the sign of the eigenvector profile at each position along the section was adjusted to produce a smoothly changing picture. Even in scaled form and with only the first mode showing some of the major oceanic feature along the section are vision, including the Subarctic Front near 60° S in the salinity, the boundary between the lower salinity central water of the subtropical gyres and the higher salinity central water of the subtropics. In the temperature vertical section, the dominant negative (blue) feature that dives down toward the north in the upper 500 m between about 10° N and 25° N, follows the isohalines in the salinity vertical section (not shown), but is perpendicular to isotherms. However, it does approximately follow a local temperature minimum that progresses downward along the temperature eigenvector feature.

5.3. Layer 3

The synthetic profiles in the deep layer, below 1000 m depth, are computed by extending the Layer 2 temperature and salinity anomaly (difference from GDEM) at 1000 m downward. The method, described in Section 4.3, is equivalent to predicting the deeper anomaly from the 1000-m anomaly using the slope computed from a linear regression of historical data between the anomalies from GDEM at 1000 m and the anomalies from GDEM at greater depths. The regressions require calculation of the variances at each depth and calculation of the correlations between anomalies at 1000 m and anomalies at deeper depths. The profile observations below 1800 m depth are too sparse in most regions to compute the correlation with adequate accuracy, and variances were also not computed below 1800 m. As a result, for consistency, once the correlations are computed from the covariances and the computed variances, we eliminated all of the variances computed as part of the regressions, and replaced them with variances from the GDEM4 climatology at all depths in Layer 3. In addition, as discussed in Section 4.3, correlations between 1000 m and the specified depths below 1800 m are assumed to have an exponential form with length scale selected to match the computed correlation between 1000 m and 1800 m depth.

The salinity analysis was performed separately from the temperature analysis. At each grid location, a search for deep temperature or salinity profiles, depending on the analysis, was made in an expanding rectangular region given by Eq. 88, until at least 50 profiles were found that extended down to the standard depth nearest (but above) the local bottom depth or to at least to a depth of 1800 m, whichever was greater. Unlike the monthly analyses performed for Layer 2, a single annual analysis, independent of time of year, was performed for Layer 3. As was done for the Layer 2 calculations, observations from a profile made at depths below the shallowest bottom bathymetry depth between the profile observation position and the grid point analysis position were excluded. Finally, each computed vertical covariance was excluded if made from fewer than 30 observation pairs. Unlike the method used for Layer 2, the Layer 3 covariances were computed using equal weights for each observation, irrespective of their distances from the grid point analysis position.

The fields of correlations and GDEM standard deviations were smoothed horizontally by 15 application of a 9-point smoothing filter (Shapiro 1970). Vertical sections in Figure 16 of the deep temperature and salinity decay functions between 1100 m and 5000 m along 215 E (145 W) longitude from the Antarctic coast to the Alaska coast. The value given by this function is multiplied times the Layer 2 temperature or salinity anomaly from GDEM at 1000 m and added to the GDEM profile at

greater depths to produce the Layer 3 temperature or salinity profile. Even though the number of profiles used in the calculations at each position are nearly the same for temperature as for salinity, the extension functions are quite different. Only the values from 1100 m and 1800 m are supported by correlations computed by observations. Below 1800 m the decay function values are computed by extending the correlations at 1800 m and using the GDEM standard deviations below 1800 m.

In order to produce two-dimensional global maps of the primary features of the three-dimensional decay functions, the vertical length scales, $\hat{L}_T$ and $\hat{L}_S$, were computed at each grid location as the depth where the decay functions, $F_T(z)$ and $F_S(z)$, respectively, in Eqs. 68, 69, 73, and 74 reached a value, e^{-1}, minus the 1000 m depth. With these length scales, decay functions can be obtained as:

$$\hat{F}_T(z) = \exp\left[-(z-1000)/\hat{L}_T\right] \quad z > 1800\,m \tag{106}$$

$$\hat{F}_S(z) = \exp\left[-(z-1000)/\hat{L}_S\right] \quad z > 1800\,m\,. \tag{107}$$

The length scales, $\hat{L}_T$ and $\hat{L}_S$, are plotted in global maps in Figures 17 and 18, respectively. The dominant features of the temperature length scale are the long (2500 m to 3000 m) scales south of the Sub Antarctic Front, the short (less than 500 m) length scales near the equator and the intermediate values in the Mediterranean, Sea of Japan, Labrador Sea, Bering Sea, Alaskan Stream, and the southeastern tropical and subtropical Atlantic. The salinity length scales have high values (greater than 2500 m) south of the Polar Front along the Antarctic coast. Values near 2000 m are found in parts of the Mediterranean, in the Bay of Bengal, In the northern Atlantic Subpolar Gyre, and in the Greenland and Norwegian Seas. Intermediate scales (1000 m to 1500 m) are found near the Sub Antarctic Front, in the tropical Pacific, the subtropical gyre of the South Atlantic, and in the Alaska Current, Alaska Stream, and the Bering Sea. Lowest values are in the equatorial Atlantic and the southern portion of the Indian Ocean subtropical gyre.

6. Validation Testing

The quality of ISOP synthetics have been evaluated globally and in two sub-regions, the Gulf of Mexico and a region around Hawaii known as the RIMPAC region. These validation experiments are direct comparisons of ISOP to in situ observations. In the scientific literature comparisons of models with observations is commonly call misfit error. The estimated synthetic error output from ISOP is used in data assimilation systems and is examined in the VTR Part II.

6.1. ISOP SETUP

There are two ISOP configurations tested in this document, the first (1) represents *ideal* ($ISOP^{Ideal}$) and the second (2) *operational* ($ISOP^{oper}$) implementations. The $ISOP^{Ideal}$ configuration provides an estimate of the best possible accuracy that can be expected, while the $ISOP^{oper}$ configuration provides accuracy that may be obtained in an operational implementation. The inputs for the $ISOP^{Ideal}$ configuration are derived from the validation profiles themselves and provide a best possible capability limit of ISOP if the inputs provided were of exceptional quality. The $ISOP^{oper}$ configuration uses inputs from legacy capabilities that are not as good as current Navy operational systems. For this reason, the error levels for the $ISOP^{oper}$ configuration are conservative, representing the lower limit for ISOP accuracy. We expect the actual operational accuracy for ISOP to be greater than the $ISOP^{oper}$ results of this analysis. The $ISOP^{Ideal}$ and $ISOP^{oper}$ validation cases provide an upper and lower bound, respectively, on the performance expected from ISOP.

In order to demonstrate the improved capability of ISOP, we have also generated MODAS synthetics and extracted GDEM 4.0 (Carnes, Helber, et al. 2010) matching profiles. To match the two ISOP configurations, we have two MODAS configurations with the same *ideal* and *operational* inputs used for ISOP. All of the synthetics and ancillary products are listed in Table 7, and are matched to the validation profile observations described in the next section.

6.2. Validation Data

For the global validation we have compiled and quality controlled a global set of observation profiles that have not been used in the making of ISOP. The dataset is therefore completely independent of the ISOP background statistics and provides ideal data set for validation purposes. The data sources are the same as described in section 3, a combination of profiles from the Navy's MOODS, the WOD through 2009, and Argo (USGODAE 2012). Only profiles having both temperature and salinity were used and the same quality control procedures are used as describe in section 3. The total collection of data from July 2008 through November 2009 contains 111,199 profiles and of those, 74,314 profiles extend to at least 1000 m. Of those, 44,118 extend down to 1800 m. We have performed two global validation cases. Test Case 1: Global Validation uses the 74,314 profiles that extend to at least 1000 m. Test Case 2: Global Deep Ocean Validation uses the 52,480 profiles that extend to at least 1800 m in order to test the quality of the Layer 3 synthetics.

For the Gulf of Mexico (GOM) region validation, we have collected 4001 paired T and S profiles mostly from the years 2010 and 2011 that have passed a minimum set of requirements within the latitude and longitude limits of 18°N to 32°N and 100°W and 77°W. All profiles retained have depth ranges from above 12 m to at least 120 m with at least 5 levels in the profile. The data source is the National Oceanographic Data Center (NODC) World Ocean Database (WOD) (NODC 2009). To evaluate ISOP in water shallower than 1000 m, a subset of 1202 profiles are used for this purpose.

For the RIMPAC region validation, we have collected 3227 paired T and S profiles that have passed the same minimum set of requirements listed for the GOM region within the latitude and longitude limits of 18°N to 23°N and 161.5°W to 153.5°W. The data source is the Navy's MOODS profile archive updates that were provided for NRL in January 2011 and April 2012. Most of the data are from 2009 through 2011 but some of the data date back to 2005.

6.3. Validation Metrics

Since a main focus of the ISOP system is to improve the representation of acoustically relevant features of ocean structure, key acoustic metrics for this analysis are Sonic Layer Depth (SLD; Helber et al. 2008), the Mixed Layer Depth (MLD; Kara et al. 2000) and Below Layer Gradient (BLG). The BLG is the change in sound speed per 100 feet in units of m/s per 100 ft. The software used for computing SLD and BLG is version controlled in the NRL's software repository (R. W. Helber 2012).

For completeness, we also validate the T, S, and Sound Speed (SSPD) profiles themselves. The validation of SLD, MLD, and BLG, however, is different from the validation of T, S, and SSPD. The difference is that there is one value for SLD, MLD, and BLG for each ocean profile. This clearly is not the case for T, S, and SSPD, since there are values at every depth in the profile. For this reason, the T, S, and SSPD difference (or misfit) of *observation – synthetic* at each depth is computed. Then the RMSE and Skill Score (SS) are computed for each difference profile to give one value, either RMSE or SS for each profile match.

6.4. Validation Methods

All comparisons of SLD, MLD, and BLG are done relative to the validation profile observations with the convention of *observation – synthetic*. For example, in each 10° by 10° region or each of the sub-regions, the mean bias (MB) is given by

$$MB = \frac{1}{n}\sum_{i=1}^{n}\left(syn_i - obs_i\right) \tag{108}$$

and the root mean square error (RMSE) is given by

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(syn_i - obs_i\right)^2}\ , \tag{109}$$

where n is the number of profiles. In cases where the MB of RMSE is computed for T, S, and SSPD for every profile over a depth range, the MB for the i[th] profiles is

$$MB_i = \frac{1}{b-(a-1)}\sum_{k=a}^{b}\left(syn_k - obs_k\right) \tag{110}$$

and the RMSE is becomes

$$RMSE_i = \sqrt{\frac{1}{b-(a-1)}\sum_{k=a}^{b}\left(syn_k - obs_k\right)^2}\ , \tag{111}$$

where *a* and *b* are the start and stop depth indexes in the profile interpolated to standard depths. To provide global or regional estimates of MB and RMSE, the median of MB_i and $RMSE_i$ are then computed. These values, along with the error values of SLD, MLD, and BLG are then binned in either 10° by 10° regions or the sub-regions in the GOM and RIMPAC.

When $RMSE$ is computed at each depth, the standard error (STDE) is computed as

$$STDE = \sqrt{\sigma_{RMSE}^2/n} \tag{112}$$

where

$$\sigma_{RMSE} = \frac{1}{n}\sum_{i=1}^{n}\left(RMSE_i - \overline{RMSE}\right) \tag{113}$$

and *n* is the number of profiles. This standard error estimate assumes that all estimates are independent. Since not all estimates are independent, STDE is an underestimate of error uncertainty. Regardless, the STDE estimate gives an approximate estimate of the uncertainty of the error estimates themselves.

Another way to consider the ISOP error levels is to compare ISOP with the GDEM climatology error levels using skill score (SS) (Murphy 1988). This is done by comparing the mean square error (MSE) of the synthetic to the MSE of the GDEM climatology using the formula

$$SS = 1 - \frac{MSE\left(syn\right)}{MSE\left(GDEM\right)}, \tag{114}$$

where

$$MSE\left(syn\right) = \frac{1}{n}\sum_{i=1}^{n}\left(syn_i - obs_i\right)^2 \tag{115}$$

and

$$MSE\left(GDEM\right) = \frac{1}{n}\sum_{i=1}^{n}\left(GDEM_i - obs_i\right)^2 . \tag{116}$$

In equations 108-111 and 114-116, syn_i, obs_i, and $GDEM_i$ are any of the metric variables SLD, MLD, or BLG for the i^{th} ISOP or MODAS synthetic, in situ observation, and GDEM climatology profile, respectively. The syn_i represents any of the ISOP or MODAS configurations (Table 7) and the i^{th} profile of each type is from a 10°by 10° region or one of the sub-regions. Following these methods, the quality of the ISOP acoustic parameters are validated relative to MODAS synthetics, GDEM climatology, and in situ observations.

6.5. Results: Test Case 1: Global Validation

The purpose of Test Case 1 is to validate the global quality of ISOP synthetics against independent observation. To present the global geographical error characteristics of the ISOP system, we first compute the SLD RMSE and MB error in 10° by 10° blocks. In each 10° by 10° geographical ocean region with at least 5 in situ observations, we compute the mean bias (MB) as the mean misfit synthetic – observation. The root mean square error (RMSE) is the root mean square of the misfit synthetic – observation. The results for SLD for the $ISOP^{Ideal}$ and $ISOP^{Oper}$ cases are shown in Figure 19 and Figure 20, respectively. The vast majority of the global ocean has a RMSE error less than 40 m and a shallow bias less than 24 m for the $ISOP^{Oper}$ case (Figure 20). Synthetic ocean profiles and climatologies have a long standing shallow bias, which is represented in Figure 20 by negative mean bias since the MB convention is the synthetic minus the observation. Given ideal inputs, $ISOP^{Ideal}$ (Figure 19) has a substantial improvement over the case with operational inputs (Figure 20).

Next, *ideal* and *operational* RMSE and MB results (Figures 21 and 22) are shown for MLD using the same color bar as for SLD. The MLD represents the ocean dynamical and turbulent response to air-sea interaction. MLD tends to be shallower with less variance than the SLD resulting in smaller error levels.

The results for BLG are shown in Figures 23 and 24. The units are m/s per 100 ft. The themocline structure along the equator is not well represented by climatologies. The seasonal cycle at the equator is heavily modulated by atmospheric mesosclae wind forcing that drives equatorial trapped Kelvin and Rossby waves. For this reason, the below layer gradient has larger error levels along the equator compared to the rest of the ocean.

To demonstrate the global relative accuracy of each ISOP configuration relative to MODAS and GDEM, we take the median RMSE and MB for SLD, MLD, and BLG for all 10° by 10° regions (Table 9). For every parameter, the Ideal ISOP case had the best performance. The next best performance is from ISOP with operation inputs.

To further test the ocean dynamical representation of ISOP we also test the accuracy of the synthetic T and S over four depth ranges. We compute the global median of MB of T, S, and SSPD for all 10° by 10° regions for each of the depth ranges 0 to 50 m, 50 to 100 m, 100 to 150 m, and 150 to 200 m (Table 10). Results indicate that the smallest errors are at the surface and largest errors occur in the upper portion of the thermocline, which occurs in the depth range of 50 to 100 m.

We show geographical plots of the T, S, and SSPD misfit errors in Figures 25 through 30. On the top panel (a) of each Figure we have the RMSE error over the upper 1000 m. The lower panel (b) is the MB for the depth range 50 to 100. These figures characterize the global error characteristics of ISOP for the ideal and operational configurations as labeled in each Figure.

The RMSE and MB error levels provide an incomplete picture of the error characteristics of ISOP. First, RMSE and MB error levels are elevated where the variance of SLD is large. For example, in the southern ocean in the vicinity of the Antarctic circumpolar current, RMSE and MB errors are

large because the depth variability of SLD is large there. Conversely, in the tropics where the SLD tends to be shallow, the errors are also shallow.

Another way to consider the ISOP error levels is to compare ISOP with the GDEM climatology error levels using skill score (SS) (Murphy 1988). This is done by comparing the mean square error (MSE) of the synthetic to the MSE of the GDEM climatology using the formula given by equation (110) described in section 6.4. Figure 31 shows the skill score for *Ideal* ISOP relative to GDEM temperature and salinity over the upper 1000 m. Warm colors indicate that ISOP has skill over climatology while grey indicates no improvement. Cool colors in figure 31 indicate regions where ISOP was worse than GDEM, which rarely occurs. The skill score for the *operational* case is in figure 32.

To demonstrate the global relative accuracy of each configuration, we take the median RMSE of the upper 1000 m and SS for T, S, and SSPD of all 10° by 10° regions (Table 10d). For every parameter, the Ideal ISOP case had the best performance except for salinity in the 0 to 50 m depth range. The next best performance is from ISOP with operation inputs.

To gain an understanding of the geographical distribution of where ISOP performs better then MODAS we have plotted the difference in T and S RMSE over the upper 1000 m depth range in Figures 33, 34, and 35.

6.6. Results: Test Case 2: Layer 3 Validation

Using the separate set of validation profiles that extend down to at least 1800 m, we evaluate the quality of the Layer 3 ISOP synthetics. The total number of profiles that extend to at least 1800 m is 52,480 from the year 2009 through 2001. We find that ISOP deep synthetic values are accurate within the expected error levels. The exception is two bad profiles, one in the south Mediterranean Sea and the other near Antarctica. To place the error levels within the context of the existing operational systems. ISOP errors are compared with error levels of MODAS and GDEM. The results suggest that while ISOP, MODAS, and GDEM are cool and fresh relative to more recent observations, data aliasing due to irregular sampling distribution could account for the bias.

The global deep validation data set initially contained 52.480 profiles with paired observed temperature (T) and salinity (S) values from the near surface (<12 m depth) and extend to at least 1800 m depth. The data set covers the time period from January 2009 through January 2011 (Figure 36) and all the profiles are independent from those used to construct the ISOP database. Of the original 52,480 profile 52,288 profiles were used to perform the validation analysis. The 192 profile (0.4%) that were dropped from the analysis either did not have a GOFS, MODAS, and GDEM profile pair, or the inputs for ISOP for those profiles did not fall within acceptable parameter ranges. The geographical distribution of the observation profiles in 1/2 ° bins is shown in Figure 37.

We evaluate the skill of operational ISOP ($ISOP^{Oper}$) compared to operational MODAS ($MODAS^{oper}$) and *GDEM* (see Table 7) in the depth range of Layer 3 in three ways. First, histograms are computed of median bias in the depth range from 1000 to 1400 m. Second, the median bias is plotted geographically also for the depth range from 1000 to 1400 m. Finally, we computed the root mean square (RMS) error versus depth below 1000 m.

For the median bias in the depth range from 1000 to 1400 m, both T and S values for each paired set observation and synthetic profiles from $ISOP^{Oper}$, $MODAS^{oper}$, and *GDEM* are included in a histogram analysis shown in Figures 38 and 39. The median is computed as the median bias over the 1000 to 1400 m depth range from each profile pair. These methods give a T and S bias number for each observation and synthetic profile pair. All of the profile pairs are then included in the histogram for each synthetic type. We find a cool and fresh bias for the synthetic profiles in this depth range. Since $ISOP^{Oper}$ has the largest number of profile in the low bias bins, $ISOP^{Oper}$ therefore has the lowest bias

error in this depth range. This can be seen more clearly in the next figures that show the geographical error of T and S bias.

The 1000 to 1400 m T in $ISOP^{Oper}$ tends to have a cool bias globally (Figure 40). The bias is particularly large in the northeast Atlantic Ocean. There is a slight warm bias in the southern Indian Ocean. The 1000 to 1400 m T bias in $MODAS^{oper}$ and *GDEM* has a similar pattern but tends to be worse (Figures 41 and 42). $MODAS^{oper}$ has a larger region of warm bias in the south Indian Ocean. All synthetics have a cool bias near the Antarctic Circumpolar Current region.

A similar bias pattern is seen in the salinity values (Figure 43, 44 and 45). Of particular interest is the larger fresh bias in the northeast Atlantic Ocean. The source of this bias is under investigation.

RMS errors with depth for temperature, salinity, and sound speed are shown in Figures 46, 47 and 48, respectively.

The slight increase in errors near 1200 m is due to a few very large outliers in two small isolated regions of the ocean. One is in the southern part of the Mediterranean Sea (Figure 49) and the other is near Antarctica (Figure 50).

6.7. Results: Gulf of Mexico Validation

This section describes experiments to validate the ISOP synthetics in the Gulf of Mexico (GOM). The observational data was made during 2010 and most was taken following the oil spill of the Deepwater Horizon drilling platform.

This report details the performance of ISOP stand-alone synthetics in the GOM region relative to 4001 in situ observation profiles from July 2008 through October 2011 (Figure 51). The latitude and longitude coordinates of the region are 18°N to 32°N and 100°W to 77°W.

The validation test cases for the stand-alone ISOP synthetics is described in Table 7. The goal is to provide an ISOP synthetic test case for and ideal configuration and a case that is characteristic of and operational implementation of ISOP. For comparison, MODAS synthetics are also provided, also designed to have the characteristics of an ideal and operational implementation of MODAS (Table 7).

Results indicate substantial improvement of ISOP over MODAS in the GOM for T (Figure 52), S (Figure 53), and sound speed (Figure 54). For each case, the standard error (equation 112) is plotted in dashed lines plus and minus on either side of the main error estimate. Since some of the observations may not be completely independent, standard error is an under-estimate of error uncertainty. Since the difference between each case are far from overlapping within the standard error estimates, Figures 52 through 54 suggests that the improvement of ISOP over MODAS accuracy is likely statistically significant.

Below 1000 m all synthetics have similar error levels. Because ISOP is designed to improve the accuracy of the shape of synthetic profiles, we compare ISOP mixed layer depth (MLD) with the observations (Figure 55). In general, ISOP has a shallow bias with MLD.

To summarize the validation results for SLD, MLD, and BLG, Tables 11, 12 and 13 show the misfit error synthetic-observation for all test cases in Table 7. Green shading indicates the smallest misfit error percentages. ISOP performed better than MODAS in all cases except for BLG median % misfit. The relative % misfit for SLD, MLD, and BLG between ISOP and MODAS and ISOP and GDEM is shown in Tables 14, 15, and 16. We find that ISOP has smaller error levels for all cases accept the median difference for SLD.

6.7.1. Profiles in Water Less Than 1000 m Deep

To test the quality of the ISOP synthetics where the water depth is shallower than ISOP layer 2, we have taken a subset of 1202 GOM validation profiles that are in water less than 1000 m (Figure 56).

Figures 57 and 58 show the RMS misfit error for T and S versus depth, respectively. For each case, the standard error is plotted in dashed lines plus or minus on either side of the main error estimate. Results indicate that ISOP accuracy is nearly equal to climatology and MODAS in shallow water. The magnitude of the error is also slightly larger in the upper 120 m. This result indicates that shallow water is difficult to reproduce synthetically due to complications in satellite altimetry near the coast and other factors.

6.8. Results: RIMPAC Validation

This describes experiments to validate the ISOP synthetics surrounding Hawaii, a region known as the RIMPAC. This region does not have a large sea surface height signal, and therefore does not have the type of dynamics that are easily accounted for in synthetic generation. Regardless of this fact, ISOP synthetics are preforming as well as to be expected in this region. Results show that ISOP performs slightly better than MODAS, which performs slightly better than GDEM climatology.

This report details the performance of ISOP stand-alone synthetics in the RIMPAC region relative to 3,227 observation profiles from 2009 through April 2012. Because, we did not have the Global Ocean Forecasting System (GOFS) for the entire time, there are only 814 profiles that were used in the validation (Figure 59). The latitude and longitude coordinates of the region are 18°N to 23°N and 161.5°W to 153.5°W.

The validation test case for the operational stand-alone ISOP synthetics is described in Table 17. The goal is to provide an ISOP synthetic test case that is characteristic of and operational implementation of ISOP. For comparison, MODAS synthetics are also provided, also designed to have the characteristics of an operational implementation of MODAS (Table 18).

The results of the validation experiment suggest that ISOP synthetics are as accurate as can be expected. The error values are slightly better than those for the Gulf of Mexico, for example. ISOP is more accurate than MODAS for temperature (T; Figure 60), salinity (S; Figure 61) and sound speed (SSPD; Figure 62) at all depths from 10m to 500 m. For each case, the standard error is plotted in dashed lines plus and minus on either side of the main error estimate. Since the difference between each case are far from overlapping within the standard error estimates, Figures 60 through 62 suggests that the improvement of ISOP over MODAS accuracy is likely statistically significant.

7. Summary and Conclusions

The Navy's global ocean forecasting system (GOFS) relies more heavily on space based observations of the ocean surface than any other in situ observation. The reason for this is the relative abundance of surface observations compared to sub-surface observations, especially in near-real-time. For this reason, the ISOP system will provide an essential capability designed to project surface ocean information into the subsurface for data assimilation.

The ISOP system relies on a large database of historically observed in situ ocean profile observations of paired temperature and salinity values. These data are used to construct both T and S climatology and a six-mode EOF representation of the vertical structure error covariance. Both the climatology and the error covariance include the vertical gradients of T and S. The vertical gradients are a unique aspect of this system that is a departure from prior art on the subject. Due to the importance of ocean vertical gradients for acoustic transmission, the U.S. Navy is highly interested in accurate vertical gradient prediction.

This document details the methods for constructing the synthetics in addition to creating the ISOP databases of vertical structure eigenvalues and eigenvectors and the T and S climatology. The equations are derived in detail with an appendix designed to help interpret the computer code

implementation of the equations. The terms of the equations are labeled for reference to computer code that computes each term in the operational prototype system.

The methods described in this document are identical to the methods used for creating ISOP synthetics within the Navy Coupled Ocean Data Assimilation (NCODA) system. Validation of ISOP cycling with a models running NCODA is contained in a separate document titled "Validation Test Report for the Improved Synthetic Ocean Profile (ISOP) System, Part II: Synthetic Profile Assimilation in Navy Ocean Forecast Systems."

The validation experiments in this report suggest that ISOP has substantial skill improvement over MODAS. In Figure 52, 53, 54, 60, 61, and 62 the standard errors of the misfit estimates themselves for each test case are shown. Since the difference in error between ISOP and MODAS far exceeds this standard error limits, the improvement of ISOP over MODAS is very likely statistically significant. The same statistical significance is likely throughout the validation analysis.

Acknowledgements

We would like to thank the Office of Naval Research for funding the Naval Research Laboratory project The Impact of Spice on Ocean Circulation, program element 61153N. The authors also thank Dennis Krynen and Frank Bub for helpful guidance and suggestions. This is a Naval Research Laboratory Memorandum Report NRL/MR/7320—13-9364.

Appendix: Cost function System of Equations

We must take the partial derivative of the cost function (13) with respect to the *m* EOF amplitudes *a*, with respect to the *m* vertical difference EOF amplitudes *b*, and with respect to the $\hat{T}'$ and $\hat{S}'$ at each of the *N* depth levels. The cost function has ten terms,

$$J = J_1 + J_2 + J_3 + J_4 + J_5 + J_6 + J_7 + J_8 + J_9 + J_{10}$$

which are

$$J_1 = \left(x - x_b\right)^T B^{-1} \left(x - x_b\right) = a^T \Lambda^{-1} a = \sum_{i=1}^{m} \frac{a_i^2}{\lambda_i}$$

$$J_2 = \left(d - d_b\right)^T B_g^{-1} \left(d - d_b\right) = \sum_{i=1}^{m} \frac{b_i^2}{\mu_i}$$

$$J_3 = \left(x_{fg} - x\right)^T R^{-1} \left(x_{fg} - x\right) = F_{fg} \sum_{i=1}^{m} \frac{\left(a_i^2 + a_{fg,i}^2 - 2a_i a_{fg}\right)}{\lambda_i}$$

$$J_4 = \left(d_{fg} - d\right)^T R_g^{-1} \left(d_{fg} - d\right) = G_{fg} \sum_{i=1}^{m} \frac{\left(b_i^2 + b_{fg,i}^2 + 2b_i b_{fg}\right)}{\mu_i}$$

$$J_5 = \sum_{i=1}^{N} \left(\frac{T_i' - \hat{T}_i'}{u_i} \right)^2 = \sum_{i=1}^{N} \left(\frac{\sum_{k=1}^{m} \left(u_i a_k \gamma_{ik}\right) - \hat{T}_i'}{u_i} \right)^2$$

$$J_6 = \sum_{i=1}^{N-1} \left(\frac{\Delta T_i' - \left(\hat{T}_{i+1}' - \hat{T}_i'\right)}{w_i} \right)^2 = \sum_{i=1}^{N-1} \left(\frac{\sum_{k=1}^{m} \left(w_i b_k \varphi_{ik}\right) - \left(\hat{T}_{i+1}' - \hat{T}_i'\right)}{w_i} \right)^2$$

$$J_7 = \sum_{i=1}^{N} \left(\frac{S_i' - \hat{S}_i'}{u_{i+N}} \right)^2 = \sum_{i=1}^{N} \left(\frac{\sum_{k=1}^{m} \left(u_{i+N} a_k \gamma_{i+Nk}\right) - \hat{S}_i'}{u_{i+N}} \right)^2$$

$$J_8 = \sum_{i=1}^{N+1} \left(\frac{\Delta S_i' - \left(\hat{S}_{i+1}' - \hat{S}_i'\right)}{w_{i+N-1}} \right)^2 = \sum_{i=1}^{N-1} \left(\frac{\sum_{k=1}^{m} \left(w_{i+N-1} b_k \varphi_{i+n-1k}\right) - \left(\hat{S}_{i+1}' - \hat{S}'\right)}{w_{i+N-1}} \right)^2$$

$$J_9 = \frac{\left(\tilde{T}_{MLD}' - \hat{T}_{MLD}'\right)^2}{\varepsilon_{SST}^2} = \frac{\left(\tilde{T}_{MLD}' - \sum_{i=1}^{N} Q_i \hat{T}_i'\right)^2}{\varepsilon_{SST}^2}$$

$$J_{10} = \frac{\left(\tilde{h}_{MLD} - \hat{h}_{MLD}\right)^2}{\varepsilon_h^2} = \frac{\left(\tilde{h}_{MLD} - \sum_{i=1}^{N} \left[\left(\delta_{b,i} + \alpha_i \hat{T}_i' + \beta_i \hat{T}_i'\right) dz_i \right]\right)^2}{\varepsilon_h^2}$$

and each of these terms can be operated on by each of the *2m+2N* partial derivatives. The minimized solution occurs when each of the *2m+2N* partial derivative equations are each equal to zero. For example, the equation for the partial derivative of *J* with respect to a_p is

$$\frac{\partial J}{\partial a_p} = \frac{\partial J_1}{\partial a_p} + \frac{\partial J_2}{\partial a_p} + \frac{\partial J_3}{\partial a_p} + \frac{\partial J_4}{\partial a_p} + \frac{\partial J_5}{\partial a_p} + \frac{\partial J_6}{\partial a_p} + \frac{\partial J_7}{\partial a_p} + \frac{\partial J_8}{\partial a_p} + \frac{\partial J_9}{\partial a_p} + \frac{\partial J_{10}}{\partial a_p}. \quad \text{(A1)}$$

Since there are *m* modes, the above expression represent *m* equations. The *m* equations for b_p are

$$\frac{\partial J}{\partial b_p} = \frac{\partial J_1}{\partial b_p} + \frac{\partial J_2}{\partial b_p} + \frac{\partial J_3}{\partial b_p} + \frac{\partial J_4}{\partial b_p} + \frac{\partial J_5}{\partial b_p} + \frac{\partial J_6}{\partial b_p} + \frac{\partial J_7}{\partial b_p} + \frac{\partial J_8}{\partial b_p} + \frac{\partial J_9}{\partial b_p} + \frac{\partial J_{10}}{\partial b_p}. \quad \text{(A2)}$$

The 2*N* equations for *T* and *S* are

$$\frac{\partial J}{\partial \hat{T}'_q} = \frac{\partial J_1}{\partial \hat{T}'_q} + \frac{\partial J_2}{\partial \hat{T}'_q} + \frac{\partial J_3}{\partial \hat{T}'_q} + \frac{\partial J_4}{\partial \hat{T}'_q} + \frac{\partial J_5}{\partial \hat{T}'_q} + \frac{\partial J_6}{\partial \hat{T}'_q} + \frac{\partial J_7}{\partial \hat{T}'_q} + \frac{\partial J_8}{\partial \hat{T}'_q} + \frac{\partial J_9}{\partial \hat{T}'_q} + \frac{\partial J_{10}}{\partial \hat{T}'_q}. \quad \text{(A3)}$$

and

$$\frac{\partial J}{\partial \hat{S}'_q} = \frac{\partial J_1}{\partial \hat{S}'_q} + \frac{\partial J_2}{\partial \hat{S}'_q} + \frac{\partial J_3}{\partial \hat{S}'_q} + \frac{\partial J_4}{\partial \hat{S}'_q} + \frac{\partial J_5}{\partial \hat{S}'_q} + \frac{\partial J_6}{\partial \hat{S}'_q} + \frac{\partial J_7}{\partial \hat{S}'_q} + \frac{\partial J_8}{\partial \hat{S}'_q} + \frac{\partial J_9}{\partial \hat{S}'_q} + \frac{\partial J_{10}}{\partial \hat{S}'_q}. \quad \text{(A4)}$$

When the *2m+2N* equations (A1, A2, A3, and A4) are all equal to zero, the solution is obtained. Below are each of the ten cost function terms differentiated with respect to each a_p, b_p, $\hat{T}'_q$, or $\hat{S}'_q$. Many of the derivatives are zero because some of the terms do not depend on some of the EOF amplitudes or $\hat{T}'_q$, or $\hat{S}'_q$. Only the non-zero terms are shown. To keep things straight, the index for the derivative mode is *p* while the cost function term index for mode is *k*. The index for the derivative depth level is q, while the index for the cost function term is i. That way we can keep track of the indexes, which is important.

