Impact of Dynamical Representational Errors on an Indian Ocean Ensemble Data Assimilation System

Item Type	Article
Authors	Sanikommu, Siva Reddy;Benerjee, Deep Sankar;Baduru, Balaji;Paul, Biswamoy;Paul, Arya;Chakraborty, Kunal;Hoteit, Ibrahim
Citation	Sanikommu, S., Benerjee, D. S., Baduru, B., Paul, B., Paul, A., Chakraborty, K., & Hoteit, I. (2019). Impact of Dynamical Representational Errors on an Indian Ocean Ensemble Data Assimilation System. Quarterly Journal of the Royal Meteorological Society. doi:10.1002/qj.3649
Eprint version	Post-print
DOI	10.1002/qj.3649
Publisher	Wiley
Journal	Quarterly Journal of the Royal Meteorological Society
Rights	Archived with thanks to Quarterly Journal of the Royal Meteorological Society
Download date	2024-03-13 09:55:59
Link to Item	http://hdl.handle.net/10754/656680

Impact of Dynamical Representational Errors on an Indian Ocean Ensemble Data Assimilation System

Sivareddy Sanikommu[1, 2] *, Deep Sankar Benerjee[1], Balaji Baduru[1], Biswamoy Paul[1],

Arya Paul[1], Kunal Chakraborty[1], Ibrahim Hoteit[2]

1. ESSO-Indian National Centre for Ocean Information Services (ESSO-INCOIS),

   Ministry of Earth Sciences, Govt. of India, Hyderabad, India

2. Division of Physical Sciences and Engineering, King Abdullah University of

   Science and Technology, Thuwal, Kingdom of Saudi Arabia

Corresponding author: sivamtech07@gmail.com

Abstract

This study investigates the impact of dynamical representational error (RE) on the analysis of an ocean ensemble Kalman filter-based data assimilation system, LETKF-ROMS (Local Ensemble Transform Kalman Filter - Regional Ocean Modeling system) configured for the Indian Ocean and assimilating *in-situ* temperature and salinity observations from Argo. Three different approaches to account for the RE are studied and inter-compared: (i) static RE (varies in horizontal and vertical direction), (ii) dynamic RE (varies in space and time) estimated from concurrent observations, and (iii) dynamic RE estimated using concurrent high resolution model outputs. RE estimated from the model outputs exhibits rich spatial and temporal variability with an estimated temporal mean RE for temperature below 0.5 °C and 0.2 °C in the surface and deep layers, respectively, and reaching up to 1°C in the thermocline layers. The region encompassing the Great Whirl displays a large seasonal variability reaching up to 0.8°C, and the South Equatorial Current (SEC)a large inter-annual variability reaching up to 0.4°C.

Neglecting such spatio-temporal variations of RE and assimilating with a static RE limited the benefits of assimilation by entertaining over-fitting issues that caused degradations in the Bay of Bengal, the western parts of the Arabian Sea, and the equatorial Indian ocean. Assimilating with the observations-based dynamic RE improved the results in these regions, but the best performances were obtained with the configuration using the model-based dynamic RE, which yielded further improvements (e.g. reduction of sea surface height root-mean-square-errors reaches 30% with respect to the observations-based dynamic RE). The latter also better handled

the rich spatial variability regions and areas not well sampled by the observations. Improved estimates of the spatial and temporal variations of RE helped to better exploit the assimilated observations and provided enhanced analyses less prone to assimilation shocks.

1. Introduction

The observation error (OE) is a critical component of any data assimilation system because the correction to the background ocean state by the observation is determined by the error covariances of the background model state and observations. In the context of data assimilation, the OE is defined as the difference between the sampled observation and the true state of the ocean projected onto the observation space (*Janjic et al., 2017*). It consists of two components, namely, (1) the Instrument Error (IE), and (2) the Representation Error (RE), which can be expressed as

$$OE = Y - H(X^t) = (Y - Y^t) + (Y^t - H(X^t)) = IE + RE,$$

where Y is the sampled observation, Y^t is the truth in the observation space, X^t is the true state in the model space, and H is the observation operator that projects the model state onto the observation space.

IE is associated with the quality of the sensor being deployed to sample the observations. This error is typically not large, particularly for *in-situ* state-of-the-art ocean observation systems. RE is known to significantly contribute to the observational error (*Wunch 1996, Hoteit et al., 2010*) and comprises (i) errors due to unresolved scales and processes, and (ii) errors in the observation operator error (*Janjic et al., 2017*). The first errors type arises from

the mismatch between the scales described in the observations and the model state. For instance, an observation may represent the value of a geophysical variable at a given point in space and time, whereas the model output is likely to represent a spatial and temporal average, depending, among others, on the discretization and resolution of the model. Observation operator errors arise from the under-representation of the observation operator H, particularly when the observed variables are not prognostic (state variables). In ocean data assimilation, the representation errors due to the observation operator errors are typically negligible in comparison to those associated with the unresolved scales and processes (*Janjic et al., 2017*). Henceforth we focus on the errors due to unresolved scales and processes and refer to it as RE.

The RE is expected to vary in space and time (*Oke and Sakov, 2008; Janjic et al., 2017*) and with the resolution of the model (e.g. *Karspeck* 2016); the coarser the resolution the larger the RE is expected to be. *Oke and Sakov (2008)* argued that smaller (larger) than realistic observational error in ocean data assimilation systems would enforce large (weak) correction with the observation. While weak corrections limit the benefits of assimilated observations, large corrections may cause assimilation shocks, which can be generated by spurious corrections. These assimilation shocks propagate as planetary waves and degrade the analysis elsewhere, unless the observation coverage is sufficient enough to suppress them (*Sivareddy et al., 2017*).The use of spatially and temporally varying (here onwards dynamic) observational error may help mitigating these adverse effects by imposing proper weighting of the innovations, and is therefore crucial to obtain reliable analysis state.

Despite a wide range of methods to estimate a dynamic RE; using for e.g. analysis increments by Desroziers et al., (2001), (2006); and the maximum likelihood method by Dee Silva (1999), methods implemented in Ocean data assimilation systems are still largely based

on the concurrent observations (e.g. *Behringer et al., 1998; Richman et al., 2005; Oke et al., 2008; Balmaseda et al., 2013; Sivareddy, 2015*). For example, *Behringer et al. (1998)*and *Balmaseda et al. (2013)* estimated a dynamic RE, that varies in space and time, for temperature based on vertical temperature gradients in the concurrent temperature profile. *Sivareddy (2015)* applied the same method to estimate a dynamic RE for salinity. *Oke and Sakov. (2008)* estimated a dynamic RE for sea surface height as a fraction of the variance from binned satellite altimeter measurements. These observations-based estimates are, however, likely not to be very robust in under-sampled areas, as demonstrated in the results section of this study. Employing high-resolution model simulations in the dynamic RE estimation procedure, as has been implemented in atmospheric assimilation studies by for e.g. *Etherton and Bishop (2004)* using differences in the state variables of a dynamical models running at two different resolutions, is a natural alternative to under-sampling issues. Such an approach is still not explored by the ocean data assimilation community, probably due to the computational overhead, which is no more a serious concern at present. Most present ocean data assimilation systems neglect the spatial and temporal variations of RE and use static (in space or time) observational errors (*Zhang et al., 2007; Hoteit et al., 2010; Penny et al., 2011; Hoteit et al., 2013*) and did not conduct a comprehensive assessment of the impact of RE on the ocean analysis.

This study aims to provide a comprehensive assessment of the impact of RE on ocean analyses using an ensemble Kalman filter-based assimilation system composed of a Local Ensemble Transform Kalman Filter (LETKF) and a Regional Ocean Model (ROMS), hereafter LETKF-ROMS. This is achieved by inter-comparing the quality of ocean analyses from three different assimilation experiments using (1) static REs, (2) observations-based dynamic REs

following *Behringer et al. (1998)*, and (3) model-based dynamic REs following *Etherton and Bishop (2004)*. We analyze the estimated spatial and temporal structures of the REs and show that model-based dynamic RE enhances the performance of the assimilation system and mitigates issues related to observations under-sampling.

