Nat. Hazards Earth Syst. Sci., 17, 17–30, 2017 www.nat-hazards-earth-syst-sci.net/17/17/2017/ doi:10.5194/nhess-17-17-2017 © Author(s) 2017. CC Attribution 3.0 License.



Data assimilation of Argo profiles in a northwestern Pacific model

Zhaoyi Wang^{1,3}, Andrea Storto², Nadia Pinardi², Guimei Liu^{1,3}, and Hui Wang^{1,3}

 ¹National Marine Environmental Forecasting Center of China (NMEFC), Beijing, 100081, China
²Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC), 40139 Bologna, Italy
³Key laboratory of Research on Marine Hazards Forecasting, National Marine Environmental Forecasting Center, Beijing, 100081, China

Correspondence to: Guimei Liu (liugm@nmefc.gov.cn)

Received: 13 February 2016 – Published in Nat. Hazards Earth Syst. Sci. Discuss.: 10 March 2016 Revised: 29 November 2016 – Accepted: 30 November 2016 – Published: 5 January 2017

Abstract. Based on a novel estimation of background-error covariances for assimilating Argo profiles, an oceanographic three-dimensional variational (3DVAR) data assimilation scheme was developed for the northwestern Pacific Ocean model (NwPM) for potential use in operational predictions and maritime safety applications. Temperature and salinity data extracted from Argo profiles from January to December 2010 were assimilated into the NwPM. The results show that the average daily temperature (salinity) root mean square error (RMSE) decreased from 0.99 °C (0.10 psu) to 0.62 °C (0.07 psu) in assimilation experiments throughout the northwestern Pacific, which represents a 37.2 % (27.6 %) reduction in the error. The temperature (salinity) RMSE decreased by ~ 0.60 °C (~ 0.05 psu) for the upper 900 m (1000 m). Sea level, temperature and salinity were in better agreement with in situ and satellite datasets after data assimilation than before. In addition, a 1-month experiment with daily analysis cycles and 5-day forecasts explored the performance of the system in an operational configuration. The results highlighted the positive impact of the 3DVAR initialization at all forecast ranges compared to the non-assimilative experiment. Therefore, the 3DVAR scheme proposed here, coupled to ROMS, shows a good predictive performance and can be used as an assimilation scheme for operational forecasting.

1 Introduction

Operational prediction systems for forecasting waves, currents and sea level variations are fundamental for maritime safety, serving a wide range of applications such as search and rescue, oil spill, tourism-oriented bulletins, climate change monitoring and many other downstream applications, eventually through downscaling the forecasts into coastal hydrodynamic models. The Chinese Global Operational Oceanography Forecasting System (CGOFS), which is run at the National Marine Environmental Forecasting Center of China (NMEFC), predicts physical properties of the global oceans, such as temperature, salinity, current, wave and sea ice. The CGOFS consists of a suite of nested model configurations. The operational northwestern Pacific Ocean model (NwPM) is a regional model at the CGOFS, which is based on the Regional Ocean Model System (ROMS), a free-surface, primitive equation ocean circulation model formulated using terrain-following coordinates. It produces daily analyses and forecasts, up to 5 days ahead, of the main ocean variables and provides boundary conditions for the East China Sea model (ECSM) and the South China Sea model (SCSM). It thus represents a fundamental ingredient in the operational marine environment and disaster forecasting chain and alert systems developed at the NMEFC.

Ocean forecasts require the specification of initial and boundary (at the surface and laterally) conditions. The accuracy of the forecasts depends on the accuracy of the initial and boundary conditions. While lateral and surface boundary conditions are usually taken from a global (or coarser resolution) ocean model and meteorological analyses and forecasts, respectively, data assimilation is a widely used and effective way of producing the best estimates of the state of the physical system to be used as initial conditions in the prognostic model.

Many data assimilation methods have been developed for combining model and observational data. These can be broadly split into three approaches: Kalman filter and derived schemes, generally known as sequential schemes (Daley, 1991); optimal interpolation; and three-dimensional and four-dimensional variational methods (3DVAR and 4DVAR; Lorenc, 1986). The variational methods are based on the minimization of a cost function that weights the differences between the analysis and the observations and the differences between the analysis and a priori knowledge of the state of the ocean, namely the background, typically from a previous forecast. The ensemble Kalman filter (EnKF) was introduced by Evensen (2003) to avoid the explicit temporal propagation of error covariances in the Kalman Filter, replacing it with covariances derived from an ensemble system. Depending on the resolution of the ocean configuration and computational resources, the EnKF may not be suitable for operational forecasting systems. As an approximation of EnKF, an ensemble optimal interpolation (EnOI) scheme was applied to ROMS to assimilate the along-track sea level anomaly (SLA) (Lyu et al., 2014). ROMS is also equipped with a 4DVAR assimilation method (Tshimanga et al., 2008; Moore et al., 2011), which might be too computationally demanding for highresolution configurations in the northwestern Pacific Ocean. 3DVAR represents a sound compromise between the sophistication and computational requirements of the assimilation scheme.

Operational oceanography has benefited from the development of data assimilation schemes such as EnOI, EnKF and 3DVAR, which have led to improved forecast skill scores. There are two main advantages of variational schemes over other methods. Firstly, the variational solution uses all observations simultaneously, whereas the EnOI technique requires the data selection into artificial subdomains. Second, balance constraints, e.g., geostrophy and hydrostatic balances, can be embedded into the definition of the balance operators that is implicit in the background-error covariance formulation (e.g., Weaver et al., 2005).

