Insights from very Large Ensemble Data Assimilation Experiments with a High Resolution General circulation model of the Red Sea

Sivareddy Sanikommu[1], Naila Raboudi[1], Mohamad El Gharampti[2], Peng Zhan[3], Bilel Hadri[1], Ibrahim Hoteit[1*]

[1] King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

[2]National Center for Atmospheric Research, USA

[3]Department of Ocean Science and Engineering, Southern University of Science and Technology, Shenzhen, China

Key points

1. Assimilation experiments with up to 5000 members are conducted with high resolution Red Sea model
2. Localization remains useful even with large ensembles
3. Accounting for model uncertainties is more beneficial than increasing the ensemble size
4. Large-ensemble forcing fields and non-Gaussian assimilation methods may help benefiting more from large ensembles

[*]**Corresponding author:** Prof. Ibrahim Hoteit

King Abdullah University of Science and Technology (KAUST)

Physical Science and Engineering Division

Thuwal 23955-6900, Saudi Arabia

E-mail: ibrahim.hoteit@kaust.edu.sa

Abstract

Ensemble Kalman Filters (EnKFs), which assimilate observations based on statistics derived from samples of ocean states called ensemble, have become the norm for ocean data assimilation (DA) and forecasting. These schemes are commonly implemented with inflation and localization techniques to increase their ensemble spread and to filter out spurious long-range correlations resulting from the limited-size ensembles imposed by computational burden constraints. Such ad hoc methods were found not necessary in ensemble DA experiments with simplified ocean/atmospheric models and large ensembles. Here, we conduct a series of 1-year-long ensemble experiments with a fully realistic EnKF-DA system in the Red Sea using tens-to-thousands of ensemble members. The system assimilates satellite and in-situ observations and accounts for model uncertainties by integrating a 4km-resolution ocean model with ECMWF atmospheric ensemble fields, perturbed internal physics and initial conditions for forecasting.

Our results indicate that accounting for model uncertainties is more beneficial than simply increasing the ensemble size, with the improvements due to large ensemble leveling off at about 250 members. Besides, and in contrast to what is commonly observed with simplified models, the investigated ensemble DA system still required localization even when implemented with thousands of members. These findings are explained by (i) amplified spurious long-range correlations produced by the low-rank nature of the ECMWF atmospheric forcing ensemble, and (ii) non-Gaussianity generated by the perturbed internal physical parameterization schemes. Large ensemble forcing fields and non-Gaussian DA methods might be needed to take full benefits from large ensembles in ocean DA.

Plain Language Summary

Data assimilation (DA) using Ensemble Kalman Filters (EnKFs) requires large ensembles to estimate robust statistics in order to efficiently spread the observation's information to all model variables, key for robust predictions of the ocean state. Until recently, only limited ensembles (~10-100 members) could be afforded in realistic ocean DA applications. With the ever-increasing computational resources, the use of larger ensembles will become possible in the near future. In this context, the present study assesses the performance of a fully realistic high-resolution ocean EnKF-DA system by systematically examining its sensitivity to ensembles composed of tens-to-thousands of members. It offers fresh perspectives on the employment of ad hoc inflation and localization methods, which have been traditionally implemented with EnKFs to compensate for the use of small ensembles and for the omittance of various model uncertainties. The results of this study suggest large-ensemble forcing fields as well as non-Gaussian DA methods may be needed to maximize the benefits of large ensembles in ocean DA.

1. Introduction

Ensemble Kalman Filter (EnKF) based data assimilation (DA) systems provide an efficient framework to update the background error covariance, a critical element for any ocean DA system to spread the observation information across model variables [*Derber and Bouttier,* 1999; *Bouttier and Courtier,* 1999, 2002; *Edwards et al.,* 2015; *Hoteit et al.,* 2018]. In ENKFs, a set of ocean states, called ensemble, is integrated using an ocean model for forecasting the first two-moments statistics of the ocean state, i.e., sample mean and covariance, which are in turn used to apply a Kalman filter to update the forecast [*Hoteit et al.,* 2015]. These DA schemes require the size of the ensemble to be large enough to estimate robust statistics [*Mitchell et al.,* 2002; *Houtekamer and Zhang,* 2016; *Lei and Whitaker,* 2017; *Leutbecher,* 2019]. Using small ensembles was shown to produce spurious long-range covariances and rapid collapse of the ensemble spread after few assimilation cycles [*Houtekamer and Mitchell*, 1998; *Anderson* 2001]. In spite of this, only limited ensembles (~10-100 members) were so far considered in real-world applications mainly to reduce the computational burden [e.g., *Hoteit et al.,* 2012; *Wang and Lei,* 2014; *Penny et al.,* 2015; *Lei and Whitaker*, 2017; *Baduru et al.,* 2019; *He et al.,* 2019; *Toye et al.,* 2020; *Sanikommu et al.,* 2023]. Ad hoc techniques such as inflation and localization, which respectively inflates ensemble covariances artificially and limits the observations influence to only within certain radius by tapering long-range correlations in the ensemble, are usually deployed as a compensation for small ensemble size [e.g., *Houtekamer and Mitchell*, 1998, 2001; *Hunt et al.,* 2007; *Bishop and Hodyss,* 2007; *Miyoshi,* 2011; *Whitaker and Hamill,* 2012; *Lee et al.,* 2017; *Luo et al.,* 2018]. With the ever-increasing computational resources, the use of large ensembles will become possible in the near future. Hence, identifying the benefits and outstanding issues associated with large ensembles is important to provide new insights into the future applications of ensemble DA methods in ocean applications.

Up to date, large ensemble experiment studies (LEEs) with realistic ocean and atmospheric DA and forecasting systems remain limited. The studies of *Miyoshi et al.* (2014), *Kondo and Miyoshi* (2016) and *Toye et al.* (2020) have suggested that with large ensembles, the performance of the ensemble DA system generally becomes less dependent on inflation and localization. However, these LEE studies were not performed with a fully realistic ensemble DA framework. For instance, the LEE (up to 10,000 members) results of the first two studies were based on a mid-complex coupled ocean-atmospheric model and assimilated pseudo-observations. *Toye et al.,* (2018) conducted LEEs, up to 1000 members, with an ocean general circulation model of the Red Sea assimilating real observations, but only accounted for uncertainties in the initial conditions. Ensembles from such mid-complex LEE systems were also investigated to (1) examine the characteristics of background model errors [*Pinardi et al.,* 2008; *Jacques and Zawadzki*, 2015], (2) understand and model sampling errors [*Necker et al.,* 2020a, 2020b], (3) assess non-Gaussianity and sensitivity to covariance localization [*Miyoshi et al.,* 2014; *Kondo and Miyoshi,* 2016; *Toye et al.,* 2018], and (4) examine the potential impact of observations [*Necker et al.,* 2020a].

Accounting for uncertainties of a forecasting ocean model via stochastic inputs and parameters has recently become popular in the EnKF DA systems [see review papers *Martin et al.,* 2015; *Houtekamer and Zhang,* 2016]. In ocean applications, uncertainties are now considered in the atmospheric forcing, boundary conditions, internal physics, bathymetry, and of course initial conditions. These studies suggested noticeable improvements, compared to those that do not account for uncertainties, in the performance of the underlying ensemble assimilation systems [e.g. *Fujita et al.,* 2007; *Bowler et al.,* 2008; *Houtekamer et al.,* 2009; *Kwon et al.,* 2016; *Penny et al.,* 2015; *Vandenbulcke and Barth,* 2015; *Sanikommu et al.,* 2017, 2019, 2020; *Baduru et al.,* 2019]. Accounting for uncertainties in model's inputs may however introduce non-Gaussian features in the distribution of the forecast ensemble due to the model's non-linearity [e.g., *Sura et al.,* 2005;

Sura and Sardeshmukh, 2008; *Sura and Hannachi*, 2015], which may limit the performances of the Gaussian-based EnKFs [e.g. *Anderson*, 2001; *Whitaker and Hamill*, 2002; *Hoteit et al., 2008, 2012; Subramanian et al., 2012*].

Here we conduct for the first time LEEs with up to 5000 members using a fully realistic high-resolution EnKF-DA system of the Red Sea (RS) and assimilating real observations accounting for uncertainties in the atmosphere and physics, in addition to the initial conditions. We assess the sensitivity of the system to increasing ensembles. We examine in particular probability distribution of the forecasted ensemble and investigate the need for localization. We further examine the sensitivity of the LEE results to accounting for the uncertainties of various inputs. Our results suggest that increasing the ensemble size seems to be beneficial up to a few hundred ensemble members, and localization is necessary even when the system is implemented with large ensembles. Accounting for uncertainties is more beneficial than increasing the ensemble size, even though it may introduce non-Gaussian features that limit the performance of our EnKF-based assimilation system. Detailed analyses and discussions of these findings are provided in the remainder of this study, which is organized as follows. Section 2 describes the RS EnKF DA system, including the ocean model. Section 3 presents the setup of the assimilation experiments, including various observational datasets used for validation. The ensemble DA experiment results are analyzed and discussed in Section 4. A summary of the main findings and a discussion on future research conclude the work in Section 5.

2. The Ensemble Ocean Data Assimilation System of the Red Sea

The Red Sea ensemble DA system is based on the Massachusetts Institute of Technology Ocean general circulation model (MITgcm; *Marshall et al.,* 1997) and an ensemble adjustment Kalman Filter (EAKF) available from the Data Assimilation Research Testbed (DART), the MITgcm-DART

[*Hoteit et al.,* 2013; Gopalakrishnan et al., 2019; *Toye et al.,* 2018, 2020, 2021; and *Sanikommu et al.,* 2020, 2023]. The MITgcm-DART configuration used here is basically the same as that of *Sanikommu et al.,* (2020), including the initial conditions, the model physics dictionary (described below), the atmospheric forcing ensemble, assimilated observations, localization, etc. The outputs of this assimilation system were validated against independent in-situ and satellite observations and were found to provide robust estimates of the Red Sea state [*Sanikommu et al.,* 2020; *Toye et al.,* 2020], significantly better than those available from the widely used global ocean reanalyses [*Sanikommu et al.,* 2023]. Below we provide a brief overview of the system components.

The MITgcm was implemented on a spherical polar grid covering the entire RS domain, including the Gulfs of Suez and Aqaba, and a part of the Gulf of Aden where an open boundary connects it to the Arabian Sea [e.g., *Krokos et al.,* 2019, 2022; *Sanikommu et al.,* 2020; and *Zhan et al.,* 2019, 2022]. The open boundary conditions for temperature, salinity, and horizontal velocity are prescribed daily from the 8km-resolution Global Ocean Reanalysis and Simulation data [GLORYS; *Parent et al.,* 2003]. The Red Sea MITgcm uses a direct space-time 3rd order scheme for tracer advection, harmonic viscosity with the coefficients of 30 m^2/s in the horizontal and $7x10^{-4}$ m^2/s in the vertical direction, implicit horizontal diffusion for both temperature and salinity, and the K-Profile Parameterization (KPP) scheme [*Large et al.,* 1994] for vertical mixing with a vertical diffusion coefficient of 10^{-5} m^2/s for both temperature and salinity. The model was spin up for 31 years starting from 1979 to 2010 using the 75 km resolution European Center for Medium Range Weather Forecast (ECMWF) reanalysis of atmospheric surface fluxes of radiation, momentum, and freshwater sampled every 6 hours [*Dee et al.,* 2011]. The MITgcm outputs have been extensively validated for the RS by earlier studies [e.g. *Yao et al.,* 2014a, 2014b; *Zhan et al.,* 2018; *Toye et al.,* 2017; *Gittings et al.,* 2019; *Krokos et al.,* 2022].

