UNIVERSIDADE FEDERAL DA BAHIA
INSTITUTO DE GEOCIÊNCIAS
CURSO DE GRADUAÇÃO EM OCEANOGRAFIA

JÚLIA PORTO SILVA CARVALHO

AVALIANDO A PREVISIBILIDADE ESTENDIDA DO HYCOM COM O SISTEMA DE ASSIMILAÇÃO DE DADOS DA REMO (RODAS) NO OCEANO ATLÂNTICO SUL

Salvador
2017

JÚLIA PORTO SILVA CARVALHO

AVALIANDO A PREVISIBILIDADE ESTENDIDA DO HYCOM COM O SISTEMA DE ASSIMILAÇÃO DE DADOS DA REMO (RODAS) NO OCEANO ATLÂNTICO SUL

Monografia apresentada ao Curso de Oceanografia, Instituto de Geociências, Universidade Federal da Bahia, como requisito parcial para obtenção do grau de Bacharel em Oceanografia. Este trabalho é apresentado na forma de um manuscrito que será submetido para a revista Journal of Operational Oceanography.

Orientador: Prof. Dr. Clemente Augusto Souza Tanajura

Salvador
2017

Modelo de ficha catalográfica fornecido pelo Sistema Universitário de Bibliotecas da UFBA para ser confeccionada pelo autor

TERMO DE APROVAÇÃO

JÚLIA PORTO SILVA CARVALHO

AVALIANDO A PREVISIBILIDADE ESTENDIDA DO HYCOM COM O SISTEMA DE ASSIMILAÇÃO DE DADOS DA REMO (RODAS) NO OCEANO ATLÂNTICO SUL

Monografia aprovada como requisito parcial para obtenção do grau de Bacharel em Oceanografia, Universidade Federal da Bahia, pela seguinte banca examinadora:

Clemente Augusto Souza Tanajura – Orientador
Doutor em Meteorologia pela Universidade de Maryland, Estados Unidos
Universidade Federal da Bahia

Janini Pereira
Doutora em Oceanografia física pela Universidade de São Paulo, USP, São Paulo, Brasil
Universidade Federal da Bahia

Raquel Leite Mello
Doutora em Oceanografia física pela Universidade de São Paulo, USP, São Paulo, Brasil
Centro de Pesquisa em Geofísica e Geologia (CPGG/UFBA) - REMO

Salvador, 25 de julho de 2017

AGRADECIMENTOS

Agradeço aos meus pais e às minhas irmãs pelo companheirismo e por sempre me apoiarem nas minhas escolhas, me ajudando a tornar tudo possível.

Ao Professor Clemente Tanajura pela orientação e por todo o aprendizado durante a realização deste trabalho. Aos demais professores do curso de oceanografia da UFBA pelos ensinamentos dentro e fora da sala de aula.

À toda a equipe da REMO pela troca de conhecimentos e pela paciência e disponibilidade sempre que precisei de ajuda. Agradeço também aos colegas de laboratório pela companhia diária, pelos momentos de descontração e por todo o apoio que deram.

Aos meus amigos do curso de oceanografia por sempre estarem ao meu lado em momentos bons e ruins e por tornarem a minha rotina muito mais leve e divertida. E aos meus familiares e amigos que mesmo longe sempre me incentivaram a seguir em frente.

RESUMO

Sistemas de previsão oceânica tem o papel fundamental na sociedade, pois fornecem informações sobre o estado do oceano que são relevantes para diversas finalidades. A evolução operacional desses sistemas tem sido extremamente significativa nos últimos anos. A Rede de Modelagem e Observação Oceanográfica (REMO) desenvolveu o Sistema de Assimilação de Dados da REMO (RODAS), que emprega o método da Interpolação ótima por conjuntos (EnOI) aplicado ao modelo HYCOM na porção oeste do Oceano Atlântico Sul. Este estudo visa investigar a previsibilidade estendida do sistema HYCOM+RODAS. A análise produzida pelo RODAS foi utilizada como condição inicial para 48 integrações de 30 dias do modelo livre (hindcasts), que foram comparadas com a persistência (condição inicial fixa em toda a janela de previsão) e com uma rodada livre do modelo inicializada sem assimilação. Os hindcasts tiveram o menor desvio quadrático médio (RMSD) e maior correlação de temperatura da superfície do mar (TSM) e anomalia da altura do nível do mar (AANM) com relação às análises do RODAS em toda a janela de 30 dias. O RMSD da persistência aumentou de 0°C para 1.09°C no último dia e de 0 m para 0.08 m no mesmo período, para TSM e AANM respectivamente. O RMSD do hindcast atingiu valores máximos de 0.46°C e 0.05 m. O RMSD (correlação) da rodada livre foi sempre maior (menor) que o do hindcast, com uma média de 0.88°C e 0.13 m. Na coluna d'água, o aumento de RMSD do hindcast foi ainda menor. Os resultados indicam que a capacidade do sistema HYCOM+RODAS em prever o estado do oceano se estende por mais de um mês e o estado termohalino do oceano melhorou de forma substancial quando comparado à rodada livre. Um estudo de caso de um evento de ressurgência costeira na costa sudeste do Brasil demonstrou que o HYCOM+RODAS foi eficientemente capaz de reproduzir este fenômeno de circulação.

Palavras-chave: Assimilação de dados oceânicos; HYCOM; Oceano Atlântico Sul

Assessing the extended-range predictability of HYCOM with the REMO Ocean Data Assimilation System (RODAS) in the South Atlantic Ocean

Abstract

Ocean forecasting systems (OFS) have a fundamental role to deliver ocean services to society and their operational evolution has been extremely significant over the last years. The Brazilian Oceanographic Modelling and Observation Network (REMO) has developed the REMO Ocean Data Assimilation System (RODAS), which is based on an Ensemble Optimal Interpotalion (EnOI) scheme applied into HYCOM over the western South Atlantic Ocean. This study aims to investigate the extended-range forecast skills of the HYCOM+RODAS system. The analysis produced by RODAS are used as initial condition for 48 HYCOM 30-day simulations, which are then compared to persistence (no change from the initial condition) and to a model free run initialized with no assimilation. The model hindcasts had the lowest root mean square difference (RMSD) and highest correlation of sea surface temperature (SST) and sea level anomaly (SLA) with respect to RODAS analysis at all lead times. Persistence RMSD increased from 0°C to 1.09°C by the 30th day and from 0 m to 0.08 m in the same period for SST and SLA, respectively, while the hindcast RMSD increased to 0.46°C and 0.05 m. The free run RMSD (correlation) was always higher (lower) than that of the hindcast with an average of 0.88°C and 0.13 m. In the subsurface, hindcast RMSD increase was even lower. The results suggest that HYCOM+RODAS predictive skill extends for more than a month and the thermohaline state of the ocean was consistently improved. A case study on a coastal upwelling event in the southeast coast of Brazil demonstrated that the HYCOM+RODAS system was efficiently able to reproduce this ocean feature.

Keywords: Ocean data assimilation; HYCOM; South Atlantic Ocean.

LISTA DE FIGURAS

LISTA DA ABREVIAÇÕES

AVISO – Archiving, Validation et Interpretation des données des Satellites Océanographique

BC – Brazil Current

BMC – Brazil-Malvinas Confluence

CENPES – Centro de pesquisas Leopoldo Américo Miguez de Mello

CF – Cabo Frio

CFSR – Climate Forecast System Reanalysis

CHM – Navy's Hydrographic Center

COADS – Comprehensive Ocean Atmosphere Data Set

EnOI – Ensemble Optimal Interpolation

ETOPO2 – Earth Topography 2

GEBCO – General Bathymetric Chart of the Oceans

GHRSST – Group of High Resolution Sea Surface Temperature

GODAE – Global Ocean Data Assimilation Experiment

GOV – GODAE Ocean View

HYCOM – Hybrid Coordinate Ocean Model

IAP/CAS – Institute of Atmospheric Physics by the Chinese Academy of Science

IOC – Intergovernmental Oceanographic Comission

MOVAR – Monitoring the Upper Ocean Transport Variability

NCEP – National Centers for Environmental Prediction

NOAA – National Oceanic and Atmospheric Admnistration

OFS – Operacional Forecasting Systems

OSTIA – Ocean Sea Surface Temperature and Sea Ice Analysis

PNBOIA – Programa Nacional de Bóias

QC – Quality Control

REMO – Rede de Modelagem e Observação Oceanográfica

RMSD – Root Mean Square Deviation

RODAS – REMO Ocean Data Assimilation System

S - Salinity

SACW – South Atlantic Central Water

SLA – Sea Level Anomaly

SSH – Sea Surface Height

SSS – Sea Surface Salinity

SST – Sea Surface Temperature

T – Temperature

TW – Tropical Water

UFBA – Federal University of Bahia

UFRJ – Federal University of Rio de Janeiro

WOA – World Ocean Atlas

Δp_{obs} – Pseudo-observed Layer Thickness

SUMÁRIO

1. Introduction

Ocean forecasting systems (OFS) have a fundamental role to deliver ocean services to society. Monitoring and forecasting the behaviour of the ocean is essential for the sustained development of the marine environment and to provide relevant information for many users such as stakeholders dealing with fisheries, oil and gas industries, maritime safety, climate change and coastal management (Schiller et al. 2015). Operational evolution of OFS has been extremely significant over the last years and real-time ocean analysis, forecasts and reanalysis products are routinely delivered to users in high quality (Dombrowsky et al. 2009).

An eddy-resolving ocean model is a key component for ocean prediction at the mesoscale (Hurlburt et al. 2009). The Hybrid Coordinate Ocean Model (HYCOM) is a sophisticated model, which incorporates numerical techniques that are optimal for dynamically different regions of the ocean. Isopycnic layers are used in the open stratified ocean, pressure-level coordinates in the surface mixed layer and terrain-following coordinates in shallow coastal regions. However, model forecasts are likely to have errors due to limitations in the numerical methods, physical parameterizations, grid resolution, atmospheric forcing and initial and boundary conditions (Bleck 2002; Chassignet at al. 2007; Chassignet et al. 2009).

Observational data can be effectively used to minimize these errors and they are fundamental to operational OFS. They constrain ocean models with data assimilation systems and are employed to validate them and assess the OFS skills. A global ocean observing system has been implemented, providing valuable information from a large set of data (Oke at al. 2015a; Oke at al. 2015b). Satellites collect real time, global, high space-time resolution observations of key ocean variables such as sea surface temperature (SST), sea surface height (SSH), sea level anomaly (SLA), sea surface salinity (SSS) and ocean colour (Le Traon et al. 2015). These measurements, however, are restricted to the surface and are insufficient to represent the subsurface variability (Chassignet et al. 2009). Therefore, a secure supply of *in situ* ocean data is important to represent the ocean interior. Vertical profiles of temperature and salinity are collected by XBTs, CTDs and the Argo profilers system. The latter has proved to be an effective tool to monitoring the manifold of the ocean, with a large number of floats distributed around the world (Legler et al. 2015). Altogether, *in situ* observations

and satellite remote sensing provide a good degree of complementary information (Lea et al. 2014). However, given the large size of the ocean and the limitations of the measurements, no feasible observation system determines the state of the ocean completely. Therefore, models are necessary to complement the basic observations by extrapolating the data information in space, time and among different properties (Fukumori 2006; Zaron 2011).