Despite such advantages, variational methods still have weaknesses. For instance, given both imperfect observations and prior (e.g., background) information as inputs to the assimilation system, the quality of the output analysis crucially depends on the appropriateness of prescribed errors, which, unlike EnKF, are usually defined as stationary errors and only include seasonal variations.

The 3DVAR data assimilation method is widely used in oceanic operational forecasting systems at both global (e.g., Storto et al., 2011; Waters et al., 2015) and regional scales (e.g., Li et al., 2008; Dobricic and Pinardi, 2008). In this study we adapted an oceanographic three-dimensional variational data assimilation scheme called OceanVar (Dobricic and Pinardi, 2008) to the ROMS model in order to assimilate temperature and salinity (T/S) measurements from Argo profiles into the CGOFS northwestern Pacific Ocean forecasting system. Although ROMS comes with a 4DVAR system, we adopted a less computationally demanding scheme, which may be inexpensively applied in the future in the production of long-term assimilation experiments (reanalyses).

The lack of tangent-linear propagation (4DVAR) in our system is compensated for by an accurate and locally varying definition of background-error covariances. The daily frequency of the assimilation cycle of 3DVAR is also expected to limit the potential advantages of a more expensive 4DVAR scheme.

To better illustrate and evaluate the performance of the assimilation scheme, the NwPM implements an eddy-resolving resolution. This system will be used in the future to increase the quality of initial conditions for daily forecasts, whose production has already started within CGOFS (v1.0).

The paper is organized as follows. Section 2 describes the components of the data assimilation scheme for assimilating Argo profiles in the northwestern Pacific. The results from data assimilation experiments are presented in Sect. 3, focusing on the performance of 3DVAR. Section 4 discusses the performance of the system in an operational configuration. Finally, Sect. 5 presents the conclusions.

2 Model configuration and data assimilation

2.1 Model configuration

The ocean model used in this work is ROMS (Schepetkin and McWilliams, 2005; Malcolm et al., 2009), a free-surface and primitive equation ocean circulation model formulated using terrain-following coordinates, which is widely used in oceanographic studies (Wang et al., 2012; Lyu et al., 2014). The model domain is the northwestern Pacific Ocean, which extends from 8° S to 52° N and from 99 to 160° E, as shown in Fig. 1a. The horizontal resolution is 1/20° in both zonal and meridional directions. There are 30 vertical sigma layers. The maximum depth is set to 7000 m for stability purposes. The bathymetry used here is derived from GEBCO (General Bathymetric Chart of the Oceans; IOC, IHO, and BODC, 2003), a global 30 arcsec gridded bathymetry, which was supplied by the Intergovernmental Oceanographic Commission and International Hydrographic Organization. In order to reduce the influence of seamounts on the model stability, the bathymetry was low-pass filtered. Three parameters can be used in the model to smooth the bathymetry. Firstly, the slope parameter $(r = \operatorname{grad}(h)h^{-1})$ maximum value for bathymetry; secondly, the number of passes of a selective filter that reduces the isolated seamounts on the deep ocean; and lastly, the number of passes of a Hanning filter at the end of the smoothing procedure to ensure no noise in the bathymetry. In our configuration, these parameter values are 0.25, 12 and 4, respectively, based on a compromise between model stability and closeness to the real topography.

The model is initialized from rest using the monthly climatological air-sea flux from the Comprehensive Ocean-Atmosphere Data Set (COADS; Clark et al., 1996) with a 10year spinup in order to obtain a fairly stable initial state. From January 1990 to December 2009, momentum and buoyancy



Figure 1. (a) Topography of the model region (depths in meters); the dotted line is the location of the 136° E section. (b) Horizontal distributions. (c) Temporal evolutions of Argo profiles used in this paper. The black line and the short dashed line represent the number of Argo floats and observations, respectively.

air-sea fluxes were derived from the 6-hourly NCEP Climate Forecast System Reanalysis (CFSR) product (Saha et al., 2010). The model configuration implements open boundary conditions. Water level, temperature, salinity and velocity at the open boundaries are derived from Simple Ocean Data Assimilation (SODA; Carton and Giese, 2008). Monthly climatological freshwater inflows from the Yangtze River, Pearl River and Mekong River with zero salinity and monthly varying temperatures are prescribed at the upstream boundaries. Considering the previous validation exercises (Wang et al., 2016), the model gives a good simulation of northwestern Pacific Ocean, especially in the subtropical Pacific region. The initial conditions for both the control (i.e., without assimilation) and the assimilation experiments are provided by the simulated ocean state valid on 1 January 2010. The control experiment for 2010 without data assimilation provides a basis for comparison.

The assimilation corrections are performed daily, using all the Argo profiles in the previous 1-day assimilation time window. The ocean model is used to bring the ocean fields 1 day forward in time. There are three main steps in the system as shown in Fig. 2: (a) preparation of temperature and salinity observations from Argo profiles; (b) integration of the NwPM model using the previous analysis increments to correct the initial conditions and calculation of misfits; (c) running the data assimilation system to produce the analysis increments for the next model integration. The misfits are computed online by the model during step (b). As the misfits come from Argo floats and are evaluated during the model integration (i.e., before being incorporated into the data assimilation system in the next analysis step), they represent a fairly independent dataset for the validation, because their subsequent measurements are sampled at different locations. Thus the temporal correlation of the observational error can be reasonably assumed to be negligible.