Term J_1*:*

$$J_1 = \left(x - x_b\right)^T B^{-1} \left(x - x_b\right) = a^T \Lambda^{-1} a = \sum_{i=1}^{m} \frac{a_i^2}{\lambda_i}$$

$$\frac{\partial J_1}{\partial a_p} = 2\frac{a_p}{\lambda_p}$$

Term J_2*:*

$$J_2 = \left(d - d_b\right)^T B_g^{-1} \left(d - d_b\right) = \sum_{i=1}^{m} \frac{b_i^2}{\mu_i}$$

$$\frac{\partial J_2}{\partial b_p} = 2\frac{b_p}{\mu_p}$$

Term J_3:

$$J_3 = \left(x_{fg} - x\right)^T R^{-1} \left(x_{fg} - x\right) = F_{fg} \sum_{i=1}^{m} \frac{\left(a_i^2 + a_{fg,i}^2 - 2a_i a_{fg,i}\right)}{\lambda_i}$$

$$\frac{\partial J_3}{\partial a_p} = F_{fg} \frac{\partial}{\partial a_p} \sum_{i=1}^{m} \frac{\left(a_i^2 + a_{fg,i}^2 - 2a_i a_{fg,i}\right)}{\lambda_i}$$

Notice that we will only get non-zero contributions when $i=p$.

$$\frac{\partial J_3}{\partial a_p} = F_{fg} \frac{\left(2a_p - 2a_{fg,p}\right)}{\lambda_p}$$

$$\frac{\partial J_3}{\partial a_p} = 2F_{fg} \frac{a_p}{\lambda_p} - 2F_{fg} \frac{a_{fg,p}}{\lambda_p}$$

Term J_4:

$$J_4 = \left(d_{fg} - d\right)^T R_g^{-1} \left(d_{fg} - d\right) = G_{fg} \sum_{i=1}^{m} \frac{\left(b_i^2 + b_{fg,i}^2 - 2b_i b_{fg,i}\right)}{\mu_i}$$

$$\frac{\partial J_4}{\partial b_p} = G_{fg} \frac{\partial}{\partial b_p} \sum_{i=1}^{m} \frac{\left(b_i^2 + b_{fg,i}^2 - 2b_i b_{fg,i}\right)}{\mu_i}$$

Notice that we will only get a non-zero contribution when $i=p$.

$$\frac{\partial J_4}{\partial b_p} = G_{fg} \frac{\left(2b_p - 2b_{fg,p}\right)}{\mu_p}$$

$$\frac{\partial J_4}{\partial b_p} = 2G_{fg} \frac{b_p}{\mu_p} - 2G_{fg} \frac{b_{fg,p}}{\mu_p}$$

Term J_5:

$$J_5 = \sum_{i=1}^{N}\left(\frac{T_i' - \hat{T}_i'}{u_i}\right) = \sum_{i=1}^{N}\left(\frac{\sum_{k=1}^{m}\left(u_i a_k \gamma_{ik}\right) - \hat{T}_i'}{u_i}\right)^2$$

$$\frac{\partial J_5}{\partial a_p} = \frac{\partial}{\partial a_p}\sum_{i=1}^{N}\left(\frac{\sum_{k=1}^{m}\left(u_i a_k \gamma_{ik}\right) - \hat{T}_i'}{u_i}\right)^2$$

$$\frac{\partial J_5}{\partial a_p} = \sum_{i=1}^{N}\frac{\partial}{\partial a_p}\left(\frac{\sum_{k=1}^{m}\left(u_i a_k \gamma_{ik}\right) - \hat{T}_i'}{u_i}\right)^2$$

$$\frac{\partial J_5}{\partial a_p} = 2\sum_{i=1}^{N}\left(\frac{\sum_{k=1}^{m}\left(u_i a_k \gamma_{ik}\right) - \hat{T}_i'}{u_i}\right)\frac{\partial}{\partial a_p}\left(\frac{\sum_{k=1}^{m}\left(u_i a_k \gamma_{ik}\right) - \hat{T}_i'}{u_i}\right)$$

Notice that there is a non-zero component only when the k sum on the right is equal to p.

$$\frac{\partial J_5}{\partial a_p} = 2\sum_{i=1}^{N}\left(\frac{\sum_{k=1}^{m}\left(u_i a_k \gamma_{ik}\right) - \hat{T}_i'}{u_i}\right)\gamma_{ip}$$

$$\frac{\partial J_5}{\partial a_p} = 2\sum_{i=1}^{N}\left(\sum_{k=1}^{m}\left(a_k \gamma_{ik}\gamma_{ip}\right) - \gamma_{ip}\frac{\hat{T}'}{u_i}\right)$$

$$\frac{\partial J_5}{\partial a_p} = 2\sum_{i=1}^{N}\left(\sum_{k=1}^{m}\left(a_k \gamma_{ik}\gamma_{ip}\right)\right) - 2\sum_{i=1}^{N}\gamma_{ip}\frac{\hat{T}'}{u_i}$$

Swap the order of the summations. That will be important for when we construct the matrix equation for solving.

$$\frac{\partial J_5}{\partial a_p} = 2\sum_{k=1}^{m}\left(a_k \sum_{i=1}^{N}\left(\gamma_{ik}\gamma_{ip}\right)\right) - 2\sum_{i=1}^{N}\gamma_{ip}\frac{\hat{T}'}{u_i}$$

Now for the J_5 term with respect to T. Start with

$$\frac{\partial J_5}{\partial \hat{T}_q'} = \sum_{i=1}^{N} \frac{\partial}{\partial \hat{T}_q'} \left(\frac{\sum_{k=1}^{m} \left(u_i a_k \gamma_{ik} \right) - \hat{T}_i'}{u_i} \right)^2$$

$$\frac{\partial J_5}{\partial \hat{T}_q'} = 2 \sum_{i=1}^{N} \left(\frac{\sum_{k=1}^{m} \left(u_i a_k \gamma_{ik} \right) - \hat{T}_i'}{u_i} \right) \frac{\partial}{\partial \hat{T}_q'} \left(\frac{\sum_{k=1}^{m} \left(u_i a_k \gamma_{ik} \right) - \hat{T}_i'}{u_i} \right)$$

Notice that we are only going to get non-zero contributions when i=q.

$$\frac{\partial J_5}{\partial \hat{T}_q'} = 2 \left(\frac{\sum_{k=1}^{m} \left(u_q a_k \gamma_{qk} \right) - \hat{T}_q'}{u_q} \right) \frac{-1}{u_q}$$

$$\frac{\partial J_5}{\partial \hat{T}_q'} = -2 \left(\frac{\sum_{k=1}^{m} \left(u_q a_k \gamma_{qk} \right) - \hat{T}_q'}{u_q^{\,2}} \right)$$

$$\frac{\partial J_5}{\partial \hat{T}_q'} = -\frac{2}{u_q} \sum_{k=1}^{m} \left(a_k \gamma_{qk} \right) + 2 \frac{\hat{T}_q'}{u_q^{\,2}}$$

Term J_6:

$$J_6 = \sum_{i=1}^{N-1}\left(\frac{\Delta T_i' - \left(\hat{T}_{i+1}' - \hat{T}_i'\right)}{w_i}\right)^2 = \sum_{i=1}^{N-1}\left(\frac{\sum_{k=1}^{m}\left(w_i b_k \phi_{ik}\right) - \left(\hat{T}_{i+1}' - \hat{T}_i'\right)}{w_i}\right)^2$$

$$\frac{\partial J_6}{\partial b_p} = \sum_{i=1}^{N-1}\frac{\partial}{\partial b_p}\left(\frac{\sum_{k=1}^{m}\left(w_i b_k \phi_{ik}\right) - \left(\hat{T}_{i+1}' - \hat{T}_i'\right)}{w_i}\right)^2$$

$$\frac{\partial J_6}{\partial b_p} = 2\sum_{i=1}^{N-1}\left(\frac{\sum_{k=1}^{m}\left(w_i b_k \phi_{ik}\right) - \left(\hat{T}_{i+1}' - \hat{T}_i'\right)}{w_i}\right)\frac{\partial}{\partial b_p}\left(\frac{\sum_{k=1}^{m}\left(w_i b_k \phi_{ik}\right) - \left(\hat{T}_{i+1}' - \hat{T}_i'\right)}{w_i}\right)$$

$$\frac{\partial J_6}{\partial b_p} = 2\sum_{i=1}^{N-1}\left(\frac{\sum_{k=1}^{m}\left(w_i b_k \phi_{ik}\right) - \left(\hat{T}_{i+1}' - \hat{T}_i'\right)}{w_i}\right)\phi_{ip}$$

$$\frac{\partial J_6}{\partial b_p} = 2\sum_{i=1}^{N-1}\left(\frac{\sum_{k=1}^{m}\left(w_i b_k \phi_{ik}\phi_{ip}\right) - \phi_{ip}\left(\hat{T}_{i+1}' - \hat{T}_i'\right)}{w_i}\right)$$

$$\frac{\partial J_6}{\partial b_p} = 2\sum_{i=1}^{N-1}\left(\sum_{k=1}^{m}\left(b_k \phi_{ik}\phi_{ip}\right) - \frac{\phi_{ip}}{w_i}\left(\hat{T}_{i+1}' - \hat{T}_i'\right)\right)$$

$$\frac{\partial J_6}{\partial b_p} = 2\sum_{k=1}^{m} b_k\left(\sum_{i=1}^{N-1}\left(\phi_{ik}\phi_{ip}\right)\right) - 2\sum_{i=1}^{N-1}\frac{\phi_{ip}}{w_i}\left(\hat{T}_{i+1}' - \hat{T}_i'\right)$$

Split out the terms so that all $\hat{T}'$ are in separate terms. That will be important for when we construct the matrix equation for solving.

$$\frac{\partial J_6}{\partial b_p} = 2\sum_{k=1}^{m} b_k\left(\sum_{i=1}^{N-1}\left(\phi_{ik}\phi_{ip}\right)\right) - 2\sum_{i=1}^{N-1}\frac{\phi_{ip}}{w_i}\hat{T}_{i+1}' + 2\sum_{i=1}^{N-1}\frac{\phi_{ip}}{w_i}\hat{T}_i'$$

Now with respect to T.

$$\frac{\partial J_6}{\partial \hat{T}'_q} = 2\sum_{i=1}^{N-1}\left(\frac{\sum_{k=1}^{m}\left(w_i b_k \phi_{ik}\right)-\left(\hat{T}'_{i+1}-\hat{T}'_i\right)}{w_i}\right)\frac{\partial}{\partial \hat{T}'_q}\left(\frac{\sum_{k=1}^{m}\left(w_i b_k \phi_{ik}\right)-\left(\hat{T}'_{i+1}-\hat{T}'_i\right)}{w_i}\right)$$

The only non-zero terms are those when i=q and i+1=q.

$$\frac{\partial J_6}{\partial \hat{T}'_q} = 2\left(\frac{\sum_{k=1}^{m}\left(w_q b_k \phi_{qk}\right)-\left(\hat{T}'_{q+1}-\hat{T}'_q\right)}{w_q}\right)\frac{\partial}{\partial \hat{T}'_q}\left(\frac{\sum_{k=1}^{m}\left(w_q b_k \phi_{qk}\right)-\left(\hat{T}'_{q+1}-\hat{T}'_q\right)}{w_q}\right)$$

$$+2\left(\frac{\sum_{k=1}^{m}\left(w_{q-1} b_k \phi_{q-1k}\right)-\left(\hat{T}'_q-\hat{T}'_{q-1}\right)}{w_{q-1}}\right)\frac{\partial}{\partial \hat{T}'_q}\left(\frac{\sum_{k=1}^{m}\left(w_{q-1} b_k \phi_{q-1k}\right)-\left(\hat{T}'_q-\hat{T}'_{q-1}\right)}{w_{q-1}}\right)$$

$$\frac{\partial J_6}{\partial \hat{T}'_q} = 2\left(\frac{\sum_{k=1}^{m}\left(w_q b_k \phi_{qk}\right)-\left(\hat{T}'_{q+1}-\hat{T}'_q\right)}{w_q}\right)\frac{1}{w_q}+2\left(\frac{\sum_{k=1}^{m}\left(w_{q-1} b_k \phi_{q-1k}\right)-\left(\hat{T}'_q-\hat{T}'_{q-1}\right)}{w_{q-1}}\right)\frac{-1}{w_{q-1}}$$

$$\frac{\partial J_6}{\partial \hat{T}'_q} = 2\left(\frac{\sum_{k=1}^{m}\left(w_q b_k \phi_{qk}\right)-\left(\hat{T}'_{q+1}-\hat{T}'_q\right)}{w_q^{\,2}}\right)-2\left(\frac{\sum_{k=1}^{m}\left(w_{q-1} b_k \phi_{q-1k}\right)-\left(\hat{T}'_q-\hat{T}'_{q-1}\right)}{w_{q-1}^{\,2}}\right)$$

$$\frac{\partial J_6}{\partial \hat{T}'_q} = \frac{2}{w_q}\sum_{k=1}^{m}\left(b_k \phi_{qk}\right)-2\frac{\left(\hat{T}'_{q+1}-\hat{T}'_q\right)}{w_q^{\,2}}-\frac{2}{w_{q-1}}\sum_{k=1}^{m}\left(b_k \phi_{q-1k}\right)+2\frac{\left(\hat{T}'_q-\hat{T}'_{q-1}\right)}{w_{q-1}^{\,2}}$$

$$\frac{\partial J_6}{\partial \hat{T}'_q} = 2\sum_{k=1}^{m} b_k\left(\frac{\phi_{qk}}{w_q}-\frac{\phi_{q-1k}}{w_{q-1}}\right)-2\frac{\left(\hat{T}'_{q+1}-\hat{T}'_q\right)}{w_q^{\,2}}+2\frac{\left(\hat{T}'_q-\hat{T}'_{q-1}\right)}{w_{q-1}^{\,2}}$$

$$\frac{\partial J_6}{\partial \hat{T}'_q} = 2\sum_{k=1}^{m} b_k\left(\frac{\phi_{qk}}{w_q}-\frac{\phi_{q-1k}}{w_{q-1}}\right)-2\frac{\hat{T}'_{q+1}}{w_q^{\,2}}+2\frac{\hat{T}'_q}{w_q^{\,2}}+2\frac{\hat{T}'_q}{w_{q-1}^{\,2}}-2\frac{\hat{T}'_{q-1}}{w_{q-1}^{\,2}}$$

Make is to that all $\hat{T}'$ are in terms by themselves. That will be important for when we construct the matrix equation for solving.

$$\frac{\partial J_6}{\partial \hat{T}'_q} = 2\sum_{k=1}^{m} b_k\left(\frac{\phi_{qk}}{w_q}-\frac{\phi_{q-1k}}{w_{q-1}}\right)-2\frac{\hat{T}'_{q+1}}{w_q^{\,2}}+2\hat{T}'_q\left(\frac{1}{w_q^{\,2}}+\frac{1}{w_{q-1}^{\,2}}\right)-2\frac{\hat{T}'_{q-1}}{w_{q-1}^{\,2}}$$

Term J_7:

$$J_7 = \sum_{i=1}^{N}\left(\frac{S_i' - \hat{S}_i'}{u_{i+N}}\right)^2 = \sum_{i=1}^{N}\left(\frac{\sum_{k=1}^{m}\left(u_{i+N}a_k\gamma_{i+Nk}\right) - \hat{S}_i'}{u_{i+N}}\right)^2$$

$$\frac{\partial J_7}{\partial a_p} = \sum_{i=1}^{N}\frac{\partial}{\partial a_p}\left(\frac{\sum_{k=1}^{m}\left(u_{i+N}a_k\gamma_{i+Nk}\right) - \hat{S}_i'}{u_{i+N}}\right)^2$$

$$\frac{\partial J_7}{\partial a_p} = 2\sum_{i=1}^{N}\left(\frac{\sum_{k=1}^{m}\left(u_{i+N}a_k\gamma_{i+Nk}\right) - \hat{S}_i'}{u_{i+N}}\right)\frac{\partial}{\partial a_p}\left(\frac{\sum_{k=1}^{m}\left(u_{i+N}a_k\gamma_{i+Nk}\right) - \hat{S}_i'}{u_{i+N}}\right)$$

$$\frac{\partial J_7}{\partial a_p} = 2\sum_{i=1}^{N}\left(\frac{\sum_{k=1}^{m}\left(u_{i+N}a_k\gamma_{i+Nk}\right) - \hat{S}_i'}{u_{i+N}}\right)\gamma_{i+Np}$$

$$\frac{\partial J_7}{\partial a_p} = 2\sum_{i=1}^{N}\left(\sum_{k=1}^{m}\left(a_k\gamma_{i+Nk}\right)\gamma_{i+Np} - \gamma_{i+Np}\frac{\hat{S}_i'}{u_{i+N}}\right)$$

$$\frac{\partial J_7}{\partial a_p} = 2\sum_{i=1}^{N}\left(\sum_{k=1}^{m}\left(a_k\gamma_{i+Nk}\right)\gamma_{i+Np}\right) - 2\sum_{i=1}^{N}\left(\gamma_{i+Np}\frac{\hat{S}_i'}{u_{i+N}}\right)$$

Re order the summations so that the summation over the *m* modes is on the outside. That will be important for when we construct the matrix equation for solving.

$$\frac{\partial J_7}{\partial a_p} = 2\sum_{k=1}^{m}a_k\sum_{i=1}^{N}\left(\gamma_{i+Nk}\gamma_{i+Np}\right) - 2\sum_{i=1}^{N}\left(\gamma_{i+Np}\frac{\hat{S}_i'}{u_{i+N}}\right)$$

Now with respect to S.

$$\frac{\partial J_7}{\partial \hat{S}'_q} = 2\sum_{i=1}^{N}\left(\frac{\sum_{k=1}^{m}\left(u_{i+N}a_k\gamma_{i+Nk}\right)-\hat{S}'_i}{u_{i+N}}\right)\frac{\partial}{\partial \hat{S}'_q}\left(\frac{\sum_{k=1}^{m}\left(u_{i+N}a_k\gamma_{i+Nk}\right)-\hat{S}'_i}{u_{i+N}}\right)$$

$$\frac{\partial J_7}{\partial \hat{S}'_q} = 2\left(\frac{\sum_{k=1}^{m}\left(u_{q+N}a_k\gamma_{q+Nk}\right)-\hat{S}'_q}{u_{q+N}}\right)\frac{-1}{u_{q+N}}$$

$$\frac{\partial J_7}{\partial \hat{S}'_q} = -2\left(\frac{\sum_{k=1}^{m}\left(u_{q+N}a_k\gamma_{q+Nk}\right)-\hat{S}'_q}{u_{q+N}^{\,2}}\right)$$

$$\frac{\partial J_7}{\partial \hat{S}'_q} = -2\left(\frac{\sum_{k=1}^{m}\left(a_k\gamma_{q+Nk}\right)}{u_{q+N}}\right)+2\frac{\hat{S}'_q}{u_{q+N}^{\,2}}$$

$$\frac{\partial J_7}{\partial \hat{S}'_q} = -2\sum_{k=1}^{m}\left(a_k\frac{\gamma_{q+Nk}}{u_{q+N}}\right)+2\frac{\hat{S}'_q}{u_{q+N}^{\,2}}$$

Term J_8:

$$J_8 = \sum_{i=1}^{N-1}\left(\frac{\Delta S_i' - \left(\hat{S}_{i+1}' - \hat{S}_i'\right)}{w_{i+N-1}}\right)^2 = \sum_{i=1}^{N-1}\left(\frac{\sum_{k=1}^{m}\left(w_{i+N-1}b_k\phi_{i+N-1k}\right) - \left(\hat{S}_{i+1}' - \hat{S}_i'\right)}{w_{i+N-1}}\right)^2$$

$$\frac{\partial J_8}{\partial b_p} = \sum_{i=1}^{N-1}\frac{\partial}{\partial b_p}\left(\frac{\sum_{k=1}^{m}\left(w_{i+N-1}b_k\phi_{i+N-1k}\right) - \left(\hat{S}_{i+1}' - \hat{S}_i'\right)}{w_{i+N-1}}\right)^2$$

$$\frac{\partial J_8}{\partial b_p} = 2\sum_{i=1}^{N-1}\left(\frac{\sum_{k=1}^{m}\left(w_{i+N-1}b_k\phi_{i+N-1k}\right) - \left(\hat{S}_{i+1}' - \hat{S}_i'\right)}{w_{i+N-1}}\right)\frac{\partial}{\partial b_p}\left(\frac{\sum_{k=1}^{m}\left(w_{i+N-1}b_k\phi_{i+N-1k}\right) - \left(\hat{S}_{i+1}' - \hat{S}_i'\right)}{w_{i+N-1}}\right)$$

$$\frac{\partial J_8}{\partial b_p} = 2\sum_{i=1}^{N-1}\left(\frac{\sum_{k=1}^{m}\left(w_{i+N-1}b_k\phi_{i+N-1k}\right) - \left(\hat{S}_{i+1}' - \hat{S}_i'\right)}{w_{i+N-1}}\right)\phi_{i+N-1p}$$

$$\frac{\partial J_8}{\partial b_p} = 2\sum_{i=1}^{N-1}\left(\frac{\sum_{k=1}^{m}\left(w_{i+N-1}b_k\phi_{i+N-1k}\phi_{i+N-1p}\right)}{w_{i+N-1}}\right) - 2\sum_{i=1}^{N-1}\left(\frac{\phi_{i+N-1p}\left(\hat{S}_{i+1}' - \hat{S}_i'\right)}{w_{i+N-1}}\right)$$

$$\frac{\partial J_8}{\partial b_p} = 2\sum_{i=1}^{N-1}\left(\sum_{k=1}^{m}\left(b_k\phi_{i+N-1k}\phi_{i+N-1p}\right)\right) - 2\sum_{i=1}^{N-1}\left(\frac{\phi_{i+N-1p}\left(\hat{S}_{i+1}' - \hat{S}_i'\right)}{w_{i+N-1}}\right)$$

$$\frac{\partial J_8}{\partial b_p} = 2\sum_{k=1}^{m}\left(b_k\sum_{i=1}^{N-1}\left(\phi_{i+N-1k}\phi_{i+N-1p}\right)\right) - 2\sum_{i=1}^{N-1}\left(\frac{\phi_{i+N-1p}\left(\hat{S}_{i+1}' - \hat{S}_i'\right)}{w_{i+N-1}}\right)$$

$$\frac{\partial J_8}{\partial b_p} = 2\sum_{k=1}^{m}\left(b_k\sum_{i=1}^{N-1}\left(\phi_{i+N-1k}\phi_{i+N-1p}\right)\right) - 2\sum_{i=1}^{N-1}\left(\frac{\phi_{i+N-1p}\hat{S}_{i+1}'}{w_{i+N-1}}\right) + 2\sum_{i=1}^{N-1}\left(\frac{\phi_{i+N-1p}\hat{S}_i'}{w_{i+N-1}}\right)$$

Now with respect to S.

$$\frac{\partial J_8}{\partial \hat{S}'_q} = \sum_{i=1}^{N-1} \frac{\partial}{\partial \hat{S}'_q} \left(\frac{\sum_{k=1}^{m} \left(w_{i+N-1} b_k \phi_{i+N-1k} \right) - \left(\hat{S}'_{i+1} - \hat{S}'_i \right)}{w_{i+N-1}} \right)^2$$

$$\frac{\partial J_8}{\partial \hat{S}'_q} = \sum_{i=1}^{N-1} 2 \left(\frac{\sum_{k=1}^{m} \left(w_{i+N-1} b_k \phi_{i+N-1k} \right) - \left(\hat{S}'_{i+1} - \hat{S}'_i \right)}{w_{i+N-1}} \right) \frac{\partial}{\partial \hat{S}'_q} \left(\frac{\sum_{k=1}^{m} \left(w_{i+N-1} b_k \phi_{i+N-1k} \right) - \left(\hat{S}'_{i+1} - \hat{S}'_i \right)}{w_{i+N-1}} \right)$$

The only terms that have a non-zero contribution are when i=q and i=q-1.

$$\frac{\partial J_8}{\partial \hat{S}'_q} = 2 \left(\frac{\sum_{k=1}^{m} \left(w_{q+N-1} b_k \phi_{q+N-1k} \right) - \left(\hat{S}'_{q+1} - \hat{S}'_q \right)}{w_{q+N-1}} \right) \frac{\partial}{\partial \hat{S}'_q} \left(\frac{\sum_{k=1}^{m} \left(w_{q+N-1} b_k \phi_{q+N-1k} \right) - \left(\hat{S}'_{q+1} - \hat{S}'_q \right)}{w_{q+N-1}} \right)$$

$$+2 \left(\frac{\sum_{k=1}^{m} \left(w_{q-1+N-1} b_k \phi_{q-1+N-1k} \right) - \left(\hat{S}'_q - \hat{S}'_{q-1} \right)}{w_{q-1+N-1}} \right) \frac{\partial}{\partial \hat{S}'_q} \left(\frac{\sum_{k=1}^{m} \left(w_{q-1+N-1} b_k \phi_{q-1+N-1k} \right) - \left(\hat{S}'_q - \hat{S}'_{q-1} \right)}{w_{q-1+N-1}} \right)$$

$$\frac{\partial J_8}{\partial \hat{S}'_q} = 2 \left(\frac{\sum_{k=1}^{m} \left(w_{q+N-1} b_k \phi_{q+N-1k} \right) - \left(\hat{S}'_{q+1} - \hat{S}'_q \right)}{w_{q+N-1}} \right) \frac{1}{w_{q+N-1}} + 2 \left(\frac{\sum_{k=1}^{m} \left(w_{q-1+N-1} b_k \phi_{q-1+N-1k} \right) - \left(\hat{S}'_q - \hat{S}'_{q-1} \right)}{w_{q-1+N-1}} \right) \frac{-1}{w_{q-1+N-1}}$$

$$\frac{\partial J_8}{\partial \hat{S}'_q} = 2 \left(\frac{\sum_{k=1}^{m} \left(w_{q+N-1} b_k \phi_{q+N-1k} \right) - \left(\hat{S}'_{q+1} - \hat{S}'_q \right)}{w_{q+N-1}^2} \right) - 2 \left(\frac{\sum_{k=1}^{m} \left(w_{q-1+N-1} b_k \phi_{q-1+N-1k} \right) - \left(\hat{S}'_q - \hat{S}'_{q-1} \right)}{w_{q-1+N-1}^2} \right)$$

$$\frac{\partial J_8}{\partial \hat{S}'_q} = 2 \left(\frac{\sum_{k=1}^{m} \left(b_k \phi_{q+N-1k} \right)}{w_{q+N-1}} - \frac{\left(\hat{S}'_{q+1} - \hat{S}'_q \right)}{w_{q+N-1}^2} \right) - 2 \left(\frac{\sum_{k=1}^{m} \left(b_k \phi_{q-1+N-1k} \right)}{w_{q-1+N-1}} - \frac{\left(\hat{S}'_q - \hat{S}'_{q-1} \right)}{w_{q-1+N-1}^2} \right)$$

$$\frac{\partial J_8}{\partial \hat{S}'_q} = 2 \sum_{k=1}^{m} b_k \left(\frac{\phi_{q+N-1k}}{w_{q+N-1}} - \frac{\phi_{q-1+N-1k}}{w_{q-1+N-1}} \right) - 2 \frac{\hat{S}'_{q+1}}{w_{q+N-1}^2} + 2 \frac{\hat{S}'_q}{w_{q+N-1}^2} + 2 \frac{\hat{S}'_q}{w_{q-1+N-1}^2} - 2 \frac{\hat{S}'_{q-1}}{w_{q-1+N-1}^2}$$

$$\frac{\partial J_8}{\partial \hat{S}'_q} = 2 \sum_{k=1}^{m} b_k \left(\frac{\phi_{q+N-1k}}{w_{q+N-1}} - \frac{\phi_{q-1+N-1k}}{w_{q-1+N-1}} \right) - 2 \frac{\hat{S}'_{q+1}}{w_{q+N-1}^2} + 2 \hat{S}'_q \left(\frac{1}{w_{q+N-1}^2} + \frac{1}{w_{q-1+N-1}^2} \right) - 2 \frac{\hat{S}'_{q-1}}{w_{q-1+N-1}^2}$$

Term J_9:

$$J_9 = \frac{\left(\tilde{T}'_{SST} - \hat{T}'_{MLD}\right)^2}{\varepsilon_{SST}^2} = \frac{\left(\tilde{T}'_{SST} - \sum_{i=1}^{N} Q_i \hat{T}'_i\right)^2}{\varepsilon_{SST}^2}$$

$$\frac{\partial J_9}{\partial \hat{T}'_q} = \frac{\partial}{\partial \hat{T}'_q} \frac{\left(\tilde{T}'_{SST} - \sum_{i=1}^{N} Q_i \hat{T}'_i\right)^2}{\varepsilon_{SST}^2}$$

$$\frac{\partial J_9}{\partial \hat{T}'_q} = \frac{2}{\varepsilon_{SST}^2}\left(\tilde{T}'_{SST} - \sum_{i=1}^{N} Q_i \hat{T}'_i\right) \frac{\partial}{\partial \hat{T}'_q}\left(\tilde{T}'_{SST} - \sum_{i=1}^{N} Q_i \hat{T}'_i\right)$$

$$\frac{\partial J_9}{\partial \hat{T}'_q} = \frac{-2}{\varepsilon_{SST}^2}\left(\tilde{T}'_{SST} - \sum_{i=1}^{N} Q_i \hat{T}'_i\right) Q_q$$

$$\frac{\partial J_9}{\partial \hat{T}'_q} = -2\frac{Q_q \tilde{T}'_{SST}}{\varepsilon_{SST}^2} + 2\frac{Q_q}{\varepsilon_{SST}^2} \sum_{i=1}^{N} Q_i \hat{T}'_i$$

Term J_{10}:

$$J_{10} = \frac{\left(\tilde{h}_{MLD} - \hat{h}_{MLD}\right)^2}{\varepsilon_h^2} = \frac{\left(\tilde{h}_{MLD} - \sum_{i=1}^{N}\left[\left(\delta_{b,i} + \alpha_i \hat{T}'_i + \beta_i \hat{S}'_i\right) dz_i\right]\right)^2}{\varepsilon_h^2}$$

$$\frac{\partial J_{10}}{\partial \hat{T}'_q} = \frac{\partial}{\partial \hat{T}'_q} \frac{\left(\tilde{h}_{MLD} - \sum_{i=1}^{N}\left[\left(\delta_{b,i} + \alpha_i \hat{T}'_i + \beta_i \hat{S}'_i\right) dz_i\right]\right)^2}{\varepsilon_h^2}$$

$$\frac{\partial J_{10}}{\partial \hat{T}'_q} = \frac{2}{\varepsilon_h^2}\left(\tilde{h}_{MLD} - \sum_{i=1}^{N}\left[\left(\delta_{b,i} + \alpha_i \hat{T}'_i + \beta_i \hat{S}'_i\right) dz_i\right]\right) \frac{\partial}{\partial \hat{T}'_q}\left(\tilde{h}_{MLD} - \sum_{i=1}^{N}\left[\left(\delta_{b,i} + \alpha_i \hat{T}'_i + \beta_i \hat{S}'_i\right) dz_i\right]\right)$$

$$\frac{\partial J_{10}}{\partial \hat{T}'_q} = \frac{2}{\varepsilon_h^2}\left(\tilde{h}_{MLD} - \sum_{i=1}^{N}\left[\left(\delta_{b,i} + \alpha_i \hat{T}'_i + \beta_i \hat{S}'_i\right) dz_i\right]\right) \alpha_q dz_q$$

$$\frac{\partial J_{10}}{\partial \hat{T}'_q} = \frac{2}{\varepsilon_h^2}\left(\tilde{h}_{MLD} - \sum_{i=1}^{N} \delta_{b,i} dz_i\right) \alpha_q dz_q - \frac{2\alpha_q dz_q}{\varepsilon_h^2} \sum_{i=1}^{N} \alpha_i \hat{T}'_i dz_i - \frac{2\alpha_q dz_q}{\varepsilon_h^2} \sum_{i=1}^{N} \beta_i \hat{S}'_i dz_i$$

Now with respect to S

$$\frac{\partial J_{10}}{\partial \hat{S}'_q} = \frac{\partial}{\partial \hat{S}'_q} \frac{\left(\tilde{h}_{MLD} - \sum_{i=1}^{N} \left[\left(\delta_{b,i} + \alpha_i \hat{T}'_i + \beta_i \hat{S}'_i \right) dz_i \right] \right)^2}{\varepsilon_h^2}$$

$$\frac{\partial J_{10}}{\partial \hat{S}'_q} = \frac{2}{\varepsilon_h^2} \left(\tilde{h}_{MLD} - \sum_{i=1}^{N} \left[\left(\delta_{b,i} + \alpha_i \hat{T}'_i + \beta_i \hat{S}'_i \right) dz_i \right] \right) \frac{\partial}{\partial \hat{S}'_q} \left(\tilde{h}_{MLD} - \sum_{i=1}^{N} \left[\left(\delta_{b,i} + \alpha_i \hat{T}'_i + \beta_i \hat{S}'_i \right) dz_i \right] \right)$$

$$\frac{\partial J_{10}}{\partial \hat{S}'_q} = \frac{2}{\varepsilon_h^2} \left(\tilde{h}_{MLD} - \sum_{i=1}^{N} \delta_{b,i} dz_i \right) \beta_q dz_q - \frac{2\beta_q dz_q}{\varepsilon_h^2} \sum_{i=1}^{N} \alpha_i \hat{T}'_i dz_i - \frac{2\beta_q dz_q}{\varepsilon_h^2} \sum_{i=1}^{N} \beta_i \hat{S}'_i dz_i$$

The terms containing temperature or salinity observations ($\tilde{T}'$ or $\tilde{S}'$) are present for each standard depth only when an observation is available at that depth.

To keep track of all the terms, we list all the derivative terms in their final form here:

$$\frac{\partial J_1}{\partial a_p} = 2\frac{a_p}{\lambda_p}$$

$$\frac{\partial J_2}{\partial b_p} = 2\frac{b_p}{\mu_p}$$

$$\frac{\partial J_3}{\partial a_p} = 2F_{fg}\frac{a_p}{\lambda_p} - 2F_{fg}\frac{a_{fg,p}}{\lambda_p}$$

$$\frac{\partial J_4}{\partial b_p} = 2G_{fg}\frac{b_p}{\mu_p} - 2G_{fg}\frac{b_{fg,p}}{\mu_p}$$

$$\frac{\partial J_5}{\partial a_p} = 2\sum_{k=1}^{m}\left(a_k\sum_{i=1}^{N}\left(\gamma_{ik}\gamma_{ip}\right)\right) - 2\sum_{i=1}^{N}\gamma_{ip}\frac{\hat{T}'}{u_i}$$

$$\frac{\partial J_5}{\partial \hat{T}'_q} = -\frac{2}{u_q}\sum_{k=1}^{m}\left(a_k\gamma_{qk}\right) + 2\frac{\hat{T}'_q}{u_q{}^2}$$

$$\frac{\partial J_6}{\partial b_p} = 2\sum_{k=1}^{m}b_k\left(\sum_{i=1}^{N-1}\left(\phi_{ik}\phi_{ip}\right)\right) - 2\sum_{i=1}^{N-1}\frac{\phi_{ip}}{w_i}\hat{T}'_{i+1} + 2\sum_{i=1}^{N-1}\frac{\phi_{ip}}{w_i}\hat{T}'_i$$

$$\frac{\partial J_6}{\partial \hat{T}'_q} = 2\sum_{k=1}^{m}b_k\left(\frac{\phi_{qk}}{w_q} - \frac{\phi_{q-1k}}{w_{q-1}}\right) - 2\frac{\hat{T}'_{q+1}}{w_q{}^2} + 2\hat{T}'_q\left(\frac{1}{w_q{}^2} + \frac{1}{w_{q-1}{}^2}\right) - 2\frac{\hat{T}'_{q-1}}{w_{q-1}{}^2}$$

$$\frac{\partial J_7}{\partial a_p} = 2\sum_{k=1}^{m}a_k\sum_{i=1}^{N}\left(\gamma_{i+Nk}\gamma_{i+Np}\right) - 2\sum_{i=1}^{N}\left(\gamma_{i+Np}\frac{\hat{S}'_i}{u_{i+N}}\right)$$

$$\frac{\partial J_7}{\partial \hat{S}'_q} = -2\sum_{k=1}^{m}\left(a_k\frac{\gamma_{q+Nk}}{u_{q+N}}\right) + 2\frac{\hat{S}'_q}{u_{q+N}{}^2}$$

$$\frac{\partial J_8}{\partial b_p} = 2\sum_{k=1}^{m}\left(b_k\sum_{i=1}^{N-1}\left(\phi_{i+N-1k}\phi_{i+N-1p}\right)\right) - 2\sum_{i=1}^{N-1}\left(\frac{\phi_{i+N-1p}\hat{S}'_{i+1}}{w_{i+N-1}}\right) + 2\sum_{i=1}^{N-1}\left(\frac{\phi_{i+N-1p}\hat{S}'_i}{w_{i+N-1}}\right)$$

$$\frac{\partial J_8}{\partial \hat{S}'_q} = 2\sum_{k=1}^{m}b_k\left(\frac{\phi_{q+N-1k}}{w_{q+N-1}} - \frac{\phi_{q-1+N-1k}}{w_{q-1+N-1}}\right) - 2\frac{\hat{S}'_{q+1}}{w_{q+N-1}{}^2} + 2\hat{S}'_q\left(\frac{1}{w_{q+N-1}{}^2} + \frac{1}{w_{q-1+N-1}{}^2}\right) - 2\frac{\hat{S}'_{q-1}}{w_{q-1+N-1}{}^2}$$

$$\frac{\partial J_9}{\partial \hat{T}'_q} = -2\frac{Q_q\tilde{T}'_{SST}}{\varepsilon_{SST}^2} + 2\frac{Q_q}{\varepsilon_{SST}^2}\sum_{i=1}^{N}Q_i\hat{T}'_i$$

$$\frac{\partial J_{10}}{\partial \hat{T}'_q} = \frac{2}{\varepsilon_h^2}\left(\tilde{h}_{MLD} - \sum_{i=1}^{N}\delta_{b,i}dz_i\right)\alpha_q dz_q - \frac{2\alpha_q dz_q}{\varepsilon_h^2}\sum_{i=1}^{N}\alpha_i\hat{T}'_i dz_i - \frac{2\alpha_q dz_q}{\varepsilon_h^2}\sum_{i=1}^{N}\beta_i\hat{S}'_i dz_i$$

$$\frac{\partial J_{10}}{\partial \hat{S}'_q} = \frac{2}{\varepsilon_h^2}\left(\tilde{h}_{MLD} - \sum_{i=1}^{N}\delta_{b,i}dz_i\right)\beta_q dz_q - \frac{2\beta_q dz_q}{\varepsilon_h^2}\sum_{i=1}^{N}\alpha_i\hat{T}'_i dz_i - \frac{2\beta_q dz_q}{\varepsilon_h^2}\sum_{i=1}^{N}\beta_i\hat{S}'_i dz_i$$

Put it all together:

Now we need to put those all together to form the 2m+2N equation. The m equations for T and S EOF amplitudes are:

$$\frac{\partial J}{\partial a_p} = 2\frac{a_p}{\lambda_p} + 2F_{fg}\frac{a_p}{\lambda_p} - 2F_{fg}\frac{a_{fg,p}}{\lambda_p} + 2\sum_{k=1}^{m}\left(a_k\sum_{i=1}^{N}\left(\gamma_{ik}\gamma_{ip}\right)\right)$$
$$-2\sum_{i=1}^{N}\gamma_{ip}\frac{\hat{T}'}{u_i} + 2\sum_{k=1}^{m}a_k\sum_{i=1}^{N}\left(\gamma_{i+Nk}\gamma_{i+Np}\right) - 2\sum_{i=1}^{N}\left(\gamma_{i+Np}\frac{\hat{S}'_i}{u_{i+N}}\right)$$

The m equations for the T and S difference EOF amplitudes are:

$$\frac{\partial J}{\partial b_p} = 2\frac{b_p}{\mu_p} + 2G_{fg}\frac{b_p}{\mu_p} - 2G_{fg}\frac{b_{fg,p}}{\mu_p} + 2\sum_{k=1}^{m}b_k\left(\sum_{i=1}^{N}\left(\phi_{ik}\phi_{ip}\right)\right) - 2\sum_{i=1}^{N-1}\frac{\phi_{ip}}{w_i}\hat{T}'_{i+1} + 2\sum_{i=1}^{N-1}\frac{\phi_{ip}}{w_i}\hat{T}'_i$$
$$+2\sum_{k=1}^{m}\left(b_k\sum_{i=1}^{N-1}\left(\phi_{i+N-1k}\phi_{i+N-1p}\right)\right) - 2\sum_{i=1}^{N-1}\left(\frac{\phi_{i+N-1p}\hat{S}'_{i+1}}{w_{i+N-1}}\right) + 2\sum_{i=1}^{N-1}\left(\frac{\phi_{i+N-1p}\hat{S}'_i}{w_{i+N-1}}\right)$$

The N equations for the T analysis derivatives are:

$$\frac{\partial J}{\partial \hat{T}'_q} = -\frac{2}{u_q}\sum_{k=1}^{m}\left(a_k\gamma_{qk}\right) + 2\frac{\hat{T}'_q}{{u_q}^2} + 2\sum_{k=1}^{m}b_k\left(\frac{\phi_{qk}}{w_q} - \frac{\phi_{q-1k}}{w_{q-1}}\right) - 2\frac{\hat{T}'_{q+1}}{{w_q}^2} + 2\hat{T}'_q\left(\frac{1}{{w_q}^2} + \frac{1}{{w_{q-1}}^2}\right) - 2\frac{\hat{T}'_{q-1}}{{w_{q-1}}^2}$$
$$-2\frac{Q_q\tilde{T}'_{SST}}{\varepsilon^2_{SST}} + 2\frac{Q_q}{\varepsilon^2_{SST}}\sum_{i=1}^{N}Q_i\hat{T}'_i + \frac{2}{\varepsilon_h^2}\left(\tilde{h}_{MLD} - \sum_{i=1}^{N}\delta_{b,i}dz_i\right)\alpha_q dz_q - \frac{2\alpha_q dz_q}{\varepsilon_h^2}\sum_{i=1}^{N}\alpha_i\hat{T}'_i dz_i - \frac{2\alpha_q dz_q}{\varepsilon_h^2}\sum_{i=1}^{N}\beta_i\hat{S}'_i dz_i$$

The N equations for the S analysis derivatives are:

$$\frac{\partial J}{\partial \hat{S}'_q} = -2\sum_{k=1}^{m}\left(a_k\frac{\gamma_{q+Nk}}{u_{q+N}}\right) + 2\frac{\hat{S}'_q}{{u_{q+N}}^2} + 2\sum_{k=1}^{m}b_k\left(\frac{\phi_{q+N-1k}}{w_{q+N-1}} - \frac{\phi_{q-1+N-1k}}{w_{q-1+N-1}}\right) - 2\frac{\hat{S}'_{q+1}}{{w_{q+N-1}}^2} + 2\hat{S}'_q\left(\frac{1}{{w_{q+N-1}}^2} + \frac{1}{{w_{q-1+N-1}}^2}\right)$$
$$-2\frac{\hat{S}'_{q-1}}{{w_{q-1+N-1}}^2} + \frac{2}{\varepsilon_h^2}\left(\tilde{h}_{MLD} - \sum_{i=1}^{N}\delta_{b,i}dz_i\right)\beta_q dz_q - \frac{2\beta_q dz_q}{\varepsilon_h^2}\sum_{i=1}^{N}\alpha_i\hat{T}'_i dz_i - \frac{2\beta_q dz_q}{\varepsilon_h^2}\sum_{i=1}^{N}\beta_i\hat{S}'_i dz_i$$