The EAKF is a deterministic square-root filter [*Anderson,* 2001; *Hoteit et al.,* 2015]. It is used here to assimilate three types of observations every 3 days, with a localization radius of 300 km. The observations include SST data extracted from a level-4 in-situ and advanced very high-resolution radiometer infrared satellite SST blended daily product available at 25 km resolution [*Reynolds et al.,* 2007]), along-track satellite level-3 merged altimeter filtered sea level anomalies (SLA), corrected for dynamic atmospheric, ocean tide, and long wavelength errors, from Copernicus Marine Environment Monitoring Service (CMEMS; *Pujol et al.,* 2018), and in-situ temperature and salinity profiles available from *Good et al.* (2013). Observational errors are assumed uncorrelated. Temporally static and spatially homogeneous observational error variance values of $(0.04\ \mathrm{m})^2$, $(0.5°\mathrm{C})^2$, and $(0.3\mathrm{psu})^2$ are prescribed for the satellite along-track SLA, and the in-situ T and S, respectively, in accordance with the suggested ranges of in-situ observational errors by earlier assimilation studies [e.g., *Richman et al.,* 2005; *Forget and Wunsch*, 2007; *Oke and Sakov*, 2008; *Karspeck,* 2016]. The specified observational error variances for SST vary between $(0.1°\mathrm{C})^2$ and $(0.6°\mathrm{C})^2$ in accordance with those of the level-4 gridded SST product of *Reynolds et al.,* (2007).

The MITgcm-DART is implemented with ROCOTO scheduler to facilitate the inclusion of uncertainties from various inputs, thereby avoiding the need of any inflation method [*Sanikommu et al.,* 2020]. Ocean hindcasts from the aforementioned model spin up are used to generate an ensemble of initial conditions to initialize the EAKF. To account for uncertainties in the atmospheric forcing fields, the MITgcm was forced with 6 hourly, 50km-resolution, 50-member ECMWF atmospheric ensemble available from The Observing System Research and Predictability Experiment Interactive Grand Global Ensemble project (TIGGE; *Bougeault et al.,* 2010; *Buizza,* 2014). Uncertainties in internal physics are accounted for by integrating each ensemble forecast model run with a set of model physics randomly selected from a predefined dictionary of model physics (MPD), specifically a model run with a certain set of model physics in a given cycle is integrated with a different set of model physics in the next cycle. The MPD consists of various vertical and horizontal mixing schemes, and different diffusion and viscosity parameters and is described in details in *Sanikommu et al.,* (2020).

3. Data Assimilation Experiments

Ten one-year long assimilation experiments are conducted starting from January 1st, 2011 to systematically investigate the behavior of the Red Sea MITgcm-DART with respect to the ensemble size and localization radius, as outlined in Table 1. The *Fexp* is a free-run in which the model was integrated with the mean of the ECMWF atmospheric ensemble. *A50exp_loc* is a 50-member assimilation experiment accounting for uncertainties in the initial conditions, internal physics, and atmospheric forcing. *A100exp_loc*, *A250exp_loc*, and *A500exp_loc* are the same as *A50exp_loc* with ensembles of 100, 250, and 500 members, respectively. Localization was used in all these experiments with a radius of 300km. *A500*exp and *A5000exp* respectively use 500 and 5000 ensemble members and are configured here without localization. The initial ensembles for all these experiments are selected from the January months of a long model hindcast run (here 20 years). The atmospheric forcing ensembles are randomly sampled from a Gaussian distribution with mean and covariance matching those of the 50-member ECMWF ensemble, using

$$X_s = \bar{X} + \frac{1}{\sqrt{k-1}} C_k u_{k\times s};$$

where k is the size of the original ensemble (50), s is the size of the intended ensemble, $\bar{X}$ is the matrix of the mean of the original ensemble, C_k is the matrix of atmospheric ensemble members anomalies, $u_{k\times s}$ is the matrix of perturbations sampled from Gaussian (0,1), and X_s is the intended ensemble.

Four more experiments, *Atm_A500exp_loc* and *Atm_A500exp,* and *Init_A500exp_loc* and *Init_A500exp* were conducted to assess the impact of different sources of uncertainties (Table 1). *Atm_A500exp_loc* and *Atm_A500exp* are the same as those of *A500exp_loc* and *A500exp* except that the uncertainties in the model physics are not accounted for. Similarly, *Init_A500exp_loc* and *Init_A500exp* are the same as those of *A500exp_loc* and *A500exp* except that the uncertainties in both model physics and atmospheric forcing are not accounted for, and uses an inflation factor of a value of 1.1 instead.

In addition to the aforementioned one-year long experiments, we conducted five single-DA EAKF runs on October 1st, 2011, to specifically investigate the influence of long-range correlations in a forecast ensemble derived from *A500exp*. These experiments, namely *SSHrun*, *SSTrun*, *SSTnrsrun*, *SSTsrsrun*, and *SSTgoarun*, assimilated a subset of observations as described in Table 2. These runs were designed to provide insights into the impact of long-range correlations on the assimilation process and its subsequent effects on the analyzed variables.

3.1. Validation Datasets

Various satellite and in-situ observations are used to evaluate the MITgcm-DART state estimates of the RS. The daily-averaged forecasts of SST and SSH are compared against the merged satellite level-3 observations from Group for High-Resolution Sea Surface Temperature (GHRSST; *EUMETSAT*, 2008) available at 5 km-resolution and merged along-track level-3 altimeter observations of SSH from CMEMS [*Pujol et al.,* 2018] available at 12 km-resolution, respectively. In the subsurface, the RS state estimates are evaluated against Conductivity Temperature and Depth (CTD) observations of temperature and salinity profiles that have been collected between 15th September and 8th October, 2011. This dataset includes 206 profiles sampled by a joint King Abdullah University of Science and Technology (KAUST) and Woods Hole Oceanography

Institute (WHOI) cruise along the eastern RS, with a horizontal spacing of 10 km [*Zhai et al.,* 2015; hereafter KAUST/WHOI summer cruise]. These observations are not assimilated in any of the experiments and are therefore considered independent observations for validation.

4. Results and Discussion

In this section, we first address the computational costs associated with the 5000-member ensemble experiment in comparison to a standard 50-member ensemble experiment. This comparison helps us gain a better understanding of the scale of the computational challenge we dealt with.

Next, we compare the outputs of all the large ensemble experiments with satellite and independent in-situ observations. This comparative analysis enables us to assess the system's sensitivity to ensemble size and localization. By examining how the assimilation experiments perform under different configurations, we can identify the impact of these factors on the quality of the assimilated data.

Additionally, we investigate the system's sensitivity to inputs stochasticity. This analysis allows us to understand how uncertainties in the initial conditions, atmospheric forcing and internal physics affect the assimilation results when large ensembles are used.

By considering these various factors, we aim to delve into the underlying reasons behind the observed improvements and degradations across different experimental settings. This analysis will provide valuable insights into the behavior of the assimilation system and contribute to a comprehensive understanding of its performance.

4.1. Computational costs

KAUST is home to SHAHEEN-II [*Hadri et al.,* 2015], a Cray XC40 supercomputer with 6,174 dual-socket compute nodes, each powered by 32 core Intel Haswell processors clocked at 2.3 GHz. Each node is equipped with 128GB of DDR4 memory running at 2300MHz. With a total of 197,568 processor cores and 790TB of aggregate memory, SHAHEEN-II can achieve a performance of 7.2Pflop/s in double precision. All experiments were conducted on this supercomputer, utilizing ROCOTO, a workflow manager to streamline our processes. This allowed us to take full advantage of SHAHEEN-II's capabilities and achieve optimal results.

To evaluate the computational costs, we present a comparison in Table 3 between the 5000-member and standard 50-member ensemble experiments, focusing on the two primary components: model integration using MITgcm and assimilation step using DART. For each ensemble member, running the 4km-resolution MITgcm to generate three days of forecasts requires three minutes with 3 nodes. Concurrently running 50 MITgcm instances necessitates 150 nodes. The standard 50-member ensemble DART process utilizes 8 nodes and takes approximately 5 minutes to assimilate all observations. In contrast, the 5000-member ensemble DART occupies 590 nodes to meet its large internal memory demand and takes around 30 minutes to finish the task. However, running the entire 5000 MITgcm instances simultaneously exceeds the available resources of SHAHEEN-II, requiring 15,000 nodes. Consequently, the 5000-member ensemble experiments demand a considerable amount of time to complete a single assimilation cycle. The computational cost for the 5000-member ensemble experiment to complete one assimilation cycle reaches 33,440 core hours, approximately 130 times higher than that of the standard 50-member ensemble experiment (261 core hours). Furthermore, the large number of input/output (IO) operations required for the 5000-member ensemble experiments significantly increases the time needed to complete a full assimilation cycle. This, in turn, increases the overall computational cost and makes it challenging to conduct various sensitivity experiments at this ensemble size. As a result, achieving optimal results can be challenging, especially that Shaheen-II is in production mode with many other users and different applications and workloads.

4.2. Comparisons against observations

Figure 1 illustrates the time series of the root-mean-square difference (RSMD) between the assimilation experiments and satellite observations for (a) SST and (b) SSH. All assimilation experiments outperform the *Fexp* baseline, consistent with the findings of *Toye et al.* (2020). Incorporating localization in the assimilation process improves the estimation of SST and SSH, with a slight enhancement observed as the ensemble size increases. However, there is a saturation point (at about 250 members) beyond which further increases in ensemble size yield marginal improvements, in agreement with Toye et al. (2020). Strikingly, not applying localization significantly degrades the assimilation solution, even with a 5000-member ensemble. This is especially notable in SSH, where the accuracy loss exceeds 3cm.

In order to gain deeper insights into these results, we analyzed the latitude-wise differences in temperature and salinity profiles between the model and independent in-situ observations, as depicted in Figure 2. The assimilation experiments without localization experience substantial deteriorations, particularly in the upper 120 m layers, with temperature differences exceeding 1°C and salinity differences exceeding 0.4psu. However, the differences between ensemble DA runs with and without localization become negligible in the deeper layers. This can be attributed to the limited observational data coverage and limited influence of atmospheric forcing ensemble and perturbed internal physics in those deeper layers, resulting in minimal analysis increments and ensemble spread (see Figure 3) at depth.

