Observing system experiments over the Atlantic Ocean with the REMO ocean data assimilation system (RODAS) into HYCOM

Clemente A. S. Tanajura[1,2] · Davi Mignac[2] · Alex N. de Santana[2] · Filipe B. Costa[2] · Leonardo N. Lima[2,3] · Konstantin P. Belyaev[2,4] · Jiang Zhu[5]

Abstract

The Oceanographic Modeling and Observation Network (REMO) focuses on scientific and technological development of operational oceanography in Brazil considering both numerical forecasting and observational systems. A key component of the forecasting system is the recently constructed REMO Ocean Data Assimilation System (RODAS) into the HYbrid Coordinate Ocean Model (HYCOM). Here, RODAS is presented for the first time with its full capability. RODAS employs a multivariate ensemble optimal interpolation scheme that is able to assimilate Argo temperature (T) and salinity (S) profiles, sea surface temperature (SST) analyses, and satellite along-track or gridded sea-level anomaly (SLA) data. RODAS is presented together with a series of Observing System Experiments (OSEs), in which components of the Global Ocean Observing System (GOOS) were systematically withheld from 3-year assimilation runs over the Atlantic Ocean. Using the same initial condition from a full assimilation run, OSEs were performed from 1 January 2010 to 31 December 2012 withholding (i) only Argo data; (ii) only UK MetOffice OSTIA SST analyses; (iii) only satellite along-track altimetry data; and (iv) all observation types. These runs were also compared with the full assimilation run and the model free run to evaluate the impact of different observations into the model state. The results show that each observation type brings complementary information into the analyses. Assimilation of SST is needed to better constrain the mixed layer temperature, while assimilation of SLA mainly improves the representation of circulation by adding mesoscale-like features, such as those found in the Gulf Stream and the Brazil-Malvinas Confluence. In the subsurface, only Argo observations are able to constrain the thermohaline state. When Argo data are withheld, the quality of S is seriously compromised and becomes worse than the free run in the upper ocean. Additionally, the run withholding all observation types shows that the model state in the surface almost reaches the free run state by the end of the third year. However, below 300 m, the memory of the Argo data assimilation is longer and the quality of T and S is only degraded by 35% in comparison with the full assimilation run.

This article is part of the Topical Collection on the *10th International Workshop on Modeling the Ocean (IWMO), Santos, Brazil, 25-28 June 2018*

Responsible Editor: Tal Ezer

✉ Clemente A. S. Tanajura
   cast@ufba.br

1 Departamento de Física da Terra e do Meio Ambiente, Instituto de Física, Universidade Federal da Bahia (UFBA), Salvador, Brazil

2 Oceanographic Modeling and Observation Network (REMO), Geophysics and Geology Research Center, Federal University of Bahia (CPGG/UFBA), Salvador, Brazil

3 Ocean Modeling and Data Assimilation Division, Euro-Mediterranean Center on Climate Change (CMCC), Bologna, Italy

4 Shirshov Institute of Oceanology, Russian Academy of Sciences (SIORAS), Moscow, Russia

5 Institute of Atmospheric Physics, Chinese Academy of Sciences (IAP/CAS), Beijing, China

1 Introduction

Data assimilation methods are key tools for scientific investigation of the ocean and atmosphere circulation, for weather and climate prediction and for operational oceanography, among other applications. Data assimilation methods seek optimal corrections of the dynamical model state by combining the model state with observations (Daley 1991). They produce the so-called analysis, which should better represent the physical state and variability of the system with smaller deviations with respect to observations than the model. The analysis has been historically used as initial condition for the atmosphere and ocean forecast models, since the model predictability strongly depends on the quality of the initial condition (Kalnay 2003). Motivations for further developments on data assimilation are climate studies, implementation of monitoring systems, identification of the relative importance of different observational data, and others (Kalnay et al. 1996; Oke and Schiller 2007).

The majority of the data assimilation methods can be divided into two large categories, namely statistical, which includes the ensemble Kalman filter and the optimal interpolation scheme (Kalnay 2003; Evensen 2006), and variational, which are represented by the three- and four-dimensional variational schemes (Weaver et al. 2003; Moore et al. 2011). However, hybrid schemes that combine ideas from both categories (Lorenc 2003; Penny et al. 2015) and coupled ocean-atmosphere data assimilation schemes are under development (Lea et al. 2015; Penny and Hamill 2017) and should open new avenues in the very active and relatively young data assimilation research area.

Despite the existence of very sophisticated data assimilation systems, including the ones running in operational mode in ocean-atmosphere forecasting centers (Bell et al. 2015; Schiller et al. 2018), the implementation from scratch and realization of basic schemes is still not trivial. In order to succeed, one should overcome difficulties related to data quality control, the large dimension of the discrete physical system, parallel programming, and the necessary numerical and computational approximations in the data assimilation algorithm to ultimately realize the assimilation into a specific numerical model. The data assimilation system should be able to produce a multivariate correction of the dynamical model in different regions with completely different physical characteristics and variability, as well as extrapolate information from regions with observations to regions without observations.

In 2007, the Oceanographic Modeling and Observation Network (in Portuguese, Rede de Modelagem e Observação Oceanográfica—REMO) was established to promote scientific and technological development on operational oceanography in Brazil considering the country forecasting and observational needs (www.rederemo.org) (Lima et al., 2013). Among REMO's main goals was the implementation of a high-quality short-range operational forecasting system based on the HYbrid-Coordinate Ocean Model (HYCOM) (Bleck 2002, Chassignet et al. 2009) and a nested grid strategy to resolve the ocean state and the circulation of the Atlantic Ocean, the South Atlantic Ocean, and sub-regions along the Brazilian shore. Positive preliminary results led to REMO's admission into the international project GODAE OceanView (Bell et al. 2015) in December 2010, the continuation of the Global Ocean Data Assimilation Experiment (GODAE) project (Bell et al. 2009). A simplified optimal interpolation scheme was implemented operationally in the Brazilian Navy Hydrography Center (CHM) in January 2010 to produce 5-day forecasts in a daily basis with focus on the South Atlantic (Melo et al. 2013). Upgrades with an ensemble optimal interpolation (EnOI) scheme were made in January 2014 and another upgrade will be implemented until the end of 2019 with a version of the system introduced here along with an improved HYCOM configuration.

This paper presents for the first time a complete description of the recently constructed REMO Ocean Data Assimilation System (RODAS) into HYCOM together with observing system experiments (OSEs). The goals of the paper are to demonstrate the skills of RODAS and to estimate the relative importance of different observations in the forecasting system with the OSEs. Due to computational constrains, the results will concentrate in the large-scale domain that covers most of the Atlantic Ocean and provides lateral boundary conditions to other higher resolution grid over the South Atlantic (Lima and Tanajura 2013). Preliminary results were presented in Tanajura et al. (2014) with a RODAS version that was not consolidated at that time. More recently, the HYCOM+ RODAS 30-day predictability was assessed in Carvalho et al. (2019) with the first RODAS version.

RODAS is based on the EnOI scheme and is designed to consider the specificities of HYCOM hybrid vertical coordinate system. RODAS is able to assimilate sea surface temperature (SST) analyses, vertical profiles of temperature (T) and salinity (S), and satellite along-track or gridded sea level anomaly (SLA) data. The RODAS component that assimilates T/S vertical profiles has been presented in details in Mignac et al. (2015). That work was the basis for the construction of RODAS and provided key information about the impact of the assimilation of Argo data in the model circulation and mean dynamic topography. Important intermediate steps towards the first version of RODAS included the construction of SLA analysis followed by the realization of the Cooper and Haines (1996) scheme to project SLA increments into the model subsurface thermohaline structure (Tanajura et al. 2013, Lima and Tanajura 2013). These efforts also dealt with the investigation a coherent SLA innovation considering the quite distinct decorrelation scales of sea surface height (SSH) found the South Atlantic, as well as the offset of the model SSH with respect to observations.

Later, the SLA assimilation was complemented by the assimilation of Argo *T/S* profiles with a statistical interpolation scheme (Costa and Tanajura 2015).

The EnOI used in the first RODAS version follows the technique presented in Evensen (2003), Oke and Schiller (2007), Xie and Zhu (2010), and Xie et al. (2011), in which the multivariate model error covariance matrix is estimated by the mean of several co-variance matrices obtained by an ensemble of model states previously calculated. The matrix is calculated for each assimilation time to capture the high-frequency variability. Its realization is much cheaper than the genuine ensemble Kalman filter, in which a new ensemble by the forecast model is produced for each assimilation step to capture the so-called error of the day (Kalnay 2003). Because of its relatively low computational cost and good quality, the EnOI is more feasible for operational purposes (Oke and Schiller 2007; Counillon and Bertino 2009). The EnOI has also been recently used by Backeberg et al. (2014) to assimilate SLA into HYCOM as part of an effort to build a regional operational forecasting system for the greater Agulhas Current System.

The present work describes details of RODAS in Section 2, along with the model configuration, the operational forecasting system, and the data employed in the assimilation runs. Here, RODAS assimilated SST analysis produced by the UK MetOffice, Argo *T/S* profiles, and satellite along-track SLA data from Archiving, Validation and Interpretation of Oceanographic Satellites (AVISO). The OSEs are described in Section 3. They cover the period from 1 January 2010 to 31 December 2012. In addition to the full assimilation run with all observations and the model free run without assimilation, the OSEs consider runs withholding SST analyses, Argo *T/S* data, and along-track SLA data. The results are discussed in Section 4 and the conclusions are in Section 5.

2 RODAS

2.1 The ocean model

RODAS was implemented into HYCOM considering its generalized vertical coordinate system. The model allows the use of three different vertical coordinates. It employs isopycnal coordinates for the open stratified ocean, which reverts to terrain-following sigma coordinates in shallow coastal regions, and z-level coordinates in the mixed layer over unstratified ocean regions. The coordinate system choice is adjusted dynamically to the best option according to the ocean characteristics in each region around a prescribed potential density reference (Bleck 2002). HYCOM solves five prognostic equations associated with the shallow water physics: two for the horizontal motion, one for mass balance, and two for the conservation of thermodynamic tracers that can be salinity,

potential temperature, and potential density. In the present work, advection and diffusion of temperature and potential density were employed, so that salinity is diagnosed from the equation of state.

The present paper describes the first version of RODAS implemented on a basin scale eddy-permitting HYCOM grid with approximately 1/4° of horizontal resolution that covers almost all the Atlantic Ocean, from 78° S to 50° N and from 98° W to 21° E, excluding the Pacific Ocean and the Mediterranean Sea. The 1/4° resolution remains constant in longitude, but varies in latitude attaining higher resolution towards the poles. The number of grid points in each of the 21 vertical layers is 480 in the zonal direction and 760 in the meridional direction. The vertical discretization is set to 19.50, 20.25, 21.00, 21.75, 22.50, 23.25, 24.00, 24.70, 25.28, 25.70, 26.18, 26.52, 26.80, 27.03, 26.22, 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 ocean surface. The vertical mixing scheme is the K-profile parameterization (KPP) (Large et al., 1994). The lateral model bathymetry was interpolated from the Earth Topography 1 (ETOPO1) with 1-min resolution. After interpolation, a few adjustments were made in the model bathymetry. The interpolated bathymetry considered the Patos-Mirim lagoon system in southern Brazil as an ocean area, and the thin strip of land that separates the lagoons from the ocean was reestablished. The other adjustment was to restore Trinidad and Tobago as an island. On the boundaries, relaxation to climatological temperature and salinity was applied considering the outermost 10 grid cells and the time scale of 30 days. 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 south of South Africa along 20° E, and eastward flux of 120 Sv from the latter region until Antarctica. This approach, in particular at the northern boundary, provided good results with isopycnal models in previous simulations (Gabioux et al. 2013). Mediterranean inflow is also simulated through a relaxation zone around Gibraltar in order to maintain salt balance in long-term simulations together with river outflow. The latter is implemented via a precipitation equivalent. A simple ice model resolves ice extent and ice thickness in the Antarctic region. More details about the model configuration can be found in Gabioux et al. (2013).

After a 30-year spin-up integration with climatological fluxes, the model free run was initialized. It was forced by the National Centers for Environmental Prediction (NCEP) Coupled Forecasting System Reanalysis (CFSR) from 1 January 2002 to 31 December 2012 at each 6 h. The forcing consisted of 2-m air temperature and mixing ratio, 10-m winds, net shortwave and longwave radiation fluxes, and precipitation. In order to minimize the deviation of the model sea

surface temperature (SST) and sea surface salinity (SSS) from climatology, which were observed to increase in preliminary runs, relaxation of SST and SSS to the World Ocean Atlas 2001 monthly mean climatology (Conkright et al., 2002) was also included in the surface forcing, with restoring time scale of 30 days.

The HYCOM configuration presented above is the large-scale component of the REMO Ocean Forecasting System that runs operationally in the Brazilian Navy Hydrographic Center (CHM) (http://www.mar.mil.br/dhn/chm/meteo/prev/modelos/hycom-v.htm). In addition to the 1/4° grid, the operational system is composed of two other nested grids with 1/12° and 1/24° resolution. They focus on the so-called Metarea V, from 35.5° S to 7° N and west of 20° W until the Brazilian coast, and on the Southwest Atlantic, from 12° S to 32° S and from 34° W to 54° W. All three grids are configured with the same vertical discretization.

2.2 The data assimilation scheme

The analysis X_a according to EnOI is given by the formula (Evensen, 2003)

$$X^a = X^b + K\left(Y - HX^b\right) \qquad (1)$$

where $X^b \in \mathbb{R}^N$ is the model background or prior state with dimension N, K is the gain matrix, Y is the vector of observations, $Y \in \mathbb{R}^{N_{\text{OBS}}}$, and HX^b is the projection of the prior onto the observational space with dimension N_{OBS} by the observational operator, H. The gain matrix, K, is calculated from the equation

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

where α is a scalar that can tune the magnitude of the analysis increment assumed here to be equal to 0.3, $\boldsymbol{\sigma}$ denotes the localization operator, the symbol ∘ denotes the Schur product, B denotes the co-variance matrix of the model error, R is the variance diagonal matrix of the observational error, and the superscript ^{T} represents the transpose of a vector or matrix.

The EnOI formulation for B is

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

where $A' = [A'^1 A'^2 ... A'^M]$, $A'^k = \left(X^k - \frac{1}{M}\sum_{m=1}^{M} X^m\right)$, $X^k \in R^N$, is the model state vector of the k-th ensemble member, $k = 1$, M, and $M = 126$ 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 describing the spatial correlations and the anisotropic nature of ocean

circulation, keeping the analysis dynamically consistent and reducing the computational cost.

The EnOI and EnKF schemes are sensitive to the ensemble size (Evensen 2003; Oke et al. 2008; Counillon and Bertino 2009). 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. Depending on the observation, B may contain co-variances among the model variables: layer thickness (Δp), zonal velocity (U), meridional velocity (V), potential temperature (T), and salinity (S) for each model vertical layer, and the barotropic zonal velocity (U_{bar}), barotropic meridional velocity (V_{bar}), barotropic pressure (P_{bar}), and SSH. Details on how B was constructed are presented below.