Data assimilation comprises a set of techniques for estimating the oceanic state by combining model predictions with observed data in an optimal manner. As a result, data assimilation produces an objective analysis in order to represent the most accurate description of the past and best initial condition for the forecast. Given that ocean model forecasts are highly sensitive to the initial condition, initializing the model with conditions that represent as accurately as possible the actual state of the ocean at eddy-resolving resolution can lead to a gain in predictability and produce more reliable predictions (Brasseur 2006; Chassignet et al. 2007).

The Global Ocean Data Assimilation Experiment (GODAE) and its continuation, GODAE Ocean View (GOV), have provided a firm base for the development of global and regional analysis and forecasting systems (Bell et al. 2015). Since the beginning of the 1990s, many systems have been developed in several countries based on different ocean models and data assimilation methods (Tonani et al. 2015). These methods range from relatively simple schemes, such as Analysis Correction and Optimal Interpolation, to more sophisticated schemes, such as variational and ensemble technique (Cummings et al. 2009).

In Brazil, oceanographic information is highly demanded for application in the petroleum industry and for military activities. In 2007, the Brazilian Oceanographic Modelling and Observation Network (REMO) was created with the main goal of developing numerical ocean models with assimilative skills for the western South Atlantic Ocean. REMO brings together different institutions, including the Federal University of Bahia (UFBA), Federal University of Rio de Janeiro (UFRJ), the Brazilian Navy's Hydrographic Center (CHM) and Centro de Pesquisas Leopoldo Américo Miguez de Mello from Petrobras (CENPES). The scientific results and products developed by REMO may reach a broad variety of scientists, students and general

users of oceanographic information so that other applications can become feasible to promote benefits to the Brazilian society (Lima et al. 2013).

In collaboration with GOV and the Institute of Atmospheric Physics by the Chinese Academy of Sciences (IAP/CAS), the REMO Ocean Data Assimilation System (RODAS) was developed at UFBA. It employs HYCOM in a nested grid system with 1/24º, 1/12° and 1/4º horizontal resolution and 21 hybrid layers in the vertical over the Atlantic Ocean with focus on the southwest Atlantic. RODAS is based on the Ensemble Optimal Interpolation (EnOI) scheme and it assimilates SST, sea level anomaly (SLA) and T/S profiles into the model (Tanajura et al. 2014; Costa and Tanajura 2015; Mignac et al. 2015). RODAS is able to constrain the ocean model towards observation using a set of different model states to estimate the model errors. It is computationally efficient and therefore, suitable for operational purposes (Xie and Zhu 2010; Tanajura et al 2014; Mignac et al. 2015). The model domain employed in the present work encompasses the Atlantic Metarea V (from 36°S to 7°N and from 20°W up to the coast of Brazil), which is an area of high interest to the Brazilian Navy and has high economic and environmental relevance due to petroleum extraction.

Surface circulation in Metarea V is highly influenced by the Subtropical Gyre. The South Equatorial Current flows westward and branches in North Brazil Current and Brazil Current (BC) (Petreson and Stramma 1991). The BC is a western boundary current that flows southward along the Brazilian coast until the confluence with the Malvinas Current. Then both currents separate from the coast and their seaward extension is the South Atlantic Current (Peterson and Stramma, 1991). The water masses in the upper layers of Metarea V include the Tropical Water (TW) (T>20°C and S>36,0) as the surface mixed layer and the South Atlantic Central Water (SACW) (5°C<T<20°C and 34.3<S<36), located underneath (Stramma and England, 1999).

High resolution operational oceanography requires accurate depiction of mesoscale ocean features such as eddies and upwelling events, which are frequently observed within Metarea V. Upwelling has a remarkable importance due to its influence on local marine ecosystems and ocean dynamics. The region around Cabo Frio (CF, 23°S) is characterized by a robust coastal upwelling system, which has received large scientific attention (Campos et al. 2000; Rodrigues and Lorenzzeti 2001; Castelao and Barth 2006; Calado et al. 2010; Castelao 2012; Soutelino and Miranda 2013; Aguiar et al.

2014; Palóczy at al. 2014). Upwelling events can be identified by a sharp SST gradient due to an intrusion of SACW into the continental shelf, possibly reaching the surface near the coast. They have a strong seasonal signal and are more likely to develop in spring and summer, when prevailing winds blow from northeast. Some current-driven mechanisms such as cyclonic eddies and interaction of the Brazil Current with bottom topography are also important contributors to the development of this coastal upwelling system (Calado et al. 2010; Rodrigues and Lorenzzeti 2001; Aguiar et al. 2014; Palóczy at al. 2014).

The present work aims at contributing to the permanent development of RODAS in order to better represent the state and circulation of the South Atlantic Ocean and serve as a component of a forecasting system. More specifically, it aims at investigating the extended-range forecast skills of the HYCOM+RODAS system. The analysis produced by RODAS are used as initial condition for 48 HYCOM 30-day simulations, which are then compared to persistence (no change from the initial condition) and to a model free run initialized with no assimilation. The simulations were forced by reanalysis fields produced by Climate Forecast System Reanalysis (CFSR) of the National Centers for Environmental Prediction (NCEP/NOAA). Forcing the model with reanalysis allows us to minimize the impacts of errors in the atmospheric forcing and to examine the performance of the ocean model and its data assimilation scheme (Zhu 2011). The atmospheric reanalysis can also be seen as the best possible atmospheric forcing the ocean forecasting system could use. Therefore, the 30-day HYCOM simulations could be considered forecasts in which the atmospheric reanalysis would lead to the lowest bound of the ocean model errors considering all possible atmospheric forcings. Lastly, HYCOM+RODAS predictive skill was assessed over the south and southeast coast of Brazil, with a focus on the upwelling system, and a case study was investigated. The HYCOM+RODAS predictability is put to proof with respect to this feature that evolves in the scale of days and is strongly dominated by the atmospheric forcing and not so much by the initial condition.

Section 2 presents the methodology, and it includes a brief description of the ocean model, some details of the data assimilation scheme, the numerical experiments and the metrics employed to assess the hindcast skills. Section 3 presents the results, and Section 4 the conclusions.

2. Methodology

2.1. The ocean model

HYCOM is a primitive equation general circulation model, which has freedom to adjust the vertical spacing of the generalized coordinate layers, simplifying the numerical implementation of several processes (Bleck 2002). In the open stratified ocean, isopycnic layers are used, but they make a smooth transition to terrain-following coordinates in shallow coastal regions and to fixed depth (z) coordinates in the surface mixed layer. The model is formulated in terms of target densities and the vertical coordinate distribution is chosen at every time step and in every grid column individually, allowing the model to optimally simulate coastal and open-ocean circulation features (Chassignet et al. 2009).

In the present work, HYCOM was configured with a horizontal resolution of 1/12° for the domain 45°S-10°N, 68°W-18°W, which contains Metarea V. It corresponds to 601 and 733 grid points in the zonal and meridional directions, respectively. The model was nested in another HYCOM configuration with horizontal resolution of 1/4° for the domain 78°S-50°N, 98°W-20°E, excluding the Mediterranean Sea and the Pacific Ocean. The horizontal resolution remains constant in longitude but varies in latitude. The vertical domain for both grids was discretized in 21 vertical layers. The chosen target potential densities are shown on Table 1. The first layers have a few light target density values that ensure a good vertical resolution in the mixed layer.

The 1/12° model bathymetry was interpolated from the Earth Topography 2 (ETOPO2) and then adjusted with bathymetric information from the Brazilian Navy. In the 1/4° grid, bathymetry was interpolated from General Bathymetric Chart of the Oceans (GEBCO), with 2 min resolution.

On the lateral boundaries of the 1/12° grid, interpolated fields of velocities, temperature, salinity and layer thicknesses from the 1/4° run were applied. Relaxation to climatological SST and SSS were added considering a timescale of 90 days. In the case of the lower resolution model run, relaxation to monthly climatological temperature, salinity and layer thicknesses from World Ocean Atlas (WOA) was applied on the lateral boundaries. On the eastern and western boundaries (South of Africa and Drake Passage, respectively), an open boundary condition was added to

the barotropic velocity. This strategy considers the Antarctic Circumpolar Current, which is necessary to realistically represent currents and water masses in the South Atlantic Ocean. Barotropic fluxes were determined as 110 Sv eastward in the Drake Passage, 10 Sv westward in 10 grid points south of South Africa along 20°E to represent Agulhas eddy and 120 Sv eastward further south to Antarctica representing the Antarctic Circumpolar Current.

Table 1. Potencial Densities of the model. To obtain the volumetric density in kg m^{-3}, 1000 should be added to each target density.

Layer	Potential Density	Layer	Potential Density	Layer	Potential Density
1	19.50	8	24.70	15	27.22
2	20.25	9	25.28	16	27.38
3	21.00	10	25.70	17	27.52
4	21.75	11	26.18	18	27.64
5	22.50	12	26.52	19	27.74
6	23.25	13	26.80	20	27.82
7	24.00	14	27.03	21	27.88

The surface atmospheric forcing employed here were the 6-hourly reanalysis synoptic fields produced by CFSR with 1/4° spatial resolution. It was composed by wind stress at 10 m, air temperature and mixing ratio at 2 m, precipitation, shortwave and longwave fluxes. Latent and sensible heat fluxes were calculated by the model from bulk formulas from CFSR fields and model SST.

Mass flux at the surface was calculated by subtracting precipitation from evaporation and adding a relaxation term to monthly climatological salinity from WOA with a timescale of 30 days. In this case, precipitation is given and evaporation is calculated by the model bulk formulas. Freshwater fluxes from the main rivers were considered using the same approach of precipitation minus evaporation.

The 1/4° HYCOM run was initialized with a climatological mass field from WOA and zero velocities. It was forced by climatological atmospheric fields from the Comprehensive Ocean Atmosphere Data Set (COADS) for 20 years. The result of this spin-up run was taken as initial condition for a forced run in which NCEP/NCAR reanalysis I was employed. On 1 January 1995, the 1/12° HYCOM was initialized by the interpolated HYCOM 1/4° output. From 1995 to 2006, both domains were forced by atmospheric fields from NCEP/NCAR reanalysis I, and from 2006 on, they were forced by CFSR. In addition to the model free runs, the assimilation runs in both grids were performed after 1 January 1998.

2.2. The ocean data assimilation system

RODAS employs the EnOI scheme, which is based on the following equation (Evensen 2003):

$$X^a = X^b + K(Y - HX^b) \tag{1}$$

where X^a represents the analysis, X^b is the model background or prior state, K is the gain matrix, Y is the observational vector, and HX^b is the projection of the prior state onto the observational space by the observational operator H. The term $(Y - HX^b)$ is called the innovation vector and the term $K(Y - HX^b)$ is the analysis increment. The gain matrix K is calculated from the equation

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

where B denotes the ensemble co-variance matrix of the model error and R denotes the diagonal co-variance matrix of the observational error. The term $\alpha \in (0, 1]$ is a scalar that can tune the magnitude of the analysis increment and σ denotes the localization operator applied over B by a Schur product represented by the symbol ∘. In the EnOI scheme, B is calculated from the equation

$$B = \frac{A'A'^T}{(N-1)} \tag{3}$$

18

In equation 3, $A' = [A'^1 A'^2, \ldots, A'^N]$, $A'^m = (X^m - \frac{1}{N} \sum_{n=1}^{N} X^n)$, where A' represents the anomalies, X^m represents the model state vector of the mth ensemble member with m varying from 1 to N, and $N = 126$ is the total number of ensemble members used in all assimilation steps in this study.

