Ocean Science

Argo data assimilation into HYCOM with an EnOI method in the Atlantic Ocean

D. Mignac[1,2]**, C. A. S. Tanajura**[2,3,4]**, A. N. Santana**[2]**, L. N. Lima**[2]**, and J. Xie**[5]

[1]Graduate Program in Geophysics, Physics Institute and Geosciences Institute, Federal University of Bahia, Salvador, Brazil
[2]Oceanographic Modeling and Observation Network, Center for Research in Geophysics and Geology,
Federal University of Bahia, Salvador, Brazil
[3]Physics Institute, Federal University of Bahia, Salvador, Brazil
[4]Ocean Sciences Department, University of California, Santa Cruz, CA, USA
[5]Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

Correspondence to: D. Mignac (davi.mignac@gmail.com)

Abstract. An ocean data assimilation system to assimilate Argo temperature (T) and salinity (S) profiles into the HYbrid Coordinate Ocean Model (HYCOM) was constructed, implemented and evaluated for the first time in the Atlantic Ocean (78° S to 50° N and 98° W to 20° E). The system is based on the ensemble optimal interpolation (EnOI) algorithm proposed by Xie and Zhu (2010), especially made to deal with the hybrid nature of the HYCOM vertical coordinate system with multiple steps. The Argo T–S profiles were projected to the model vertical space to create pseudo-observed layer thicknesses (Δp_{obs}), which correspond to the model target densities. The first step was to assimilate Δp_{obs} considering the sub-state vector composed by the model layer thickness (Δp) and the baroclinic velocity components. After that, T and S were assimilated separately. Finally, T was diagnosed below the mixed layer to preserve the density of the model isopycnal layers. Five experiments were performed from 1 January 2010 to 31 December 2012: a control run without assimilation, and four assimilation runs considering the different vertical localizations of T, S and Δp. The assimilation experiments were able to significantly improve the thermohaline structure produced by the control run. They reduced the root mean square deviation (RMSD) of T and S calculated with respect to Argo independent data in 34 and 44 %, respectively, in comparison to the control run. In some regions, such as the western North Atlantic, substantial corrections in the 20 °C isotherm depth and the upper ocean heat content towards climatological states were achieved.

The runs with a vertical localization of Δp showed positive impacts in the correction of the thermohaline structure and reduced the RMSD of T (S) from 0.993 °C (0.149 psu) to 0.905 °C (0.138 psu) for the whole domain with respect to the other assimilation runs.

1 Introduction

Many ocean circulation models include highly sophisticated numerical schemes and a large set of physical parameterizations. However, these are approximations of the governing equations and therefore they are sources of errors or discrepancies with respect to observations. Also, errors may be produced due to inaccuracies of the initial conditions, atmospheric forcing and lateral boundary conditions (Kalnay et al., 1996; Chassignet et al., 2009). For these reasons, data assimilation methods are important scientific tools in oceanography and other fields. They combine model outputs with observational data in a mathematically optimal or sub-optimal way and produce an objective analysis with smaller errors than the model output alone (Daley, 1991; Kalnay, 2003). The analysis is used as the model initial condition for weather and climate forecasts (Kalnay, 2003), and more recently for ocean weather forecasts under the framework of operational oceanography (Dombrowsky et al., 2009; Chassignet et al., 2009; Schiller and Brassington, 2011). Data assimilation methods are also applied to produce long-term series of

analyses for climate diagnostics studies and they contribute to a better understanding of the physical mechanisms that are responsible for the ocean variability. For example, the depth of the mixed layer and the heat content can be better represented by analyses than by model simulations without assimilation (Carton and Giese, 2008).

A major obstacle in ocean data assimilation is the relatively small volume of observed ocean data available for assimilation and validation. Most of the spatial and high-frequency temporal variability from the ocean surface is acquired by satellite measurements such as the sea surface height (SSH) and sea surface temperature (SST). However, these observations are available for only a few decades and they are insufficient to determine the sub-surface variability (Ezer and Mellor, 1994; Chassignet et al., 2006). Therefore, the implementation of the Argo network with more than 3300 profilers freely reporting temperature and salinity data down to 2000 m has transformed the in situ ocean observing system in the new millennium (Schiller and Brassington, 2011). For the first time it is possible to have continuous measurements of the temperature, salinity and velocity of the upper ocean which makes the Argo network indispensable for any global or regional data assimilation system (Chassignet et al., 2007; Oke et al., 2008; Xie and Zhu, 2010). For example, Oke and Schiller (2007) showed that the assimilation of SST and SSH should be complemented with Argo profiles since the assimilation of this component plays a crucial role in improving the model thermohaline state, especially for salinity.

Among all data assimilation methods, the ensemble-based methods use a set of different model states to estimate the model errors (Evensen, 2003; Oke et al., 2005). One widely used is the ensemble optimal interpolation (EnOI) (Oke et al., 2005) in which the ensemble members are derived from an existing model run. This reduces the computational cost of the assimilation and makes this method suitable for operational purposes. It is already verified that the EnOI is able to effectively constrain the model towards observations. For instance, it was successfully applied to assimilate Argo data in the Pacific Ocean (Xie and Zhu, 2010), sea level anomaly (SLA) data in the South China Sea (Xie et al., 2011) and in the Gulf of Mexico (Counillon and Bertino, 2009), and SLA, SST and Argo data in the Australian region (Oke et al., 2008) and in the Indo-Pacific Ocean (Yan et al., 2010).

Considering the importance of the model in the construction of the analysis, several state-of-the-art, publically available ocean circulation models should be evaluated in order to develop an operational ocean forecasting system with data assimilation. One choice is the HYbrid Coordinate Ocean Model (HYCOM). It is formulated in terms of target densities and employs a hybrid vertical coordinate system to combine the best features of each vertical coordinate in specific oceanic regions (Bleck, 2002). The model has fixed z-levels to better represent the mixed layer, isopycnal layers to discretize the deep stratified ocean and σ-levels to better reproduce the bathymetry in shallow areas. Because of the hybrid

nature of the model, the best way to assimilate vertical profile data into HYCOM is still an open question. A choice of the model prognostic variables should be made a priori since only two of the three state variables – temperature, salinity and potential density – are independent. Furthermore, the layer thickness, which is a key model variable, varies spatially and temporally according to the evolution of temperature, salinity and density.

Taking into account these characteristics, Thacker and Esenkov (2002; hereafter TE) proposed a method to assimilate expendable bathythermographic (XBT) data into HYCOM with a three-dimensional variational scheme (3-D Var). In this work, the XBT profiles are converted into observed layer thicknesses that respect the target densities of each model layer, and the temperature and salinity of the XBTs are projected to each observed layer thickness. Xie and Zhu (2010; hereafter XZ) used an EnOI scheme to assimilate Argo data, and showed that the TE approach produced significant improvements in relation to simpler schemes, in which the innovation is calculated in the observational space.

Very little has been published to evaluate the impact of in situ profile data assimilation into HYCOM with a focus on the Atlantic Ocean (e.g., Thacker et al., 2004; Belyaev et al., 2012). In the present work, a data assimilation system for HYCOM was constructed, implemented and realized for the Atlantic Ocean for the first time. The data assimilation algorithm follows the EnOI scheme suggested by XZ very closely. The present system was developed within the Brazilian Oceanographic Modeling and Observation Network (REMO) to be a component of an operational ocean forecasting system for the Atlantic Ocean (www.rederemo. org) (Tanajura and Belyaev, 2009; Lima et al., 2013; Tanajura et al., 2013). In this paper, the focus is on the impact of Argo data assimilation over the Atlantic Ocean and on a sensitivity study of the analysis run, considering different vertical localizations of the model error co-variance matrix involving temperature, salinity and especially the layer thickness. The REMO forecasting system uses a nested model approach based on HYCOM. The present work deals with the construction of large-scale analyses over almost the whole Atlantic Ocean. This domain was conceived and configured to provide reasonable boundary conditions to higher-resolution grids of greater interest to REMO over the South Atlantic. In the near future, the present data assimilation methodology will be used in the Atlantic domain, and in the higher-resolution grids over the Metarea V (from 35.5° S to 7° N, west of 20° W to Brazil) and sub-regions off the Brazilian coast of particular interest to the Brazilian Navy and the active petroleum industry located there.

This paper is organized as follows. In Sect. 2, HYCOM and the configuration used in this study are briefly described. In Sect. 3, the EnOI scheme to assimilate Argo data is presented. Section 4 shows the design of the assimilation experiments and Sect. 5 presents their results. Section 6 contains discussion and conclusions.

2 HYCOM and its configuration

HYCOM is a primitive equation general circulation model, which has evolved from the Miami Isopycnic Coordinate Ocean Model (MICOM) (Bleck and Smith, 1990). The main advantage of the isopycnic coordinate is its ability to maintain the properties of water masses which do not communicate directly with the mixed layer. In HYCOM, with the advection of layer thicknesses by the continuity equation, the isopycnal coordinates smoothly transform into the z-coordinate in the weakly stratified upper ocean mixed layer and into the terrain-following σ-coordinate in the shallow water regions (Bleck, 2002; Chassignet et al., 2007). The freedom to adjust the vertical spacing of the coordinate surfaces in HYCOM simplifies the numerical implementation of several physical processes, such as turbulence in the mixed layer, detrainment and convective adjustment. Also, the capability of assigning additional coordinate surfaces to the oceanic mixed layer in HYCOM allows the option of implementing sophisticated vertical mixing turbulence closure schemes (Halliwell, 2004). Hence, HYCOM is considered to be a suitable model for operational ocean forecasting systems and climate studies (Chassignet et al., 2007, 2009). In this work, the version 2.2.14 of HYCOM was used.

The model grid in the present configuration has 760×480 horizontal grid points, with a spatial resolution of $0.25°$, which remains constant in longitude, but varies in latitude attaining higher resolution towards the poles. The computational model domain covered almost all the Atlantic Ocean from $78°$S to $50°$N and from $100°$W to $20°$E, excluding the Pacific Ocean, the Mediterranean Sea and the North Atlantic subpolar region. The choice for the northern limit at $50°$N was based on two facts. First, the surface currents are primarily zonal at this latitude (Gabioux et al., 2013). Second, the purpose of this grid is to provide reasonable boundary conditions to higher-resolution grids that will soon be configured along the Brazilian coast, which is the area of main interest to REMO and far away from the northern boundary. On the lateral boundaries, relaxation to monthly climatological temperature and salinity from Levitus (1982) was applied considering the outermost 10 grid cells and the timescale of 30 days. This approach attempts to preserve climatological shear through geostrophic adjustment and has been successfully used in previous works (e.g., Paiva and Chassignet, 2001; Gabioux et al., 2013). Constant barotropic volume fluxes were imposed: zero flux in the north, eastward flux of 110 Sv in the Drake Passage, westward flux of 10 Sv in 12 grid points adjacent to South Africa along $20°$E, and eastward flux of 120 Sv further south to Antarctica.

The model was configured to use surface pressure as reference for potential density, aiming for improved representation of near-surface fields, at the cost of not representing accurately the Antarctic Bottom Water (Chassignet et al., 2003). The vertical domain was discretized in 21 vertical layers. The chosen target potential densities were 19.50, 20.25, 21.00, 21.75, 22.50, 23.25, 24.00, 24.70, 25.28, 25.77, 26.18, 26.52, 26.80, 27.03, 27.22, 27.38, 27.52, 27.64, 27.74, 27.82 and 27.88. To obtain the volumetric density in kg m^{-3}, 1000 should be added to each target density. The first layers have a few light target density values that ensure a minimum of three fixed-depth layers near the surface of the ocean.

The vertical mixing scheme is the K-profile parameterization (KPP) (Large et al., 1994). The model bathymetry was interpolated from the Earth Topography 1 (ETOPO1) with 1 min resolution. The model was initialized from the state of rest with climatological thermohaline structure, and a 30-year spin-up was performed using monthly climatological forcing fields from the Comprehensive Ocean and Atmosphere data set (COADS) (Woodruff et al., 1987). Then, from January 1995 to December 2009, the model was forced on the ocean surface with 6-hourly atmospheric reanalysis version 1 by the National Centers for Environmental Prediction/National Centers for Atmospheric Research (NCEP/NCAR) (Kalnay et al., 1996), including precipitation, wind speed at 10 m, short- and long-wave radiation fluxes at the surface, air temperature and humidity at 2 m. The result of this simulation was used as the initial condition for the Argo data assimilation experiments starting on 1 January 2010.