The remainder of the paper is organized as follows. Section 2 briefly presents the LETKF-ROMS assimilation system and assimilated observations. Section 3 describes the methodologies to estimate dynamic REs. Section 4 discusses the spatial and temporal structures of the estimated dynamic REs. The impact of the different REs on the ocean analyses is analyzed in Section 5. Section 6 summarizes the results and discusses the computational feasibility of the model-based dynamic REs.

2. The assimilation system.

We used an in-house regional ocean data assimilation LETKF-ROMS system (*Balaji et al., 2018*). The ocean model, ROMS, is a free surface, terrain following general circulation model. It solves the primitive equations on an orthogonal curvilinear coordinate grid*(Song et al., 1994; Haidvogel et al., 2000; Shchepetkin., 2005)*. It has been deliberately implemented at a relatively low resolution, ¼° x ¼° horizontal grid interval, for assimilation to investigate the relevance of the estimated REs. The model domain covers the Indian Ocean (IO) region from 30°E to 120°E in the east-west direction and from 30°S to 30°N in the north-south direction. The ocean depth is discretized into 40 vertical sigma levels. The vertical stretching parameters are chosen in such a way that the vertical resolution is highest in the upper ocean. The lateral boundaries in the east and south are treated as open, where the tracer and momentum fields are nudged to 10-day mean fields derived from the ocean analysis of the Global Ocean Data Assimilation System, available from the Indian National Centre for Ocean Information

Services (here after INCOIS-GODAS; *Siva Reddy, 2015)*. The western and northern boundaries are solid walls with no-slip conditions. The model uses the K-profile parameterization scheme *(Large et al., 1994)* to parameterize the vertical mixing. Smagorinsky type bi-harmonic diffusion (50 m^2s^{-1}) and viscosity (300 m^2s^{-1}) schemes are used for horizontal mixing and a bulk parameterization scheme *(Fairall et al., 1996; Griffies and Hallberg, 2000)* is selected for the computation of air–sea fluxes of heat *(Jithin et al., 2017)*. Salinity in the model's top layer is relaxed to monthly climatology derived from the World Ocean Atlas (WOA) climatology *(Antonov et al., 2009; Locarnini et al., 2009)*.

The assimilation code of this regional ocean data assimilation system, LETKF, was originally developed for the modular ocean model (MOM) by *Penny (2011)* and has been adopted to suit ROMS *(Balaji et al., 2018)*.   The LETKF is implemented with 32-member ensemble, with capabilities to assimilate sea level anomaly, temperature and salinity profiles. The choice of 32 ensemble size is to balance computational cost and results. As demonstrated by earlier studies, (e.g. Xu et al., 2013; Xu and Oey, 2014), a 32 ensemble size is a reasonable choice for LETKF based ocean assimilation system to capture oceanic features. In the present study, we assimilate *in-situ* Argo temperature and salinity profiles as made available by *Ingleby and Huddleston (2009).* Argo is a global network of freely moving profiling floats that sample the ocean column from surface to 2000m depth *(Roemmich et al., 2009)*. It shows good spatial coverage during 2008-2010 in the Indian Ocean (see Figure 1) except in the regions dominated by horizontal divergence and upwelling (e.g. South-western parts of the Indian Ocean around Madagaskar, South-eastern parts of the Arabian Sea covering Lakshadweep). The observational errors for temperature (salinity) vary between 0.5-3°C (0.3–3 psu) in the present study, in line with the suggested ranges of observational errors by earlier assimilation studies with coarse

resolution ocean models (e.g., *Richman et al.,* 2005; *Forget and Wunsch,* 2007; *Oke and Sakov,* 2008; *Karspeck,* 2016). We further assume that the observational errors are uncorrelated, a reasonable assumption given that we assimilate relatively distanced *in-situ* observations (e.g., *Wunsch and Heimbach,* 2007; *Zhang et al.,* 2007; *Giese and Ray,* 2011; *Balmaseda et al.,* 2013; *Karspeck et al.,* 2013). The observational error covariance matrix is therefore diagonal with the variances of the OEs as diagonal elements. A localization radius of 360 km is implemented to avoid spurious ensemble correlations between remote locations. The lack of the ensemble spread can be an issue for ensemble Kalman filter-based data assimilation systems. This was tackled by setting an inflation factor to 1.05 and also by perturbing the atmospheric forcing. The ensemble ROMS forecasts, including assimilation-free experiments, are forced using a 32-member ensemble atmospheric fluxes generated, as in *Penny et al. (2015)*, by adding ensemble perturbations from the Twentieth Century Reanalysis Project (20CR) daily averaged 3-h forecasts (*Compo et al. 2006, 2011; Whitaker et al. 2004*) to NCEP-R2 (*Kanamitsu et al., 2002*). All the model experiments start from the same initial ensemble (generated from a hindcast run of ROMS; *Jithin et al., 2017*) of 1st January, 2008 and ran up to 31st December, 2010.

. Estimation of dynamic REs

Two types of dynamic REs are estimated in the present study; observations-based dynamic RE and model-based dynamic RE following *Behringer et al. (1998)* and *Etherton and Bishop (2004)*, respectively.

3.1 Observations-based dynamic RE

The observations-based dynamic RE is estimated at every observation location (x)and analysis time (t)using the following expression, as suggested by *Behringer et al.* (1998)

$$SF \times \left(\frac{\frac{dT(x,t)}{dz} - \frac{dT(x,t)}{dz_{min}}}{\frac{dT(x,t)}{dz_{max}}} \right) \qquad (1)$$

where SF is scaling factor, $\frac{dT(x,t)}{dz}$ is the smoothened (in the vertical direction) tracer (temperature/salinity) gradient of the concurrent observed tracer profile, $\frac{dT(x,t)}{dz_{min}}$ and $\frac{dT(x,t)}{dz_{max}}$ are respectively the minimum and maximum vertical tracer gradient in the observed tracer gradient profile. Following *Behringer et al. (1998)*, the RE so estimated is added to uniform Standard Error (SE) to obtain the total observational error for the corresponding tracer profile. The chosen SE encompasses IE and sets a lower limit for the total observational error. As can be inferred from the expression *(1)*, the term within the open brackets vary between 0 and 1; the larger the vertical gradient the closer the term to unity. The parameters SF and SE enforces the OE to vary between SE and (SE+SF).

3.2 Model-based dynamic RE

The model-based dynamic RE is estimated on a coarse resolution grid (*m x m*) using high resolution (*nxn*) model outputs. The steps to estimate the model-based dynamic RE at any analysis time *t* can be summarized as follows:

Step 1: Take the high-resolution (*n x n*) assimilation-free model state, *Y*, which plays the role of the true state of the ocean in the observation space. We implicitly assume that the chosen high resolution assimilation-free model resolves the variability of the Indian Ocean at all scales.

Step 2: Choose a lower resolution grid ($m \times m$), such that m/n is an integer. Compute the spatial mean of Y on this lower resolution grid to filter out the sub-grid scale variability and estimate the ocean state (Y_c) at the lower resolution grid (typical regridding).

Step 3: Apply a simple bilinear interpolation operator P to map Y_c on the high-resolution observational space of Y.

Step 4: Take the Root-Mean-Square (RMS) of [Y-$P(Y_c)$] using the $(m/n)^2$ points corresponding to each lower resolution grid point to estimate RE at the low resolution $m \times m$ grid.

The method is straightforward to apply and naturally introduces spatio-temporal variability into RE. It is similar to that of *Etherton and Bishop (2004)* except that we use only one dynamical model for the estimation of RE. The coarse resolution grid $m \times m$ is then considered as a proxy for the coarse resolution model in *Etherton and Bishop (2004)*. Considering the fact that the model with a certain resolution can only represent the waves of wave lengths four times larger than its grid spacing (e.g. *Pielke* 1984; *Grasso,* 2000), one may choose the coarse resolution grid $m \times m$ four times coarser than the targeted assimilation grid, as a reasonable alternative for setting a dedicated coarse resolution model. In addition to this, our approach has an advantage of increasing the number of samples from $(m/n)^2$ to $(4m/n)^2$ for the estimation of RE.