2.2 Observational data for assimilation

Argo is a global array of free-drifting profiling floats that measures the temperature and salinity of the upper 2000 m of the ocean. Figure 1b and c show the horizontal distributions and temporal evolutions of the Argo profiles in 2010, respectively. The profiles are quality-controlled and disseminated



Figure 2. Flow chart of the system: T_n represents the system running date and T_{n+1} represents the next system running date.

by Ifremer/CORIOLIS (Cabanes et al., 2013). The Argo network provides a fair coverage in the northwestern Pacific region and south of the Sea of Japan. Only a few Argo profiles are available in the northern region of the South China Sea. From January to December 2010, there were 9101 (1011064) T/S profiles (observations) available in the northwestern Pacific domain. In order to control the quality of observations, maximum thresholds for misfits are set up, equal to 5.0 °C and 0.5 psu for temperature and salinity observations, respectively. In the future, nonuniform thresholds may be implemented to account for the nonhomogeneous background errors in the NwPM domain.

2.3 Assimilation algorithms

The basic goal of the 3DVAR system is to provide an "optimal" estimate of the true oceanic state at analysis time through solving the assimilation problem by minimizing the following cost function (e.g., Ide et al., 1997):

$$J(x) = J_{b} + J_{o}$$
(1)
= $\frac{1}{2} (x - x^{b})^{T} B^{-1} (x - x^{b}) + \frac{1}{2} [H(x) - y^{o}]^{T} R^{-1} [H(x) - y^{o}]$

where x is the unknown ocean state, equal to the analysis x^a at the minimum of J; x^b is the background, which is an a priori estimate of the state of the ocean (in our case a oneday forecast from the previous analysis and forecast cycle); y^o is the vector of the observations; H(x) is the (nonlinear) observation function that projects the model field x to the observation space; and **B** and **R** are the covariance matrices of the background and observation errors, respectively. Equation (1) is linearized around the background state (e.g., Lorenc, 1997) into the following form:

$$J(\delta x) = \frac{1}{2} \delta x^T B^{-1} \delta x + \frac{1}{2} (\mathbf{H} \delta x - d)^T \mathbf{R}^{-1} (\mathbf{H} \delta x - d), \quad (2)$$

where $d = [y^0 - H(x_b)]$ is the misfit, **H** is the tangentlinear version of H(x), linearized around x_b , and $\delta x = x - x_b$ calculates the analysis increments. The linearization of H(x) in Eq. (2) ensures that the cost function is quadratic. The misfits are evaluated online by the model, taking the background at the same time as the observation, namely using the so-called first guess at appropriate time (FGAT) scheme. The minimization is performed over the model space defined with depth levels, thus requiring interpolation from – and to – the vertical sigma coordinates used by ROMS before – and after – the analysis step. The quasi-Newton L-BFGS algorithm (limited-memory Broyden–Fletcher–Goldfarb–Shanno; Byrd et al., 1995) is used to minimize the cost function, thus producing the optimal analysis.

The background term of the cost function is preconditioned via a control variable transformation (Lorenc, 1988); i.e., the cost function is minimized over the control variable u, with $\delta x = \mathbf{U}u$ and $\mathbf{B} = \mathbf{U}\mathbf{U}^T$. The transformation also avoids the explicit inversion of **B**. The cost function thus becomes

$$J = \frac{1}{m2}u^{T}u + \frac{1}{2}(\mathbf{H}\mathbf{U}u - d)^{T}\mathbf{R}^{-1}(\mathbf{H}\mathbf{U}u - d).$$
 (3)

The formulation of the background-error covariance **B** is thus simplified to the definition of the square-root background-error covariance operator \mathbf{U} (see next section).

2.4 Background-error covariance matrix

The specification of the background-error covariance matrix is one of the most important aspects affecting the performance of any variational data assimilation system. Therefore, an appropriate background-error covariance matrix is crucial for our 3DVAR system. The formulation of the background term of the cost function is described in Dobricic and Pinardi (2008). The background-error covariance matrix is decomposed into horizontal correlations and vertical covariances, which are assumed to be independent of each other, i.e., separable, namely $\mathbf{U} = \mathbf{U}_h \mathbf{U}_v$.

For the vertical component of the background-error covariance matrix, monthly multivariate empirical orthogonal functions (EOFs) are used as in Barker et al. (2004), namely $\mathbf{U}_v = \mathbf{E} \mathbf{\Lambda}^{1/2}$, where columns of \mathbf{E} contain multivariate eigenvectors and $\mathbf{\Lambda}$ is a diagonal matrix containing the associated eigenvalues. The EOFs are calculated from the model daily means of a full-resolution simulation covering 1995–2005 and contain covariances of sea level, temperature and salinity.

For the northwestern Pacific region, seasonal differences of the model errors are large. Therefore, we adopted monthly sets of EOFs to construct the vertical background-error co-



Figure 3. Distribution of yearly mean background-error standard deviation reconstructed from the EOFs: (a) sea level (m), (b) sea surface temperature ($^{\circ}$ C) and (c) sea surface salinity (psu). Areas shallower than 200 m are masked out; (d) eigenvalues corresponding to the first 20 EOFs.

variance matrix. Each monthly set consists of 20 EOFs with 100 *z* levels in the vertical. To compute EOFs over depth levels, we interpolate the EOFs computed on sigma levels to depth levels. This strategy resulted in cheaper calculations and less noisy background-error profiles than the eigen decomposition over model fields previously interpolated onto depth levels. The horizontal correlation operator (U_h) assumes a Gaussian shape for the correlation, with a uniform correlation radius that is given as an input parameter and equal to 4.2 km based on a set of sensitivity experiments. These experiments demonstrated that 4.2 km provides better skill scores than shorter or longer correlation radii (not shown). A first-order recursive filter is used to approximate the Gaussian shape horizontal covariances.