In RODAS, the assimilation steps are taken sequentially and independently according to equations 1, 2 and 3. First, SST is assimilated at 00 UTC in order to mostly constrain the mixed layer. Three hours later, at 03 UTC, T/S profiles are assimilated aiming to correct the model termohaline structure around the observations. Finally, at 06 UTC, SLA is assimilated to in order to mainly correct the model mesoscale circulation. This assimilation cycle is performed every 3 days. This strategy reduces the computational cost and allows the model to better diffuse the analysis increments along the integration.

For SST assimilation, daily gridded fields from the Ocean Sea Surface Temperature and Sea Ice Analysis (OSTIA) with 1/20° horizontal resolution were used. These data were made available by the Group for High Resolution Sea Surface Temperature (GHRSST) OSTIA SST data were only assimilated in regions with depths equal to or greater than 30 m.

Different *in situ* observations were used to assimilate T/S profiles. Argo system has provided measurements down to 2000m depth in different regions of Metarea V. Morever, XBT profiles from the NOAA line AX97 associated with the Monitoring the Upper Ocean transport Variability in the Western South Atlantic (MOVAR) project were assimilated. This is a high-resolution line located between Rio de Janeiro and Trindade Island. Finally, CTD data provided by the Brazilian Centro de Pesquisas Leopoldo Américo Miguez de Mello from Petrobras (CENPES) were assimilated. They were collected along the south and southeast coast of Brazil. A 3-day observational window was considered in order to select all valid profiles collected up to 3 days before the assimilation day.

Since HYCOM is formulated with hybrid vertical coordinates, which are not fixed in time and space, assimilating T/S profiles is not trivial. In RODAS, the z-level profiles are projected into the model vertical space to create pseudo-observed layer thicknesses (Δp_{obs}), following the strategy adopted by Thacker and Esenkov (2002),

Xie and Zhu (2010) and Mignac et al. (2015). Thus, each profile is processed as follows. Potential density profiles are calculated by an equation of state for seawater (Brydon et al. 1999) based on a pair of profiles of potential temperature and salinity. The estimated surface density from the observed profile is compared to the top layer target density from the model. If the first density of the profile is higher than the density of the first model layer, a minimum thickness is assigned to the first pseudo-observed layer and temperature and salinity average values are computed. Once water with the target density is found, the remainder of the potential density profile can be partitioned and the layer averages will correspond to the model target densities down to the maximum depth of the profile. The step functions of T, S and Δp_{obs} are the data that will be actually assimilated by the EnOI scheme.

The Argo data were required to step into a data quality control (QC) procedure developed by REMO together with the Brazilian Navy. The date, the location and the temperature and salinity of each Argo profile previously collected were tested according to all criteria stablished by the Global Temperature and Salinity Pilot Program from the Intergovernmental Oceanographic Commission (IOC 1990). The CTD profiles were carefully analysed by CENPES before assimilation and the MOVAR profiles had already passed by a specific data QC realized by National Oceanic and Atmospheric Administration (NOAA).

In the assimilation of SLA, RODAS uses gridded data from the satellites Envisat, Jason-1, Jason-2 and Cryosat distributed by the French *Archiving, Validation et Interpretation des données des Satellites Océanographique* (AVISO). The delayed time product was chosen, which goes through a better data processing and quality control. The spatial resolution of the data is 1/4°. The model SLA was calculated based on the mean sea surface height (SSH) from a six-year period (2008 to 2013) and the observed SLA is based on the mean SSH from a 20-year period (1993 to 2012). Due to this temporal incompatibility, the model SLA presented significant discrepancies with respect to the AVISO SLA. Therefore, in each SLA assimilation step, an offset was calculated by the difference between the area averaged SLA from AVISO and from the background. This offset was used to level observed SLA according to the model. Regions shallower than 300 m were not assimilated owing to altimeters uncertainties.

As RODAS is based on a multivariate EnOI scheme, i.e, the variable that is directly assimilated in each assimilation step will modify other variables in the model state vector. The co-variances between different parameters are calculated from the ensemble members. In the SST assimilation, the analysis update is carried out for the state vector (Δp_i, U_i, V_i, T_i, S_i); i = 1, … , nz, where Δp_i represents layer thickness, U_i and V_i represents the zonal and meridional velocities, T_i and S_i are temperature and salinity defined in the i-th model layer, and nz is the number of model layers. Therefore, assimilation of SST modifies all the baroclinic variables in the model. For T/S profiles, the assimilation is carried out in different steps. First, Δp_{obs} is assimilated, modifying the state vector (Δp_i, U_i, V_i); i = 1, … , nz. Next, temperature and salinity are assimilated in separate steps and in a univariate way, but with previously adjusted layer thicknesses by the assimilation of Δp_{obs}, as in Xie and Zhu (2010). The state vector for SLA assimilation is (Δp_i, U_i, V_i); i = 1, … , nz. In HYCOM, SSH is a diagnosed variable which depends on the layer. Barotropic variables were not altered in any assimilation step.

According to Thacker and Esenkov (2002), within the context of HYCOM, when layer interfaces are altered, potential density should be kept unchanged, and when density is corrected, the layer interfaces should be kept unchanged. Considering that Δp is modified in all assimilation steps, it is necessary to keep the densities unaltered. Therefore, T or S are always diagnosed below the mixed layer by the sea water equation of state in order to keep the potential density constant in the ocean simulated with isopycnal coordinates. For SST assimilation, salinity is diagnosed below the mixed layer. The same goes for XBT profiles in the western South Atlantic, given that temperature is the only observed variable and salinity is indirectly estimated by polynomial functions. In the assimilation of Argo and CTD profiles, temperature was chosen to be diagnosed because in the experiments of Xie and Zhu (2010), most of the temperature corrections were due to changes in the layer thicknesses by the assimilation of Δp_{obs}, while salinity correction is more effective when this variable is assimilated. In the case of SLA assimilation, T and S are not modified, keeping densities unchanged in an isopycnal ocean.

To avoid the analysed layer thicknesses to occasionally become negative, a computationally efficient scheme is performed. If the layer thickness becomes

negative, it is reset to zero and the thickness deficit is added to the neighboring layers. Finally, the sum of the layer thicknesses should be equal to the local depth. For this, the deficit (surplus) is added to (taken from) the last layer of the model.

In the EnOI scheme, the propagation of the observational information is highly dependent on the size and the quality of the ensemble, as the final analysis can be regarded as a combination of the ensemble anomalies whose relative weight is determined by the co-variances. In this work, assimilation of SST, SLA and T/S profiles was performed with 126 ensemble members. They were selected from a long-term model run, corresponding to a 6-year period from 01/01/2008 to 31/12/2013 with assimilation of Argo profiles and SST. In each assimilation step, a different model co-variance matrix was calculated by taking 21 members at 00 UTC for each year of the 6-year period around the date of the corresponding assimilation day, with 3 days between each member.

The localization technique is applied to delete long-distance correlations that may appear in the gain matrix. In this work, the localization operator was separated into a horizontal component (σ_h) and a vertical component (σ_v), and it is defined as $\sigma = \sigma_h \sigma_v$. In order to define the horizontal correlation matrix σ_h, a fifth order function is used:

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

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

In this function, I_{ij} is defined as the Euclidean distance between any two arbitrary points in the horizontal space and L is the horizontal scale of influence. Equation 4 can be considered as a quasi-Gaussian function, forcing the model error co-variance matrix to decrease to zero when the distance I_{ij} reaches $2L$, which corresponds to the radius of localization. In the assimilation of SST and SLA, the radius of localization is 30 km, in the case of Argo and CTD profiles it is 150 km and for XBT data, it is 50 km.

In this first version of RODAS, vertical localization is only applied in the assimilation of T/S profiles and only Δp is localized according to Mignac et al. (2015).The correlation matrix σ_v is defined by the function

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

where $\Delta\rho_{(i,j)}$ is the density difference between two layers i and j and L_ρ is a vertical scale factor defined as 0.5 kg m^{-3},

The diagonal co-variance matrix of the observational error (R) in equation 2 depends on the observation type. The SST and SLA data comes with an observational error field. This data is squared to generate the variances in R. In the case of T/S profiles, the observational errors of T and S in the model layers are calculated as a function of the depth D in meters, respectively, by the following equations (Thacker e Esenkov, 2002):

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

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

The standard deviations of the observational errors of T vary in the vertical from 0.5 °C on the surface to 0.05 °C in the deep ocean. For S, they vary from 0.12 to 0.02. In the case of layer thickness, the standard deviation of Δp_{obs} is calculated depending on the oceanic region. For instance, in the mixed layer, the layer thickness is assigned to the minimum layer thickness allowed by the model configuration, and the standard deviation is defined as 0.05 Δp_k where Δp_k represents the layer thickness at the kth layer calculated from the observed profiles. In the isopycnal layers, the standard deviation is defined by the formula proposed by Thacker and Esenkov (2002):

$$SD(DP_k) = max\left\{0.5\delta p_k, max\left[0.05DP_k, DP_k\left(0.05 + (0.5 - 0.05)\frac{sd_{\sigma(k)}}{SD_{\sigma(k)}}\right)\right]\right\} \tag{8}$$

Where δp_k is the minimum layer thickness specified by the model configuration for the kth layer, $sd_{\sigma(k)}$ is the minimum standard deviation of the potential density defined as 0.001 kg m^{-3}, and $SD_{\sigma(k)}$ is the standard deviation of the potential density from observations. The latter should be small when the potential density from the observed

layer thickness has values close to its target density. $SD(DP_k)$, $SD_T(D)$ and $SD_S(D)$ are squared to generate the co-variances in R.

Concerning the α term in equation 2, sensitivity tests were performed in each assimilation step. The highest value ($\alpha=1$) was stablished for SLA assimilation in order to avoid K producing too much weight on the background, since the observational error can be ten times higher than the instrumentation error (~0.03 m) in some regions. For assimilation of SST and XBT profiles, $\alpha=0.3$ and for the remaining T/S profiles, $\alpha=0.5$.

2.3. Numerical experiments and evaluation metrics

In order to investigate the impact of RODAS on the quality of the forecast, two integrations were performed from 1 January 2011 to 31 December 2012. The first one was a control run without assimilation, hereafter called free run, and the second one was an assimilation run using RODAS. The experiments were objectively evaluated by the root mean square deviation (RMSD) and correlation based on observational data from OSTIA, AVISO and Argo.

After that, an assessment of the extended-range predictive skill of the system was carried out considering three different experiments on the same 1/12° grid. They were: i) 48 HYCOM 30-day hindcasts initialized from RODAS analysed fields. The initial condition was taken from the assimilation run described above, at 06 UTC right after assimilation and the model was integrated for 30 days with no assimilation. From 1 January 2011 to 11 December 2012, a new 30-day hindcast was performed every 15 days, totalizing 48 cycles (i.e.: from 01/01/2011 to 31/01/2011; from 16/01/2011 to 15/02/2011 and so on); ii) the free run described above, which had no assimilation at all; and iii) persistence, which represents a forecast of no change from the initial condition of the hindcasts. It was included because it enables us to assess whether, given the initial conditions, the model adds any skill during the following 30 days.

The 2-year RODAS assimilation run was used as reference to evaluate the three experiments. They were compared against the analysed fields that are valid for each day of the 30-day cycle and the RMSD and correlation between these fields were calculated. The average of the 48 cycles was plotted as a function of lead time.

To compare the hindcast and persistence, a skill metric was used, whereby the hindcast provides useful skill if the skill score is positive.

Skill score = 1 − (RMSE$_{hindcast}$/RMSE$_{persistence}$)

We must take into account that the RODAS analysis errors were not included in the evaluation. Therefore, these statistics do not give the overall magnitude of the errors, but they can give a bound on the expected accuracy of the hindcast and provide information about the evolution of errors in time, as in Chassignet et al. (2009), Hurlburt et al. (2009), Martin (2011) and Hernandez et al. (2015).