Put in form Dg=e

Set to zero and put in the Da=e form

$$\frac{a_p}{\lambda_p}+F_{fg}\frac{a_p}{\lambda_p}+\sum_{k=1}^{m}\left(a_k\sum_{i=1}^{N}\left(\gamma_{ik}\gamma_{ip}\right)\right)-\sum_{i=1}^{N}\gamma_{ip}\frac{\hat{T}'}{u_i}+$$

$$\sum_{k=1}^{m}a_k\sum_{i=1}^{N}\left(\gamma_{i+Nk}\gamma_{i+Np}\right)-\sum_{i=1}^{N}\left(\gamma_{i+Np}\frac{\hat{S}'_i}{u_{i+N}}\right)=F_{fg}\frac{a_{fg,p}}{\lambda_p}$$

$$\boxed{\begin{aligned}&\sum_{k=1}^{m}\left(a_k\sum_{i=1}^{N}\left(\gamma_{ik}\gamma_{ip}\right)\right)+\sum_{k=1}^{m}\left(a_k\sum_{i=1}^{N}\left(\gamma_{i+Nk}\gamma_{i+Np}\right)\right)\\&-\sum_{i=1}^{N}\left(\gamma_{ip}\frac{\hat{T}'_i}{u_i}\right)-\sum_{i=1}^{N}\left(\gamma_{i+Np}\frac{\hat{S}'_i}{u_{i+N}}\right)+\frac{a_p}{\lambda_p}+F_{fg}\frac{a_p}{\lambda_p}=F_{fg}\frac{a_{fg,p}}{\lambda_p}\end{aligned}}$$

Next m equations:

$$2\frac{b_p}{\mu_p}+2G_{fg}\frac{b_p}{\mu_p}+2\sum_{k=1}^{m}b_k\left(\sum_{i=1}^{N}\left(\phi_{ik}\phi_{ip}\right)\right)-2\sum_{i=1}^{N-1}\frac{\phi_{ip}}{w_i}\hat{T}'_{i+1}+2\sum_{i=1}^{N-1}\frac{\phi_{ip}}{w_i}\hat{T}'_i$$

$$+2\sum_{k=1}^{m}\left(b_k\sum_{i=1}^{N-1}\left(\phi_{i+N-1k}\phi_{i+N-1p}\right)\right)-2\sum_{i=1}^{N-1}\left(\frac{\phi_{i+N-1p}\hat{S}'_{i+1}}{w_{i+N-1}}\right)+2\sum_{i=1}^{N-1}\left(\frac{\phi_{i+N-1p}\hat{S}'_i}{w_{i+N-1}}\right)=2G_{fg}\frac{b_{fg,p}}{\mu_p}$$

$$\frac{b_p}{\mu_p}+G_{fg}\frac{b_p}{\mu_p}+\sum_{k=1}^{m}b_k\left(\sum_{i=1}^{N}\left(\phi_{ik}\phi_{ip}\right)\right)-\sum_{i=1}^{N-1}\frac{\phi_{ip}}{w_i}\hat{T}'_{i+1}+\sum_{i=1}^{N-1}\frac{\phi_{ip}}{w_i}\hat{T}'_i$$

$$+\sum_{k=1}^{m}\left(b_k\sum_{i=1}^{N-1}\left(\phi_{i+N-1k}\phi_{i+N-1p}\right)\right)-\sum_{i=1}^{N-1}\left(\frac{\phi_{i+N-1p}\hat{S}'_{i+1}}{w_{i+N-1}}\right)+\sum_{i=1}^{N-1}\left(\frac{\phi_{i+N-1p}\hat{S}'_i}{w_{i+N-1}}\right)=G_{fg}\frac{b_{fg,p}}{\mu_p}$$

$$\boxed{\begin{aligned}&+\sum_{k=1}^{m}b_k\left(\sum_{i=1}^{N}\left(\phi_{ik}\phi_{ip}\right)\right)+\sum_{k=1}^{m}\left(b_k\sum_{i=1}^{N-1}\left(\phi_{i+N-1k}\phi_{i+N-1p}\right)\right)+\frac{b_p}{\mu_p}+G_{fg}\frac{b_p}{\mu_p}\\&-\sum_{i=1}^{N-1}\frac{\phi_{ip}}{w_i}\hat{T}'_{i+1}+\sum_{i=1}^{N-1}\frac{\phi_{ip}}{w_i}\hat{T}'_i-\sum_{i=1}^{N-1}\left(\frac{\phi_{i+N-1p}\hat{S}'_{i+1}}{w_{i+N-1}}\right)+\sum_{i=1}^{N-1}\left(\frac{\phi_{i+N-1p}\hat{S}'_i}{w_{i+N-1}}\right)=G_{fg}\frac{b_{fg,p}}{\mu_p}\end{aligned}}$$

Next N equations:

$$-\frac{1}{u_q}\sum_{k=1}^{m}\left(a_k\gamma_{qk}\right)+\frac{\hat{T}'_q}{u_q^{\,2}}+\sum_{k=1}^{m}b_k\left(\frac{\phi_{qk}}{w_q}-\frac{\phi_{q-1k}}{w_{q-1}}\right)-\frac{\hat{T}'_{q+1}}{w_q^{\,2}}+\hat{T}'_q\left(\frac{1}{w_q^{\,2}}+\frac{1}{w_{q-1}^{\,2}}\right)-\frac{\hat{T}'_{q-1}}{w_{q-1}^{\,2}}$$

$$-\frac{\alpha_q dz_q}{\varepsilon_h^2}\sum_{i=1}^{N}\alpha_i\hat{T}'_i dz_i-\frac{\alpha_q dz_q}{\varepsilon_h^2}\sum_{i=1}^{N}\beta_i\hat{S}'_i dz_i=\frac{Q_q\tilde{T}'_{SST}}{\varepsilon_{SST}^2}-\frac{Q_q}{\varepsilon_{SST}^2}\sum_{i=1}^{N}Q_i\hat{T}'_i-\frac{1}{\varepsilon_h^2}\left(\tilde{h}_{MLD}-\sum_{i=1}^{N}\delta_{b,i}dz_i\right)\alpha_q dz_q$$

$$-\frac{1}{u_q}\sum_{k=1}^{m}\left(a_k\gamma_{qk}\right)+\sum_{k=1}^{m}b_k\left(\frac{\phi_{qk}}{w_q}-\frac{\phi_{q-1k}}{w_{q-1}}\right)+\frac{\hat{T}'_q}{u_q^{\,2}}-\frac{\hat{T}'_{q+1}}{w_q^{\,2}}+\hat{T}'_q\left(\frac{1}{w_q^{\,2}}+\frac{1}{w_{q-1}^{\,2}}\right)-\frac{\hat{T}'_{q-1}}{w_{q-1}^{\,2}}$$

$$-\frac{Q_q}{\varepsilon_{SST}^2}\sum_{i=1}^{N}Q_i\hat{T}'_i-\frac{\alpha_q dz_q}{\varepsilon_h^2}\sum_{i=1}^{N}\alpha_i\hat{T}'_i dz_i-\frac{\alpha_q dz_q}{\varepsilon_h^2}\sum_{i=1}^{N}\beta_i\hat{S}'_i dz_i=\frac{Q_q\tilde{T}'_{SST}}{\varepsilon_{SST}^2}-\frac{1}{\varepsilon_h^2}\left(\tilde{h}_{MLD}-\sum_{i=1}^{N}\delta_{b,i}dz_i\right)\alpha_q dz_q$$

Final N equations

$$-2\sum_{k=1}^{m}\left(a_k\frac{\gamma_{q+Nk}}{u_{q+N}}\right)+2\frac{\hat{S}'_q}{u_{q+N}^{\,2}}+2\sum_{k=1}^{m}b_k\left(\frac{\phi_{q+N-1k}}{w_{q+N-1}}-\frac{\phi_{q-1+N-1k}}{w_{q-1+N-1}}\right)-2\frac{\hat{S}'_{q+1}}{w_{q+N-1}^{\,2}}+2\hat{S}'_q\left(\frac{1}{w_{q+N-1}^{\,2}}+\frac{1}{w_{q-1+N-1}^{\,2}}\right)$$

$$-2\frac{\hat{S}'_{q-1}}{w_{q-1+N-1}^{\,2}}-\frac{2\beta_q dz_q}{\varepsilon_h^2}\sum_{i=1}^{N}\alpha_i\hat{T}'_i dz_i-\frac{2\beta_q dz_q}{\varepsilon_h^2}\sum_{i=1}^{N}\beta_i\hat{S}'_i dz_i=-\frac{2}{\varepsilon_h^2}\left(\tilde{h}_{MLD}-\sum_{i=1}^{N}\delta_{b,i}dz_i\right)\beta_q dz_q$$

$$-\sum_{k=1}^{m}\left(a_k\frac{\gamma_{q+Nk}}{u_{q+N}}\right)+\sum_{k=1}^{m}b_k\left(\frac{\phi_{q+N-1k}}{w_{q+N-1}}-\frac{\phi_{q-1+N-1k}}{w_{q-1+N-1}}\right)+\frac{\hat{S}'_q}{u_{q+N}^{\,2}}-\frac{\hat{S}'_{q+1}}{w_{q+N-1}^{\,2}}+\hat{S}'_q\left(\frac{1}{w_{q+N-1}^{\,2}}+\frac{1}{w_{q-1+N-1}^{\,2}}\right)$$

$$-\frac{\hat{S}'_{q-1}}{w_{q-1+N-1}^{\,2}}-\frac{\beta_q dz_q}{\varepsilon_h^2}\sum_{i=1}^{N}\alpha_i\hat{T}'_i dz_i-\frac{\beta_q dz_q}{\varepsilon_h^2}\sum_{i=1}^{N}\beta_i\hat{S}'_i dz_i=-\frac{1}{\varepsilon_h^2}\left(\tilde{h}_{MLD}-\sum_{i=1}^{N}\delta_{b,i}dz_i\right)\beta_q dz_q$$

The equations used to code the minimization:

$$\sum_{k=1}^{m}\left(a_k\sum_{i=1}^{N}\left(\gamma_{ik}\gamma_{ip}\right)\right)+\sum_{k=1}^{m}\left(a_k\sum_{i=1}^{N}\left(\gamma_{i+Nk}\gamma_{i+Np}\right)\right)$$
$$-\sum_{i=1}^{N}\left(\gamma_{ip}\frac{\hat{T}_i'}{u_i}\right)-\sum_{i=1}^{N}\left(\gamma_{i+Np}\frac{\hat{S}_i'}{u_{i+N}}\right)+\frac{a_p}{\lambda_p}+F_{fg}\frac{a_p}{\lambda_p}=F_{fg}\frac{a_{fg,p}}{\lambda_p} \quad \text{(A1)}$$

$$+\sum_{k=1}^{m}b_k\left(\sum_{i=1}^{N}\left(\phi_{ik}\phi_{ip}\right)\right)+\sum_{k=1}^{m}\left(b_k\sum_{i=1}^{N-1}\left(\phi_{i+N-1k}\phi_{i+N-1p}\right)\right)+\frac{b_p}{\mu_p}+G_{fg}\frac{b_p}{\mu_p}$$
$$-\sum_{i=1}^{N-1}\frac{\phi_{ip}}{w_i}\hat{T}_{i+1}'+\sum_{i=1}^{N-1}\frac{\phi_{ip}}{w_i}\hat{T}_i'-\sum_{i=1}^{N-1}\left(\frac{\phi_{i+N-1p}\hat{S}_{i+1}'}{w_{i+N-1}}\right)+\sum_{i=1}^{N-1}\left(\frac{\phi_{i+N-1p}\hat{S}_i'}{w_{i+N-1}}\right)=G_{fg}\frac{b_{fg,p}}{\mu_p} \quad \text{(A2)}$$

$$-\frac{1}{u_q}\sum_{k=1}^{m}\left(a_k\gamma_{qk}\right)+\sum_{k=1}^{m}b_k\left(\frac{\phi_{qk}}{w_q}-\frac{\phi_{q-1k}}{w_{q-1}}\right)+\frac{\hat{T}_q'}{u_q{}^2}-\frac{\hat{T}_{q+1}'}{w_q{}^2}+\hat{T}_q'\left(\frac{1}{w_q{}^2}+\frac{1}{w_{q-1}{}^2}\right)-\frac{\hat{T}_{q-1}'}{w_{q-1}{}^2}$$
$$-\frac{Q_q}{\varepsilon_{SST}^2}\sum_{i=1}^{N}Q_i\hat{T}_i'-\frac{\alpha_q dz_q}{\varepsilon_h^2}\sum_{i=1}^{N}\alpha_i\hat{T}_i'dz_i-\frac{\alpha_q dz_q}{\varepsilon_h^2}\sum_{i=1}^{N}\beta_i\hat{S}_i'dz_i=\frac{Q_q\tilde{T}_{SST}'}{\varepsilon_{SST}^2}-\frac{1}{\varepsilon_h^2}\left(\tilde{h}_{MLD}-\sum_{i=1}^{N}\delta_{b,i}dz_i\right)\alpha_q dz_q \quad \text{(A3)}$$

$$-\sum_{k=1}^{m}\left(a_k\frac{\gamma_{q+Nk}}{u_{q+N}}\right)+\sum_{k=1}^{m}b_k\left(\frac{\phi_{q+N-1k}}{w_{q+N-1}}-\frac{\phi_{q-1+N-1k}}{w_{q-1+N-1}}\right)+\frac{\hat{S}_q'}{u_{q+N}{}^2}-\frac{\hat{S}_{q+1}'}{w_{q+N-1}{}^2}+\hat{S}_q'\left(\frac{1}{w_{q+N-1}{}^2}+\frac{1}{w_{q-1+N-1}{}^2}\right)$$
$$-\frac{\hat{S}_{q-1}'}{w_{q-1+N-1}{}^2}-\frac{\beta_q dz_q}{\varepsilon_h^2}\sum_{i=1}^{N}\alpha_i\hat{T}_i'dz_i-\frac{\beta_q dz_q}{\varepsilon_h^2}\sum_{i=1}^{N}\beta_i\hat{S}_i'dz_i=-\frac{1}{\varepsilon_h^2}\left(\tilde{h}_{MLD}-\sum_{i=1}^{N}\delta_{b,i}dz_i\right)\beta_q dz_q \quad \text{(A4)}$$

These equations (A1, A2, A3, and A4) represent the 2m+2N equations that make up the system of equations that are solved using the "Numerical Recipes" (1986) LUDCMP and LUBKSB functions for performing the LU decomposition. The equations need to be put in the form

$$\mathbf{D} \cdot \mathbf{g} = \mathbf{e} \tag{A5}$$

and are solved with calls to the functions like this:

call LUDCMP(**D**,N,NP,INDX,**S**)

call LUBKSB(**D**,N,NP,INDX,**e**)

At the start **D** and **e** are from the corresponding variables in equation A5. After the calls are made the solution g is returned in **e**.

For the case of ISOP, the order of the terms is defined by

$$\mathbf{g}^T = \left[a_1, a_2, \cdots, a_m, b_1, b_2, \cdots, b_m, \hat{T}'_1, \hat{T}'_2, \cdots, \hat{T}'_N, \hat{S}'_1, \hat{S}'_2, \cdots \hat{S}'_N \right].$$

Part of the solution is the EOF amplitudes. These amplitudes can be used to reconstruct the EOF representation of the synthetic. The right-hand-side of equation A5**, e**, is then

$$\mathbf{e} = \begin{bmatrix} F_{fg} a_{fg,1} / \lambda_1 \\ \vdots \\ F_{fg} a_{fg,m} / \lambda_m \\ G_{fg} b_{fg,1} / \mu_1 \\ \vdots \\ G_{fg} b_{fg,m} / \mu_m \\ \dfrac{Q_1 \tilde{T}'_{SST}}{\varepsilon^2_{SST}} - \dfrac{1}{\varepsilon^2_h} \left(\tilde{h}_{MLD} - \sum_{i=1}^{N} \delta_{b,i} dz_i \right) \alpha_1 dz_1 \\ \vdots \\ \dfrac{Q_N \tilde{T}'_{SST}}{\varepsilon^2_{SST}} - \dfrac{1}{\varepsilon^2_h} \left(\tilde{h}_{MLD} - \sum_{i=1}^{N} \delta_{b,i} dz_i \right) \alpha_N dz_N \\ -\dfrac{1}{\varepsilon^2_h} \left(\tilde{h}_{MLD} - \sum_{i=1}^{N} \delta_{b,i} dz_i \right) \beta_1 dz_1 \\ \vdots \\ -\dfrac{1}{\varepsilon^2_h} \left(\tilde{h}_{MLD} - \sum_{i=1}^{N} \delta_{b,i} dz_i \right) \beta_N dz_N \end{bmatrix}$$

For constructing the matrix **A**, let us start with A1 for p=1:

$$\sum_{k=1}^{m}\left(a_k\sum_{i=1}^{N}\left(\gamma_{ik}\gamma_{i1}\right)\right)+\sum_{k=1}^{m}\left(a_k\sum_{i=1}^{N}\left(\gamma_{i+Nk}\gamma_{i+N1}\right)\right)$$

$$-\sum_{i=1}^{N}\left(\gamma_{i1}\frac{\hat{T}_i'}{u_i}\right)+-\sum_{i=1}^{N}\left(\gamma_{i+N1}\frac{\hat{S}_i'}{u_{i+N}}\right)+\frac{a_1}{\lambda_1}+F_{fg}\frac{a_1}{\lambda_1}=F_{fg}\frac{a_{fg,1}}{\lambda_1}$$

This is the equation for the first row of the matrix.

$$\frac{1}{\lambda_1}+F_{fg}\frac{1}{\lambda_1}+\sum_{i=1}^{N}\left(\gamma_{i1}\gamma_{i1}\right)+\sum_{i=1}^{N}\left(\gamma_{i+N1}\gamma_{i+N1}\right),\sum_{i=1}^{N}\left(\gamma_{i2}\gamma_{i1}\right)+\sum_{i=1}^{N}\left(\gamma_{i+N2}\gamma_{i+N1}\right),\cdots,\sum_{i=1}^{N}\left(\gamma_{im}\gamma_{i1}\right)+\sum_{i=1}^{N}\left(\gamma_{i+Nm}\gamma_{i+N1}\right),$$

$$\gamma_{11}\frac{\hat{T}_1'}{u_1},\cdots,\gamma_{N1}\frac{\hat{T}_N'}{u_N},\gamma_{1+N1}\frac{\hat{S}_1'}{u_{1+N}},\cdots,\gamma_{N+N1}\frac{\hat{S}_N'}{u_{N+N}}$$

This is a lengthy process but I am unsure if it will help much. Instead, I'm going to label each term of equations A1, A2, A3, and A4. Then we can match each term up with the computer code.

The equations used to code the minimization with labels:

$$\overbrace{\sum_{k=1}^{m}\left(a_k\sum_{i=1}^{N}\left(\gamma_{ik}\gamma_{ip}\right)\right)}^{1.1}+\overbrace{\sum_{k=1}^{m}\left(a_k\sum_{i=1}^{N}\left(\gamma_{i+Nk}\gamma_{i+Np}\right)\right)}^{1.2}$$
$$\underbrace{-\sum_{i=1}^{N}\left(\gamma_{ip}\frac{\hat{T}_i'}{u_i}\right)}_{1.3}\underbrace{-\sum_{i=1}^{N}\left(\gamma_{i+Np}\frac{\hat{S}_i'}{u_{i+N}}\right)}_{1.4}\underbrace{+\frac{a_p}{\lambda_p}}_{1.5}\underbrace{+F_{fg}\frac{a_p}{\lambda_p}}_{1.6}=\underbrace{F_{fg}\frac{a_{fg,p}}{\lambda_p}}_{1.r.1} \quad \text{(A1)}$$

$$\overbrace{\sum_{k=1}^{m}b_k\left(\sum_{i=1}^{N}\left(\phi_{ik}\phi_{ip}\right)\right)}^{2.1}+\overbrace{\sum_{k=1}^{m}\left(b_k\sum_{i=1}^{N-1}\left(\phi_{i+N-1k}\phi_{i+N-1p}\right)\right)}^{2.2}+\overbrace{\frac{b_p}{\mu_p}}^{2.3}+\overbrace{G_{fg}\frac{b_p}{\mu_p}}^{2.4}$$
$$\underbrace{-\sum_{i=1}^{N-1}\frac{\phi_{ip}}{w_i}\hat{T}_{i+1}'}_{2.5}+\underbrace{\sum_{i=1}^{N-1}\frac{\phi_{ip}}{w_i}\hat{T}_i'}_{2.6}\underbrace{-\sum_{i=1}^{N-1}\left(\frac{\phi_{i+N-1p}\hat{S}_{i+1}'}{w_{i+N-1}}\right)}_{2.7}+\underbrace{\sum_{i=1}^{N-1}\left(\frac{\phi_{i+N-1p}\hat{S}_i'}{w_{i+N-1}}\right)}_{2.8}=\underbrace{G_{fg}\frac{b_{fg,p}}{\mu_p}}_{2.r.1} \quad \text{(A2)}$$

$$\overbrace{-\frac{1}{u_q}\sum_{k=1}^{m}\left(a_k\gamma_{qk}\right)}^{3.1}+\overbrace{\sum_{k=1}^{m}b_k\left(\frac{\phi_{qk}}{w_q}-\frac{\phi_{q-1k}}{w_{q-1}}\right)}^{3.2}+\overbrace{\frac{\hat{T}_q'}{u_q^{\,2}}}^{3.3}-\overbrace{\frac{\hat{T}_{q+1}'}{w_q^{\,2}}}^{3.4}+\overbrace{\hat{T}_q'\left(\frac{1}{w_q^{\,2}}+\frac{1}{w_{q-1}^{\,2}}\right)}^{3.5}-\overbrace{\frac{\hat{T}_{q-1}'}{w_{q-1}^{\,2}}}^{3.6}$$
$$\underbrace{-\frac{Q_q}{\varepsilon_{SST}^2}\sum_{i=1}^{N}Q_i\hat{T}_i'}_{3.7}\underbrace{-\frac{\alpha_q dz_q}{\varepsilon_h^2}\sum_{i=1}^{N}\alpha_i\hat{T}_i'dz_i}_{3.8}\underbrace{-\frac{\alpha_q dz_q}{\varepsilon_h^2}\sum_{i=1}^{N}\beta_i\hat{S}_i'dz_i}_{3.9}=\underbrace{\frac{Q_q\tilde{T}_{SST}'}{\varepsilon_{SST}^2}}_{3.r.1}\underbrace{-\frac{1}{\varepsilon_h^2}\left(\tilde{h}_{MLD}-\sum_{i=1}^{N}\delta_{b,i}dz_i\right)\alpha_q dz_q}_{3.r.2} \quad \text{(A3)}$$

$$\overbrace{-\sum_{k=1}^{m}\left(a_k\frac{\gamma_{q+Nk}}{u_{q+N}}\right)}^{4.1}+\overbrace{\sum_{k=1}^{m}b_k\left(\frac{\phi_{q+N-1k}}{w_{q+N-1}}-\frac{\phi_{q-1+N-1k}}{w_{q-1+N-1}}\right)}^{4.2}+\overbrace{\frac{\hat{S}_q'}{u_{q+N}^{\,2}}}^{4.3}-\overbrace{\frac{\hat{S}_{q+1}'}{w_{q+N-1}^{\,2}}}^{4.4}+\overbrace{\hat{S}_q'\left(\frac{1}{w_{q+N-1}^{\,2}}+\frac{1}{w_{q-1+N-1}^{\,2}}\right)}^{4.5}$$
$$\underbrace{-\frac{\hat{S}_{q-1}'}{w_{q-1+N-1}^{\,2}}}_{4.6}\underbrace{-\frac{\beta_q dz_q}{\varepsilon_h^2}\sum_{i=1}^{N}\alpha_i\hat{T}_i'dz_i}_{4.7}\underbrace{-\frac{\beta_q dz_q}{\varepsilon_h^2}\sum_{i=1}^{N}\beta_i\hat{S}_i'dz_i}_{4.8}=\underbrace{-\frac{1}{\varepsilon_h^2}\left(\tilde{h}_{MLD}-\sum_{i=1}^{N}\delta_{b,i}dz_i\right)\beta_q dz_q}_{4.r.1} \quad \text{(A4)}$$

These terms are matched up with the computer code in isop_syn_compute.f for solving these equations. The comments in isop_syn_compute.f match the labels of the terms in A1, A2, A3, and A4. In the computer code the eigenvectors are split up into variables for the T-vectors and a separate set for the S-vectors. In the equation above, the T-eigenvectors are listed 1 to N and the S-eigenvectors are listed N+1 to 2N.

References

Barron, C. N., A. B. Kara, P. J. Martin, R. C. Rhodes, and L. F. Smedstad. "Formulation, implementation and examination of vertical coordinate choices in the global Navy Coastal Ocean Model (NCOM)." *Ocean Modeling*, 2006: 347-375.

Bauer, R. "Functional Description: Master Oceanographic Data Set (MOODS), Compass Systems Inc." 1982.

Boyer, T. P., et al. *World Ocean Database 2005.* S. Levitus, Ed., NOAA Atlas NESDIS 60, Washington, D.C.: U.S. Government Printing Office, 2006, 190 pp.

Carnes, M. R. *Descsription and Evaluation of GDEM-V 3.0.* Memorandum Report, 7330--09-9165, NRL, 2009.

Carnes, M. R., J. L. Mitchell, and P. W. DeWitt. "Synthetic Temperature Profiles Derived From Geosat Altimetry: Comparison With Air-Dropped Expendable Bathythermograph Profiles." *J. Geophys. Res.*, 1990: 17,979-17,992.

Carnes, M. R., R. W. Helber, C. N. Barron, and J. M. Dastugue. *Validation Test Report for GDEM4.* Memorandum Report, 7330--10-9271, MRL, 2010.

Carnes, M. R., W. J. Teague, and J. L. Mitchell. "Inference of Subsurface Thermohaline Strucure from Fields Mearuable by Satellite." *J. Atmos. Ocean. Tech.*, 1994: 551-566.

Cummings, J. A. "Operational multivariate ocean data assimilation." *Q. J. R. Meteorol. Soc.*, 2005: 3583–3604.

Ezer, T., and G. L. Mellor. "Continuous Assimilation of Geosat Altimeter Data into a Three-Dimensional Primitive Equation Gulf Stream Model." *J. Phys. Oceangr.*, 1994: 832-847.

Ezer, T., and G. L. Mellor. "Data Assimilation Experiments in the Gulf Stream Region: How Useful Are Satellite-Derived Surface Data for Nowcasting the Subsurface Fields." *J. Atmos. Oceanic Tech.*, 1997: 1379-1391.

Fischer, M. "Multivariate projection of ocean surface data onto subsurface sections." *Geophys. Res. Lett.*, 2000: 755-757.

Fox, D. N., W. J. Teague, C. N. Barron, M. R. Carnes, and C. M. Lee. "The Modular Ocean Data Assimilation System (MODAS)." *J. Atmos. Oceanic. Tech.*, 2002: 240-252.

Fritsch, F. N., and R. E. Carlson. "Monotone piecewise cubic interpolation." *SIAM J. Numer. Anal.*, 1980: 238-246.

Fujii, Y., and M. Kamachi. "A Reconstruction of Observed Profiles in the Sea East of Japan Using Vertical Coupled Temperature-Salinity EOF Modes." *J. Oceanogr.*, 2003a: 173-186.

Fujii, Y., and M. Kamachi. "Three-dimensional analysis of temperature and salinity in the equatorial Pacific using a variational method with vertical coupled temperature-salinity empirical orthogonal function modes." *J. Geophys. Res.*, 2003b: 10.1029/2002JC001745.

Gilson, J., D. Roemmich, B. Cornuelle, and L.-L. Fu. "Relationship of TOPEX/Poseidon altimetric height to steric height and circulation in the North Pacific." *J. Geophys. Res.*, 1998: 27,947-27,965.

Guinehut, S., P. Y. Le Traon, G. Larnicol, and S. Philipps. "Combining Argo and remote-sensing data to estimate the ocean three-dimensional temperature fields—a first approach based on simulated observations." *J. Mar. Syst.*, 2004: 85-98.

Helber, R. W, M. A. Carnes, and C. N. Barron. Automated System and Methods for Vertical Gradeint Correciton. USA. October 4, 2011.

Helber, R. W. *Profile Acoustic Parameters, NRL Software Repository.* 2012. https://www7320.nrlssc.navy.mil/svn/repos/ProfParam.

Jugan, M. J., and H. Beresford. "Editing Approach for the Navy's Master Oceanographic Observation Data Set." *Proceedings of MTS 1991, An Ocean Cooperative: Industry, Government, and Academia, Vol. II.* 1991.

Kara, A. B., P. A. Rochford, and H. E. Hurlburt. "An optimal definition for ocean mixed layer depth." *J. Geophys. Res.*, 2000: 16,803-16,821.

Ko, D. D. *DBDB2 v3.0 Global 2-Minute Topography, Naval Research Laboratory.* 2009. http://www7320.nrlssc.navy.mil/DBDB3_WWW.

Masina, S., N. Pinardi, and A. Navarra. "A global ocean temperature and altimeter data assimilation system for studies of climate variablity." *Clim. Dyn.*, 2001: 687-700.

Mellor, G. L., and T. Ezer. "A Gulf Stream Model and an Altimetry Assimilation Scheme." *J. Geophys. Res.*, 1991: 8779-8795.

Metzger, E. J., et al. *Validation Test Report for the Global Ocean Forecast System V3.0 - 1/12° HYCOM/NCODA: Phase II.* Memorandum Report, 7320--10-9236, NRL, 2010.

Murphy, A. H. "Skill Scores Based on the Mean Square Error and Their Relationships to the Correlation Coefficient." *Mon. Weather Rev.*, 1988: 2417-2424.

NODC, NOAA. *World Ocean Database Select and Search.* 2009. http://www.nodc.noaa.gov/OC5/SELECT/dbsearch/dbsearch.html.

Pascual, A., and D. Gomis. "Use of Surface Data to Estimate Geostrophic Transport." *J. Atmos. Ocean Tech.*, 2003: 912-926.

Roemmich, D., and J. Gilson. "The 2004–2008 mean and annual cycle of temperature, salinity, and steric height in the global ocean from the Argo Program." *Prog. Oceanogr.*, 2009: 81-100.

Shapiro, R. "Smoothing, Filtering, and Boundary Effects." *Rev. Geophys. Spce Phys.*, 1970: 359-387.

Smith, S., et al. *Validation Test Report for the Navy Coupled Ocean Data Assimilation 3D Variational Analysis (NCODA-VAR) System, Version 3.43.* Memorandum Report, 7320--12-9363, NRL, 2012.

Teague, W. J., M. J. Carron, and P. J. Hogan. "A COMPARISON BETWEEN THE GENERALIZED DIGITAL ENVIRONMENTAL-MODEL AND LEVITUS CLIMATOLOGIES." *J. Geophys. Res.*, 1990: 7167-7183.

USGODAE. *USGODAE Argo Page.* 2012. http://www.usgodae.org/argo/argo.html.

Tables

Table 1. Instrument Description

The description of the measurement instrument for each instrument type number.

Instrument Type Number	Description of Instrument Type
1	Message data (regardless of instrument)
4	SEABIRD CTD Time Series
5	Air-deployed CTD (AXCTD)
19	Submarine deployed CTD
25	Hydrocast: bottles and reversing thermometers
30	Unknown electronic salinity, temp., depth instruments
31	Salinity, temperature and depth probe (STD)
32	Low-resolution STD from NODC
	Conductivity, temperature, depth probe (CTD)
	CTD time series (YO-YO)
33	Sound velocity, salinity, temp, and depth (SVSTD)
35	Sippican XCTD
38	Seabird CTD SBE-19 (Seacat profiler)
41	Seabird CTD SBE-911 (deep ocean)
42	Seabird CTD SBE-25
43	Falmouth CTD
51	Argo/Palace floats
62	tesac RTDHS message data
64	Palace/Apex floats RTDHS message data
65	Glider RTDHS message data
71	Undulating ocean recorder (UOR)
73	Unknown instrument

Table 2. Number of Observations

The number of profiles of each instrument type before and after each step in removal. The "Instrument Type Number" refers to the description of each type of instrument in Table 1. "Original Number" is the number of temperature and salinity profiles of raw unedited data. The third column, "Number Removed", is the number of profiles removed either because they were removed during manual quality control, because the number of observation depths is less than 2, and the fourth column lists the number of profiles remaining after the removal. The fifth column list the number of profiles removed because the bottom bathymetry depth was greater than or equal to 400 m, but the greatest depth of the profiles was less than 200 m. The sixth column lists the final number of temperature/salinity profiles remaining in the final profile data used in the construction of ISOP. The totals from columns two through six are listed in the bottom row.

Instrument Type Num.	Original Number	Number Removed	Number Remaining	Number Removed	Final Number
1	682052	-52513	629539	-85212	38623
4	3130	-4	3126	0	3126
5	83	-9	74	0	74
19	38	-1	37	0	37
25	1708992	-437824	1271168	-41393	1173847
30	34128	-3056	31072	-2897	28076
31	13274	-2122	11152	-183	10254
32	656	-181	475	-4	469
33	494042	-70821	423221	-12725	403331
35	1289	-80	1209	-93	1116
36	7737	-1348	6389	-207	6182
37	1962	-1311	651	-3	592
38	5392	-377	5015	-50	4965
41	7388	-139	7249	-41	7208
42	128	-1	127	-1	126
43	1138	-108	1030	-17	1013
51	290107	-55095	235012	-2012	215169
62	9685	-1397	8288	-90	6477
64	81940	-30468	51472	-478	41695
65	8443	-79	8364	-623	7741
71	21225	-1083	20142	-5448	14573
73	133667	-10399	123268	-20628	4386
Total	3506496	-668416	2838080	-172105	1969081

Table 3. Standard Depths
The 78 standard depths listed by depth ranges with the same level spacing.

every 2 m from 0 m to 10 m	0, 2, 4, 6, 8, 10
every 5 m from 10 m to 100 m	15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100
every 10 m from 100 m to 200 m	110, 120, 130, 140, 150, 160, 170, 180, 190, 200
every 20 m from 200 m to 300 m	220, 240, 260, 280, 300
every 50 m from 300 m to 400 m	350, 400
every 100 m from 400 m to 1600 m	500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600
every 200 m from 1600 m to 6600 m	1800, 2000, 2200, 2400, 2600, 2800, 3000, 3200, 3400, 3600, 3800, 4000, 4200, 4400, 4600, 4800, 5000, 5200, 5400, 5600, 5800, 6000, 6200, 6400, 6600

Table 4 ISOP Inputs

Name		Label	Description	Source
Sea surface temperature		SST	Observations or estimates of the sea surface temperature	Satellite observations, OI SST products, model forecasts
Sea surface temperature uncertainty		SST_err	The uncertainty estimate for the input SST	Depending on source, observation error estimate, OI or model uncertainty
Sea surface height anomaly		SSHA	Sea surface height difference from the annual mean.	Satellite Sea surface height difference from the record length mean, steric height difference from the annual mean steric heght*
Sea surface height anomaly uncertainty		SSHA_err	The uncertainty estimate of SSHA	Depending on source, derived from observation error estimate
Mixed layer depth		MLD	Estimated depth of the mixed layer	Model forecast, climatology estiamte
First-guess estimate, subsurface profiles	Temperature	T_fg	Subsurface temperature profile	Model forecast, climatology
	Temperature uncertainty	T_fg_err	Uncertainty of Subsurface temperature profile	Model uncertainty, climatological standard deviation
	Salinity	S_fg	Subsurface salinity profile	Model forecast, climatology
	Salinity uncertainty	S_fg_err	Uncertainty of Subsurface salinity profile	Model uncertainty, climatological standard deviation

* Sea surface height estimates for ISOP have representation error associated with them because it is not possible to remotely measure steric height referenced to 1000 m. Correction factors are applied in attempt to reduce representation errors.

Table 5. ISOP Output

Name	Label	Description
Temperature profile	T	Synthetic temperature profile estimate
Temperature profile uncertainty	T_err	Synthetic temperature profile uncertainty estimate
Salinity profile	S	Synthetic salinity profile estimate
Salinity profile uncertainty	S_err	Synthetic salinity profile uncertainty estimate

Table 6. List of Cost Function Symbols

variable/symbol	Description
x	Analysis vector of the temperature and salinity solution being sought
x_{cl}	Climatology temperature and salinity vector
x_{fg}	First-guess (model forecast) temperature and salinity vector
d	Analysis vector of the difference of adjacent depth levels
d_{cl}	Climatology temperature and salinity difference of adjacent depth levels vector
d_{fg}	First guess temperature and salinity difference of adjacent depth levels vector
B	Climatological error covariance matrix
B_g	Climatological vertical difference error covariance matrix
R	First-guess error covariance matrix
R_g	Vertical difference first-guess error covariance matrix
$()^T$	Matrix transpose
$()^{-1}$	Matrix inverse
N	Number of depth levels in analysis profile
m	Number of EOF modes retained (prototype $m=6$)
T_i	m-mode eigenfunction derived temperature at the depth level i
$\hat{T}_i$	Analysis temperature solution being sought at depth level i
$T_{cl,i}$	Climatological temperature at depth level i
$T_{fg,i}$	First-guess temperature at depth level i
T_i'	$T_i' = T_i - T_{cl,i}$
$\hat{T}_i'$	$\hat{T}_i' = \hat{T}_i - T_{cl,i}$
$\Delta T_i'$	$\Delta T_i' = T_{i+1}' - T_i'$
S_i	m-mode eigenfunction derived salinity at the depth level i
$\hat{S}_i$	Analysis salinity solution being sought at depth level i
$S_{cl,i}$	Climatological salinity at depth level i
$S_{fg,i}$	First-guess salinity at depth level i
S_i'	$S_i' = S_i - S_{cl,i}$
$\hat{S}_i'$	$\hat{S}_i' = \hat{S}_i - S_{cl,i}$
$\Delta S_i'$	$\Delta S_i' = S_{i+1}' - S_i'$
u_i	Standard deviation of T_i and S_i at the depth level I, elements of U
w_i	Vertical difference standard deviation of T_i and S_i at the depth level I, elements of W

$\tilde{T}'_{MLD}$	Observed temperature anomaly at the MLD
$\hat{T}'_{MLD}$	Analysis temperature anomaly at the MLD
ε_{SST}	The SST error variance.
$\tilde{h}_{MLD}$	Observed estimate of steric height anomaly at the MLD.
$\hat{h}_{MLD}$	Analysis steric height anomaly at the MLD referenced to layer 2 bottom depth.
ε_{MLD}	The MLD error variance.
U	Vector of T and S standard deviations [1 by N]
C	Vertical correlation matrix, [N by N]
Γ	Eigenvectors for T and S, [N by m]
Λ	Eigenvalues for T and S, [m by m]
γ_{ik}	Eigenvector elements of Γ
λ_i	Eigenvalue elements of Λ
a_i	ith T and S EOF amplitude, solution being sought
$a_{fg,i}$	ith first-guess T and S EOF amplitude, computed from first guess input
W	Vector of T and S vertical difference standard deviations [1 by N]
Φ	Eigenvectors for T and S vertical difference, [N by m]
M	Eigenvalues for T and S vertical difference, [m by m]
ϕ_{ik}	Eigenvector elements of Φ
μ_i	Eigenvalue elements of M
b_i	ith vertical difference T and S EOF amplitude, solution being sought
$b_{fg,i}$	ith first-guess vertical difference T and S EOF amplitude, computed from first guess input
δ	normalized specific volume anomaly
α	Thermal expansion coefficient for density
β	Haline expansion coefficient for density
$\delta_{cl,i}$	ith climatology normalized specific volume anomaly
F_{fg}	Error covariance proportional constant, $F_{fg} = B / R$
G_{fg}	Error covariance proportional constant, $G_{fg} = B_g / R_g$

Table 7. Validation Cases
The ISOP validation cases with their labels and input configurations.