Figure 3 shows the map of the yearly mean backgrounderror standard deviation reconstructed from the EOFs, where Fig. 3a refers to sea level (m), Fig. 3b to temperature (°C) at surface and Fig. 3c to salinity (psu) at the surface. Background errors peak east of Japan in the Kuroshio area for all ocean variables. In terms of salinity, notable errors are also visible in the tropical part of the northwestern Pacific domain, which are linked to the large variability in the air– sea water fluxes. Figure 3d shows the yearly mean eigenvalues of all EOFs, where the first and second EOFs account for 33.7 and 15.0 % of the background-error matrix, respectively. The figure exemplifies the strong nonhomogeneity of background-error variances, leading to the need for our point-wise definition of background-error covariances.

3 Model validation

This section reports the results of the assimilation of in situ data from January to December 2010. We discuss the validation of the assimilation experiment and simulation (or control) experiment, where the simulated fields and the analysis fields are called SFs and AFs, respectively.



Figure 4. Vertical distribution of misfits for temperature (\mathbf{a} , in °C) and salinity (\mathbf{b} , in psu): the red line represents the AF bias, the black line represents the SF bias, the orange dotted line represents the misfits of AF and the gray dotted line represents the misfits of SF.

3.1 Comparison with profiles

To validate the multivariate assimilation scheme, Argo profiles are used for validation. Argo misfits are independent observation-model departures because they are evaluated before being incorporated into the assimilation system.

3.1.1 Vertical distribution of *T*/*S* errors

To validate the performance of the assimilation system, the vertical distribution of T/S errors are compared. Figure 4 shows the vertical distribution of T/S misfits computed with Argo profiles and model simulations from the SFs and AFs. The results show that the AF run significantly reduces the biases in SF, from more than $0.7 \,^{\circ}$ C (0.02 psu) to less than $0.2 \,^{\circ}$ C (0.01 psu) for temperature (salinity) at depths below 300 m, especially around the ocean mixed layer depth. T/S biases decrease with the depth: in the water deeper than 1000 m, biases become very small, as expected from the decreased variability of T/S.

In order to further validate the vertical distribution of T/S misfits, we compared the vertical distribution of T/S root mean square error (RMSE) and mean absolute error (AE). Figure 5 shows the RMSE and AE of temperature and salinity misfits for AFs and SFs. They have a similar vertical structure: the AF run has a significantly lower RMSE and AE than the SF run. In terms of temperature, as shown in Fig. 5a, the maximum RMSE of temperature errors is around a depth

of 150 m, which corresponds approximately to the depth of the mixed layer. The improvement rate (RMSE decrease) after assimilation can reach up to 50% at depths of around 10 m and higher than 30% at depths shallower than 600 m. The RMSE decreased by ~ 0.60 °C for the upper 900 m, but a slight deviation of ~ 0.015 appears from 900 to 1500 m. For salinity, as shown in Fig. 5b, the RMSE of misfits decrease considerably after assimilation, especially at depths shallower than 600 m. The RMSE decreased by ~ 0.040 psu for the upper 1000 m, but a slight deviation of ~ 0.002 appears from 1000 to 1500 m.

3.1.2 Temporal evolution of the RMSE of temperature and salinity

To investigate the performance of the assimilation system over time, we compared the temporal evolution of T/S RMSE. Figure 6 shows the time evolution of T/S RMSE calculated from AF and SF. Figure 6a shows that the reduction in temperature RMSE grows with time and become increasingly smaller for the AF run. On average, the RMSE decreases from 0.988 °C in the SF run to 0.620 °C in the AF run, i.e., a 37 % reduction occurs. Figure 6b shows that the reduction in salinity RMSE is quite stable. On average, the RMSE decreases from 0.098 in the SF run to 0.071 psu in the AF run, i.e., a 28 % reduction.



Figure 5. The vertical RMSE and AE for temperature (\mathbf{a} , in °C) and salinity (\mathbf{b} , in psu): the black solid line represents the RMSE of AF, the gray solid line represents the RMSE of SF, the black dotted line represents the AE of AF and the gray dotted line represents the AE of SF.

3.2 Comparison with satellite data

3.2.1 Comparison with temperature

To evaluate the performance of the assimilation system on temperature, independent satellite sea surface temperature data (OISST, $1/4^{\circ} \times 1/4^{\circ}$; Reynolds et al., 2007) are used for validation. Figure 7 shows the monthly mean average SST differences from OISST calculated from SF (Fig. 7a) and AF (Fig. 7b), where 1, 2, 3 and 4 represent January, April, July and October, respectively. The results show that the AF run reduces the SST bias from 1.21 to $1.07 \,^{\circ}$ C in most of the northwestern Pacific, especially the tropical zone and south and east of Japan. Also in the Sea of Japan, there is a visible improvement. However, there are still regions with biases larger than $1.5 \,^{\circ}$ C, especially around the Kuroshio extension, likely due to the inaccuracies in the air–sea fluxes in this region.

