

1 **Insights from very Large Ensemble Data Assimilation Experiments with a High Resolution**

2 **General circulation model of the Red Sea**

3

4 Sivareddy Sanikommu¹, Naila Raboudi¹, Mohamad El Gharampti², Peng Zhan³, Bilel Hadri¹,
5 Ibrahim Hoteit^{1*}

6

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

8 ²National Center for Atmospheric Research, USA

9 ³Department of Ocean Science and Engineering, Southern University of Science and Technology,
10 Shenzhen, China

11

12 **Key points**

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

19

20

***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

21 **Abstract**

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

33 Our results indicate that accounting for model uncertainties is more beneficial than simply
34 increasing the ensemble size, with the improvements due to large ensemble leveling off at about
35 250 members. Besides, and in contrast to what is commonly observed with simplified models, the
36 investigated ensemble DA system still required localization even when implemented with thousands
37 of members. These findings are explained by (i) amplified spurious long-range correlations
38 produced by the low-rank nature of the ECMWF atmospheric forcing ensemble, and (ii) non-
39 Gaussianity generated by the perturbed internal physical parameterization schemes. Large ensemble
40 forcing fields and non-Gaussian DA methods might be needed to take full benefits from large
41 ensembles in ocean DA.

44 **Plain Language Summary**

45 Data assimilation (DA) using Ensemble Kalman Filters (EnKFs) requires large ensembles to
46 estimate robust statistics in order to efficiently spread the observation's information to all model
47 variables, key for robust predictions of the ocean state. Until recently, only limited ensembles (~10-
48 100 members) could be afforded in realistic ocean DA applications. With the ever-increasing
49 computational resources, the use of larger ensembles will become possible in the near future. In this
50 context, the present study assesses the performance of a fully realistic high-resolution ocean EnKF-
51 DA system by systematically examining its sensitivity to ensembles composed of tens-to-thousands
52 of members. It offers fresh perspectives on the employment of ad hoc inflation and localization
53 methods, which have been traditionally implemented with EnKFs to compensate for the use of small
54 ensembles and for the omittance of various model uncertainties. The results of this study suggest
55 large-ensemble forcing fields as well as non-Gaussian DA methods may be needed to maximize the
56 benefits of large ensembles in ocean DA.

57

58

59

60

61

62

63

64 **1. Introduction**

65 Ensemble Kalman Filter (EnKF) based data assimilation (DA) systems provide an efficient
66 framework to update the background error covariance, a critical element for any ocean DA system
67 to spread the observation information across model variables [Derber and Bouttier, 1999; Bouttier
68 and Courtier, 1999, 2002; Edwards et al., 2015; Hoteit et al., 2018]. In ENKFs, a set of ocean
69 states, called ensemble, is integrated using an ocean model for forecasting the first two-moments
70 statistics of the ocean state, i.e., sample mean and covariance, which are in turn used to apply a
71 Kalman filter to update the forecast [Hoteit et al., 2015]. These DA schemes require the size of the
72 ensemble to be large enough to estimate robust statistics [Mitchell et al., 2002; Houtekamer and
73 Zhang, 2016; Lei and Whitaker, 2017; Leutbecher, 2019]. Using small ensembles was shown to
74 produce spurious long-range covariances and rapid collapse of the ensemble spread after few
75 assimilation cycles [Houtekamer and Mitchell, 1998; Anderson 2001]. In spite of this, only limited
76 ensembles (~10-100 members) were so far considered in real-world applications mainly to reduce
77 the computational burden [e.g., Hoteit et al., 2012; Wang and Lei, 2014; Penny et al., 2015; Lei and
78 Whitaker, 2017; Baduru et al., 2019; He et al., 2019; Toye et al., 2020; Sanikommu et al., 2023].
79 Ad hoc techniques such as inflation and localization, which respectively inflates ensemble
80 covariances artificially and limits the observations influence to only within certain radius by
81 tapering long-range correlations in the ensemble, are usually deployed as a compensation for small
82 ensemble size [e.g., Houtekamer and Mitchell, 1998, 2001; Hunt et al., 2007; Bishop and Hodyss,
83 2007; Miyoshi, 2011; Whitaker and Hamill, 2012; Lee et al., 2017; Luo et al., 2018]. With the ever-
84 increasing computational resources, the use of large ensembles will become possible in the near
85 future. Hence, identifying the benefits and outstanding issues associated with large ensembles is
86 important to provide new insights into the future applications of ensemble DA methods in ocean
87 applications.