Jithin et al. (2017) demonstrated that the ROMS $1/12° \times 1/12°$ covering the Indian Ocean domain closely reproduces the observed spatial and temporal variability. To estimate the model-based dynamic RE in the present study, we adopted this $1/12° \times 1/12°$($n \times n$) ROMS using the same boundary, forcing (mean NCEP-R2) and vertical resolution as the low resolution forecast ROMS model. The RE is calculated on the1°x1° grid ($m \times m$) that is 4 times

coarser than the targeted assimilation model's grid. The model-based dynamic RE so derived is interpolated on the observations locations and then added it to a uniform SE to obtain OE, as in the observations-based method.

4. Spatial and temporal structures of the estimated dynamic RE

The estimated model-based dynamic RE for sea surface temperature (SST) exhibits a significant spatio-temporal variability compared to the sea surface salinity (SSS) RE counterpart. The mean (taken over the period 2008-2010) spatial features of SST RE and SSS RE are displayed in Figures 2a and 2b ,respectively. The significant spatial variations of SST RE is evident from Figure 2a; below 0.2 °C in the equatorial regions and reaching up to 0.5 °C in the western parts of the Arabian Sea and southern parts of the Indian Ocean owing to a strong eddy activity. The spatio-temporal variability is less prominent for SSS RE compared to SST RE.; itis quite weak except in the Head Bay of Bengal, the region influenced by adjacent river discharges, where it reaches 0.5 *psu*.

Temperature and salinity changes in the whole water column largely influence the Sea Surface Height (SSH) variations (e.g. *Teague et al., 1995; Backer-Yeboah et al., 2009*). In order to identify the regions with large temperature and salinity REs in the whole water olumn, we also assessed the SSH RE. Figure 3 plots (a) SSH RE averaged over the period 2008-2010, and(b) time series of SSH RE averaged over different regions encompassing the Great whirl (black; 50°E-60°E & 5°N-10°N; GW), the South Equatorial Current (Red; 60°E-80°E & 10°S-20°S; SEC), the Head Bay of Bengal (Blue; 85°E-95°E & 15°N-20°N; HBoB) and the Southern part of Mozambique current (green; 40°E-50°E & 25°S-30°S; SMC).The SSH RE exhibits rich spatio-temporal variability. For instance, while the REs of the regions encompassing the east India coastal current, south Equatorial current, Indonesian through flow,

and Agulhas current vary between 2-4 cm, the eddy dominant regions, such as the Great whirl, the North and South branches of the Mozambique current have large REs, reaching up to 7 cm. These regions are also distinguishable in terms of seasonal and inter-annual scales (Figure 3b). For instance, the seasonal cycle dominates the temporal variations of SSH RE in the GW region, being large during summer (June-September) and small during winter (December-March) whereas the inter-annual variability is prominent in the SMC region. The differences in the spatio-temporal patterns of SSH RE and SST and SSS REs indicate important variations of temperature and salinity REs in the sub-surface layers.

Analyzing the vertical cross-sections of temperature and salinity reveals large temperature and salinity REs, reaching up to 1°C and 0.75 psu in the thermocline and halocline regions, respectively. These are also the layers that exhibit important spatio-temporal variability. Figure 4 plots the spatially averaged depth-time sections of REs for temperature for three different regions (a) GW, (b) SEC, and (c) SMC. These regions are selected based on the magnitude of SSH RE (marked in Figure 3a). One can see from Figure 4a-c that there is a great deal of prominent spatial variability at the thermocline depth. For example, the GW region shows large seasonality with the amplitude reaching 1°C, while SMC exhibits the least seasonality with an amplitude less than 0.2°C. Similarly, SEC exhibits large inter-annual variability, whereas GW has the least. Compared to temperature, the spatial RE variations for salinity are negligible with the only noticeable variations found in the Head Bay of Bengal (BoB). The magnitude of the salinity RE reaches 0.5 psu at the end of the summer monsoon season (Figure not shown), owing to the variations in fresh water discharges from the neighboring rivers.

The binned (over month and region) average temperature RE estimated from the observations-based method, as in (1) with *SF=1.5*, is plotted in Figures 4d, 4e, and 4f for GW, SEC, and SMC, respectively. The mean of dynamic RE estimated from 3 years of data is also shown in Figure 4g for all three regions. Comparing this observations-based dynamic RE with that of the above model-based dynamic RE indicates a larger mean for the former by 0.2-1°C, particularly in the thermocline layers (Figure 4g). The spatial and temporal variations of the observations-based RE are weaker compared to those of the model-based RE. For instance, the model-based temperature RE shows a strong seasonal cycle in GW (Figure 4b) and moderate inter-annual variability in SEC (Figure 4c) in the sub-surface layers, whereas they are quite weak in the observations-based RE (Figure 4d and 4e). Likewise, the distinct patterns estimated by the model-based temperature RE across the regions is weaker in the observations-based RE (Figure 4g). Similarly, the observations-based salinity RE also exhibits large RE with weaker spatial and temporal variability particularly in BoB (figure not shown).

5. Impact of the estimated dynamic REs on the Ocean analyses

Four different experiments are conducted using LETKF-ROMS to assess the impact of dynamical representational error on the ocean analyses of the Indian Ocean ensemble data assimilation system.

(i)   *Free – run* without assimilation.

(ii)   *Assim_DynObs* – assimilates *in-situ* temperature and salinity profiles from Argo using observations-based dynamic REs (with *SF* = 1.5) as *OEs* after adding uniform SEs of 0.5°C and 0.5*psu* for temperature and salinity, respectively. These coefficients are set as in the operational INCOIS-GODAS, the quality of which was assessed on various occasions (e.g. *Ravichandran, et al., 2013*; *Sivareddy et al., 2015*; *Sivareddy 2015*;

Sivareddy et al., 2017; *Pothapakula et al., 2017*). We are also motivated by the weak sensitivity of the dynamic RE experiments results to different values of this coefficient, as discussed in the Supplementary Information (Section 1 & 2).

(iii)   *Assim_DynMod* – same as *Assim_DynObs* but uses model-based dynamic REs as OEs after adding the uniform SEs of 0.75°C and 0.35psu for temperature and salinity, respectively. The choice of these SEs are motivated by two factors: (i) the insignificant sensitivity of *Assim_DynMod* to the choice of SEs (as discussed in Section 2 of the Supplementary Materials), and (ii), the choice of 0.75°C and 0.35psu in *Assim_DynMod* matches (at least to a reasonable extent, as it is not possible to achieve perfect match between them due to the pronounced spatial variations estimated by *DynMod*) the mean OEs used in *Assim_DynObs* to enable a straightforward comparison between *the two strategies*.

(iv)   *Assim_Static* – same as *Assim_DynObs* except that it uses static OEs. Different choices of static OEs were tested (supplementary section 3), and none of them found to be better than either *Assim_DynObs* or *Assim_DynMod*. Here we considered the static OE that varies in the horizontal and vertical directions by simply considering the temporal mean of the OEs used in *Assim_DynMod*. The results of this experiment are used to highlight the shortcomings of using static OEs compared to dynamically varying OEs.

Comparing the outputs from the *Free-run* and the assimilation runs with respect to independent observations suggests overall improvements with data assimilation, including for the non-assimilated state variables. Figures 5a & 5b compares time series of root-mean-square differences (RMSD) between independent *in-situ* temperature and salinity observations from RAMA (Research Moored Array for African-Asian-Australian Monsoon Analysis and

Prediction; *McPhaden et al., 1998*) buoys in the entire Indian Ocean and *Free-run*, *Assim_Static*, *Assim_DynObs*, and *Assim_DynMod*. The *Free-run* temperature and salinity fields exhibit RMSDs larger than 2°C and 0.4 psu, respectively. The assimilation with dynamic RE (both *Assim_DynObs* and *Assim_DynMod*) significantly reduces the RMSDs to 1.5°C and 0.35 psu, respectively. Time series comparisons of the RMSDs for the non-assimilated sea surface height anomaly (SSHA) and currents, with respect to altimeter based merged sea level product (AVISO 2009) and Ocean Surface Current Analysis Realtime (OSCAR; *Bonjean and Largerloef, 2002*) are shown in Figure 6. The assimilation of temperature and salinity with dynamic RE also respectively improves the estimation of SSH and currents. The improvements respectively reach up to 5 cm and 10 cm/s for SSHA and surface currents compared to *Free-run*. Assimilation with static RE also improves the estimation of all the ocean variables with respect to *Free-run*, but to a much lesser degree than the assimilation runs with dynamic RE (Figures 5 & 6).