Configuration Label	First-Guess	SSHA	SST	MLD
$ISOP^{Ideal}$	GOFS	Obs	Obs	Obs
$ISOP^{Oper}$	GOFS	MODAS 2D	MODAS 2D	GOFS
$MODAS^{Ideal}$	NA	Obs	Obs	NA
$MODAS^{Oper}$	NA	MODAS 2D	MODAS 2D	NA
GDEM	NA	NA	NA	NA

Table 8. Global Median RMSE and MB

Global median RMSE for parameters SLD, MLD, and BLG. The smallest error levels are shaded dark green and the next smallest in light green			
Synthetic	RMSE		
	SLD	MLD	BLG
$ISOP^{Ideal}$	27.75	25.48	0.97
$ISOP^{Oper}$	32.39	27.40	1.03
$MODAS^{Ideal}$	40.31	37.05	1.12
$MODAS^{Oper}$	41.18	36.77	1.27
GDEM	35.95	29.62	1.24
Global median MB for parameters SLD, MLD, and BLG.			
Synthetic	MB		
	SLD	MLD	BLG
$ISOP^{Ideal}$	-10.00	-5.67	-0.37
$ISOP^{Oper}$	-15.00	-7.03	-0.45
$MODAS^{Ideal}$	-30.00	-20.61	-0.56
$MODAS^{Oper}$	-25.00	-20.03	-0.59
GDEM	-20.00	-11.29	-0.45

Table 9. Global Errors with Depth

a) Global median for MB of temperature in four depth ranges.

Product	T MB			
	0-50 m	50-100 m	100-150 m	150-200 m
$ISOP^{Ideal}$	-0.04	-0.36	-0.22	-0.28
$ISOP^{Oper}$	-0.12	-0.48	-0.33	-0.41
$MODAS^{Ideal}$	-0.21	-0.77	-0.56	-0.65
$MODAS^{Oper}$	-0.22	-0.74	-0.53	-0.62
GDEM	-0.19	-0.71	-0.53	-0.64

b) Global median for MB of salinity in four depth ranges.

Product	S MB			
	0-50 m	50-100 m	100-150 m	150-200 m
$ISOP^{Ideal}$	-0.04	-0.09	-0.07	-0.09
$ISOP^{Oper}$	-0.04	-0.09	-0.08	-0.09
$MODAS^{Ideal}$	-0.06	-0.13	-0.11	-0.12
$MODAS^{Oper}$	-0.06	-0.13	-0.11	-0.13
GDEM	-0.03	-0.10	-0.08	-0.10

c) Global median for MB of sound speed in four depth ranges.

Product	SSPD MB			
	0-50 m	50-100 m	100-150 m	150-200 m
$ISOP^{Ideal}$	-0.17	-1.15	-0.63	-0.89
$ISOP^{Oper}$	-0.35	-1.53	-0.97	-1.25
$MODAS^{Ideal}$	-0.64	-2.39	-1.73	-2.07
$MODAS^{Oper}$	-0.67	-2.32	-1.56	-1.91
GDEM	-0.58	-2.07	-1.57	-1.90

d) Global median RMSE and SS for temperature, salinity, and sound speed over the upper 1000 m.

	RMSE			SS		
	T	S	SSPD	T	S	SSPD
$ISOP^{Ideal}$	0.39	0.10	1.40	0.69	0.28	0.66
$ISOP^{Oper}$	0.49	0.10	1.74	0.53	0.26	0.49
$MODAS^{Ideal}$	0.56	0.13	2.04	0.39	-0.26	0.34
$MODAS^{Oper}$	0.66	0.14	2.36	0.17	-0.35	0.08
GDEM	0.77	0.12	2.58	NA	NA	NA

Table 10. GOM SLD % Misfit Statistics

Case	Median	Mean	RMS
ISOP^{Ideal}	0	-4.3	19.7
ISOP^{Oper}	-7.7	-11.4	21.2
MODAS^{Ideal}	0	-8.1	25.3
MODAS^{Oper}	-1.2	-9.6	26.2
GDEM	-12.5	-14.2	22.8

Table 11. GOM MLD% Misfit Statistics

Case	Median	Mean	RMS
ISOP^{Ideal}	-1.6	-4.5	17.0
ISOP^{Oper}	0	-3.9	24.4
MODAS^{Ideal}	-27.0	-28.5	27.3
MODAS^{Oper}	-27.8	-29.1	27.0
GDEM	-19.6	-22.6	28.3

Table 12. GOM BLG % misfit Statistics

Case	Median	Mean	RMS
ISOP^{Ideal}	-2.3	-2.4	9.1
ISOP^{Oper}	-2.4	-1.9	9.6
MODAS^{Ideal}	-1.6	-2.0	11.3
MODAS^{Oper}	-3.0	3.4	12.4
GDEM	-8.8	-9.7	12.0

Table 13. GOM SLD Relative % Misfit Statistics
The mean and median difference between percent misfit from observations.

Case	Median Difference	Mean Difference
$ISOP^{Ideal}\_MODAS^{Ideal}$	0	-8.1
$ISOP^{Oper}\_MODAS^{Oper}$	0	-4.6
$ISOP^{Ideal}\_GDEM$	0	-8.0
$ISOP^{Oper}\_GDEM$	0	-3.8

Table 14. GOM MLD Relative % Misfit Statistics
The mean and median difference between percent misfit from observations.

Case	Median Difference	Mean Difference
$ISOP^{Ideal}\_MODAS^{Ideal}$	-15.8	-19.3
$ISOP^{Oper}\_MODAS^{Oper}$	-12.2	-13.1
$ISOP^{Ideal}\_GDEM$	-12.1	-16.2
$ISOP^{Oper}\_GDEM$	-10.0	-9.7

Table 15. GOM BLG Relative % Misfit Statistics
The mean and median difference between percent misfit from observations.

Case	Median Difference	Mean Difference
$ISOP^{Ideal}\_MODAS^{Ideal}$	-1.4	-1.9
$ISOP^{Oper}\_MODAS^{Oper}$	-1.2	-2.4
$ISOP^{Ideal}\_GDEM$	-4.4	-6.1
$ISOP^{Oper}\_GDEM$	-4.2	-5.9

Table 16 RIMPAC Validation ISOP Cases

ISOP Inputs	Input Sources
First guess	GOFS Analysis
SST	MODAS 2D
SSHA	MODAS 2D
MLD	GOFS Analysis

Table 17 RIMPAC Validation MODAS Cases

MODAS Inputs	Input Sources
SST	MODAS 2D
SSHA	MODAS 2D

Figures

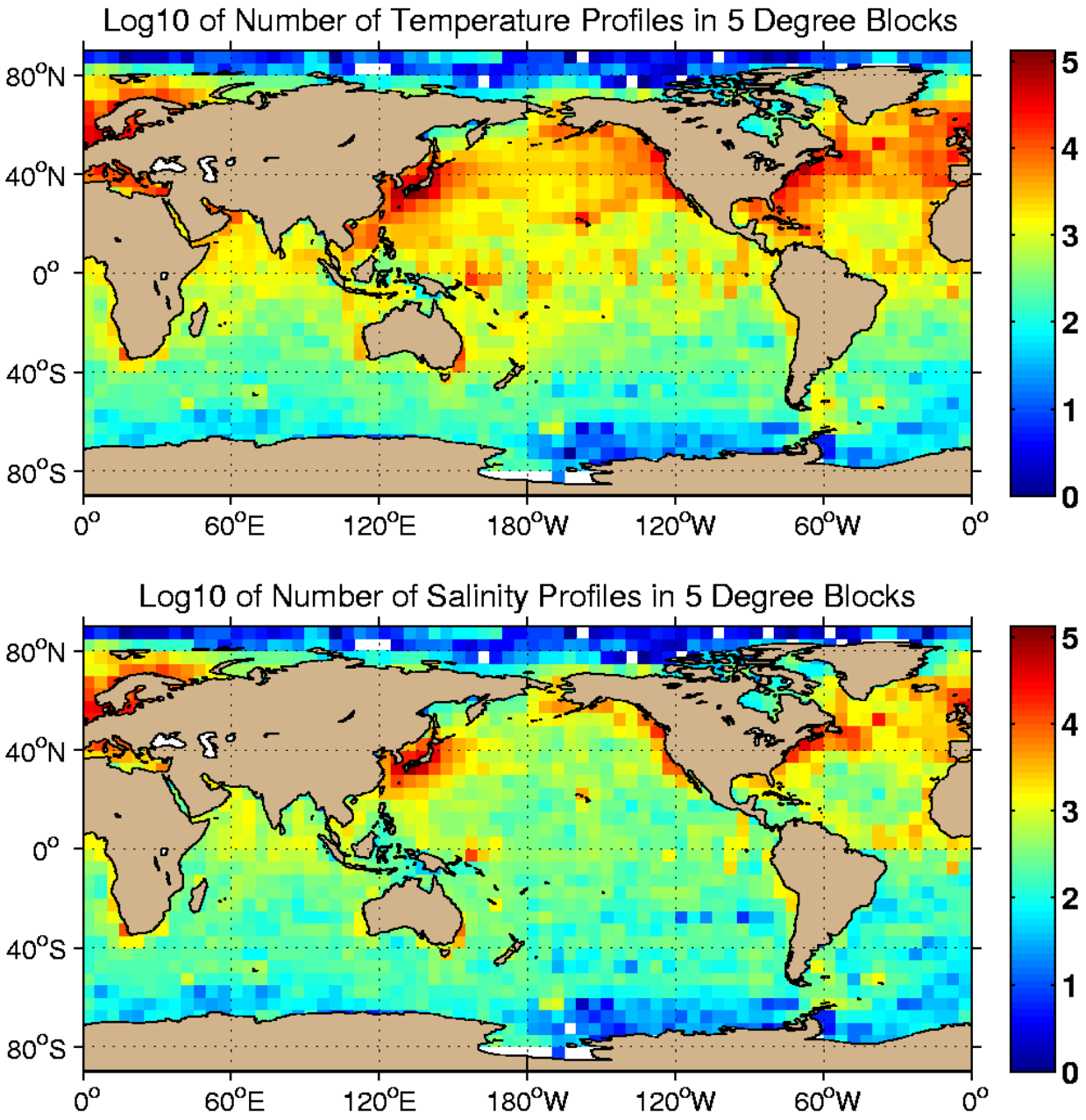


Figure 1. The $\log_{10}$ of the total number of (a) temperature and (b) salinity profiles binned in 5°by 5°degree geographical blocks.

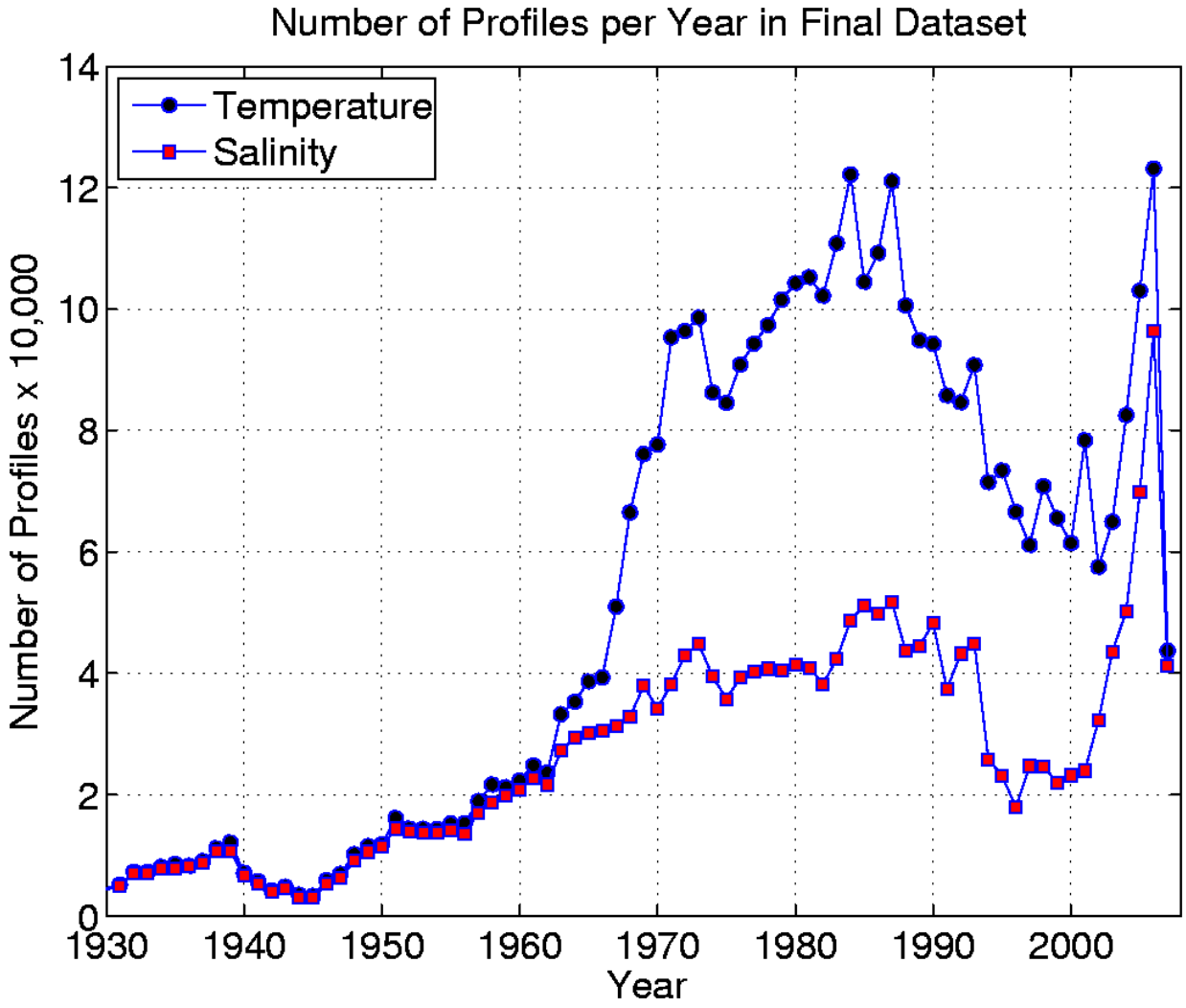


Figure 2. The total number of temperature and salinity profiles binned per year in the ISOP database.

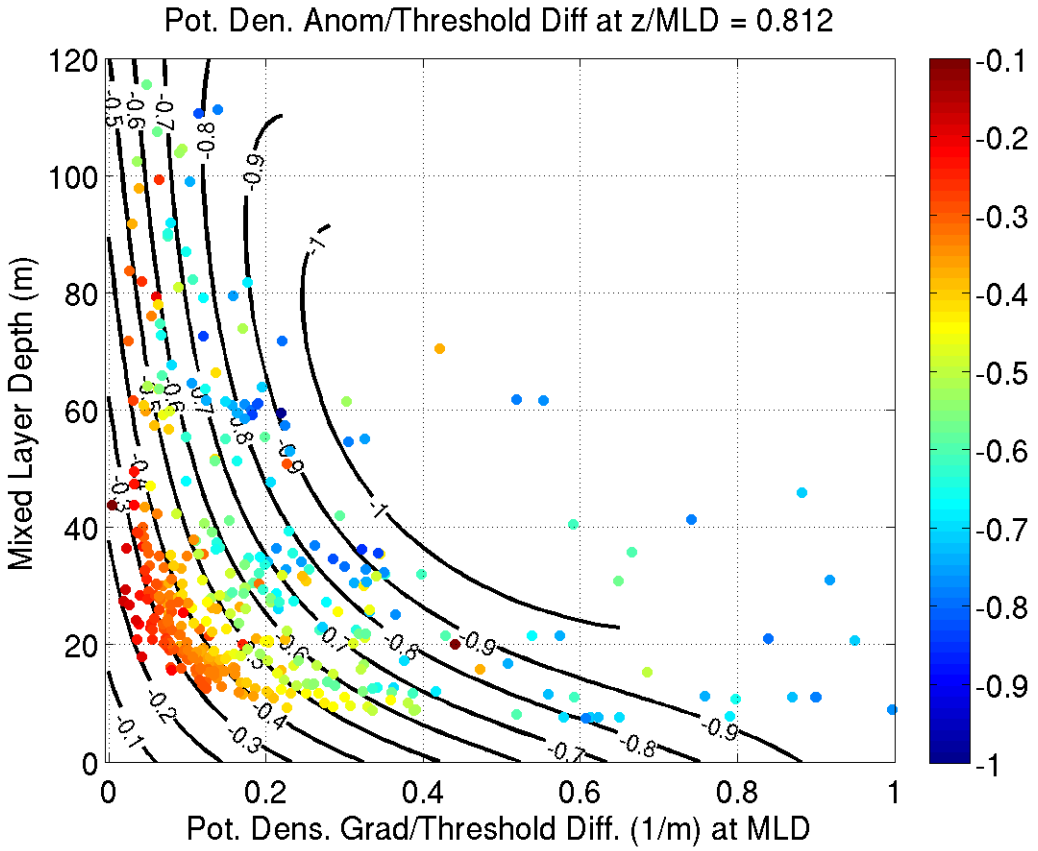


Figure 3. Values of scaled potential density anomaly at a scaled depth of z/MLD = 0.812 as a function of MLD and the scaled potential density vertical gradient at the MLD. The colored circles are value computed from profiles surround the analysis position at 30° N and 140° E. The contours are computed from Equation 3 fit to the observations.

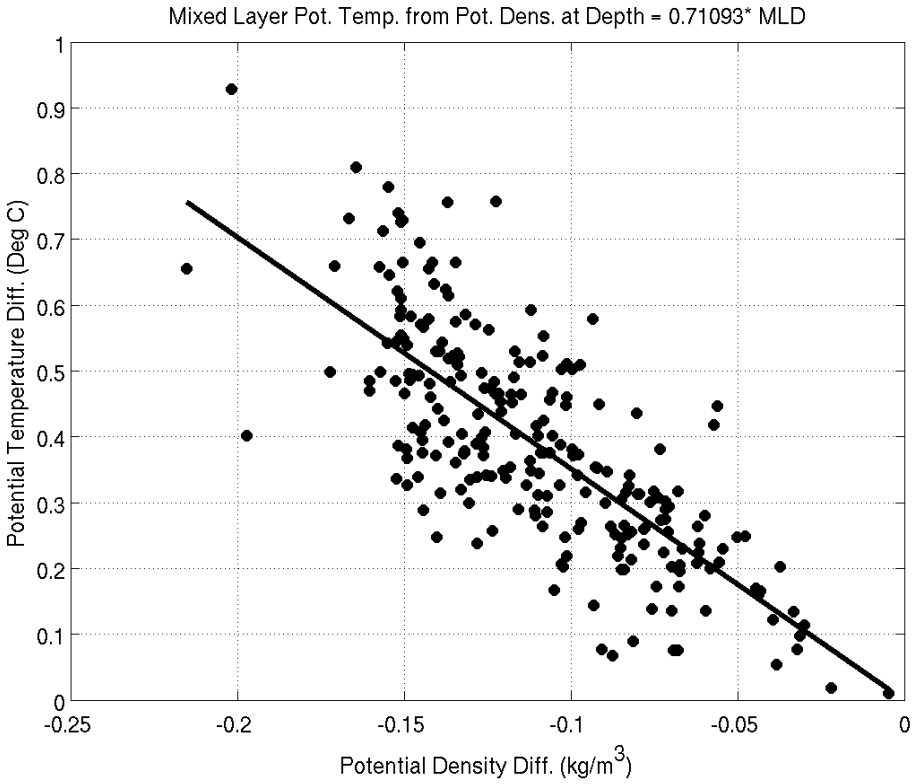


Figure 4a Relationship between potential temperature and potential density anomalies (difference from the values at the MLD) at a scaled depth of 0.71093 times the MLD determined from observations near at 30° N and 140° E.

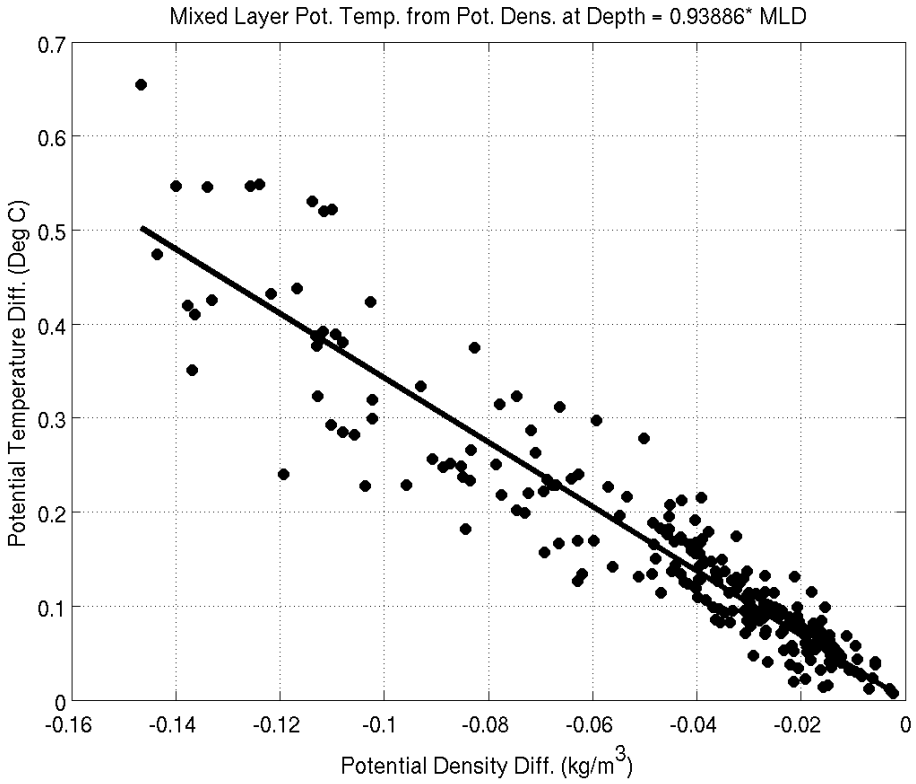


Figure 4b Same as for Figure 2a, but at a scaled depth of 0.93866 times the MLD.

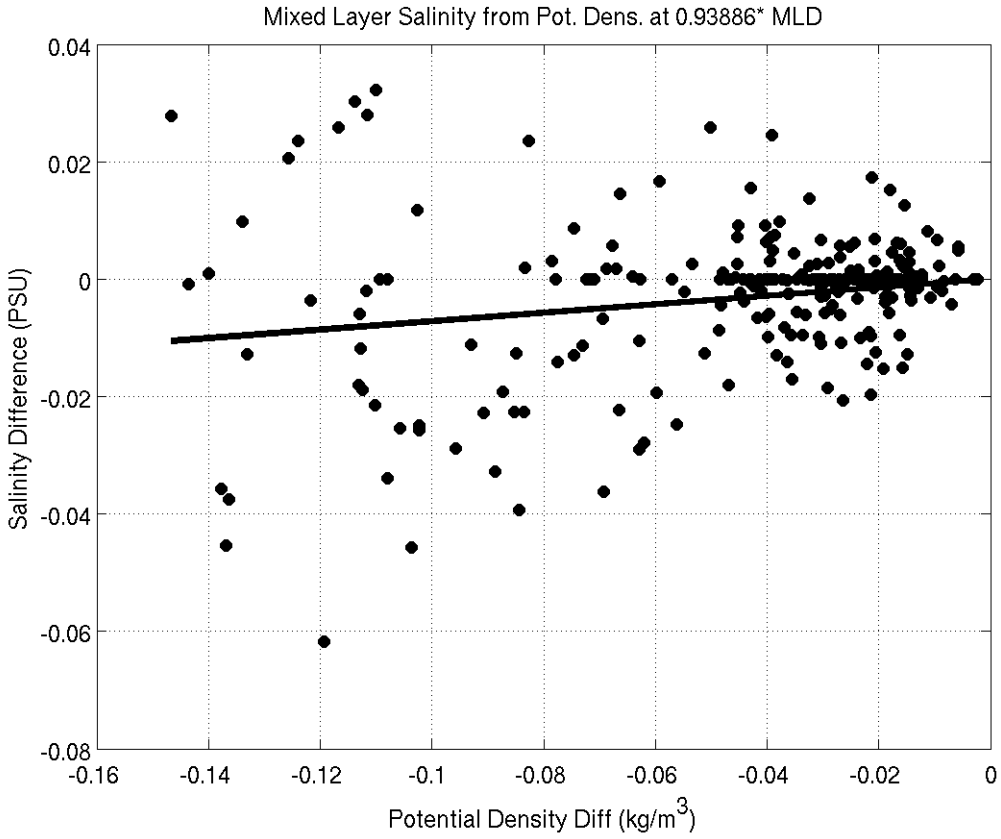


Figure 5 Relationship between salinity and potential density anomalies (difference from the values at the MLD) at a scaled depth of 0.93866 times the MLD determined from observations near at 30° N and 140° E

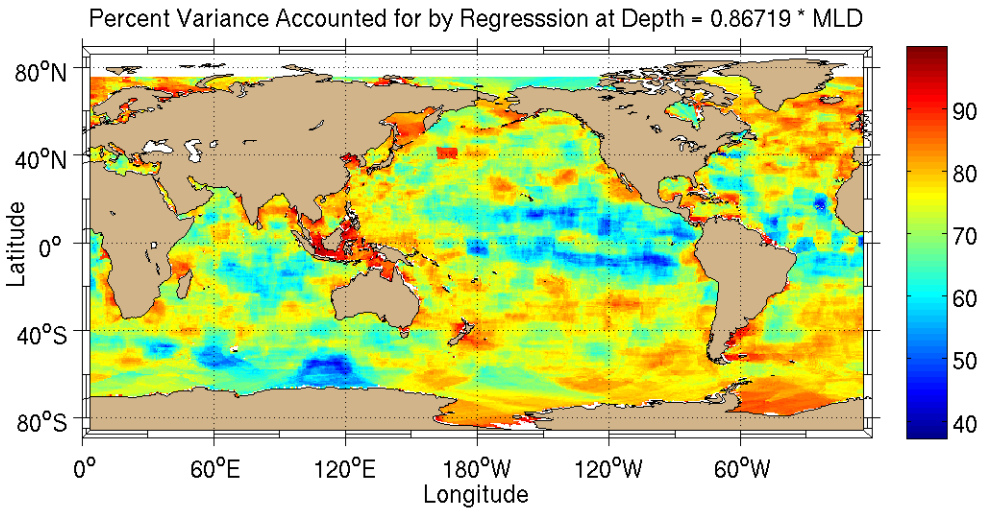


Figure 6 Percent of the potential density variance at scaled depth of 0.867 * MLD accounted for by prediction from regression equation (Equation 3) as function of the MLD, the threshold potential density difference, and the potential density vertical gradient at the MLD.

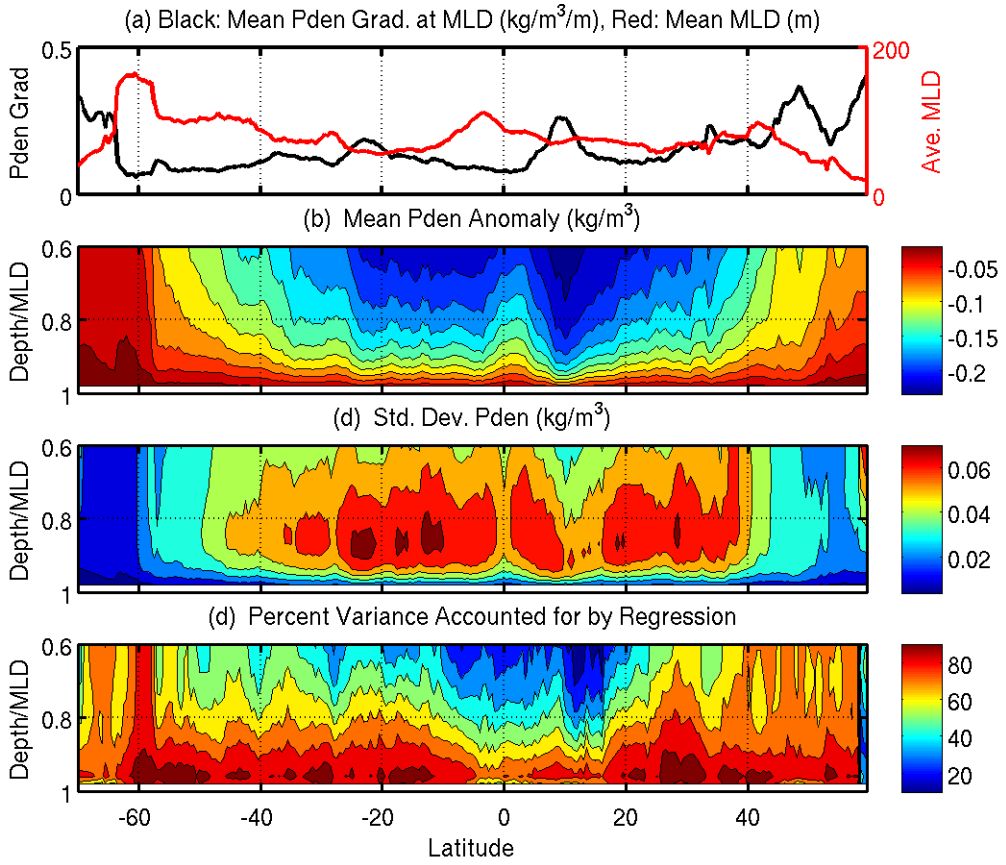


Figure 7. Top frame is the average MLD and the average potential density gradient at the MLD. The second frame is the average potential density versus scaled depth (depth/MLD) versus latitude. The third frame is the standard deviation of the potential density, and the bottom frame is the percent of variance (about the mean potential density) accounted for by the potential density mixed layer model.

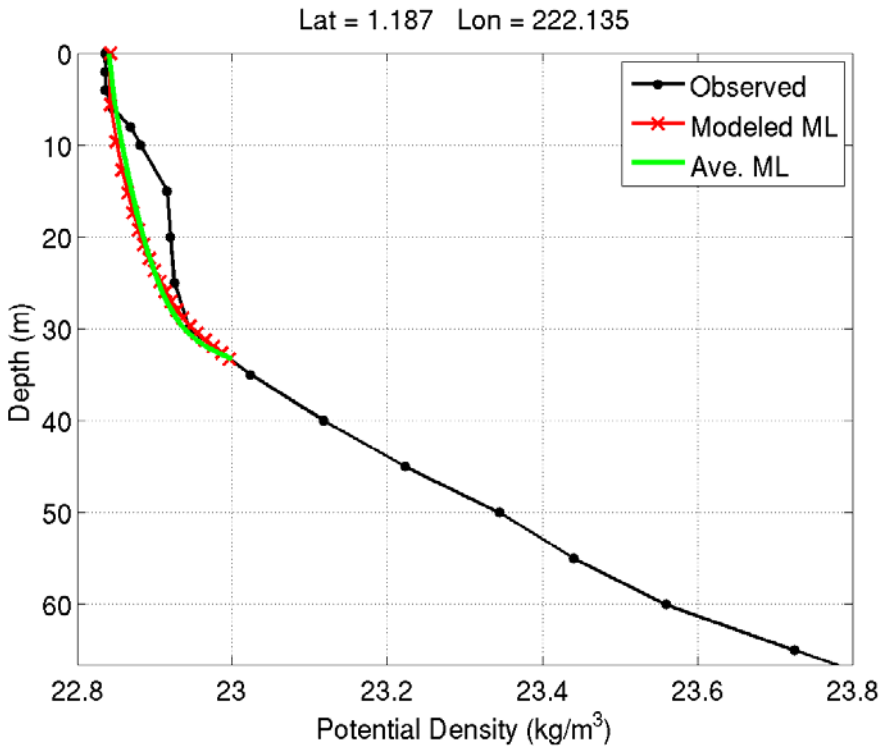


Figure 8 Comparison of observed profile (black dotted line) at 1.18° N and 222.14° E to the annual mean potential density profile in the mixed layer (green line), and to the modeled profile computed from the observed profile's MLD, potential density threshold value (0.15 kg/m**3), and potential density gradient at the MLD.

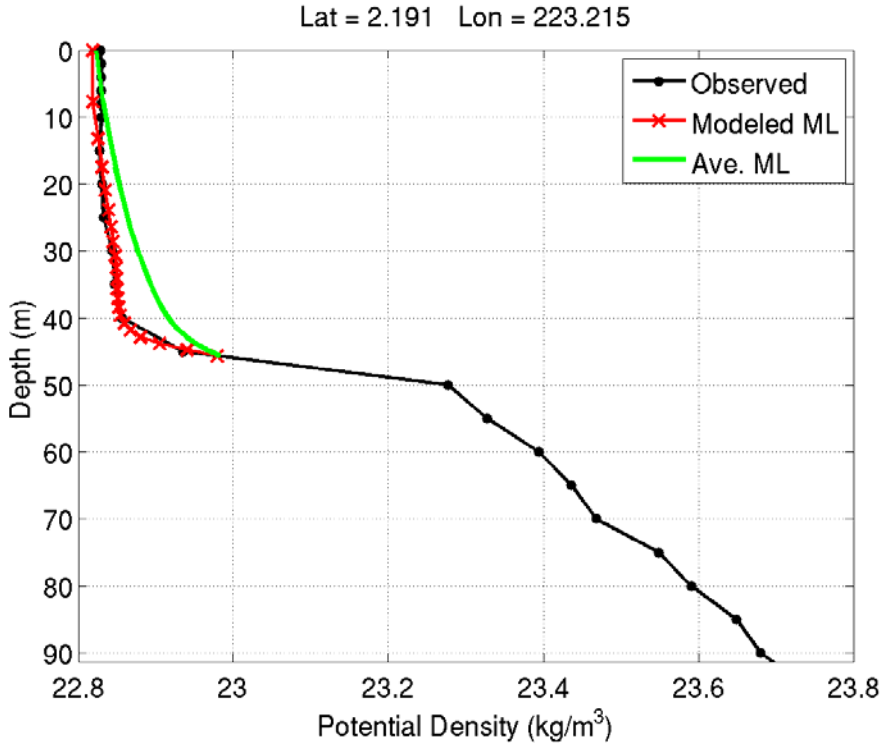


Figure 9. Same as Figure 6, but for independent profile at 2.19° N and 223.21° E.

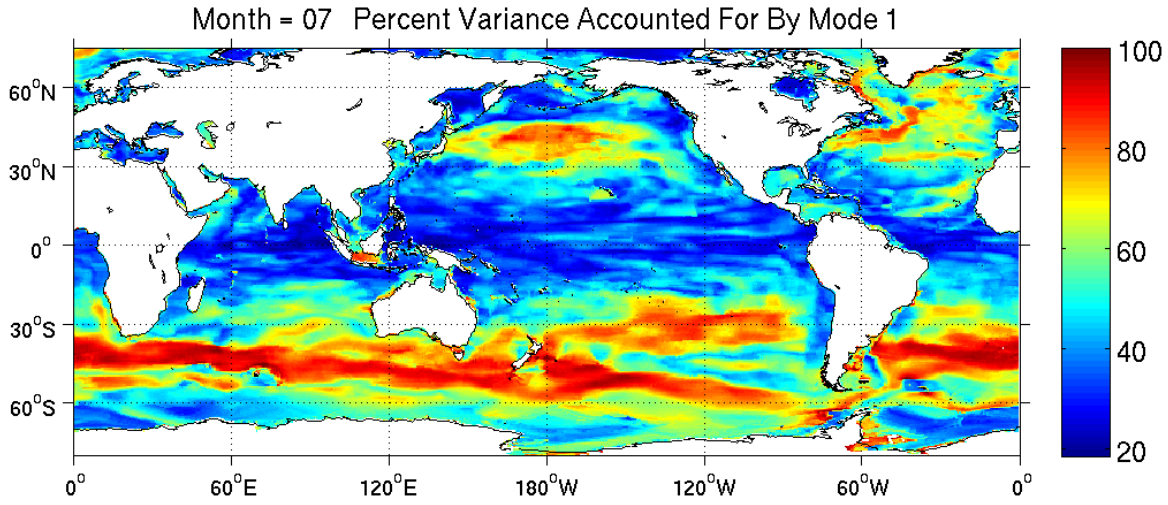


Figure 10. The percent of variance accounted for by the first mode eigenvectors.