Lastly, in order to assess the model skill in predicting mesoscale circulation features in Metarea V, a focus on coastal upwelling was given. A subregion encompassing the upwelling area near CF (19°S to 28°S and 36°W up to the Brazilian coast) was evaluated with respect to OSTIA SST in spring and summer months (September to March), when coastal upwelling is more likely to occur. The hindcast, persistence and free run RMSD and correlation were calculated over the 30-day cycles initialized in the upwelling-favourable months of 2011 and 2012. The average of 26 cycles was plotted against lead time. After that, a case study was carried out in order to assess the hindcast skill in a qualitative manner. A well-developed upwelling event was chosen based on OSTIA SST fields and the overall structure of the event was investigated in the hindcast and free run.

3. Results and discussion

3.1. Quantitative assessment

The assimilation impact was evaluated by calculating the RMSD of SST with respect to OSTIA and the SSH RMSD and SLA correlation with respect to AVISO over Metarea V. Figure 1 shows the average of RODAS analysis and free run from 1 January 2011 to 31 December 2012. A substantial reduction of the SST RMSD is observed in the analysis when compared to the free run in almost all Metarea V. The area averaged RMSD was 0.88°C and 0.43°C in the free run and in the RODAS run, respectively. In some regions, such as around the Brazil-Malvinas Confluence (BMC), the RMSD was reduced by more than 75%, from 1.63°C to 0.38°C. In the northern region of the

domain, the free run produced large errors. This is possibly associated with a model temperature warm bias, which is more pronounced near the Equator. RODAS was able to effectively constrain the temperature in this region, so that the RMSD reduction corresponded to 79%, from 1.90°C to 0.39°C. Near the coast, the analysis obtained the highest RMSD values, probably due to the lack of assimilation over small depths. However, RODAS was still able to reduce the error by more than 25% in most of the area along the coast, which indicates that the model is able to extrapolate assimilation information.

Concerning the SSH RMSD, Figures 1.c and 1.d show that overall RODAS attained a smaller error than the free run, reducing the mean RMSD from 0.07 m to 0.05 m. For instance, near the BMC, the RMSD reduction was approximately 76%, from 0.18 m to 0.04 m, and north of 3°N the error was reduced by about 70%, from 0.14 m to 0.04 m. Near the coast, the error was relatively high in both experiments. Around river mouths, the analysis RMSD was greater than 0.25 m. However, it must be taken into account that due to altimeters uncertainties, no altimetric data is assimilated in regions shallower than 300 m.

The SLA correlation map shows that assimilation had a significant impact in the analysis, increasing the overall correlation from 0.44 in the free run to 0.77 in the RODAS run [Figure 1 (e) and (f)]. Free run correlation is fairly low and sometimes negative. Only in areas of low variability, free run correlation is higher than 0.6, indicating that the model by itself is unable to reproduce mesoscale circulation at the surface. When RODAS was applied, SLA correlation substantially increased. It is greater than 0.6 in most of Metarea V and in some regions it is close to 1. Near the coast, correlation remains relatively low, which can be associated with the lack of data assimilation in shallow regions. In addition, the region around the BMC also obtained small analysis correlation values, since this is an area of high variability and remains a challenge to RODAS. Still, the analysis correlation was much higher than the free run correlation, evidencing that this lower skill did not compromise the substantial positive impact of the assimilation in representing SLA variability.
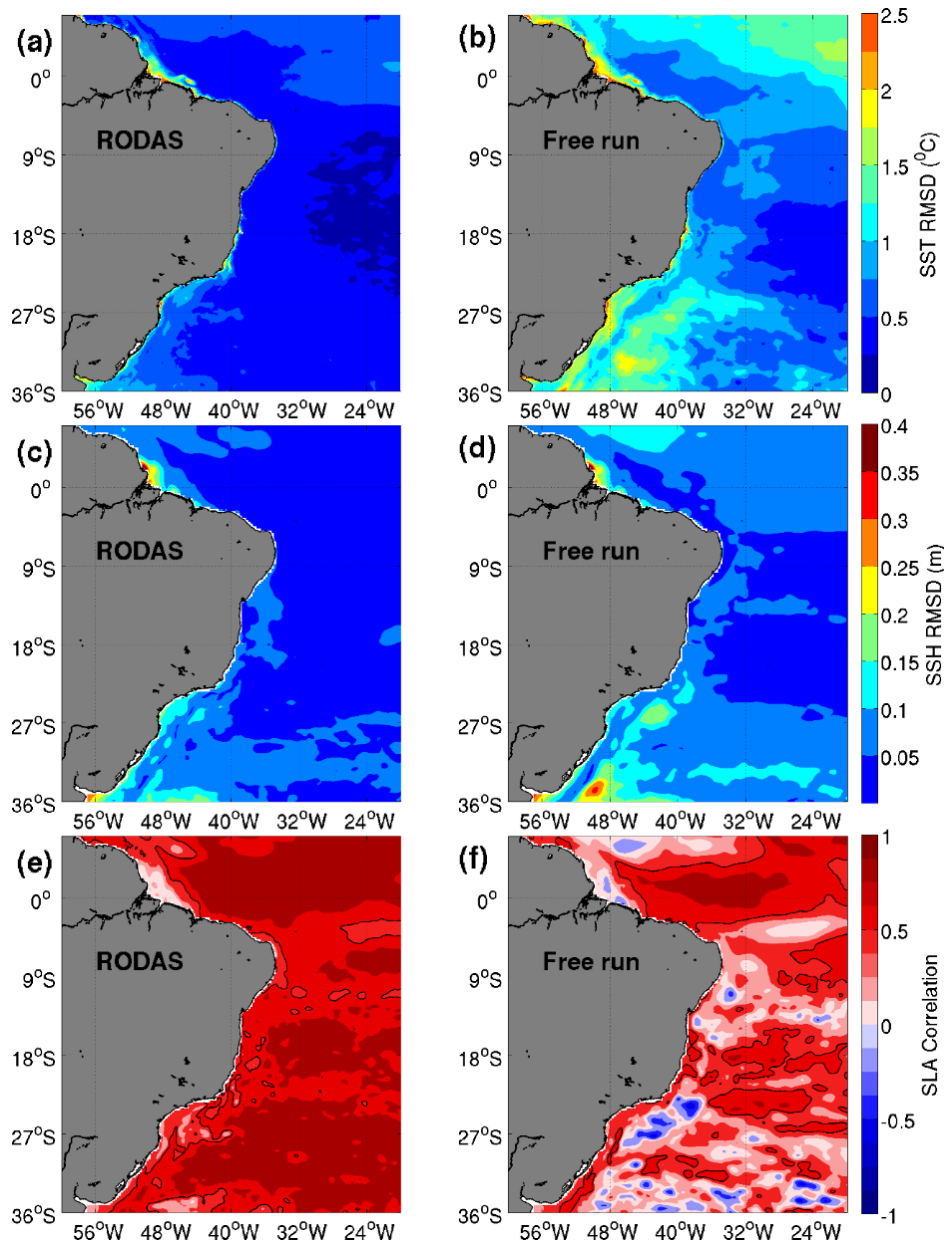
Figure 1. SST RMSD (°C) with respect to OSTIA [(a) and (b)], SSH RMSD (m) with respect to AVISO [(c) and (d)] and SLA correlation with respect to AVISO [(e) and (f)] for RODAS analysis and the free run from 1 January 2011 to 31 December 2012. The black contour represents 0.6 correlation

To assess the assimilation impact in the subsurface temperature and salinity, the vertical mean profile of RMSD with respect to Argo T/S data was calculated for RODAS and the free run from 1 January 2011 to 31 December 2012 (Figure 2). The average temperature RMSD in the profile decreased from 1.03°C in the free run to 0.73°C in the analysis. This indicates a positive assimilation impact in the subsurface, which was more pronounced in the top 200 m. The largest RMSD decrease was around 75 m depth, where the error was reduced by 39%, from 2.39°C in the free run to 1.46°C in RODAS. This is also where the highest RMSD was observed in both experiments. It is associated with the thermocline regions of sharp vertical gradients, which models have difficulties to represent (Xie and Zhu 2010; Costa and Tanajura 2015; Mignac et al. 2015), particularly the present HYCOM configuration with only 21 layers. Below 1750 m the RMSD was 20% higher in the analysis than in the free run. This increase might be due to the smaller amount of data at this depth, the absence of data below 2000 m and uncertainties in the calculated Δp_{obs}.

In the subsurface salinity, the analysis RMSD was reduced in the whole profile compared to the free run. The mean RMSD was 0.18 and 0.13 in the free run and RODAS, respectively. The biggest impact was around 150 m, where the RMSD reduction corresponded to 29%, from 0.32 in the free run to 0.22 in the analysis. As mentioned above, models have difficulties in reproducing sharp gradients and this is also associated with the halocline. Differently from the temperature, there is no assimilation of sea surface salinity. Therefore, there is a high analysis error at the surface and down to 150 m, where it oscillates between 0.22 and 0.24. The free run RMSD oscillates between 0.28 and 0.32 at the same depth range. Below that, the RMSD gradually decreases in both experiments down to 900 m depth, where RODAS reaches 0.05 and the free run reaches 0.10. RODAS surface salinity might decrease even more if SSS from SMOS and Aquarius satellites are assimilated. Overall, data assimilation effectively constrains the model towards observation and it has been demonstrated that RODAS has a relatively good skill in reproducing the ocean thermohaline state.
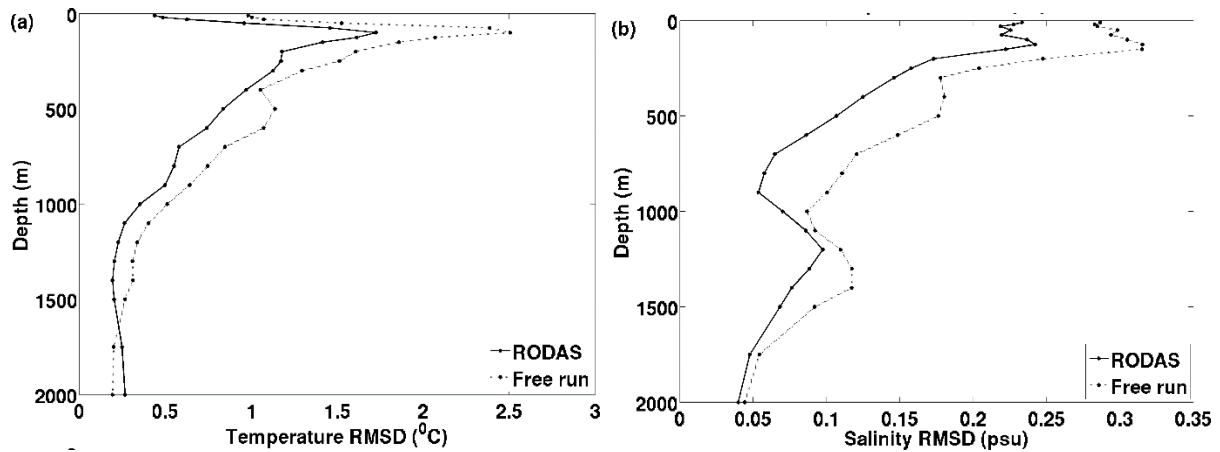
Figure 2. Vertical mean profiles of RMSD with respect to Argo T/S data from 1 January 2011 to 31 December 2012 for (a) temperature (°C) and (b) salinity over Metarea V.