Figure 1 shows the mean state of temperature and salinity simulated by HYCOM from 1 January 1997 to 31 December 2008 and its comparison with the World Ocean Atlas 2013 Climatology (WOA13) (Boyer and Mishonov, 2013) along $25°$W for the upper 1000 m. WOA13 climatology shown in Fig. 1 spans the period from 1995 to 2012. In general, the pattern of the simulated temperature and salinity is similar to the WOA13 climatology, particularly in the South Atlantic, which is the main target area for REMO. In the North Atlantic large differences are seen between 25 and $50°$N below 400 m. The Mediterranean water (MW) is more saline, warmer and found further north in comparison with WOA13. The Mediterranean Sea is not simulated in this grid and is imposed as a relaxation towards monthly climatological temperature and salinity at the boundary without mass flux. In addition, the simulated temperature is higher than WOA13 in the upper 300 m of the equatorial region, while the values of high-salinity cores in the subtropical gyres are smaller than the values found in the climatology. It is worth noting that this resolution of $1/4°$ is not enough for HYCOM to properly solve the Gulf Stream and its associated dynamics (Hulburt and Hogan, 2000), which may also contribute to some of the model discrepancies in comparison to WOA13 in the mid-latitudes of the North Atlantic. It is expected that assimilation of Argo data will improve the model state and reduce the existing differences in the thermohaline structure with respect to the WOA13 climatology.
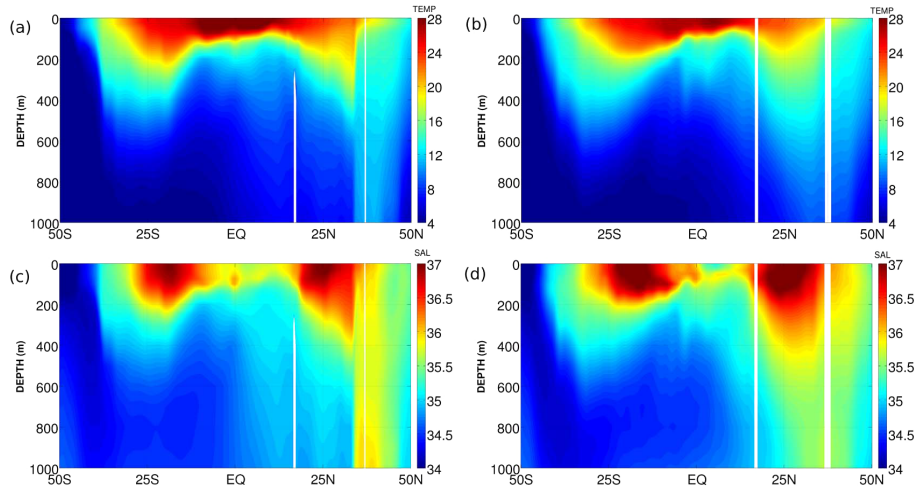
Figure 1. Mean temperature (°C) and salinity (psu) in the upper 1000 m along 25° W for **(a)** averaged simulated temperature from January 1997 to December 2008, **(b)** climatological temperature from WOA13, **(c)** averaged simulated salinity from January 1997 to December 2008, and **(d)** climatological salinity from WOA13.

3   The data assimilation scheme

The analysis (X^a) according to the EnOI scheme is given by the formula (Evensen, 2003)

$$X^a = X^b + \mathbf{K}(Y - \mathbf{H}X^b),\qquad(1)$$

where $X^b \in \mathbb{R}^N$ is the model background state or the prior, $\mathbf{K}$ is the gain matrix, Y is the vector of observations, $Y \in \mathbb{R}^{N_{\mathrm{OBS}}}$, and $\mathbf{H}X^b$ is the projection of the prior onto the observational space by the observational operator $\mathbf{H}$. The term $(Y - \mathbf{H}X^b)$ is called the innovation vector and the term $\mathbf{K}(Y - \mathbf{H}X^b)$ is the analysis increment. The gain matrix $\mathbf{K}$ is calculated from the equation

$$\mathbf{K} = \alpha(\boldsymbol{\sigma} \circ \mathbf{B})\mathbf{H}^T[\alpha\mathbf{H}(\boldsymbol{\sigma} \circ \mathbf{B})\mathbf{H}^T + \mathbf{R}]^{-1},\qquad(2)$$

where $\mathbf{B}$ is the co-variance matrix of the model errors and $\mathbf{R}$ is the diagonal variance matrix of the observational error. The term $\alpha \in (0, 1]$ is a scalar that can tune the magnitude of the analysis increment and $\boldsymbol{\sigma}$ denotes the localization operator applied over $\mathbf{B}$ by a Schur product represented by the symbol $\circ$. Equation (2) is used by many data assimilation schemes such as the optimal interpolation, the ensemble Kalman filter (EnKF) or the EnOI. In the EnOI scheme, $\mathbf{B}$ is estimated from the equation

$$\mathbf{B} = \frac{A'A'^T}{(M-1)},\qquad(3)$$

where $A' = [A'^1 A'^2, \ldots, A'^M]$, $A'^k = (X^k - \frac{1}{M}\sum_{m=1}^M X^m)$, $X^k \in \mathbb{R}^N$ is the model state vector of the kth ensemble member with k varying from 1 to M, and $M = 132$ is the number of ensemble members used in all assimilation steps in this study. This ensemble of model anomalies can be taken from

a long-term model run (Evensen, 2003) or a spin-up run (Oke et al., 2008) in order to capture the model variability at certain scales. Thus, even being stationary in time, this ensemble of model anomalies allows for the description of the spatial correlations and the anisotropic nature of ocean circulation, keeping the analysis dynamically consistent and substantially reducing the computational cost. Details on how the $\mathbf{B}$ matrix was calculated here regarding the high-frequency variability of the model are described below.

3.1   Calculation of the innovation vector

Basically, there are two ways to calculate the innovation vector. The first one projects the model state vector into the observational space. In this case, the temperature, salinity and layer thicknesses of HYCOM are projected into the vertical levels of the Argo profiles, which provide almost vertically continuous measurements of temperature and salinity, ranging from near surface to 2000 m. This procedure makes the $\mathbf{H}$ operator become complex and non-linear. Since Eq. (1) is linear, the use of a non-linear operator may cause problems in the linear analysis update and may contribute to a sub-optimal assimilation performance (Xie and Zhu, 2010). A second way to calculate the innovation follows the strategy adopted by TE and by XZ for HYCOM. In this approach, temperature (T) and salinity (S) data profiles are projected into the model vertical space to create pseudo-observed layer thicknesses (Δp_{obs}). This is done following the hybrid nature of the model's layers: each layer is required to have a minimum thickness and, after that requirement is satisfied, it should be as close as possible to its specified target value of potential density. Thus, each Argo profile is processed as follows. Based on a pair of profiles of potential temperature and salinity, the profile of potential density can be calculated
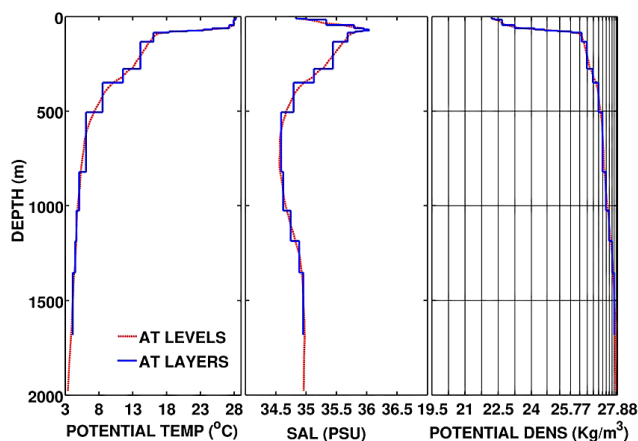
Figure 2. Profile of potential temperature (°C), salinity (psu) and potential density (kg m^{-3}) from an Argo float located at 4.04° N and 23° W on 1 January 2010 plotted against its approximation for HYCOM as layer averages. The model target density is indicated by the solid black vertical lines.

by an equation of state for seawater (Brydon et al., 1999). The estimated surface density from the Argo profile is compared to the top layer target density to decide whether any sufficiently low-density water was observed. If not, the minimum thickness is assigned to the layer and the procedure is repeated for the layer below. Once water with the target density is encountered, the remainder of the potential density profile can be partitioned and the layer averages will correspond to the model target densities down to the maximum depth of the Argo profile.

Figure 2 shows the vertical profiles of potential temperature, salinity and potential density at z-levels and the new synthetic observations defined at model layers for an Argo float located at 4.04° N and 23° W on 1 January 2010. In Fig. 2, each Δp_{obs} respects the target densities of the model as soon as the first target density is found in the potential density profile of the Argo data. Also, the averages for all the observational variables are computed for each layer, as shown by the discretized profiles of potential temperature and salinity in the model layers. The step functions of T, S and Δp_{obs} are the data that will be actually assimilated by the EnOI scheme.

3.2 The modified EnOI to assimilate profile data into HYCOM

After the observation is defined in the HYCOM layers, the observed layer thicknesses are assimilated in a first step by Eqs. (1) and (2), and the analysis update is carried out for the model control state vector $(\Delta p_i, U_i, V_i)$; $i = 1, \ldots, nz$, where Δp_i, U_i and V_i are the layer thicknesses and the baroclinic velocity components, respectively, defined at the nz model layers for a single time step. To avoid the analyzed layer thicknesses occasionally becoming negative, a compu-

tationally efficient scheme based on Sakov et al. (2012) is used. If the thickness of a layer becomes negative, it is reset to zero and the thickness deficit is added to the neighboring layers. The layers are traversed twice, once from top to bottom, and a second time from bottom to top. Finally, the sum of the layer thicknesses should be equal to the initial bottom pressure (or local depth).

In the next step, temperature and salinity are assimilated separately and in a univariate way also according to Eqs. (1) and (2), but now with the previously adjusted layer thicknesses. Finally, T or S is diagnosed below the mixed layer by the seawater equation of state. According to TE, "within the context of HYCOM, when correcting temperature, it is necessary to decide whether to move interfaces, keeping potential densities of the layers unchanged, or to correct the densities, leaving the interfaces unchanged". Therefore, when the assimilation of layer thicknesses is performed, the potential density should be kept constant in the ocean simulated with isopycnal coordinates. Considering that most of the T corrections in the experiments of XZ were due to changes in the layer thicknesses by the assimilation of Δp_{obs}, T was chosen to be diagnosed below the mixed layer in the present work, instead of S.

3.3 Generation of a running ensemble

Many authors have shown how sensitive the EnOI and EnKF schemes are to the ensemble size (Mitchell et al., 2002; Evensen, 2003; Oke et al., 2007). In the EnOI scheme, the propagation of the observational information is highly dependent on the size and the quality of the ensemble, because the final analysis can be regarded as a combination of the ensemble anomalies whose relative weight is determined by the co-variances. In this work, 132 ensemble members were used. They were selected from the model free run for each assimilation day, regarding the intra-seasonal variability and the high-frequency model dynamics. This number of ensemble members was chosen after a few sensitivity experiments considering a reasonable representation of the model's anomalies without high computational cost, and is in agreement with the numbers used in recent works (e.g., Counillon and Bertino, 2009; Xie and Zhu, 2010; Xie et al., 2011).

The long-term model run that was used to select the ensemble members corresponded to the 12-year period from 1 January 1997 to 31 December 2008. For each assimilation step, a different model co-variance matrix was calculated. Considering the assimilation day, 11 ensemble members for each year of the 12-year period were selected around the date of the corresponding assimilation day. For instance, to perform assimilation on 15 March 2010, 11 members centered on 15 March of each year from 1997 to 2008 were taken 8 days apart, which gives a time window of 80 days for each year. However, the computational code developed here is flexible to use another number of ensemble members and to select different intervals between each ensemble member.

3.4 Localization

The localization technique is a feasible solution to reduce the effect of the sampling error in the ensemble-based methods, especially when the ensemble size is small (Hamill et al., 2001; Oke et al., 2007). The significance range of a measurement is a critical question in assimilation. In the present case, it should be unreasonable that a measurement in the Gulf Stream contributes to resolving the mesoscale features of the circulation of the Brazil Current. Therefore, the localization aims to delete long-distance correlations that may appear in the gain matrix and to limit the influence of a single observation by the Kalman update equation within a fixed region around the observation location. However, a drawback of the localization is that it can breakdown the geostrophic balance. Oke et al. (2007) showed that localization conserves the geostrophic balance when the radius of the localization is equal to or larger than the radius of decorrelation, which is the scale in which the correlations become negligible.