To further investigate the aforementioned results, we conducted an analysis of the ocean state estimates from various assimilation experiments: *Atm_A500exp_loc*, *Atm_A500exp*, *Init_A500exp_loc*, and *Init_A500exp*. These experiments share a similar configuration to *A500exp_loc* and *A500exp*, with the main difference being the systematic reduction of ensemble stochasticity by limiting the sources of uncertainty. Figure 4 illustrates the temporal evolution of the difference in RMSD in SST and SSH between the assimilation experiments with and without localization, compared against satellite observations. Positive values indicate improved performance with localization. In Figure 5, we focus on salinity differences between the model and in-situ observations, specifically for the *Atm_A500exp_loc*, *Atm_A500exp*, *Init_A500exp_loc*, and *Init_A500exp* experiments. The assimilation results of *Atm_A500exp_loc* and *Atm_A500exp* demonstrate significant deterioration when localization is not applied, particularly evident in SSH (Figure 4b) and salinity (Figures 5c-d), consistent with the findings discussed earlier for *A500exp_loc* and *A500exp*. Conversely, the results of *Init_A500exp_loc* and *Init_A500exp* do not exhibit a consistent pattern across the variables. SSTs degrade without localization, while SSH (Figure 4b) and salinity (comparing Figure 5a and 5b) show improvements. However, the differences between observations and estimated salinity in both *Init_A500exp_loc* and *Init_A500exp* (Figures 5a-b) are considerable, raising concerns about the suitability of employing EnKF DA systems that neglect important sources of uncertainty. It is worth mentioning that the ocean state estimates from the DA system are significantly improved when uncertainties in atmospheric forcing are accounted for by integrating the ocean model with an ensemble of atmospheric forcing (comparing Figures 5a-b with Figures 5c-d). Furthermore, considering uncertainties in internal physics (Figures 5c-d vs Figures 2l-m) further enhances the quality of system outputs, albeit to a lesser extent, in line with the findings reported by *Sanikommu et al.* (2020).

The deterioration in the assimilation system performance after the removal of localization, even at the largest ensemble size, indicate that there are potential sources of sampling errors other than the size of the ensemble.

4.3. Long-range ensemble correlations

To assess the deteriorations associated with the removal of localization in the assimilation experiments, we analyzed the long-range correlations in the forecast ensembles of different assimilation experiments. Figures 6a-d depict the spatial cross-correlations on October 1st, 2011, between sea surface salinity (SSS) at a location in the northern Red Sea (NRS) and sea surface temperature (SST) in the rest of the domain for *A50exp_loc*, *A500exp_loc*, *A500exp*, and *A5000exp*. Figures 6e-h show the same cross-correlations for another location in the southern Red Sea (SRS). These locations were selected based on prominent salinity differences observed in the assimilation experiments without localization (Figure 2m, 2n, and 5d). Overall, increasing the ensemble size from 50 to 500 smoothens the long-range correlations in the ensemble DA experiments with localization, consistent with previous studies on localization-related error covariance inflation. The differences in correlations between A500exp_loc and A500exp are notable, despite being driven by the same perturbed physics and atmospheric forcing. All ensemble data assimilation experiments indicate the presence of long-range correlations, even with 5000 ensemble members. However, some of these correlations appear to be spurious. For example, SSS in the NRS exhibits a strong correlation with SST in the Gulf of Aden (GoA) while showing weaker local correlations (Figure 6d). Similarly, SSS in the SRS strongly correlates with isolated patches of SST in the NRS, with less significant local correlations (Figure 6h).

To investigate whether these spurious long-range correlations have the potential to generate predominant increments, we conducted various single-data assimilation runs using the forecast ensemble of *A500exp* on October 1st, 2011, and examined the resulting analysis increments. Comparing the results obtained from *A500exp* (Figure 7a), *SSHrun* (Figure 7b), and *SSTrun* (Figure 7c), we observed that the salinity increments are primarily influenced by the assimilation of SST observations rather than SSH observations. For example, *A500exp* shows strong negative salinity analysis increments (up to 0.2psu) in the northeastern parts of the RS and moderate positive salinity analysis increments in the western parts of the RS and the GoA. The spatial patterns and magnitudes of salinity increments closely resemble those obtained from *SSTrun* compared to *SSHrun*. In Figure 7d and 7e, we present salinity increments obtained from the assimilation of SST observations at two different locations: one in the northern RS and another in the GoA. The results from these runs indicate that the salinity increments are prominent not only in the nearby region of the SST observation location but also in farther regions. For instance, *SSTnrsrun* yields large salinity increments in the GoA, while *SSTsrsrun* generates larger salinity increments in the NRS. These findings clearly indicate that the spurious long-range cross-correlations indeed have the potential to generate predominant increments, thereby confirming that the presence of these spurious long-range correlations is behind the deteriorations observed in ensemble DA runs without localization.

4.3.1 The sources of spurious long-range ensemble correlations

Previous studies have noted that the perturbed atmospheric forcing ensemble from ECMWF contains non-Gaussianity and long-range correlations due to its low rank [*Bertossa et al.,* 2021]. When such atmospheric ensemble is used to force complex nonlinear systems like the ocean model, the non-Gaussianity is expected to amplify [*Miyoshi et al.,* 2014]. Additionally, the nonlinear nature of the model itself can generate non-Gaussianity even when perturbing linear model physics parameters [*Sura and Hannachi*, 2015].

To assess any systematic increase in the long-range correlations associated with atmospheric ensemble forcing and non-linearity of the perturbed model physics, we examined the forecast ensembles of *Init_A500exp*, *Atm_A500exp,* and *A500exp* (Figure 8, 6g and Table 4). Figure 8 shows cross-correlations between SSS at a location in the SRS and SST in the rest of the domain for the ensemble of (a) *Init_A500exp*, and (b) *Atm_A500exp*. In the ensemble of *Init_A500exp*, the correlations are limited to the local neighborhoods, similar to previous findings of *Toye et al.* (2018)

with 1000-member ensemble. On the other hand, the ensemble of *Atm_A500exp* exhibits long-range correlations, with SSS in the SRS correlating more strongly with SST in the NRS than with its closer neighborhoods (Figure 8b). The long-range correlations in the ensemble of *A500exp* are even more prominent (Figure 6g).

To examine non-Gaussianity, we analyzed skewness (s) and kurtosis (k) of the Probability Density Function (PDF) of the forecast ensemble from *Init_A500exp*, *Atm_A500exp*, and *A500exp*. Table 4 summarizes the percent occurrence of non-Gaussianity in the whole domain for different variables based on the ensembles of *Init_A500exp*, *Atm_A500exp*, and *A500exp*. The PDF is considered non-Gaussian when $|s| > 0.3$ or $|k| > 0.6$. The presence of non-Gaussianity is relatively low for sea surface height (SSH) across all experiments. However, for SST and SSS, non-Gaussianity is notable. It is around 14% in *Init_A500exp*, but increases to 38% in *Atm_A500exp* and 86% in *A500exp*. A similar systematic increase in non-Gaussianity is observed for SST as well.

To illustrate how the bimodality and long-range correlations in the forecast ensemble caused deteriorations independently from each other, we present the scatter between SST and SSS (Figure 9). The scatter plot between SST in the GoA and SSS in the NRS (Figure 9a) suggests a prominent bimodality in the SST ensemble, with distinct warm and cold groups. The correlations computed for each of these groups within the same ensemble vary significantly from each other, indicating that the bimodal correlations cannot be well described by linear correlations commonly used in Kalman-like updates [*Anderson et al.,* 2001; *Hoteit et al.,* 2015]. In contrast, the long-range correlation between SST in the NRS and SSS in the SRS (Figure 9b) is strong without clear signs of bimodality. Note that the magnitude of the long-range correlation is stronger than the local correlations between SST and SSS in the SRS (Figure 6h). This result suggests that there are instances/places where degradations were resulted only from the spurious long-rage correlations too.

Be it influenced by spurious features arising from the size-constrained atmospheric forcing ensemble or the non-Gaussian effects introduced by stochastic perturbations, our EnKF-based assimilation system's utilization of long-range correlations may prove unreliable. Consequently, the necessity for localization remains, even when the system is implemented large ensembles. Ensemble data assimilation systems that account for uncertainties in atmospheric forcing and internal physics have been shown to be beneficial. To further enhance performance of the DA systems, more advanced assimilation schemes that account for non-Gaussianity and mitigate long-range correlations should be incorporated.

5. Summary and future directions

This study focused on conducting large ensemble experiments using a real ocean data assimilation system implemented in the Red Sea. The ensemble sizes ranged from 50 to 5000 members, which significantly increased the computational cost by 130 times compared to smaller ensembles. By utilizing these large ensembles, we aimed to gain insights into the validity of the Gaussian assumption in the Ensemble Kalman Filter (EnKF) and evaluate the relative importance of addressing uncertainties in different inputs versus mitigating long-range correlations through the utilization of a large ensemble. Another objective was to examine the system's sensitivity to ensemble size and the significance of localization in a large ensemble context. Through these investigations, we aimed to enhance our understanding of the behavior and performance of the data assimilation system under different conditions.

Our results indicate that increasing the ensemble size leads to improved ocean state estimates and reduced long-range correlations. However, we observed that the these improvements seize well below an ensemble size of 500 members. Localization continues to play a crucial role even in the context of large ensembles. We also observed that accounting for uncertainties in various inputs yields more benefits than simply increasing the ensemble size, despite the presence of amplified long-range correlations and non-Gaussianity.

The presence of non-Gaussianity and long-range correlations even in large ensembles underscores the ongoing need for localization. Interestingly, our results are not in full agreement with previous studies [*Kondo and Miyoshi*, 2016; *Toye et al.,* 2018], which suggested that large ensembles lead to smoother correlations and improved assimilation solutions while being less reliant on localization. Thus, the findings presented in this study can change scientific perspectives regarding the benefit of large ensembles in data assimilation.

Based on the improvements in the ensemble DA experiments achieved after incorporating uncertainties in atmospheric forcing and internal physics, we envision that their capabilities could be further enhanced by employing more sophisticated non-Gaussian data assimilation schemes. Increasing the rank of the atmospheric forcing ensemble and implementing non-Gaussian ensemble filters such as ensemble Gaussian Mixture filters (*Anderson* 2010; *Hoteit et al.,* 2008, 2012, 2015; *Fletcher et al.,* 2023) are potential avenues to explore. As computational power advances, increasing the rank of the atmospheric and oceanic ensembles will become more affordable in the future. KAUST is acquiring its new supercomputing system SHAHEEN-III which is set to be 20 times faster than the current supercomputer SHAHEEN-II thanks to its more efficient computing resources. This will offer us new perspectives to develop and test non-Gaussian ensemble data assimilation schemes for improved ocean analyses and forecasting.

Acknowledgments

This work was funded by the Office of Vice President of Research at King Abdullah University of Science and Technology (KAUST) under the Virtual Red Sea Initiative (Grant #REP/1/3268-01-01), and the Saudi ARAMCO Marine Environmental Centre at KAUST. All the model experiments were run on the KAUST supercomputing facility, SHAHEEN-II. The support of the KAUST supercomputing team is highly acknowledged.

Data Availability statement

Data used for this study will be made available in a global archive after the acceptance of the paper.