88 Up to date, large ensemble experiment studies (LEEs) with realistic ocean and atmospheric
89 DA and forecasting systems remain limited. The studies of *Miyoshi et al.* (2014), *Kondo and*
90 *Miyoshi* (2016) and *Toye et al.* (2020) have suggested that with large ensembles, the performance
91 of the ensemble DA system generally becomes less dependent on inflation and localization.
92 However, these LEE studies were not performed with a fully realistic ensemble DA framework. For
93 instance, the LEE (up to 10,000 members) results of the first two studies were based on a mid-
94 complex coupled ocean-atmospheric model and assimilated pseudo-observations. *Toye et al.*,
95 (2018) conducted LEEs, up to 1000 members, with an ocean general circulation model of the Red
96 Sea assimilating real observations, but only accounted for uncertainties in the initial conditions.
97 Ensembles from such mid-complex LEE systems were also investigated to (1) examine the
98 characteristics of background model errors [*Pinardi et al.*, 2008; *Jacques and Zawadzki*, 2015], (2)
99 understand and model sampling errors [*Necker et al.*, 2020a, 2020b], (3) assess non-Gaussianity
100 and sensitivity to covariance localization [*Miyoshi et al.*, 2014; *Kondo and Miyoshi*, 2016; *Toye et*
101 *al.*, 2018], and (4) examine the potential impact of observations [*Necker et al.*, 2020a].

102 Accounting for uncertainties of a forecasting ocean model via stochastic inputs and
103 parameters has recently become popular in the EnKF DA systems [see review papers *Martin et al.*,
104 2015; *Houtekamer and Zhang*, 2016]. In ocean applications, uncertainties are now considered in
105 the atmospheric forcing, boundary conditions, internal physics, bathymetry, and of course initial
106 conditions. These studies suggested noticeable improvements, compared to those that do not
107 account for uncertainties, in the performance of the underlying ensemble assimilation systems [e.g.
108 *Fujita et al.*, 2007; *Bowler et al.*, 2008; *Houtekamer et al.*, 2009; *Kwon et al.*, 2016; *Penny et al.*,
109 2015; *Vandenbulcke and Barth*, 2015; *Sanikommu et al.*, 2017, 2019, 2020; *Baduru et al.*, 2019].
110 Accounting for uncertainties in model's inputs may however introduce non-Gaussian features in
111 the distribution of the forecast ensemble due to the model's non-linearity [e.g., *Sura et al.*, 2005;

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

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

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

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

135 [*Hoteit et al.*, 2013; Gopalakrishnan et al., 2019; *Toye et al.*, 2018, 2020, 2021; and *Sanikommu et*
136 *al.*, 2020, 2023]. The MITgcm-DART configuration used here is basically the same as that of
137 *Sanikommu et al.*, (2020), including the initial conditions, the model physics dictionary (described
138 below), the atmospheric forcing ensemble, assimilated observations, localization, etc. The outputs of
139 this assimilation system were validated against independent in-situ and satellite observations and
140 were found to provide robust estimates of the Red Sea state [*Sanikommu et al.*, 2020; *Toye et al.*,
141 2020], significantly better than those available from the widely used global ocean reanalyses
142 [*Sanikommu et al.*, 2023]. Below we provide a brief overview of the system components.

143 The MITgcm was implemented on a spherical polar grid covering the entire RS domain,
144 including the Gulfs of Suez and Aqaba, and a part of the Gulf of Aden where an open boundary
145 connects it to the Arabian Sea [e.g., *Krokos et al.*, 2019, 2022; *Sanikommu et al.*, 2020; and *Zhan et*
146 *al.*, 2019, 2022]. The open boundary conditions for temperature, salinity, and horizontal velocity are
147 prescribed daily from the 8km-resolution Global Ocean Reanalysis and Simulation data [GLORYS;
148 *Parent et al.*, 2003]. The Red Sea MITgcm uses a direct space-time 3rd order scheme for tracer
149 advection, harmonic viscosity with the coefficients of 30 m²/s in the horizontal and 7x10⁻⁴ m²/s in
150 the vertical direction, implicit horizontal diffusion for both temperature and salinity, and the K-
151 Profile Parameterization (KPP) scheme [*Large et al.*, 1994] for vertical mixing with a vertical
152 diffusion coefficient of 10⁻⁵ m²/s for both temperature and salinity. The model was spin up for 31
153 years starting from 1979 to 2010 using the 75 km resolution European Center for Medium Range
154 Weather Forecast (ECMWF) reanalysis of atmospheric surface fluxes of radiation, momentum, and
155 freshwater sampled every 6 hours [*Dee et al.*, 2011]. The MITgcm outputs have been extensively
156 validated for the RS by earlier studies [e.g. *Yao et al.*, 2014a, 2014b; *Zhan et al.*, 2018; *Toye et al.*,
157 2017; *Gittings et al.*, 2019; *Krokos et al.*, 2022].