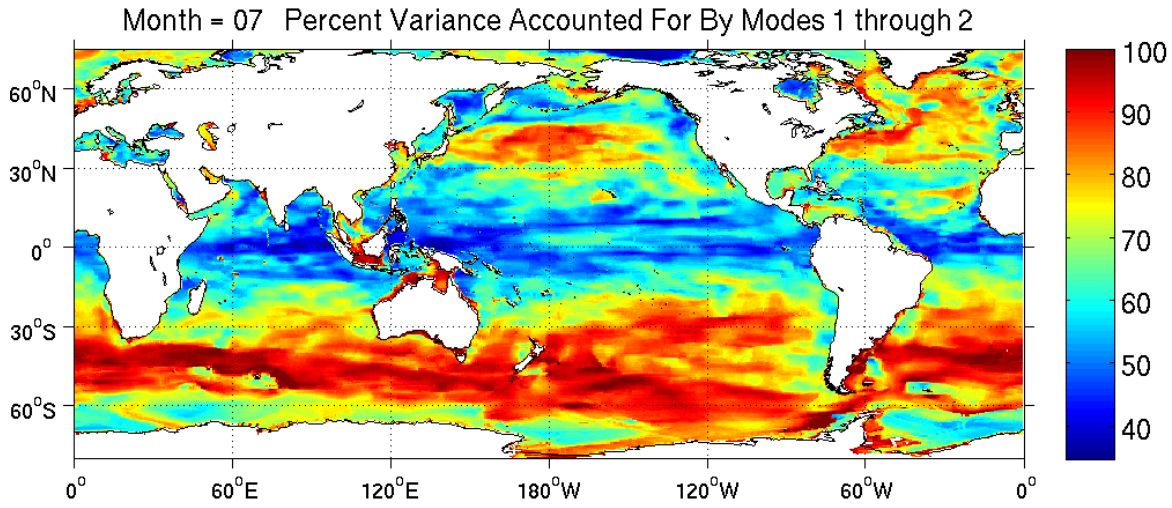


Figure 11. The percent of variance accounted for by the first and second mode eigenvectors.

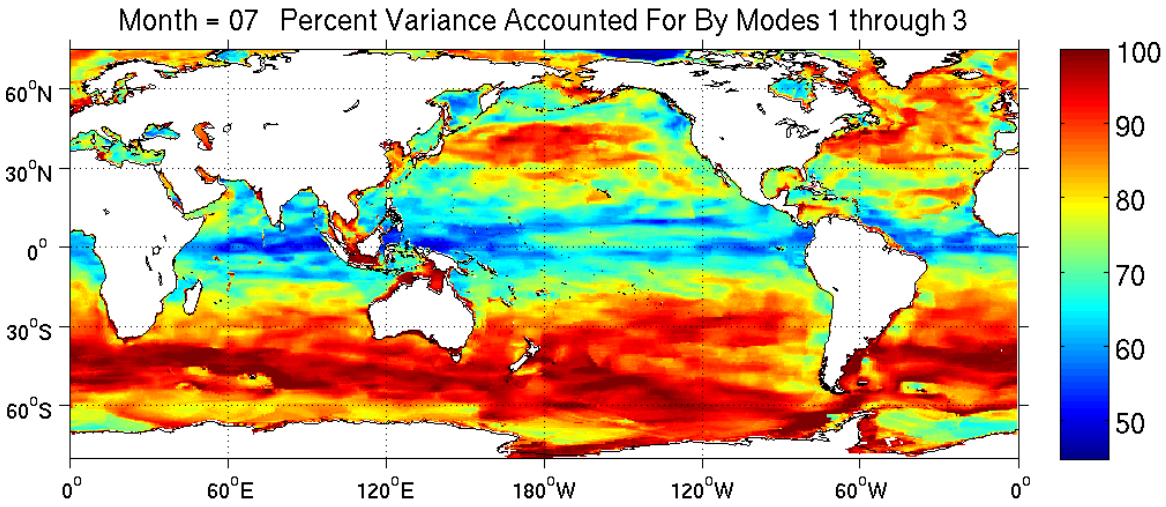


Figure 12. The percent of variance accounted for by the first, second, and third mode eigenvectors.

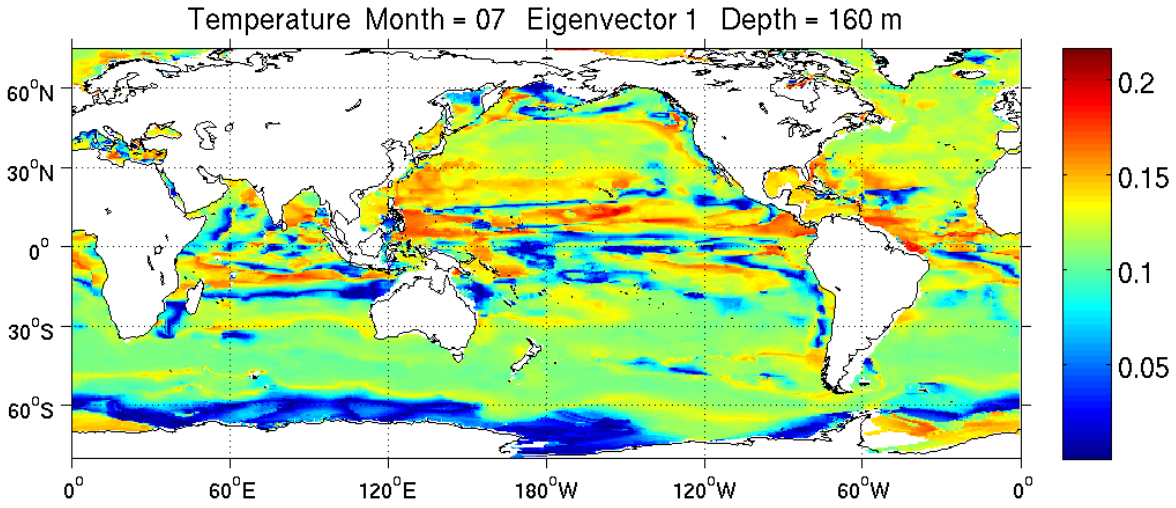


Figure 13 Absolute value of the mode 1 temperature eigenvector for June at 160 m depth.

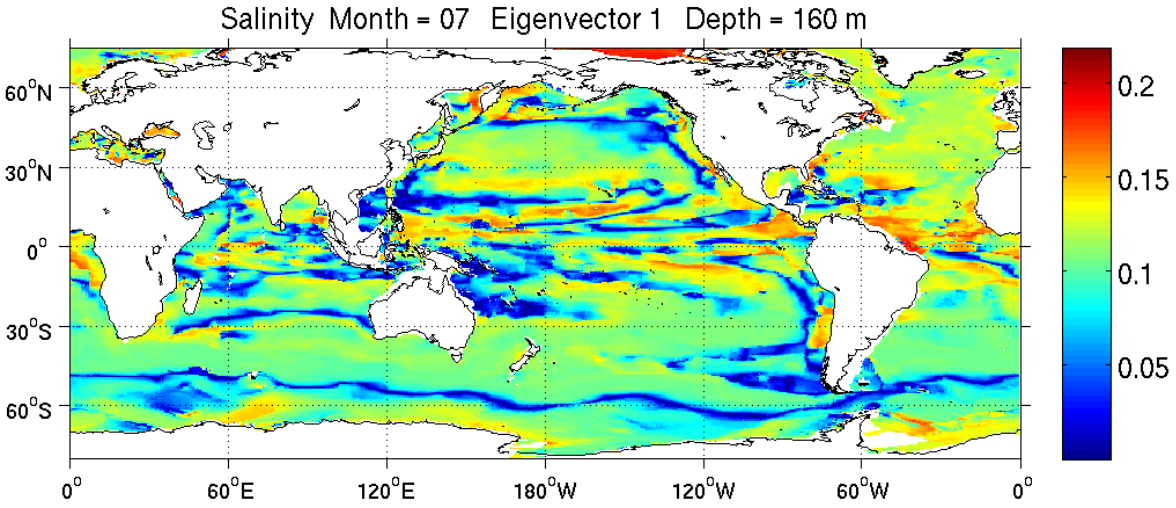


Figure 14 Same as Figure 4, but for salinity.

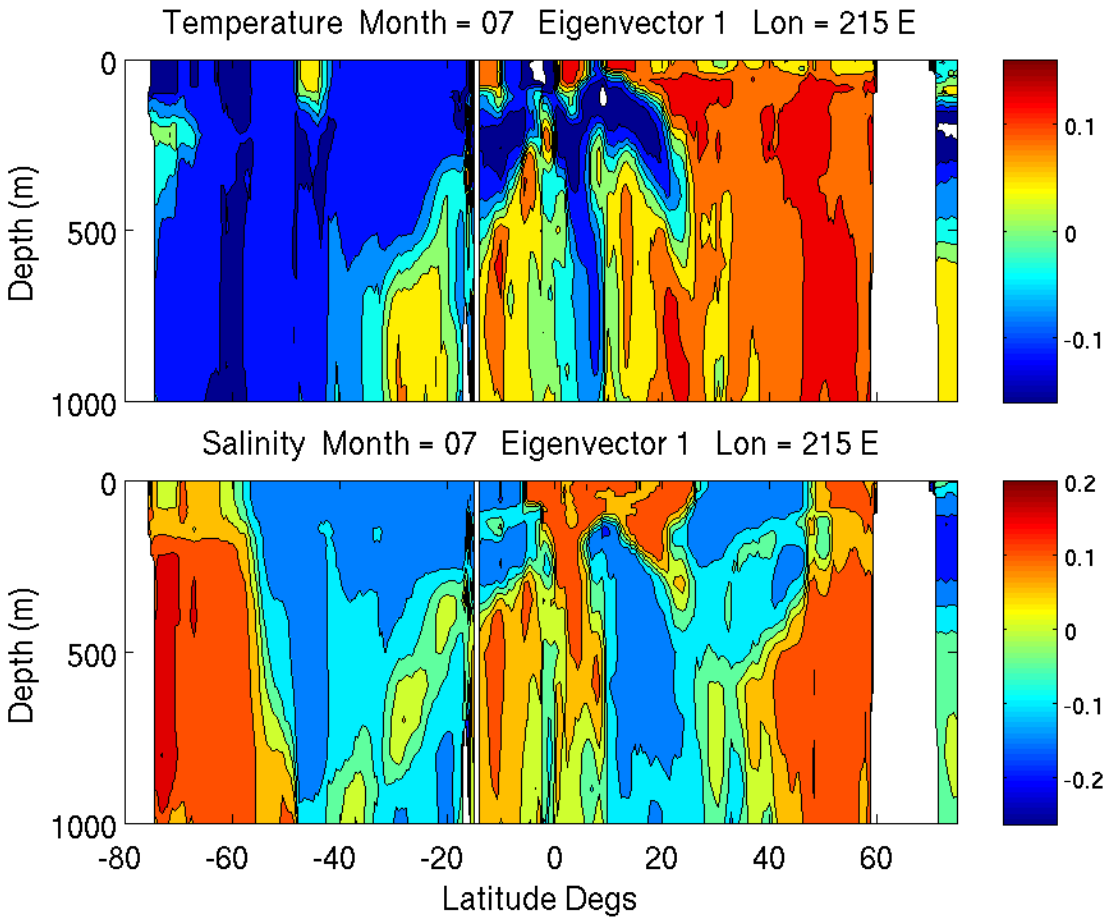


Figure 15. Vertical sections of first mode eigenvectors of the scaled temperature and salinity for June in the Pacific Ocean along 215° E (145° W) longitude from the coast of Antarctica to the coast of Alaska.

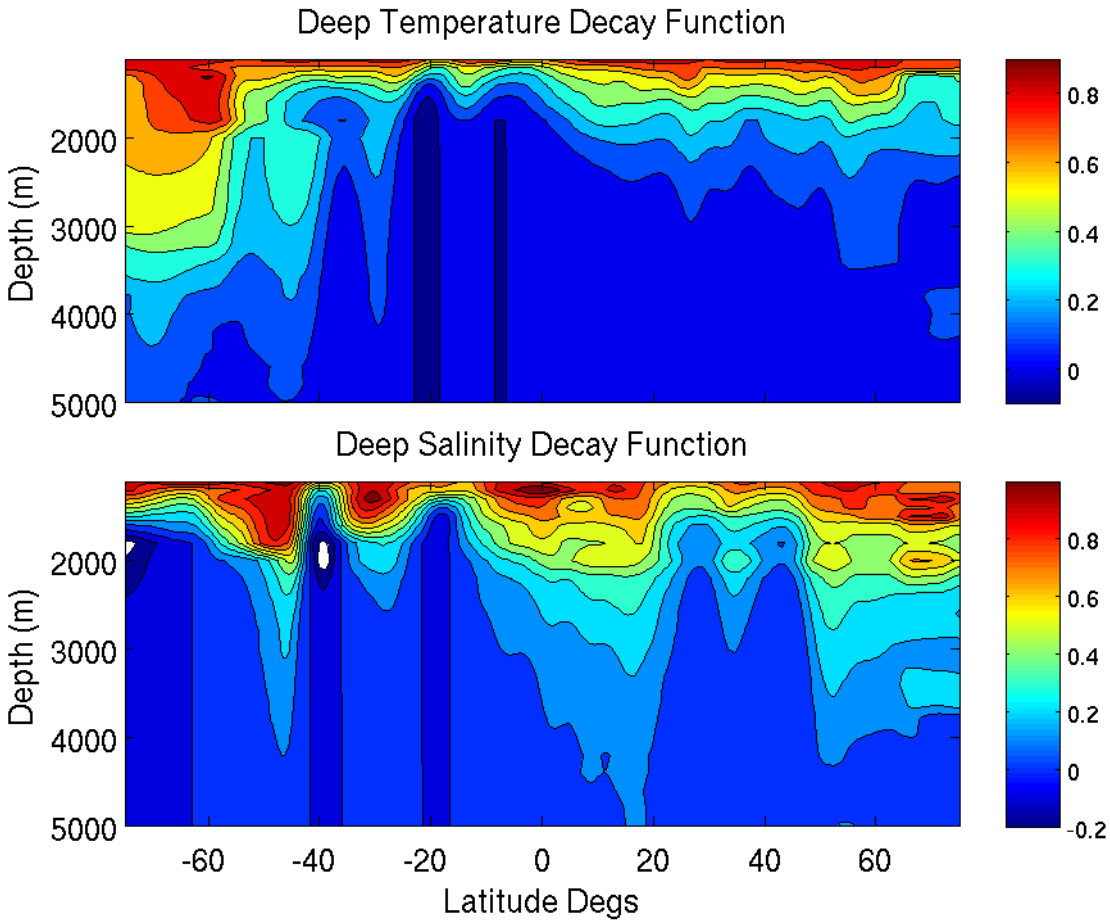


Figure 16. Vertical sections between 1100 m and 5000 m along 215 E (145 W) longitude from the Antarctic coast to the Alaska coast.

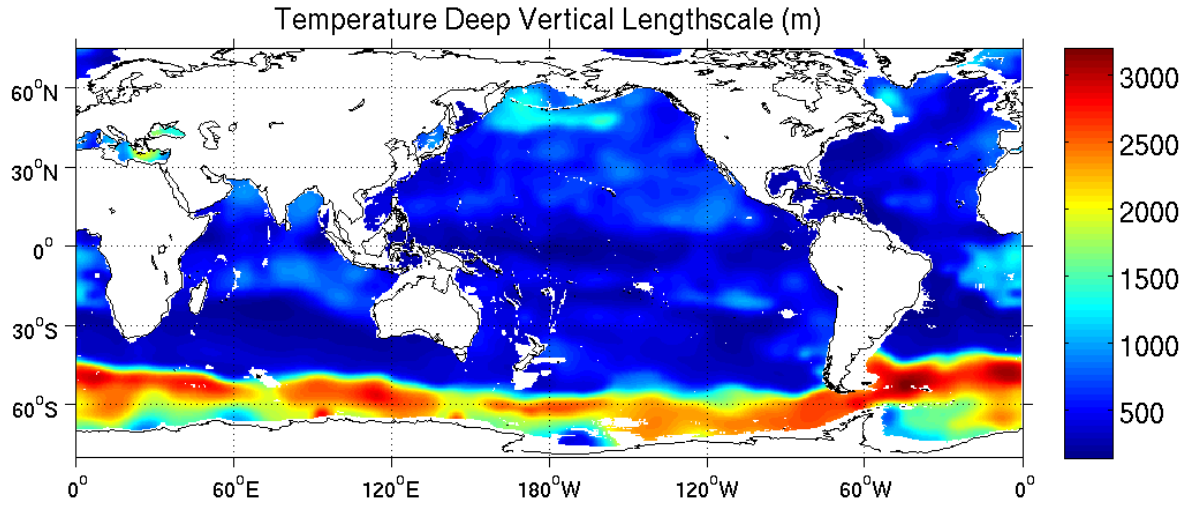


Figure 17. Global distribution of the temperature vertical length scale $\hat{L}_T$ for the Layer 3 relaxation to climatology.

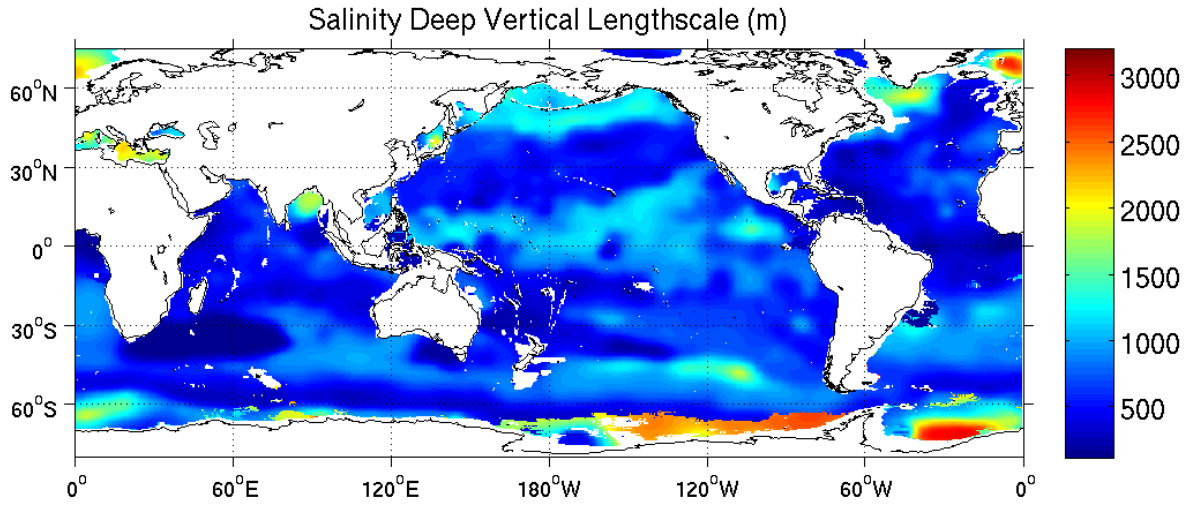


Figure 18. Global distribution of the salinity vertical length scale $\hat{L}_S$ for the Layer 3 relaxation to climatology.

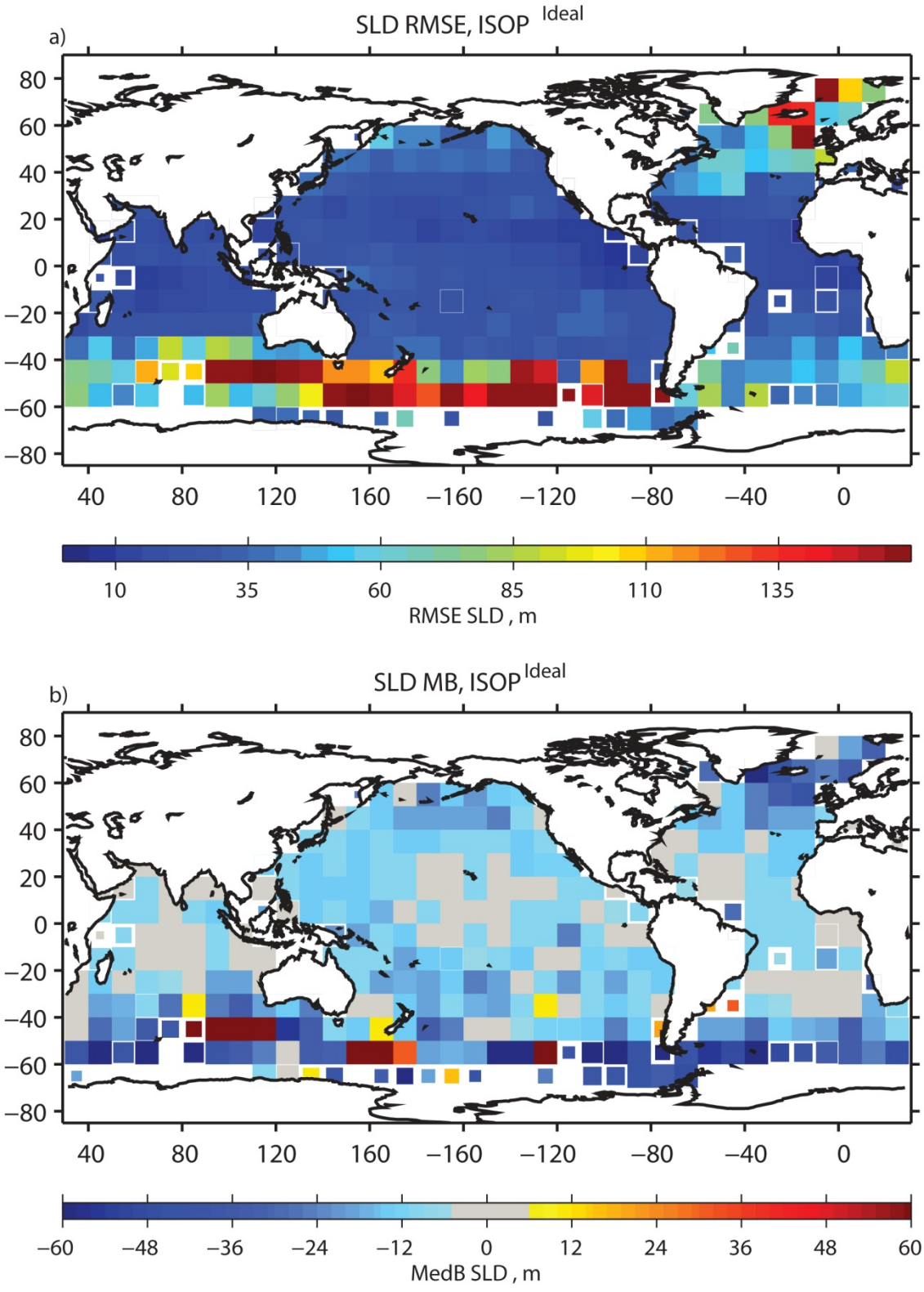


Figure 19. The *ideal* ISOP SLD (a) RMSE and (b) MB error binned in 10° by 10° blocks. The units are in meters and each non-white block in (a) has at least 5 validation profiles. The convention of MB is synthetic – observation.

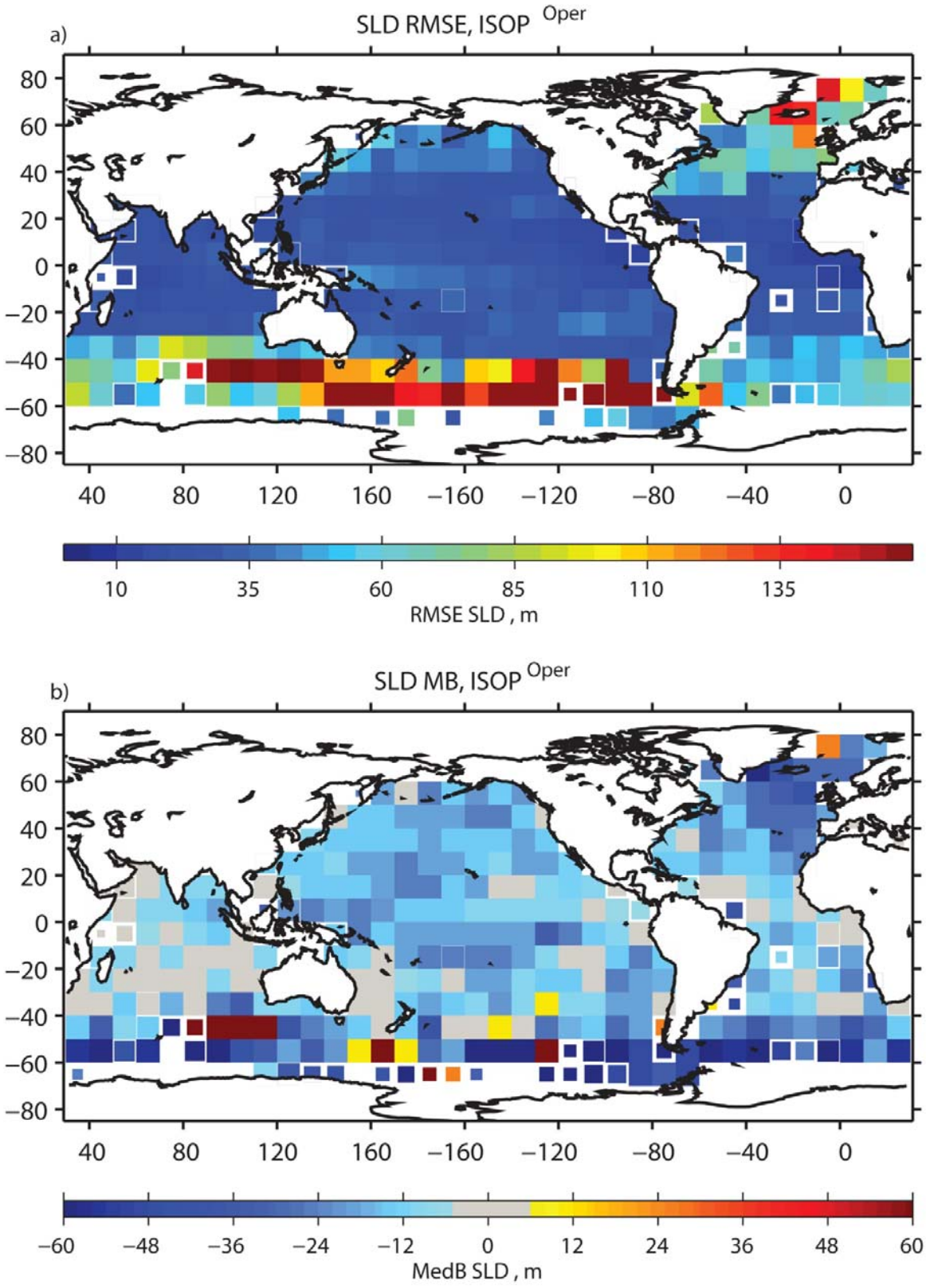


Figure 20. The *operational* ISOP SLD (a) RMSE and (b) MB error binned in 10° by 10° blocks. The units are in meters and each non-white block in (a) has at least 5 validation profiles. The convention of MB is synthetic – observation.

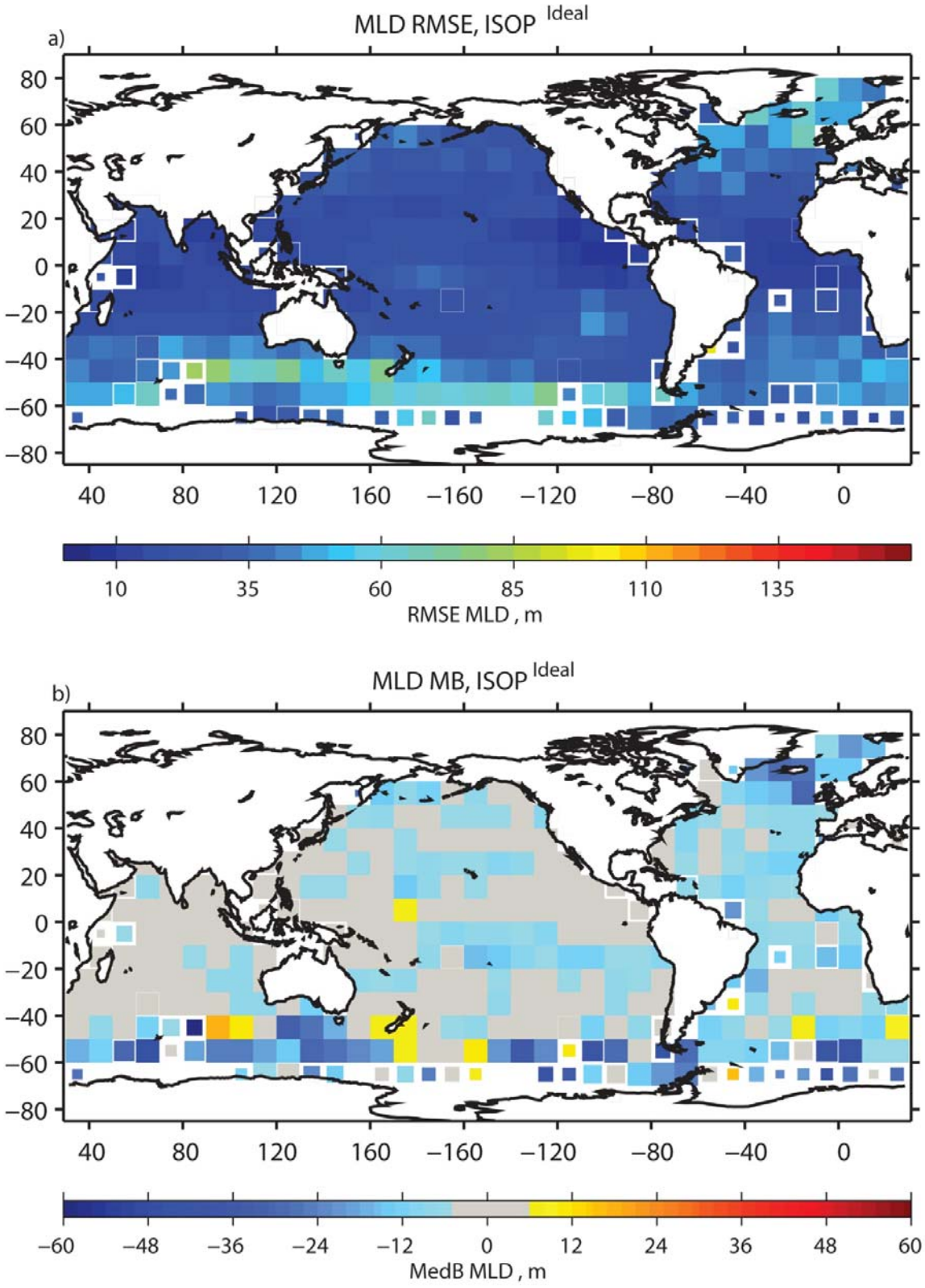


Figure 21. The *ideal* ISOP MLD (a) RMSE and (b) MB error binned in 10° by 10° blocks. The units are in meters and each non-white block in (a) has at least 5 validation profiles. The convention of MB is synthetic – observation.

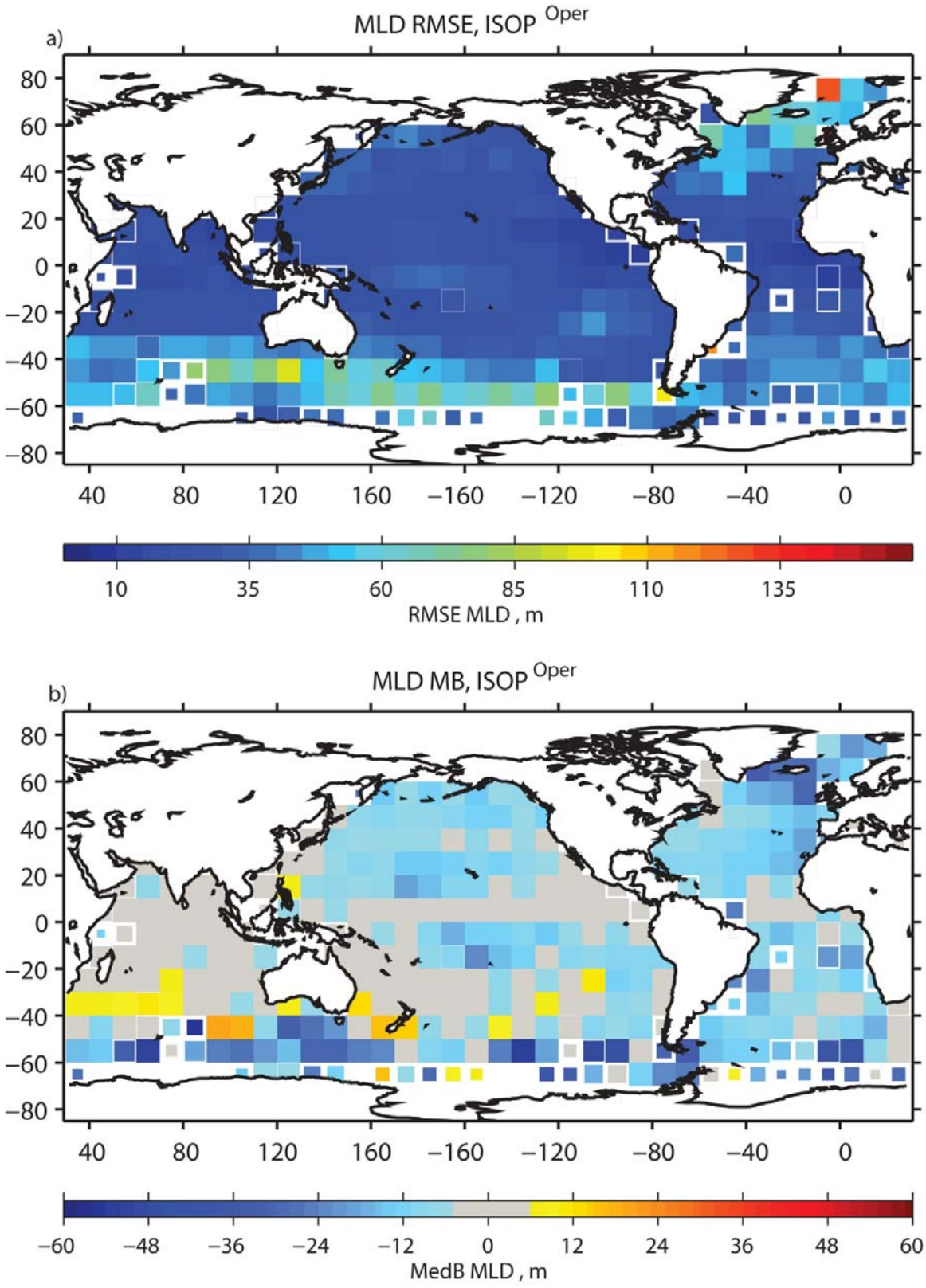


Figure 22. The *operational* ISOP MLD (a) RMSE and (b) MB error binned in 10° by 10° blocks. The units are in meters and each non-white block in (a) has at least 5 validation profiles. The convention of MB is synthetic – observation.

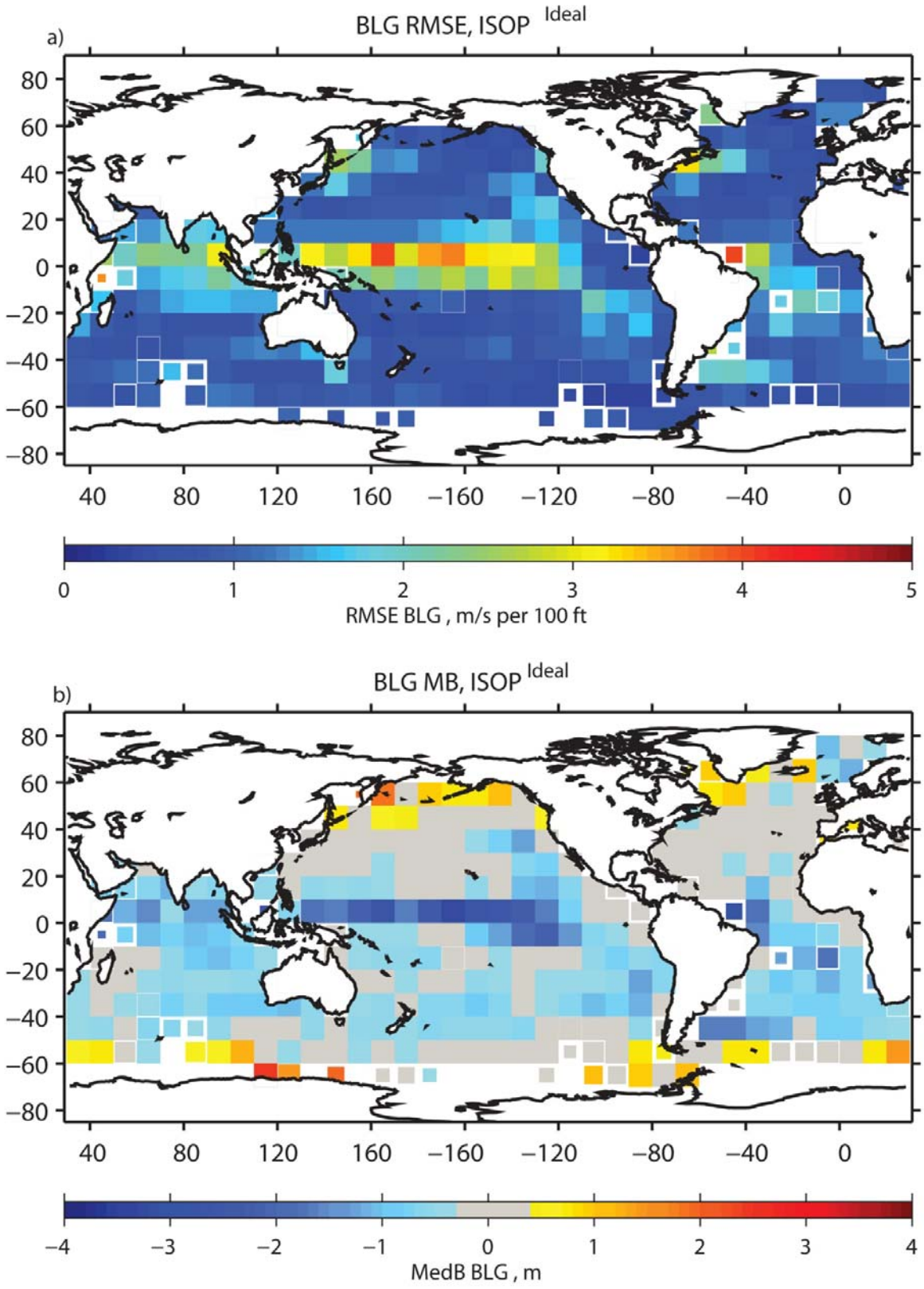


Figure 23. The *ideal* ISOP BLG (a) RMSE and (b) MB error binned in 10° by 10° blocks. The units are in meters and each non-white block in (a) has at least 5 validation profiles. The convention of MB is synthetic – observation.