3.2 Assessment of the extended-range predictive skill

To provide an estimate of the decrease of the HYCOM+RODAS extended forecast skill, Figure 3 shows the RMSD and correlation of SST and SLA averaged over all 48 30-day hindcasts with respect to the RODAS run over Metarea V. It also shows the skill of persistence and of the free run. For both SST and SLA, the model hindcast had the lowest RMSD and the highest correlation throughout the 30-day windows. The evolution of SST RMSD shows that the hindcast slowly degrades, while persistence does it much more quickly, reaching the free run values by the 22nd day. By the 30th day, the persistence error reaches 1.09°C, while the hindcast RMSD is 0.46°C. The free run error is almost constant and the RMSD values are about 0.88°C at all lead times. The hindcast SST correlation remains high until the last day, ranging from 1 to 0.94 and persistence correlation rapidly decreases from 1 to 0.75, taking up 11 days to reach the free run values. In the latter, average correlation is 0.89. Thus, the HYCOM+RODAS hindcast typically provides a reliable SST estimate over a 30-day window.

In the case of SLA, the sharpest RMSD increase was in the first 24h, when the hindcast value grew up to 0.02 m and the persistence error increased to 0.05 m. Thereafter, the RMSDs increased at a similar rate, reaching 0.05 m in the hindcast and 0.08 m in persistence by the 30th day. The free run had the highest RMSD at all lead times, with

29

an average of 0.13 m. The hindcast SLA correlation is consistently higher than persistence correlation, ranging from 1 to 0.79, while the latter ranges from 1 to 0.41. Once again, persistence degrades fairly quickly in the first 24 h, when SLA correlation reaches 0.84, but it keeps degrading faster than the hindcast until the end of the cycle. By the 16th day, persistence SLA correlation is less than 60% and by the 29th day, it is less than the free run correlation. The latter in turn, oscillates between 0.44 and 0.48. In general, the SLA errors grow more quickly than the SST errors. Although different assimilation systems and ocean models might have very different results, previous studies have had similar results, such as in Oke et al. (2015a), where observations are withheld from the forecast system to provide an estimate of the decrease in analysis and forecast skill.

Since the hindcast was better than persistence at all lead times, the results suggest that the model adds some skill to the initialised state over a 30-day forecast. Also, the errors of the SST and SLA hindcasts are smaller than the errors of the free run. As expected, it demonstrates the benefit of data assimilation on the forecast skill and shows that an accurate initial condition considerably enhances predictability. Moreover, it gives an indication of the system tolerance to observation dropouts, as the hindcast error takes more than a month to saturate and reach the free run values. However, it must not be forgotten that the model was forced by atmospheric reanalysis fields. Therefore, it is not expected that the quality of these fields degrade over each forecast cycle as they would if the atmospheric forcing was a true weather forecast.
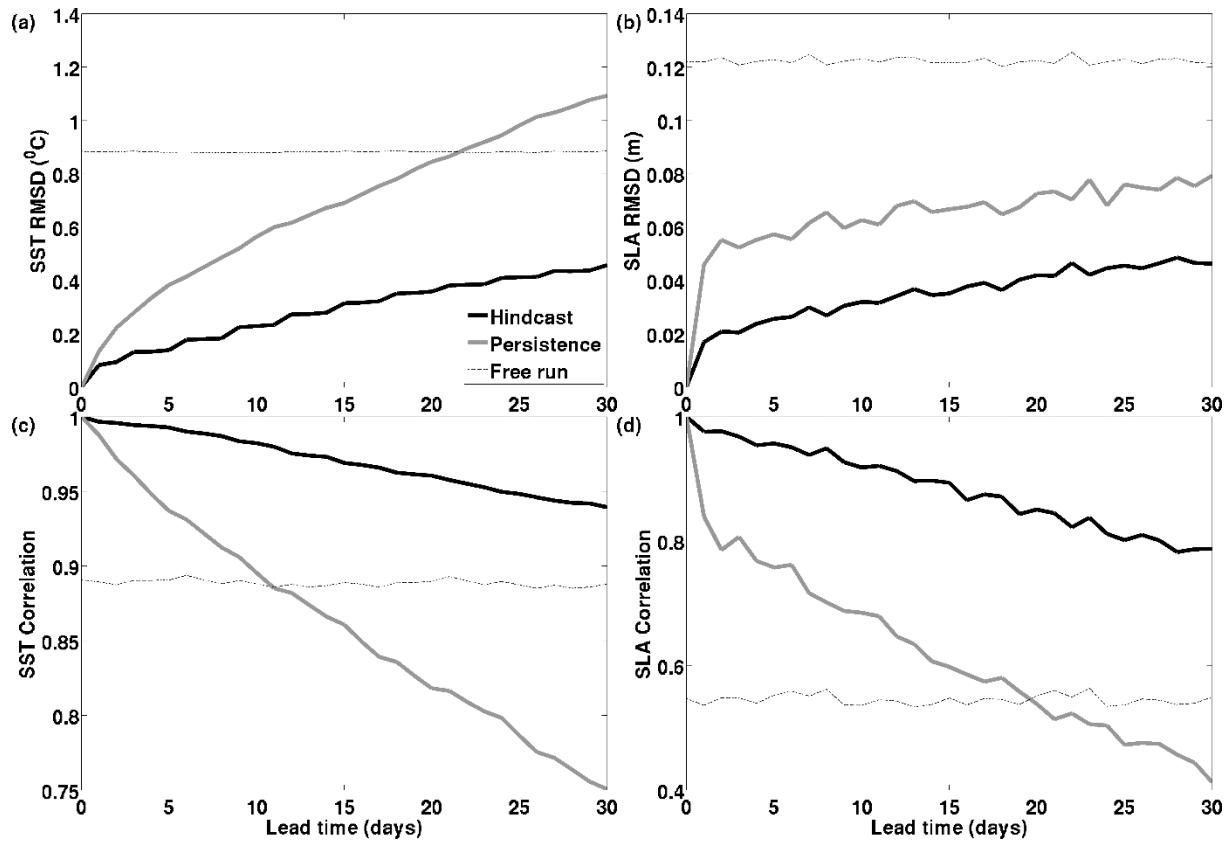
Figure 3. RMSD [(a) and (b)] and correlation [(c) and (d)] of SST and SLA with respect to RODAS analysis over 48 30-day cycles averaged over Metarea V.

Predictability was also assessed using a skill score calculated for SST and SLA over Metarea V (Figure 4). The hindcast has a strong positive skill against persistence for SST. Even for 30-day lead time, the hindcast performs better than persistence in most part of Metarea V and fails only in the north of the equator. This is the area where the model has a stronger temperature bias, but RODAS efficiently corrects it [Figure 1(a)]. As the lead time increases, the bias becomes more evident, increasing the hindcast RMSD and consequently, reducing its skill compared to persistence. Nonetheless, in 10-day lead time, the area with negative skill score is very small and even in the 30th day, the average skill score is positive. Thus, SST hindcast initialized with RODAS is both highly accurate and skilful.

The SLA skill score is positive overall, but some differences are observed when looking at specific regions. For instance, in day 10 the skill score is slightly negative in small areas east of 32°W and in day 30 the negative values are also observed around the BMC. In this region, the hindcast RMSD is not lower than that of persistence due to the fact that the flow is dominated by mesoscale flow instabilities, rather than by

31

atmospheric forcing. As explained above, this region has high variability and represents a challenge for the model. However, as shown in Figure 1, RODAS is able substantially reduce the error around the BMC, and the hindcast retains some skill out to at least 10 days after initialization. Near the south and southeast coast of Brazil, a highly positive skill was obtained at all lead times. As observed in previous studies, in shallow waters and in the surface mixed layer, the ocean state is very sensitive to atmospheric forcing, rather than to the initial condition, therefore persistence quickly loses skill (Zhu 2011; Martin 2011; Hurlburt et al. 2009).
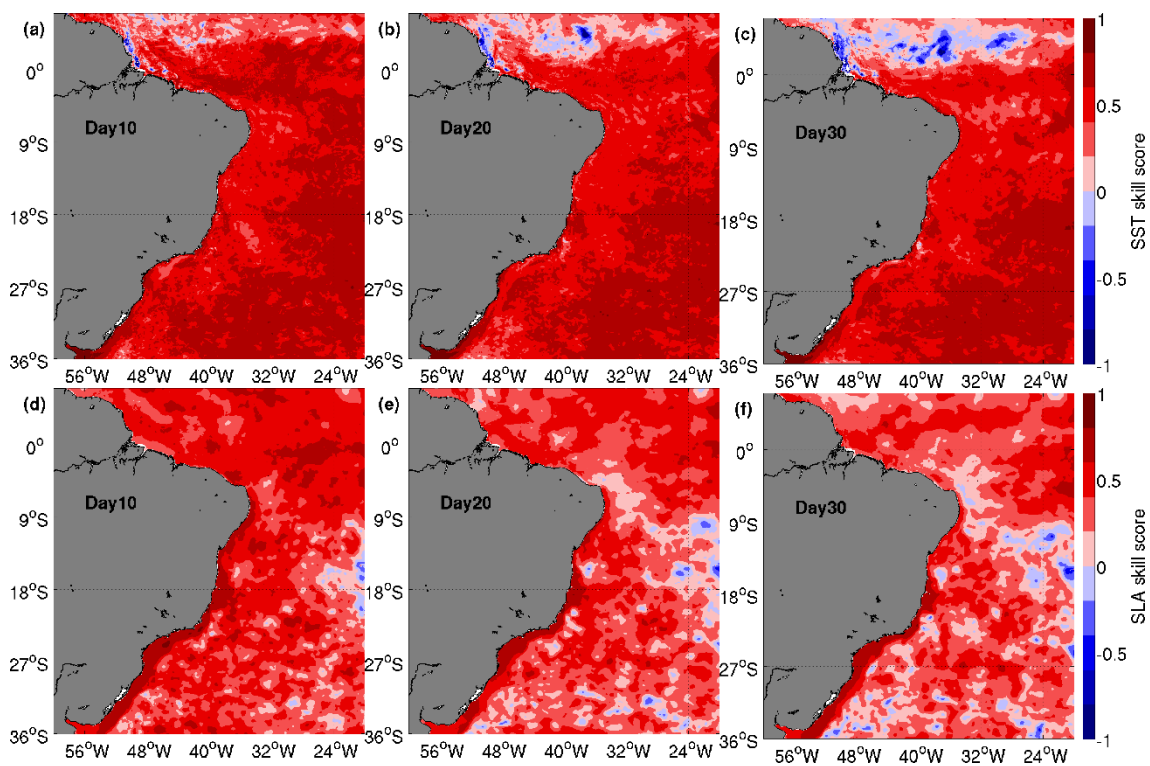

Figure 4. Skill score comparing hindcast and persistence RMSD statistics against RODAS analysis for SST [(a), (b) and (c)] and SLA [(d), (e) and (f)] in 10-, 20- and 30-day lead times.