2.2.1 Calculation of the innovation

The innovation of SST was calculated by linearly interpolating the model SST to the observation location. The analysis increment was calculated for the model state vector composed of (U, V, T, S) in the z-coordinate layers contained in the mixed layers; (Δp, U, V, T) in all model isopycnal layers; and U_{bar}, V_{bar}, and P_{bar}. In this approach, temperature is explicitly modified by the analysis increment and salinity is diagnosed in the isopycnal layers using the seawater state equation in order to preserve the potential density.

The assimilation of Argo data followed Thacker and Esenkov (2002) and Xie and Zhu (2010). In this approach, the T/S profile data at z-levels are projected into the model vertical space and a pseudo-observed layer thickness ($\triangle p_{obs}$) is created for each potential density specified in the model vertical discretization. The T/S data are also projected into each of these pseudo-observed layers. 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. Each Argo profile is processed as follows. Based on a pair of T/S profiles, the profile of potential density can be calculated by an equation of state for seawater. The estimated surface density from the Argo profile is compared with 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 question is repeated to the layer below. Once water with the target density is encountered, the remainder of the potential density profile can be partitioned, so that layer averages correspond to target densities until the maximum depth of the Argo profile is reached. The step functions created for $\triangle p_{\text{obs}}$, T, and S were employed in the assimilation. First, the innovation of Δp is calculated and $\triangle p_{\text{obs}}$ is assimilated to create analysis increments in (Δp, U, V) for all model layers. Second, T and S in the mixed layer

and S in the isopycnic layers were assimilated separately in a univariate way. Previous works showed that most of the correction of the T profiles was done when $\triangle p_{obs}$ was assimilated, and that assimilating S was the only effective way to constrain it (Oke and Schiller, 2007; Xie and Zhu, 2010; Tanajura et al. 2014, Mignac et al. 2015).

The innovation of along-track SLA data also employed a similar strategy to SST. However, a procedure was necessary before the calculation of the SLA innovation. The model SLA had to be first extracted by subtracting the model SSH at the analysis time from a model SSH mean. The latter was calculated from an assimilation run that considered only OSTIA SST and Argo T/S data from 1 January 2002 to 31 December 2007. As mentioned above, Mignac et al. (2015) observed a substantial reduction of the model SSH when assimilating Argo T/S data. The model had a warm bias so that the assimilation of Argo data reduced the model SSH mean in the entire domain by more than 0.1 m. After 3 years of Argo data assimilation, the system produced a stable SSH. In the present work, the SST and Argo data assimilation run before the assimilation of SLA was performed with this purpose of obtaining a more accurate SSH and large-scale circulation as well as a stable SSH to enable a reliable estimate of the model SLA and of the innovation. A second step was also imposed before the calculation of the SLA innovation to consider possible offsets between the model SLA and the observed SLA. This is expected because the model SSH mean was calculated for a period that did not coincide with the much longer 20-year period (1993–2012) employed by AVISO to estimate the observed mean dynamic topography (Rio et al. 2014). A strategy described and tested in Tanajura et al. (2013) and Lima and Tanajura (2013) was employed. First, the model SLA is interpolated to each observed along-track independently and the average along the track is taken for the observations and for the model. The offset was calculated by subtracting the observational average from the model average. Then, the model SLA was adjusted by subtracting the offset from the model SLA. The innovation was calculated with the adjusted model SLA, so that the mesoscale-like features were highlighted. However, it should be mentioned that the offset was in general quite small in magnitude, about 5 cm, because the previous assimilation of SST and Argo data produced a good adjustment of the model SSH to the AVISO mean dynamical topography. This strategy was also used in ECMWF to improve the model SSH before assimilating SLA (Martin et al. 2015).

2.2.2 The ensemble members and model error co-variance matrix

Two sets of ensemble members were employed in the present work to calculate $\boldsymbol{B}$, so that the OSE initial condition resulted from two different preliminary assimilation runs. The first preliminary run assimilated OSTIA and Argo data every 3

days from 1 January 2002 to 31 December 2007 considering an observational window of 1 day for SST and 3 days for the Argo data. The ensemble members employed in this assimilation were composed of outputs of the model free run. The choice of the ensemble members considered the intra-seasonal variability and the high-frequency model dynamics, as in Xie and Zhu (2010), Xie et al. (2011), and Mignac et al. (2015). For instance, to perform assimilation on 15 March 2002, 21 members centered on 15 March of each year from 2002 to 2007 were taken with 3 days between each member. However, RODAS is flexible to use any number of ensemble members and to select different intervals between each ensemble member. The number of ensemble members was chosen after a few sensitivity experiments considering a reasonable representation of the model anomalies without high computational cost, and is in agreement with the numbers used in other works (Counillon and Bertino 2009; Xie and Zhu 2010; Xie et al., 2011, Mignac et al. 2015).

The goals of the first preliminary assimilation run with only OSTIA and Argo T/S data were to reconstruct the model thermohaline structure, obtain a stable model SSH and a reliable large-scale circulation to allow assimilation of SLA data, and provide the ensemble members for the realization of SLA assimilation. This approach is also used in the European Center for Medium-Range to Weather Forecasts (ECMWF) to insert mesoscale features into the underlying model rather than attempting to adjust the large-scale patterns (Balmaseda et al. 2013, Martin et al. 2015). A similar strategy was also employed by Castruccio et al. (2008) to assimilate absolute dynamic topography. They performed a short spin-up run with relaxation of T/S to climatology to avoid drift of the model mean thermohaline structure and better face the assimilation of altimetric data.

It is fair to consider that the goals of the preliminary assimilation run were achieved as demonstrated by comparisons of the model results with the AVISO mean dynamic topography (MDT) and the World Ocean Atlas climatology (WOA13) Boyer et al. (2013) temperature shown in Fig. 1. To compare the AVISO MDT with the model SSH, an offset of approximately 0.19 m and 0.08 m was added to the assimilation run and the free run, respectively. The position of the Gulf Stream and its extension towards Europe characterized by a sharp SSH gradient in the northwest region of the North Atlantic are much better represented by the assimilation run than the model free run. Due to constrain imposed by model resolution and lateral boundary conditions, the Gulf Stream in the free run was more diffuse in the mid-latitudes. Also, a large ridge was produced in the west associated with a zonal extension along approximately 35° N. Close to South America, the representation of the Brazil-Malvinas Confluence around 45° S, 50° W was very much improved by the assimilation run. The latter could capture the penetration of the Malvinas Current until close to the mouth of the Plata River. Regarding the
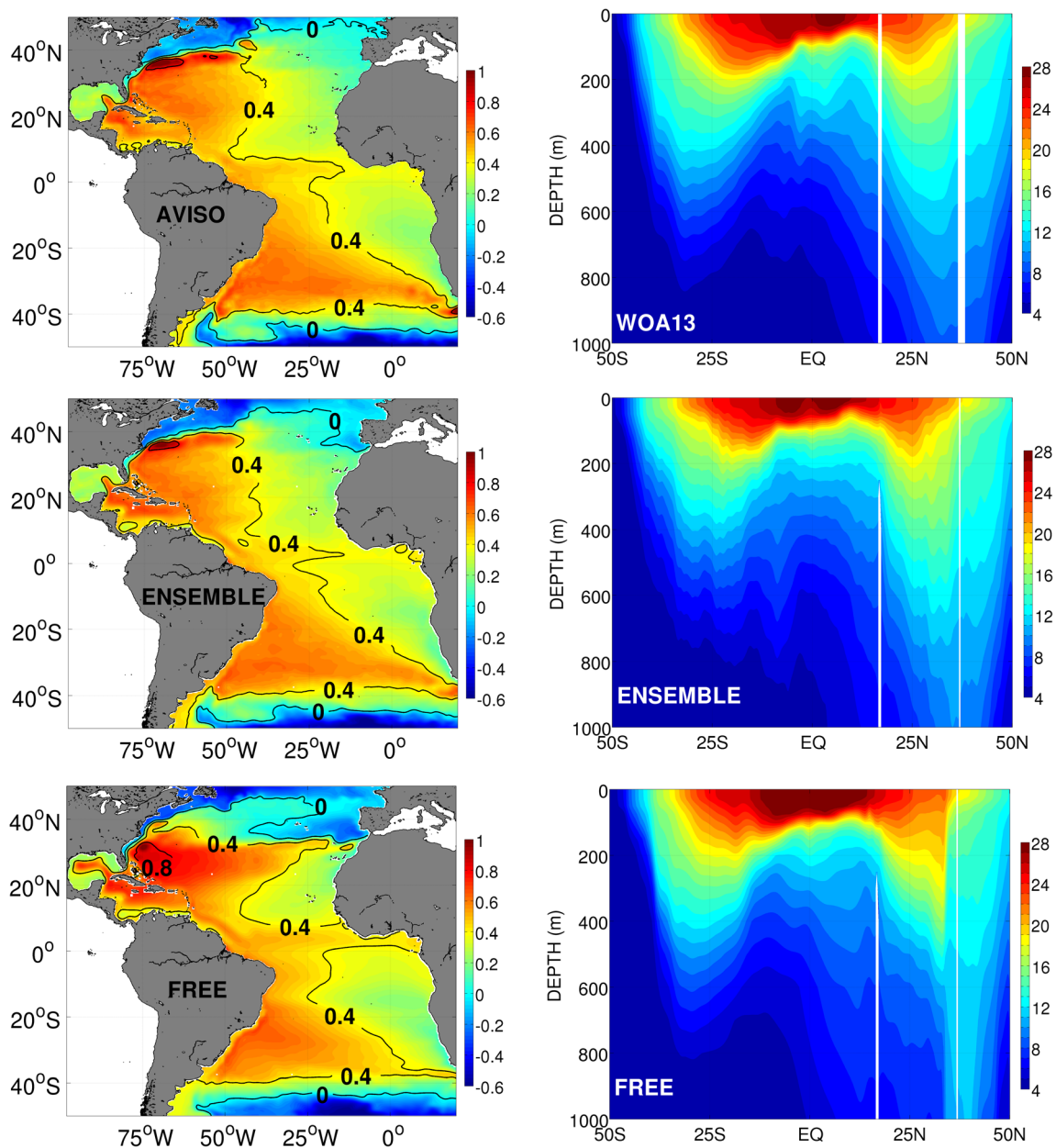
Fig. 1 The left column shows the AVISO mean dynamic topography (m) in the top from WOA13, the model SSH mean produced by the OSTIA and Argo data assimilation run in the middle, and the model SSH mean produced by the free run in the bottom. The period employed for the AVISO mean dynamic topography was from 1993 to 2012, and for the model runs was from 1 January 2002 to 31 December 2007. An offset equal to 0.19 m and 0.08 m was subtracted from the model assimilation run and free run, respectively. The right column shows the vertical section of the annual mean temperature (°C) from surface until 1000 m along 28° W according to the World Ocean Atlas (2003) climatology in the top, the model temperature mean produced by the OSTIA and Argo data assimilation run in the middle, the free run. The outputs of the OSTIA and Argo data assimilation run compose the ensemble members for the OSEs assimilation runs

thermal structure, it can be seen that the assimilation produced a more accurate mixed layer depth and thermocline particularly in the equatorial region, where the free run presents a strong warm bias in the surface. Around 35° N, the free run temperature contains the signal of the inaccurate representation of the Gulf Stream extension towards the east. There is almost a discontinuity of the isotherms in the upper 400 m at this latitude.

Considering the positive results of the OSTIA and Argo data assimilation run, an assimilation run was performed from 1 January 2008 to 31 December 2009. It was initialized with the last output of the previous run and now included assimilation of along-track SLA data in addition to OSTIA and Argo data. The goal of this assimilation run was to include higher variability structure and prepare the initial condition for the OSEs on 1 January 2010.

2.2.3 Localization

The localization operator was only applied in the horizontal domain, according to the formula by Gaspari and Cohn (1999) employed also in Xie and Zhu (2010) and Mignac et al. (2015). Here the horizontal scale of influence was defined as 75 km for all the assimilated variables so that the localization operator σ forces the model error co-variance matrix B to decrease to zero when the distance between points is 150 km or more.

Over the Metarea V, a specific study on the decorrelation scale for SLA was conducted by Lima and Tanajura (2013) to assimilate along-track SLA data. The spatial scale of decorrelation—here defined as the length in which correlation dropped to about 36%—varied from approximately 120 km in the high variability region of the Brazil Current–Malvinas (Falkland) Current Confluence (BMC) to 440 km in the central equatorial Atlantic. In the North Atlantic, based on observations and model results on the Gulf Stream, Ezer and Mellor (1997) imposed a correlation operator in which points more than 300 km apart had almost null correlation. Therefore, in the present study, the choice of null covariances for points separated by more than 150 km over the whole domain was conservative and privileged the spatial scales of high variability areas. In future works, localization radius with spatial dependence may be employed.

2.2.4 Data and observational errors

The data employed in the assimilation runs were (i) daily SST analyses from the Ocean Sea Surface Temperature and Sea Ice Analysis (OSTIA) with 1/20° resolution [available at ftp://data.ncof.co.uk/ostia_reanalysis/]; (ii) 46,788 vertical profiles of *T/S* from Argo [available at ftp://ftp.ifremer.fr/ifremer/argo/geo/atlantic_ocean/]; and (iii) along-track SLA data from the satellites Jason-1, Jason-2, and Envisat from AVISO [available at ftp://ftp.myocean.sltac.cls.fr/Core/]. The Argo *T/S* data was quality controlled considering the tests recommended by the Global Temperature-Salinity Profile Programme (GTSPP, 2010). They included tests for location, date, impossible parameter values, increasing depth, spike, gradient, temperature inversion, and climatology. Only the pair *T/S* was assimilated, i.e., if one these variables was not qualified the quality control, the pair was rejected. OSTIA SST and AVISO SLA data were not quality controlled by our system.

The observational error covariance matrix $\boldsymbol{R}$ in Eq. (2) is assumed as diagonal. It depends on the observation type, and in general includes information about the instrumentation error ($\mathcal{E}_I$), the representativeness error ($\mathcal{E}_R$) due to unresolved scales, and an error associated with the relative "age" of each observation ($\mathcal{E}_A$). The latter is necessary in order to consider the relative importance between observations made few days before the analysis time and observations at or closer to the analysis time. Therefore, the observation error variance $\mathcal{E}_O^2$ of a single observation is given by $\mathcal{E}_O^2 = \mathcal{E}_I^2 + \mathcal{E}_R^2 + \mathcal{E}_A^2$.