For many EnKF and EnOI schemes, localization is only applied in the horizontal direction (Oke et al., 2008; Sakov et al., 2012). However, some works have already investigated the vertical localization and its impact on the analysis, such as in XZ. They presented a vertical localization scheme for T and S when assimilating Argo profiles into HYCOM. Here, localization in the vertical direction is also investigated, but differently from XZ, as the focus is on Δp.

The operator $\sigma \circ$ in Eq. (2) defines the implementation of the localization by a Schur product, i.e., a product between elements with the same index in the arrays. The notation $\sigma \circ \mathbf{B}$ denotes the Schur product of a correlation matrix σ with the $\mathbf{B}$ matrix, and this approach is used in many works (Oke et al., 2007, 2008; Xie and Zhu, 2010). Here, the localization operator is separated into a horizontal component (σ_h) and a vertical component (σ_v), and it is defined as $\sigma = \sigma_\mathrm{h}\sigma_\mathrm{v}$.

3.4.1 Horizontal localization

In order to define the horizontal correlation matrix σ_h, a fifth-order function is used as in Gaspari and Cohn (1999):

$$\sigma_h\left(I_{ij}, L\right) = \tag{4}$$

$$\begin{cases} -\frac{1}{4}\left(\frac{I_{ij}}{L}\right)^5 + \frac{1}{2}\left(\frac{I_{ij}}{L}\right)^4 + \frac{5}{8}\left(\frac{I_{ij}}{L}\right)^3 - \frac{5}{3}\left(\frac{I_{ij}}{L}\right)^2 + 1, & 0 \le I_{ij} \le L \\ \frac{1}{12}\left(\frac{I_{ij}}{L}\right)^5 - \frac{1}{2}\left(\frac{I_{ij}}{L}\right)^4 + \frac{5}{8}\left(\frac{I_{ij}}{L}\right)^3 + \frac{5}{3}\left(\frac{I_{ij}}{L}\right)^2 - 5\left(\frac{I_{ij}}{L}\right) + 4 - \frac{2}{3}\left(\frac{I_{ij}}{L}\right)^{-1}, & L < I_{ij} \le 2L \\ 0, & I_{ij} > 2L \end{cases}$$

In this function, I_{ij} is defined as the Euclidean distance between any two arbitrary points in the horizontal space and L is the horizontal scale of influence defined as 150 km for all the assimilated variables. It is similar to a Gaussian function in physical space but more compact. The correlation function σ_h forces the model error co-variance matrix B to decrease to zero when I_{ij} reaches 300 km. Thus, the radius of localization is defined as 300 km.

The horizontal scale of influence was chosen to be 150 km because this number roughly matches the scales of boundary currents (Ezer and Mellor, 1994; Carton and Giese, 2008). Since the present work was the first REMO effort to implement an EnOI scheme to assimilate Argo data into HYCOM, the horizontal scale of influence was fixed for the entire model domain and for all depths. In this case, the choice of 150 km was conservative in order to avoid capturing co-variances that could damage the analysis in high-variability regions, such as the Gulf Stream and the Brazil–Malvinas Confluence.

3.4.2 Vertical localization

Concerning the vertical localization, XZ found considerable off-diagonal correlations for temperature and salinity between different HYCOM layers. However, there was no significant impact in the analysis when those co-variances were filtered and no significant differences in the experiments with and without vertical localization were reported. However, in the present work the vertical localization of co-variances between the layer thicknesses was investigated. They showed considerable off-diagonal correlations of above 0.4 or below -0.4 from the seventh layer to the bottom (Fig. 3). In the formulation of the vertical localization operator, the vertical distance between the layers was measured by the water column stratification, rather than by the Euclidian distance. Thus, in order to define the correlation matrix σ_v, the following function was used:

$$\sigma_{\mathrm{v}(i,j)} = \exp\left[-(\Delta\rho_{(i,j)}/L_\rho)^2\right], \tag{5}$$

where $\Delta\rho_{(i,j)}$ is the density difference between the HYCOM layers i and j, and L_ρ is a vertical-scale factor defined as $0.5\,\mathrm{kg\,m}^{-3}$ according to XZ. As shown by Fig. 3, when applying the vertical localization, the elements a few entries away from the diagonal in the co-variance matrix are almost canceled, but the correlations between the adjacent layers remain.

3.5 Observational errors

Since the observations that are actually assimilated are defined in the model vertical space, the observational errors of T and S in the model layers are calculated as a function of the depth D in meters, respectively, as in XZ:

$$\mathrm{SD_T}(D) = 0.05 + 0.45\exp(-0.002D), \tag{6}$$

$$\mathrm{SD_S}(D) = 0.02 + 0.10\exp(-0.008D). \tag{7}$$

The standard deviations of the observational errors of T vary in the vertical from $0.5\,°\mathrm{C}$ on the surface to $0.05\,°\mathrm{C}$ in the deep ocean. For S, they vary from 0.12 to 0.02 psu. According to Eqs. (6) and (7), the observational errors are assumed Gaussian with zero mean and uncorrelated.

In case of layer thickness, according to TE, the standard deviation of Δp_obs is calculated depending on the oceanic region. For example, in the mixed layer, the layer thickness
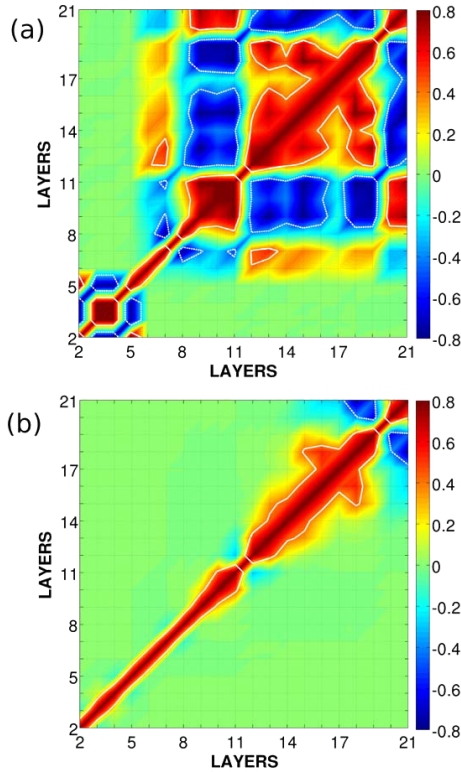
Figure 3. Vertical correlations of layer thickness between different layers at the location of an Argo float (69.95° W and 30.14° N) in the model domain: **(a)** calculated directly from the model ensembles ($M = 132$) and **(b)** after applying the vertical localization of layer thickness according to Eq. (5) and L_ρ defined as $0.5 \, \text{kg} \, \text{m}^{-3}$. The white lines denote correlations above 0.4 and below −0.4.

is assigned to the minimum layer thickness allowed by the model configuration, and the standard deviation is defined as $0.05\Delta p_k$, where Δp_k represents the layer thickness at the kth layer calculated from the observed profiles. In the isopycnal layers, the standard deviation is defined by the formula

$$\text{SD}(\Delta p_k) = \max \tag{8}$$
$$\left\{ 0.5\delta p_k, \max \left[0.05\Delta p_k, \Delta p_k \left(0.05 + (0.5 - 0.05) \frac{\text{sd}_{\sigma(k)}}{\text{SD}_{\sigma(k)}} \right) \right] \right\},$$

where δp_k is the minimum layer thickness specified by the model configuration for the kth layer, $\text{sd}_{\sigma(k)}$ is the minimum standard deviation of the potential density defined as $0.001 \, \text{kg} \, \text{m}^{-3}$, and $\text{SD}_{\sigma(k)}$ is the standard deviation of the potential density from observations. The latter should be small when the potential density from the observed layer thickness has values close to its target density.

4 Assimilation experiments

4.1 Argo data and quality control

The Argo data employed in the assimilation were collected from a global data center (available at: ftp://ftp.ifremer.fr/ifremer/argo/geo/atlantic_ocean/). All those observations were required to step into a data quality control (QC) procedure, which is an essential part to any oceanic data assimilation system since spurious data can compromise the analysis quality and introduce artificial trends in the assimilation results (Yan and Zhu, 2010). The QC used in this work was both developed by REMO together with the Brazilian Navy and tests the date, the location and the temperature and salinity of each Argo profile previously collected. The validation of T and S profiles was made according to all criteria and references established by the Global Temperature and Salinity Pilot Program from the Intergovernmental Oceanographic Commission (IOC, 1990) and from the database of the National Oceanographic Data Center (NODC).

From January 2010 to December 2012, 47 999 valid Argo profiles were assimilated into HYCOM in the Atlantic Ocean. These profilers covered almost the entire model domain, and were especially dense in the North Atlantic, as shown in Fig. 4.

4.2 Configuration and evaluation of the assimilation experiments

Five integrations were performed from 1 January 2010 to 31 December 2012 to evaluate the impact of the Argo data assimilation and the vertical localization in the correction of the ocean state and circulation. The first one was a control run without assimilation (CTL). The other runs were assimilation runs, namely, (i) assimilation without vertical localization (ASSIM), (ii) assimilation with vertical localization of layer thickness (VL_DP), (iii) assimilation with vertical localization of temperature and salinity (VL_TS) and (iv) assimilation with vertical localization of all the assimilated variables (VL_DPTS). In all experiments with data assimilation, a 3-day observational window was considered in order to select all the valid profiles collected 3 days before the assimilation day. The interval between each assimilation step was also 3 days and the scalar α in Eq. (2) was defined as 0.3.

The Argo daily data were also used to validate the results of the experiments on a daily basis, despite the fact that assimilation was performed only every 3 days. However, the prior state was always considered in the evaluation of the assimilation runs. The prior is the model state immediately before the assimilation step. Considering each assimilation cycle, the ocean states at 24, 48 and 72 h after assimilation were assessed in the assimilation runs. Therefore, the validation of the prior is done with independent data, since all the Argo profiles employed in the validation were used only in the next assimilation cycle. This procedure is analogous
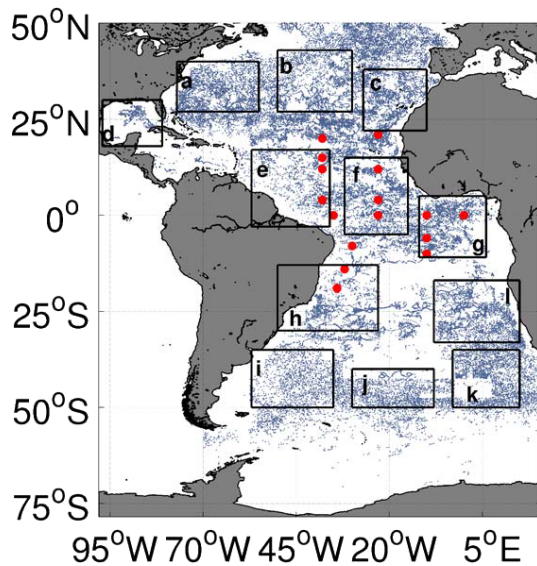
Figure 4. Locations of the 47 999 valid Argo profiles (blue dots) assimilated and also used to validate the prior state of each experiment from 2010 to 2012 in the model domain (100° W–20° E, 78° S–50° N), which excludes the Pacific Ocean, the Mediterranean Sea, and the North Atlantic subpolar region. The 16 fixed moorings from PIRATA used in the experiment's validation are represented by red dots. The numbers of valid Argo profiles in the 12 sub-regions (a–l) were 3085, 2352, 2522, 540, 1463, 2750, 2475, 1849, 1560, 1597, 2536 and 3067.

to the evaluation of forecasts. Moreover, 16 fixed moorings from the Prediction and Research Moored Array in the Tropical Atlantic (PIRATA) were used as another independent data set for validation. Their locations are represented by red dots in Fig. 4. Also, the domain was split into 12 sub-regions from a to l, as shown in Fig. 4, in order to evaluate the regional impact of the assimilation. These sub-regions and their coordinates were selected taking into account the spatial distribution of the surface mean kinetic energy and the mean standard deviations of SSH, SST and sea surface salinity (SSS) from 1997 to 2008. The numbers of Argo profiles assimilated in the experiments in the 12 sub-regions a–l were 3085, 2352, 2522, 540, 1463, 2750, 2475, 1849, 1560, 1597, 2536 and 3067.