3.2.2 Comparison with sea level anomaly

To evaluate the performance of the assimilation system on sea level, independent Archiving, Validation and Interpretation of Satellite Oceanographic data (AVISO, $1/4^{\circ} \times 1/4^{\circ}$; Le Traon et al., 2003) were used for validation. Figure 8 shows monthly mean average SLA differences from AVISO calculated from SF (Fig. 8a) and AF (Fig. 8b), the numbers have the same meanings as in Fig. 7. The results show that the in situ observation assimilation (AF run) reduces the SLA bias from 11.8 to 11.1 cm in the northwestern Pacific. As Fig. 8 shows, the SLA bias is reduced after data assimilation in

Extension, which are difficult to correct without directly assimilating altimetry data. Figure 9 shows the time evolution of SLA RMSE calculated from AF and SF. On average, the RMSE decreases from

lated from AF and SF. On average, the RMSE decreases from 13.1 cm in the SF run to 12.8 cm in the AF run. Because the greatest errors occur in the region of the Kuroshio Extension as shown in Fig. 8, the RMSE decreases from 10.6 cm in the SF run to 9.6 cm in the AF run, i.e., a 9.4 % reduction, when this region (30–40° N, 135–160° E) is masked out.

most of the northwestern Pacific, especially the tropical area. However, there are large biases in the region of the Kuroshio

3.3 Comparison with reprocessed datasets

Consistency checks were carried out by comparing the SF and AF monthly mean temperature and salinity with the EN4.0.2 ($1^{\circ} \times 1^{\circ}$; Good et al., 2013) datasets, which are objective analyses that incorporate all in situ hydrographic profiles. In this section, the salinity error calculated by the model output for AF and SF and the in situ reprocessed dataset for EN4.0.2 will first be compared at the surface. Secondly, the *T/S* error will be validated at the 137° E section, which has a long time series of in situ observations.

3.3.1 Comparison with salinity at surface

To validate the sea surface salinity (SSS), Fig. 10 shows the monthly mean average SSS bias computed from SF (Fig. 10a) and AF (Fig. 10b) with respect to EN4.0.2, where the numbers have the same meanings as in Fig. 7. As shown in Fig. 10, there are many regions, especially near the Equa-



Figure 6. Temporal evolution of temperature (\mathbf{a} , in °C) and salinity (\mathbf{b} , in psu). The RMSE is computed against the Argo profiles for the AF (black solid line) and SF (gray dotted line).



Figure 7. Monthly mean temperature bias at surface (in $^{\circ}$ C) for (1) January, (2) April, (3) July and (4) October: (a) SST error calculated by SF and OISST; (b) SST error calculated by AF and OISST. Depths below 200m are masked out.

tor, with a bias larger than 0.3 for both the SF run and AF run. The results show that the reduction in SSS bias is, on average, from 0.363 to 0.308 psu (i.e., a reduction of 15.2%) in October (Fig. 10a4 and b4). Due to the lack of observations, no significant improvements occur in the South China Sea.

3.3.2 Comparison at 137° E section

To validate the vertical properties of the system, temperature and salinity sections of 137° E are presented. The section is chosen because of the long time series of in situ observations. Figures 11 and 12 show the temperature and salinity distributions of AF (Figs. 11a, 12a) and EN4.0.2 (Figs. 11b, 12b), wherein the numbers have the same meanings as in Fig. 7 for the upper 1000 m. For temperature (Fig. 11), the vertical distribution of the AF run has a structure similar to the reprocessed dataset. The April and July penetrations of the 18° isotherm in AF have the same depth as EN4.0.2, while SF does not reproduce these features. Also for salinity (Fig. 12), during the spring and summer seasons the AF run has a better performance in reproducing the vertical shapes of salinity, especially in the northernmost area of the domain.



Figure 8. Monthly mean sea level anomaly (SLA) bias (in meter) for (1) January, (2) April, (3) July and (4) October: (a) SLA bias calculated by SF and AVISO; (b) SLA bias calculated by AF and AVISO. Depths below 200 m are masked out.



Figure 9. Temporal evolution of SLA RMSE (in meters). RMSE computed with the AVISO dataset for the AF (solid line) and SF (dotted line). The black line represents the RMSE in the entire domain and the gray line represents the RMSE in the region without the Kuroshio Extension (NKE).

4 Discussion

4.1 Fitting the analyses with observations

Following the assimilation scheme in Sect. 2.1, the Argo misfits are independent observations that can be used for validation. In order to evaluate the greatest potential for the system, another analysis and forecast scheme was set up, illustrated by the dotted line in Fig. 2. Compared to the original scheme (AF), the analysis increments are used to correct the initial conditions of the model at the beginning of the assimilation time window. Therefore, two sets of misfits are available, before (bDA) or after (aDA) the data assimilation correction, respectively, i.e., before or after being incorporated into the system. To investigate the performance of this scheme, we assimilated Argo profiles in December 2010. Meanwhile, a control run (CTRL) without data assimilation was performed for validation purposes.

Figure 13a, b show the RMSE vertical profiles of temperature and salinity for the AF, bDA, aDA and CTRL runs. Here the profiles are not independent because they have been already assimilated into the model. The aDA therefore provides a consistency check of the assimilation system rather than independent validation metrics. For temperature, the data assimilation led to a large improvement within the top 800 m. On average, the RMSE for temperature was 0.58 °C for AF, 0.60 °C for bDA, 0.40 °C for aDA and 0.68 °C for CTRL. For salinity, the RMSE of aDA decreased, particularly in the top 600 m. The aDA run performed better than the other runs, but there were slight deviations for water deeper than 1000 m. On average, the RMSE for salinity was 0.069 psu for AF, 0.070 psu for bDA, 0.056 psu for aDA and 0.081 psu for CTRL. By comparing the two schemes, skill score improvements were not fully satisfactory for the regions shallower than 1000 m. In the future, we plan to retune the vertical EOFs in the shallow areas.