The instrumentation error $\mathcal{E}_I$ for SST, SLA, and the vertical profiles of T and S was assumed to be equal to 0.25 °C, 0.04 m, 0.05 °C, and 0.01 psu, respectively. The error $\mathcal{E}_R$ was considered only for SST and SLA. They were calculated according to Oke and Sakov (2008) with the construction of super-observation (superob). The value of the OSTIA SST superob was estimated considering a degraded model resolution so that a single superob value was taken for a set of 2×2 (4) grid points. This strategy was mainly used to reduce the dimension of the observational space and, consequently, the computational cost. Each SST superob value was given by the weighted average of SST data as a function of the distance to the center of the four model grid points. The value of the SLA superob for each model grid cell was given by the weighted average of the original along-track SLA data as a function of the distance to the analysis grid point. The SST and SLA $\mathcal{E}_R$ were calculated from the standard deviation of the difference between the original data and the weighted average. The ranges of the $\mathcal{E}_R$ in the present work for SST and SLA were, respectively, from 0.0014 °C and 0.0001 m in the center of the South Atlantic gyre to 2.8 °C and 0.36 m in the Gulf Stream and the Brazil-Malvinas Confluence. The error $\mathcal{E}_A$ was calculated according to Oke et al. (2008) by the formula

$$\mathcal{E}_A = \mathrm{RMS}_{\mathrm{mod}} \left[1 - \exp\left(-\frac{|t^a - t^o|}{2t_{\mathrm{ef}}} \right) \right]$$

where $\mathrm{RMS}_{\mathrm{mod}}$ is the root-mean-squared deviation of the model value around its seasonal cycle produced by the spin-sup run, t^a is the analysis time, t^o is the observation time, and t_{ef} is an e-folding time scale. This error was considered only for the Argo and the SLA data since the observational time of the OSTIA SST data roughly coincided with the assimilation time. A 3-day observational window covering the analysis day and 2 days before was employed for the Argo and the SLA data, and t_{ef} was set to 3 days.

3 The numerical experiments

In the present work, assimilation was performed every 3 days from 1 January 2010 to 31 December 2012. Assimilation of OSTIA, Argo *T/S* profiles, and along-track SLA data were realized in separate steps in the full assimilation run. The first analysis increment in this run was produced by the assimilation of only SST at 03 UTC on 1 January 2010. Three hours later, only Argo *T/S* data were assimilated, and 3 h later, only SLA data were assimilated. The same cycle was repeated each 3 days considering a 3-day observational window for the Argo *T/S* and the SLA data, and a 1-day window for OSTIA. The use of distinct steps to assimilate the available data has been

already employed (e.g., Balmaseda et al., 2008, Yan et al., 2010), but future work will attempt to produce a single analysis increment by assimilating all observations at the same time. This will require a substantial modification in RODAS strategy to assimilate Argo *T/S* data, so that assimilation of the pseudo-observed layer thickness would not be performed, and innovation would be actually calculated in the observation space.

The OSE was composed of five integrations from 1 January 2010 to 31 December 2012 forced with the NCEP CFSR atmosphere at each 6 h. They were conceived to allow investigating the influence of each observation type in the analysis and "forecasts." The runs were initialized with the same initial condition at 00 UTC 1 January 2010 produced at the end of the assimilation run from 1 January 2008 to 31 December 2009 described above, in which OSTIA SST, Argo *T/S* profiles, and along-track SLA data were assimilated. The first integration of the OSE was the full assimilation run, in which all data was assimilated each 3 days. This run utilized RODAS full capability, and it will be here called as the ALL run. The other integrations denied observation types in the assimilation algorithm. The runs without OSTIA, without Argo, without altimetry data, and without any data were called as NOOSTIA, NOARGO, NOALTIM, and NOASSIM runs, respectively. In addition to these five runs, the model free run (FREE) was also used to help evaluate the impact of the observations in the model skills. The difference between the NOASSIM run and the FREE run is basically the initial condition. The FREE run was initialized on 1 January 2008 and it was integrated until 31 December 2012 without assimilation. The NOASSIM did not perceive assimilation as well, except on its initial condition on 1 January 2010.

4 Results

The experiments were evaluated against WOA13 climatology, Argo *T/S* data, OSTIA, SLA gridded data, and absolute dynamical topography from AVISO. The altimetric data employed in the evaluation of the assimilation experiments were in delayed mode and they had the same model horizontal resolution (1/4°). Since the gridded data are also a product of assimilation, with emphasis on the observational information, no additional smoothing or filtering was used in the data for the experiments evaluation. It was assumed that both the outputs of the assimilation runs and the AVISO gridded data could approximately resolve the same features.

It should be mentioned that the analyses produced by the OSEs were not explicitly evaluated here. As mentioned above, assimilation was performed every 3 days from 1 January 2010 to 31 December 2012. On the assimilation day, assimilation of SST was performed at 00Z, Argo *T/S* at 03Z, and SLA at 06Z. The model outputs at 00Z on the following 3 days, i.e., 18 h,

42 h, and 66 h after SLA assimilation, were employed to compare the OSE results against climatology and observational data. The model output 66 h after SLA assimilation was used as background for the next SST assimilation. If atmospheric forecasts had been used in the place of the CFSR atmospheric forcing fields, the true HYCOM+RODAS short-term predictability would have been investigated by this strategy. It is expected that the errors of CFSR atmospheric fields are smaller than or equal to the errors of an atmospheric forecast, since the latter is not constrained by observations in the forecast window. In this sense, the CFSR atmospheric fields may be considered as the best possible forecast that can be produced by the Coupled Forecasting System. Therefore, HYCOM+RODAS would develop smaller errors when forced by CFSR atmospheric fields than when forced by a true forecast. Considering the quality of the atmospheric forcing, the results presented below may represent a lower bound of the HYCOM+RODAS forecast errors, since the sources of errors from the atmospheric forcing would be minimum.

4.1 Impact on temperature and salinity

The impact in the mean thermohaline structure is assessed considering the difference of the OSE runs means with respect to WOA13 climatology. The OSE runs means were taken over the 3-year period from 2010 to 2012. The difference (model minus WOA13) of the model mean temperature and salinity vertical cross section along 28° W from 50° S to 50° N up to 1000 m depth with respect to the WOA13 climatology is presented in Fig. 2 for each run. The ALL run presents the largest temperature deviations in the thermocline region in the tropics. Substantial temperature deviations larger than 1 °C are also found below 300 m in the tropics of the South Atlantic and in the mid-latitudes of the North Atlantic. The largest salinity deviations are negative and occur around 200 m depth in the tropics. In most of the domain, deviations are smaller than 1 °C and 0.1 psu. When Argo data is denied (NOARGO), there is a substantial increase in the deviations of *T* and *S* with respect to climatology in almost all latitudes. The deviation patterns are similar to the ALL run, but the magnitudes are larger. The temperature below 300 m tends to be colder than climatology by more than 1 °C from the equator until about 25 °N and from about 15° S to 50° S. The salinity tends to be smaller than climatology everywhere in the upper ocean as well as below 300 m in subtropical regions of the North and South Atlantic. The impact of the Argo data has different aspects from the ones observed by Mignac et al. (2015) and Tanajura et al. (2014), in which only Argo data was assimilated in an experiment initialized by a model free run state. In those experiments, the model free run had in general a warm bias in the top 600 m, so that the corrections imposed by Argo data assimilation cooled the model upper ocean and increased
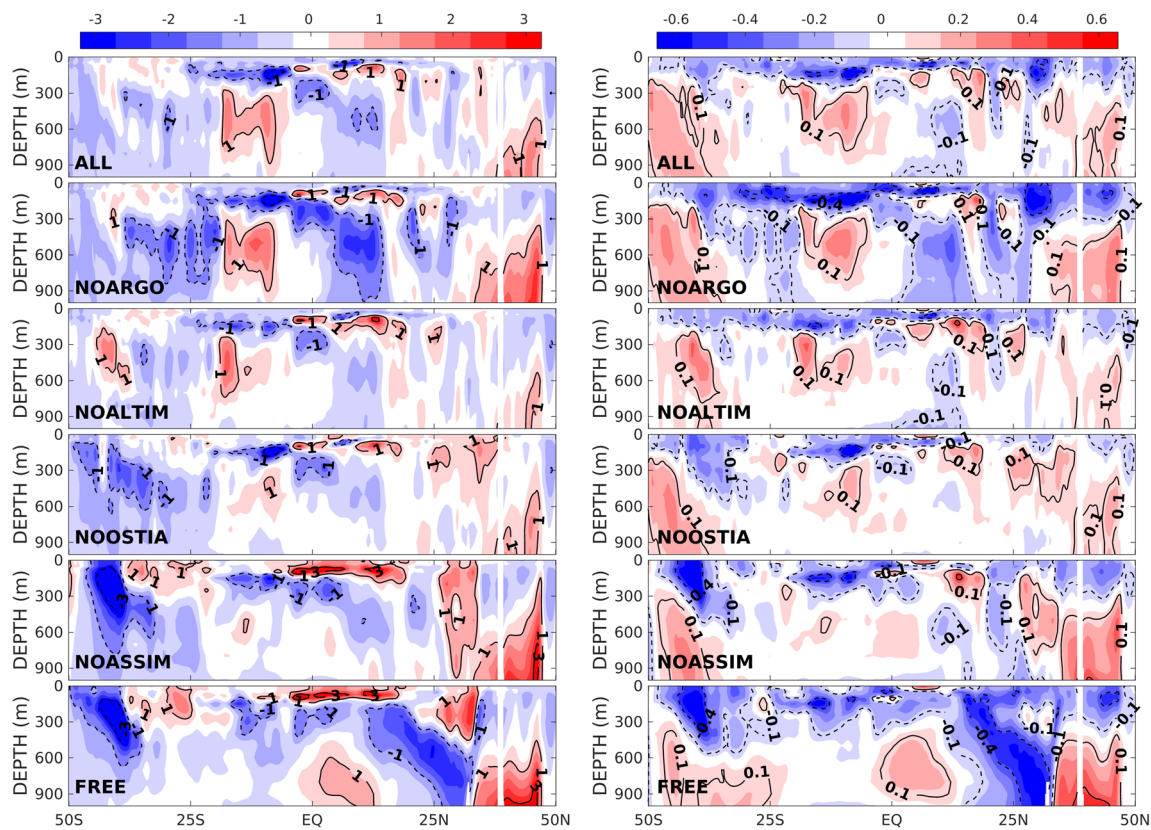
Fig. 2 Vertical cross section along 28° W in the top 1000 m from 50° S to 50° N for the difference between the mean temperature (°C) (left column) and salinity (psu) (right column) from 1 January 2010 to 31 December 2012 for each OSE run and the WOA13 climatology (OSE run minus WOA). White areas range ± 0.25 °C in the left column and ± 0.05 psu in the right column

the model mean density. Here, when Argo data was denied, temperature tended in general to decrease, except in deeper regions of the North Atlantic. Denying altimetry data and OSTIA produced similar modifications, except in the Southern Ocean, where the NOALTIM run produced an increase in temperature and the NOOSTIA run a decrease by more than 1 °C. The deviations of the NOALTIM and NOOSTIA runs are, however, smaller than the ones attained by the NOARGO run and by the ALL run in some regions, particularly for S in the upper 300 m. The importance of Argo data to constrain the model thermohaline structure is expected, but better behavior of the NOALTIM and NOOSTIA runs with respect to the ALL run in T below the mixed layer and in S in the mixed layer and below is not. This result was observed in vast regions of the domain, i.e., it is not a feature captured only by this specific cross section. This clearly shows that the assimilation of OSTIA and SLA are imposing errors in thermohaline structure that can be counteracted only by the assimilation of Argo data. When all data are denied (NOASSIM), the deviations with respect to climatology are substantially increased. The upper 100 m in the equatorial region develops a strong warm bias of more than 3 °C, and the upper 400 m in the southern ocean a cold bias of more 3 °C below the surface. In a vast region in the South Atlantic mid-

latitude, a strong cold bias and freshening is produced in the NOASSIM run. As expected, the NOASSIM state tends to the FREE run state.

A more detailed assessment of RODAS quality and observation impact in the upper ocean can be seen in Fig. 3, which shows zonal vertical cross sections of T and S mean down to 300 m along 30° N. ALL is able to substantially reduce temperature with respect to the FREE run and correct the warm bias towards climatology. Assimilation of OSTIA in the NOARGO and NOALTIM runs constrains the mixed layer temperature and the deepening of warm surface waters, which is observed when OSTIA is denied in the NOOSTIA and NOASSIM runs. The importance of SST assimilation to constrain the mixed layer temperature has been already observed (e.g., Oke and Schiller 2007; Tanajura et al. 2014, Oke et al., 2015a, b).

The impact in S is more complex. The ALL run is not able to reproduce the high salinity subtropical water, except in regions west of 40° W where values greater than 36.7 psu are observed. Also, a relatively low salinity region with values smaller than 36.2 psu below 100 m around 25° W in the ALL run is not observed in the NOALTIM, NOOSTIA, and NOASSIM runs. This is a clear indication that the assimilation of OSTIA and SLA data is damaging the salinity field in the
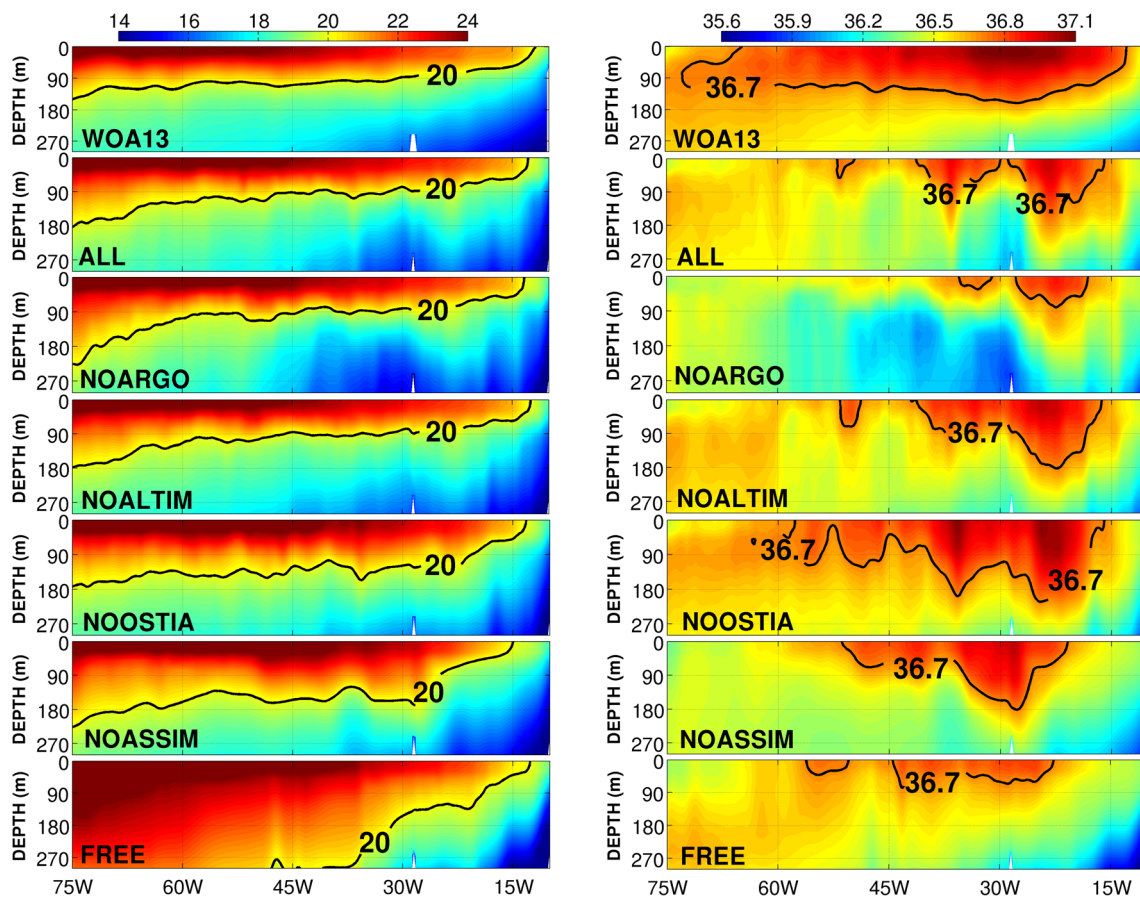
Fig. 3 Vertical cross section along 30° N in the top 300 m from 75° W to 12° W for the mean temperature (°C) (left column) and salinity (psu) (right column) according to the WOA13 climatology, the OSE runs, and the FREE run

mixed layer and below. This behavior is also verified in Fig. 2, but not in all latitudes. The EnOI produces the analysis increments depending on the quality of the ensemble co-variance. An investigation of the correlation between SST and S taken from the ensemble members showed values larger than 60% along 30° N in the surface and in regions below 100 m. For comparison, the same correlation was calculated for the HYCOM+NCODA analyses. Much weaker values were verified in this system below 100 m, except in a region between 100 and 200 m depth, and 30 and 20° W. The strong correlation between SST and S contained in the ensemble is not accurate, since there is no physical mechanism that in general links these vari-ables explicitly. In future work, this covariance should be damped by using vertical localization.