Outputs from the HYCOM Navy Coupled Ocean Data Assimilation (HYCOM + NCODA) (Chassignet et al., 2007, 2009) system available in z-levels and fields from the Ocean Surface Current Analyses Real Time (OSCAR) (Johnson et al., 2007) were employed to compare the velocity fields produced by the assimilation runs. Also, WOA13 climatology for the period from 2005 to 2012 was used to evaluate the mean state of T, S and heat content of the upper 300 m (HC300) in all experiments. In this period, WOA13 includes the global coverage of Argo floats to compose the climatological gridded fields of T and S in 1/4° of spatial resolution.

5 Results

5.1 Comparison of mean states

The first comparison is conducted with the WOA13 climatology along 25° W for the upper 1000 m (Fig. 5). It was already verified that the free model run mean from 1997 to 2008 showed substantial differences with respect to the WOA13 climatology (Fig. 1). According to Chen et al. (2000), ocean circulation models can naturally present large and systematic biases in T and S due to their initialization and configuration. This is also shown by the control run, particularly in the North Atlantic. As mentioned before, to simulate the MW in the Atlantic, relaxation of T and S in the boundary condition is imposed without mass flux. Moreover, the resolution of 1/4° could also be a limitation to accurately solve the Gulf Stream and its associated dynamics in the mid-latitudes of the North Atlantic (Hulburt and Hogan, 2000). The main purpose of this grid is to provide boundary conditions for higher-resolution grids focusing on the Metarea V. However, even with these limitations, the assimilation schemes are able to substantially reduce these differences and correct the ocean state towards WOA13. For example, the positive temperature bias up to 3 °C and the negative salinity bias up to 0.4 psu in the upper 300 m of the control run are not produced in any assimilation run. Moreover, all the discrepancies found in the MW are remarkably diminished in the data assimilation runs, which are able to decrease the differences in 0.6 psu and more than 2 °C towards WOA13 between 20 and 40° N in the sub-surface, especially below 400 m. This correction is very effective in the VL_DP and VL_DPTS runs, which adopt the vertical localization of Δp. The VL_DPTS run is able to more efficiently improve the mean T and S states in the North Atlantic along the entire water column by further reducing the differences of 1 °C and 0.1 psu found in the other assimilation runs with respect to WOA13. Also, at 300 m near 25° S, only the VL_DPTS run decreases the differences towards WOA13 from more than 3 to almost 1 °C, while the ASSIM and VL_TS runs are not able to reduce these differences with respect to the control run. But, south of 30° S, the improvements of T and S in the VL_DPTS are considerably smaller than those obtained by the other assimilation runs.

Near 50° S, the control run does not well represent the formation of the Antarctic Intermediate Water Mass (AAIW). The experiments with data assimilation can correct the negative biases shown by the control run by up to 2 °C and 0.3 psu, except for in experiment VL_DPTS. Since this is a region of water mass formation, it seems that the vertical correlations between model layers are very important in order to represent physical processes. In the experiments with vertical localization of Δp, much of this information between layers is lost (Fig. 3), and the experiment VL_DPTS was not efficient at improving the model state for this region in particular.
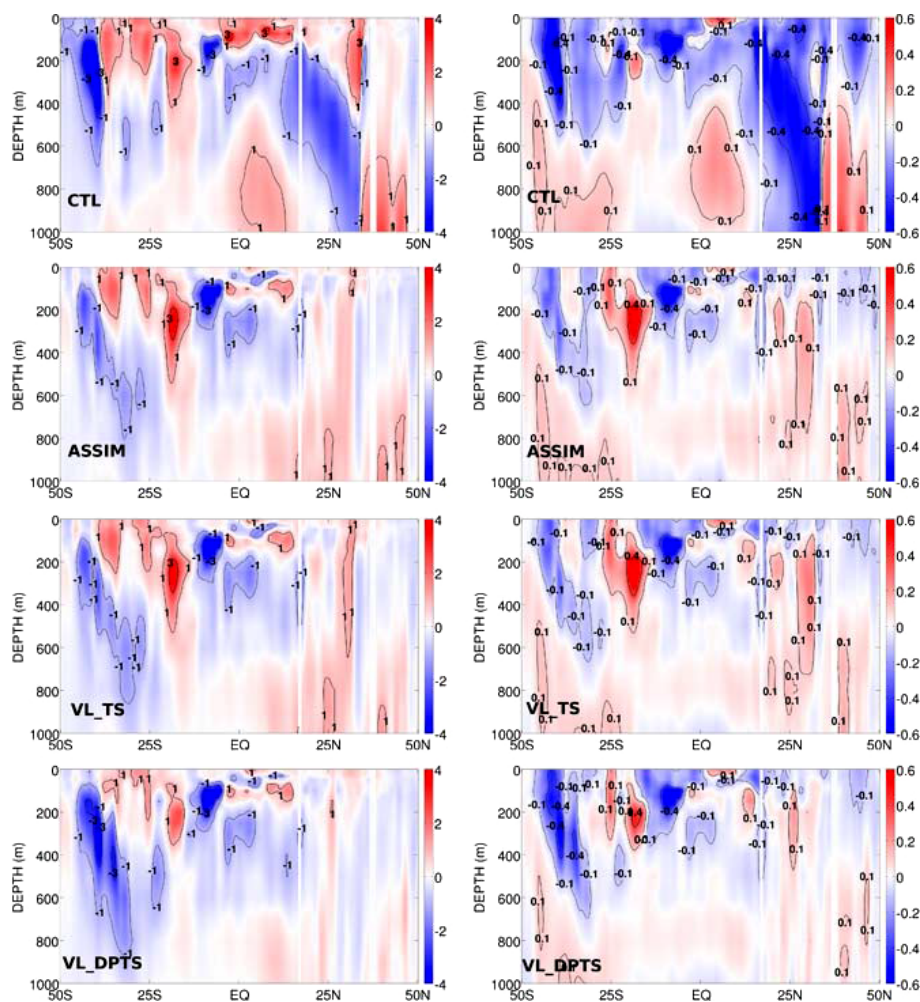
Figure 5. The model mean minus WOA13 along 25° W in the upper 1000 m for temperature (°C) on the left and salinity (psu) on the right from 1 January 2010 to 31 December 2012 according to the CTL run and the prior state of the ASSIM, VL_TS and VL_DPTS runs.

In order to investigate how stratification was modified by Argo data assimilation in the upper 300 m, Fig. 6 shows the meridional section along 25° W of the mean potential density and the position of the layers interfaces for the experiments CTL, ASSIM and VL_DPTS. In general, as already shown in Fig. 5, the assimilation runs tend to decrease the temperature and increase the salinity in the upper 300 m, driving the model towards WOA13. This is associated with a general increase of the mean potential density in the sub-surface in all the assimilation experiments, especially between 30° S and 30° N. This increase of the potential density can reach more than $0.5 \, \mathrm{kg \, m^{-3}}$ near the equator and in the subtropical gyres. According to Bleck (2002), the implementation of the isopycnal vertical coordinate in HYCOM follows the theoretical formulation that each isopycnal layer will move to be as close as possible to its target density. Therefore, in the assimilation experiments the deeper layers will move towards the surface to encompass the density increase produced by the T and S increments and satisfy their target den-

sities. For example, near the equator and in the subtropical gyre of the South Atlantic, there is a displacement of 25 m or more of the deeper layer interfaces towards the surface. In the North Atlantic, this displacement is even larger and it can reach 150 m, as can be seen by the behavior of the lower-most white dashed line in Fig. 6 that represents the interface between the 11th and 12th layer. Large et al. (1997) showed that the warm temperature bias near the equator is a common feature of coarse-resolution models and it is related to weak zonal currents and weak zonal slopes of the isotherms. However, all the data assimilation experiments are able to force the present low-resolution model to stack the layer interfaces in the upper 300 m and correct the model bias of temperature and salinity not only at the equator but also at all latitudes, especially between 30° S and 30° N.

The depth of the 20 °C isotherm is also evaluated for the experiments and the WOA13 climatology in two longitudinal sections, one along the equator and the other along 30° N (Fig. 7). In the WOA13, the depth of 20 °C isotherm along

equator is about 120 m in the west and it gets shallower in the east with a value of almost 50 m. The control run is clearly warmer than WOA13 in the upper ocean with temperatures higher than 28 °C in the first 100 m, so that the position of 20 °C isotherm is deeper, especially in the east of the section with values close to 100 m. All the assimilation experiments reduce the warm bias in the upper ocean, and the depth of 20 °C isotherm is lifted to a shallower position in much better agreement with the WOA13 climatology. In the North Atlantic, the control run has an even stronger bias of more than 4 °C in the upper 300 m. Hence, the depth of 20 °C isotherm of the control run is deeper than WOA13, attaining more than 300 m in the west of the section. Again, the experiments with data assimilation are able to significantly reduce the warm bias by moving the position of the 20 °C isotherm more than 150 m towards the surface. This agrees very well with the large displacement of the layer interfaces in the North Atlantic shown in Fig. 6, and also with the WOA13 climatology. Along 30° N, between 75 and 65° W and between 40 and 25° W, the experiment VL_DPTS is able to bring the 20 °C isotherm up to 50 m closer to the surface in comparison with the other runs without the vertical localization of layer thickness. On the other hand, the experiment VL_DPTS is the only assimilation experiment that does not represent very well the upwelling in the east margin of the Atlantic Ocean at 30° N in Fig. 7. Probably, this is because there is a strong vertical correlation among several model layers in the upwelling region that is not considered by the vertical localization of layer thickness in the VL_DPTS run. Similar behavior was observed in the region of AAIW formation shown in Fig. 5.

To further assess the impact of the assimilation in the upper ocean thermal structure, the spatial distribution of the HC300 between 50° S and 50° N is presented in Fig. 8 for the WOA13 together with the differences (model mean minus WOA13) for each run. In general, the control run has a larger HC300 in comparison with the WOA13 climatology. Vast areas of the North and South Atlantic have more than 50 MJ m^{-2} of HC300 excess over the climatology. The largest differences are attained in the Gulf Stream region (about 400 MJ m^{-2}), since it is misplaced northward with respect to observations, and in the southwest Atlantic around 40° S, associated with a misrepresentation of the Brazil–Malvinas Confluence. When Argo data are assimilated, the model temperature is constrained towards WOA13 and there is a strong impact on the HC300. All the assimilation experiments decrease the HC300 difference with respect to WOA13 by more than 150 MJ m^{-2} for large areas of the subtropical North Atlantic and 50 MJ m^{-2} in the equatorial and South Atlantic. However, large differences still remain in the simulated Gulf Stream region. No Argo data were available in this shelf region – to the north of the observed Gulf Stream – to constrain the model (Fig. 4); therefore, the assimilation runs could not correct this discrepancy. Also, no SST and SLA were assimilated and the assimilation of Argo
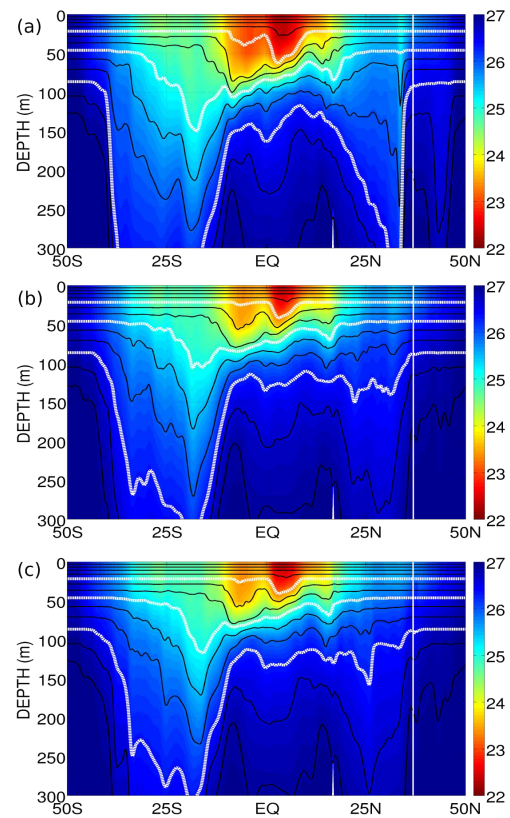

Figure 6. Potential density (kg m^{-3}) and the position of the layer interfaces along 25° W in the upper 300 m for the **(a)** CTL run and the prior state of the **(b)** ASSIM and **(c)** VL_DPTS runs from 1 January 2010 to 31 December 2012. The white dashed lines represent the interface between the 5th and the 6th, between the 8th and the 9th, and between the 11th and the 12th isopycnal layers, from top to bottom.

only could not completely modify this important circulation feature.