Figure 5 represents the evolution of the temperature and salinity RMSD in the subsurface for hindcast and persistence. The larger errors were in the upper ocean associated with both the mixed layer and the thermocline/halocline, and there is a gradual decrease in errors with depth, in line with a decrease in variability in the deeper

ocean. This behaviour is in accordance with previous studies (Divakaran et al. 2015). As expected, persistence degrades much more quickly than hindcast near the surface. Around 10 m depth, persistence had the highest temperature RMSD values and reached 0.76°C in day 30, when the hindcast RMSD reached 0.16°C. It is evident that persistence rapidly loses skill near the surface, corroborating with earlier results. Around 100 m depth the hindcast had the largest errors, reaching 0.19°C in day 28. This is associated with the thermocline and is in accordance with what was shown in Figure 2. Persistence took up 17 days to reach the same RMSD value at this depth, indicating a positive hindcast skill even in the thermocline. In the deep ocean, the temperature RMSD decreases very slowly in both experiments. In 1000 m depth, persistence RMSD took up 8 days to reach 0.01°C and the hindcast took up 13 days to reach the same value (not shown). As described by Hurlburt et al. (2009), the deep ocean variability is non-deterministic with respect to atmospheric forcing, and the time scale for predictive skill depends mostly on the quality of the initial state, the accuracy of the model dynamics and the flow instability. In this region, the flow is dominated by the termohaline circulation, which evolves on longer time scales. The free run in turn kept the RMSD values almost constant throughout the 30 days, with the highest values (1.81°C) around 100 m depth, near the thermocline. At all depths the free run error was higher than those of the hindcast, with an average of 0.63°C at 10 m depth and 0.25°C at 1000 m depth (not shown). Thus, HYCOM+RODAS skill in predicting temperature is not restricted to the surface, but extends to the deep ocean as well.

Similarly to the temperature, salinity RMSD in persistence was highest near the surface, where the values reached 0.06 in 30-day lead time. In the hindcast the error was not higher than 0.01 in any lead time at the surface. Between 150 m and 200 m depth, hindcast produced the largest error reaching 0.02 in the 26th day and persistence reached the same value in the 17th day. This depth is associated with the halocline and, as for the temperature, this is in agreement with what was observed in Figure 2. Below this depth, the error was fairly small in both experiments. At 500 m depth, both hindcast and persistence RMSD took up 16 days to reach 0.01. In the free run, salinity RMSD was almost constant through time, but varied with depth, ranging from 0.01 at 5000 m to 0.11 at 300 m (not shown). These are reasonable results as salinity is a tracer quantity, which evolves on longer timescales than temperature in large parts of the ocean (Ryan et al. 2015).
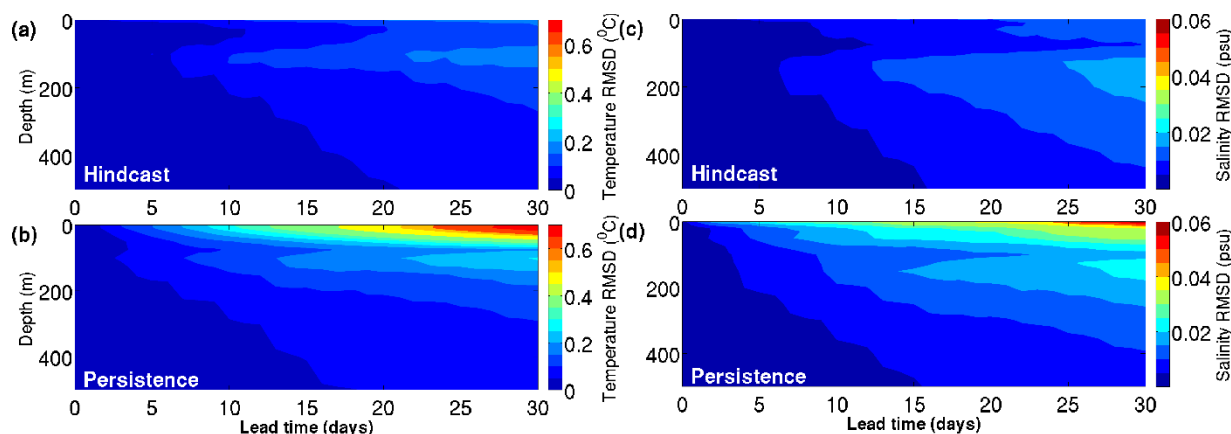
Figure 5. RMSD of the subsurface temperature (°C) [(a) and (b)] and salinity [(c) and (d)] with respect to RODAS over 48 30-day cycles averaged in Metarea V.

3.3. Assessment of HYCOM+RODAS in the coastal upwelling subregion and case study

Daily OSTIA SST fields were analysed from 01 January 2011 to 31 December 2012 in the upwelling subregion. During spring and summer months, coastal upwelling was frequently observed, in either intense and long-lived events or just ephemeral ones (not shown). Figure 6 shows the average SST RMSD and correlation of 26 30-day cycles initialized in the upwelling-favourable months (September to March) with respect to OSTIA over the upwelling subregion. Hindcast and persistence start with a RMSD of 0.51°C in the initial condition, while the free run starts with a RMSD of 1.08°C. During the first 17 days, persistence and hindcast RMSDs remain close, but after that the rate of RMSD increase by the hindcast diminishes. After 30 days, hindcast and persistence RMSD reach 0.94°C and 1.09°C, respectively. The free run RMSD remains nearly constant around 1.04°C. For the SST correlation, hindcast and persistence start at 0.97 and after 30 days it decreases to 0.94 and 0.86, respectively. From the 5th day on, the hindcast correlation is higher than that of persistence and it takes up 13 days for persistence to reach free run values. The latter corresponded to 0.94 on average. Persistence correlation quickly decreases as it is unable to represent the development and evolution of the upwelling processes that occurred during the spring and summer.

Since coastal upwelling is primarily induced by wind-driven mechanisms, the model representation of these events is strongly dependent on atmospheric forcing. Northeasterly winds are highly favourable for the occurrence of upwelling in this region and consequently cause SST variations (Rodrigues and Lorenzzetti 2001; Palóczy et al. 2010; Aguiar et al. 2014). Thus, both hindcast and free run have good correlation values (higher than 0.9) and they vary in a very similar way. This is a good indication that overall the model is capable of capturing and reproducing coastal upwelling events at the surface. There is a decrease in RMSD from the hindcast to the free run, suggesting that the first is more efficient than the latter.
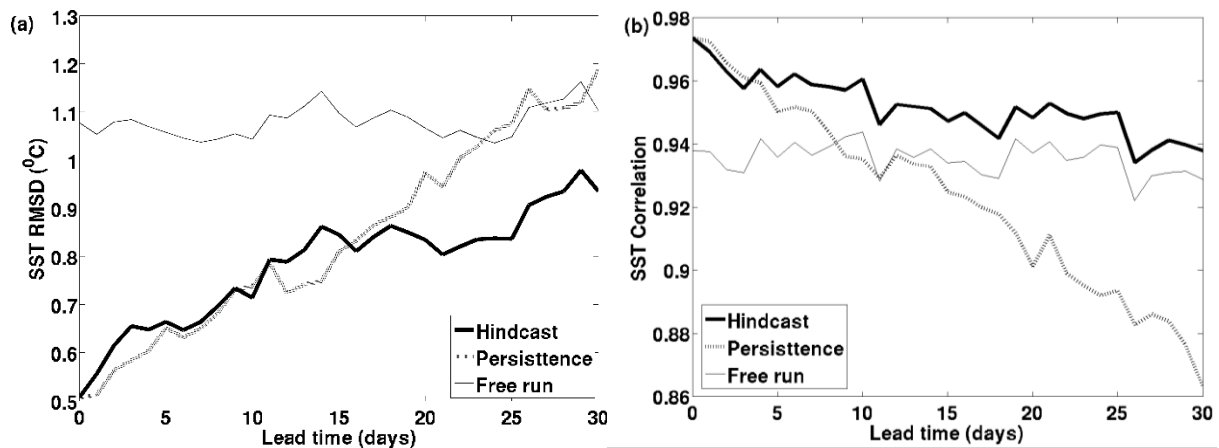

Figure 6. SST RMSD (^{o}C) (a) and correlation (b) with respect to OSTIA over the upwelling subregion for the hindcast, persistence and free run initialized in spring and summer months of 2011 and 2012.

For a more detailed and qualitative assessment, a specific coastal upwelling event that occurred on 23 February 2011 was investigated. Figure 7 shows the SST fields from OSTIA, free run and two hindcasts, one initialized on 15 February 2011 and the other on 31 January 2011, corresponding to 8-day and 23-day lead time, respectively. Generally, the modelled SST fields compare well with the observation as they all capture a conspicuous surface signal of the event. The upwelling plume in OSTIA roughly follows the shelf break and extends to approximately 45°W, with a minimum temperature of 24.28°C. In the model fields, the event is more intense, with minimum temperatures of 20.34°C in the 8-day hindcast, 19.51°C in the 23-day hindcast and

35

19.22°C in free run. This might suggest that the model possibly overestimates the upwelling event, with temperatures colder than the observed ones. However, *in situ* data provided by the Brazilian National Buoy Program (Programa Nacional de Bóias - PNBOIA) between 13 March 2012 and 26 February 2013 were analysed. The data was collected by the buoy located near CF, at 22.98°S, 42.10°W. Figure 8 shows the 3-day mean SST of the buoy, OSTIA, free run and RODAS assimilation run in the same location and period. OSTIA SST was on average 2.76°C higher than *in situ* measurements. Therefore, it is reasonable to assume that OSTIA satellites are often unable to capture the real magnitude of the upwelling events at that location. In addition, as the wind is the primary forcing of coastal upwelling, the event is well developed in the modelled SST fields. Recall that the model was forced by reanalysis fields. Therefore these simulations represent the best possible predictions, considering all available atmospheric forcings. The importance of the ocean initial condition is associated with a good representation of the themohaline structure to produce an accurate temperature of the subsurface water that would be upwelled.