The impact of the observations in temperature and salinity can also be investigated using the root mean squared deviation (RMSD) of the daily outputs of each 3-year run with respect to the Argo T/S profiles. In this evaluation, data from 46.788 Argo profilers were employed. The daily HYCOM+RODAS results were horizontally interpolated to the position of the observational data. Then, both data and model results were vertically interpolated to the standard vertical levels employed

in the World Ocean Atlas 2009 (Locarnini et al. 2010). The squared deviation between model and observation and its mean were calculated for each level separately for all available data in the model domain, before the square root was taken. The largest T and S RMSDs in the FREE run are found around 100 m and 600 m depths (Fig. 4). Global and regional models have difficulties to represent the sharp vertical gradient of temperature found in the thermocline, since they tend to over-estimate the vertical heat diffusion and/or to misrepresent the mixed layer depth (e.g., Ezer and Mellor 1997; Xie and Zhu 2010; Balmaseda et al. 2013; Oke et al. 2015a, b; Carvalho et al. 2019). The largest RMSD for T in global and regional models with data assimilation commonly varies between 1 and 2 °C around 100–150 m depth. The largest RMSDs for S are commonly between 0.15 and 1 psu at the surface or in the mixed layer associated with erroneous freshwater fluxes and vertical diffusion (e.g., Balmaseda et al. 2013; Oke et al. 2015a, b; Tranchant et al. 2019). In the HYCOM+RODAS system, relaxation of sea surface salinity to climatology is employed in all runs, and it controls the errors in the surface. In Mignac et al. (2015), the FREE run was evaluated along with the impact of Argo data assimilation. It was observed that the FREE run could not properly capture the Mediterranean

Fig. 4 Vertical profiles of RMSD of temperature (°C) (left) and salinity (psu) (right) for the OSE runs and the FREE run with respect to daily Argo *T/S* data from surface to 1400 m. A total of 46,788 profiles were employed
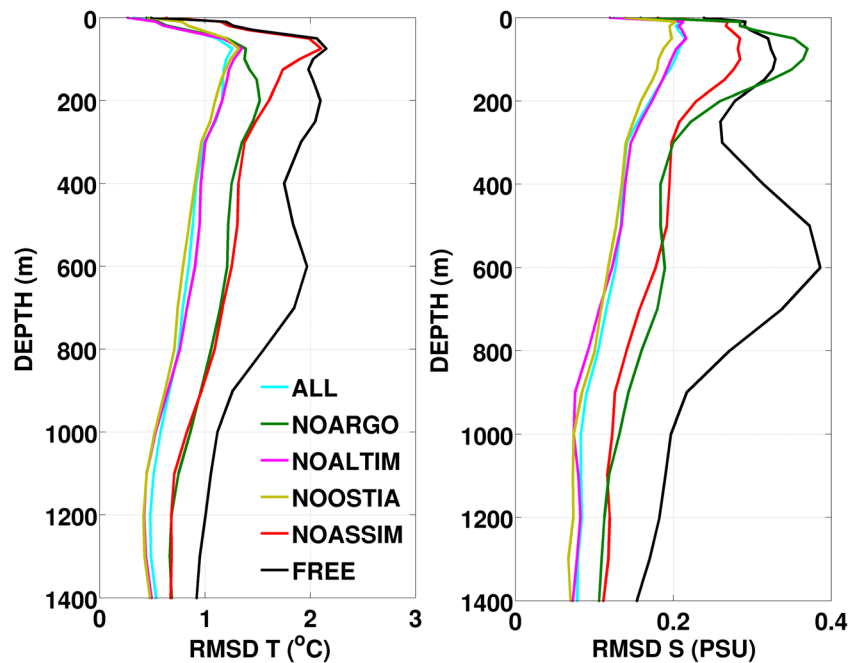
Water, and it negatively impacted the model mean RMSD for *T* and *S* around 600 m depth.

In the present work, all runs that include Argo data produce similar RMSD profiles with maxima values of about 1.2 °C and 0.2 psu at 100 m depth and minimum values of about 0.5 °C and 0.07 psu below 1000 m (Fig. 4). When Argo data is withheld, there is a substantial increase in RMSD in both *T* and *S* in the whole profile, except in *T* in the upper 100 m, mainly because of the positive influence of the SST assimilation in the mixed layer temperature. A slight increase in RMSD of *T* in the NOOSTIA run is also observed. The RSMD of *S* in the NOARGO run around 100 m depth is larger than the values attained by the NOASSIM and the FREE runs and almost reaches the maximum RMSD of about 0.4 psu produced by the FREE run at 600 m depth. This corroborates the results shown in Figs. 2 and 3, in which assimilation of SLA and SST is producing wrong corrections in *S*. As already observed by Oke and Schiller (2007), Lea et al. (2014), Tanajura et al. (2014), and Oke et al. (2015a, b), it seems that only data of vertical profiles of *S* are able to effectively constrain this variable in the models today. This highlights the importance of the Argo observational system as well as other systems that include salinity. It can also be inferred from the RMSD profiles that the NOASSIM run tends to the FREE run state in the upper ocean, but below 200 m, it roughly coincides with the NOARGO run. This shows that the influence of the initial condition on 1 January 2010 is still strong below 200 m in the 3-year mean.

Considering the vertically averaged RMSD of T and S with respect to Argo and PIRATA *T/S* data, the substantial improvement of the ALL run with respect to the FREE run is highlighted in Table 1. The vertically averaged RMSD of *T*

with respect to PIRATA and Argo drops 20% (from 1.69 to 1.34 °C) and 47% (from 1.53 to 0.81 °C), respectively. The vertically averaged RMSD of *S* with respect to PIRATA and Argo drops 6% (from 0.30 to 0.28 psu) and 46% (from 0.27 to 0.15 psu), respectively. The absence of Argo data (NOARGO) produces RMSD growth of 22% for *T* and 42% for *S* with respect to ALL. When all data is denied (NOASSIM), the errors with respect to PIRATA data attain values close to the FREE run, but the errors with respect to Argo data remain in between the ALL and the FREE. It shows that some of the memory from initial condition on 1 January 2010 remains in the thermohaline structure for 3 years. This indicates that there is potential for HYCOM+ RODAS to be used in climate studies.

4.2 Impact on SST

The ALL, NOARGO, and NOALTIM runs produce very similar RMSDs with respect to OSTIA, as shown in Fig. 5 and Table 1. The area averaged SST RSMD for the ALL and NOALTIM runs is 0.67 °C, and for the NOARGO run is 0.68 °C. This corresponds to a decrease of 50% with respect to the FREE run. In most of the domain, the deviations are less than 0.5 °C in both the North and South Atlantic, and in only very few points, the deviations reach 1 °C. The largest values are attained in the high variability areas dominated by the Gulf Stream (GS) and Brazil-Malvinas Confluence (BMC). There is a small negative impact of SLA assimilation in SST in the North Atlantic, since RMSDs are slightly smaller in the NOALTIM run than in the ALL and NOARGO runs, but this is not observed in the South Atlantic. The small degradation of SST due to SLA assimilation was also

Table 1 For each OSE run, this table presents the root mean squared deviation (RMSD) of SST with respect to OSTIA SST analyses, correlation (CORR) of SLA with respect to AVISO gridded data, and the vertically averaged RMSD of the vertical profiles of *T/S* with respect to PIRATA and Argo data

	RMSD OSTIA (°C)	CORR AVISO	RMSD PIRATA *T* (0–500 m)/*S* (0–120 m)		RMSD Argo (0–1400 m)	
			T (°C)	*S* (PSU)	*T* (°C)	*S* (PSU)
ALL	0.67	0.61	1.34	0.28	0.81	0.15
NOARGO	0.68	0.60	1.43	0.35	0.99	0.21
NOALTIM	0.67	0.47	1.33	0.28	0.81	0.14
NOOSTIA	0.93	0.61	1.39	0.27	0.82	0.14
NOASSIM	1.19	0.42	1.69	0.29	1.15	0.19
FREE	1.36	0.41	1.69	0.30	1.53	0.27
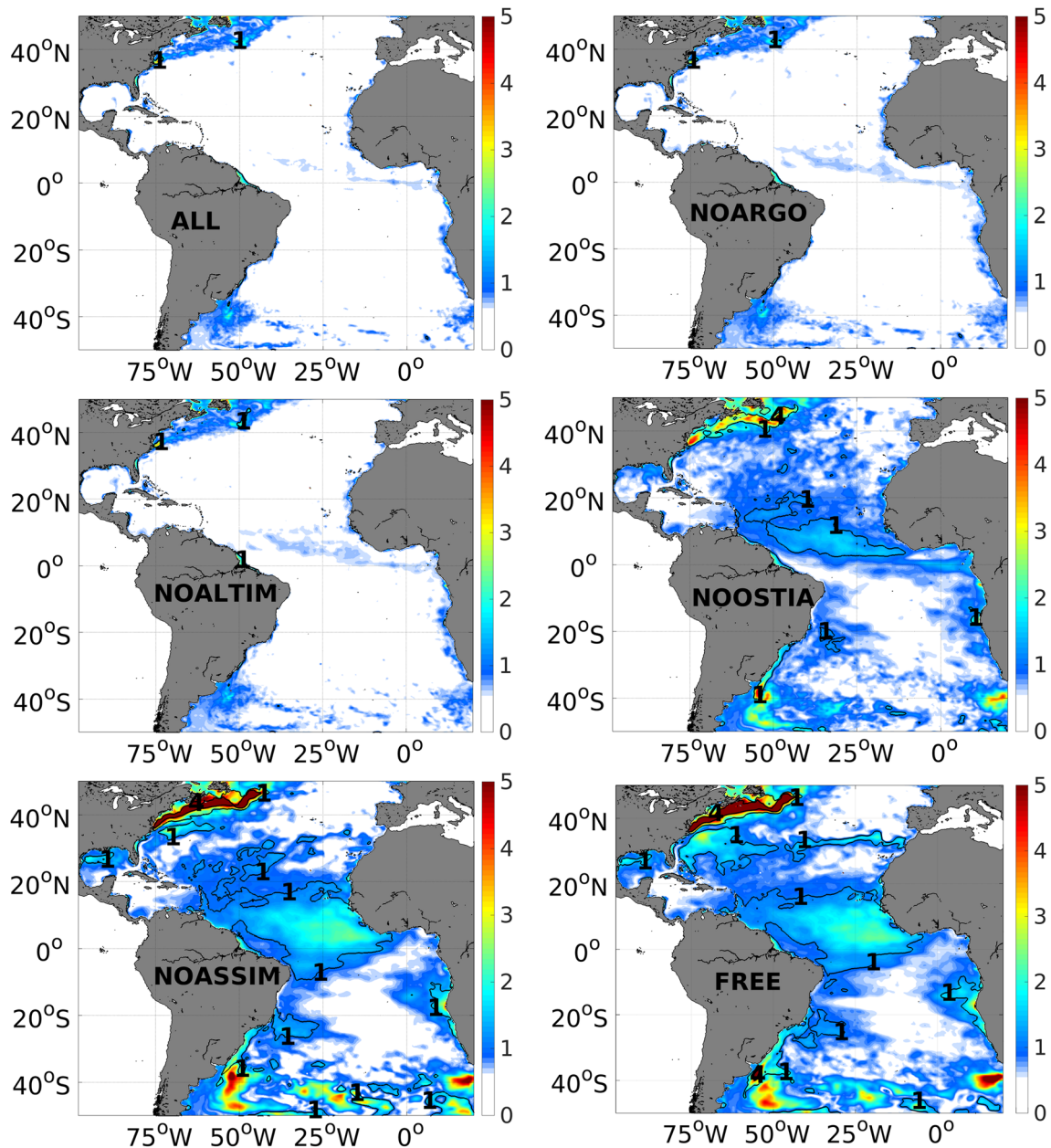
Fig. 5 RMSD of SST (°C) for each OSE run and the FREE run with respect to OSTIA. The 1 °C and 4 °C contour lines are also presented

observed by Backeberg et al. (2014) in the Agulhas Current System. Denying SST produces a substantial increase of SST RMSD in almost all the domain, including the equatorial region. But the largest deviations are observed in the GS, the BMC, and in the Agulhas Current region, with values larger than 3–4 °C. The area averaged SST RMSD for the NOOSTIA run is 0.93 °C (Table 1), which corresponds to an increase of 39% with respect to the ALL run. The RMSDs obtained by the NOASSIM and FREE runs are very similar in pattern and much larger than the NOOSTIA run. It means that the SST of the NOASSIM run converged very quickly to a pattern similar to the FREE run state. Also, assimilation of Argo data positively contributed to correct SST, despite the local aspect of the analysis increments.

The time evolution of the SST RMSD averaged between 50° S and 50° N for all runs contains a strong seasonal cycle with peak in February–March as shown by Fig. 6. The FREE run deviation oscillates between about 1.1 and 1.7 °C, while the ALL, NOARGO, and NOALTIM deviations oscillate between 0.5 and 0.9 °C, i.e., the magnitude and the range of the deviations are substantially reduced when SST is assimilated. The deviations are dominated by values attained in the GS, BMC, and Agulhas Current, but the maxima are attained in the northwest Atlantic, where the model FREE run is not able to simulate the main position and variability of the GS. When evaluating the area averaged RMSD for the FREE run SST for the North and South Atlantic separately, the deviation varies along the year from about 1.3 °C in the boreal summer to 3.4 °C in the boreal winter for the North Atlantic and from about 1.3 °C in the austral summer to 2.1 °C in the austral winter for the South Atlantic (not shown). This indicates the model FREE run has difficulties to simulate the boreal and austral winters, when the large-scale meridional SST gradients are greater than in the summer, and the western boundary currents intensify and reach higher latitudes. When area average is taken over the whole domain, the GS region imposes the peak of the SST RSMD seasonal variability. This issue will be further discussed below along with aspects of the simulated circulation.