5.2 Comparison with Argo profilers and PIRATA moorings

The root mean square deviation (RMSD) is calculated on a daily basis against Argo and PIRATA observations to objectively evaluate the temperature and salinity produced by the experiments. This comparison is done with independent data, since the 24 h, the 48 h and the 72 h simulations after each analysis cycle is assessed. The 72 h simulations correspond to the prior or background states used in the assimilation steps.

The depth averaged RMSD of T and S above 2000 m against the Argo profilers in the model domain from 1 January 2010 to 31 December 2012 is shown in Fig. 9. The RMSDs of the ASSIM and VL_DPTS runs are substantially reduced from 1.75 °C and 0.30 psu in the beginning of the runs to 1 °C and 0.14 psu by the end of the runs. In Fig. 9,
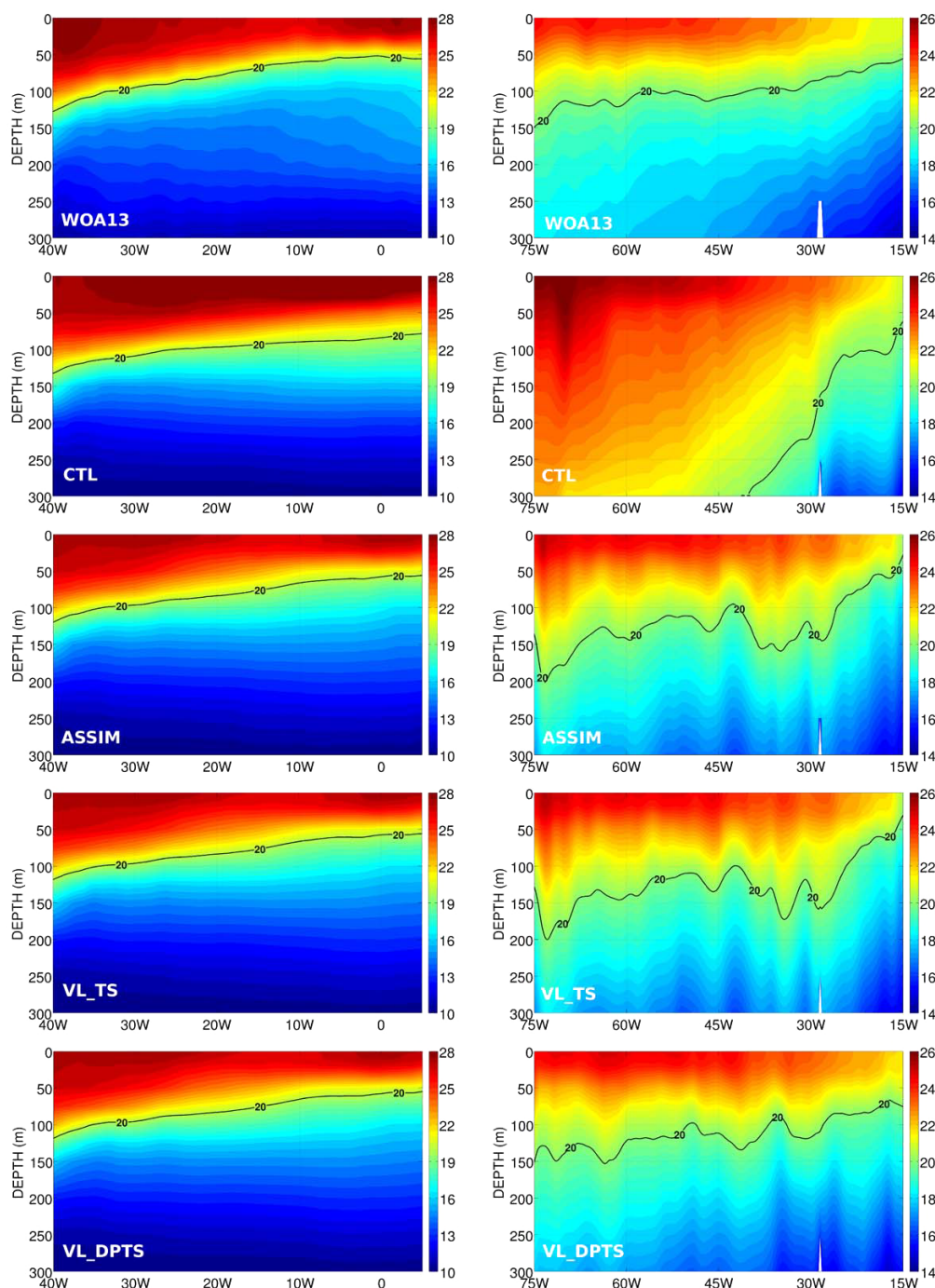
Figure 7. Temperature (°C) along the Equator (left) and along 30° N (right) in the upper 300 m for the WOA13 climatology, the CTL run and the prior state of the ASSIM, VL_TS and VL_DPTS runs from 1 January 2010 to 31 December 2012.

the only assimilation runs presented are the ASSIM and VL_DPTS, since the other two assimilation runs produced similar curves. The first year of integration has the greatest RMSD reduction for the assimilation experiments. From the second year on, especially in the last year, the RMSD reduction is smaller and the RMSD values tend to oscillate

around a stable mean. This indicates that the analysis error is close to saturation and that changes in the assimilation parameters such as α may be necessary to better constrain the model towards observations. Here, α is equal to 0.3. However, care should be taken when increasing α. The relatively small value chosen here considered two aspects. First, a small
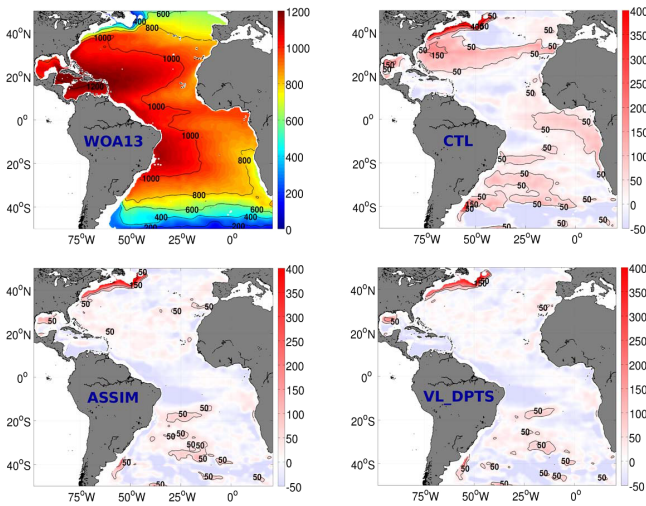
Figure 8. Heat content (MJ m^{-2}) to 300 m of the WOA13 climatology and the model mean minus WOA13 for the CTL run and the prior state of the ASSIM and VL_DPTS runs over the period from 1 January 2010 to 31 December 2012. Regions shallower than 300 m were not considered.

value of α helps avoiding abrupt changes in the model state and numerical instability considering the relatively large discrepancies between the model background and observations in certain regions. Also, the model resolution is relatively low and the use of high values of α could overfit the data and try to introduce mesoscale patterns that cannot be resolved by the model.

The VL_DPTS run produces smaller RMSDs than the ASSIM run. The RMSD of the VL_DPTS is about 86 % (69 %) of the ASSIM RMSD for temperature (salinity). However, this improvement is mainly due to the vertical localization of the layer thickness. It was found that when applying the vertical localization of the layer thickness, the number of grid points with negative Δp after assimilation and before the post-processing was greatly reduced. In the 365 assimilation steps from 1 January 2010 to 31 December 2012, more than 235 000 grid points with negative Δp were generated and needed post-processing in the ASSIM run. This number was decreased to less than 80 000 in the VL_DP and VL_DPTS runs, which means a reduction of 66 %. The improvement caused by the vertical localization of Δp, rather than the vertical localization of T and S, will be discussed in detail below.

The vertical profiles of the mean RMSD of all runs against Argo and PIRATA data over the whole model domain are presented in Fig. 10. The largest temperature and salinity RMSD of the control run are attained in the top 700 m with respect to Argo data and in the top 200 m with respect to PIRATA data. This is associated with difficulties that models have with representing the thermocline and pycnocline regions of sharp vertical gradients (Oke and Schiller, 2007; Xie and Zhu, 2010). When Argo data are assimilated, the vertical
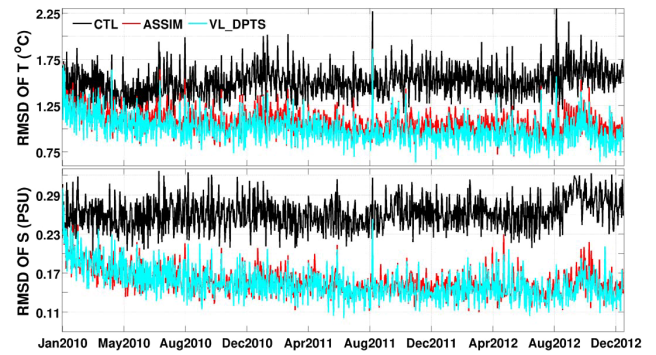

Figure 9. Depth-averaged RMSD of temperature (°C) and salinity (psu) to 2000 m against 47 999 Argo profilers in the model domain (100° W–20° E, 78° S–50° N) from 1 January 2010 to 31 December 2012 according to the control run (black) and the prior state of the experiment ASSIM (red) and VL_DPTS (light blue).

thermohaline structure is very much improved. With respect to Argo data, the experiment ASSIM decreases the RMSD of the control run in the top 600 m from 2 to 1.3 °C and from almost 0.4 to 0.19 psu. Regarding PIRATA data, the experiment ASSIM reduces the RMSD values of the control run by up to 1 °C and 0.13 psu in the first 120 m. Gains with the vertical localization of Δp are seen from 200 to 800 m and from 1200 to 2000 m. In these ranges, the experiment VL_DPTS is able to further decrease the RMSD by 0.15 °C and 0.025 psu with respect to the other assimilation runs. The vertical localization of only T and S had almost no positive impacts on the RMSD, and contributed to slightly degrading S near the surface when the runs were compared to PIRATA data. This is mainly because the most important HYCOM variable to determine the thermohaline structure is the model layer thickness (Thacker and Esenkov, 2002; Thacker et al., 2004; Xie and Zhu, 2010). For instance, according to HYCOM formulation, if the analysis increments of T and S decrease the density of a specific layer in comparison to its target density, the model attempts to move the lower interface downward, so that the flux of denser water across this interface increases the layer density (Bleck, 2002). The opposite happens if there is an increase of the layer density compared to its target density. Therefore, the analysis increments of T and S will try to indirectly change the model layer thicknesses in order to adjust the layer densities to their target densities. However, in the EnOI scheme proposed here, the assimilation of Δp_{obs} is realized before the assimilation of T and S, and is very effective in changing the model layer thicknesses. This may explain why the model thermohaline structure is much more sensitive to the vertical localization of Δp rather than to the vertical localization of T and S. The latter was studied by XZ and they also did not find any significant impact in their assimilation experiments.

According to Fig. 10, the vertical localization of Δp seems to be a good approach to better improve the model state,
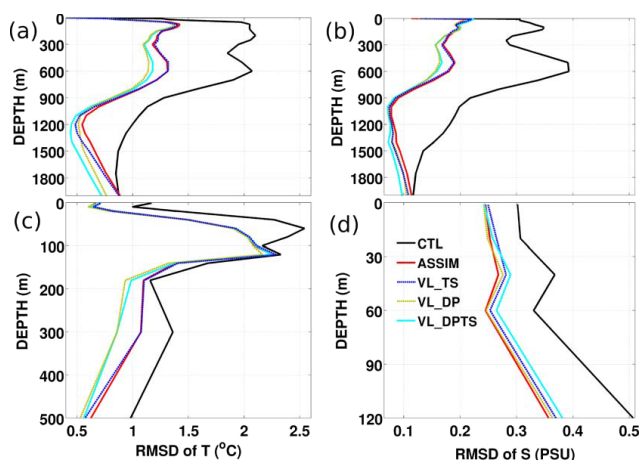
Figure 10. Vertical distribution of the RMSD of temperature (°C) and salinity (psu) against 47 999 Argo profilers and 16 PIRATA fixed moorings in the model domain (100° W–20° E, 78° S–50° N) from 1 January 2010 to 31 December 2012 according to the control run (black) and the prior state of the experiment ASSIM (red), VL_TS (dark dashed blue), VL_DP (dark dashed yellow) and VL_DPTS (light blue). **(a)** and **(b)** correspond to the RMSD of T and S with respect to Argo observations while **(c)** and **(d)** correspond to the RMSD of T and S with respect to PIRATA data, respectively.

especially when correlations of the ensemble members come from a free model run that contains large discrepancies in certain regions in comparison with climatology and observations. Since the co-variance matrix of the model errors in the EnOI scheme is calculated by many snapshots of the model state, the performance of the assimilation is quite dependent on the accuracy of the error co-variances (Oke et al., 2005; Xie and Zhu, 2010). Due to the biases of the ensemble members used in this work, many vertical correlations might not be well represented, especially between distant layers. Thus, since the vertical localization of layer thickness mostly keeps the correlations between adjacent layers, this strategy avoids the generation of many grid points with negative Δp during the assimilation process and makes this approach a safer way to produce a more physically consistent and reliable analysis.