Furthermore, free run shows an event with larger horizontal dimension than OSTIA while in the hindcasts, the shape of the plume is more constrained towards observation. As expected, the upwelling plume in the 23-day hindcast is slightly larger and colder than it is in the 8-day hindcast. However, the 23-day hindcast is still closer to the observation than the free run, suggesting that the model is able to retain some skill for more than 23 days. In addition, both hindcasts present an offshore stretched cold water filament at 40°W, which indicates a cyclonic circulation pattern. In AVISO SLA field, a cyclonic eddy is present at the same location [Figure 7(e)]. The hindcasts were able to capture a signal of this feature, in both SST and SLA fields, while free run did not capture it at all. It is well established in the literature that coastal and oceanic systems do interact and the presence of cyclonic meanders/eddies can cause or enhance coastal upwelling (Campos et al. 2000; Calado et al. 2010; Aguiar et al. 2014; Palóczy et al. 2014). It is also of note the clear warm bias the model develops, since the model SST in the 23-day lead time between 36° and 39°W is higher than in the 8-day lead time and the model free run is much higher than both hindcasts and observations.
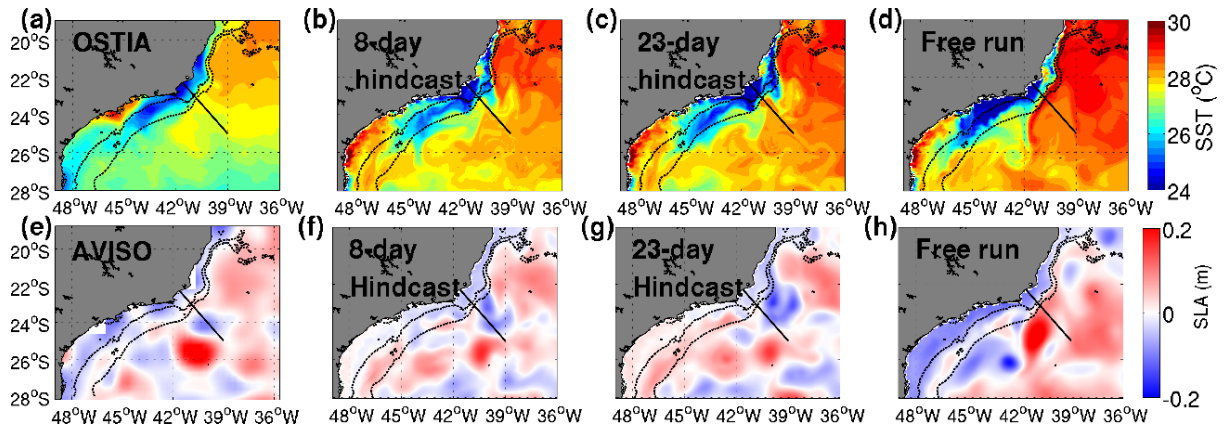
Figure 7. SST (°C) [(a) to (d)] and SLA (m) [(e) to (h)] fields during the upwelling case study on 23 February 2011. The solid black line represents the position of the vertical section. The black dashed lines represent the 100 m and 1000 m isobaths.
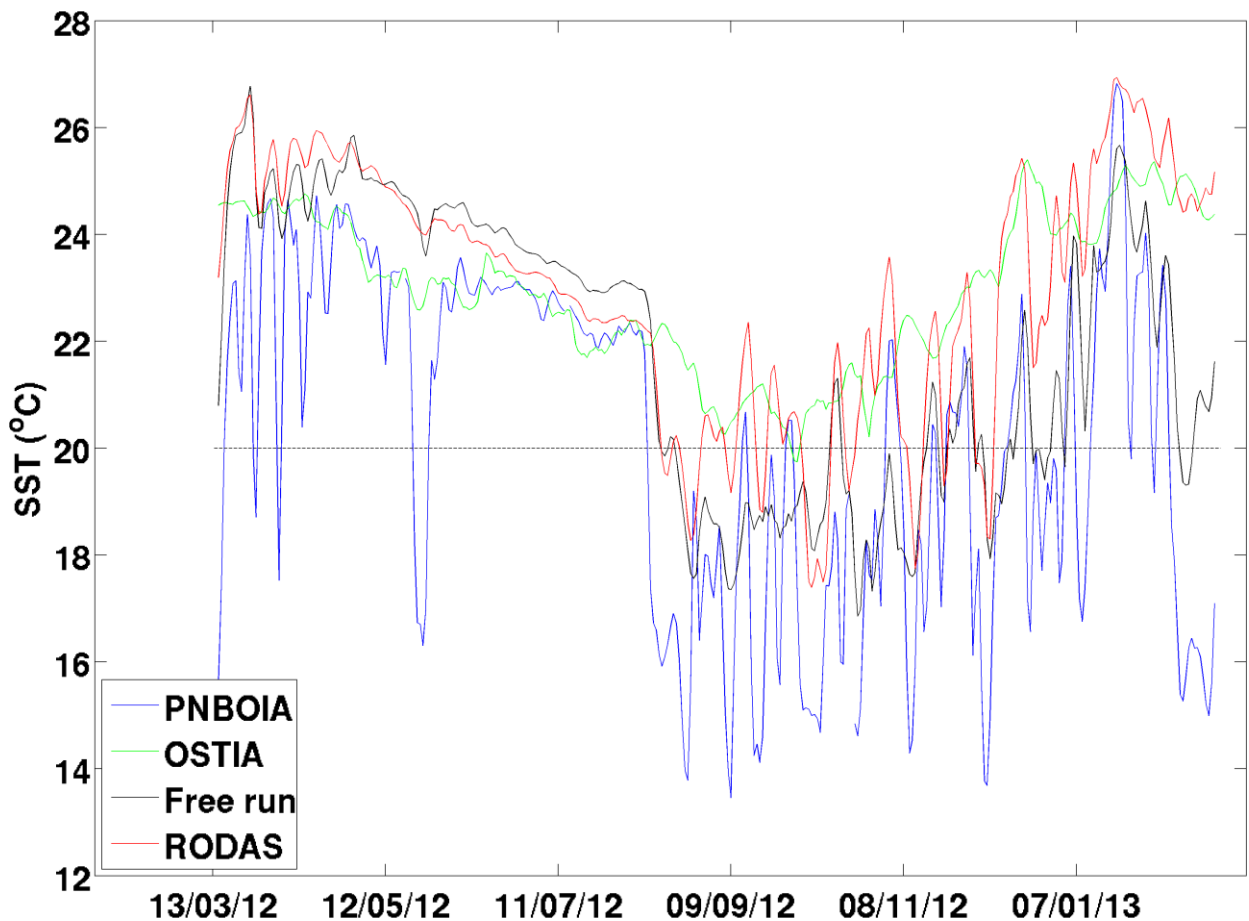

Figure 8. 3-day mean SST (°C) at the buoy location (22.98°S, 42.10°W)

In the subsurface, the model simulations show slope water intrusions in the continental shelf, as the 20°C isotherm - an indication of the upper limit of SACW - is uplifted to depths shallower than 50 m (Figure 9). This suggests that shelf-break upwelling was taking place. According to Aguiar et al. (2014), this is a permanent feature throughout the year in the region and it is an important current-driven upwelling mechanism. The interaction between the BC and the continental shelf enhances shelf-ocean exchange and preconditions the isotherms as a complementary part of the coastal upwelling process (Aguiar et al. 2014, Palóczy et al. 2014). Moreover, cyclonic eddies also contribute to this process as they raise the isotherms, through upward Ekman pumping due to surface divergence (Campos et al. 2000; Calado et al. 2010; Aguiar et al. 2014; Palóczy et al. 2014).

In the vertical sections of both hindcasts, 16°C temperatures are observed at 50 m depth, suggesting that SACW was highly available at the continental shelf. Furthermore, the cyclonic eddy is evident from the uplift of the isotherms near 40°W, matching the position in AVISO SLA field [Figure 7(d)]. In the 23-day hindcast the eddy is about 0.3° further west than in the 8-day hindcast. Still, both sections are very similar, reinforcing that the hindcast has a longer predictive skill in the subsurface. In the free run section, the 18°C isotherm barely reaches the shelf and the 16°C isotherm is only raised to 150 m depth. Once again, this simulation did not capture the cyclonic eddy. Moreover, in the free run the mixed layer depth is deeper and the thermocline is much more diffuse than in the hindcasts. As indicated above by the vertical profiles of temperature RMSDs in the free run and in the HYCOM+RODAS run (Figure 2), the assimilation system consistently improves the termohaline state of the ocean and in Figure 9 its impact on the quality of the hindcast is evident.

Even though free run is capable of reproducing coastal upwelling at the surface, it is not very efficient at the subsurface, leaving out some important current-driven mechanisms. In contrast, the hindcast has a good representation of oceanic and coastal processes acting together during a coastal upwelling event. Similar results were obtained when investigating other upwelling events in the region, such as the ones on 24 February 2012, 27 October 2012 and 29 November 2012 (not shown). Therefore, it can be concluded that overall the hindcasts initialized by RODAS are

efficiently able to reproduce coastal upwelling as a whole, and the upwelling predictive skill for the HYCOM+RODAS system in this region extends up to a month.
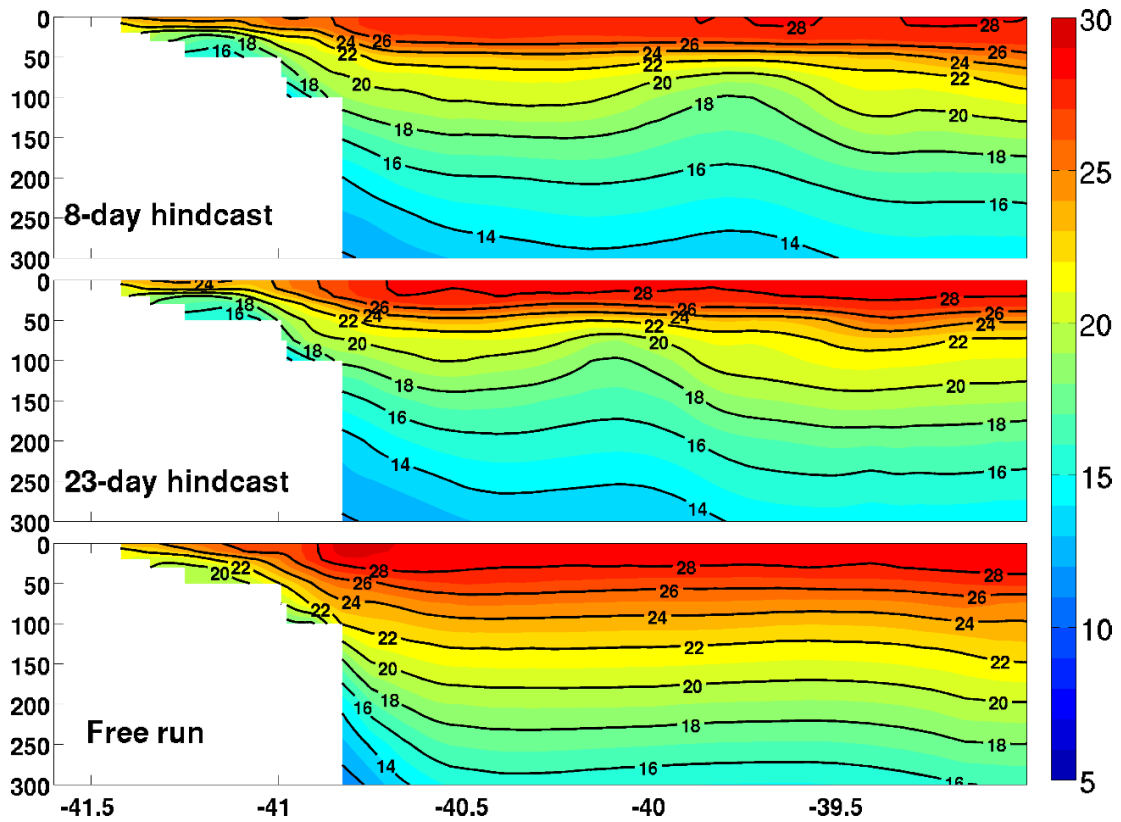

Figure 9. Vertical sections of temperature (°C) during the case study upwelling event on 23 February 2011. The location of the section is shown in Figure 7.

4. Conclusion

HYCOM+RODAS system was quantitatively assessed and compared to free run over the Atlantic Ocean Metarea V in 2011 and 2012. The results showed that overall data assimilation had a positive impact in the representation of the ocean state, reducing the RMSD of SST and SSH with respect to observations by 48% and 13%, respectively, and increasing the overall SLA correlation by 65%. A large improvement was obtained around the BMC, which is an area of high variability, hence large forecast errors. In addition, a substantial positive impact was attained in the northern region of the domain, due to a model temperature bias, which RODAS was able to correct. Near the coast, the analysis obtained the highest RMSD and lowest correlation values, and this is probably associated with the lack of assimilated data in shallow regions. Nevertheless, RODAS was still able to reduce the SST error and increase the SLA

correlation in coastal regions compared to the free run, suggesting that the assimilation system is able to extrapolate observation information. The vertical profiles showed that down to 2000 m depth the average RMSD of temperature and salinity was reduced by 29% and 28%, respectively, from the free run to the analysis. The largest errors, as well as the largest corrections were at the thermocline and halocline regions. Overall, it has been demonstrated that HYCOM+RODAS system was able to efficiently reproduce the ocean state over Metarea V.