Figure 13c, d show the time evolution of T/S RMSE for the AF, bDA, aDA and CTRL runs. The difference between the simulation with and without assimilation grows with time. In terms of temperature (Fig. 13c), on 31 December 2010, the RMSE decreased from 0.97 °C in the CTRL run to 0.73 °C in the AF run, 0.75 °C in the bDA run and 0.46 °C in the aDA run. Regarding salinity (Fig. 13d), the RMSE de-



Figure 10. Monthly mean salinity bias at surface (in 1): (1) January, (2) April, (3) July and (4) October: (a) SSS error calculated by SF and EN4.0.2, (b) SSS error calculated by AF and EN4.0.4. Depths below 200 m are masked out.



Figure 11. Distribution of temperature (in $^{\circ}$ C) at transect of 137 $^{\circ}$ E: (a) AF, (b) SF and (c) EN4.0.2 for (1) January, (2) April, (3) July and (4) October.



Figure 12. Distribution of salinity (in psu) at transect of 137° E: (a) AF, (b) SF and (c) EN4.0.2 for (1) January, (2) April, (3) July and (4) October.

creased from 0.095 psu in the CTRL run to 0.081 psu in the AF run, 0.077 psu in the bDA run and 0.063 psu in the aDA run. Overall, the improvement in the data assimilation system was $\sim 0.46 \,^{\circ}$ C (0.06 psu) for temperature (salinity).

4.2 Operational forecast experiment

In order to discuss the performance of data assimilation in operational configurations, an experiment was set up for 2-31 December 2010. The analysis frequency was daily, with daily forecasts of up to 5 days. Figure 14 shows the temporal evolution of T/S RMSE in comparison with a control run without data assimilation. The results show that the reduction in T/S RMSE with respect to the control experiment was increasingly less significant with the increase in forecast time, but it was smaller than the CTRL run RMSE in most cases, especially in the second half of the experiment. In terms of temperature, on average, the RMSE was 0.761, 0.773, 0.781, 0.782 and 0.801 °C for the first, second, third, fourth and fifth day forecast, respectively, i.e., always smaller than the RMSE of the CTRL run, which was equal to 0.873 °C. Regarding salinity, on average, the RMSE was 0.0846, 0.0858, 0.0863, 0.0867 and 0.0875 psu for the five forecast days, respectively, i.e., performing better than the CTRL run (with a RMSE of 0.0926 psu). We can thus conclude that the data assimilation system significantly improves the prediction results for up to 5 days. Overall, the 3DVAR assimilation system can be used as an assimilation scheme for operational forecasting.

5 Conclusions

We have implemented a 3DVAR scheme in ROMS that assimilates temperature and salinity observations from Argo profiles. This work represents a first step towards a fully operational analysis and forecast system developed at NMEFC for use in maritime safety applications. The data assimilation system was implemented in an eddy-resolving configuration of the northwestern Pacific from January to December 2010. A specific feature of our 3DVAR system is the separation of the background-error covariance matrix into vertical and horizontal modes in order to reduce the size of the data assimilation problem. Horizontal correlations are modeled as Gaussian functions through a first-order recursive filter, while vertical covariances are estimated from a long-term model simulation and formulated as monthly sets of EOFs.

After assimilating the Argo profiles, the average daily temperature (salinity) RMSE decreased from $0.988 \,^{\circ}C$ (0.098 psu) to $0.620 \,^{\circ}C$ (0.071 psu) in the assimilation experiment throughout the northwestern Pacific Ocean. The temperature RMSE decreased by $\sim 0.60 \,^{\circ}C$ for the upper 900 m. In addition, the salinity RMSE decreased by ~ 0.040 psu for the upper 1000 m.



Figure 13. Vertical (a, b) and temporal (c, d) evolution of T/S RMSE: green solid line represents the RMSE of the AF run, the blue solid line represents the RMSE of bDA run, the red dotted line represents the RMSE of the aDA run and the black dotted line represents the RMSE of CTRL run.



Figure 14. Temporal evolution of temperature (a) and salinity (b) RMSE with 5-day forecasting experiment: dark blue bar represents the RMSE of first-day forecast, the light blue bar represents the RMSE of the second-day forecast, the emerald bar represents the RMSE of third-day forecast, the orange bar represents the RMSE of fourth-day forecast, the red bar represents the RMSE of fifth-day forecast and the solid line represents the RMSE of CTRL run.

The OISST satellite-derived datasets (SST) and AVISO (sea level anomaly) and temperature and salinity objective analyses from EN4.0.2 were collected for validation. A comparison of these datasets showed that the data assimilation provides a beneficial effect for the sea level, temperature and salinity at the surface in the model region. By comparing the assimilation experiment with the reprocessed dataset, the data assimilation provided a good reproduction of the vertical structure across the data-rich transect at 137° E. This led to a correct representation of the penetration of the 18° isotherm in the northern part of the domain during the spring and summer seasons.