158 The EAKF is a deterministic square-root filter [Anderson, 2001; Hoteit et al., 2015]. It is used
159 here to assimilate three types of observations every 3 days, with a localization radius of 300 km. The
160 observations include SST data extracted from a level-4 in-situ and advanced very high-resolution
161 radiometer infrared satellite SST blended daily product available at 25 km resolution [Reynolds et
162 al., 2007]), along-track satellite level-3 merged altimeter filtered sea level anomalies (SLA),
163 corrected for dynamic atmospheric, ocean tide, and long wavelength errors, from Copernicus Marine
164 Environment Monitoring Service (CMEMS; Pujol et al., 2018), and in-situ temperature and salinity
165 profiles available from Good et al. (2013). Observational errors are assumed uncorrelated.
166 Temporally static and spatially homogeneous observational error variance values of $(0.04 \text{ m})^2$,
167 $(0.5^\circ\text{C})^2$, and $(0.3\text{psu})^2$ are prescribed for the satellite along-track SLA, and the in-situ T and S,
168 respectively, in accordance with the suggested ranges of in-situ observational errors by earlier
169 assimilation studies [e.g., Richman et al., 2005; Forget and Wunsch, 2007; Oke and Sakov, 2008;
170 Karspeck, 2016]. The specified observational error variances for SST vary between $(0.1^\circ\text{C})^2$ and
171 $(0.6^\circ\text{C})^2$ in accordance with those of the level-4 gridded SST product of Reynolds et al., (2007).

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

182 model physics in the next cycle. The MPD consists of various vertical and horizontal mixing
183 schemes, and different diffusion and viscosity parameters and is described in details in *Sanikommu*
184 *et al.*, (2020).

185 **3. Data Assimilation Experiments**

186 Ten one-year long assimilation experiments are conducted starting from January 1st, 2011 to
187 systematically investigate the behavior of the Red Sea MITgcm-DART with respect to the ensemble
188 size and localization radius, as outlined in Table 1. The *Fexp* is a free-run in which the model was
189 integrated with the mean of the ECMWF atmospheric ensemble. *A50exp_loc* is a 50-member
190 assimilation experiment accounting for uncertainties in the initial conditions, internal physics, and
191 atmospheric forcing. *A100exp_loc*, *A250exp_loc*, and *A500exp_loc* are the same as *A50exp_loc*
192 with ensembles of 100, 250, and 500 members, respectively. Localization was used in all these
193 experiments with a radius of 300km. *A500exp* and *A5000exp* respectively use 500 and 5000
194 ensemble members and are configured here without localization. The initial ensembles for all these
195 experiments are selected from the January months of a long model hindcast run (here 20 years). The
196 atmospheric forcing ensembles are randomly sampled from a Gaussian distribution with mean and
197 covariance matching those of the 50-member ECMWF ensemble, using

$$198 \quad X_s = \bar{X} + \frac{1}{\sqrt{k-1}} C_k u_{k \times s};$$

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

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

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

216 3.1. Validation Datasets

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

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

229 **4. Results and Discussion**

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

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

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

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

245 *4.1. Computational costs*

246 KAUST is home to SHAHEEN-II [Hadri *et al.*, 2015], a Cray XC40 supercomputer with 6,174
247 dual-socket compute nodes, each powered by 32 core Intel Haswell processors clocked at 2.3 GHz.
248 Each node is equipped with 128GB of DDR4 memory running at 2300MHz. With a total of 197,568

249 processor cores and 790TB of aggregate memory, SHAHEEN-II can achieve a performance of
250 7.2Pflop/s in double precision. All experiments were conducted on this supercomputer, utilizing
251 ROCOTO, a workflow manager to streamline our processes. This allowed us to take full advantage
252 of SHAHEEN-II's capabilities and achieve optimal results.