Figure 6 shows that the SST RMSD increases very quickly from about 0.7 to 1.1 °C in 3 months when OSTIA is withheld, and from about 0.7 to 1.2 °C when all data are withheld. Part of this increase is due to the strong seasonal cycle of the error presented in all runs. After 1 year, the RMSD of the NOASSIM run gets very close to the FREE run. It is worth noting that SST RMSD in the NOOSTIA run lies in between the ALL and the NOASSIM runs along the whole integration period. This result is consistent with Fig. 5 and the conclusion that Argo data produced a general positive correction of SST. This reinforces results obtained by other works (e.g., Oke and Schiller 2007; Costa and Tanajura 2015).

4.3 Impact on SSH

Impact on SSH is evaluated by calculating correlations between the OSEs and the AVISO absolute dynamic topography. This metric has the advantage to avoid the offsets between the model SSH mean and the MDT from AVISO, and focus on the SLA mesoscale-like variability relative to each reference surface. The ALL, NOARGO, and NOOSTIA runs correlations are very similar, 0.61, 0.60, and 0.61, respectively, as shown in Fig. 7 and Table 1, and about 48% larger than the correlation of the FREE and NOAASSIM runs. This indicates that some of the observed mesoscale activity was effectively introduced in the model by assimilation of SLA. High correlation values are attained in large parts of the domain, particularly in low variability regions of the tropical Atlantic, as well as in higher latitudes in the eastern North Atlantic. Lower correlations were produced in regions of high mesoscale activity of the GS, BC, and in the mid-latitudes of the whole South Atlantic. This is expected, since the model resolution is eddy-permitting, but not eddy-resolving. The smallest wavelengths resolved by the present model configuration are about 125 km in the tropics and 100 km in the mid-latitudes. However, considering the first baroclinic Rossby radius of deformation as a reference to ocean mesoscale, mesoscale eddy wavelengths vary from more than 230 km in the equatorial region to less than 20 km in the mid-latitudes. Therefore,

Fig. 6 Time series of the daily SST RMSD (°C) for each OSE run and the FREE run with respect to OSTIA from 1 January 2010 to 31 December 2012 considering the model domain from 50° S to 50° N
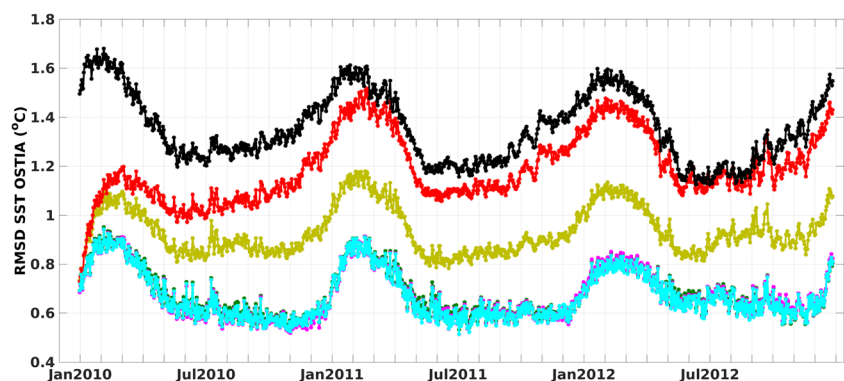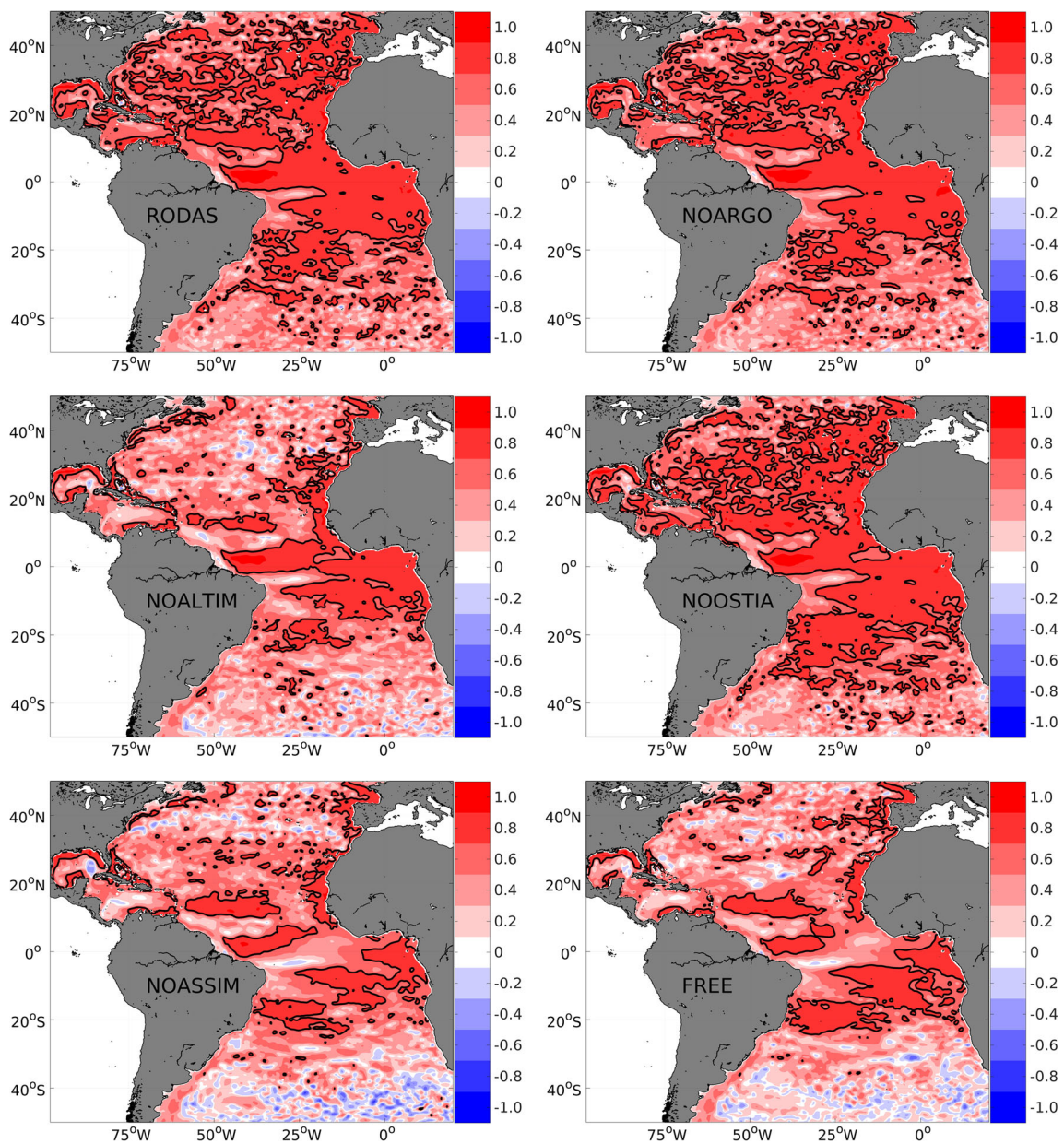
Fig. 7 Correlation between SSH from AVISO and each OSE run for the period 1 January 2010 to 31 December 2012 considering the model domain from 50° S to 50° N. The black contour line corresponds to correlation equal to 0.7

mesoscale eddies are not resolved by the present model configuration. This is a strong limitation, because important processes controlled by mesoscale and submesoscale eddies are completely absent, and this leads to systematic errors in the model mean state and variability. For instance, Su et al. (2018) show strong evidence based on global ocean model simulations that only high-resolution models—resolving submesoscale eddies with wavelengths between 10 and 50 km—can accurately produce upward heat transport from the cold deep waters to the warm surface waters, particularly in the mid-latitudes. Submesoscale and mesoscale eddies are not only a key component of the global heat transport. They are also crucial, for instance, in the cascade of energy by

transforming available potential energy into eddy kinetic energy, and in the formation of water masses by modulating isopycnic surfaces motions as a direct response to wind stress and Ekman pumping (Su et al. 2014, Su and Ingersoll, 2016).

Despite the lack of mesoscale eddies in the present work, in comparison with the NOALTIM run, the runs that included SLA assimilation provide much better results in the high variability regions and highlight the importance of this observation to correctly constrain the mesoscale-like variability. This impact is also clearly observed by other OSE studies, such as Oke and Schiller (2007) with focus in the Australian region and Lea et al. (2014) over the globe. In the present work, when altimetry data is denied, the high correlation regions
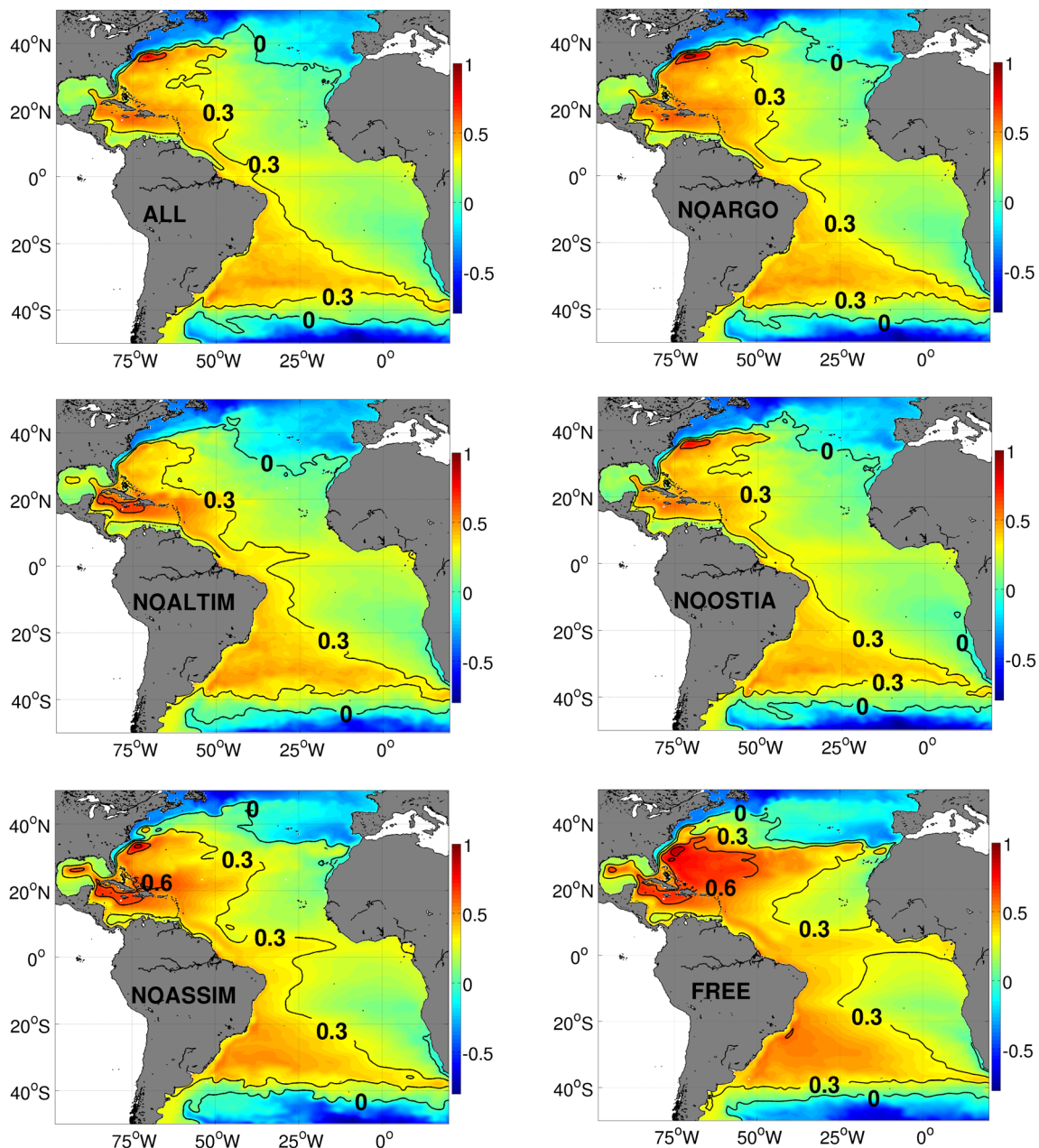
Fig. 8 Mean SSH (m) for each OSE run and the FREE run for the period 1 January 2010 to 31 December 2012

substantially diminish, particularly in the North Atlantic, and negative correlations show up in the BMC, the SAC, and the GS extension. The NOALTIM and NOASSIM runs are similar to each other, except in the higher latitudes of the South Atlantic. It shows that some positive influence of Argo data and/or OSTIA is occurring in the region.

Despite similarities of the correlation patterns produced by the ALL and the NOARGO runs, there are important differences in SSH over the western tropical North Atlantic. Figure 8 presents the mean SSH for all OSEs and the FREE run. This figure could be compared with the AVISO mean dynamic topography (MDT) presented in Fig. 1, if the offsets

among the reference surfaces are considered. For instance, the offset of approximately 0.19 m was found between the ALL run and the AVISO MDT. Figure 8 shows that denying Argo data developed higher SSHs west of 40° W from about 10° N to 35° N. It is a signal of the model tendency to overestimate SSH in this region, as presented in the NOASSIM and FREE runs. These runs simulate values greater than 0.6 m in the western North Atlantic than the ALL run. This behavior was clearly identified by Mignac et al. (2015). In this work, a similar model free run configuration showed a strong warm bias in the North Atlantic in SST and in the upper 300 m associated with SSH positive bias. This bias was remarkably

Fig 9 Time series of the daily area averaged difference of SSH (m) between each OSE run and the ALL run (OSE runs minus ALL) considering the model domain from 50° S to 50° N
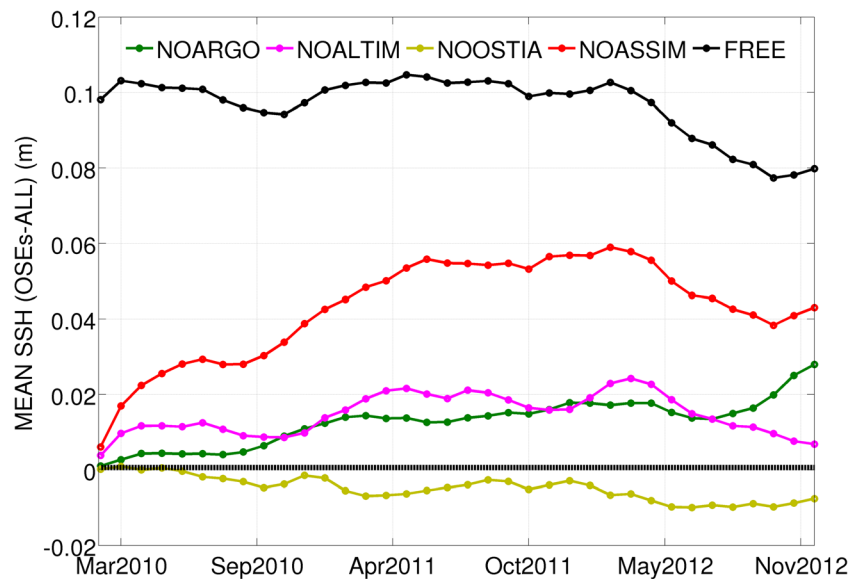
reduced when Argo data was assimilated, particularly due to relatively high number of Argo profilers in the region. Denying SLA data does not substantially change the SSH large-scale pattern as desired, but it does contribute to smooth the SSH gradient in the GS and BMC regions. Assimilation of SLA is crucial to maintain the SSH gradient and, consequently, the magnitude of the surface velocity, as presented below. The NOASSIM run in comparison with the ALL run shows a substantial drift of the SSH mean towards the model FREE run. For instance, the sharpest SSH gradient associated with the GS shifts towards the south and the high SSH region in the equator extends further east.