Figure 7 & 8 display the spatial structures of various statistical parameters for SSHA, including (7a) observed standard deviation, (7b) RMSD of *Free-run* with respect to AVISO, and difference in RMSD between (8a) *Assim_static* and *Free-run* (8b) *Assim_DynObs* and *Free-run* and (8c) *Assim_DynMod* and *Free-run*. In Figure 8, a positive value indicates degradation and negative scale indicates improvement after assimilation. The *Free-run* shows large RMSD in the southern Indian Ocean, reaching up to 18 cm that is mostly larger than STD. Assimilation using *Assim_static* effectively reduces these errors to 16 cm, but degrades the quality of SSHA elsewhere. Compared to *Assim_Static*, the improvements are larger and relatively homogeneous in *Assim_DynObs*. However, the best performance is obtained by assimilating observations with the model-based RE in *Assim_DynMod*. For instance, as can be

seen in Figure 5, although not as clear for salinity, the RMSDs for temperature in *Assim_DynMod* are always lower than those of *Assim_DynObs,* with improvements reaching up to 0.5°C. The *Assim_DynMod* yields SSHA improvements in the BoB, in contrast to net degradations in *Assim_Static* and in *Assim_DynObs* (comparing Figure 8c with 8a and 8b). Also, the SSHA improvements resulting from *Assim_DynMod* in the southern Indian Ocean are larger and more homogeneous than those of the improvements obtained in *Assim_DynObs.* SSHA RMSDs in *Assim_DynMod* (figure not shown) are further always below the observed standard deviations, which was not always true for *Assim_DynObs* and *Assim_Static*. Similar results were also obtained from the comparison of the model estimated velocity fields with the OSCAR fields (results not shown).

In order to provide further insights on the relative improvements of *Assim_DynMod*, we examined the evolution of SSHA RMSDs along 10°S as they result from (a) *Free,* (b) *Assim_Static,* (c) *Assim_DynObs,* and (d) *Assim_DynMod* (Figure 9). While the RMSDs are mostly below 8cm in both *Free* (Figure 9a) and *Assim_DynMod* (Figure 9d), they reach ~20cm and ~12cm, a degradation of ~12cm and ~4cm, respectively in *Assim_Static* (Figure 9b) *and Assim_DynObs* (Figure 9c) during March-August, 2009 between 50°E-90°E. This large increase in SSHA RMSDs, compared to the robustness in *Assim_DynMod*, suggests that *Assim_Static* and *Assim_DynObs* are over-fitting the observations in these areas/periods. The *Assim_DynMod* is not only more robust in terms of observation fitting, but seems also to exploit more efficiently the information from the observations. This can be seen from the spatio-temporal extent of improvement in SSHA RMSDs in *Assim_DynObs* and *Assim_Static,* which are always less pronounced than those of *Assim_DynMod.*

Prescribing realistic values of OEs in the assimilation system is important to minimize the adverse effects of over-fitting/under-fitting (e.g. *Oke and Sakov*, 2008). Given that the OEs in *Assim_DynMod* and *Assim_DynObs* were already scaled to similar means, the better results of *Assim_DynMod* can be attributed to the more realistic spatio-temporal variations of RE. Relying purely on observational profiles that are often sparse may not always provide reliable estimates of the statistics of the REs in *DynObs*. The southern Indian Ocean is one such a region where spatial and year-to-year variations of RE are large (Figure 3 & Figure 4), and the Argo coverage is sparse (Figure 1). This explains why *Assim_DynMod* leads to considerable improvements with respect to *Assim_DynObs* in this region (comparing Figure 8c with 8b).

6. Summary and Conclusions

The impact of using dynamical Representation Errors (RE) on the ocean analyses of an ensemble data assimilation system is studied using three different approaches, an observations-based dynamic RE, model-based dynamic RE, and static RE. The first is estimated from the local vertical gradients of concurrent temperature/salinity profiles, the second is estimated from the outputs of a high-resolution ocean model, and the third is estimated by using the temporal mean of OE from the second approach. The estimated REs for SSH, temperature and salinity exhibit significant spatial and temporal variability that are much more pronounced when estimated from the model outputs.

Inter-comparing the ocean analyses estimated using three different REs; static RE (*Assim_Static)*, observations-based dynamic RE (*Assim_DynObs)*, and model-based dynamic RE (*Assim_DynMod*), with independent observations in the Indian Ocean shows that *Assim_Static* was the least performant. It limited the benefits of assimilating observations due to the issues over-fitting/under-fitting. *Assim_DynMod* obtained significant improvements. The

reduction in RMSDs for temperature and salinity reaches 50% and 30%, respectively. Comparisons between *Assim_DynMod* and *Assim_DynObs* demonstrates the relevance of the model-based dynamic RE. The improvements of *Assim_DynMod* over *Assim_DynObs* reach up to 0.5°C and 0.2psu (5 cm and 5 cm/s) for temperature and salinity (SSH and currents), respectively. The reduction in SSH RMSD further reaches 30%, particularly in the south Indian Ocean. Accurate estimation of OEs in both space and time enhances the assimilation analyses by limiting the occurrence of degradations from over-fitting. Using such improved representation of spatio-temporal variations of REs, *Assim_DynMod* was able to better exploit the observations information than *Assim_DynObs*. It further helped *Assim_DynMod* to better handle the strongly variable regions and those with sparse observations coverage.

The estimation of the model-based dynamic RE involves running an assimilation-free high-resolution ocean model in conjunction to running the assimilation system. This is an additional computational requirement compared to the observations-based dynamic RE. Considering, however, the positive impact of the model-based dynamic RE method and the continuous progress in computational resources, this method offers a viable alternative forocean data assimilation systems.

References:

Antonov, J. I. et al.(2009). World Ocean Atlas 2009. Volume 2: Salinity. (ed. Levitus, S.), *NOAA Atlas NESDIS* 69, US Government Printing Office, Washington, DC, 2010, p. 184.20.

AVISO, Ssalto/Duacs User Handbook.(2009). (M)SLA and (M)ADT Near-Real Time and Delayed Time Products. Reference, CLS-DOS-NT-06.034, AVISO Altimetry, France, 51.

Backer-Yeboah, S., Watts, D.R., and Byrne, D.A.(2009). Measurements of Sea Surface Height Variability in the Eastern South Atlantic from Pressure Sensor–Equipped Inverted Echo Sounders:Baroclinic and Barotropic Components. *Journal of Atm. Ocean. Technology*, Vol. 26, 2593-2609, DOI: 10.1175/2009JTECHO659.1

Balaji, B., Deepsankar, B., Paul, B., Sivareddy, S., and Arya Paul. (2018). LETKF-ROMS: An improved predictability system for the Indian Ocean. *Technical Report, ESSO-INCOIS-MDG-TR-03*, available at http://moeseprints.incois.gov.in/id/eprint/4572.

Balmaseda, M. A., Mogensen, K., and Weaver, T. (2013). Evaluation of the ECMWF ocean reanalysis system ORAS4. Q. . *J. R.Meteorol. Soc*. 139: 1132-1161. DOI:10.1002/qj.2063

Behringer, D. W., Ji, M. & Leetmaa, A. (1998). An improved coupled model for ENSO

prediction and implications for ocean initialization. Part I: The ocean data assimilation system. *Monthly Weather Review*. 126(4), pp. 1013–1021.

Bonjean, F., Lagerloef, G.S.E.(2002). Diagnostic model and analysis of the surface currents in the tropical Pacific Ocean. *J. Phys. Oceanogr.* 32, 2938–2954.

Bormann N, Bonavita M, Dragani R, Eresmaa R, Matricardi M, McNally A. (2016). Enhancing the impact of IASI observations through an updated observation-error covariance matrix. *Q. J. R. Meteorol. Soc.* **142**: 1767 – 1780. https://doi.org/10.1002/qj.2774.

Compo, G. P., J. S. Whitaker, and P. D. Sardeshmukh. (2006). Feasibility of a 100-year reanalysis using only surface pressure data. *Bull. Amer. Meteor. Soc.,* 87, 175-190, doi:10.1175/BAMS-87-2-175.