253 To evaluate the computational costs, we present a comparison in Table 3 between the 5000-
254 member and standard 50-member ensemble experiments, focusing on the two primary components:
255 model integration using MITgcm and assimilation step using DART. For each ensemble member,
256 running the 4km-resolution MITgcm to generate three days of forecasts requires three minutes with
257 3 nodes. Concurrently running 50 MITgcm instances necessitates 150 nodes. The standard 50-
258 member ensemble DART process utilizes 8 nodes and takes approximately 5 minutes to assimilate
259 all observations. In contrast, the 5000-member ensemble DART occupies 590 nodes to meet its
260 large internal memory demand and takes around 30 minutes to finish the task. However, running
261 the entire 5000 MITgcm instances simultaneously exceeds the available resources of SHAHEEN-
262 II, requiring 15,000 nodes. Consequently, the 5000-member ensemble experiments demand a
263 considerable amount of time to complete a single assimilation cycle. The computational cost for the
264 5000-member ensemble experiment to complete one assimilation cycle reaches 33,440 core hours,
265 approximately 130 times higher than that of the standard 50-member ensemble experiment (261
266 core hours). Furthermore, the large number of input/output (IO) operations required for the 5000-
267 member ensemble experiments significantly increases the time needed to complete a full
268 assimilation cycle. This, in turn, increases the overall computational cost and makes it challenging
269 to conduct various sensitivity experiments at this ensemble size. As a result, achieving optimal
270 results can be challenging, especially that Shaheen-II is in production mode with many other users
271 and different applications and workloads.

272 4.2. Comparisons against observations

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

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

291 To further investigate the aforementioned results, we conducted an analysis of the ocean
292 state estimates from various assimilation experiments: *Atm_A500exp_loc*, *Atm_A500exp*,
293 *Init_A500exp_loc*, and *Init_A500exp*. These experiments share a similar configuration to
294 *A500exp_loc* and *A500exp*, with the main difference being the systematic reduction of ensemble

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

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

317 4.3. Long-range ensemble correlations

318 To assess the deteriorations associated with the removal of localization in the assimilation
319 experiments, we analyzed the long-range correlations in the forecast ensembles of different
320 assimilation experiments. Figures 6a-d depict the spatial cross-correlations on October 1st, 2011,
321 between sea surface salinity (SSS) at a location in the northern Red Sea (NRS) and sea surface
322 temperature (SST) in the rest of the domain for *A50exp_loc*, *A500exp_loc*, *A500exp*, and *A5000exp*.
323 Figures 6e-h show the same cross-correlations for another location in the southern Red Sea (SRS).
324 These locations were selected based on prominent salinity differences observed in the assimilation
325 experiments without localization (Figure 2m, 2n, and 5d). Overall, increasing the ensemble size
326 from 50 to 500 smoothens the long-range correlations in the ensemble DA experiments with
327 localization, consistent with previous studies on localization-related error covariance inflation. The
328 differences in correlations between *A500exp_loc* and *A500exp* are notable, despite being driven by
329 the same perturbed physics and atmospheric forcing. All ensemble data assimilation experiments
330 indicate the presence of long-range correlations, even with 5000 ensemble members. However,
331 some of these correlations appear to be spurious. For example, SSS in the NRS exhibits a strong
332 correlation with SST in the Gulf of Aden (GoA) while showing weaker local correlations (Figure
333 6d). Similarly, SSS in the SRS strongly correlates with isolated patches of SST in the NRS, with
334 less significant local correlations (Figure 6h).