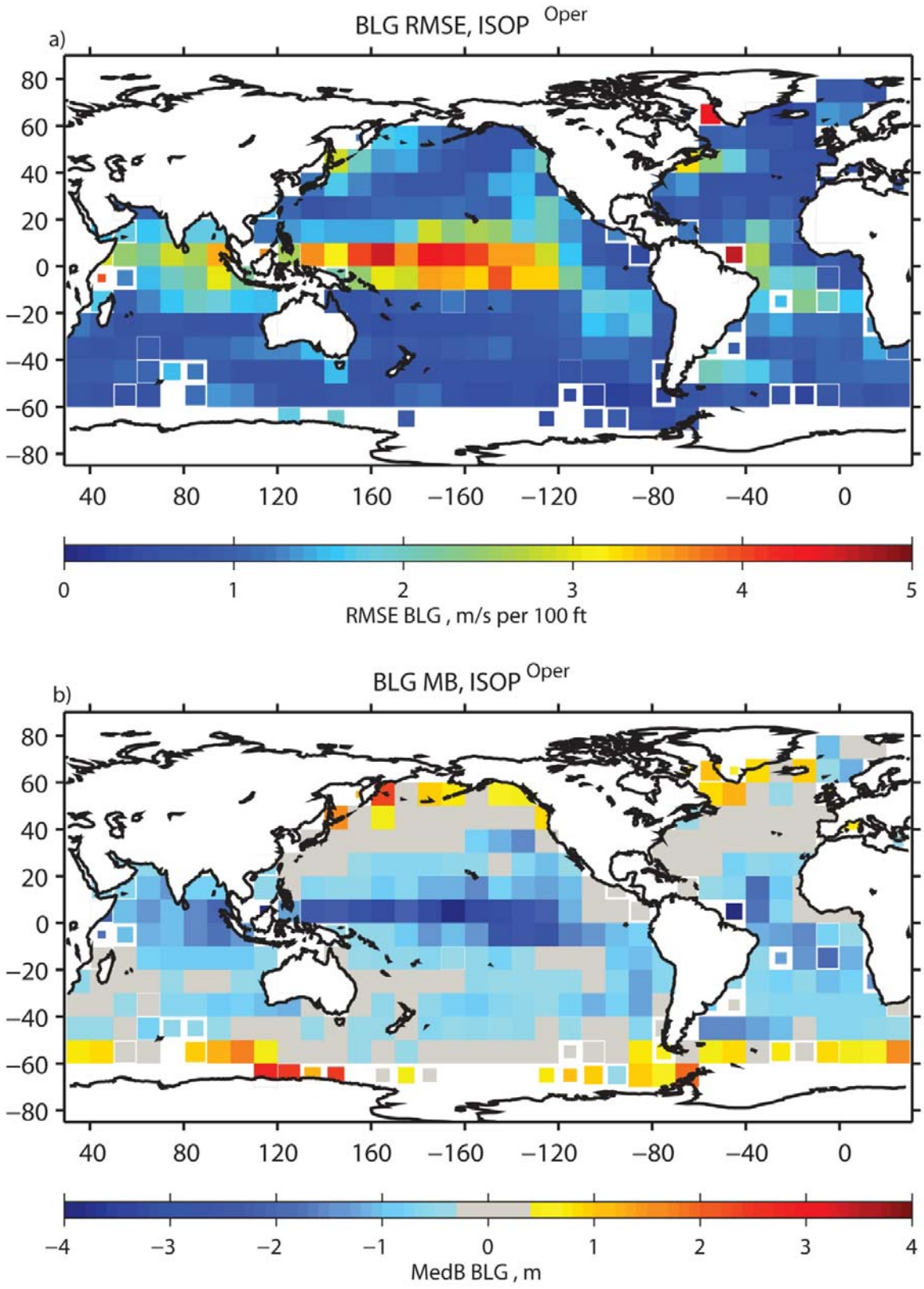


Figure 24. The *operational* ISOP BLG (a) RMSE and (b) MB error binned in 10° by 10° blocks. The units are in meters and each non-white block in (a) has at least 5 validation profiles. The convention of MB is synthetic – observation.

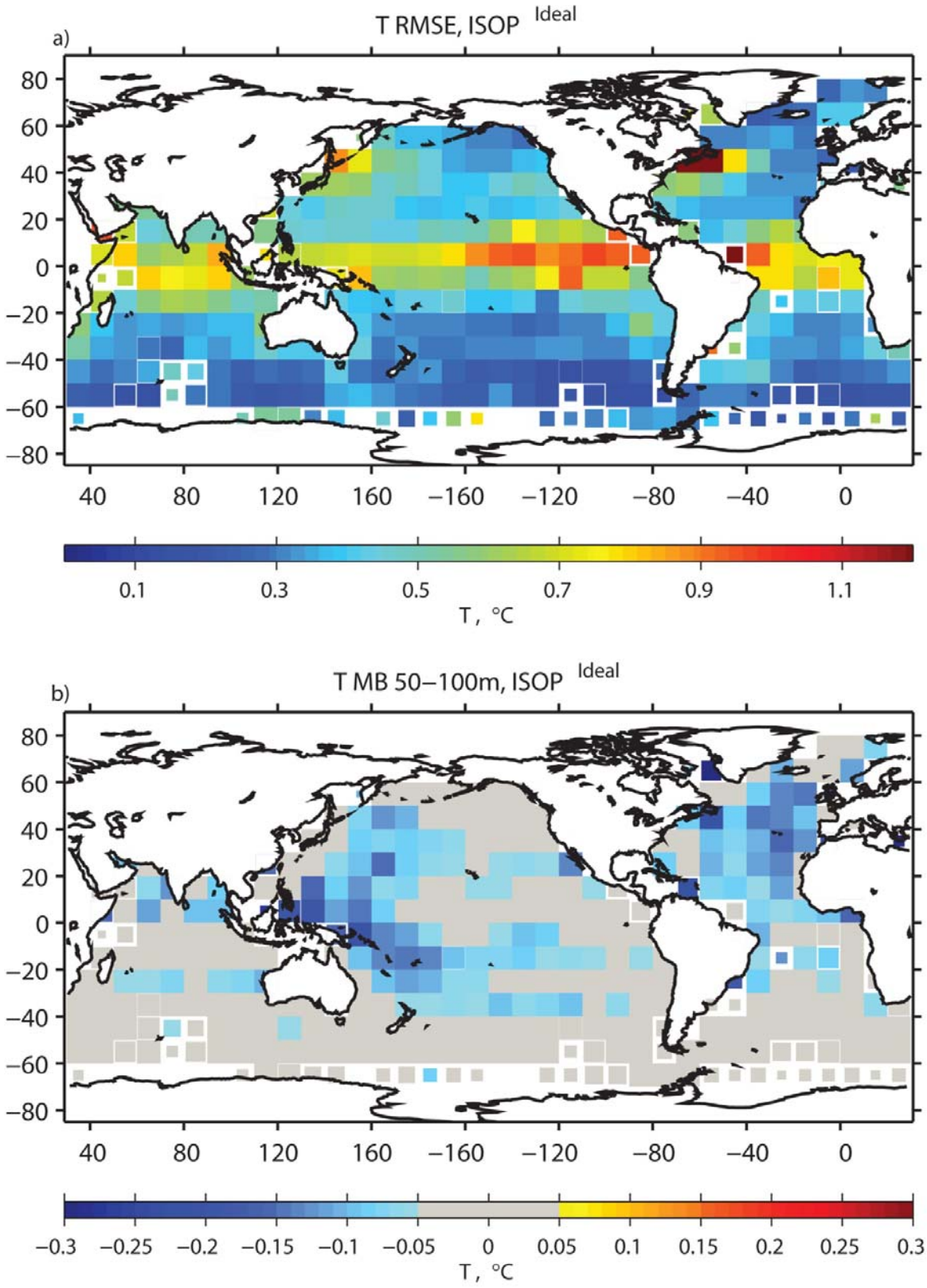


Figure 25. The median $ISOP^{Ideal}$ T (a) RMSE and (b) MB error in the 50 to 100 m depth range binned in 10° by 10° blocks.

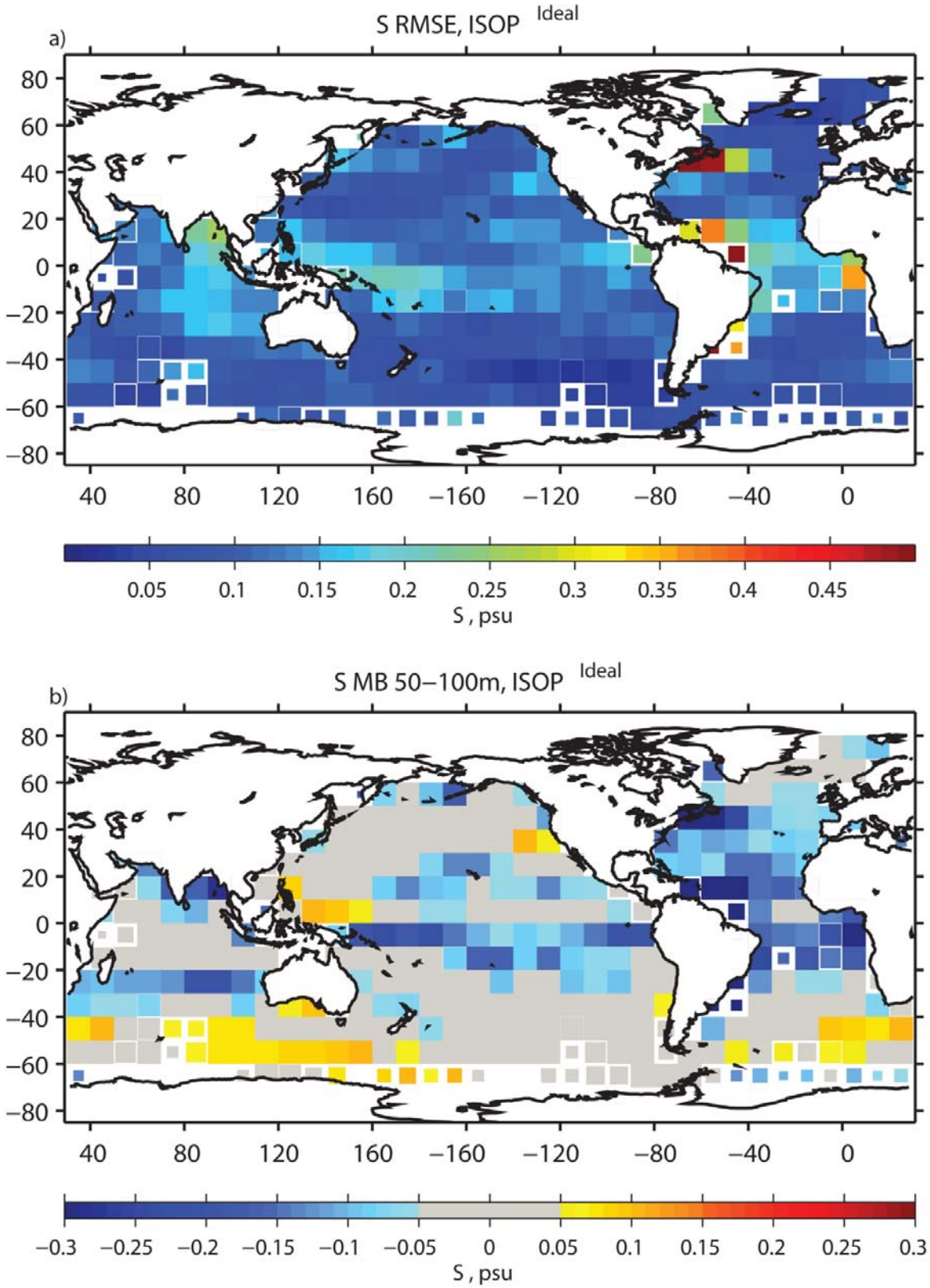


Figure 26. The median $ISOP^{Ideal}$ S (a) RMSE and (b) MB error in the 50 to 100 m depth range binned in 10° by 10° blocks.

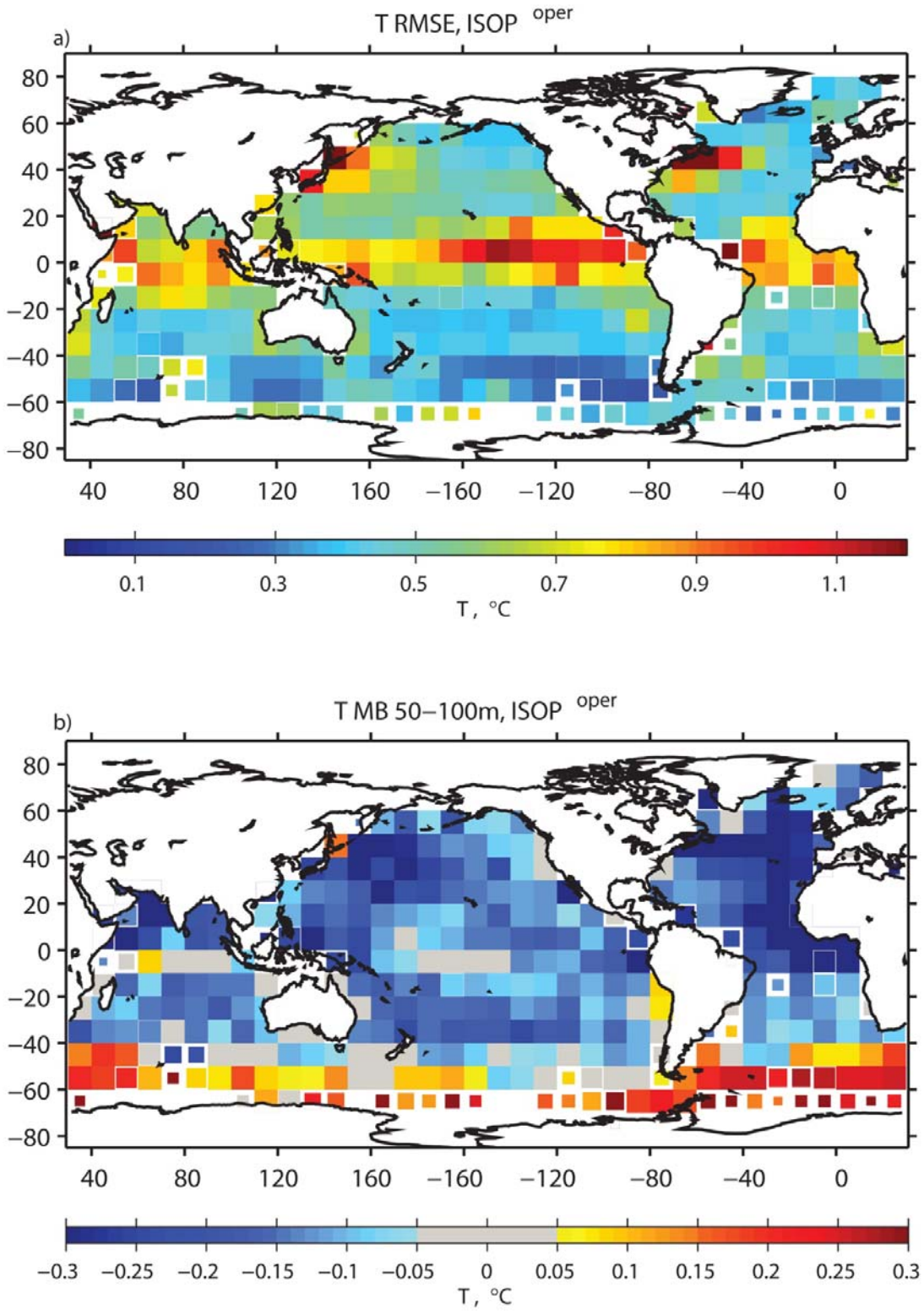


Figure 27. The median *ISOP*^{Oper} T (a) RMSE and (b) MB error in the 50 to 100 m depth range binned in 10° by 10° blocks.

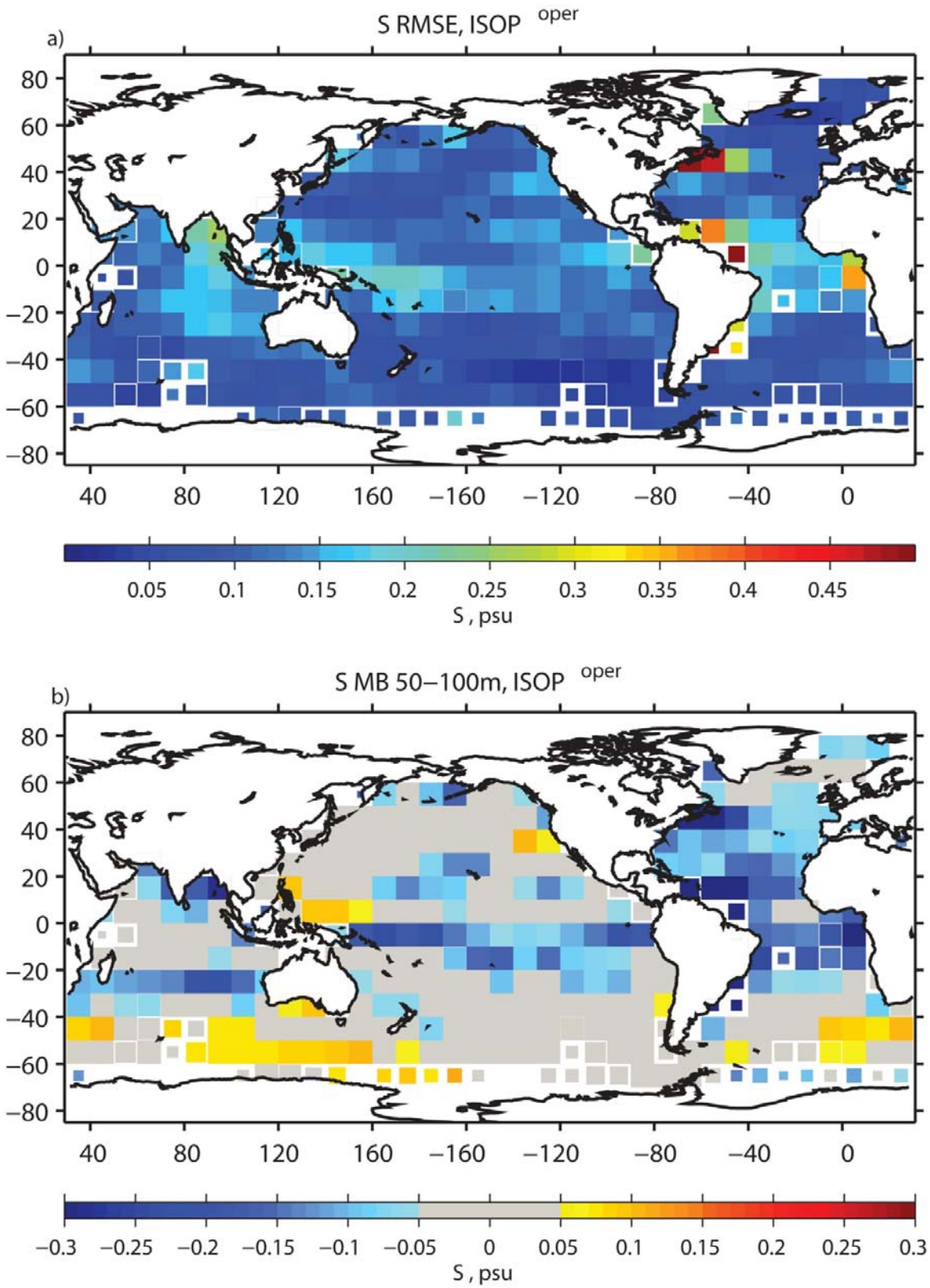


Figure 28. The median $ISOP^{Oper}$ S (a) RMSE and (b) MB error in the 50 to 100 m depth range binned in 10° by 10° blocks.

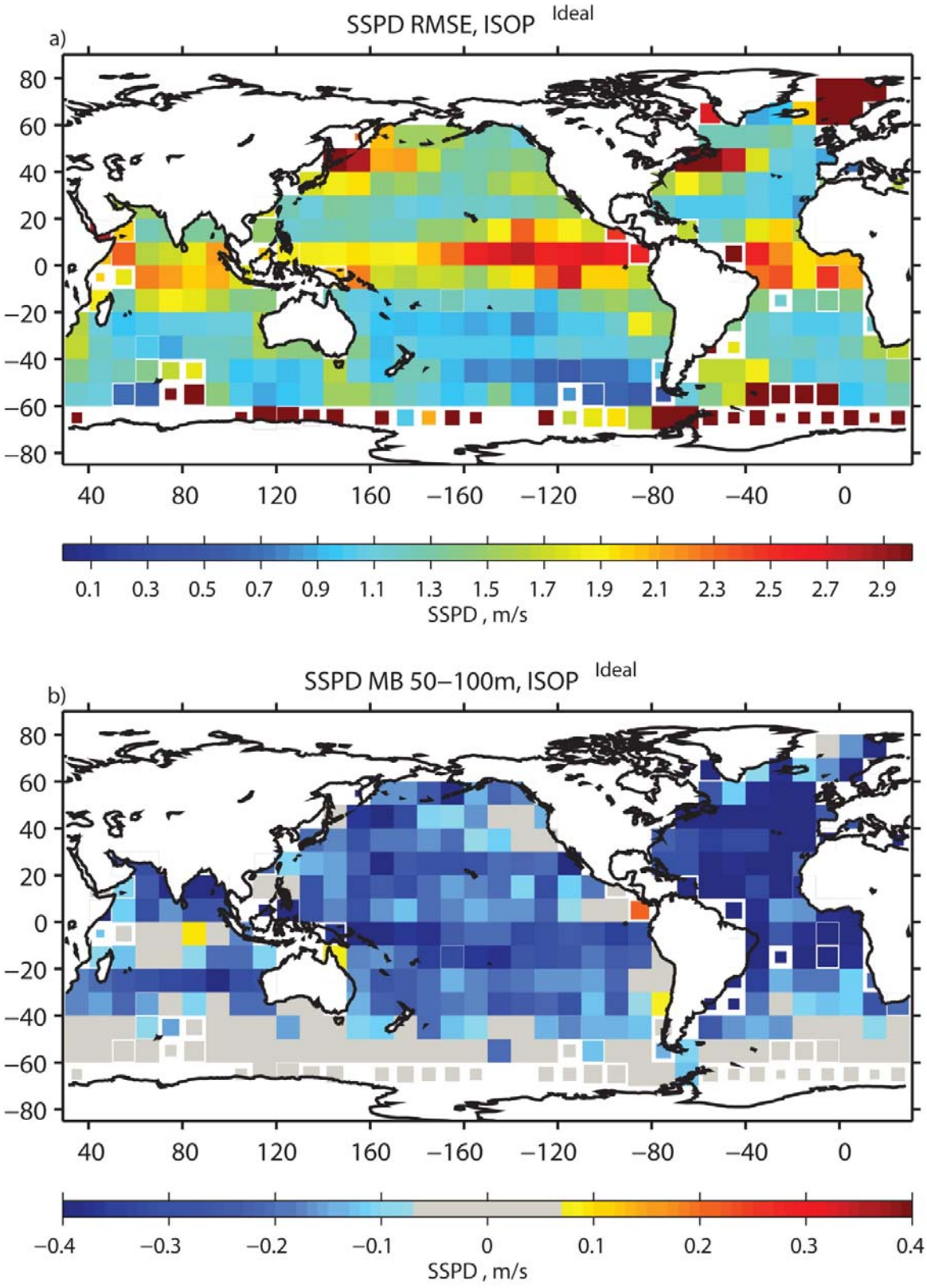


Figure 29. The median $ISOP^{Ideal}$ SSPD (a) RMSE and (b) MB error in the 50 to 100 m depth range binned in 10° by 10° blocks.

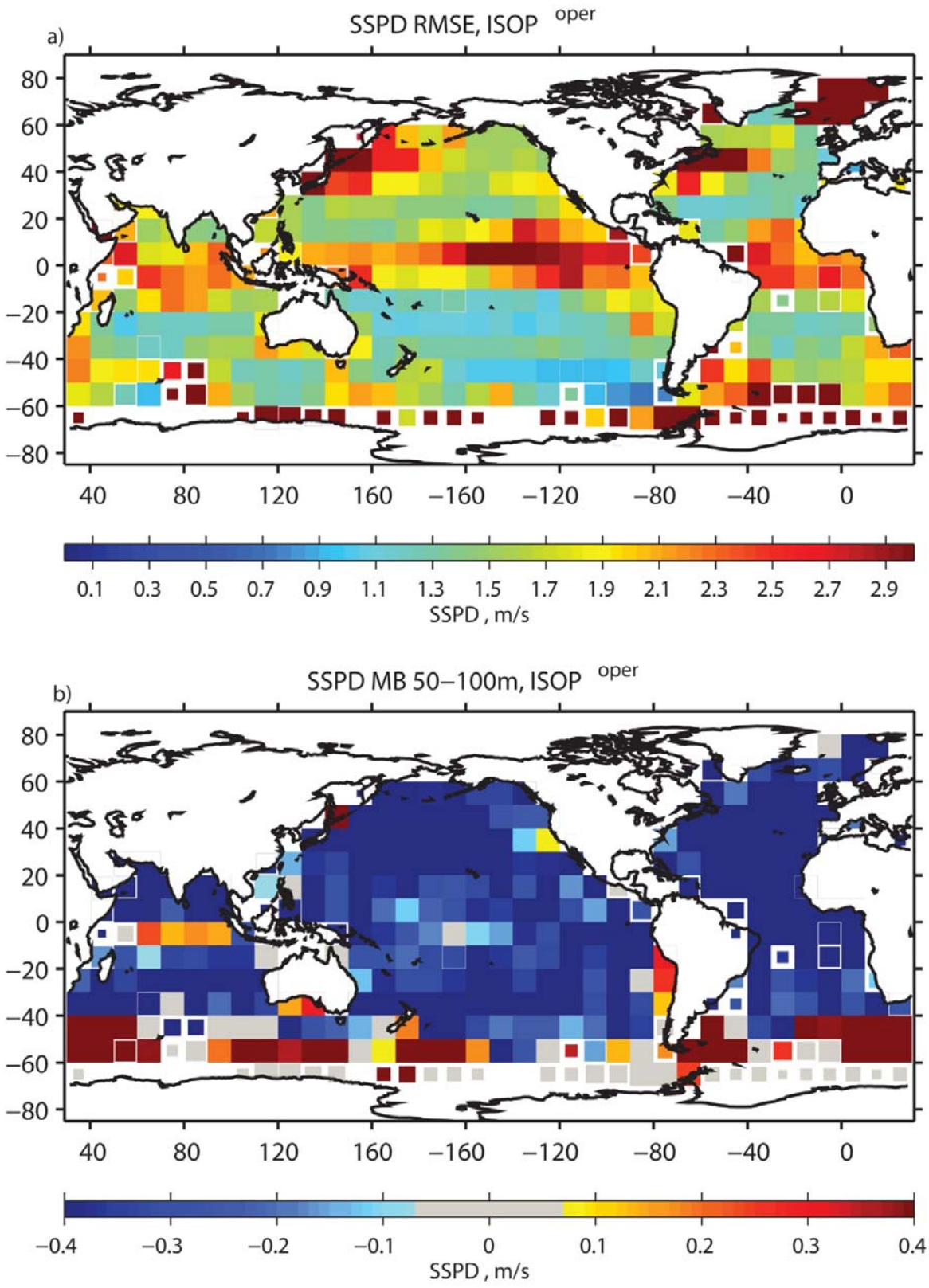


Figure 30. The median $ISOP^{Oper.}$ (a) RMSE and (b) MB binned in 10° by 10° blocks for sound speed (SSPD) in the 50 to 100 m depth range.

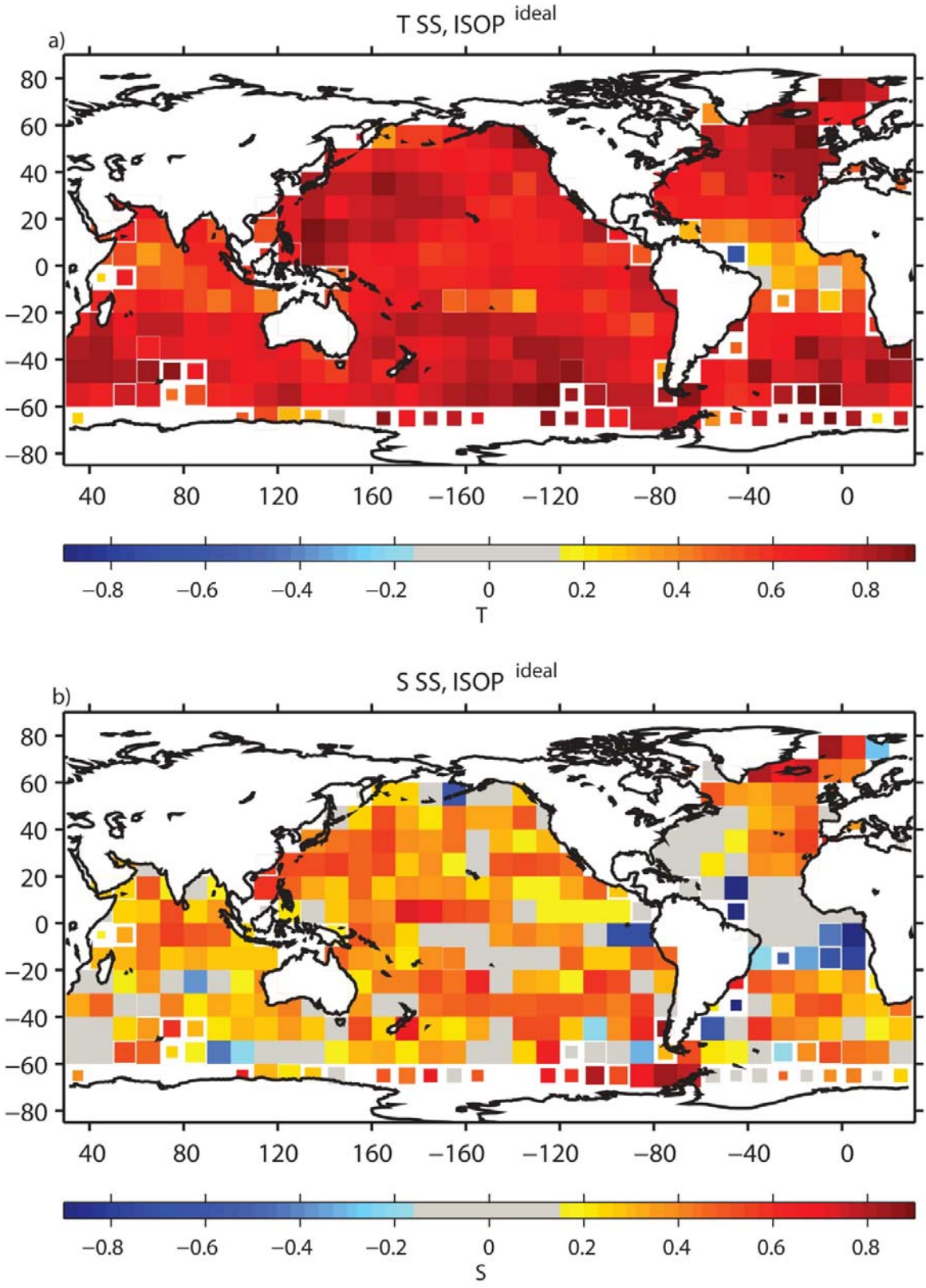


Figure 31. The $ISOP^{Ideal}$ median skill score in each 10°by 10° bin for (a) temperature and (b) salinity over the upper 1000 m.

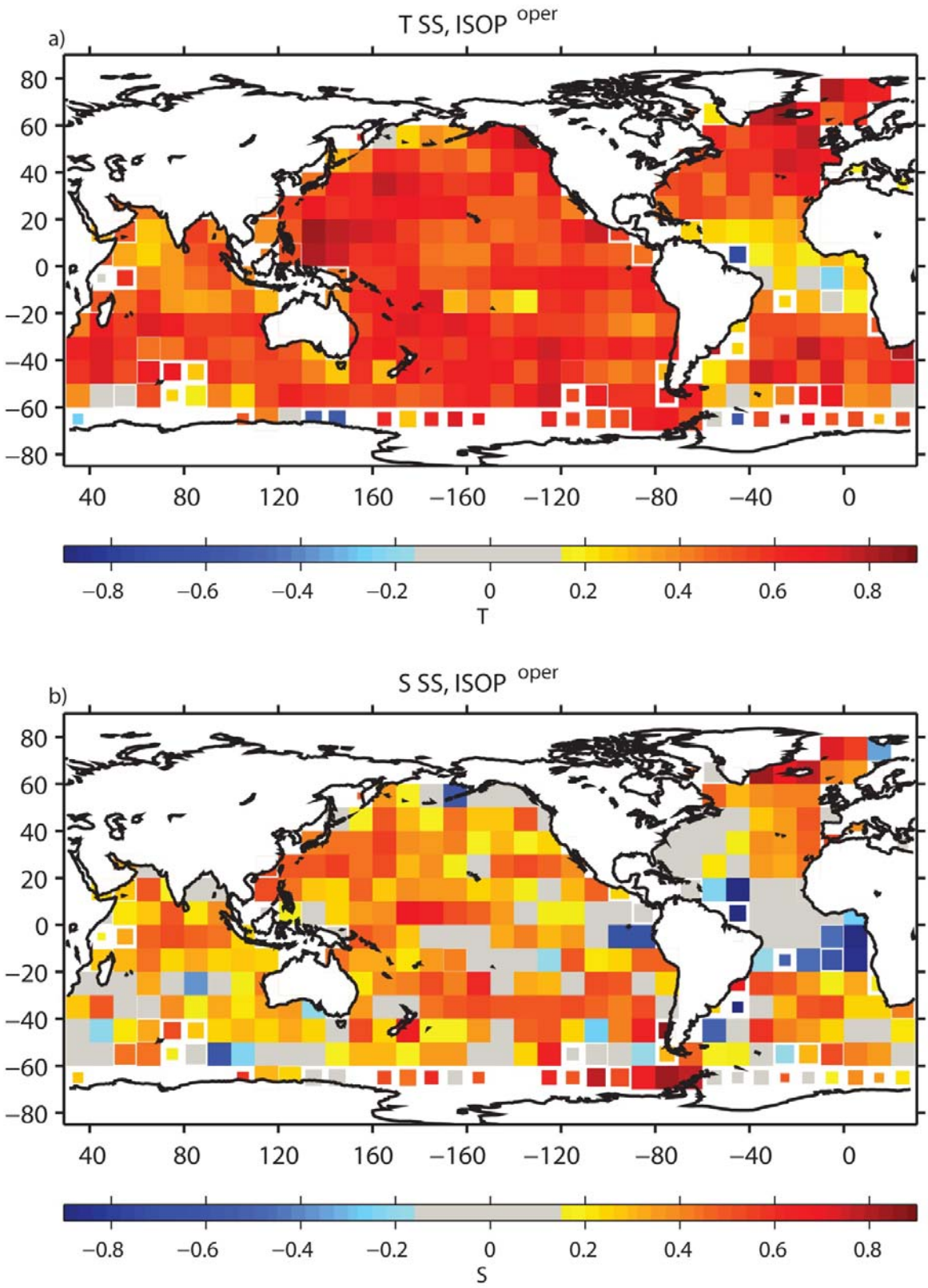


Figure 32. The $ISOP^{Oper}$ median skill score in each 10°by 10° bin for (a) temperature and (b) salinity over the upper 1000 m.

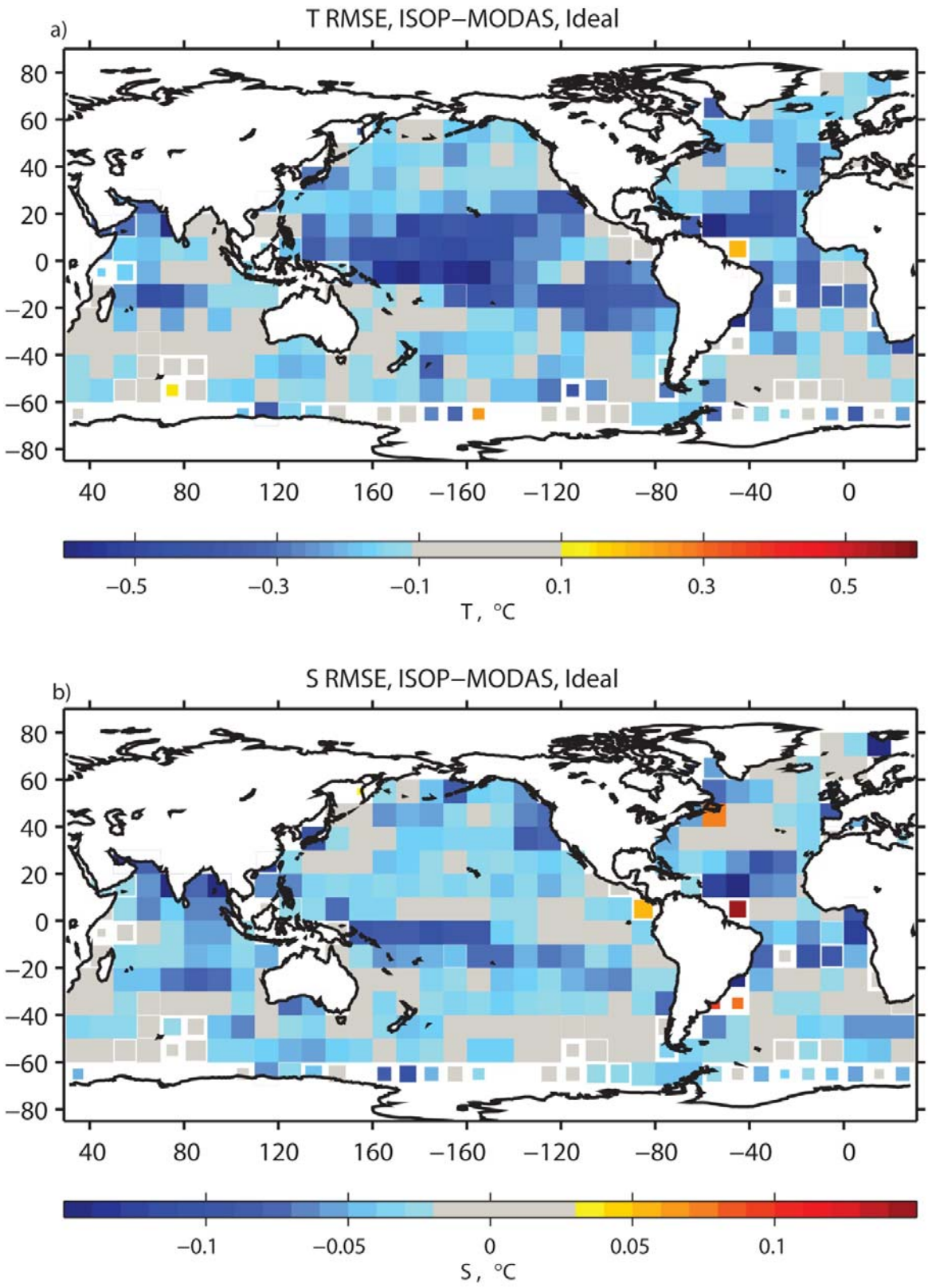


Figure 33. The geographical distribution of the difference in RMSE for (a) T and (b) S for $ISOP^{Ideal} - MODAS^{Ideal}$.