References

Anderson, J.L., 2010. A non-Gaussian ensemble filter update for data assimilation. Mon Weather Rev 138. https://doi.org/10.1175/2010MWR3253.1

Anderson, J.L., 2001. An ensemble adjustment Kalman filter for data assimilation. Mon Weather Rev 129. https://doi.org/10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2

Bishop, C. H., & Hodyss, D. (2007). Flow-adaptive moderation of spurious ensemble correlations and its use in ensemble-based data assimilation. Quarterly Journal of the Royal Meteorological Society, 133(629 B). https://doi.org/10.1002/qj.169

Baduru, B., Paul, B., Banerjee, D.S., Sanikommu, S., Paul, A., 2019. Ensemble based regional ocean data assimilation system for the Indian Ocean: Implementation and evaluation. Ocean Model (Oxf) 143. https://doi.org/10.1016/j.ocemod.2019.101470

Bertossa, C., Hitchcock, P., Degaetano, A., Plougonven, R., 2021. Bimodality in Ensemble Forecasts of 2-Meter Temperature: Identification. Weather and Climate Dynamics 1–23. https://doi.org/10.5194/wcd-2021-33

Bougeault, P., Toth, Z., Bishop, C., Brown, B., Burridge, D., de Chen, H., Ebert, B., Fuentes, M., Hamill, T.M., Mylne, K., Nicolau, J., Paccagnella, T., Park, Y.Y., Parsons, D., Raoult, B., Schuster, D., Dias, P.S., Swinbank, R., Takeuchi, Y., Tennant, W., Wilson, L., Worley, S., 2010. The thorpex interactive grand global ensemble. Bull Am Meteorol Soc 91. https://doi.org/10.1175/2010BAMS2853.1

Bouttier, F., and Courtier, P., 2002. Data assimilation concepts and methods March 1999. Training.

Bouttier, F., and Courtier, P., 1999. Data assimilation concepts and methods, in: ECMWF Meteorological Training Course Lecture Series.

Bowler, N.E., Arribas, A., Mylne, K.R., Robertson, K.B., Beare, S.E., 2008. The MOGREPS short-range ensemble prediction system. Quarterly Journal of the Royal Meteorological Society 134. https://doi.org/10.1002/qj.234

Buizza, R., 2014. The TIGGE global, medium-range ensembles, ECMWF Technical Memoranda.

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E. v., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., Mcnally, A.P., Monge-Sanz, B.M., Morcrette, J.J., Park, B.K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.N., Vitart, F., 2011. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. Quarterly Journal of the Royal Meteorological Society 137. https://doi.org/10.1002/qj.828

Derber, J., and Bouttier, F., 1999. A reformulation of the background error covariance in the ECMWF global data assimilation system. Tellus, Series A: Dynamic Meteorology and Oceanography 51. https://doi.org/10.3402/tellusa.v51i2.12316

Edwards, C.A., Moore, A.M., Hoteit, I., and Cornuelle, B.D. (2015). Regional ocean data assimilation. Annual Review of Marine Science, 7, 21–42, doi:10.1146/annurev-marine-010814-015821.

EUMETSAT/OSI-SAF, GHRSST level 3P global subskin sea surface temperature from the Advanced Very High Resolution Radiometer (AVHRR) on the MetOp-A satellite, 2008, http://dx.doi.org/10.5067/GHGMT-3PE01, Ver. 1. PO.DAAC, CA, USA. Dataset accessed 2018.10.01.

Forget, G., Wunsch, C., 2007. Estimated global hydrographic variability. J Phys Oceanogr 37. https://doi.org/10.1175/JPO3072.1

Fletcher, S. J., M. Zupanski, M. R. Goodliff, A. J. Kliewer, A. S. Jones, J. M. Forsythe, T.-C. Wu, M. J. Hossen and S. Van Loon, 2023. Lognormal and Mixed Gaussian-lognormal based Kalman Filters. Monthly Weather Review, 151, 761-774. https://doi.org/10.1175/MWR-D-22-0072.1

Fujita, T., Stensrud, D.J., Dowell, D.C., 2007. Surface data assimilation using an ensemble Kalman filter approach with initial condition and model physics uncertainties. Mon Weather Rev 135. https://doi.org/10.1175/MWR3391.1

Gittings, J.A., Raitsos, D.E., Kheireddine, M., Racault, M.F., Claustre, H., Hoteit, I., 2019. Evaluating tropical phytoplankton phenology metrics using contemporary tools. Sci Rep 9. https://doi.org/10.1038/s41598-018-37370-4

Gopalakrishnan, G., Hoteit, I., Cornuelle, B.D. and Rudnick, D.L., 2019. Comparison of 4DVAR and EnKF state estimates and forecasts in the Gulf of Mexico. Quarterly Journal of the Royal Meteorological Society, 145(721), pp.1354-1376. https://doi.org/10.1002/qj.349

Good, S.A., Martin, M.J., Rayner, N.A., 2013. EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. J Geophys Res Oceans 118. https://doi.org/10.1002/2013JC009067

Hadri, B., Kortas, S., Feki, S., Khurram, R., and Newby, G., 2015. Overview of the KAUST's Cray X40 system–Shaheen II. Proc. 2015 Cray User Group, Chicago, IL, Cray User Group, 7 pp., https://cug.org/proceedings/cug2015_proceedings/includes/files/pap129.pdf.

He, J., Zhang, F., Chen, X., Bao, X., Chen, D., Kim, H.M., Lai, H.W., Leung, L.R., Ma, X., Meng, Z., Ou, T., Xiao, Z., Yang, E.G., Yang, K., 2019. Development and Evaluation of an Ensemble-Based Data Assimilation System for Regional Reanalysis Over the Tibetan Plateau and Surrounding Regions. J Adv Model Earth Syst 11. https://doi.org/10.1029/2019MS001665

Hoteit, I., Pham, D.T., Triantafyllou, G. and Korres, G., 2008. A new approximate solution of the optimal nonlinear filter for data assimilation in meteorology and oceanography. Monthly Weather Review, 136(1), pp.317-334. https://doi.org/10.1175/2007MWR1927.1

Hoteit, I., Luo, X. and Pham, D.T., 2012. Particle Kalman filtering: A nonlinear Bayesian framework for ensemble Kalman filters. Monthly weather review, 140(2), pp.528-542. https://doi.org/10.1175/2011MWR3640.1

Hoteit, I., Hoar, T., Gopalakrishnan, G., Collins, N., Anderson, J., Cornuelle, B., Köhl, A., Heimbach, P., 2013. A MITgcm/DART ensemble analysis and prediction system with application to the Gulf of Mexico. Dynamics of Atmospheres and Oceans 63. https://doi.org/10.1016/j.dynatmoce.2013.03.002

Hoteit, I., Pham, D.T., Gharamti, M.E. and Luo, X., 2015. Mitigating observation perturbation sampling errors in the stochastic EnKF. Monthly Weather Review, 143(7), pp.2918-2936. https://doi.org/10.1175/MWR-D-14-00088.1

Hoteit, I., Luo, X., Bocquet, M., Köhl, A., Ait-El-Fquih, B., 2018. Data Assimilation in Oceanography: Current Status and New Directions, in: New Frontiers in Operational Oceanography. GODAE OceanView. https://doi.org/10.17125/gov2018.ch17

Houtekamer, P. L., and Mitchell, H. L. (1998). Data Assimilation Using an Ensemble Kalman Filter Technique. Monthly Weather Review, 126(3), 796–811. https://doi.org/10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2

Houtekamer, P. L., & Mitchell, H. L. (2001). A sequential ensemble Kalman filter for atmospheric data assimilation. Monthly Weather Review, 129(1). https://doi.org/10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2

Houtekamer, P.L., Mitchell, H.L., Deng, X., 2009. Model error representation in an operational ensemble Kalman filter. Mon Weather Rev 137. https://doi.org/10.1175/2008MWR2737.1

Houtekamer, P.L., Zhang, F., 2016. Review of the ensemble Kalman filter for atmospheric data assimilation. Mon Weather Rev. https://doi.org/10.1175/MWR-D-15-0440.1

Hunt, B.R., Kostelich, E.J., Szunyogh, I., 2007. Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. Physica D 230, 112–126. https://doi.org/10.1016/j.physd.2006.11.008

Jacques, D., Zawadzki, I., 2015. The impacts of representing the correlation of errors in radar data assimilation. Part II: Model output as background estimates. Mon Weather Rev 143, 2637–2656. https://doi.org/10.1175/MWR-D-14-00243.1

Karspeck, A.R., 2016. An ensemble approach for the estimation of observational error illustrated for a nominal 1° global ocean model. Mon Weather Rev 144, 1713–1728. https://doi.org/10.1175/MWR-D-14-00336.1

Kondo, K., Miyoshi, T., 2016. Impact of removing covariance localization in an ensemble Kalman Filter: Experiments with 10 240 members using an intermediate AGCM. Mon Weather Rev 144, 4849–4865. https://doi.org/10.1175/MWR-D-15-0388.1

Krokos, G., Cerovečki, I., Papadopoulos, V.P., Hendershott, M.C., Hoteit, I., 2022. Processes Governing the Seasonal Evolution of Mixed Layers in the Red Sea. J Geophys Res Oceans 127. https://doi.org/10.1029/2021JC017369

Krokos, G., Papadopoulos, V.P., Sofianos, S.S., Ombao, H., Dybczak, P., Hoteit, I., 2019. Natural Climate Oscillations may Counteract Red Sea Warming Over the Coming Decades. Geophys Res Lett 46. https://doi.org/10.1029/2018GL081397

Kwon, K.M., Choi, B.J., Lee, S.H., Kim, Y.H., Seo, G.H., Cho, Y.K., 2016. Effect of model error representation in the Yellow and East China Sea modeling system based on the ensemble Kalman filter. Ocean Dyn 66. https://doi.org/10.1007/s10236-015-0909-8

Large, W.G., Mcwilliams, J.C., Doney, S.C., 1994. OCEANIC VERTICAL MIXING: A REVIEW AND A MODEL WITH A NONLOCAL BOUNDARY LAYER PARAMETERIZATION.