To obtain a broader view of the observations impacts on SSH, the daily area averaged SSH was assessed. The SSH produced by the ALL run was subtracted from each OSE along the experiment period, and the difference is presented in Fig. 9. The FREE run creates higher SSHs than the ALL run by about 0.09 m along most of the integration. This is caused by the warm bias of the FREE run, mainly in the western North Atlantic, and constrains imposed by Argo data assimilation to substantially reduce this bias. The integration window is relatively short to infer about the long-term impact of withholding part of the observations. However, for the window of few months, the results show quite distinct impacts of SST, Argo, and SLA in SSH. Denying altimetry data (NOALTIM) produces a relatively sharp increase of about 0.01 m in the first few months of integration, but SSH growth seems to stabilize after that and in the end of the integration the difference is relatively small. Therefore, it seems that SLA is not substantially changing SSH in the integration period. When Argo data is withheld, there is a smooth SSH increase relative to the ALL run in the first year. After that, the NOARGO run reaches the NOALTIM SSH, and remains smaller than or equal to this SSH. However, the slow SSH

increase in the NOARGO run is maintained during most of the integration and the raise is sharper in the last months, reaching almost 0.03 m. This behavior shows again the importance of Argo data to constrain the model SSH along with its thermohaline structure and corroborates the strategy to perform assimilation of Argo and SST before SLA to produce a stable and more accurate SSH mean. Denying SST slightly reduces the area averaged SSH in the first 2 years, and after that the reduction is more pronounced, but still relatively small. This shows that SST has a small influence in SSH when SLA and Argo data are assimilated.

4.4 Impact on circulation

The strong influence of Argo data on SSH anticipates their importance in the representation of the mean surface currents. The GS is among the features that are not well represented by the FREE run, mainly because of model limitations associated with low horizontal resolution and simplified lateral boundary conditions. Therefore, it is expected that RODAS will be able to produce large positive corrections there. This expectation is fulfilled as shown by Fig. 10. It contains the mean surface currents in the top 30 m according to the Ocean Surface Current Analyses Real-time (OSCAR)—produced by the Earth and Space Research Institute (www.esr.org)—and to the OSEs and the free run. The GS veers towards the east/northeast around 37° N, meanders around 40° N, and diverges around 45° W, 43° N. Similar patterns are reproduced by the ALL, NOARGO, NOALTIM, and NOOSTIA runs, but there is clear reduction of the velocity magnitude in the NOALTIM run. This is consistent with the reduction of the magnitude of the SSH gradient around 35° N, 65° W with respect to the ALL run. The mean surface currents of the NOASSIM run display two branches of the GS close to the
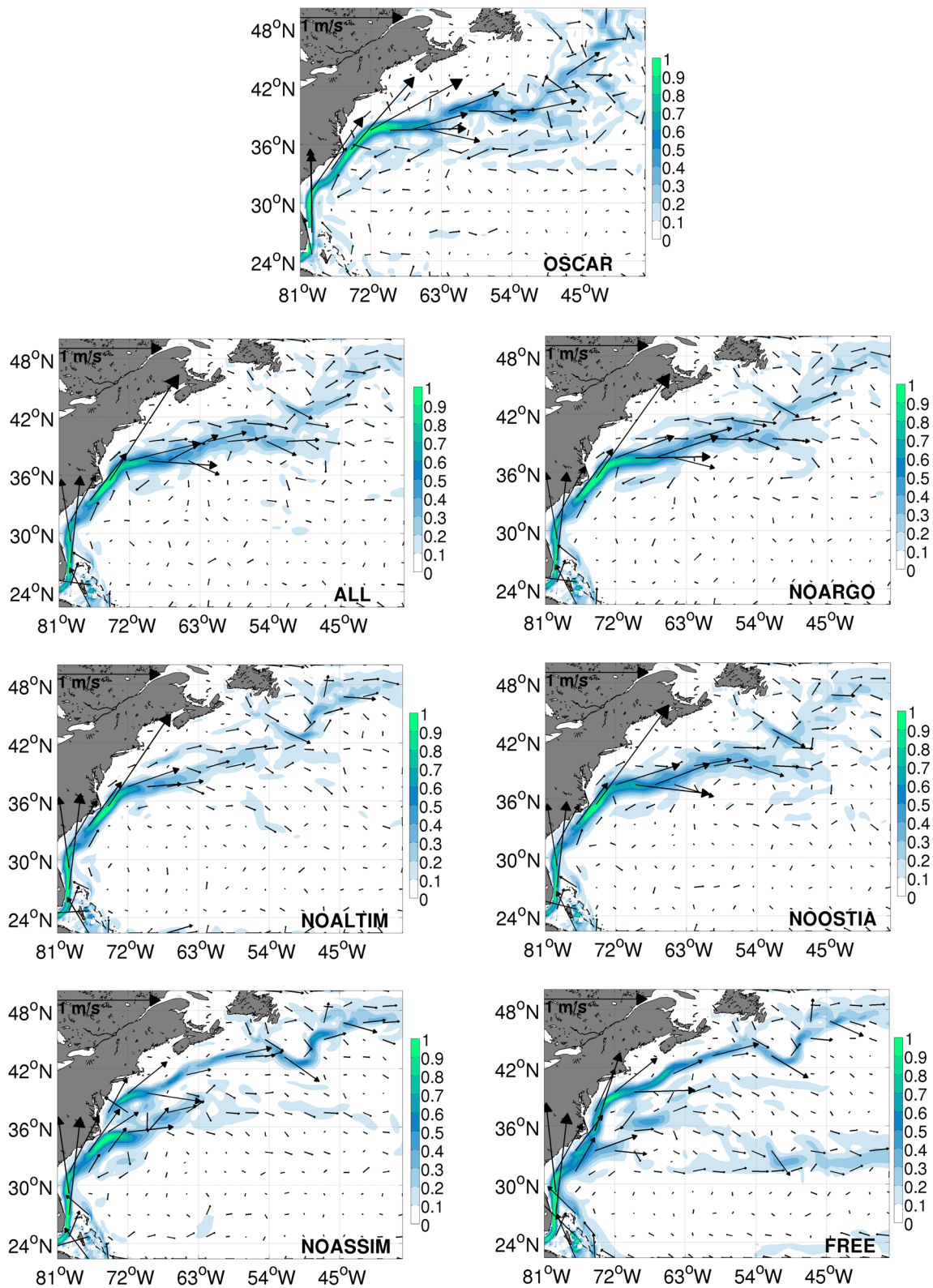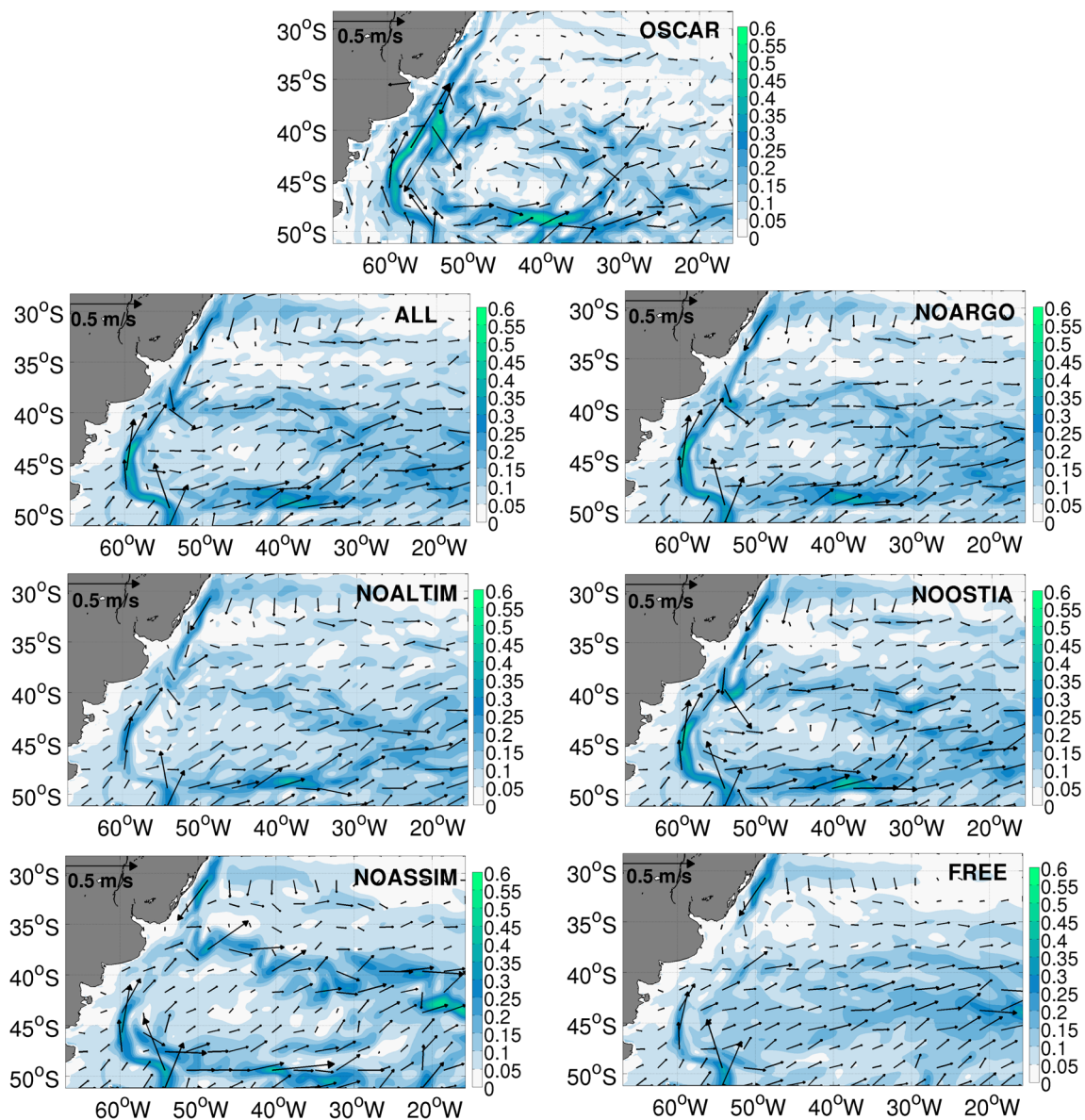
Fig. 10 Mean surface currents in the top 15 m in the Gulf Stream and adjacent areas from 1 January 2010 to 31 December 2012 according to OSCAR and each OSE run. The color bar represents the speed (m/s). The velocity vector scale is 1 m/s

coast in which the northernmost branch flows along the North America coast until higher latitudes and veers towards

northeast following very closely the continental shelf break south of the Canadian Island of Newfoundland. The

Fig. 11 Mean surface currents in the top 15 m in the Brazil-Malvinas Confluence and adjacent areas from 1 January 2010 to 31 December 2012 according to OSCAR and each OSE run. The color bar represents the speed (m/s). The velocity vector scale is 0.5 m/s

NOASSIM run shows a pattern that is clearly tending to the FREE run inaccurate representation. It is interesting to notice that despite the reduction of the GS flow produced by the NOALTIM run, all observations positively contribute to the accuracy of the large-scale pattern of the GS. As in Oke and Schiller (2007), the Argo and SST data are able to constrain large-scale circulation, but only SLA is able to contribute to the accurate representation of velocities in regions dominated by mesoscale activity.

Similar results to the GS region are obtained in the BMC region (Fig. 11). The OSE runs have a more accurate representation of the BCM than the FREE run, and the magnitude of the NOALTIM surface currents is smaller than the ALL, NOARGO, and NOOSTIA runs. The northward Malvinas

current flowing along the Argentinean shelf break and the southward BC converge around 40° S, 55° W in the ALL, NOARGO, NOALTIM, and NOOSTIA runs. This convergence is close to the position displayed by OSCAR. However, the latter shows more complex structures than the model runs, such as the zonal shear of velocities produced by the Malvinas Current flowing to the west of the BC around 40° S, 55° W and the closed circulation of the Zapiola Anticyclone. RODAS was applied into a relatively low-resolution model and is unable to produce these fine scale structures. The FREE run has a broader BMC region and no signal of the Zapiola circulation.