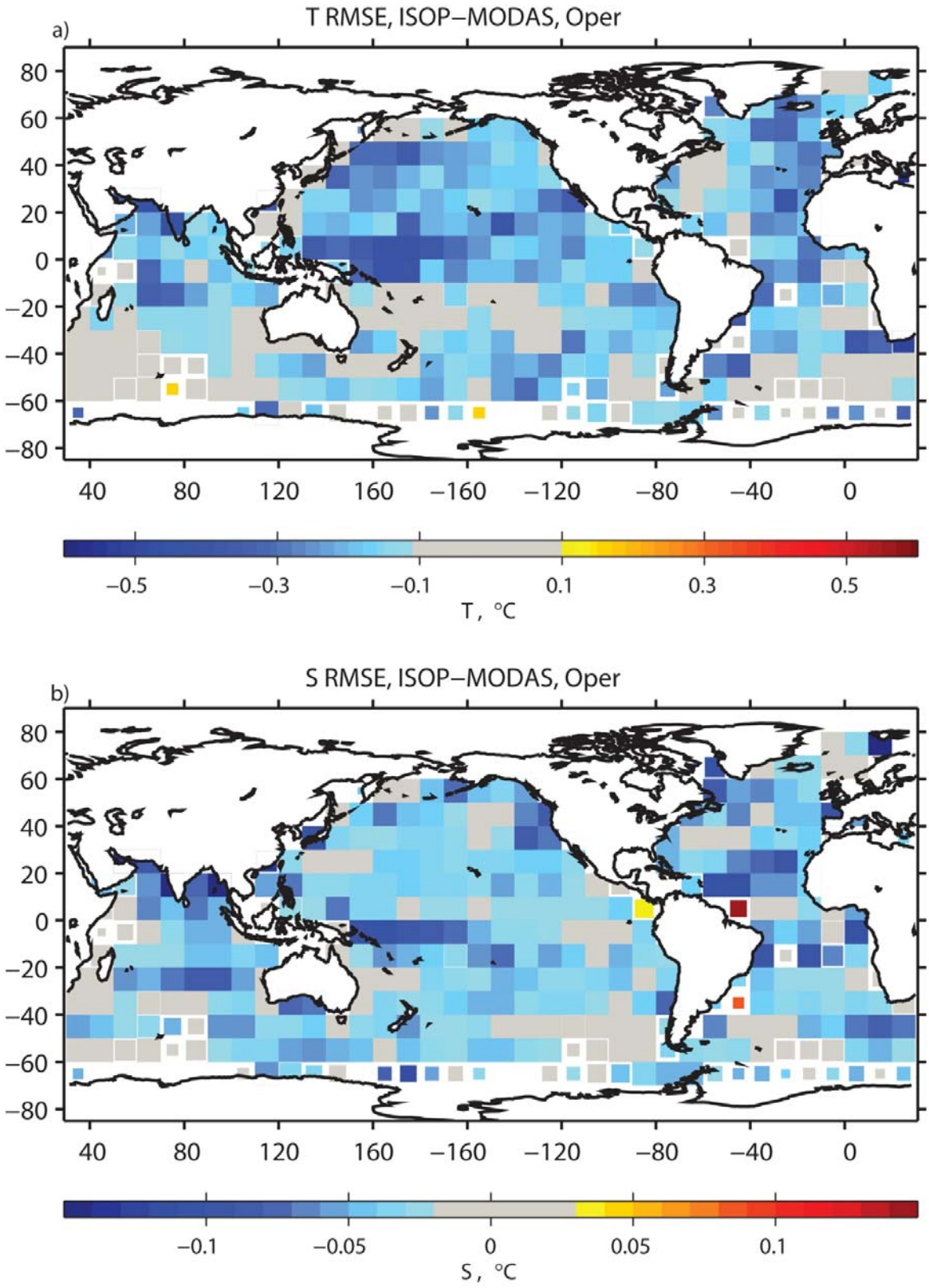


Figure 34. The geographical distribution of the difference in RMSE for (a) T and (b) S for $ISOP^{Oper} - MODAS^{Oper}$.

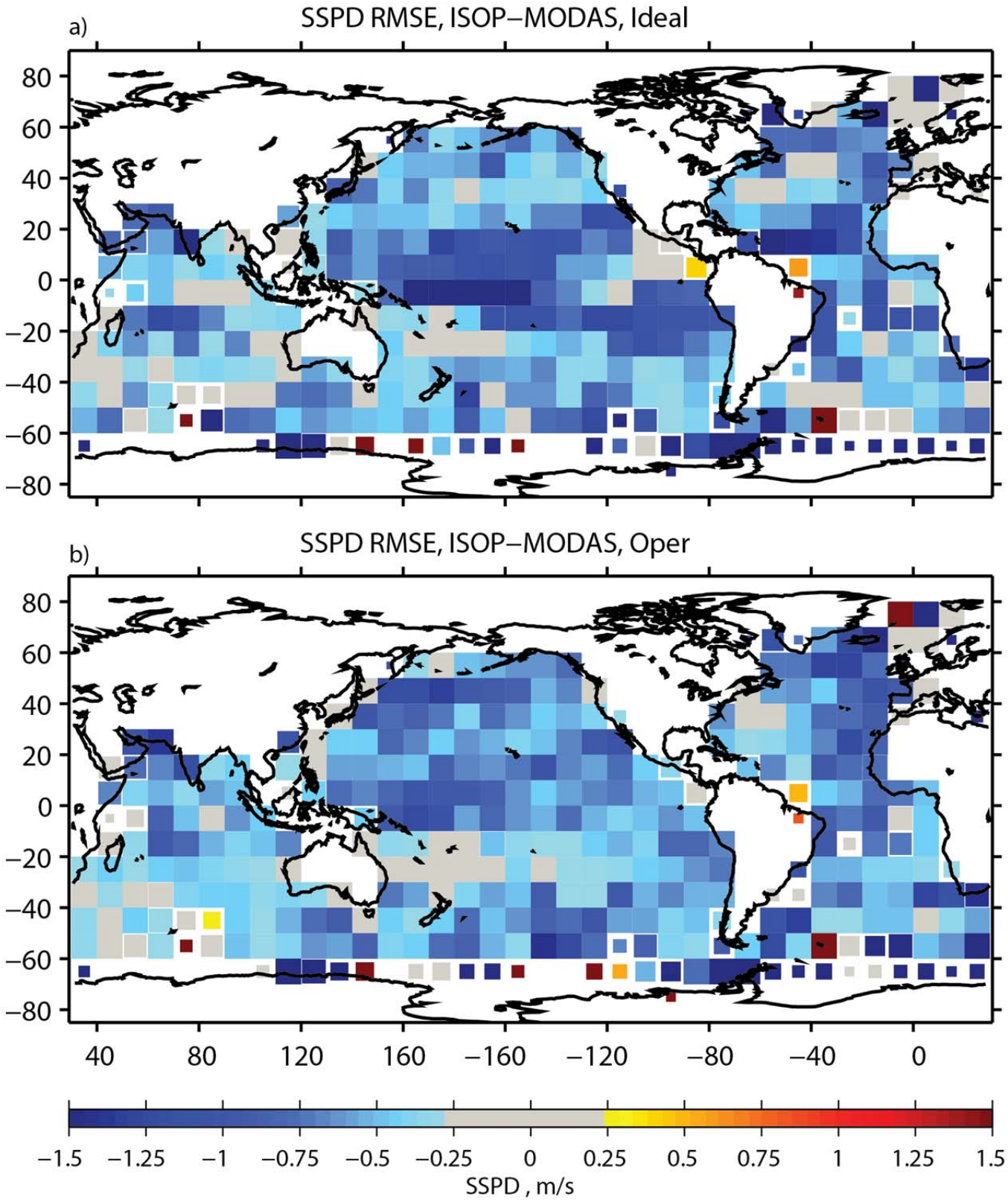


Figure 35. The geographical distribution of the difference in SSPD RMSE for (a) $ISOP^{Ideal} - MODAS^{Ideal}$ and (b) $ISOP^{Oper} - MODAS^{Oper}$.

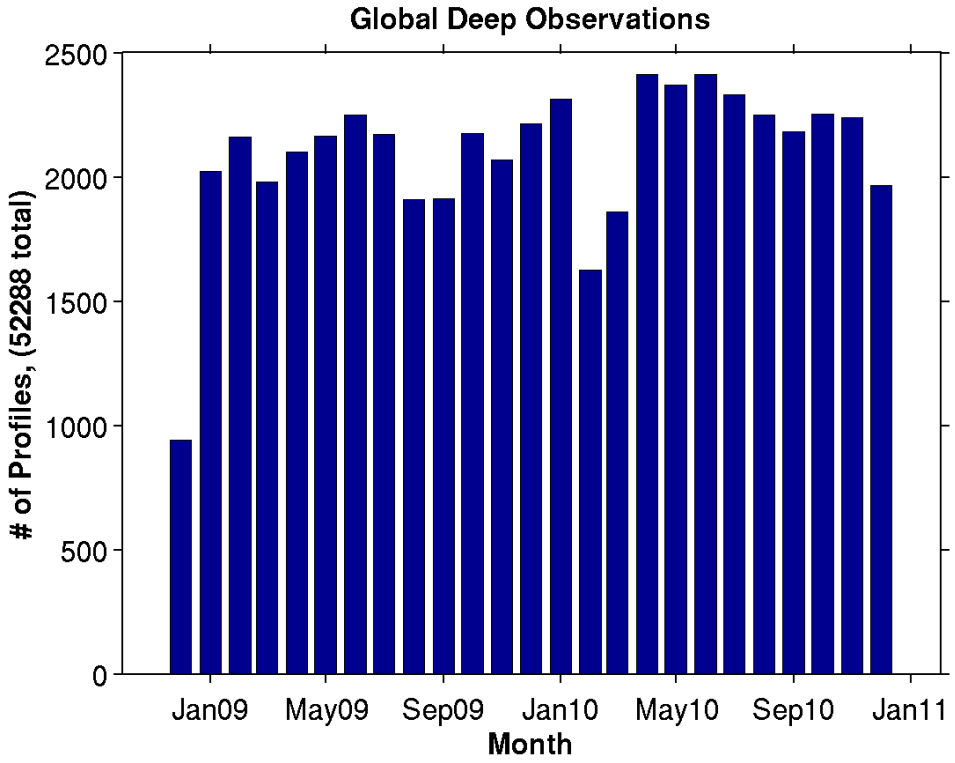


Figure 36. The number of profiles verses month in the global deep validation data set that are used in the validation analysis. There are a total of 52,288 profiles.

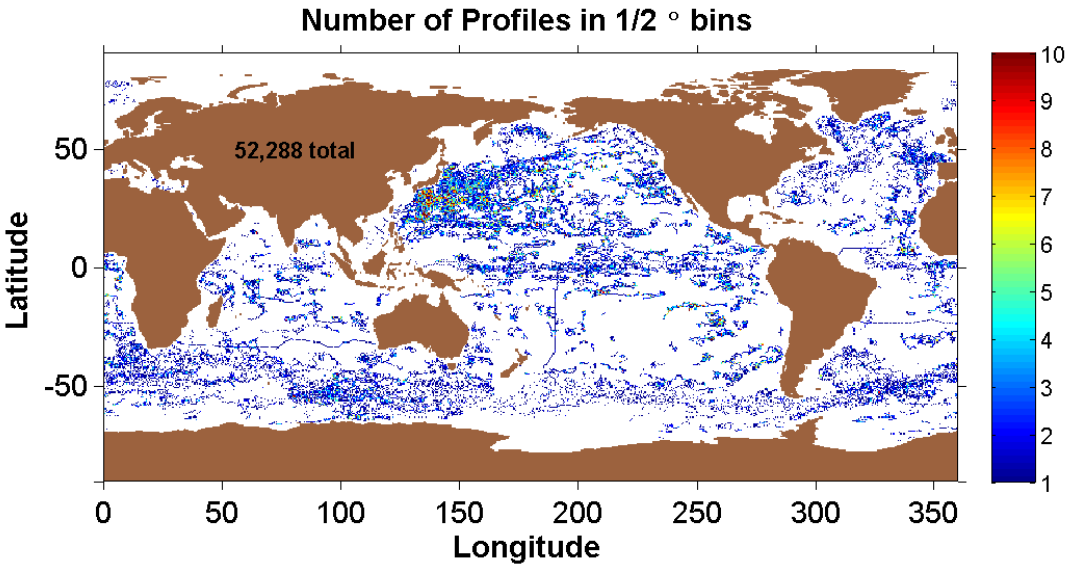


Figure 37. The number of profiles plotted geographically in ½°bins for the global ocean.

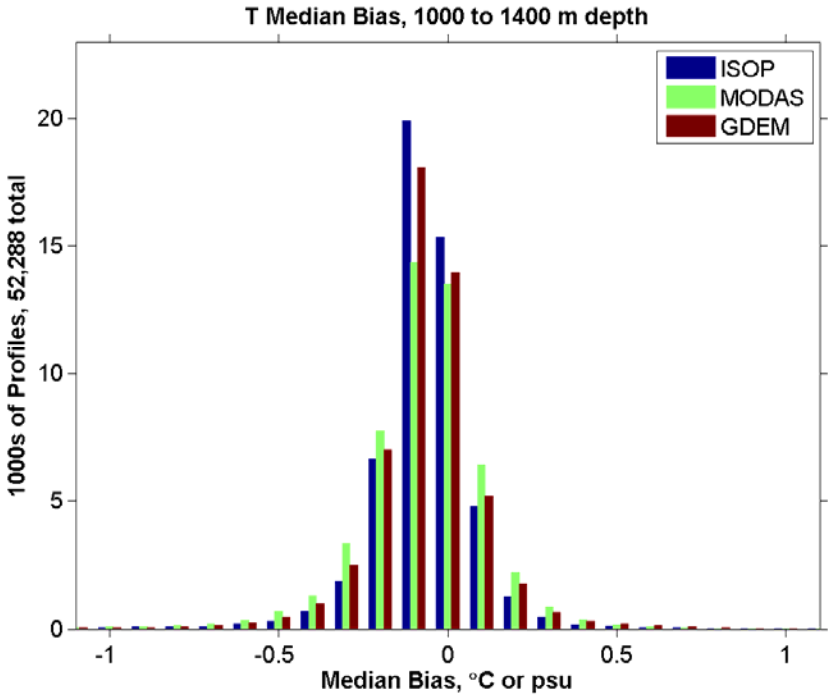


Figure 38. The histogram of temperature (T) bias for $ISOP^{Oper}$, $MODAS^{oper}$, and *GDEM* in the depth range from 1000 to 1400 m.

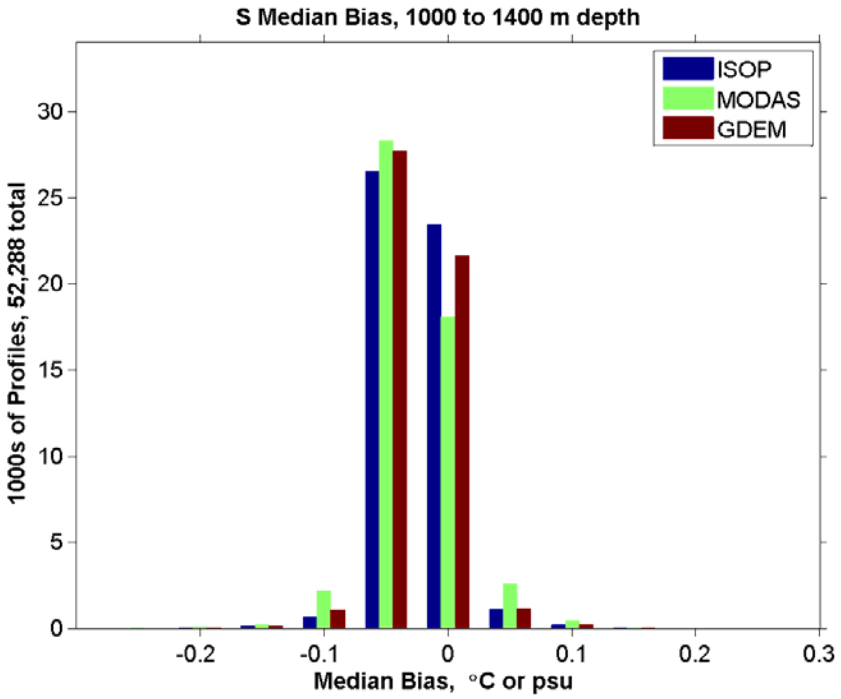


Figure 39. The histogram of temperature (T) bias for $ISOP^{Oper}$, $MODAS^{oper}$, and *GDEM* in the depth range from 1000 to 1400 m.

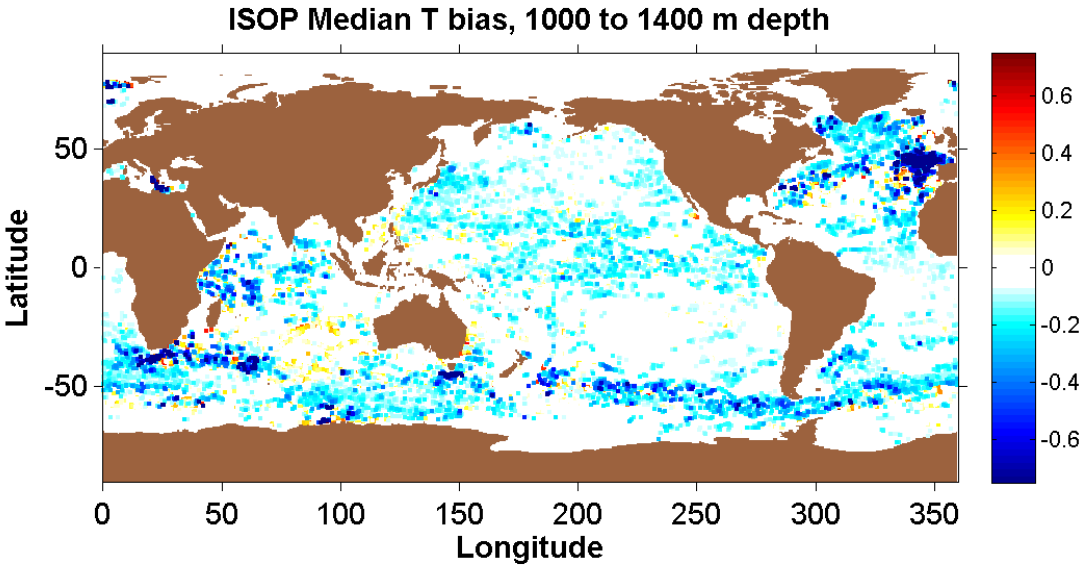


Figure 40. The geographical distribution of $ISOP^{Oper}$ T bias error in the 1000 to 1400 m depth range.

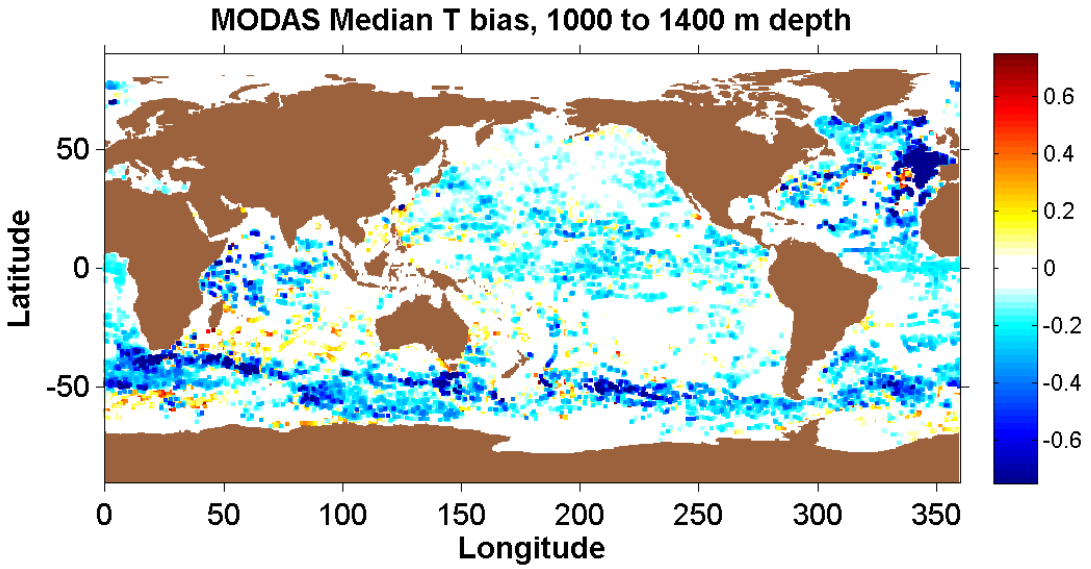


Figure 41. The geographical distribution of $MODAS^{oper}$ T bias error in the 1000 to 1400 m depth range.

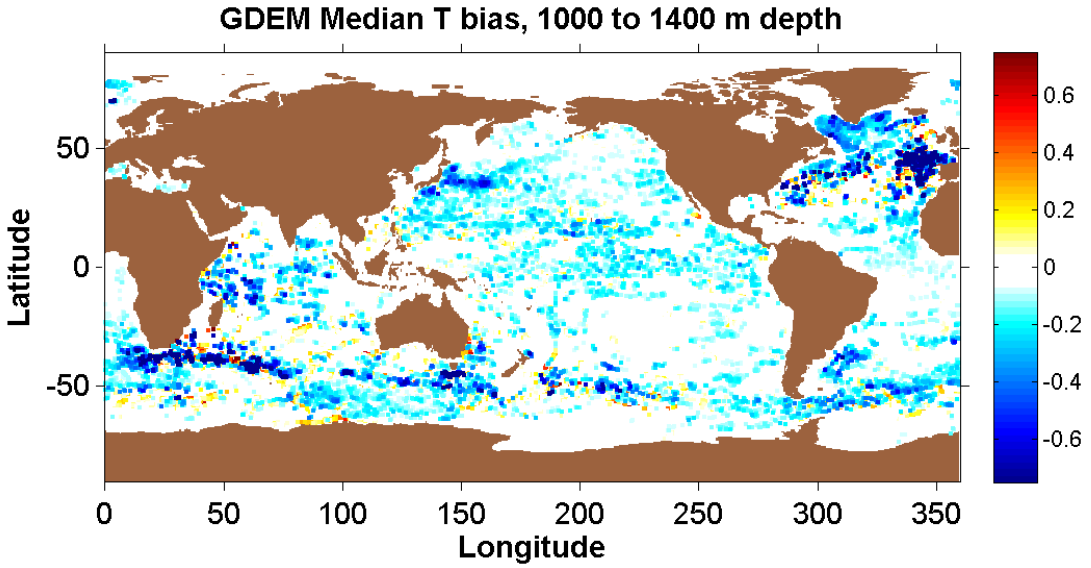


Figure 42. The geographical distribution of *GDEM* T bias error in the 1000 to 1400 m depth range.

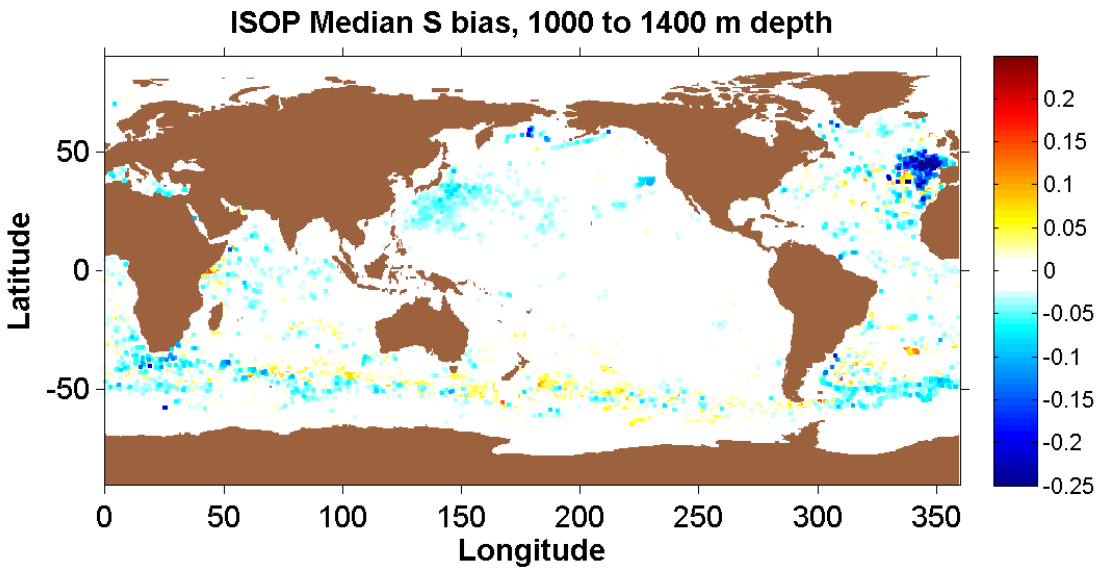


Figure 43. The geographical distribution of $ISOP^{Oper}$ S bias error in the 1000 to 1400 m depth range.

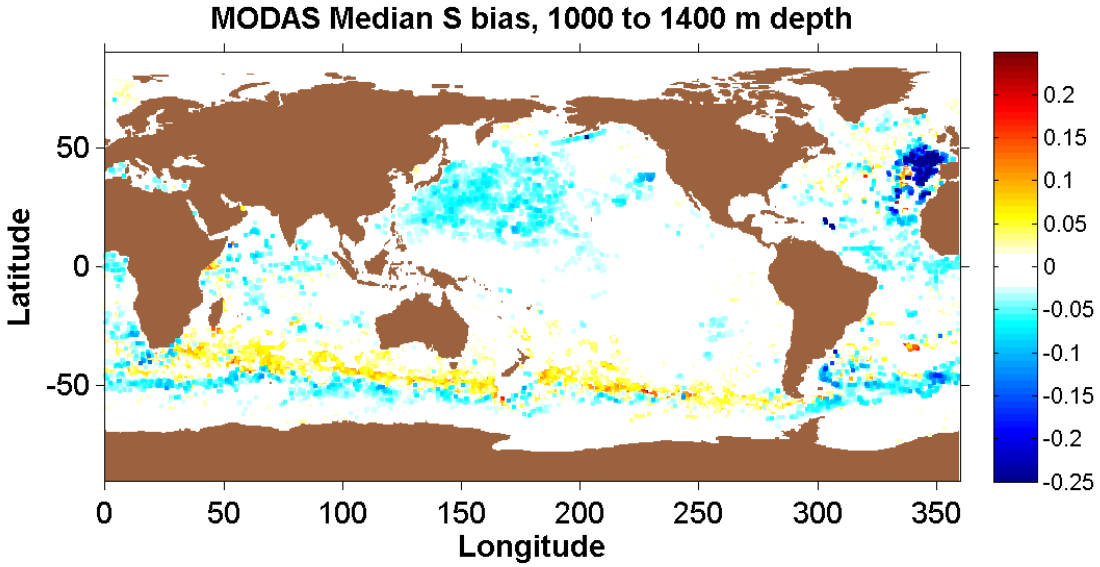


Figure 44. The geographical distribution of $MODAS^{oper}$ S bias error in the 1000 to 1400 m depth range.

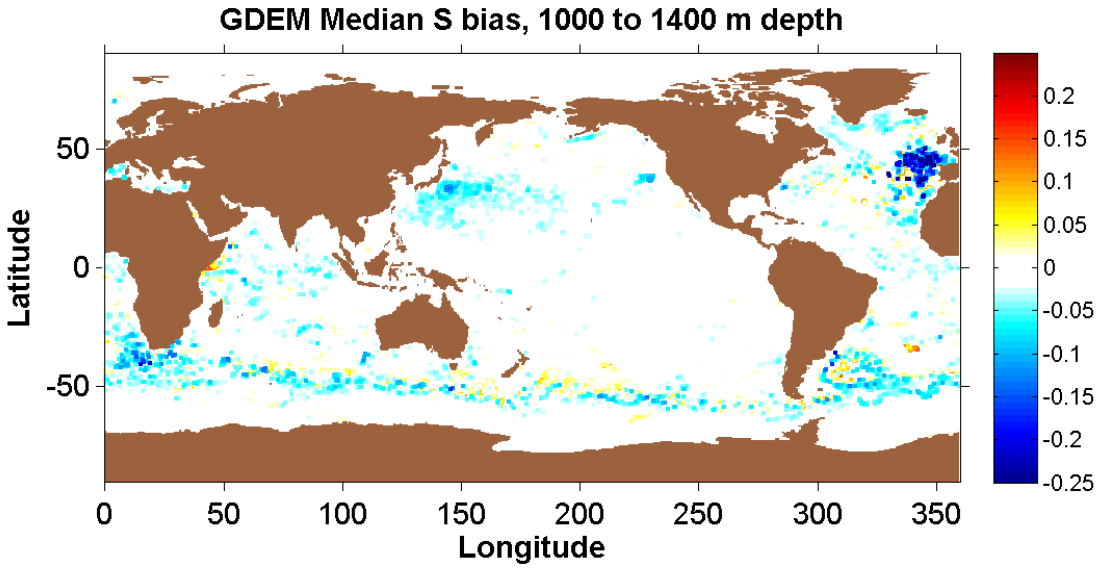


Figure 45. The geographical distribution of *GDEM* S bias error in the 1000 to 1400 m depth range.

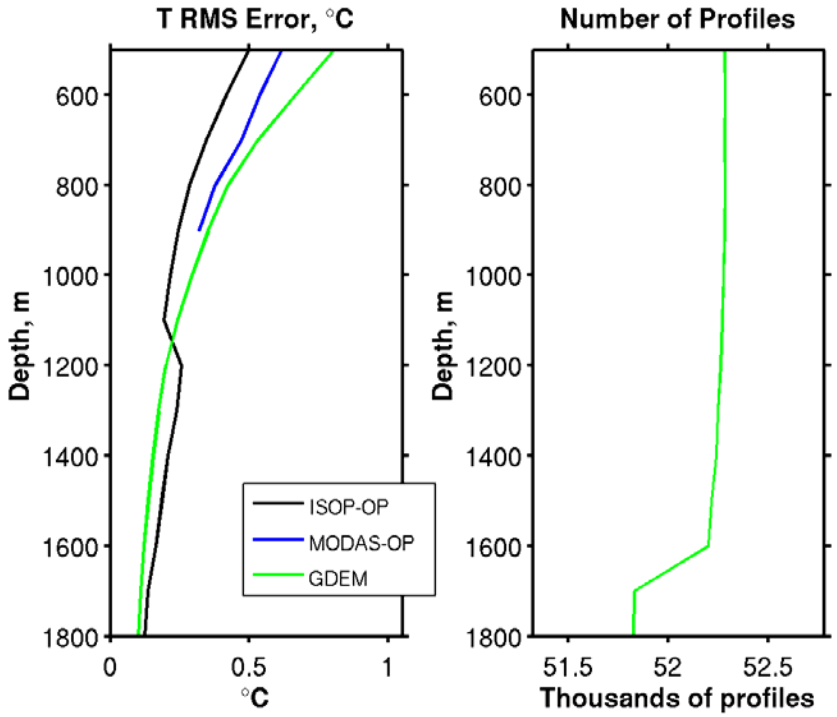


Figure 46. This figure is of RMS Error for temperature for $ISOP^{Oper}$, $MODAS^{oper}$, *GDEM* versus depth.

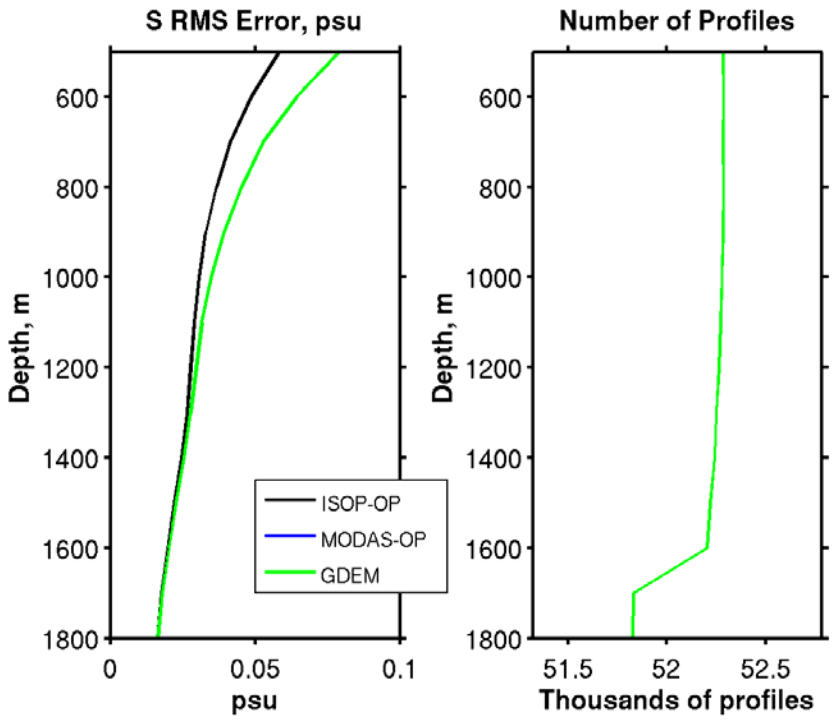


Figure 47. This figure is of RMS Error for salinity for $ISOP^{Oper}$, $MODAS^{oper}$, *GDEM* versus depth.

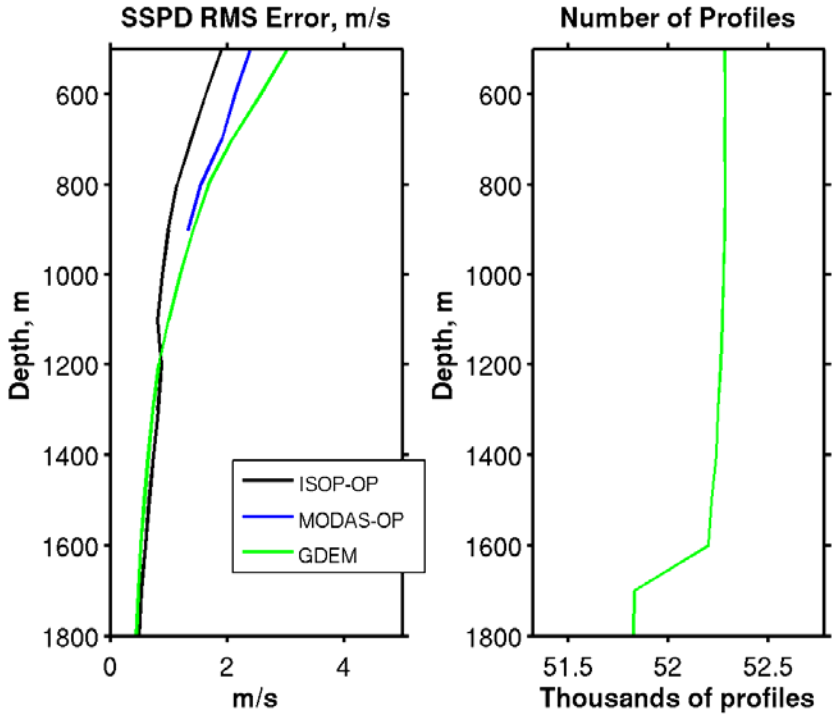


Figure 48. This figure is of RMS Error for sound speed for $ISOP^{Oper}$, $MODAS^{oper}$, *GDEM* versus depth.

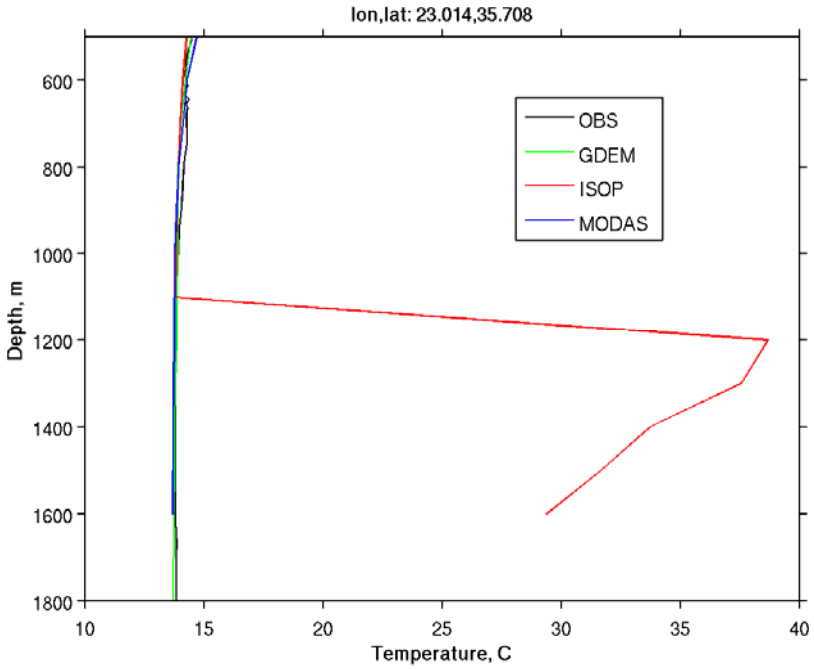


Figure 49. Bad ISOP synthetic from the Mediterranean Sea located at 35.7N, 23.0E.