To investigate the spatial distribution of the RMSD of T and S with respect to Argo data above 2000 m, Table 1 contains the deviations for all the 12 sub-regions previously defined in Fig. 4. In the control run, the largest RMSDs of T and S are attained in the sub-regions a, b and d with values greater than 1.6 °C and 0.27 psu. All those sub-regions are found in the North Atlantic, where the model has the largest differences in comparison with the WOA13 climatology and with Argo observations. This is particularly clear for sub-region a corresponding to the Gulf Stream region with RMSD of 2.3 °C, and for sub-region d in the Gulf of Mexico with RMSD of 0.368 psu. It should be reinforced that this grid configuration was prepared to provide boundary conditions

for higher-resolution grids focusing on the Metarea V. Hence, the control run is better adjusted with smaller RMSDs in the South Atlantic than in the North Atlantic as shown in Table 1. The maximum RMSD in the South Atlantic is attained in the sub-region i corresponding to the Brazil–Malvinas Confluence with values of 1.29 °C and 0.207 psu. For all the sub-regions, the different EnOI runs are able to reduce the RMSD values of T and S in comparison with the control run. The greatest impact of data assimilation is in the North Atlantic, where approximately 60 % of the assimilated Argo profiles are found and where the control run has its largest RMSDs. For example, the experiment ASSIM decreases the RMSD of the control run about 1 °C in the sub-regions a and b, and more than 0.17 psu for the sub-regions b, c and d. In the Gulf of Mexico, this reduction is up to 0.2 psu. On other hand, the RMSD reductions by the ASSIM run in the South Atlantic are only about 0.2 °C and 0.04 psu. Even with the large impact of the Argo data assimilation, some regions still remain with large RMSDs. This is the case with the Gulf Stream, the Gulf of Mexico and the Brazil–Malvinas Confluence in the sub-regions a, d and i, respectively, where the experiments still have RMSDs greater than 1 °C and 0.145 psu. These are regions of strong gradients, variability and mesoscale eddy activity. The Gulf Stream is highly influenced by mesoscale circulation and the growth of baroclinic instability after its separation from the North American coast (Robinson et al., 1989; Spall and Robinson, 1990; Lee and Mellor, 2003). Also, the Brazil–Malvinas Confluence is characterized by a weak and warm southward Brazil Current meeting a strong, cold and less saline northward Malvinas Current, which results in large contrasts in stratification and strong eddy activity (Gordon, 1989; Garzoli and Garrafo, 1989; Goni et al., 1996). Thus, the large RMSD still found in the assimilation runs in these regions can be due to lack of data, inaccuracies in the ensemble members, and limitation of the model resolution to solve the mesoscale circulation patterns. Similarly, XZ found that the assimilation of Argo profiles did not capture the mesoscale activities in the Pacific Ocean due to the coarse model resolution, and the Kuroshio System remained with large RMSDs in their assimilation run.

In Table 1, the assimilation experiments with the best performances are the ones with vertical localization of layer thickness. This is particularly clear in the sub-regions a, b and c, where the experiments VL_DPTS and VL_DP are able to reduce the RMSD values up to 0.12 °C and 0.018 psu with respect to the ASSIM and VL_TS runs. These three sub-regions are located in the North Atlantic, where the model free run has its largest bias in comparison with WOA13 (Fig. 1). They are sub-regions where the ensemble vertical correlations might not be well represented. Therefore, relaxing the vertical constraints by vertical localization should improve the model state. Small gains can also be seen in the South Atlantic when applying the vertical localization of layer thickness, for example, in the sub-region h corresponding to the Brazil Current. The only regions where

Table 1. RMSD of temperature (°C) and salinity (psu) to 2000 m with respect to Argo data for the prior state of each experiment from 1 January 2010 to 31 December 2012. The calculations were performed for the whole model domain (100° W–20° E, 78° S–50° N) and for each one of the 12 sub-regions previously defined in Fig. 4.

	RMSD									
	CTL		ASSIM		VL_DPTS		VL_DP		VL_TS	
	T (°C)	S (psu)	T (°C)	S (psu)	T (°C)	S (psu)	T (°C)	S (psu)	T (°C)	S (psu)
A	2.302	0.279	1.300	0.166	1.180	0.148	1.153	0.147	1.268	0.157
B	1.722	0.283	0.728	0.113	0.647	0.107	0.674	0.110	0.707	0.109
C	1.214	0.287	0.813	0.116	0.665	0.112	0.652	0.111	0.815	0.142
D	1.663	0.368	1.089	0.160	0.995	0.175	1.016	0.170	1.034	0.175
E	0.972	0.227	0.678	0.119	0.655	0.115	0.633	0.114	0.662	0.119
F	1.177	0.163	0.761	0.097	0.729	0.094	0.729	0.094	0.765	0.098
G	1.139	0.161	0.800	0.107	0.759	0.105	0.759	0.105	0.764	0.106
H	0.756	0.166	0.633	0.125	0.550	0.115	0.554	0.109	0.651	0.132
I	1.290	0.207	1.048	0.179	1.035	0.179	1.038	0.178	1.036	0.180
J	0.920	0.158	0.670	0.124	0.700	0.127	0.709	0.129	0.673	0.124
K	0.845	0.151	0.594	0.106	0.596	0.113	0.599	0.112	0.638	0.110
L	0.786	0.140	0.575	0.090	0.549	0.082	0.548	0.079	0.595	0.092
All	1.507	0.264	0.993	0.149	0.905	0.139	0.915	0.138	0.972	0.147

the experiments VL_DP and VL_DPTS are just a little bit worse than the experiment ASSIM is the Gulf of Mexico, sub-region d, for salinity, and the sub-regions j and k in the mid-latitudes close to the AAIW formation. In general, the pairs of experiments ASSIM and VL_TS, and VL_DP and VL_DPTS, produced similar results in all sub-regions, and the strategies with vertical localization of layer thickness attained smaller RMSDs of T and S. It corroborates the conclusion that the vertical localization of Δp was more important than the vertical localization of T and S for the present experiments.

5.3 Adjustment of the altimetry and velocity fields

As shown above, the EnOI scheme used in this work adjusts the baroclinic velocity fields via their co-variance with layer thickness in the analysis increments. This is a physically consistent adjustment, since the horizontal velocity fields are dominated by the slope of the isopycnal layers, which in turn generates the pressure gradient force. XZ showed that around the assimilated Argo profiles in the mid-latitudes, cyclonic and anti-cyclonic circulations develop, consistent with the large local modifications of the model layer thicknesses and geostrophy. This is due to the nature of the co-variances, which come from an ensemble of model states and allow for the description of the anisotropic patterns of the circulation (Oke et al., 2005, 2008; Xie and Zhu, 2010). Figure 11 shows the analysis increment of layer thickness and the circulation originating in the 12th layer around an Argo profile located in the mid-latitudes of the North Atlantic on 1 January 2010 regarding the experiments ASSIM
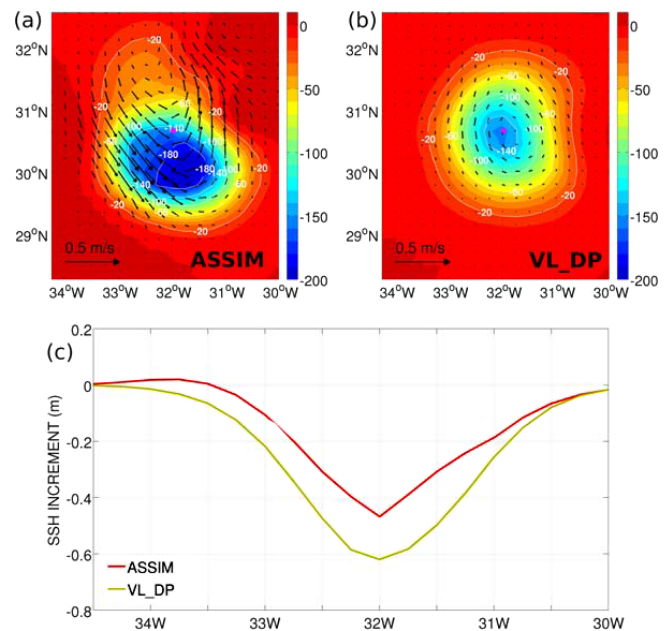

Figure 11. Analysis increments after assimilating an Argo profile located at 30.72° N and 31.98° W on 1 January 2010 of: **(a)** layer thickness (unit: m) and current components (unit: m s^{-1}) in the 12th layer for the experiment ASSIM, **(b)** layer thickness (unit: m) and current components (unit: m s^{-1}) in the 12th layer for the experiment VL_DP and **(c)** SSH (unit: m) along 30.72° N for the experiment ASSIM (red) and VL_DP (dark dashed yellow). The pink dot represents the location of the Argo float.
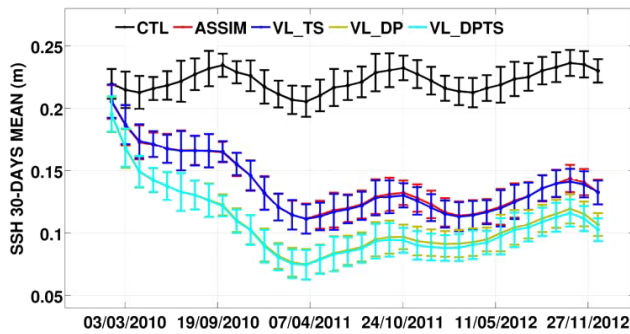
Figure 12. 30-day running SSH mean (m) of the control run (black) and the prior state of the experiment ASSIM (red), VL_TS (dark blue), VL_DP (dark yellow) and VL_DPTS (light blue) from 1 January 2010 to 31 December 2012. The markers represent the 30-day running mean values of SSH, while the bars correspond to the standard deviations of each run.

and VL_DP. The experiment ASSIM robustly induces thinning of layer thickness up to 200 m, and then a stronger and more anisotropic cyclonic pattern is imposed with velocity increments reaching $0.5 \, \mathrm{m \, s^{-1}}$. Both experiments develop a cyclonic circulation around this mid-latitude Argo profile. This is coherent with the negative analysis increment of layer thickness and the negative SSH increment also shown in Fig. 11. Thus, the geostrophic balance was preserved in both experiments. Since the experiment VL_DP has limited influence between distant layers, the analysis increments are smoothed and smaller velocity increments of about $0.1 \, \mathrm{m \, s^{-1}}$ are produced. The cyclonic gyre is better represented in the VL_DP and VL_DPTS (not shown) runs, which show a more negative and more continuous increment of SSH than the ASSIM run. The analysis constructed by this increment may be more suitable to initialize a model than the analysis by the ASSIM strategy considering the present background state and ensemble members. However, further adjustment of the velocity fields is made by the model itself during further integration to balance the analysis increments after assimilation of Δp, T and S.