Dee D, da Silva A. (1999). Maximum-likelihood estimation of forecast and observatrion-error covariance parameters. Part 1: Methodology. *Mon. Weather Rev*. 127: 1822-1834.

Desroziers, G., O. Brachemi, and B. Hamadache. (2001). Estimation of the representativeness error caused by the incremental formulation of variational data assimilation. *Quart. J. Roy. Meteor. Soc.,* 127, 1775–1794.

Desroziers, G., Berre, L., Chapnik, B. and Poli, P. (2006). Diagnosis of observation, background and analysis-error statistics in observation space. *Q. J. Roy. Meteorol. Soc.* 126, 3385-3396.

Etherton BJ, and Bishop CH. (2004). Resilience of hybrid ensemble/3DVAR analysisschemes to model error and ensemble covariance error. *Mon. Weather Rev.132*: 1065–1080.

Fairall, C. W., Bradley, E. F., Rogers, D. P., Edson, J. B. and Young, G. S.(1996). Bulk

parameterization of air–sea fluxes for TOGA COARE. *Journal of Geophysical Research.,* 101, 3747–3764.

Forget, G. and C. Wunsch. (2007). Estimated global hydrographic variability. *J. Phys. Oceanogr.* 37, 19997-2008, doi:10.1175/JPO3072.1.

Francis, P.A., Vinayachandran, P. N., and Shenoi, S. S. C.(2013). The Indian Ocean Forecast System. *Current Science.*, Vol. 104, No. 10, 25,1354-1368p.

Giese, B.S., and S. Ray, (2011). ElNino variability in simple ocean data assimilation (SODA), 1871-2008. *J. Geophys. Res.,* 116, C02024, doi:10.1029/2010jc006695.

Grasso, L.D. (2000). The differentiation between Grid spacing and resolution and their application to numerical modeling. *Bulletin of American Meteorological Society*, No. 3, 81, 580-581p.

Griffies, S., and Hallberg, R.W.(2000). Biharmonic friction with a Smagorinsky-like viscosity for use in large-scale eddy-permitting ocean models. *Mon. Weather Rev*. 128 (8), 2935–2946.

Haidvogel, D. B., Arango, H. G., Hedstrom, K., Beckmann, A., Malanotte-Rizzoli, P. and Shchepetkin, A. F., (2000). Model evaluation experiments in the North Atlantic Basin: simulations in nonlinear terrain-following coordinates. *Dyn. Atmos. Oceans*. 32, 239–281.

Hoteit, I., Cornuelle, B., and Heimbach, P. (2010). An eddy-perimitting, dynamically consistent adjoint-based assimilation system for the tropical Pacific: Hindcast experiments in 2000, *J. Geophys. Res.,* 115, C03011, doi: 10.1029/2009JC005437.

Hoteit, I., Hoar, T., Gopalakrishnan, G., Collins, N., Anderson, J., Cornuelle, B., Kohl, A., and Heimbach, P. (2013). A MITgcm/DART ensemble analysis and prediction system with application to the Gulf of Mexico. *Dynamics of Atmospheres and oceans* 63, 1-23, http://dx.doi.org/10.1016/j.dynatmoce.2013.03.002

Ingleby, B., and Huddleston, B. (2007). Quality control of ocean temperature and salinity profiles – historical and real-time data. *Journal of Marine systems*. 65, 158-175, 10.1016/j.jmarsys.2005.11.019.

Janjic, T., Bormann, N., Bocquet, M., Carton, J. A., Cohn, S.E., Dance, S.L., Losa, S.N., Nichols, N.K., Potthast, R., Waller, J.A., and Weston, P. (2017). On the representation error in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, doi:10.1002/qj.3130

Jithin, A.K., Francis, P.A., Abhisek Chatterjee, Suprit K., and Vijayan Fernando. (2017). Validation of the simulations by the High resolution operational ocean forecast and reanalysis system (HOOFS) for the Bay of Bengal. *MoEsTechnical Report,* No ESSO/INCOIS/ISG/TR/01

Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S.H., Hnilo, J.J., Fiorino, M., and Potter, G.L. (2002). NCEP-DOE AMIP-II Reanalysis (R-2). *Bulletin of the American Meteorological Society*,1631-1643, *https://doi.org/10.1175/BAMS-83-11-1631*

Karspeck, A. R. (2016). An Ensemble approach for the estimation of observational error illustrated for a nominal 1 global ocean model. *Monthly Weather Review,* 144, p1713-1728, Doi: 10.1175/MWR-D-14-00336.1

Karspeck, A. R. Yeager, G. Danabasoglu, T. Hoar, N. Collins, K. Raeder, J. Anderson, and J.

Tribbia (2013). An ensemble adjustment Kalman filter for the CCSM4 ocean component. *J. Climate*, 26, 7392-7413, doi:10.1175/JCLI-D-12-00402.1.

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

Locarnini, R. A. et al.(2009). World Ocean Atlas 2009, Volume 1: Temperature (ed. Levitus, S.), NOAA Atlas NESDIS 68, US Government Printing Office, Washington, DC, 2010, p. 184.

McPhaden, M. J. *et al.* (1998). The Tropical Ocean-Global Atmosphere observing system: A decade of progress. *Journal of Geophysical Research: Oceans*. **103(C7)**, pp.14169–14240.

Oke, P.R., Brassington, G. B., Griffin, D.A., and Schiller, A. (2008). The Bluelink ocean data assimilation system (BODAS). *Ocean Modelling*, 21, 46-70, https://doi.org/10.1016/j.ocemod.2007.11.002

Oke, P.R., Sakov P. (2008). Representation error of oceanic observations for data assimilation. *J. Atmos. Oceanic Technol. 25:* 1004–1017. https://doi.org/10.1175/2007JTECHO558.1.

Penny, S. G. (2011). Data assimilation of the global ocean using the 4D local ensemble transform Kalman filter (4D-LETKF) and the Modular Ocean Model. *Ph.D. dissertation, University of Maryland*, College Park, 141 pp.

Penny, S.G., Behringer, D. W., Carton, J. A., and Kalnay, E. (2015). A Hybrid Global ocean data assimilation system at NCEP. *Monthly weather Review*, 143, p4660-4677, doi:10.1175/MWR-D-14-00376.1

Pielke, R.A., 1984: Mesoscale Meteorological Modeling. *Academic Press*, 612 pp.

Pothapakula, P.K., Osuri, K.K., Pattanayak, S., Mohanty, U.C., Sil, S. and Nadimpalli, R.(2017). Observational perspective of SST changes during life cycle of tropical cyclones over Bay of Bengal. *Natural Hazards*, 88(3), pp.1769-1787.

Ravichandran, M., Behringer, D., Sivareddy, S., Girishkumar, M.S., Neethu Chacko., and Harikumar, R. (2013). Evaluation of the Global Ocean Data Assimilation System at INCOIS: The Tropical Indian Ocean. *Ocean modelling*, 69, p123-135, doi:http://dx.doi.org/10.1016/j.oceamod.2013.05.003

Richman, J. G., Miller, R. N., and Splitz, Y. H. (2005). Error estimates for assimilation of sea surface temperature data in ocean climate models. *Geophysical Research Letters*, 32, L18608, doi:10.1029/2005GL023591.

Roemmich, D., Johnson, G. C., Riser, S., Davis, R., Gilson, J., Owens, W. B., Garzoli, S. L., Schmid, C., and Ignaszewski, M. (2009). Observing the Global Ocean with Profiling Floats. Oceanography. 22, 2, 34-43.

Schott, F.A., Xie, S.-P., and McCreary, P. Jr. (2009). Indian Ocean circulation and climate variability. *Rev. Geophys.,* 47, RG1002, doi:10.1029/2007RG000245.

Shchepetkin, A. F., and McWilliams, J. C.(2005). The regional ocean modeling system (ROMS): A split-explicit, free-surface, topography-following-coordinate oceanic model. *Ocean Modelling*., 2005,9/4, 347–404; doi: 10.1016/j.ocemod.2004.08.002.