Lee, Y., Majda, A. J., & Qi, D. (2017). Preventing Catastrophic Filter Divergence Using Adaptive Additive Inflation for Baroclinic Turbulence. Monthly Weather Review, 145(2). https://doi.org/10.1175/mwr-d-16-0121.1

Lei, L., Whitaker, J.S., 2017. Evaluating the trade-offs between ensemble size and ensemble resolution in an ensemble-variational data assimilation system. J Adv Model Earth Syst 9. https://doi.org/10.1002/2016MS000864

Leutbecher, M., 2019. Ensemble size: How suboptimal is less than infinity? Quarterly Journal of the Royal Meteorological Society 145, 107–128. https://doi.org/10.1002/qj.3387

Luo, X., Bhakta, T., & Nævdal, G. (2018). Correlation-based adaptive localization with applications to ensemble-based 4D-seismic history matching. SPE Journal, 23(2). https://doi.org/10.2118/185936-pa

Marshall, J., Adcroft, A., Hill, C., Perelman, L., Heisey, C., 1997. A finite-volume, incompressible navier stokes model for, studies of the ocean on parallel computers. J Geophys Res Oceans 102. https://doi.org/10.1029/96JC02775

Martin, M.J., Balmaseda, M., Bertino, L., Brasseur, P., Brassington, G., Cummings, J., Fujii, Y., Lea, D.J., Lellouche, J.M., Mogensen, K., Oke, P.R., Smith, G.C., Testut, C.E., Waagbø, G.A., Waters, J., Weaver, A.T., 2015. Status and future of data assimilation in operational oceanography. Journal of Operational Oceanography 8, s28–s48. https://doi.org/10.1080/1755876X.2015.1022055

Mitchell, H.L., Houtekamer, P.L., Pellerin, G., 2002. Ensemble size, balance, and model-error representation in an ensemble Kalman filter. Mon Weather Rev 130. https://doi.org/10.1175/1520-0493(2002)130<2791:ESBAME>2.0.CO;2

Miyoshi, T. (2011). The gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform Kalman filter. Monthly Weather Review, 139(5). https://doi.org/10.1175/2010MWR3570.1

Miyoshi, T., Kondo, K., Imamura, T., 2014. The 10,240-member ensemble Kalman filtering with an intermediate AGCM. Geophys Res Lett 41, 5264–5271. https://doi.org/10.1002/2014GL060863

Necker, T., Geiss, S., Weissmann, M., Ruiz, J., Miyoshi, T., Lien, G.Y., 2020a. A convective-scale 1,000-member ensemble simulation and potential applications. Quarterly Journal of the Royal Meteorological Society 146, 1423–1442. https://doi.org/10.1002/qj.3744

Necker, T., Weissmann, M., Ruckstuhl, Y., Anderson, J., Miyoshi, T., 2020b. Sampling error correction evaluated using a convective-scale 1000-member ensemble. Mon Weather Rev 148, 1229–1249. https://doi.org/10.1175/MWR-D-19-0154.1

Oke, P.R., Sakov, P., 2008. Representation error of oceanic observations for data assimilation. J Atmos Ocean Technol 25. https://doi.org/10.1175/2007JTECHO558.1

Parent, L., Testut, C.E., Brankart, J.M., Verron, J., Brasseur, P., Gourdeau, L., 2003. Comparative assimilation of Topex/Poseidon and ERS altimeter data and of TAO temperature data in the tropical Pacific ocean during 1994-1998, and the mean sea-surface height issue, in: Journal of Marine Systems. Elsevier, pp. 381–401. https://doi.org/10.1016/S0924-7963(03)00026-5

Penny, S.G., Behringer, D.W., Carton, J.A., Kalnay, E., 2015. A hybrid global ocean data assimilation system at NCEP. Mon Weather Rev 143, 4660–4677. https://doi.org/10.1175/MWR-D-14-00376.1

Pinardi, N., Bonazzi, A., Scoccimarro, E., Dobricic, S., Navarra, A., Ghiselli, A., Veronesi, P., 2008. Very large ensemble ocean forecasting experiment using the grid computing infrastructure. Bull Am Meteorol Soc 89, 799–804. https://doi.org/10.1175/2008BAMS2511.1

Pujol, M., Faug, Y., Date, A., 2018. Product user manual For Sea Level SLA products. Cmems.

Reynolds, R.W., Smith, T.M., Liu, C., Chelton, D.B., Casey, K.S., Schlax, M.G., 2007. Daily high-resolution-blended analyses for sea surface temperature. J Clim 20, 5473–5496. https://doi.org/10.1175/2007JCLI1824.1

Richman, J.G., Miller, R.N., Spitz, Y.H., 2005. Error estimates for assimilation of satellite sea surface temperature data in ocean climate models. Geophys Res Lett 32. https://doi.org/10.1029/2005GL023591

Sanikommu, S., Banerjee, 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 145. https://doi.org/10.1002/qj.3649

Sanikommu, S., Langodan, S., Dasari, H.P., Zhan, P., Krokos, G., Abualnaja, Y.O., Asfahani, K., Hoteit, I., 2023. Making the case for high resolution regional reanalyses: An example with the Red Sea. Bulletin of American Meteorological Society. E1241-E1264, https://doi.org/10.1175/BAMS-D-21-0287.1

Sanikommu, S., Paul, A., Sluka, T., Ravichandran, Muthalugu, Kalnay, E., 2017. The pre-Argo ocean reanalyses may be seriously affected by the spatial coverage of moored buoys. Sci Rep 7, 46685. https://doi.org/10.1038/srep46685

Sanikommu, S., Toye, H., Zhan, P., Langodan, S., Krokos, G., Knio, O., Hoteit, I., 2020a. Impact of Atmospheric and Model Physics Perturbations on a High-Resolution Ensemble Data Assimilation System of the Red Sea. J Geophys Res Oceans 125, e2019JC015611. https://doi.org/10.1029/2019JC015611

Subramanian, A.C., Hoteit, I., Cornuelle, B., Miller, A.J. and Song, H., 2012. Linear versus nonlinear filtering with scale-selective corrections for balanced dynamics in a simple atmospheric model. Journal of the atmospheric sciences, 69(11), pp.3405-3419. https://doi.org/10.1175/JAS-D-11-0332.1

Sura, P., Hannachi, A., 2015. Perspectives of non-gaussianity in atmospheric synoptic and low-frequency variability. J Clim 28. https://doi.org/10.1175/JCLI-D-14-00572.1

Sura, P., Newman, M., Penland, C., Sardeshmukh, P., 2005. Multiplicative noise and non-Gaussianity: A paradigm for atmospheric regimes? J Atmos Sci 62. https://doi.org/10.1175/JAS3408.1

Sura, P., Sardeshmukh, P.D., 2008. A global view of non-Gaussian SST variability. J Phys Oceanogr 38. https://doi.org/10.1175/2007JPO3761.1

Toye, H., Kortas, S., Zhan, P., Hoteit, I., 2018. A fault-tolerant HPC scheduler extension for large and operational ensemble data assimilation: Application to the Red Sea. J Comput Sci 27, 46–56. https://doi.org/10.1016/j.jocs.2018.04.018

Toye, H., Sanikommu, S., Raboudi, N.F., Hoteit, I., 2020. A hybrid ensemble adjustment Kalman filter based high-resolution data assimilation system for the Red Sea: Implementation and evaluation. Quarterly Journal of the Royal Meteorological Society 146. https://doi.org/10.1002/qj.3894

Toye, H., Zhan, P., Gopalakrishnan, G., Kartadikaria, A.R., Huang, H., Knio, O., Hoteit, I., 2017. Ensemble data assimilation in the Red Sea: sensitivity to ensemble selection and atmospheric forcing. Ocean Dyn 67, 915–933. https://doi.org/10.1007/s10236-017-1064-1

Toye, H., Zhan, P., Sana, F., Sanikommu, S., Raboudi, N., Hoteit, I., 2021. Adaptive ensemble optimal interpolation for efficient data assimilation in the red sea. J Comput Sci 51. https://doi.org/10.1016/j.jocs.2021.101317

Vandenbulcke, L., Barth, A., 2015. A stochastic operational forecasting system of the Black Sea: Technique and validation. Ocean Model (Oxf) 93. https://doi.org/10.1016/j.ocemod.2015.07.010

Wang, X., Lei, T., 2014. GSI-based four-dimensional ensemble-variational (4DEnsVar) data assimilation: Formulation and single-resolution experiments with real data for NCEP global forecast system. Mon Weather Rev 142. https://doi.org/10.1175/MWR-D-13-00303.1

Whitaker, J.S., Hamill, T.M., 2002. Ensemble data assimilation without perturbed observations. Mon Weather Rev 130. https://doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2

Whitaker, J. S., & Hamill, T. M. (2012). Evaluating methods to account for system errors in ensemble data assimilation. Monthly Weather Review, 140(9). https://doi.org/10.1175/MWR-D-11-00276.1

Yao, F., Hoteit, I., Pratt, L.J., Bower, A.S., Köhl, A., Gopalakrishnan, G., Rivas, D., 2014a. Seasonal overturning circulation in the Red Sea: 2. Winter circulation. J Geophys Res Oceans 119, 2263–2289. https://doi.org/10.1002/2013JC009331

Yao, F., Hoteit, I., Pratt, L.J., Bower, A.S., Zhai, P., Köhl, A., Gopalakrishnan, G., 2014b. Seasonal overturning circulation in the Red Sea: 1. Model validation and summer circulation. J Geophys Res Oceans 119, 2238–2262. https://doi.org/10.1002/2013JC009004

Zhai, P., Bower, A.S., Smethie, W.M., Pratt, L.J., 2015. Formation and spreading of Red Sea Outflow Water in the Red Sea. J Geophys Res Oceans 120. https://doi.org/10.1002/2015JC010751

Zhan, P., Gopalakrishnan, G., Subramanian, A.C., Guo, D., Hoteit, I., 2018. Sensitivity Studies of the Red Sea Eddies Using Adjoint Method. J Geophys Res Oceans 123, 8329–8345. https://doi.org/10.1029/2018JC014531

Zhan, P., Guo, D., Krokos, G., Dong, J., Duran, R., Hoteit, I., 2022. Submesoscale Processes in the Upper Red Sea. J Geophys Res Oceans 127. https://doi.org/10.1029/2021JC018015

Zhan, P., Krokos, G., Guo, D., Hoteit, I., 2019. Three-Dimensional Signature of the Red Sea Eddies and Eddy-Induced Transport. Geophys Res Lett 46. https://doi.org/10.1029/2018GL081387

Tables

Table 1. Summary of the experiments conducted. In the table "Unperturbed" refers to the default configuration adopted from the deterministic model. "Perturbed" model physics refers to the use of a time-varying ensemble of physics during the model integration of each ensemble member for forecasting.

Experiment	**Initial ensemble**	**Atmospheric Ensemble**	**Model physics**	**Assimilation**	**Localization**
Fexp	Unperturbed	Unperturbed	Unperturbed	NA	NA
A50exp_loc	50	50	Perturbed	Yes	300km
A100exp_loc	100	100	Perturbed	Yes	300km
A250exp_loc	250	250	Perturbed	Yes	300km
A500exp_loc	500	500	Perturbed	Yes	300km
A500exp	500	500	Perturbed	Yes	Not used
A5000exp	5000	5000	Perturbed	Yes	Not used
Atm_A500exp_loc	500	500	Unperturbed	Yes	300km
Atm_A500exp	500	500	Unperturbed	Yes	Not used
Init_A500exp_loc	500	Unperturbed	Unperturbed	Yes	300km
Init_A500exp	500	Unperturbed	Unperturbed	Yes	Not used

Table 2. Summary of the single-DA cylce EAKF runs conducted using the forecast ensemble of A500exp on 1st October, 2011. In the table, NRS, SRS, and GoA corresponds to observations in the entire northern RS (32°E-38°E & 24°N-30°N), southern RS (38°E-42°E & 12°N-18°N), and Gulf-of-Aden (42°E-50°E & 12°N-16°N), respectively.

Assimilation Experiment	**SSH**	**SST**	**T&S profiles**
SSHrun	Yes; whole domain	No	No
SSTrun	No	Yes; whole domain	No
SSTnrsrun	No	Yes; only in the NRS	No
SSTsrsrun	No	Yes; only in the SRS	No
SSTgoarun	No	Yes;only in the GoA	No

Table 3. Computational costs for one assimilation cycle in 50-member and 5000-member ensemble experiments, where assimilation is performed after integrating model for 3 days. The total core hours are calculated based on 32 cores per node.

Experiment	**Ensemble of Models**		**Assimilation**		**Total core hours**
	Nodes	**Wall time**	**Nodes**	**Wall time**	
A50exp_loc	150	3 minutes	8	5 minutes	261
A5000exp	15000	3 minutes	590	30 minutes	33440

Table 4. The percentage occurrence of non-Gaussianity resulted from different experiments in different surface variables.