The trend of SSH reduction by Argo data assimilation in the present experiments is confirmed in Fig. 12. It shows the 30-day SSH running mean of all experiments considering the model domain between 50° S and 50° N. From the beginning of the integration to December 2012, SSH is reduced from 0.22 to 0.12 m in the experiments ASSIM and VL_TS, which is compatible with the largest impact found in the first year of assimilation in Fig. 9. The VL_DP and VL_DPTS runs produce a larger reduction from 0.22 to 0.08 m. From this period on, the SSH of all assimilation experiments stabilize around these values and different SSH means are achieved in the experiments with and without vertical localization of layer thickness, but with variability very close to the control run. The diagnostic model SSH varies due to the barotropic pressure mode and especially due to the deviations in tem-

perature and salinity caused by changes in the structure of the layer thicknesses (Chin et al., 2002). Hence, this lower SSH mean in the assimilation experiments reflects the new thermohaline state and stratification achieved when Argo data are assimilated. The thermal expansion of seawater constitutes a very important component of SSH and then the reduction of the temperature and the heat content in the assimilation experiments contribute strongly to the SSH decrease. For example, the heat content is positively correlated with SSH (Chambers et al., 1998; Willis et al., 2004; Dong et al., 2007). In this work, mean correlation values of 0.802, 0.815 and 0.812 between the HC and SSH are found for the experiments CTL, ASSIM and VL_DPTS, respectively. When comparing the monthly means of all the experiments with the WOA13 monthly means over the entire domain (not shown), the experiments VL_DP and VL_DPTS are able to decrease the HC by more than $7 \, \mathrm{MJ \, m^{-2}}$ towards the climatology. This could explain why there is a larger decrease of the mean SSH in the experiments with vertical localization of layer thickness.

In order to evaluate how the local changes in layer thickness, SSH and velocity fields affect the large-scale circulation, the zonal velocity along 25° W in the upper 300 m for all the experiments and HYCOM + NCODA analysis is shown in Fig. 13. Some observed features are clearly represented in the HYCOM + NCODA analysis, such as the South Atlantic Current (SAC) south of 40° S, the branches of the westward South Equatorial Current (SEC), the eastward equatorial undercurrent (EUC) and the North Equatorial Countercurrent (NECC). Modifications with respect to the control run were imposed by assimilation. However, some can be considered improvements towards the HYCOM + NCODA analysis, and some cannot. For example, there is a strong zonal current in the control run up to $0.6 \, \mathrm{m \, s^{-1}}$ between 30 and 40° N associated with the Gulf Stream recirculation path, which is much more intense than in the HYCOM + NCODA analysis. This zonal current is reduced to 0.1–$0.2 \, \mathrm{m \, s^{-1}}$ in the experiments VL_TS and VL_DPTS. However, it is displaced further south in comparison with HYCOM + NCODA analysis, especially in the VL_TS run. In the experiment ASSIM, this recirculation path is even weaker and merges in the broader eastward flow. Also, in all assimilation experiments the EUC velocity decreases from 0.6 to $0.4 \, \mathrm{m \, s^{-1}}$ and gets closer to the magnitude of the HYCOM + NCODA analysis, but the shape of its core remains the same as in the control run and is not as elongated as in the HYCOM + NCODA analysis. On the other hand, all data assimilation runs constrain the SAC near to the surface, while it reaches more than 300 m in the HYCOM + NCODA analysis. Finally, assimilation is not able to reduce the magnitude of the SEC and to increase the intensity of the NECC simulated by the control run.

XZ could observe modifications in the equatorial Pacific current system due to Argo data assimilation, but they were not always positive when compared to the Tropical Atmosphere Ocean (TAO) array data. For instance, a too deep
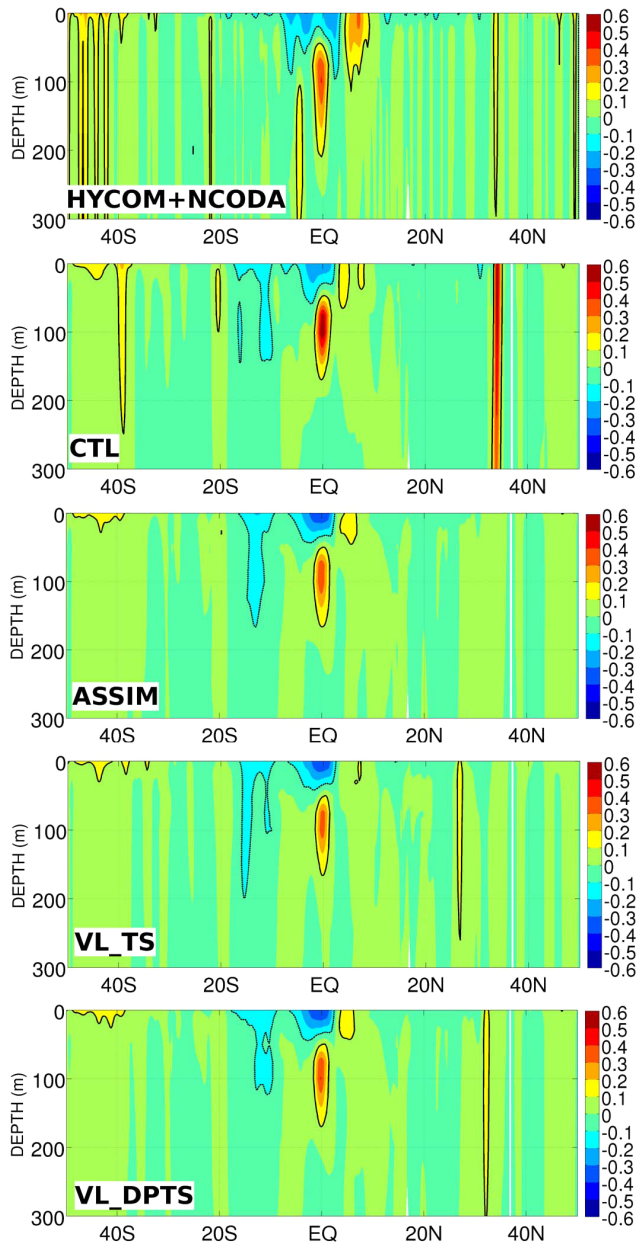
Figure 13. Zonal velocity fields ($\mathrm{m\,s^{-1}}$) in the upper 300 m along 25° W of the HYCOM + NCODA analysis, the CTL run and the prior state of the ASSIM, VL_TS and VL_DPTS runs from 1 January 2010 to 31 December 2012. The black solid and black dashed lines denote the contours of $0.1\,\mathrm{m\,s^{-1}}$ and $-0.1\,\mathrm{m\,s^{-1}}$, respectively.

undercurrent maximum in the east and a too thick and strong westward current in the west were produced. In the present work, a few modifications were observed in the equatorial Atlantic region as well, but with smaller intensity than in the mid-latitudes. It should be noted that the geostrophic balance that holds in most of the ocean interior loses importance close to the equator. Therefore, the equatorial current system should be less sensitive to analysis increments by Argo data

assimilation, so that changes in this region may be mostly attributed to remote impacts from the assimilation in extra-equatorial regions.

In the present work, the discrepancies of the model free run and its variability with respect to climatology and observations are an important limitation for the assimilation to produce the correct analysis increments, particularly for the large-scale circulation. For this reason, many works point out that the model biases should be considered during the assimilation process (Reynolds et al., 1996; Dee and Silva, 1998; Bell et al., 2004; Dee, 2005; Xie and Zhu, 2010). Also, improving the co-variances of the ensemble members may lead to more accurate analyses (Oke et al., 2008; Xie and Zhu, 2010). In addition to the comparison with the HYCOM + NCODA analysis, the surface velocity fields of the assimilation runs were compared to the OSCAR fields. Again, the results did not provide a clear signal about the impact of Argo data assimilation on the surface currents. The RMSD of U and V with respect to OSCAR were reduced by less than 5 % in comparison to the control run.

6   Conclusions and discussion

In this work, an EnOI scheme to assimilate Argo data into HYCOM was successfully constructed and implemented. The first results were evaluated against observations and analyses over the Atlantic Ocean. The EnOI scheme was based mostly on the work of XZ. A key variable in the assimilation algorithm was the observed model layer thickness, Δp_{obs}, constructed from the Argo temperature (T) and salinity (S) profiles, according to TE. First, Δp_{obs} was assimilated and the analysis state vector considered not only the model layer thickness, but also the baroclinic velocities. This procedure is physically consistent, since the horizontal velocity fields are dominated by the difference between the layer depths, which will be the responsible for the pressure gradient force. After that, with the previously adjusted model layer thicknesses, T and S were assimilated in separate steps. Finally, T was diagnosed below the mixed layer through the seawater equation of state following the best results found in XZ. Thus, this EnOI scheme respects the hybrid nature of the HYCOM vertical coordinate system and allows for the restructuring of the isopycnal layer thicknesses and velocities with the assimilation of Δp_{obs}. Also, a sensitivity study was performed considering different vertical localizations of the model error co-variance matrix, involving the variables T, S and especially Δp. Five integrations were realized from 1 January 2010 to 31 December 2012. The study of the impact of the vertical localization of Δp is an original contribution of this work.

The thermohaline structure of the experiments with assimilation was significantly improved above 2000 m, the maximum depth of the Argo data. The RMSDs with respect to Argo observations in the assimilation runs were reduced in

at least 34 % (44 %) for T (S), regarding the control run over the whole domain. Spatially, the RMSD of T and S decreased for all the 12 selected sub-regions of the domain with remarkable corrections in the depth of the 20 °C isotherm and heat content in the upper 300 m towards WOA13, particularly in the North Atlantic. Also, the reorganization of the isopycnal layers by the assimilation of Δp_{obs} has provided a large reduction of the diagnosed SSH in the model, reflecting the new thermohaline state achieved by Argo data assimilation. Indeed, the correction of layer thickness played a role in correcting the model thermohaline structure and stratification, as stated in previous data assimilation works with HYCOM (Thacker and Esenkov, 2002; Chin et al., 2002; Thacker et al., 2004; Xie and Zhu, 2010).

The vertical localization of T and S did not produce any significant impact in the assimilation experiments, which is consistent with the results found in XZ. However, the experiments VL_DP and VL_DPTS with vertical localization of Δp were able to decrease the RMSD of T (S) from 0.993 °C (0.149 psu) to 0.905 °C (0.138 psu) in the whole model domain with respect to the other assimilation runs. This improvement was especially seen in the North Atlantic, where the ensemble members and background had their largest biases with respect to observations and climatology. The experiments with vertical localization of Δp decreased by 66 % the number of grid points with negative Δp generated during the assimilation process by constraining the vertical co-variances between more distant layers, and then reducing the degrees of freedom offered by the ensemble members. This result reinforces how important the quality of the ensemble members is in order to improve the performance of the assimilation, particularly in the EnOI scheme, in which the ensemble is stationary in time and does not evolve with the model integration as in the EnKF (Evensen, 2003; Oke et al., 2005, 2008). In future works with EnOI, the quality of the ensemble members should be taken into account. For example, long reanalysis with Argo data assimilation should be performed first to provide new and better ensemble members and, therefore, improve the accuracy of the model error co-variance to be employed in a new assimilation run. If more accurate ensemble members are employed, it is expected that the strong vertical localization of Δp will not lead to improvements in the analysis.

Despite the evidence that the assimilation experiments caused a strong SSH reduction in the model domain and that the analysis increments of Δp and velocities were locally consistent, there was not a clear improvement in the large-scale circulation. The assimilation schemes were simply bias blind here. The model biases should be considered in future studies to better analyze the impact of Argo data assimilation in the large-scale circulation. According to Oke and Schiller (2007), the role of Argo data assimilation is mainly to constrain the thermohaline structure of the model, especially for salinity, and this was obtained in the present work. To constrain the SST and T in the mixed layer, the assimi-

lation of SST should be performed, and to correct the upper mesoscale circulation, the assimilation of SLA is needed.

The present work was a key step forward in two major directions. First, it was the basis for the future implementation of assimilation in the other REMO higher-resolution grids. For the current domain with 760 × 480 horizontal grid points, the Argo data assimilation code took around 12 min of CPU on 32 2 GHz processors to assimilate approximately 150 Argo T–S profiles. Therefore, it can be easily used in operational mode. Second, it served as the backbone to construct the assimilation code for along-track satellite altimetry data and SST analyses that will form the REMO Ocean Data Assimilation System into HYCOM (RODAS_H) for operational and research purposes (Tanajura et al., 2014).

References

Bell, M. J., Martin, M. J., and Nichols, N. K.: Assimilation of data into an ocean model with systematic errors near the equator, Q. J. Roy. Meteor. Soc., 130, 873–893, 2004.

Belyaev, K. P., Tanajura, C. A. S., and Tuchkova, N. P.: Comparison of Methods for Argo Drifters Data Assimilation into a Hydrodynamical Model of the Ocean, Oceanology, 52, 593–603, 2012.

Bleck, R.: An oceanic general circulation model framed in hybrid isopycnic-Cartesian coordinates, Ocean Modell., 4, 55–88, 2002.

Bleck, R. and Smith, L. T.: A wind-driven isopycnic coordinate model of the north and equatorial atlantic ocean. 1. Model development and supporting experiments, J. Geophys. Res., 95, 3273–3285, 1990.