Sivareddy, S. (2015). A study on global ocean analysis from an ocean data assimilation system and its sensitivity to observations and forcing fields, *Ph.D. thesis, Andhra University*. Available at http://www.incois.gov.in/documents/PhDThesis_Sivareddy.pdf.

Sivareddy, S., Arya Paul, Sluka, T., Ravichandran, M., and Kalnay, E. (2017). The pre-Argo ocean reanalyses may be seriously affected by the spatial coverage of moored buoys. *Scientific Reports*, Vol. 7, 46685, doi:10.1038/srep46685

Sivareddy, S., Ravichandran, M., Grishkumar, M.S., and Prasad, K.V.S.R. (2013). Assessing the impact of various wind forcing on INCOIS-GODAS simulated ocean currents in the equatorial Indian Ocean. *Ocean Dynamics*, 65, 1235-1247, doi 10.1007/s10236-015-0870-6

Song, Y. and Haidvogel, D. B. (1994). A semi-implicit ocean circulation model using a generalized topography-following coordinate system. *Journal of Computational Physics.*, 115, 228–244.

Teague, W. J., Z. R. Hallock, G. A. Jacobs, and J. L. Mitchell. (1995). Kuroshio sea surface height fluctuations observed simultaneously with inverted echo sounders and TOPEX/POSEIDON. *J. Geophys. Res.,*100, 24 987–24 994.

Whiltaker, J. S., G. P. Compo, X. Wei, and T. M. Hamil. (2004). Reanalysis without radiosondes using ensemble data assimilation. *Mon. Wea. Rev.,* 132, 1190-1200, doi: 10.1175/1520-0493(2004) 132<1190:RWRUED>2.0.CO;2.

Wunch, C. (1996). The ocean circulation Inverse problem, Cambridge Univ. Press, New York.

Wunsch, C., and P. Heimbach (2007). Practical global oceanic state estimation. *Physica D*, 230, 197-208. Doi:10.1016/j.physd.2006.09.040.

Xu, F.-H., Oey, L.-Y., Miyazawa, Y., and Hamilton, P., (2013). Hindcasts and forecasts of Loop currents and eddies in the Gulf of Mexico using local ensemble transform Kalman filter and optimum-interpolation assimilation schemes. *Ocean modelling*, 69, 22-23, doi:http://dx.doi.org/10.1016/j.ocemod.2013.05.002.

Xu, Fang-Hua, and Oey, Lie-Yauw. (2014). State analysis using the Local Ensemble Transform Kalman Filter (LETKF) and the three-layer circulation structure of the Luzon strait and the South China Sea. *Ocean Dynamics*, 64, 905-923, doi: 10.1007/s10236-014-0720-y.