The predictive skill of 30-day hindcasts initialized with RODAS analysis was evaluated against the analysed fields and compared to persistence and free run. The results revealed that for SST and SLA, the model hindcast had the lowest RMSD and highest correlation throughout the 30 days. It was observed that persistence degrades much more quickly than the hindcast, with a RMSD increase of 1.09°C for SST and 0.08 m for SLA over 30 days, while the hindcast had an increase of 0.46°C and 0.05 m over the same period. This suggests that the model typically adds some skill to the initialised state over a 30-day forecast. The free run RMSD (correlation) was always higher (lower) than that of the hindcast, demonstrating that an accurate initial condition considerably enhances predictability and the hindcast error takes more than a month to saturate.

The subsurface temperature and salinity RMSD have shown that in small depths persistence RMSD rapidly increases to 0.76°C and 0.06 over 30 days, while hindcast RMSD increases to 0.16°C and 0.01. In contrast, the deep ocean temperature and salinity error was fairly low, reaching a RMSD of 0.01°C and 0.01 after 30 days in both simulations. These results are in good agreement with previous studies, which point out that in shallow waters and in the surface mixed layer, the ocean state is very sensitive to atmospheric forcing, therefore persistence quickly loses skill. However, in the deep ocean the time scale for predictive skill is not so dependent on atmospheric forcing, but rather on the quality of the initial state, the accuracy of the model dynamics and the time scale of the flow instability (Zhu 2011; Martin 2011; Hurlburt et al. 2009). Overall, we conclude that the HYCOM hindcast initialized with RODAS analysis typically provides a good quantitative estimate of the ocean state over a 30-day window.

When investigating the upwelling subregion off the Brazilian southeast coast in spring and summer, it was observed that the hindcast and free run have high SST correlation (>0.9) with respect to OSTIA. Since coastal upwelling is primarily induced by the wind, atmospheric forcing provides the model a good skill in reproducing these events, whereas persistence is unable to reproduce their development. The SST RMSD was reduced from the free run to the hindcast by 0.57°C in the initial condition and by 0.14°C in 30-day lead time, demonstrating the positive impact of data assimilation also in this specific application.

The case study has shown that, due to atmospheric forcing, free run is able to consistently reproduce the coastal upwelling event at the surface. However, it is well stablished in literature that coastal upwelling usually results from the combination of wind-driven and current-driven mechanisms. When looking at the subsurface, the event is very subtle in the free run. In contrast, both 8- and 23- day hindcasts represent a well-developed event in the surface and subsurface, and capture a signal of a cyclonic eddy, which is in accordance with AVISO SLA field. In the hindcasts, the termohaline state of the ocean was consistently improved and the interaction of coastal and oceanic processes during the event is clear. It demonstrates that HYCOM+RODAS system efficiently reproduces coastal upwelling as a whole and it has a good predictive skill. However, these simulations represent the lowest bound of the ocean model errors considering all possible atmospheric forcings. For a real predictability assessment, true weather forecasts should be employed instead of reanalysis fields.

Acknowledgements

References

Aguiar A, Cirano M, Pereira J, Marta-Almeida M. 2014. Upwelling processes along a western boundary current in the Abrolhos–Campos region of Brazil. *Continental Shelf Research* 85: 42-59.

Bell M, Schiller A, Le Traon P, Smith N, Dombrowsky E, Wilmer-Becker K. 2015. An introduction to GODAE OceanView. *Journal of Operational Oceanography* 8: s2-s11.

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

Brydon D, Sun S, Bleck R. 1999. A new approximation of the equation of state for seawater, suitable for numerical ocean models. *Journal of Geophysical Research: Oceans* 104: 1537-1540.

Brasseur P. 2006. Ocean data assimilation using sequential methods based on the Kalman filter. In: Ocean weather Forecast. Springer; p.271-316.

Calado L, da Silveira I, Gangopadhyay A, de Castro B. 2010. Eddy-induced upwelling off Cape São Tomé (22°S, Brazil). *Continental Shelf Research* 30: 1181-1188.

Campos E, Velhote D, da Silveira I. 2000. Shelf break upwelling driven by Brazil Current Cyclonic Meanders. *Geophysical Research Letters* 27: 751-754.

Castelao R, Barth J. 2006. Upwelling around Cabo Frio, Brazil: The importance of wind stress curl. *Geophysical Research Letters* 33.

Castelao R. 2012. Sea Surface Temperature and Wind Stress Curl Variability near a Cape. *Journal of Physical Oceanography* 42: 2073-2087.

Chassignet E, Hurlburt H, Metzger E, Smedstad O, Cummings J, Halliwell G, Bleck R, Baraille R, Wallcraft A, Lozano C et al. 2009. US GODAE: Global Ocean Prediction with the HYbrid Coordinate Ocean Model (HYCOM). *Oceanography* 22: 64-75.

Chassignet E, Hurlburt H, Smedstad O, Halliwell G, Hogan P, Wallcraft A, Baraille R, Bleck R. 2007. The HYCOM (HYbrid Coordinate Ocean Model) data assimilative system. *Journal of Marine Systems* 65: 60-83.

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. *Journal of Operational Oceanography* 8: 52-62.

Cummings J, Bertino L, Brasseur P, Fukumori I, Kamachi M, Martin M, Mogensen K, Oke P, Testut C, Verron J et al. 2009. Ocean Data Assimilation Systems for GODAE. *Oceanography* 22: 96-109.

Divakaran P, Brassington G, Ryan A, Regnier C, Spindler T, Mehra A, Hernandez F, Smith G, Liu Y, Davidson F. 2015. GODAE OceanView Inter-comparison for the Australian Region. *Journal of Operational Oceanography* 8: s112-s126.

Dombrowsky E, Bertino L, Brassington G, Chassignet E, Davidson F, Hurlburt H, Kamachi M, Lee T, Martin M, Mei S et al. 2009. GODAE Systems in Operation. *Oceanography* 22: 80-95.

Evensen G. 2003. The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynamics* 53: 343-367.

Fukumori I. 2006. What is data assimilation really solving, and how is the calculation actually done?. In: Ocean weather Forecast. Springer; p.317-342.

Hernandez F, Blockley E, Brassington G, Davidson F, Divakaran P, Drévillon M, Ishizaki S, Garcia-Sotillo M, Hogan P, Lagemaa P et al. 2015. Recent progress in performance evaluations and near real-time assessment of operational ocean products. *Journal of Operational Oceanography* 8: s221-s238.

Hurlburt H, Brassington G, Drillet Y, Kamachi M, Benkiran M, Bourdallé-Badie R, Chassignet E, Jacobs G, Galloudec O, Lellouche J et al. 2009. High-Resolution Global and Basin-Scale Ocean Analyses and Forecasts. *Oceanography* 22: 110-127.

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

Le Traon P, Antoine D, Bentamy A, Bonekamp H, Breivik L, Chapron B, Corlett G, Dibarboure G, DiGiacomo P, Donlon C et al. 2015. Use of satellite observations for operational oceanography: recent achievements and future prospects. *Journal of Operational Oceanography* 8: s12-s27.

Lea D, Martin M, Oke P. 2014. Demonstrating the complementarity of observations in an operational ocean forecasting system. *Quarterly Journal of the Royal Meteorological Society* 140: 2037-2049.

Legler D, Freeland H, Lumpkin R, Ball G, McPhaden M, North S, Crowley R, Goni G, Send U, Merrifield M. 2015. The current status of the real-timein situGlobal Ocean Observing System for operational oceanography. *Journal of Operational Oceanography* 8: s189-s200.

Lima J, Martins R, Tanajura C, Paiva A, Cirano M, Campos E, Soares I, França G, Obino R, Alvarenga J. 2013. Design and implementation of the oceanographic modeling and observation network (REMO) for operational oceanography and ocean forecasting. *Revista Brasileira de Geofísica* 31: 209-228.

Martin M. 2011. Ocean forecasting systems: Product evaluation and skill. In: Schiller A, Brassington GB, editors. Operational oceanography in the 21st century. Springer Science+Business Media B. V.; p.611-632. doi:10.1007/978-94-007-0332-2_22

Mignac D, Tanajura C, Santana A, Lima L, Xie J. 2015. Argo data assimilation into HYCOM with an EnOI method in the Atlantic Ocean. *Ocean Science* 11: 195-213.

Oke P, Larnicol G, Fujii Y, Smith G, Lea D, Guinehut S, Remy E, Balmaseda M, Rykova T, Surcel-Colan D et al. 2015. Assessing the impact of observations on ocean forecasts and reanalyses: Part 1, Global studies. *Journal of Operational Oceanography* 8: s49-s62.

Oke P, Larnicol G, Jones E, Kourafalou V, Sperrevik A, Carse F, Tanajura C, Mourre B, Tonani M, Brassington G et al. 2015. Assessing the impact of observations on

ocean forecasts and reanalyses: Part 2, Regional applications. *Journal of Operational Oceanography* 8: s63-s79.

Palóczy A, da Silveira I, Castro B, Calado L. 2014. Coastal upwelling off Cape São Tomé (22°S, Brazil): The supporting role of deep ocean processes. *Continental Shelf Research* 89: 38-50.

Peterson R, Stramma L. 1991. Upper-level circulation in the South Atlantic Ocean. *Progress in Oceanography* 26: 1-73.

Rodrigues R, Lorenzzetti J. 2001. A numerical study of the effects of bottom topography and coastline geometry on the Southeast Brazilian coastal upwelling. *Continental Shelf Research* 21: 371-394.

Ryan A, Regnier C, Divakaran P, Spindler T, Mehra A, Smith G, Davidson F, Hernandez F, Maksymczuk J, Liu Y. 2015. GODAE OceanView Class 4 forecast verification framework: global ocean inter-comparison. *Journal of Operational Oceanography* 8: s98-s111.

Schiller A, Bell M, Brassington G, Brasseur P, Barciela R, De Mey P, Dombrowsky E, Gehlen M, Hernandez F, Kourafalou V et al. 2015. Synthesis of new scientific challenges for GODAE OceanView. *Journal of Operational Oceanography* 8: s259-s271.

Soutelino R, Miranda J. 2013. A semi-idealized study to evaluate the role of topography on the coastal upwelling at Cabo Frio, RJ. *Boletim do Observatório Ambiental Alberto Ribeiro Lamego* 7: 101-114.

Stramma L, England M. 1999. On the water masses and mean circulation of the South Atlantic Ocean. *Journal of Geophysical Research: Oceans* 104: 20863-20883.

Tanajura C, Costa F, 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. *Revista Brasileira de Geofísica* 31: 257-270.

Tanajura C, 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. *Atmospheric and Oceanic Science Letters* 7: 464-470.

Thacker W, Esenkov O. 2002. Assimilating XBT Data into HYCOM. *Journal of Atmospheric and Oceanic Technology* 19: 709-724.

Tonani M, Balmaseda M, Bertino L, Blockley E, Brassington G, Davidson F, Drillet Y, Hogan P, Kuragano T, Lee T et al. 2015. Status and future of global and regional ocean prediction systems. *Journal of Operational Oceanography* 8: s201-s220.

Xie J, Zhu J. 2010. Ensemble optimal interpolation schemes for assimilating Argo profiles into a hybrid coordinate ocean model. *Ocean Modelling* 33: 283-298.

Zaron E. 2011. Introduction to Ocean Data Assimilation. In: Schiller A, Brassington GB, editors. Operational oceanography in the 21st century. Springer Science+Business Media B. V.; p.321-350.

Zhu J. 2011. Overview of Regional and Coastal Systems. In: Schiller A, Brassington GB, editors. Operational oceanography in the 21st century. Springer Science+Business Media B. V.; p.413-439.