The potential of the data assimilation system was also discussed by assessing the assimilation experiments with validating observations before and after their ingestion in the system. The results show that the minimum RMSE the assimilation system is able to reach is ~ 0.46 °C for temperature and ~ 0.06 for salinity. However, the improvements in the data assimilation system are not completely satisfactory below 1000 m. In order to better simulate the T/S in the depths below 1000 m, a further retuning of the vertical EOFs may be necessary to avoid spurious analysis increments at depth.

The assimilation system was also tested in an operational framework for a 1-month period, where daily analysis cycles and 5-day forecasts were produced. The 3DVAR initialization improved the short-term predictability in the northwestern Pacific Ocean. It led to skill scores that beat those of a non-assimilative experiment for all 5 forecast days.

Overall, the 3DVAR assimilation system performed well in the assimilation experiment. All these results encourage the implementation of the system in an operational environment for maritime safety applications. In further experiments, we plan to extend the assimilated observing networks to sea level anomaly and sea surface temperature data from satellites. This may alleviate the biases occurring in the mesoscale active region of the Kuroshio extension due to inaccuracies in the air–sea exchange fluxes and limitations in capturing the eddy-dominated ocean dynamics in that region.

6 Data availability

The Argo profiles, which are quality controlled and disseminated by Ifremer/CORIOLIS, are described in Cabanes et al. (2013) and cover from 1995 to now. The atmosphere forcing data are from CFSR/NECP (Saha et al., 2010) and cover 1979 to now. The SODA dataset is described in Carton and Giese (2008) and is available from 1970 to 2010. The bathymetry data are obtained from the British Oceanographic Data Center (BODC). This version of the GEBCO_08 grid was released in November 2010 with a version code of 20100927. Information on the dataset is given by Hall (2002). The satellite daily OISST is described in Reynolds et al. (2007). The MGDSST is provided by JMA (Japan Meteorological Agency) and is described in Kurihara et al. (2006). The title of the reprocessed dataset is EN4.0.2, which is described in Good et al. (2013) and covers 1900 to 2015. The SLA dataset, which is provided by AVISO, is described in Le Traon et al. (2003).

Acknowledgements. This work was supported by the National Natural Science Foundation of China under contract nos. 41222038 and 41206023, the National Basic Research Program of China (973 program) under contract no. 2011CB403606, and the Strategic Priority Research Program of the Chinese Academy of Sciences through grant no. XDA1102010403.

Edited by: A. Olita Reviewed by: B. Powell and three anonymous referees

References

- Barker, D., Huang, W., Guo, Y., Bourgeois, A., and Xiao, Q.: A Three-Dimensional Variational Data Assimilation System for MM5: Implementation and Initial Results, Mon. Weather Rev., 132, 897–914, 2004.
- Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C.: A limited memory algorithm for bound constrained optimization, SIAM J. Sci. Comput., 16, 1190–1208, 1995.
- Cabanes, C., Grouazel, A., von Schuckmann, K., Hamon, M., Turpin, V., Coatanoan, C., Paris, F., Guinehut, S., Boone, C., Ferry, N., de Boyer Montégut, C., Carval, T., Reverdin, G., Pouliquen, S., and Le Traon, P.-Y.: The CORA dataset: validation and diagnostics of in-situ ocean temperature and salinity measurements, Ocean Sci., 9, 1–18, doi:10.5194/os-9-1-2013, 2013 (data available at: ftp://ftp.ifremer.fr/ifremer/argo/geo/pacific_ ocean/).
- Carton, J. A. and Giese, B.: A Reanalysis of Ocean Climate Using Simple Ocean Data Assimilation (SODA), Mon. Weather Rev., 136, 2999–3017, doi:10.1175/2007MWR1978.1, 2008 (data available at: http://dsrs.atmos.umd.edu).
- Clark, M. P., Serreze, M. C., and Barry, R. G.: Characteristics of Arctic Ocean Climate based on COADS data, 1980–1993, Geophys. Res. Lett., 23, 1953–1956, 1996.
- Daley, R.: Atmosferic data Analysis, Cambridge University Press, Cambridge, UK, 471 pp., 1991.
- Dobricic, S. and Pinardi, N.: An oceanographic three-dimensional variational data assimilation scheme, Ocean Model., 22, 89–105, 2008.
- Evensen G.: The Ensemble Kalman Filter: theoretical formulation and practical implementation, Ocean Dynam., 53, 343–367, 2003
- Good, S. A., Martin, M. J., and Rayner, N. A.: EN4: quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates, J. Geophys. Res.-Oceans, 118, 6704–6716, doi:10.1002/2013JC009067, 2013 (data available at: http://hadobs.metoffice.com/en4/download-en4-0-2.html).
- Hall, J. K.: Bathymetric compilations of the seas around Israel I: The Caspian and Black Seas, Geological Survey of Israel, Current Research, Vol. 13, available at: http://www.gebco.net/data_ and_products/gridded_bathymetry_data/version_20100927/, 2002.