	SST	**SSS**	**SSH**
Init_A500exp	26	14	4
Atm_A500exp	27	38	3
A500exp	59	86	4

Figures

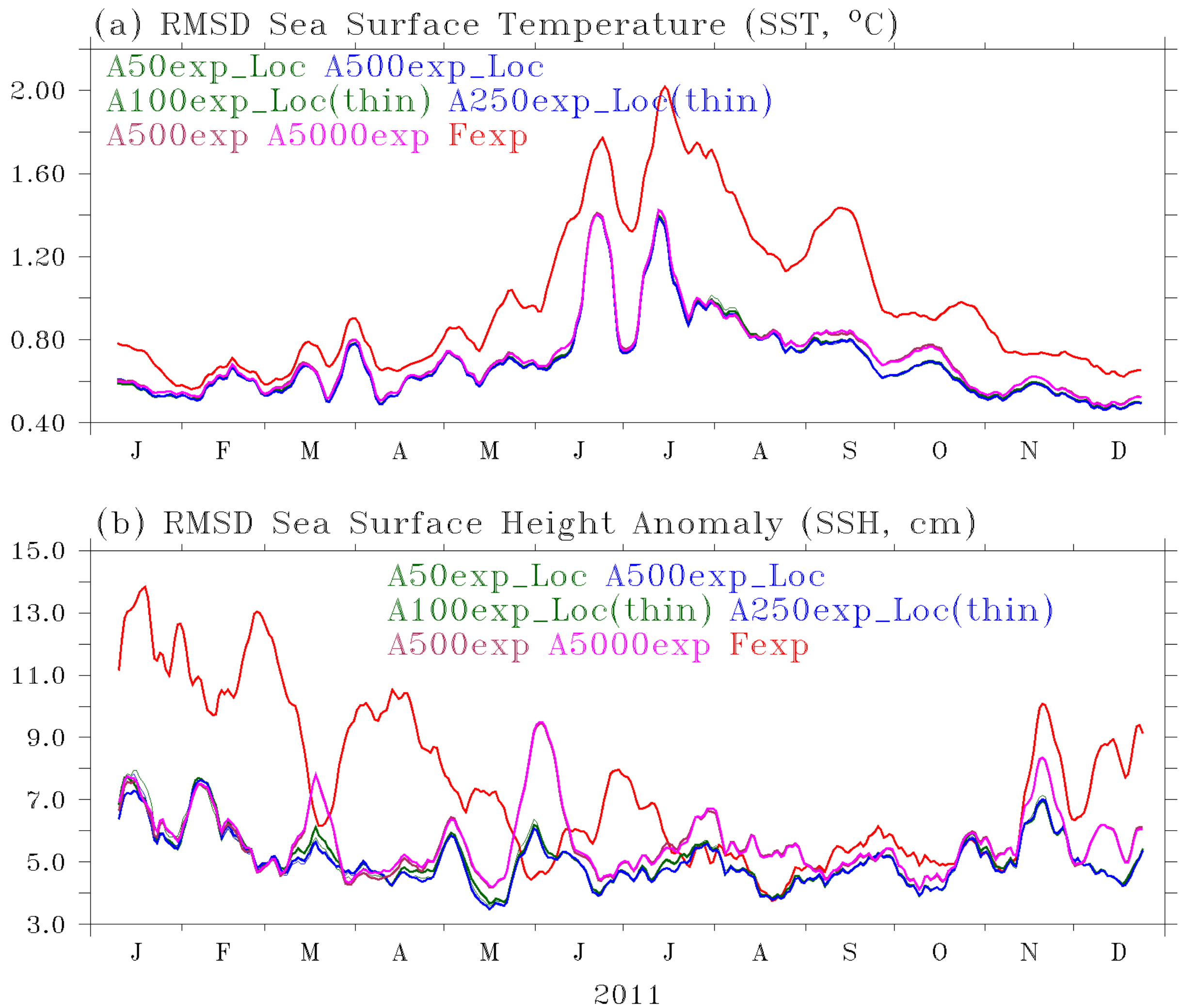

Figure 1. Time series of RMSD for daily averaged (a) SST (°C), and (b) SSH (cm) from *Fexp* (red), *A50exp_loc* (green), *A100exp_loc* (green thin line) *A250exp_loc* (blue thin line), *A500exp_loc* (blue), *A500exp* (magenta), *A5000exp* (pink). RMSDs of SST and SSH are computed by collocating the daily averaged model forecasts onto level-3 GHRSST, and level-3 altimeter observations, respectively. 10-day smoothing is applied to better visualize the differences amongst the reanalyses.

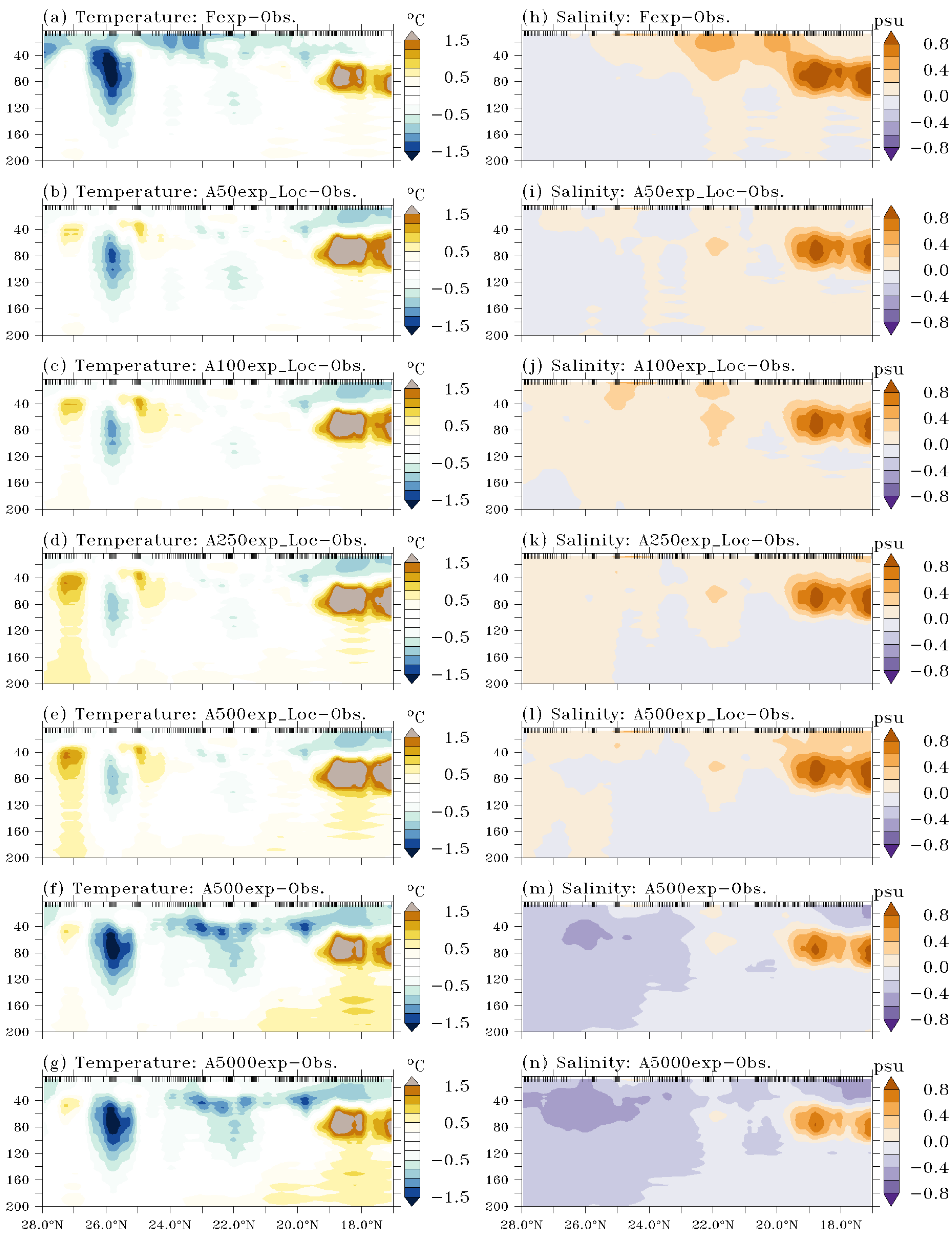

Figure 2. Collocated (in space and time) subsurface temperature (a) and salinity (b) differences between the model outputs and *in-situ* CTD observations collected during the KAUST/WHOI summer cruise conducted from 15th September – 8th October 2011. Panels a-b, c-d, e-f, g-h, i-j, k-l, and m-n show results for *Fexp, A50exp_loc, A100exp_loc, A250exp_loc, A500exp_loc, A500exp,* and *A5000exp*, respectively. Temperature and salinity observations are smoothed by 1° in the latitudinal direction and 10m in the vertical to emphasize subsurface features. Latitudes corresponding to observation locations are indicated as black vertical dashes at the top of each panel.

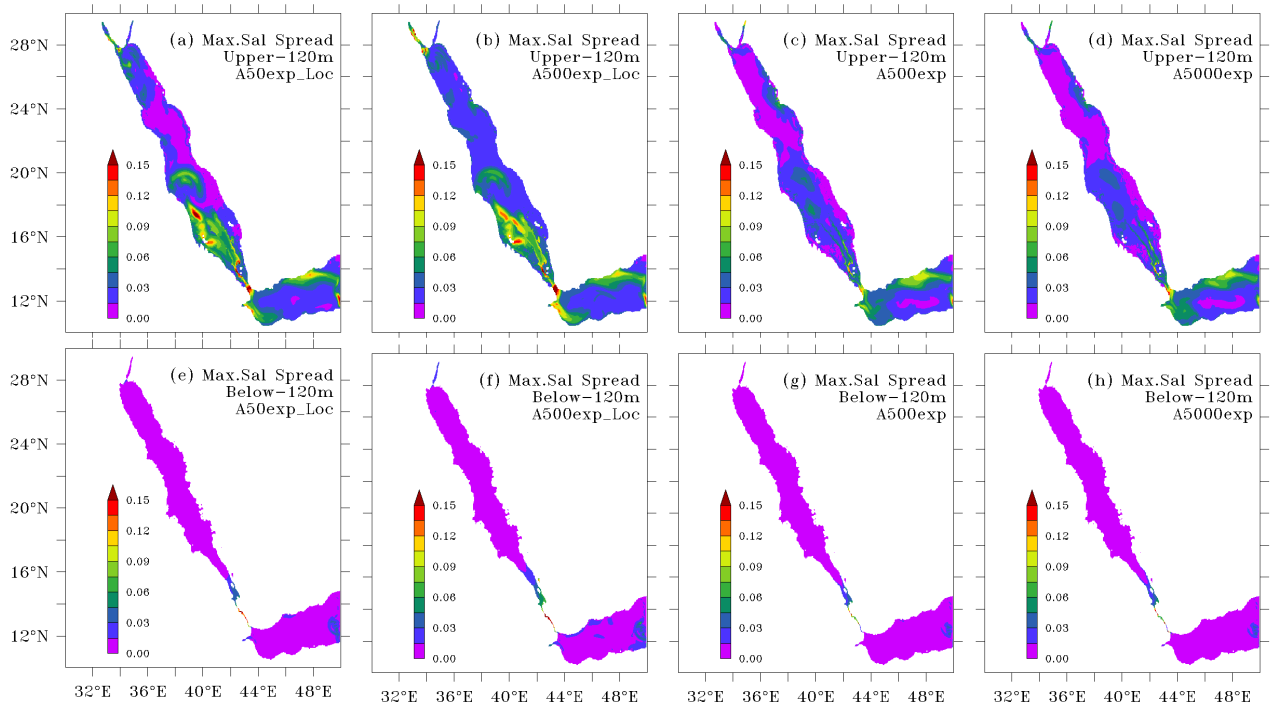

Figure 3. Maximum ensemble salinity spread in the (a-d) upper 120m and (e-h) below 120m. Results are shown for (a,e) *A50exp_loc*, (b,f) *A500exp_loc*, (c,g) *A500exp*, (d,h) *A5000exp* on 1st October, 2011.