Zhang, S., Harrison, M. J., Rosati,A., and Wittenberg, A. (2007). System design and evaluation of coupled ensemble data assimilation for global oceanic climate studies. *Monthly Weather Review*, 135(10), 3541-3564.
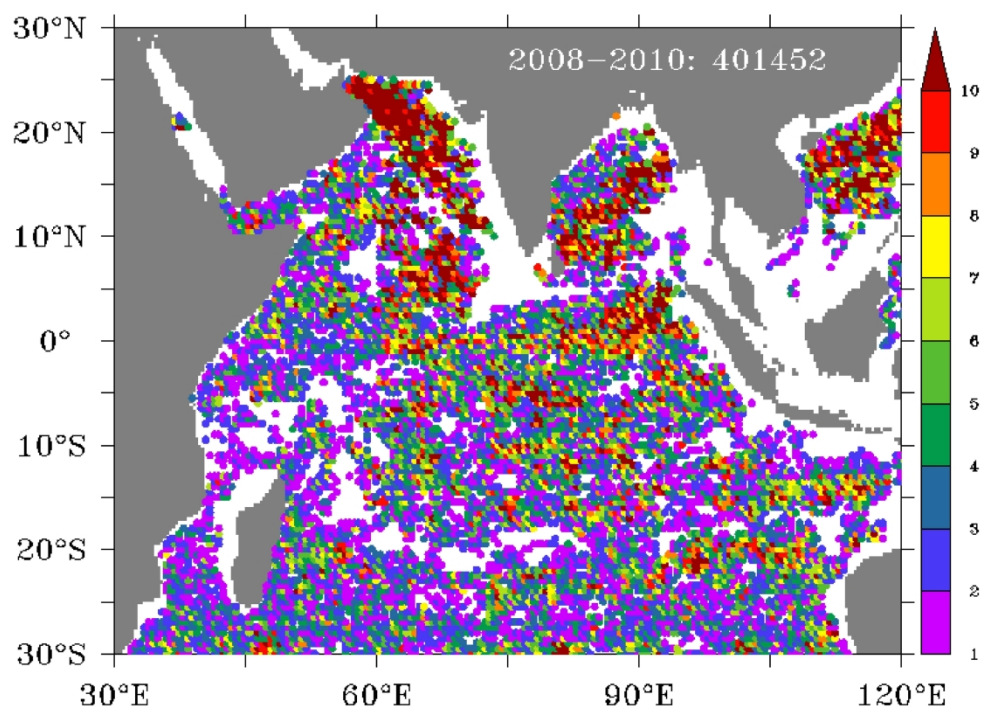
Figure 1. Spatial coverage of Argo during 2008 to 2010 on 0.5 x 0.5 degree grid. Total number of profiles taken in the whole domain is also indicated in the figure.
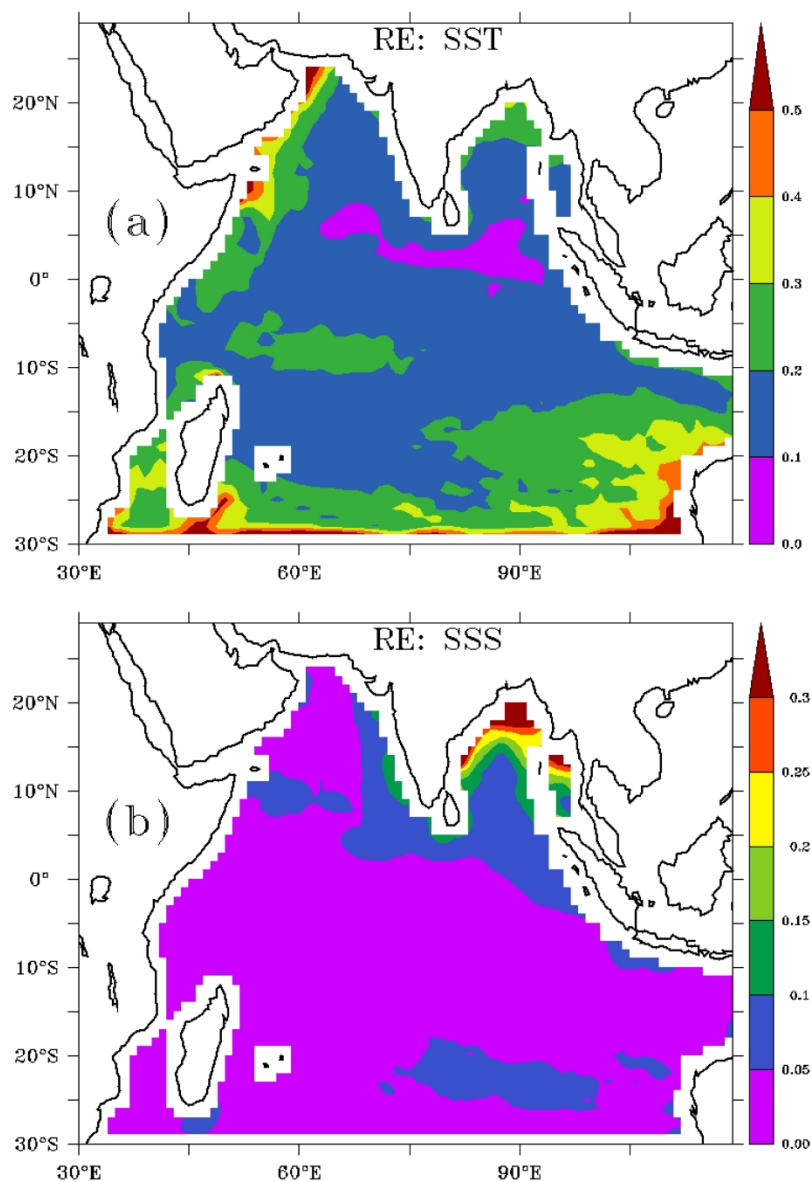
Figure 2. Model-based dynamic Representation Error (RE) for (a) sea surface temperature (SST; °C), and (b) sea surface salinity (SSS; psu), averaged over the period 2008-2010.
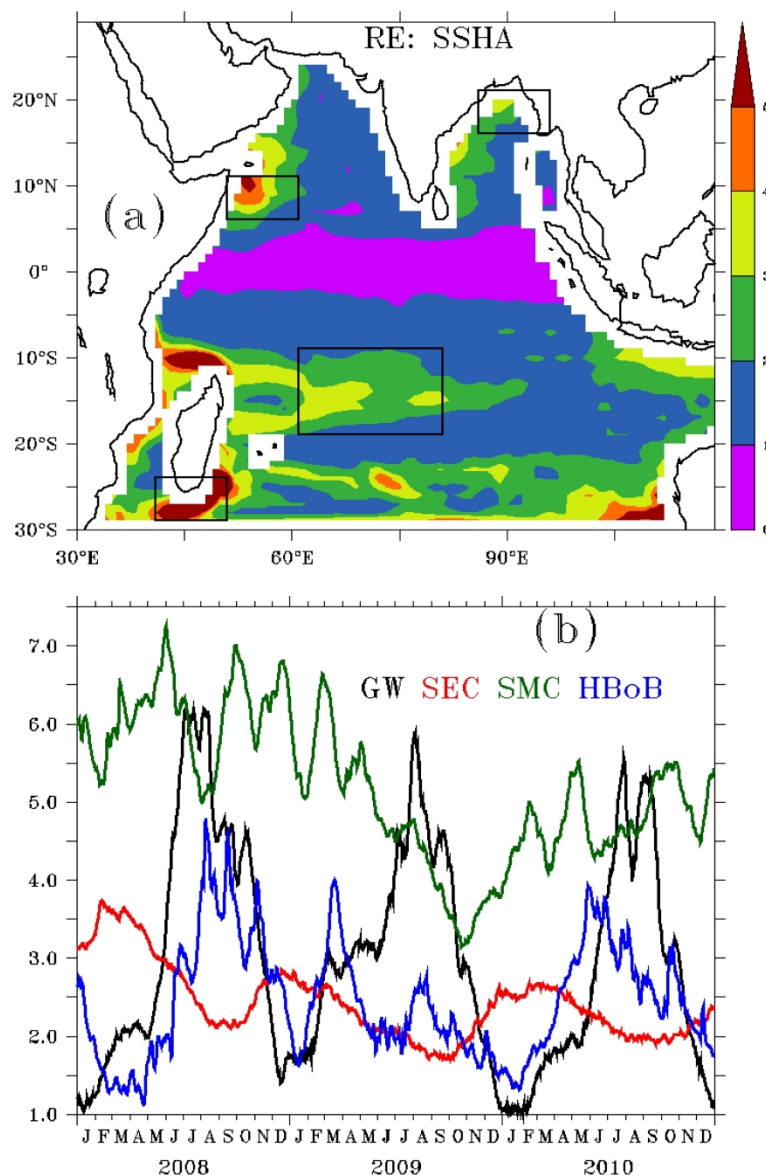
Figure 3. Model based dynamic Representation Error (RE) for (a) sea surface height (SSH; cm) averaged over the period 2008-2010, and (b) region averaged time series of SSH RE corresponding to the Great whirl (black; 50°E-60°E & 5°N-10°N; GW), the South Equatorial Current (Red; 60°E-80°E & 10°S-20°S; SEC), the head Bay of Bengal (Blue; 85°E-95°E & 15°N-20°N; HBoB), and the Southern part of Mozambique current (green; 40°E-50°E & 25°S-30°S; SMC). Panel a indicates the selected boxes on geographical map.
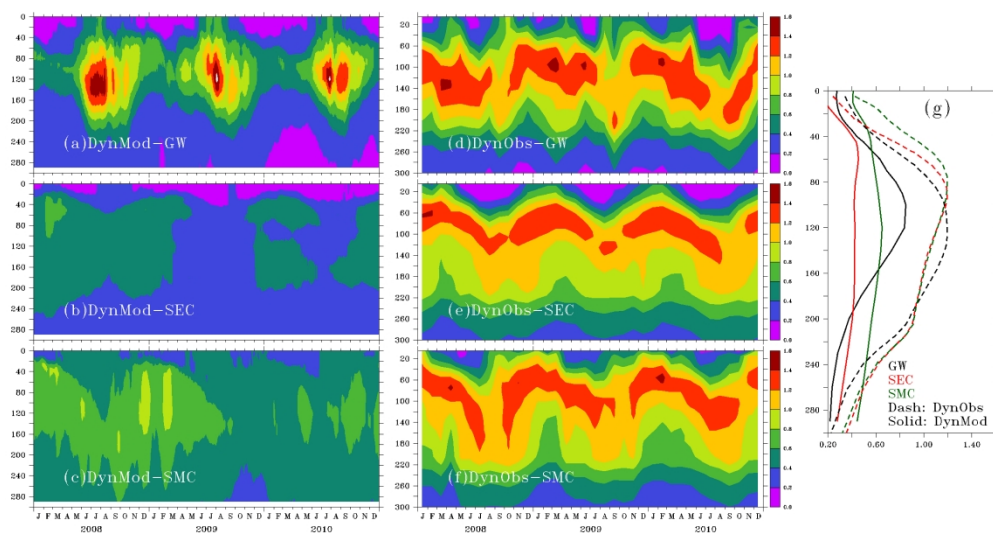
Figure 4. Depth-time sections of model-based dynamic RE for temperature (°C) averaged over three different boxes, (a) GW, (b) SEC, and (c) SMC. Panels d, e, and f show the observations-based dynamic RE binned over the corresponding regions and months. Mean REs corresponding to the analysis period 2008-2010 for all the regions from both observations-based and model-based dynamic RE are shown in panel g.
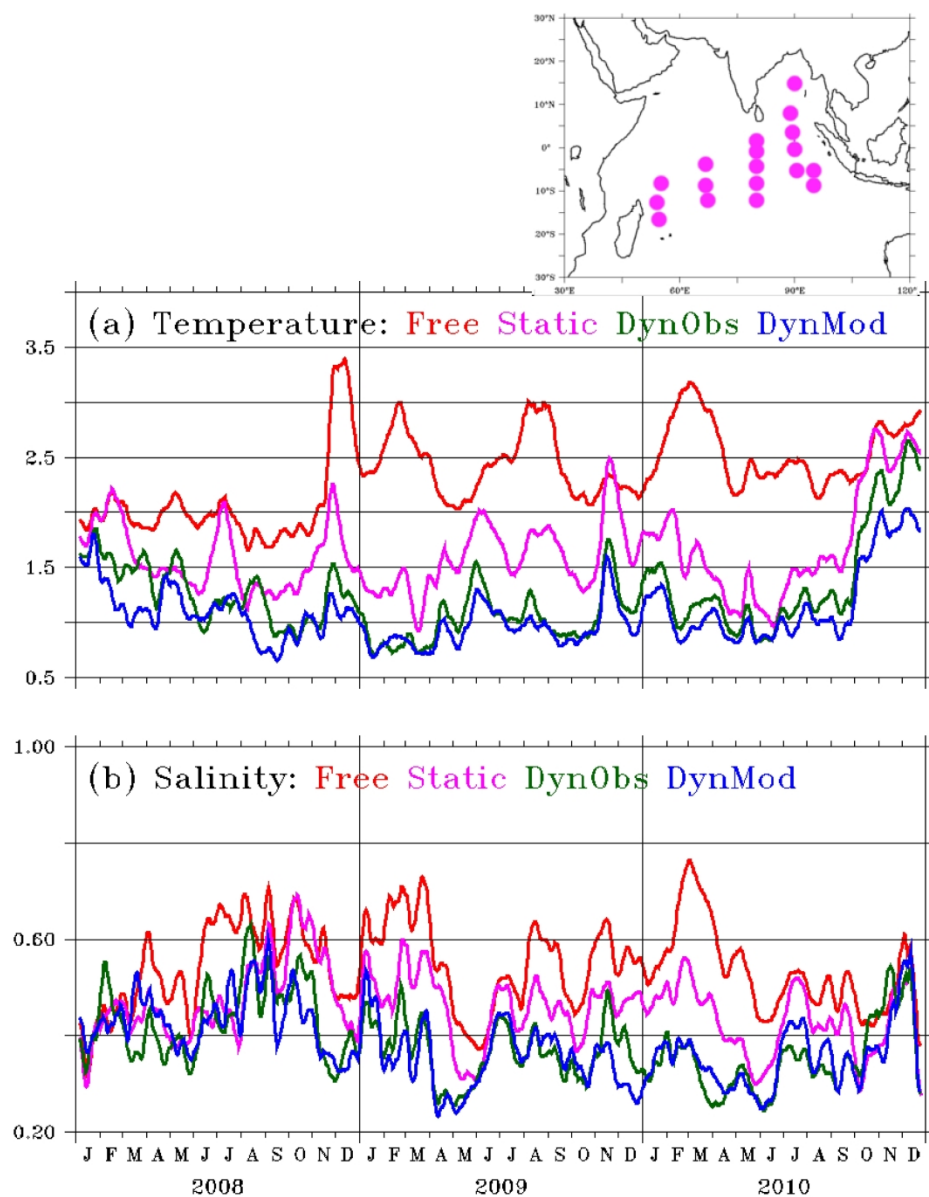