Observation impacts in the GS and in the BMC regions are complemented by assessing the vertical structure of the zonal
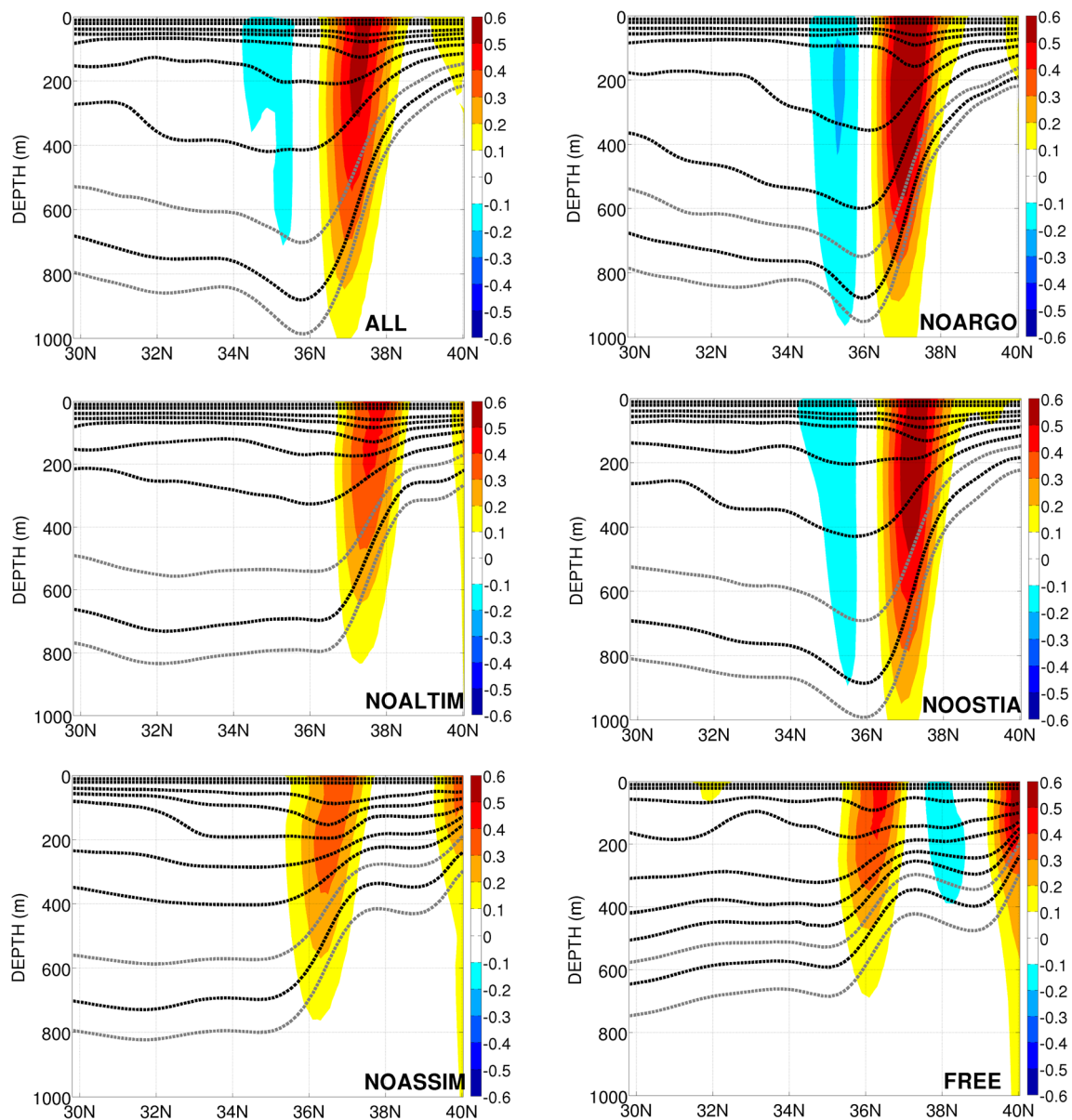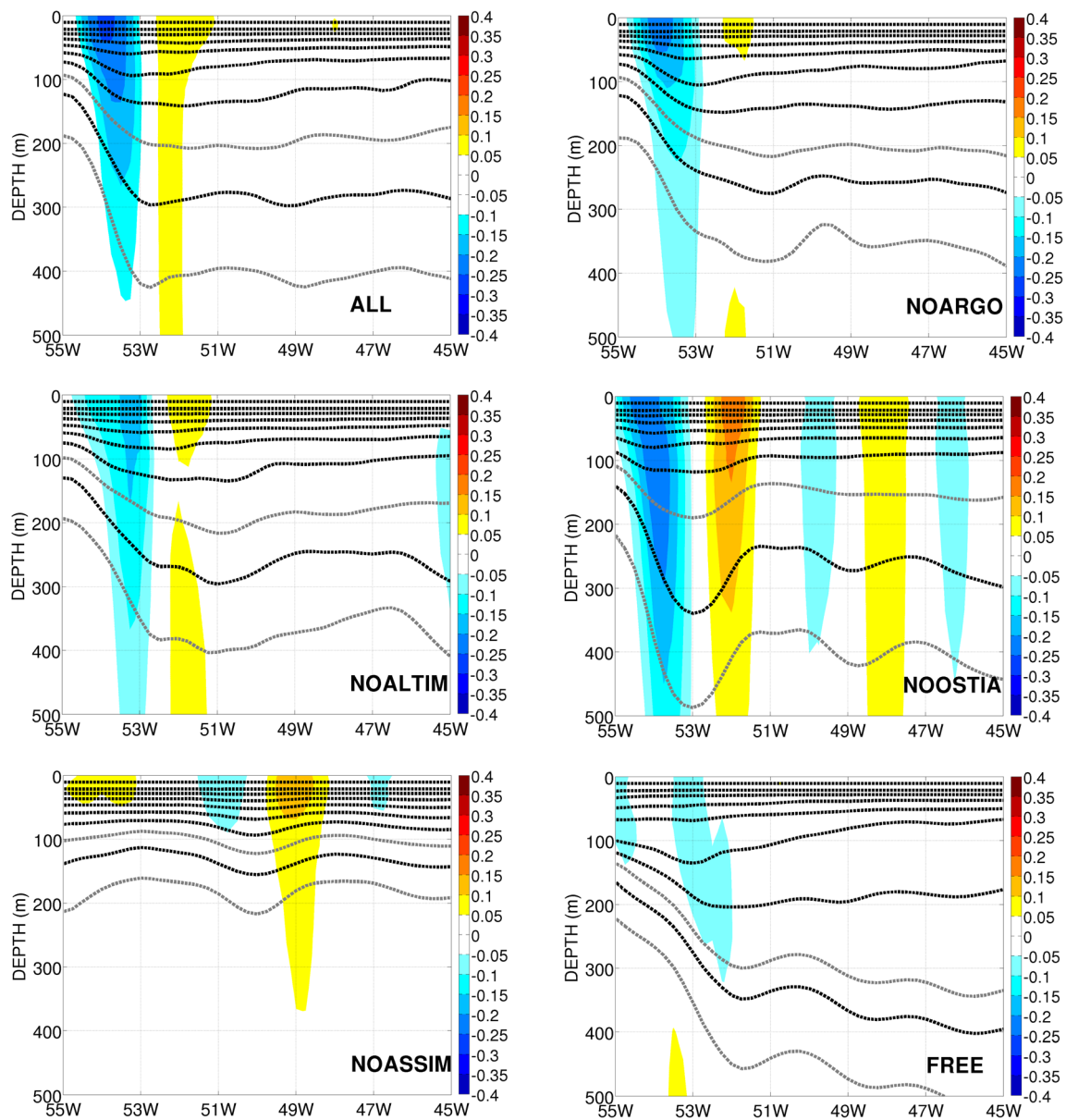
Fig. 12 Vertical cross section of the mean zonal velocity (m/s) from 1 January 2010 to 31 December 2012 along 68° W from the surface to 1000 m for each OSE run. The contour lines indicate the depth of the interfaces

of some model layers, so that the first gray line from top to bottom is the depth of the lower boundary of 12th layer, and the second is the lower boundary of the depth of the 14th layer

and meridional velocities, respectively, and the depth of the isopycnal layers in the top 1000 m. Considering the vertical cross section along 68° W shown in Fig. 12, the ALL run presents zonal velocities of the GS flowing eastward with maximum velocity core greater than 0.5 m/s around 37° N in the top 300 m. A weaker return flow to the south of the GS is also reproduced. The maximum mean velocity and structure in the upper 400 m is in good agreement with Eulerian observations, despite the magnitude of the rotated downstream velocity being much larger (e.g., Johns et al. 1995). The FREE run presents the eastward flow separated in two branches, one to the south and the other to the north of the expected GS mean position. Isopycnal layers are much

closer to the surface showing the FREE run representation of water masses in the region is also poor. When Argo is denied, there is intensification of the GS core and of the associated recirculation with respect to the ALL run. This intensification is consistent with the slope increase and deepening of the layers 10, 11, and 12 around 36° N. The position of the GS core is not altered when Argo data is withheld. When altimetry data is denied, the GS core and recirculation are damped along with a raise of the isopycnal layers around 36° N and a slope decrease. This is consistent with the weakening of the SSH gradient (Fig. 8) in the region. The influence of SST is similar to the influence of Argo data, but it is weaker and close to the ALL run. It shows that SLA and Argo data

Fig. 13 Vertical cross section of the mean meridional velocity (m/s) from 1 January 2010 to 31 December 2012 along 38° S from the surface to 1000 m for each OSE run. The contour lines indicate the depth of the interfaces of some model layers. The first gray line from top to bottom is the depth of the lower boundary of 11th layer, and the second is the lower boundary of the depth of the 13th layer

are able to constrain the thermohaline structure of the upper ocean and that SST is not affecting too much circulation in the region. The NOASSIM run is close to the FREE run state with the two-branch structure of the GS and the low depth of layers 14 to 7.

The meridional current along 38° S close to the BMC region is presented in Fig. 13. In comparison with the FREE run, the ALL run imposes a sharper slope of several model layers close to 53° W, particularly of the layers 11, 12, and 13. It produces a good representation of the southward BC with velocity core of about − 0.4 m/s. This value is very close to observations (Evans and Signorini 1985). When SST is denied, there is an increase of the BC depth as well as of the

depth and magnitude of the associated northward recirculation around 52° W. Also, there is an intensification of meanders characterized by a sequence of northward and southward flows along this latitude. This indicates that the assimilation of SST may be able to constrain circulation in the BC region more efficiently than in the GS region. When Argo is denied and when SLA is denied, the BC core is damped and the slope of some layers is smoothed. However, the NOALTIM run shows the BC is shifted by a couple degrees to the east. Overall, this is an indication that all data are important to produce a more accurate representation of the BC with the current HYCOM+RODAS system. The NOASSIM run produces a scenario that is not close to any run, including the

Table 2 For each OSE run, this table presents the estimated transport associated with the Gulf Stream (GS) and the Brazil Current (BC) in different longitudes or latitudes

	Gulf Stream (Sv)		Brazil Current (Sv)	
	58° W	68° W	22° S	38° S
ALL	75.8 ± 22.6	89.2 ± 36.6	− 3.2 ± 1.9	− 13.8 ± 2.3
NOARGO	74.9 ± 21.4	114.7 ± 30.7	− 2.6 ± 1.8	− 11.0 ± 2.2
NOALTIM	46.7 ± 18.1	69.1 ± 26.3	− 3.6 ± 2.2	− 8.0 ± 2.1
NOOSTIA	73.7 ± 19.4	95.7 ± 22.6	− 2.3 ± 1.5	− 14.6 ± 2.2
NOASSIM	18.3 ± 5.3	62.8 ± 19.1	− 3.5 ± 1.8	− 3.6 ± 1.0
FREE	19.6 ± 6.6	25.3 ± 8.3	− 4.3 ± 1.7	− 3.9 ± 1.3

FREE run. It means that the BC in this NOASSIM run is placed northward of the expected position and a diffuse Malvinas Current is dominating the circulation west of 53° W. It is expected that a longer integration of the NOASSIM run will converge to the FREE run.

The results presented above show the importance of assimilation, particularly of SLA, to improve the position and intensity of the GS and the BC. This has a direct impact in the simulated volume transport. The mean GS and BC transport and standard deviation at different locations are shown in Table 2 for each OSE member. The magnitudes of the transports simulated by the FREE run for the GS at 58° W and 68° W are about 25% and 28%, respectively, of the transports produced by the ALL run. Similar behavior is observed for the BC at 38° S. When no SLA data is assimilated, the GS transport drops substantially, indicating the importance of altimetry data to improve the representation of the western boundary currents. The BC transport at 22° S does not respond the same way. As mentioned above, at this location, the BC is more sensitive to SST data. The standard deviation of the GS and BC transports is also more strongly reduced in the NOALTIM run than in the NOARGO and the NOOSTIA runs, except for the BC at 22° S. The transports of the GS and the BC by the assimilation runs are relatively close to the ones found in the literature. Rossby et al. (2014) estimated GS transports of about 94.5 Sv combining Pegasus data at 73° W with 20 years of ADCP data from a cruise line between New Jersey and Bermuda. Hogg (1992) and Johns et al. (1995) estimated transports of about 93.7 Sv along 58° W and 95.5 Sv along 68° W for the GS, while HYCOM+RODAS simulated about 75.8 Sv and 89.2 Sv, respectively. The BC transport southward of 20° S has been measured and estimated by several authors. Its transport ranges from − 5.0 to − 10.0 Sv between 20° S and 25° S (Silveira et al. 2004) according to ADCP measurements and geostrophic estimates (Evans and Signorini 1985; Stramma 1991; Lima 1997). It may reach − 18 Sv at 31° S (Garfield 1990) and − 20 Sv at 38° S Garzoli (1993). The ALL run produced − 3.2 Sv along

22° S, and − 13.8 Sv along 38° S, smaller than observations. The cause for this weaker BC transport, particularly in higher latitudes, is explained by the anomalously intense northward Intermediate Western Boundary Current (not shown). This model deficiency could not be corrected by assimilation. The model configuration employed here has small horizontal and vertical resolution. It was construct to be the part of the first operational ocean weather forecasting system in Brazil, over which experience would be gained for future improvements. A decision was made to take the surface as the reference level for the potential density in model hybrid coordinate system since the major interest is in the ocean surface fields. These limitations may have negatively affected the formation and representation of Antarctic Intermediate Water. It has been shown that high model resolution may greatly influence the air-sea heat fluxes and the formation of crucial to correctly simulate the formation of deep and intermediate waters (e.g., Su et al. 2018). Improved model configuration with higher horizontal and vertical layers will be pursued in a near future.

5 Conclusions

The Oceanographic Modeling and Observation Network (REMO) developed an ocean data assimilation system (RODAS) to produce reanalyses and initial condition for a short-range ocean forecast system employing the ocean model HYCOM. The system based on the ensemble optimal interpolation scheme was presented here. It is able to assimilate SST fields, vertical T/S profiles, and along-track or gridded SLA data. The model error covariance matrix estimated by RODAS is not static, since it considers the intraseasonal and seasonal variability by selecting ensemble members from a long-term run according to the assimilation day. This may be considered as an advantage with respect to other EnOI applications (Counillon and Bertino 2009; Oke et al. 2013). However, more sophisticated strategies should be pursued, despite the increased computational cost. Data assimilation algorithms based on the EnKF, which estimates the model error covariance matrix from expensive ensemble forecasting system (Kalnay 2003; Lima et al. 2019), have the potential to reduce the analysis errors and, consequently, the forecasting errors.

It was shown that RODAS was able to substantially improve the model errors with respect to free run. The results of the assimilation run, in which all data were assimilated (ALL), could reduce SST errors by 50%, reduce vertical T and S errors by 47% and 46%, respectively, and improve SLA correlation by 50% with respect to FREE run. The errors attained by HYCOM+RODAS are comparable with other GODAE OceanView systems. The Australian Bluelink forecast system (Oke et al. 2013; Oke et al., 2015a, b) attains RMSD of T and S of 0.8 °C and 0.14 psu in the top 500 m, and SLA correlation

of 67%. The UK Met's Office FOAM v12 system also presented in Oke et al. (2015b) attains RMSD of T and S of 0.58 °C and 0.16 psu in the top 2000 m in the Southern Ocean and 0.61 °C and 0.14 psu in the Global Ocean. The HYCOM+ RODAS system attains RMSD of T and S of 0.93 °C and 0.18 psu in the top 500 m, and 0.81 °C and 0.15 psu in the top 1400 m, respectively. SLA correlation is 61%.

It should be highlighted that the RMSDs of the runs were calculated with respect to observational data before they were assimilated, i.e., they are not the analysis errors, but the mean of the almost 3-day window after each assimilation cycle. Therefore, these errors could be considered as a measure of the HYCOM+RODAS short-range predictability.

RODAS computational cost is relatively low, when applied over the model configuration presented here with 480 × 760 × 21 grid points and 126 ensemble members. Assimilation of SST, SLA, and T/S Argo took about 10 min, 9 min, and 5 min, respectively, while 1 day of the free model run took about 3 min, when 4 computational nodes were used, each node with 16 Intel Xeon 2.4 GHz processors with 64 GB of memory. These times include the reading and interpolation of the ensemble members. Therefore, considering HYCOM+RODAS skills and computational cost, we believe it is suitable to be employed in longer reanalysis runs and operational short-range forecasts, among other applications, despite of the need for further improvements.