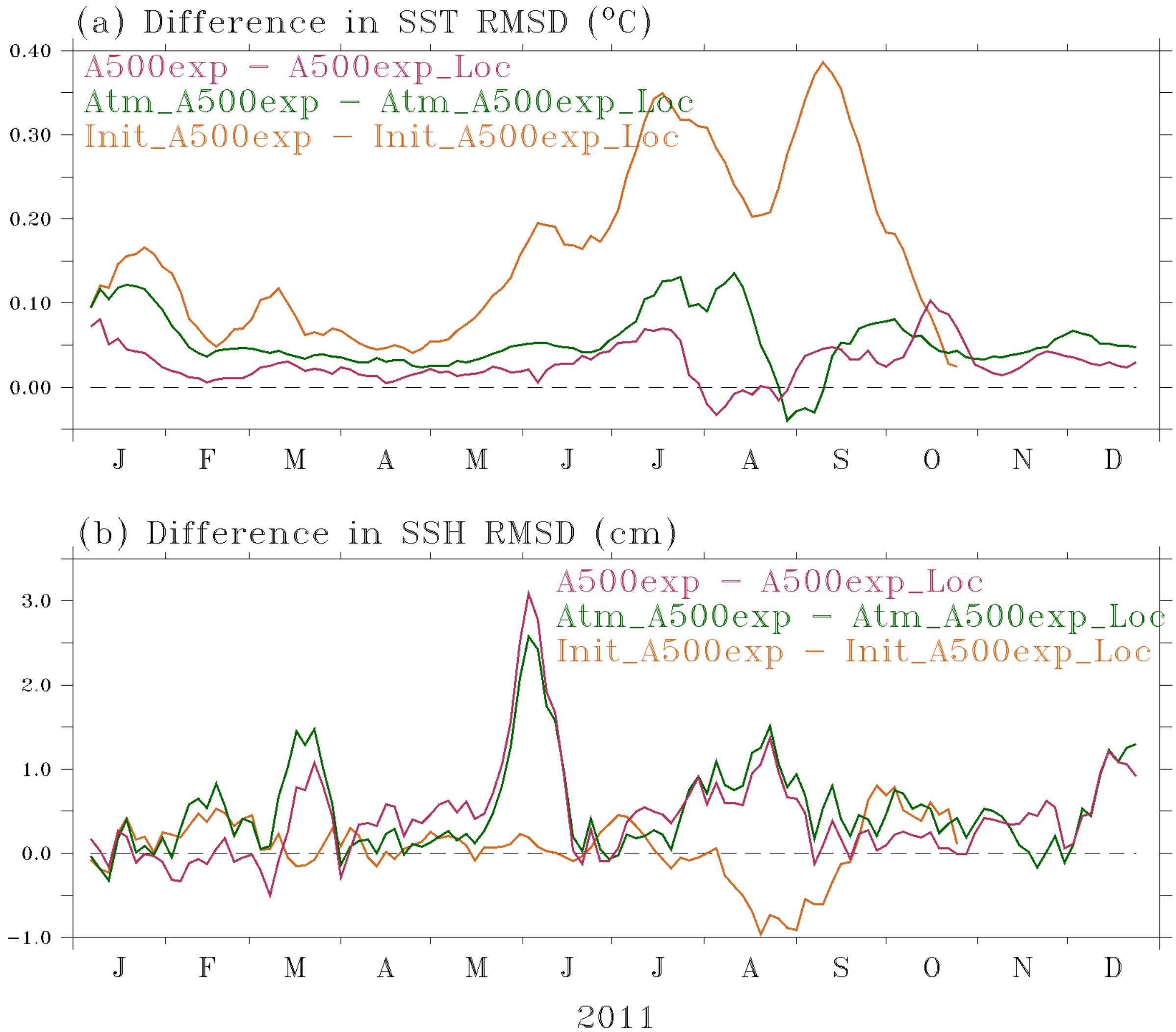

Figure 4. (a) Difference in SST RMSD (°C) between non-localization and localization experiments. Results from six different experiments are represented; *Init_A500exp - InitA500exp_loc* (orange), *Atm_A500exp - Atm_A500exp_loc* (green), *A500exp - A500exp* (maroon). Panel b is same as that of a except for SSH (cm).

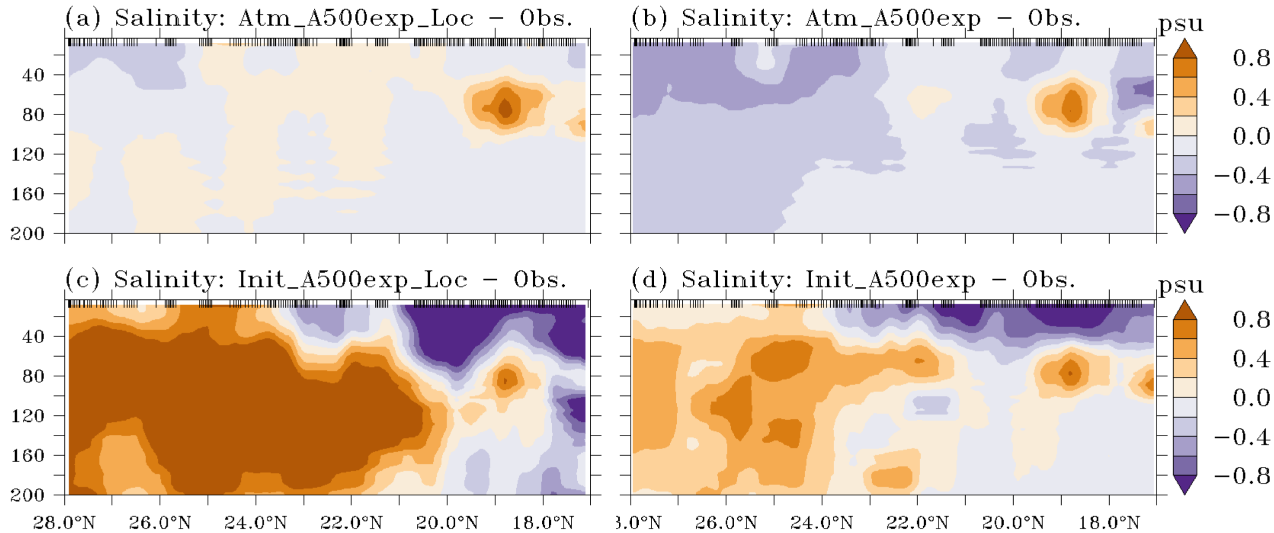

Figure 5. Same as Figure 2 except that the results are shown only for salinity. The salinity differences are between (a) *Atm_A500exp_loc* and observations, (b) *Atm_A500exp* and observations, (c) *Init_A500exp_loc* and observations, (d) *Init_A500exp* and observations.

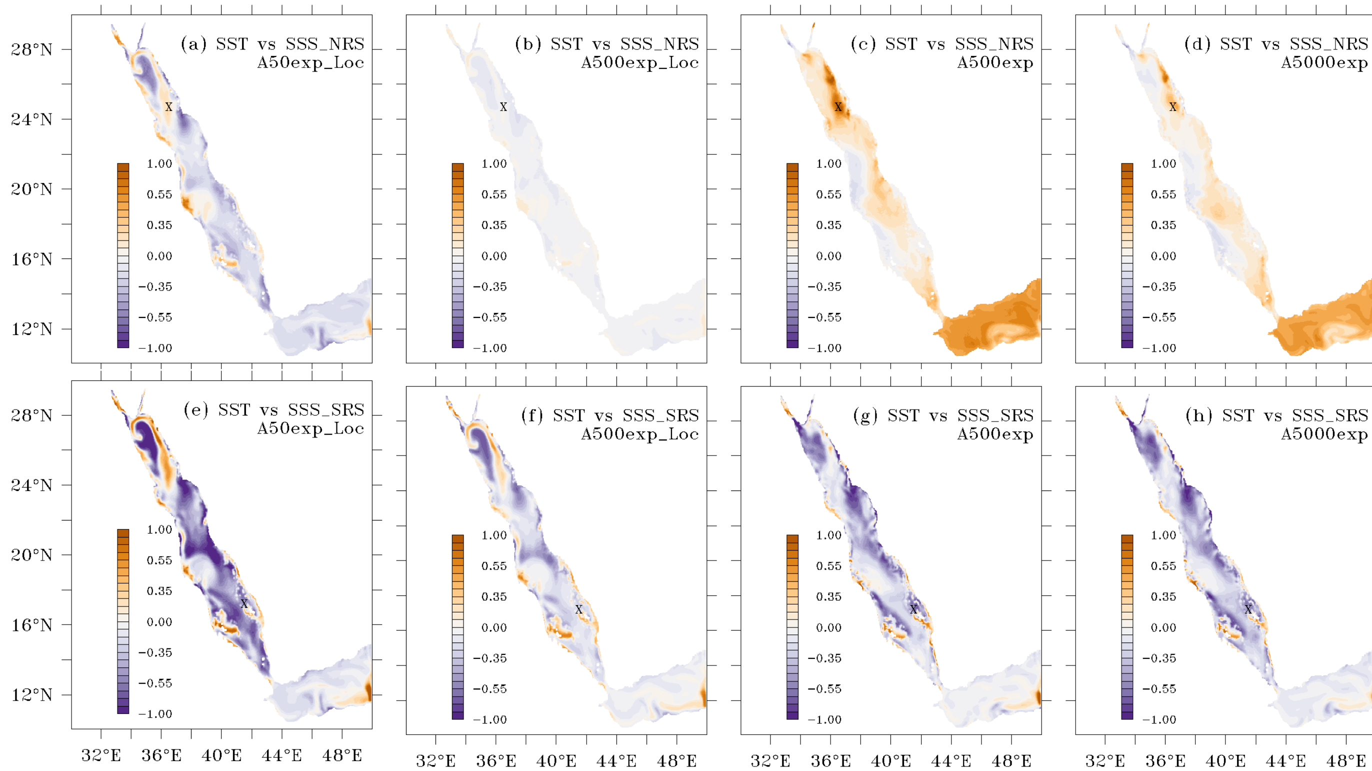

Figure 6: Ensemble cross-correlations, between SST and SSS at a point location, in (a,e) *A50exp_loc*, (b,f) *A500exp_loc*, (c,g) *A500exp*, (d,h) *A5000exp* on 1st October, 2011. The cross-correlations are shown for two different point locations (indicated as a solid black dot), one in the NRS (upper panels) and another in the SRS (bottom panels).