Boyer, T. and Mishonov, A.: World Ocean Atlas 2013 Product Documentation, Technical Ed. NOAA Atlas NEDSIS, 14 pp., 2013.

Brydon, D., Sun, S., and Bleck, R.: A new approximation of the equation of state for sea water, suitable for numerical ocean models, J. Geophys. Res., 104, 1537–1540, 1999.

Carton, J. A. and Giese, B. S.: A Reanalysis of Ocean Climate Using Simple Ocean Data Assimilation (SODA), Mon. Weather Rev., 136, 2999–3017, 2008.

Chambers, D., Tapley, B., and Stewart, R.: Measuring heat storage changes in the equatorial Pacific: A comparison between TOPEX altimetry and tropical atmosphere-ocean buoys, J. Geophys. Res., 103, 591–597, 1998.

Chassignet, E. P., Smith, L. T., and Halliwell, G. R.: North Atlantic Simulations with the Hybrid Coordinate Ocean Model (HYCOM): Impact of the vertical coordinate choice, reference

pressure, and termobaricity, J. Phys. Oceanogr., 33, 2504–2526, 2003.

Chassignet, E. P., Hulburt, H. E., Smedstad, O. M., Halliwell, G. R., Hogan, P. J., Wallcraft, A. J., and Bleck, R.: Generalized Vertical Coordinates For Eddy-Resolving Global and Coastal Ocean Forecasts, Oceanography, 19, 20–31, 2006.

Chassignet, E. P., Hurlburt, H. E., Smedstad, O. M., Halliwell, G. R., Hogan, P. J., Wallcraft, A. J., Baraille, R., and Bleck, R.: The HYCOM (HYbrid Coordinate Ocean Model) data assimilative system, J. Marine Syst., 65, 60–83, 2007.

Chassignet, E. P., Hurlburt, H. E., Metzger, E. J., Smedstad, O. M., Cummings, J., Halliwell, G. R., Bleck, R., Baraille, R., Wallcraft, A. J., Lozano C., Tolman, H. L., Srinivasan, A., Hankin, S., Cornillon, P., Weisberg, R., Barth, A., He, R., Werner, F., and Wilkin, J.: US GODAE: Global Ocean Prediction with the HYbrid Coordinate Ocean Model (HYCOM), Oceanography, 22, 64–75, 2009.

Chen, D., Cane, M. A., Zebiak, S. E., Cañizares, R., and Kaplan, A.: Bias correction of an ocean-atmosphere coupled model, Geophys. Res. Lett., 27, 2585–2588, 2000.

Chin, T. M., Haza, A. C., and Mariano, A. J.: A Reduced-Order Information Filter for Multilayer Shallow-Water Models: Profiling and Assimilation of Sea Surface Height, J. Atmos. Ocean. Technol., 19, 517–533, 2002.

Counillon, F. and Bertino, L.: Ensemble Optimal Interpolation: multivariate properties in the Gulf of Mexico, Tellus A, 61, 296–308, 2009.

Daley, R.: Atmospheric Data Analysis, Cambridge University Press, Cambridge, UK, New York, USA, 457 pp., 1991.

Dee, D. P.: Bias and data assimilation, Q. J. Roy. Meteor. Soc., 131, 3323–3343, 2005.

Dee, D. P. and da Silva, A. M.: Data assimilation in the presence of forecast bias, Q. J. Roy. Meteor. Soc., 124, 269–295, 1998.

Dombrowsky, E., Bertino, L., Brassington, G. B., Chassignet, E. P., Davidson, F., Hurlburt, H. E., Kamachi, M., Lee, T., Martin, M. J., Mei, S., and Tonani, M.: GODAE systems in operation, Oceanography, 22, 80–95, 2009.

Dong, S., Hautala, S., and Kelly, K.: Interannual variations in the upper-ocean heat content and heat transport convergence in the Western North Atlantic, J. Phys. Oceanogr., 37, 682–697, 2007.

Evensen, G.: The ensemble Kalman filter: theoretical formulation and practical implementation, Ocean Dyn., 53, 343–367, 2003.

Ezer, T. and Mellor, G. L.: Continuous assimilation of Geosat altimeter data into a three-dimensional primitive equation Gulf Stream model, J. Phys. Oceanogr., 24, 832–847, 1994.

Gabioux, M. G., Costa, V. S., Souza, J. M. A. C., Oliveira, B. F., and Paiva, A. M.: Modeling the South Atlantic Ocean from medium to high-resolution, Rev. Bras. Geo., 31, 229–242, 2013.

Garzoli, S. L. and Garrafo, Z.: Transports, frontal motions and eddies at the Brazil-Malvinas Currents Confluence, Deep-Sea Res., 36, 681–703, 1989.

Gaspari, G. and Cohn, S. E.: Construction of correlation functions in two and three dimensions, Q. J. Roy. Meteor. Soc., 125, 723–757, 1999.

Goni, G., Kamholz, S., Garzoli, S., and Olson, D.: Dynamics of the Brazil-Malvinas Confluence based on inverted echo sounders and altimetry, J. Geophys. Res., 101, 16273–16289, 1996.

Gordon, A. L.: Brazil-Malvinas Confluence-1984, Deep-Sea Res., 36, 359–384, 1989.

Halliwell, G. R.: Evaluation of vertical coordinate and vertical mixing algorithms in the HYbrid-Coordinate Ocean Model (HYCOM), Ocean Modell., 7, 285–322, 2004.

Hamill, T. M., Whitaker, J. S., and Snyder, C.: Distance-dependent filtering of background error co-variances in an ensemble Kalman Filter, Mon. Weather Rev., 129, 2776–2790, 2001.

Hulburt, H. E. and Hogan, P. J.: Impact of 1/8° to 1/64° resolution on Gulf Stream model-data comparisons in basin-scale subtropical Atlantic Ocean models, Dynam. Atmos. Oceans., 32, 283–329, 2000.

Intergovernmental Oceanographic Commission (IOC): GTSPP REAL-TIME Quality Control Manual, IOC Manuals and Guides 22, 128 pp., 1990.

Johnson, E. S., Bonjean, F., Lagerloef, G. S. E., Gunn, J. T., and Mitchum, G. T.: Validation and Error Analysis of OSCAR Sea Surface Currents, J. Atmos. Ocean. Tech., 24, 688–701, 2007.

Kalnay, E.: Atmospheric Modeling, Data Assimilation and Predictability, Cambridge University Press, Cambridge, UK, New York, USA, 341 pp., 2003.

Kalnay, E., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., Wollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-year reanalysis project, B. Am. Meteorol. Soc., 77, 437–472, 1996.

Large, W. G., McWilliams, J. C., and Doney, S. C.: Oceanic vertical mixing: a review and a model with a nonlocal boundary layer parameterization, Rev. Geophys., 32, 363–403, 1994.

Large, W. G., Danabasoglu, G., and Doney, S. C.: Sensitivity to Surface Forcing and Boundary Layer Mixing in a Global Ocean Model: Annual-Mean Climatology, J. Phys. Oceanogr., 27, 2418–2447, 1997.

Lee, H. and Mellor, G. L.: Numerical Simulation of the Gulf Stream and the Deep Circulation, J. Oceanogr., 59, 343–357, 2003.

Levitus, S. E.: Climatological atlas of the world ocean, NOAA Professional Paper, 13, 1982.

Lima, J. A., Martins, R. P., Tanajura, C. A. S., Paiva, A. M., Cirano, M., Campos, E. J. D., Soares, I. D., França, G. B., Obino, R. S., and Alvarenga, J. B. R.: Design and implementation of the Oceanographic Modeling and Observation Network (REMO) for operational oceanography and ocean forecasting, Rev. Bras. Geofis., 31, 209–228, 2013.

Mitchell, M. J., Houtekamer, P., and Pellerin, G.: Ensemble size, balance and model-error representation in an ensemble Kalman filter, Mon. Weather Rev., 130, 2791–2808, 2002.

Oke, P. R. and Schiller, A.: Impact of Argo, SST, and altimeter data on an eddy-resolving ocean reanalysis, Geophys. Res. Lett., 34, L19601, doi:10.1029/2007GL031549, 2007.

Oke, P. R., Schiller, A., Griffin, D. A., and Brassington, G. B.: Ensemble data assimilation for an eddy-resolving ocean model of the Australian region, Q. J. Roy. Meteor. Soc., 131, 3301–3311, 2005.

Oke, P. R., Sakov, P., and Corney, S. P.: Impacts of localization in the EnKF and EnOI: experiments with a small model, Ocean Dyn., 57, 32–45, 2007.

Oke, P. R., Brassington, G. B., Griffin, D. A., and Schiller, A.: The Bluelink ocean data assimilation system, Ocean Modell., 21, 46–70, 2008.

Paiva, A. M. and Chassignet, E. P.: The Impact of Surface Flux Parameterizations on the Modeling of the North Atlantic Ocean, J. Phys. Oceanogr., 31, 1860–1879, 2001.

Reynolds, C., Gelaro, R., and Murphree, T.: Observed and simulated Northern Hemisphere intraseasonal circulation anomalies and the influence of model bias, Mon. Weather Rev., 124, 1100–1118, 1996.

Robinson, A. R., Glenn, S. M., Spall, M. A., Walstad, L. J., Gardner, G. M., and Leslie, W. G.: Forecasting Gulf Stream meanders and rings, Am. Geophys. Union, 70, 1464–1473, 1989.

Sakov, P., Counillon, F., Bertino, L., Liseter, K. A., Oke, P. R., and Korablev, A.: TOPAZ4: an ocean-sea ice data assimilation system for the North Atlantic and Arctic, Ocean Sci., 8, 633–656, 2012,
http://www.ocean-sci.net/8/633/2012/.

Schiller, A. and Brassington, G. (Eds.): Operational Oceanography in the 21st Century, Springer Science & Business Media, 745 pp., 2011.

Spall, M. A. and Robinson, A. R.: Regional Primitive Equation Studies of the Gulf Stream Meander and Ring Formation Region, J. Phys. Oceanogr., 20, 985–1016, 1990.

Tanajura, C. A. S. and Belyaev, K.: A sequential data assimilation method based on the properties of a diffusion-type process, Appl. Math. Model., 33, 2165–2174, 2009.

Tanajura, C. A. S., Costa, F. B., da Silva, R. R., Ruggiero, G. A., and Daher, V. B.: Assimilation of sea surface height anomalies into HYCOM with an optimal interpolation scheme over the Atlantic Ocean Metarea V, Rev. Bras. Geofis., 31, 257–270, 2013.

Tanajura, C. A. S., Santana, A. N., Mignac, D., Lima, L. N., Belyaev, K., and Xie, J.: The REMO Ocean Data Assimilation System into HYCOM (RODAS_H): General Description and Preliminary Results, Atmos. Ocean. Sci. Lett., 7, 464–470, 2014.

Thacker, W. C. and Esenkov, O. E.: Assimilating XBT Data into HYCOM, J. Atmos. Ocean. Technol., 19, 709–724, 2002.

Thacker, W. C., Lee, S. K., and Halliwell, G. R.: Assimilating 20 years of Atlantic XBT data into HYCOM: a first look, Ocean Modell., 7, 183–210, 2004.

Willis, J., Roemmich, D., and Cornuelle, B.: Interannual variability in upper ocean heat content, temperature and thermosteric expansion on global scales, J. Geophys. Res., 109, C12037, doi:10.1029/2003JC002260, 2004.

Woodruff, S. D., Slutz, R. J., Jenne, R. L., and Steurer, P. M.: A comprehensive ocean-atmosphere data set, B. Am. Meteorol. Soc., 68, 1239–1258, 1987.

Xie, J. and Zhu, J.: Ensemble optimal interpolation schemes for Assimilating Argo profiles into a hybrid coordinate ocean model, Ocean Modell., 33, 283–298, 2010.

Xie, J., Counillon, F., Zhu, J., and Bertino, L.: An eddy resolving tidal-driven model of the South China Sea assimilating along-track SLA data using the EnOI, Ocean Sci., 7, 609–627, 2011, http://www.ocean-sci.net/7/609/2011/.

Yan, C. and Zhu, J.: The Impact of "Bad" Argo Profiles on Ocean Data Assimilation, Atmos. Ocean. Sci. Lett., 3, 59–63, 2010.

Yan, C., Zhu, J., and Xie, J.: An Ocean Reanalysis System for the Joining Area of Asia and Indian-Pacific Ocean, Atmos. Ocean. Sci. Lett., 3, 81–86, 2010.