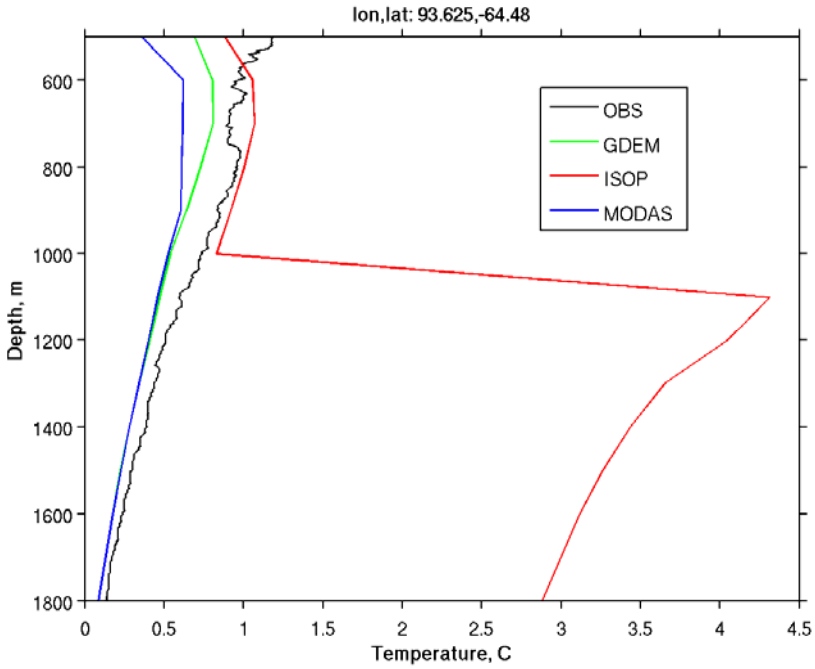


Figure 50. Bad ISOP synthetic near Antarctica located at 35.7N, 23.0E.

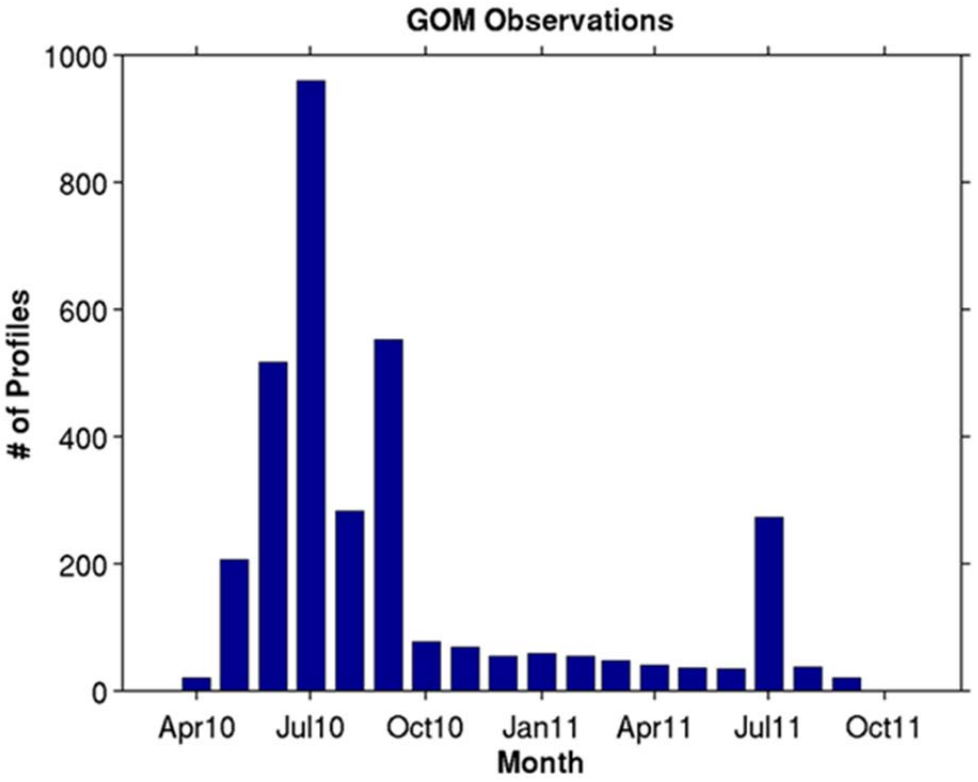


Figure 51. A histogram of the number of profiles used in the validation listed by month. A total of 4001 profiles were used.

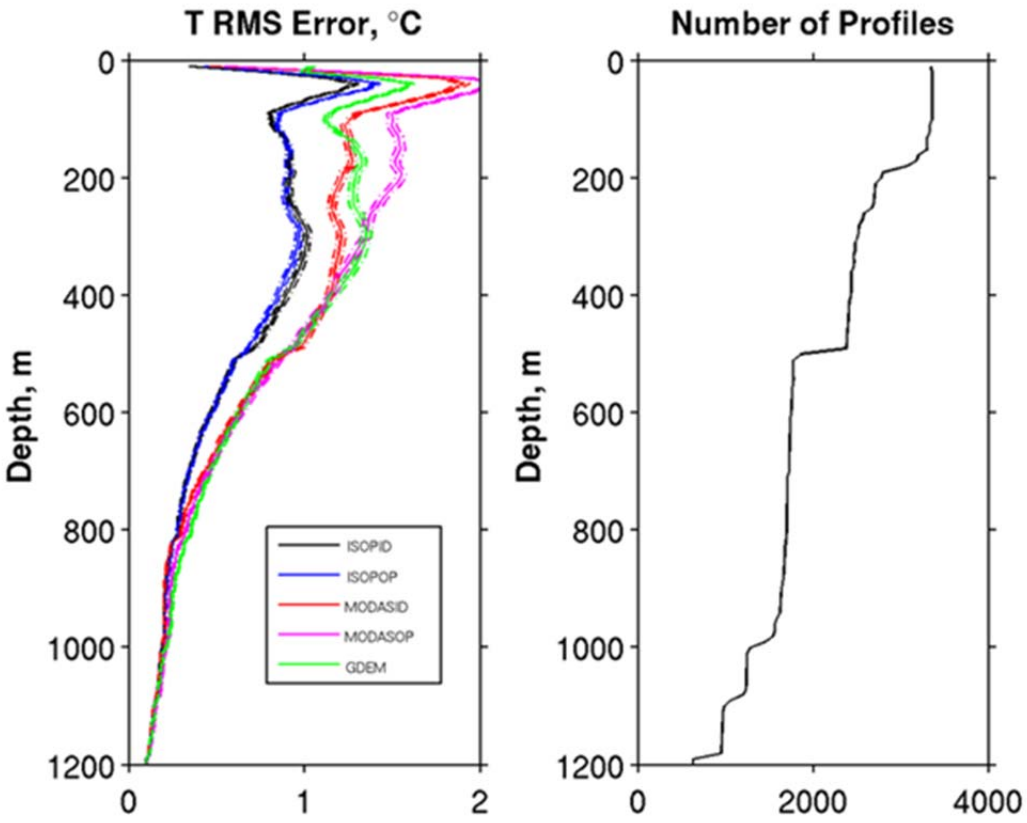
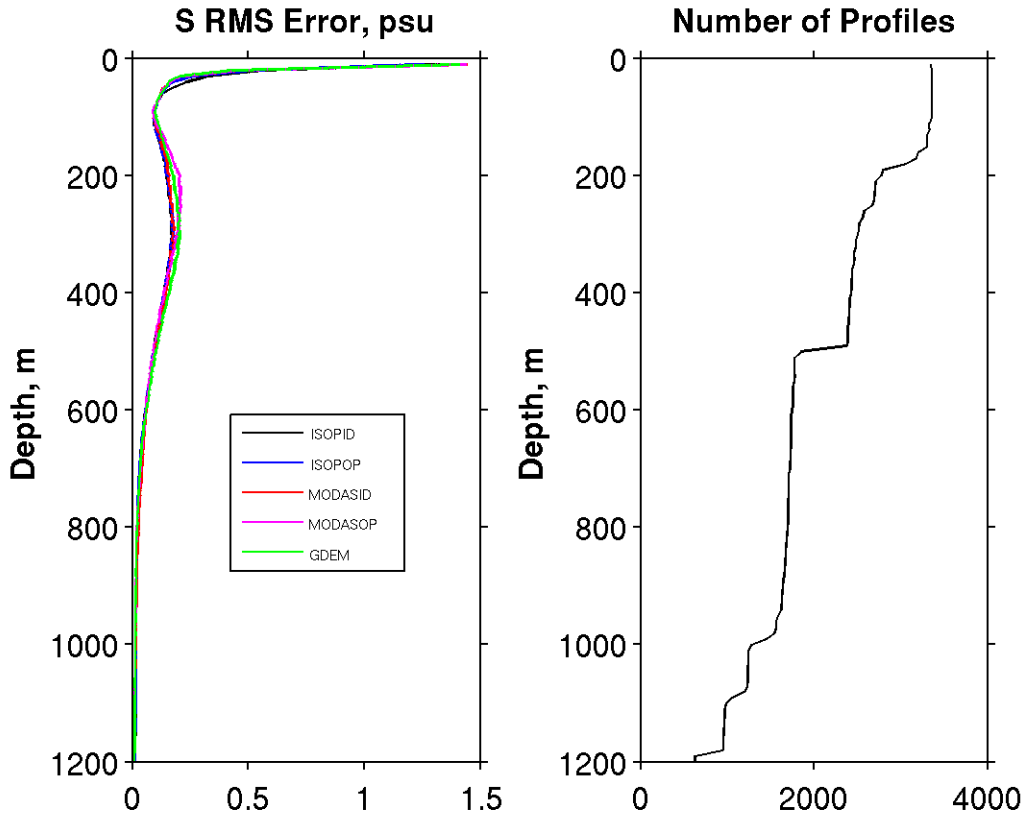


Figure 52. RMS Errors for temperature (left) of $ISOP^{Ideal}$, $ISOP^{Oper}$, $MODAS^{Ideal}$, $MODAS^{oper}$, and *GDEM* versus depth. The number of profiles used in the error analysis versus depth (right). Since not all profiles extend to 1200 m, the number of profiles in the analysis decreases with depth. Also, many profiles start below 10 m, reducing the number of profiles in the 0 m bin. Dashed lines represent the standard error limits for each case.



Depths.

Figure 53. RMS Errors for salinity (left) of $ISOP^{Ideal}$, $ISOP^{Oper}$, $MODAS^{Ideal}$, $MODAS^{oper}$, and *GDEM* versus depth. The number of profiles used in the error analysis versus depth (right). Dashed lines represent the standard error limits for each case.

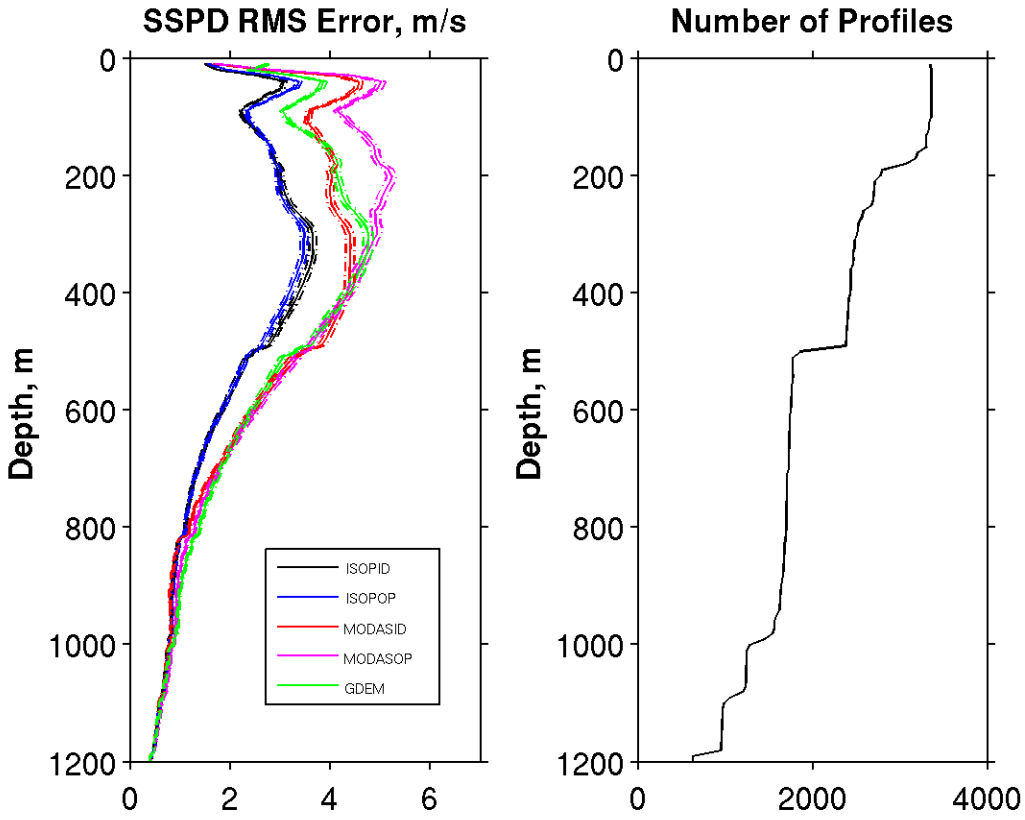


Figure 54. RMS Errors for sound speed (left) of $ISOP^{Ideal}$, $ISOP^{Oper}$, $MODAS^{Ideal}$, $MODAS^{oper}$, and *GDEM* versus depth. The number of profiles used in the error analysis versus depth (right). Dashed lines represent the standard error limits for each case.

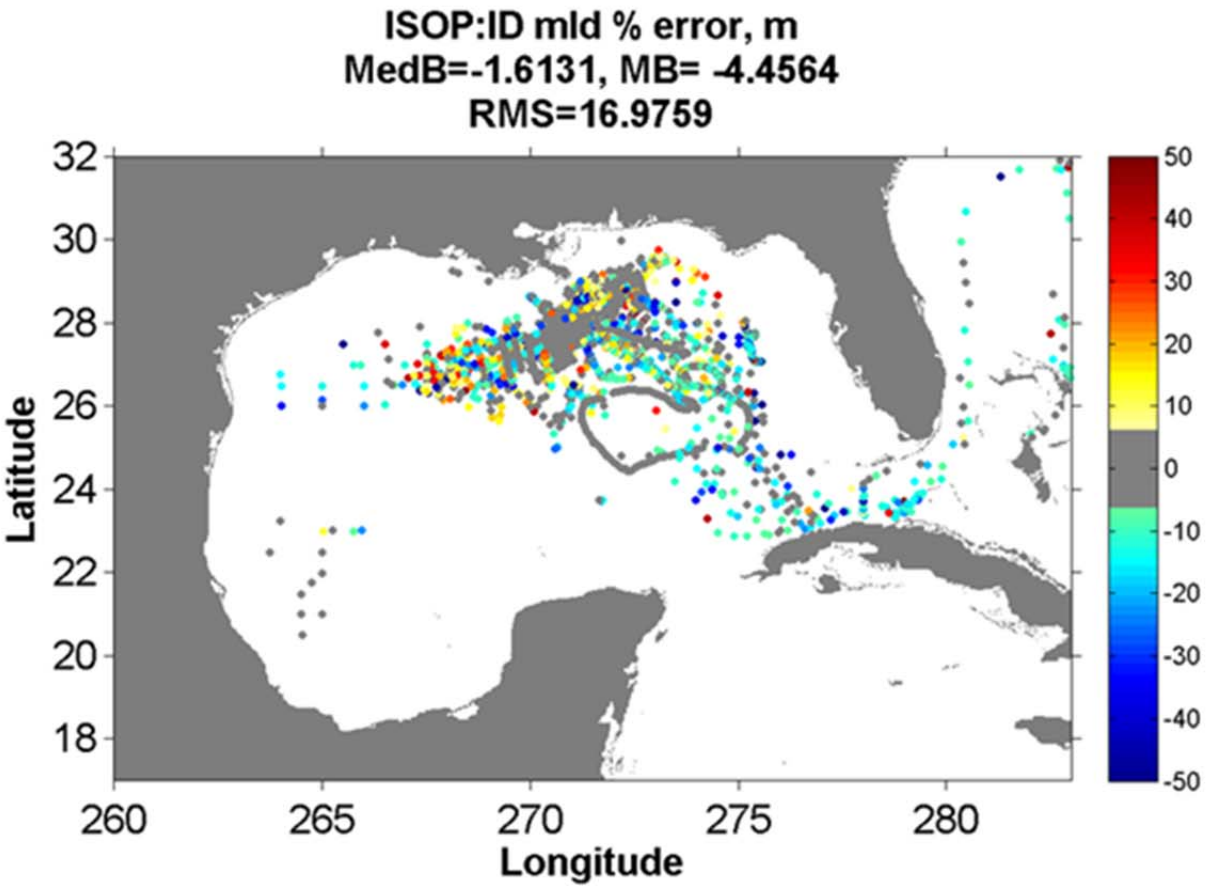


Figure 55. The mixed layer depth % misfit error estimates for $ISOP^{Ideal}$ the Gulf of Mexico. Units are in m.

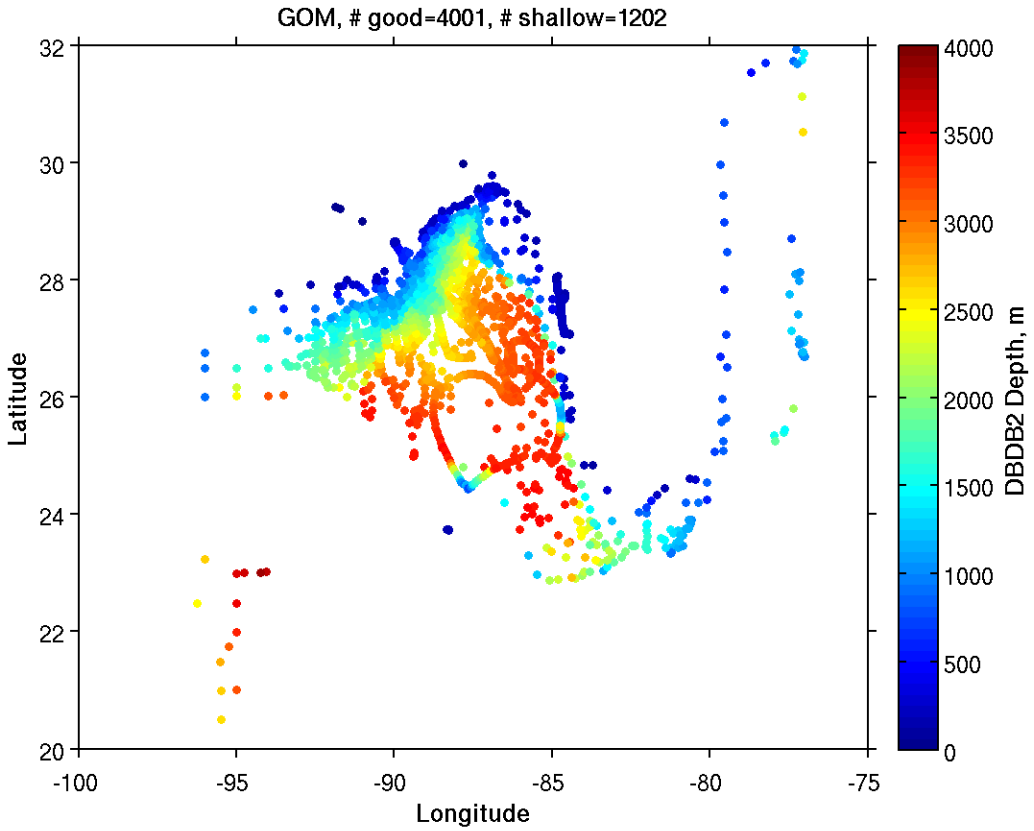


Figure 56. The location of all validation profiles color coded for the Navy's DBDB2 bathymetry depths at the same location.

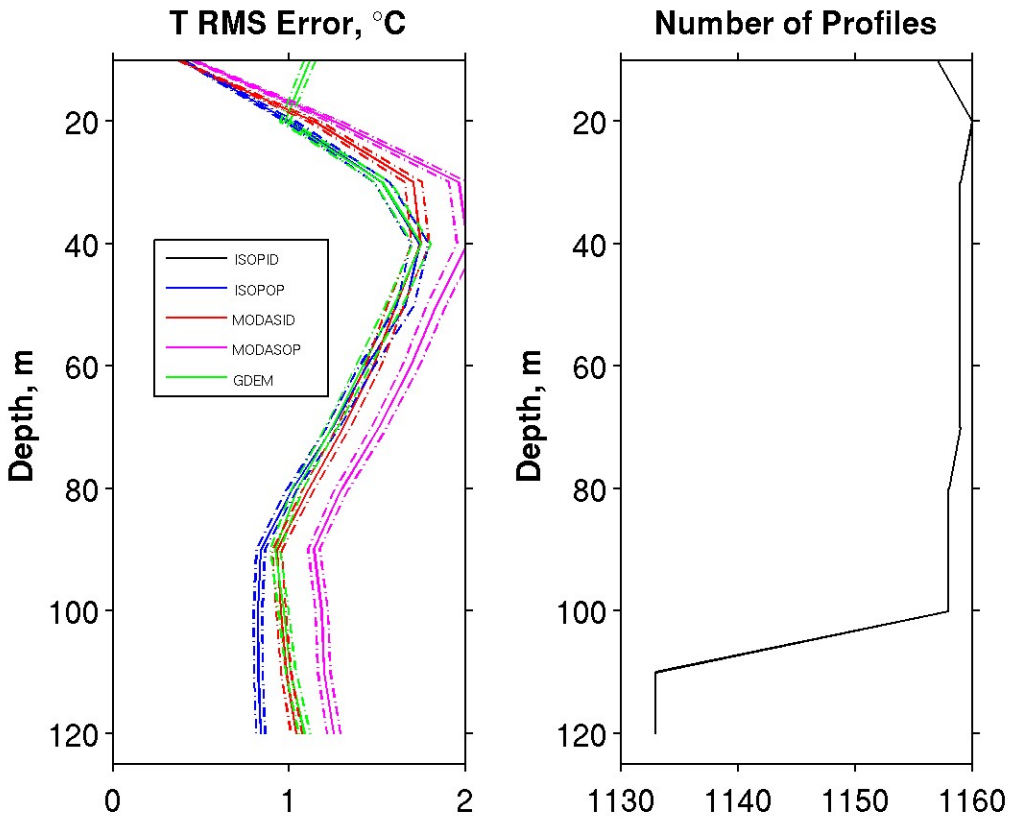


Figure 57. The RMS Errors for temperature (left) of $ISOP^{Ideal}$, $ISOP^{Oper}$, $MODAS^{Ideal}$, $MODAS^{oper}$, and *GDEM* versus depth for only the 1202 profiles in water shallower than 1000 m. Dashed lines represent the standard error limits for each case.

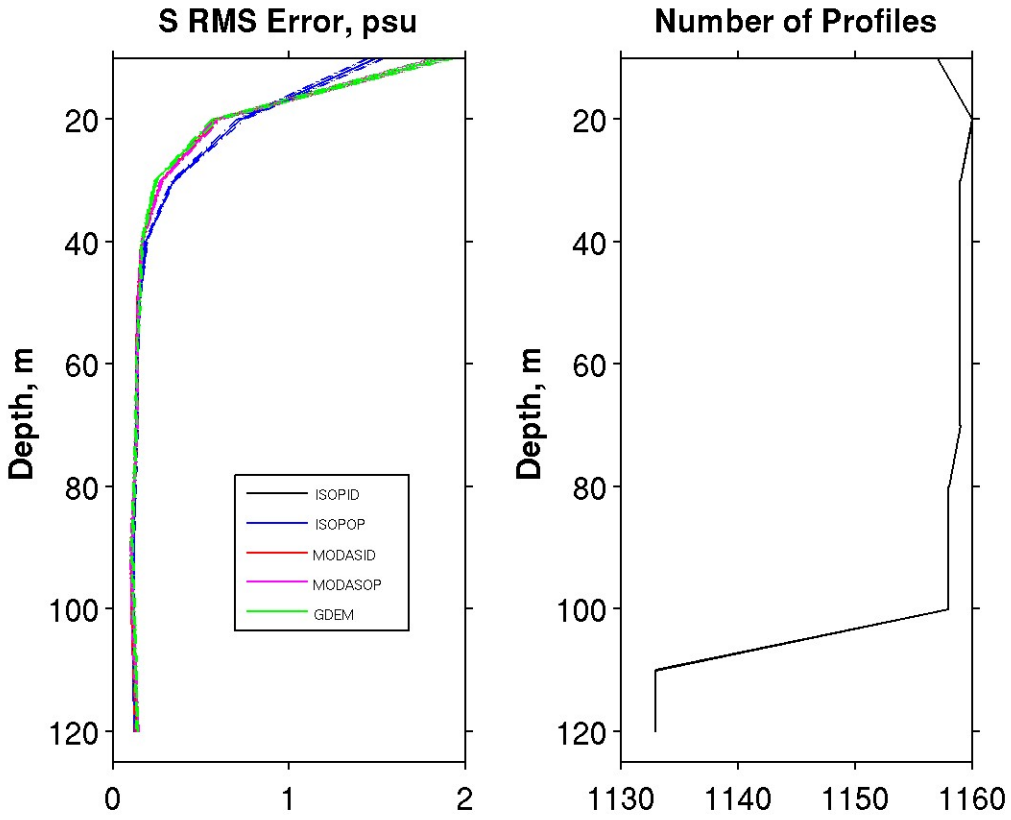


Figure 58. The RMS Errors for salinity (left) of $ISOP^{Ideal}$, $ISOP^{Oper}$, $MODAS^{Ideal}$, $MODAS^{oper}$, and *GDEM* versus depth for only the 1202 profiles in water shallower than 1000 m. Dashed lines represent the standard error limits for each case.

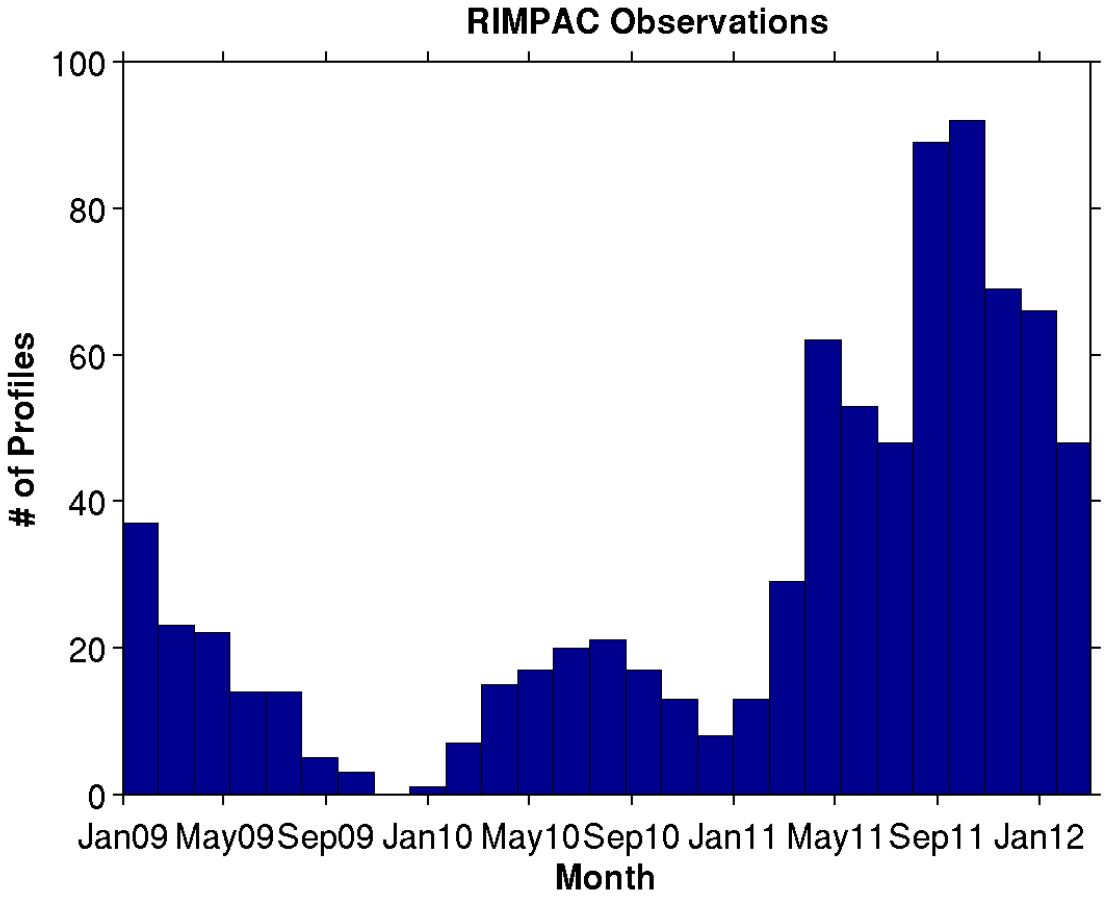


Figure 59. A histogram of the number of profiles in the RIMPAC region used in the validation listed by date. A total of 814 profiles were used.

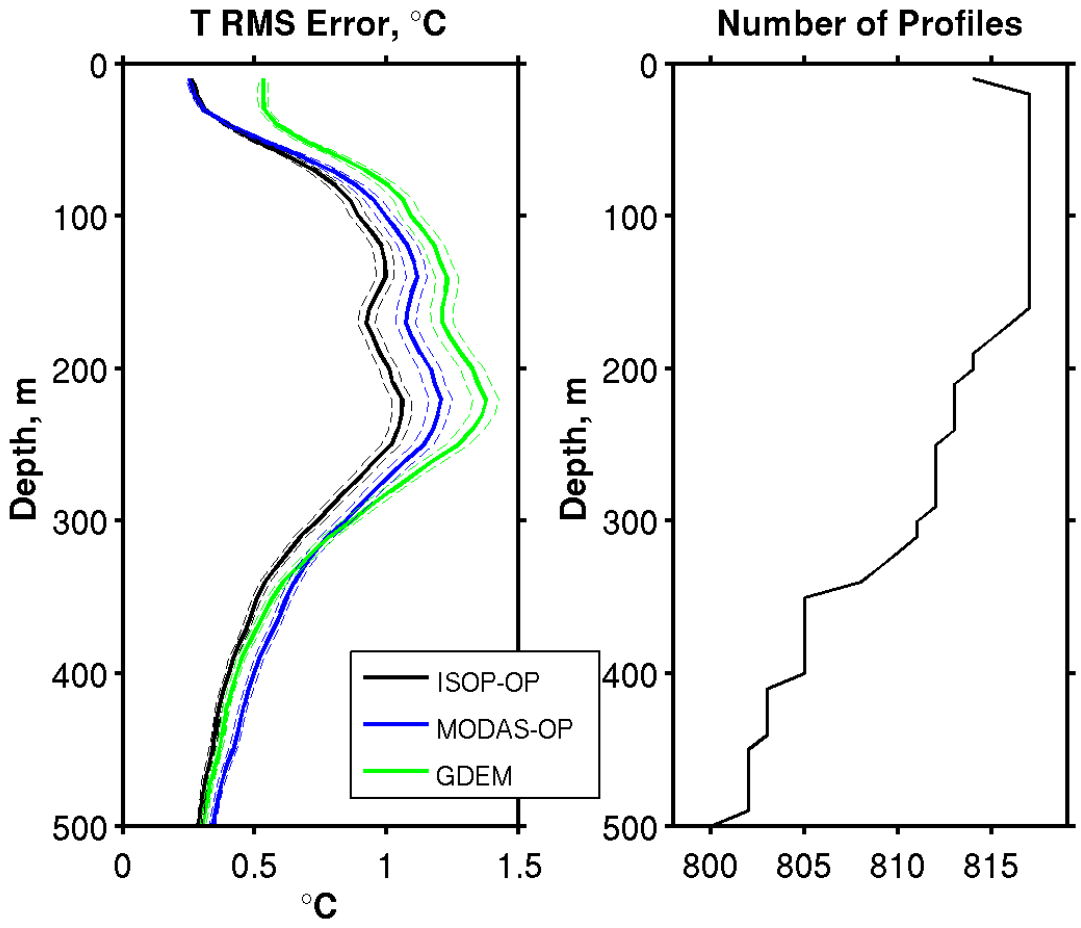


Figure 60. On the left we have temperature RMS Error versus depth. On the right we have the number of validation profiles that were used at each depth, which were computed every ten meters starting at depth zero. Since fewer profiles have an observation at depth zero, the number of profiles is lower at the surface. Dashed lines represent the standard error limits for each case.

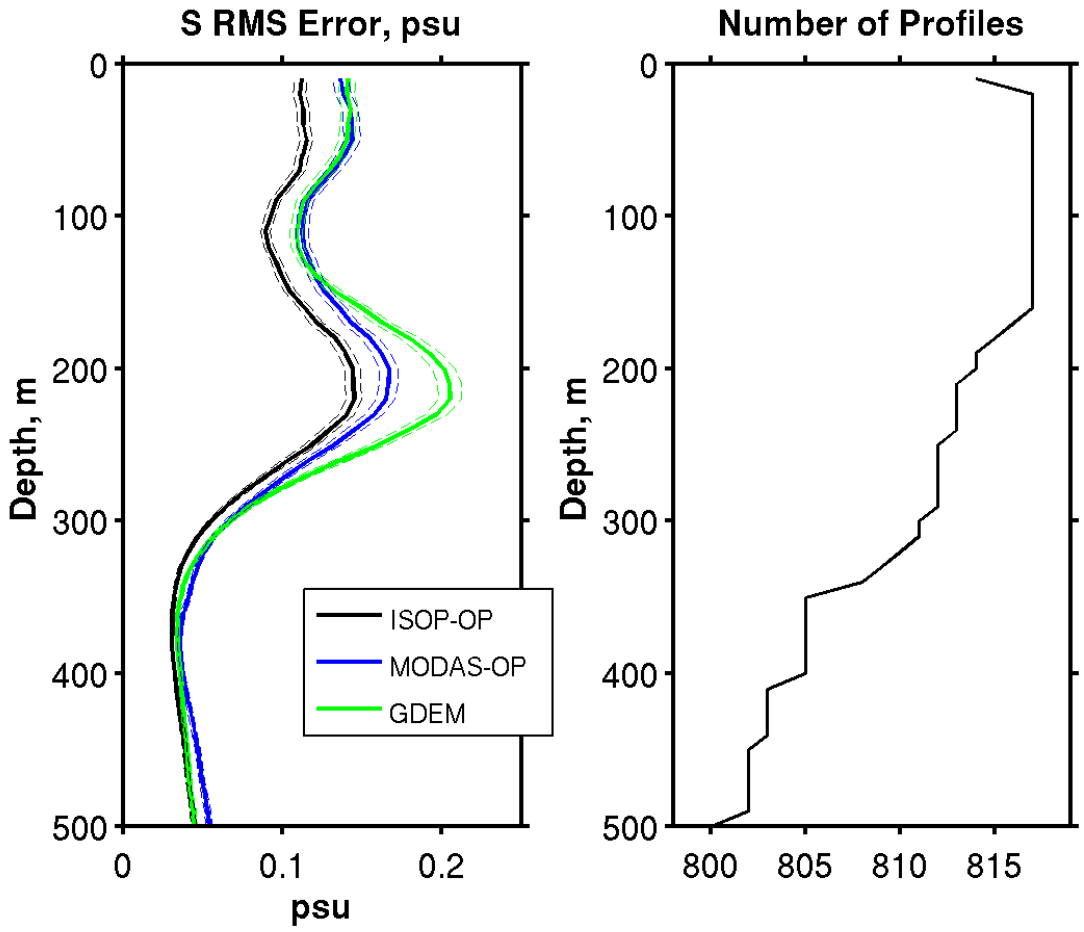


Figure 61. On the left we have salinity RMS Error versus depth. On the right we have the number of validation profiles that were used at each depth. Dashed lines represent the standard error limits for each case.

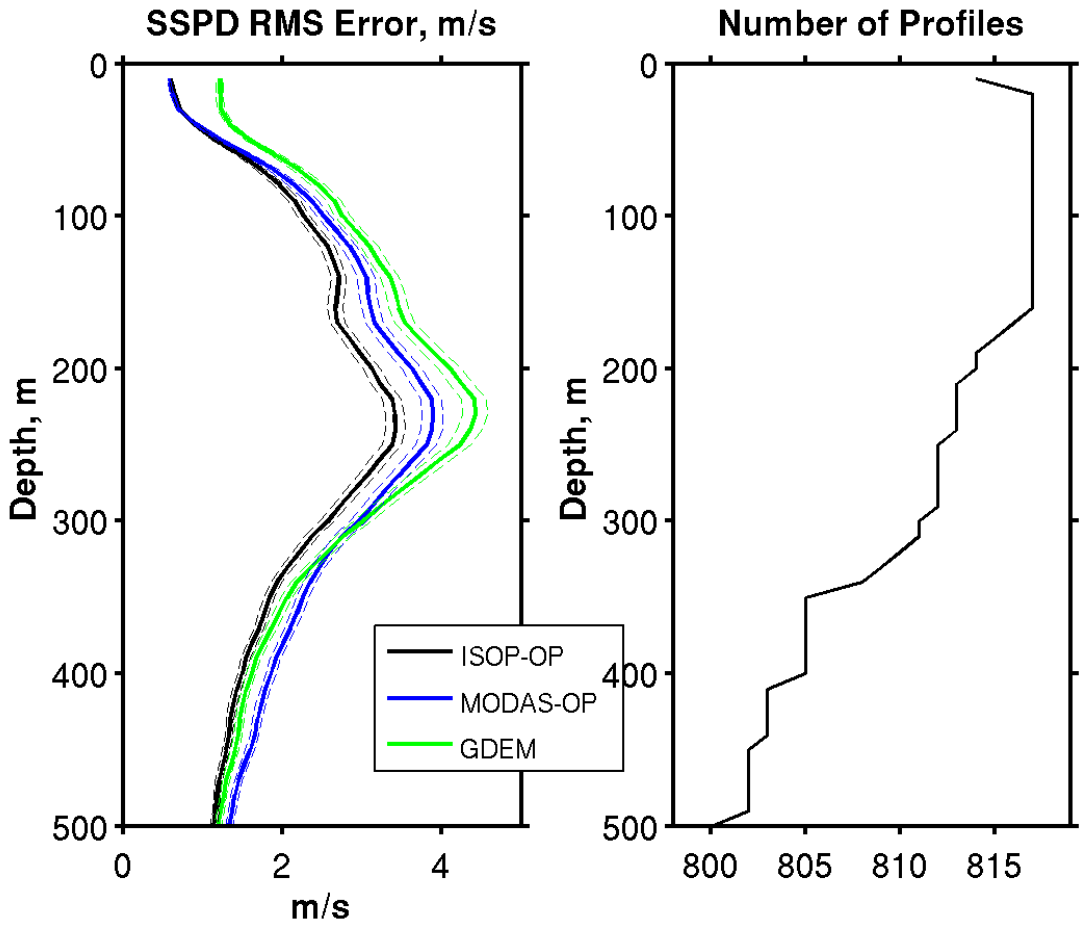


Figure 62. On the left we have sound speed RMS Error versus depth. On the right we have the number of validation profiles that were used at each depth. Dashed lines represent the standard error limits for each case.