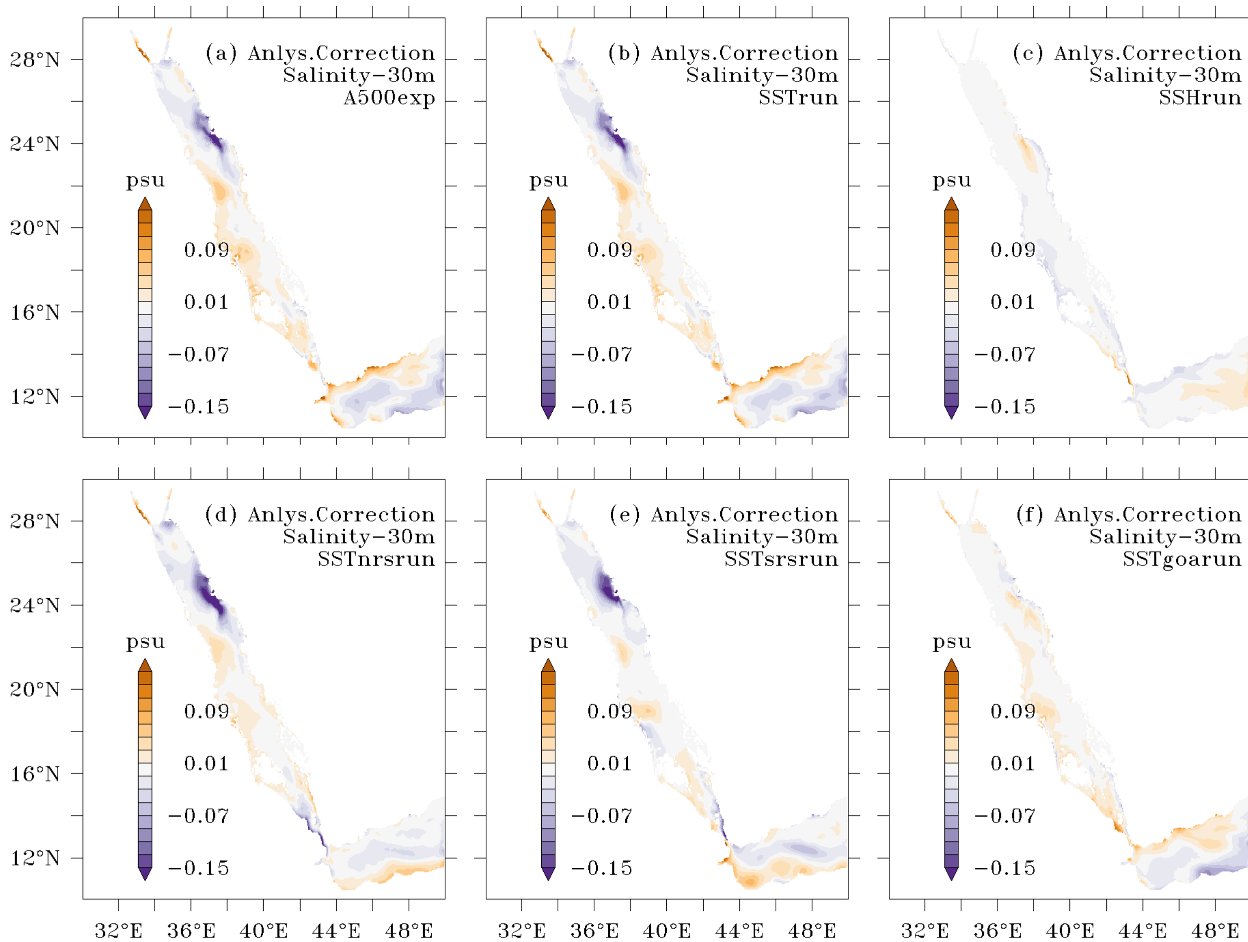

Figure 7. Analysis corrections of salinity (psu) at 30m depth on 1st October, 2011, as resulted from (a) *A500exp*, (b) *SSTrun* (c) *SSHrun*, (d) *SSTnrsrun*, (e) *SSTsrsrun*, and (f) *SSTgoarun*.

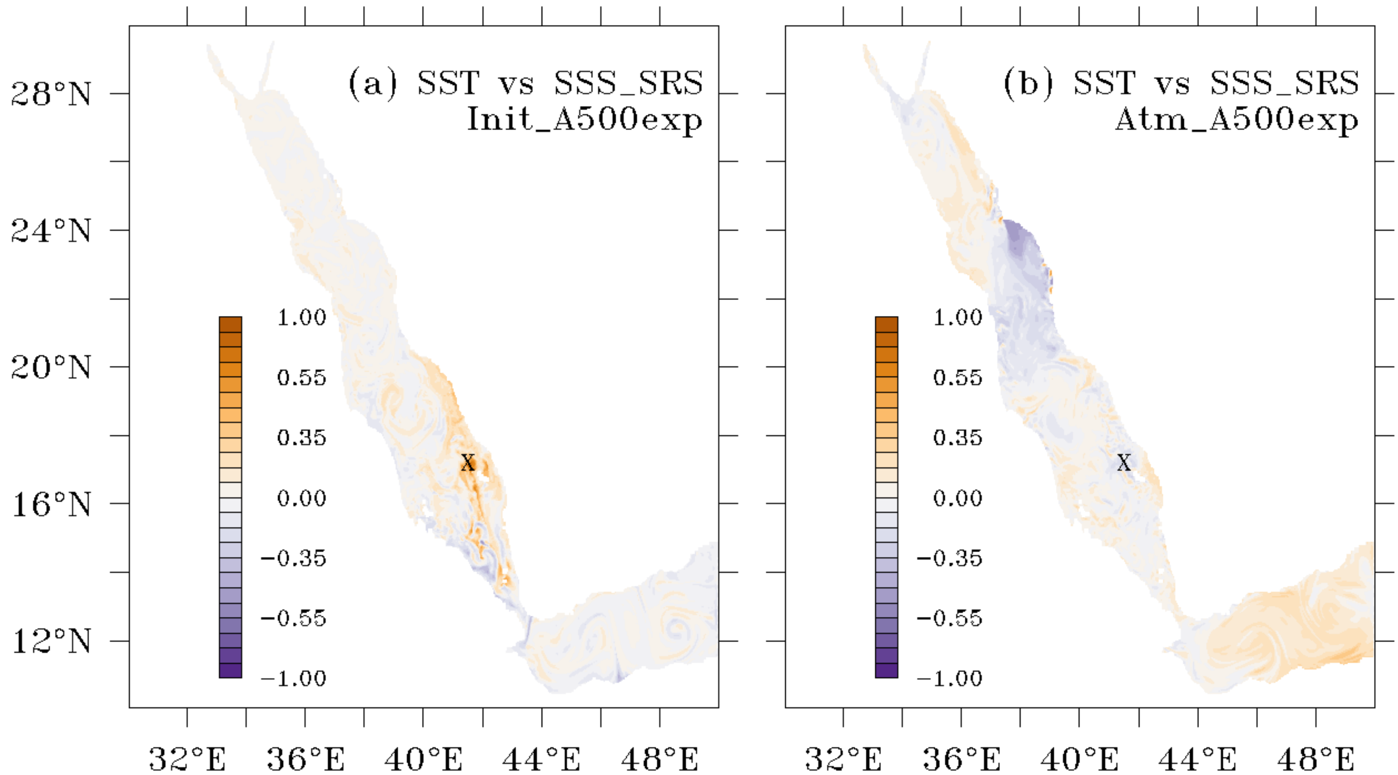

Figure 8. Ensemble cross-correlations, between SST and SSS at a point location in the SRS for (a) *Init_A500exp* and (b) *Atm_A500exp*.

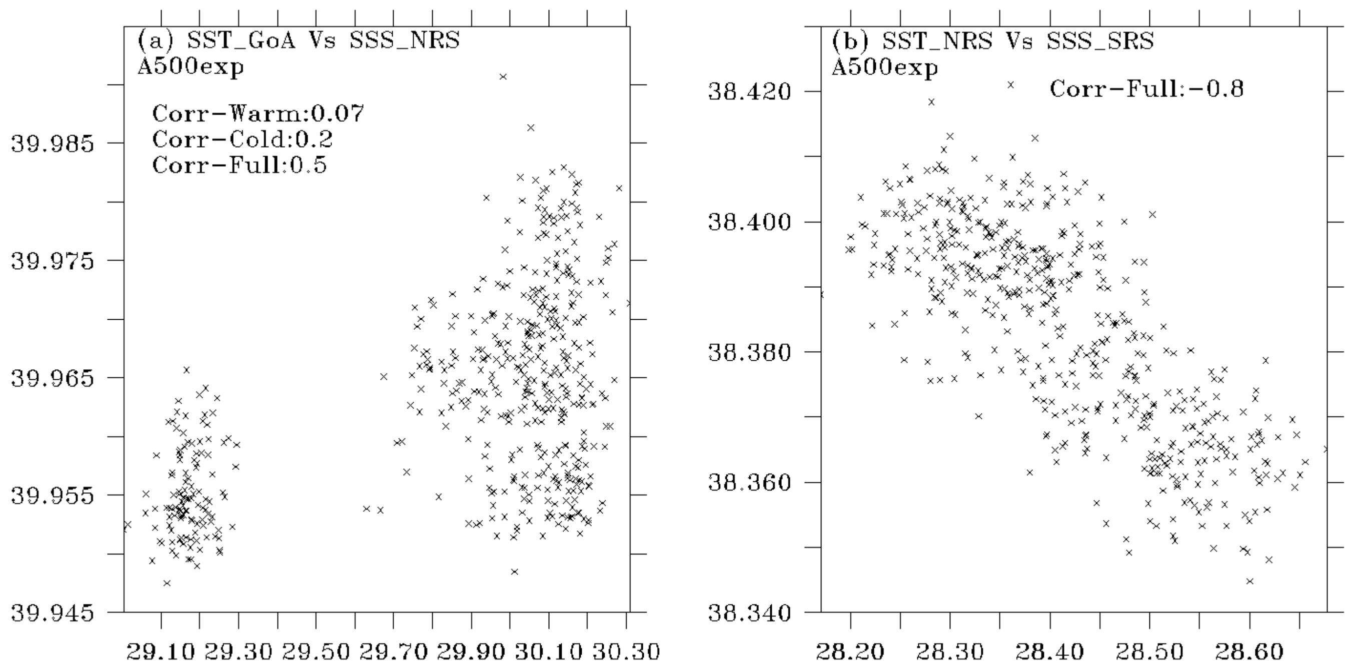

Figure 9. Scatter between SST and SSS in the *A500exp* ensemble corresponds to 1st October 2011. Scatter is between (a) SST in the GoA and SSS in the NRS, and (b) SST in the NRS and SSS in the SRS.