Together with the presentation of the forecast/reanalysis system, OSEs were conducted. It was shown that OSIIA SST was important to constrain SST and T in the ocean mixed layer, by avoiding an erroneous deepening of the thermocline in a vast region of the subtropical North Atlantic. Argo data is crucial to produce a more accurate thermohaline structure. Particularly for S, only Argo data could reduce large model biases in the upper 900 m. The lack of altimetry data produces a light positive impact in T and S in the subsurface, showing that vertical localization associated with SLA increments may improve the system skills. Also, increasing vertical resolution may help producing better covariances among SLA, model layer thicknesses, T, and S. On the other hand, altimetry data is very important to ocean circulation and together with Argo data could produce the best representation of the GS. For the BC, all data were important.

Many improvements in the current HYCOM+RODAS are necessary. A crucial improvement should come from the model free run in order to produce smaller biases and ease the correction by data assimilation. A new HYCOM configuration is under preparation with 1/12° of horizontal resolution for the current HYCOM 1/4° domain and 32 hybrid layers. Also, a single analysis increment will be sought to improve corrections with strong baroclinic characteristics. Assimilation of sea surface salinity and data from gliders, XBTs, and instrumented marine mammals should be also incorporated into the system, as part of the work towards the best possible representation of the ocean state.

References

Backeberg, BC, ·Counillon F, Johannessen JA, Pujol MI (2014) Assimilating along-track SLA data using the EnOI in an eddy resolving model of the Agulhas system. Ocean Dyn 64:1121-1136. https://doi.org/10.1007/s10236-014-0717-6

Balmaseda M, Vidard A, Anderson DLT (2008) The ECMWF ocean analysis system: ORA-S3. Mon Weather Rev 136:3018–3034. https://doi.org/10.1175/2008MWR2433.1

Balmaseda M, Mogensen K, Weaver A (2013) Evaluation of the ECMWF ocean reanalysis system ORAS4. Q J R Meteorol Soc 139:1132–1161. https://doi.org/10.1002/qj.2063

Bell MJ, Lefebvre M, Le Traon P-Y, Smith N, Wilmer-Becker K (2009) GODAE: the Global Ocean Data Assimilation Experiment. Oceanogr 22:14–21. https://doi.org/10.5670/oceanog.2009.62

Bell MJ, Schiller A, Le Traon P-Y, Smith NR, Dombrowsky E, Wilmer-Becker K (2015) An introduction to GODAE OceanView. J Oper Oceanogr 8(sup1):s2–s11. https://doi.org/10.1080/1755876X.2015.1022041

Bleck R (2002) An oceanic general circulation model framed in hybrid isopycnic-Cartesian coordinates. Ocean Model 37:55–88

Boyer, TP, Antonov JI, Baranova OK, Coleman C, Garcia HE, Grodsky A, Johnson DR, O'Brien TD, Paver CR, Locarnini RA, Mishonov AV, Reagan JR, Seidov D, Smolyar IV, Zweng MM, 2013. World Ocean Database 2013. S. Levitus, Ed., A. Mishonov Tech. Ed., NOAA Atlas NESDIS 72, 209pp

Carvalho JPS, Costa FB, Mignac D, Tanajura CAS (2019) Assessing the extended-range predictability of the ocean model HYCOM with the REMO ocean data assimilation system (RODAS) in the South Atlantic, J Oper Oceanogr. https://doi.org/10.1080/1755876X.2019.1606880

Castruccio F, Verron J, Gourdeau L, Brankart JM, Brasseur P (2008) Joint altimetric and in-situ data assimilation using the GRACE mean dynamic topography: a 1993–1998 hindcast experiment in the Tropical Pacific Ocean. Ocean Dyn 58:43–63. https://doi.org/10.1007/s10236-007-0131-4

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

Conkright ME, Locarnini RA, Garcia HE, O'Brien TD, Boyer TP, Stephens C, Antonov JI (2002) World Ocean Atlas 2001: objective analyses, data statistics, and figures. CD-ROM documentation. National Oceanographic Data Center, Silver Spring, MD

Cooper M, Haines K (1996) Altimetric assimilation with water property conservation. J Geophys Res 101(C1):1059–1077

Costa F, Tanajura C (2015) Assimilation of sea-level anomalies and Argo data into HYCOM and its impact on the 24 hour forecasts in the western tropical and South Atlantic. J Oper Oceanogr 8:52–62. https://doi.org/10.1080/1755876X.2015.1014646

Counillon F, Bertino L (2009) High-resolution ensemble forecasting for the Gulf of Mexico eddies and fronts. Ocean Dyn 59:83–95

Daley R (1991) Atmospheric data analysis. Cambridge Univ. Press, Cambridge, 457 pp

Evans DL, Signorini SR (1985) Vertical structure of the Brazil Current. Nature 315:48–50

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

Evensen G (2006) Data assimilation: The ensemble Kalman filter. Springer, New York

Ezer T, Mellor GL (1997) Data assimilation experiments in the Gulf Stream region: how useful are satellite-derived surface data for nowcasting the subsurface fields? J Atmosph Oceanogr Tech 14: 1379–1391

Gabioux M, Costa VS, Souza JMAC, Oliveira BF, Paiva AM (2013) Modeling the South Atlantic Ocean from medium to high-resolution. Rev Bras Geofis 31:229–242

Garfield N (1990) The Brazil Current at subtropical latitudes. Ph.D. Thesis, University of Rhode Island, 121 pp

Garzoli SL (1993) Geostrophic velocity and transport variability in the Brazil/Malvinas confluence. Deep-Sea Res 40:1379–1403

Gaspari G, Cohn SE (1999) Construction of correlation functions in two and three dimensions. Q J R Meteorol Soc 125:723–757

GTSPP 2010 Real-time quality control manual, First Revised Edition. UNESCO-IOC 2010, IOC Manuals and Guides No. 22, Revised Edition. https://www.nodc.noaa.gov/GTSPP/document/qcmans/MG22rev1.pdf. Accessed 6 June 2019

Hogg NG (1992) On the transport of the Gulf Stream between Cape Hatteras and the Grand Banks. Deep Sea Research Part A: Oceanographic Research Papers 39:1231–1246. https://doi.org/10.1016/0198-0149(92)90066-3

Johns WE, Shay TJ, Bane JM, Watts DR (1995) Gulf Stream structure, transport, and recirculation near 68°W. J Geophys Res:Oceans 100: 817–838. https://doi.org/10.1029/94JC02497

Kalnay E (2003) Atmospheric modeling, data assimilation and predictability. Cambridge University Press, New York

Kalnay E, Kanamitsu M, Kistler R, Collins W et al (1996) The NCEP/NCAR 40-year reanalysis project. Bull Amer Meteor Soc 77:437–472

Large WG, Mcwilliams JC, Doney SC (1994) Oceanic vertical mixing: a review and a model with a nonlocal boundary layer parameterization. Rev Geophys 32:363–403

Lea DJ, Martin MJ, Oke PR (2014) Demonstrating the complementarity of observations in an operational ocean forecasting system. Q J R Meteorol Soc 140:2037–2049. https://doi.org/10.1002/qj.2281

Lea DJ, Mirouze I, Martin MJ, King RR, Hines A, Walters D, Thurlow M (2015) Assessing a new coupled data assimilation system based on the Met Office coupled atmosphere–land–ocean–sea ice model. Mon Weather Rev 143:4678–4694. https://doi.org/10.1175/MWR-D-15-0174.1

Lima JAM (1997) Oceanic circulation on the Brazilian shelf break and slope at 22°S, Ph.D. In: Thesis. University of New South Wales, Australia

Lima LN, Tanajura CAS (2013) A study of the impact of altimetry data assimilation on short-term predictability of the HYCOM ocean model in regions of the Tropical and South Atlantic Ocean. Rev Bras Geofis 31:271–288

Lima JAM, Martins RP, Tanajura CAS et al (2013) Design and implementation of the Oceanographic Modeling and Observation Network (REMO) for operational oceanography and ocean forecasting. Rev Bras Geofis 31:209–228

Lima LN, Pezzi LP, Penny SG, Tanajura CAS (2019) An investigation of ocean model uncertainties through ensemble forecast experiments in the southwest Atlantic Ocean. J Geophys Res: Oceans 124:432–452. https://doi.org/10.1029/2018JC013919

Locarnini RA, Mishonov AV, Antonov JI, Boyer TP, Garcia HE, Baranova OK, Zweng MM, Johnson DR (2010) World Ocean Atlas 2009, volume 1: temperature. In: Levitus S (ed) NOAA Atlas NESDIS, vol 68. U.S. Government Printing Office, Washington, p 184

Lorenc AC (2003) The potential of the ensemble Kalman filter for NWP – a comparison with 4D-Var. Q J R Meteorol Soc 129:3183–3203

Martin MJ, Balmaseda M, Bertino L (2015) Status and future of data assimilation in operational oceanography. J Oper Oceanogr 8:S28–S48. https://doi.org/10.1080/1755876X.2015.1022055

Melo RL, Freitas ACN, Russo L, Oliveira JF, Tanajura CAS, Alvarenga JBR (2013) Ocean forecasts in the southwestern Atlantic: impact of different sources of sea surface height in data assimilation. Rev Bras Geofis 31:243–255

Mignac D, Tanajura CAS, Santana A, Lima LN, Xie J (2015) Argo data assimilation into HYCOM with an EnOI method in the Atlantic Ocean. Ocean Sci 11:195–213. https://doi.org/10.5194/os-11-195-2015

Moore A, Arango H, Broquet G, Powell B, Weaver AT, Zavala-Garay J (2011) The Regional Ocean Modeling System (ROMS) 4-dimensional variational data assimilation systems: I - system overview and formulation. Progress Oceanogr 91:34–49

Oke PR, Sakov P (2008) Representation error of oceanic observations for data assimilation. J Atmos Ocean Tech 25:1004–1017

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

Oke PR, Brassington G, Griffin DA, Schiller A (2008) The Bluelink ocean data assimilation system (BODAS). Ocean Model 21:46–70

Oke PR, Sakov P, Cahill ML, Dunn JR, Fiedler R, Griffin DA, Mansbridge JV, Ridgway KR, Schiller A (2013) Towards a dynamically balanced eddy-resolving ocean reanalysis: BRAN3. Ocean Model 67:52–70. https://doi.org/10.1016/j.ocemod.2013.03.008

Oke P, Larnicol G, Fujii Y, Smith G, Lea D, Guinehut S, Remy E, Balmaseda M, Rykova T, Surcel-Colan D et al (2015a) Assessing the impact of observations on ocean forecasts and reanalyses: Part 1, Global studies. J Oper Oceanogr 8(spu1):s49–s62. https://doi.org/10.1080/1755876X.2015.1022067

Oke PR, Larnicol G, Jones EM, Kourafalou V, Sperrevik AK, Carse F, Tanajura CAS, Mourre B, Tomani M, Brassington G, Le Henaff M, Halliwell GR, Atlas R, Moore AM, Edwards CA, Martin MJ, Sellar AA, Alvarez A, De Mey P, Iskandarani M (2015b) Assessing the impact of observations on ocean forecasts and reanalyses: part 2, regional applications. J Oper Oceanogr 8(sup1):s63–s79. https://doi.org/10.1080/1755876X.2015.1022080

Penny SG, Hamill TM (2017) Coupled data assimilation for integrated earth system analysis and prediction. Bull Amer Meteor Soc 98: ES169–ES172. https://doi.org/10.1175/BAMS-D-17-0036.1

Penny SG, Behringer DW, Carto JA, Kalnay E (2015) A hybrid global ocean data assimilation system at NCEP. Mon Weather Rev 143: 4660–4677. https://doi.org/10.1175/MWR-D-14-00376.1

Rio MH, Mulet S, Picot N (2014) Beyond GOCE for the ocean circulation estimate: synergetic use of altimetry, gravimetry, and in situ data provides new insight into geostrophic and Ekman currents. Geophys Res Lett 41:8918–8925. https://doi.org/10.1002/2014GL061773

Rossby T, Flagg CN, Donohue K, Sanchez-Franks A, Lillibridge J (2014) On the long-term stability of Gulf Stream transport based on 20 years of direct measurements. Geophys Res Lett 41:114–120. https://doi.org/10.1002/2013GL058636

Schiller A, Mourre B, Drillet Y, Brassington G (2018) An overview of operational oceanography. In Chassignet E, Pascual A, Tintoré J, Verron J (ed) New frontiers in operational oceanography:1–26. https://doi.org/10.17125/gov2018.ch01

Silveira IC, Calado L, Castro BM, Cirano M, Lima JAM, Mascarenhas AS (2004) On the baroclinic structure of the Brazil Current-Intermediate Western Boundary Current system at 22°S-23°S. Geophys Res Lett 31:L14308. https://doi.org/10.1029/2004GL020036

Stramma L (1991) Geostrophic transport of the South Equatorial Current in the Atlantic. J Mar Res 49:281–294

Su Z, Ingersoll AP (2016) On the minimum potential energy state and the eddy-size-constrained APE density. J Phys Oceanogr 46:2663–2674. https://doi.org/10.1175/JPO-D-16-0074.1

Su Z, Stewart AL, Thompson AF (2014) An idealized model of Weddell Gyre export variability. J Phys Oceanogr 44:1671–1688. https://doi.org/10.1175/JPO-D-13-0263.1

Su Z, Wang J, Klein P, Thompson AF, Menemenlis D (2018) Ocean submesoscales as a key component of the global heat budget. Nat Commun 9:775. https://doi.org/10.1038/s41467-018-02983-w

Tanajura C, Costa F, Ramos da Silva R, Ruggiero G, Daher V (2013) 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

Tanajura CAS, Santana A, Mignac D, Lima L, Belyaev K, Ji-Ping X (2014) The REMO Ocean Data Assimilation System into HYCOM (RODAS_H): general description and preliminary results. J Atmos Oceanic Sci Lett 7:464–470. http://159.226.119.58/aosl/EN/10.3878/j.issn.1674-2834.14.0011

Thacker W, Esenkov OE (2002) Assimilating XBT data into HYCOM. J Atmos Ocean Tech 19:709–724

Tranchant B, Remy E, Greiner E, Legalloudec O (2019) Data assimilation of Soil Moisture and Ocean Salinity (SMOS) observations into the Mercator Ocean operational system: focus on the El Niño 2015 event. Ocean Sci 15:543–563. https://doi.org/10.5194/os-15-543-2019

Weaver AT, Vialard J, Anderson DLT (2003) Three- and four-dimensional variational assimilation with a general circulation model of the tropical Pacific Ocean. Part I: formulation, internal diagnostics, and consistency check. Mon Weather Rev 131:1360–1378

Xie J, Zhu J (2010) Ensemble optimal interpolation schemes for assimilating Argo profiles into a hybrid coordinate ocean model. Ocean Model 33:283–298

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

Yan CX, Zhu J, Xie JP (2010) An ocean reanalysis system for the joining area of Asia and Indian-Pacific Ocean. J Atmos Ocean Sci Lett 3:81–86