Figure 5. Time series of RMSD for (a) temperature (°C) and (b) salinity (psu) from Free (red), Assim_Static (pink), Assim_DynObs (green), and Assim_DynMod (blue) experiments. RMSD is computed by interpolating the model outputs on available independent RAMA observations over the entire Indian Ocean Domain. The RAMA locations are indicated in the inset with pink dots. 10-day smoothing is applied to better highlight the features.
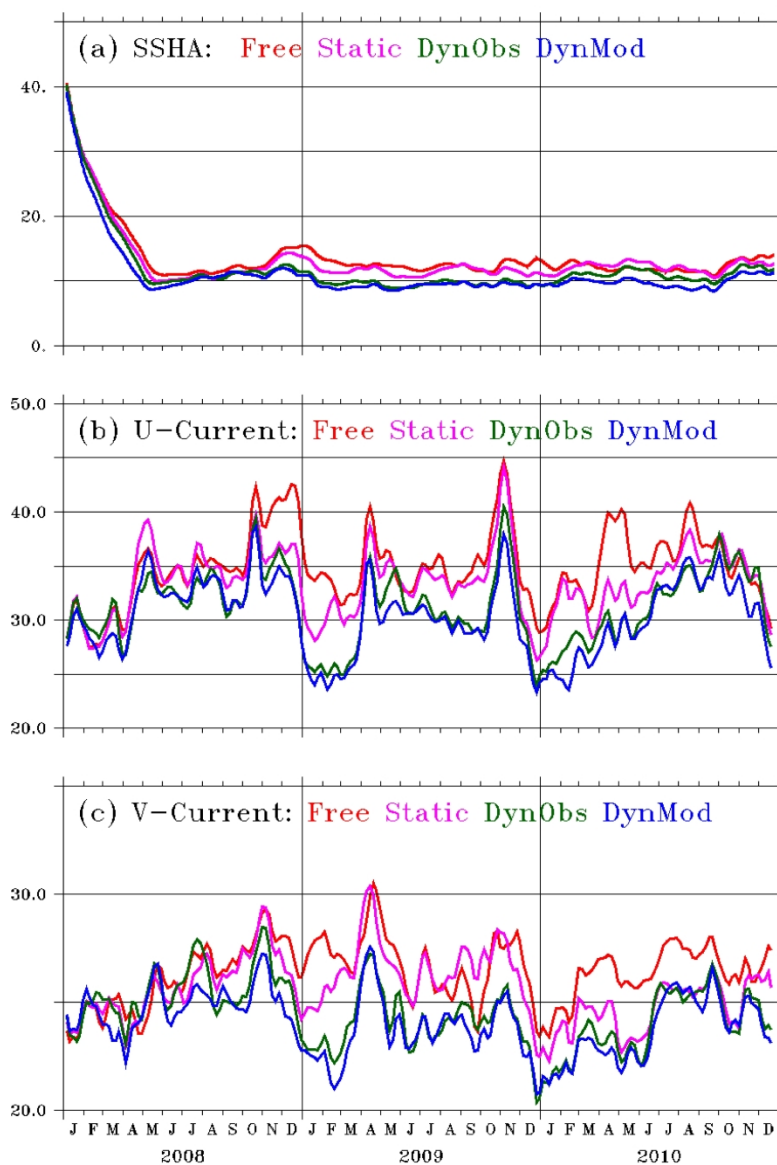
Figure 6. Time series of RMSD for non-assimilated parameters (a) sea surface height anomaly (cm), zonal surface current (cm/s), and (c) meridional surface currents (cm/s) from Free (red), Assim_Static (pink), Assim_DynObs (green), and Assim_DynMod (blue) experiments. RMSD is computed by collocating the model simulations onto the gridded products (AVISO for SSHA and OSCAR for surface currents). 10-day smoothing is applied to better highlight the features.
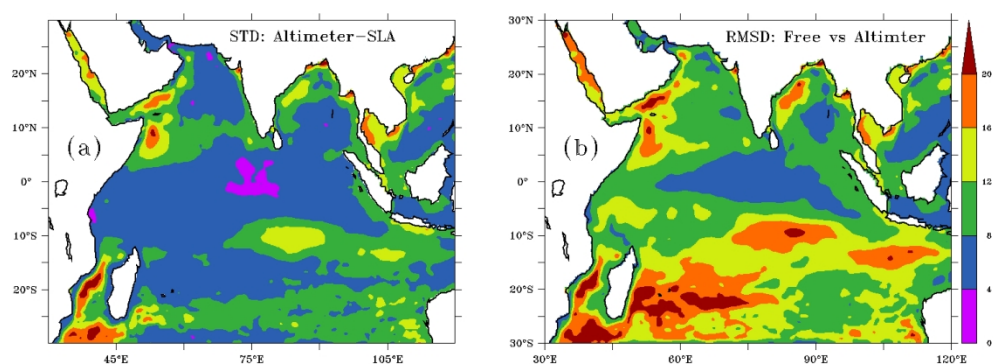
Figure 7. Spatial structures of (a) observed Standard deviation in SSHA from AVISO product, (b) RMSD of Free_run with respect to AVISO SSHA. Units are in cm. Statistics are based on 1st July, 2008 to 31st December, 2010.
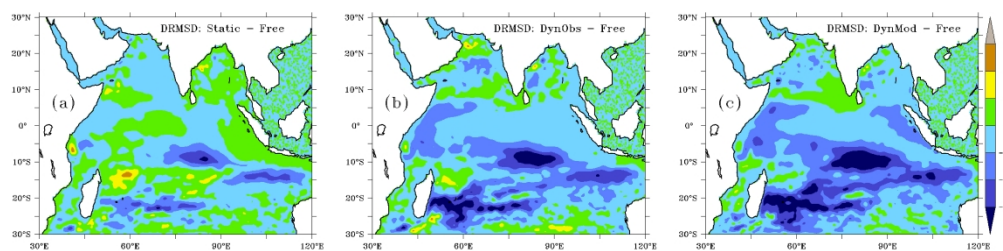
Figure 8. Difference in RMSD between (a) Assim_Static and Free_run, (b) Assim_DynObs and Free_run, and (c) Assim_DynMod and Free_run. Units are in cm. Positive values indicate degradation and negative values indicate improvement after assimilation. Statistics are based on 1st July, 2008 to 31st December, 2010.

Figure 9. Temporal evolution of the root-mean-square-differences (RMSD) of SSHA (cm) in (a) Free, (b) Assim_Static, (c) Assim_DynObs, and (d) Assim_DynMod along the latitudinal band of 10°S. RMSD is calculated by comparing model outputs to merged-altimeter-AVISO SSHA, between 8°S-12°S. 10-day smoothing is applied to better highlight the features.