- Ide, K., Courtier, P., Ghil, M., and Lorenc, A. C.: Unified notation for data assimilation: Operational, sequential and variational, J. Met. Soc. Japan, 75, 181–189, 1997.
- IOC, IHO, and BODC: Centenary Edition of the GEBCO Digital Atlas, published on CD-ROM on behalf of the Intergovernmental Oceanographic Commission and the International Hydrographic Organization as part of the General Bathymetric Chart of the Oceans, British Oceanographic Data Centre, Liverpool, UK, 2003.
- Kurihara, Y., Sakurai, T., and Kuragano, T.: Global daily sea surface temperature analysis using data from satellite microwave radiometer, satellite infrared radiometer and insitu observations, Sokko-Jihou, 73, S1–S18, 2006 (in Japanese) (data available at: http://ds.data.jma.go.jp/gmd/goos/data/rrtdb/file_list.php#0).
- Li, Z., Chao, Y., Mcwilliams, J. C., and Ide, K.: A three-dimensional variational data assimilation scheme for the regional ocean modeling system, J. Atmos. Ocean. Tech., 25, 2074–2090, 2008.
- Le Traon, P. Y., Faugère, Y., Hernandez, F., Dorandeu, J., Mertz, F., and Ablain, M.: Can We Merge GEOSAT Follow-On with TOPEX/Poseidon and ERS-2 for an Improved Description of the Ocean Circulation?, J. Atmos. Ocean. Tech., 20, 889–895, doi:10.1175/1520-0426(2003)020<0889:CWMGFW>2.0.CO;2, 2003 (data available at: http://www.aviso.altimetry.fr/en/data.html).
- Lorenc, A. C.: Analysis methods for numerical weather prediction, Q. J. Roy. Meteor. Soc., 112, 1177–1194, 1986.
- Lorenc, A. C.: Optimal nonlinear objective analysis, Q. J. Roy. Meteor. Soc., 114, 205–240, doi:10.1002/qj.49711447911, 1988.
- Lorenc, A. C.: Development of an operational variational assimilation scheme, J. Meteorol. Soc. Jpn., 75, 339–346, 1997.
- Lyu, G., Wang, H., Zhu, J., Wang, D., Xie, J., and Liu, G.: Assimilating the along-track sea level anomaly into the regional ocean modeling system using the ensemble optimal Interpolation, Acta Oceanol. Sin., 33, 72–82, 2014.
- Moore, A. M., Arango, H., Broquet, G., Powell, B., Weaver, A., and Garay, J.: The Regional Ocean Modeling System (ROMS) 4dimensional variational data assimilation systems Part I – System overview and formulation, Prog. Oceanogr., 91. 34–49, 2011.
- Malcolm, J., Roberts, M. J., Clayton, A., Demory, M., Donners, J., Vidale, P. L., Norton, W., Shaffrey, L., Stevens, D. P., Stevens, I., Wood, R. A., and Slingo, J.: Impact of resolution on the Tropical Circulation in a matrix of coupled models, J. Climate, 22, 2541– 2556, 2009.
- Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K. S., and Schlax, M. G.: Daily High-Resolution-Blended Analyses for Sea Surface Temperature, J. Climate, 20, 5473–5496, doi:10.1175/2007JCLI1824.1, 2007.

- Saha, S., Moorthi, S., Pan, H., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, J., Hou, Y., Chuang, H., Juang, H., Sela, J., Iredell, M., Treadon, R., Kleist, D., Delst, P., Keyser, D., Derber, J., Ek, M., Meng, J., Wei, H., Yang, R., Lord, S., Doll, H., Kumar, A., Wang, W., Long, C., Chellian, M., Xue, Y., Huang, B., Schemm, J., Eisuzaki, W., Lin, R., Xie, P., Han, Y., Cucurull, L., Reynolds, R., Rutledge, G., and Goldberg, M.: The NCEP Climate Forecast System Reanalysis, B. Am. Meteorol. Soc., 91, 1015–1057, doi:10.1175/2010BAMS3001.1, 2010 (data available at: http://rda.ucar.edu/datasets/ds094.1).
- Schepetkin, A. F. and McWilliams J. C.: The regional ocean modeling system (ROMS): A split-explicit, free-surface, topographyfollowing-coordinate oceanic model, Ocean Modell., 9, 347– 404, 2005.
- Storto, A., Dobricic, S., Masina, S., and Di Pietro, P.: Assimilating Along-Track Altimetric Observations through Local Hydrostatic Adjustment in a Global Ocean Variational Assimilation System, Mon. Weather Rev., 139, 738–754, doi:10.1175/2010MWR3350.1, 2011.
- Tshimanga, J., Gratton, S., Weaver, A. T., and Sartenaer, A.: Limited-memory preconditioners with application to incremental variational data assimilation, Q. J. Roy. Meteor. Soc., 134, 751–769, 2008.
- Wang, H., Wang, Z. Y., Zhu, X. M., Wang, D. K., and Liu, G. M.: Numerical study and prediction of nuclear contaminant transport from Fukushima Daiichi nuclear power plant in the North Pacific Ocean, Chin. Sci. Bull., 57, 3518–3524, 2012.
- Wang, Z., Liu, G., Wang, H., and Wang, D.: Numerical study of seasonal and interannual variation of circulation and water transports in the Luzon Strait[J], Acta Oceanol. Sin., 38, 1–13, doi:10.3969/j.issn, 2016 (in Chinese).
- Waters, J., Lea, D. J., Martin, M. J., Mirouze, I., Weaver, A., and While, J.: Implementing a variational data assimilation system in an operational 1/4 degree global ocean model, Q. J. Roy. Meteor. Soc., 141, 333–349, doi:10.1002/qj.2388, 2015.
- Weaver, A. T., Deltel, C., Machu, E., Ricci, S., and Daget, N.: A multivariate balance operator for variational ocean data assimilation, Q. J. Roy. Meteor. Soc., 131, 3605–3625, doi:10.1256/qj.05.119, 2005.