335 To investigate whether these spurious long-range correlations have the potential to generate
336 predominant increments, we conducted various single-data assimilation runs using the forecast
337 ensemble of *A500exp* on October 1st, 2011, and examined the resulting analysis increments.
338 Comparing the results obtained from *A500exp* (Figure 7a), *SSHrun* (Figure 7b), and *SSTrun* (Figure
339 7c), we observed that the salinity increments are primarily influenced by the assimilation of SST
340 observations rather than SSH observations. For example, *A500exp* shows strong negative salinity
341 analysis increments (up to 0.2psu) in the northeastern parts of the RS and moderate positive salinity

342 analysis increments in the western parts of the RS and the GoA. The spatial patterns and magnitudes
343 of salinity increments closely resemble those obtained from *SSTrun* compared to *SSHrun*. In Figure
344 7d and 7e, we present salinity increments obtained from the assimilation of SST observations at two
345 different locations: one in the northern RS and another in the GoA. The results from these runs
346 indicate that the salinity increments are prominent not only in the nearby region of the SST
347 observation location but also in farther regions. For instance, *SSTnrsrcrun* yields large salinity
348 increments in the GoA, while *SSTsrsrcrun* generates larger salinity increments in the NRS. These
349 findings clearly indicate that the spurious long-range cross-correlations indeed have the potential to
350 generate predominant increments, thereby confirming that the presence of these spurious long-range
351 correlations is behind the deteriorations observed in ensemble DA runs without localization.

352 *4.3.1 The sources of spurious long-range ensemble correlations*

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

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

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

369 To examine non-Gaussianity, we analyzed skewness (s) and kurtosis (k) of the Probability
370 Density Function (PDF) of the forecast ensemble from *Init_A500exp*, *Atm_A500exp*, and *A500exp*.
371 Table 4 summarizes the percent occurrence of non-Gaussianity in the whole domain for different
372 variables based on the ensembles of *Init_A500exp*, *Atm_A500exp*, and *A500exp*. The PDF is
373 considered non-Gaussian when $|s| > 0.3$ or $|k| > 0.6$. The presence of non-Gaussianity is relatively
374 low for sea surface height (SSH) across all experiments. However, for SST and SSS, non-
375 Gaussianity is notable. It is around 14% in *Init_A500exp*, but increases to 38% in *Atm_A500exp* and
376 86% in *A500exp*. A similar systematic increase in non-Gaussianity is observed for SST as well.

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

387 instances/places where degradations were resulted only from the spurious long-range correlations
388 too.

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

397 **5. Summary and future directions**

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

408 Our results indicate that increasing the ensemble size leads to improved ocean state
409 estimates and reduced long-range correlations. However, we observed that the these improvements

410 seize well below an ensemble size of 500 members. Localization continues to play a crucial role
411 even in the context of large ensembles. We also observed that accounting for uncertainties in various
412 inputs yields more benefits than simply increasing the ensemble size, despite the presence of
413 amplified long-range correlations and non-Gaussianity.

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

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

431 **Acknowledgments**

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

437 **Data Availability statement**

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

439

440

441

442

443

444

445

446

447

448

449

450

451

452

453

454

455 **References**

- 456 Anderson, J.L., 2010. A non-Gaussian ensemble filter update for data assimilation. *Mon Weather*
457 *Rev* 138. <https://doi.org/10.1175/2010MWR3253.1>
- 458 Anderson, J.L., 2001. An ensemble adjustment Kalman filter for data assimilation. *Mon Weather*
459 *Rev* 129. [https://doi.org/10.1175/1520-0493\(2001\)129<2884:AEAKFF>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2)
- 460 Bishop, C. H., & Hodyss, D. (2007). Flow-adaptive moderation of spurious ensemble correlations
461 and its use in ensemble-based data assimilation. *Quarterly Journal of the Royal*
462 *Meteorological Society*, 133(629 B). <https://doi.org/10.1002/qj.169>
- 463 Baduru, B., Paul, B., Banerjee, D.S., Sanikommu, S., Paul, A., 2019. Ensemble based regional ocean
464 data assimilation system for the Indian Ocean: Implementation and evaluation. *Ocean Model*
465 (Oxf) 143. <https://doi.org/10.1016/j.ocemod.2019.101470>
- 466 Bertossa, C., Hitchcock, P., Degaetano, A., Plougonven, R., 2021. Bimodality in Ensemble Forecasts
467 of 2-Meter Temperature: Identification. *Weather and Climate Dynamics* 1–23.
468 <https://doi.org/10.5194/wcd-2021-33>
- 469 Bougeault, P., Toth, Z., Bishop, C., Brown, B., Burridge, D., de Chen, H., Ebert, B., Fuentes, M.,
470 Hamill, T.M., Mylne, K., Nicolau, J., Paccagnella, T., Park, Y.Y., Parsons, D., Raoult, B.,
471 Schuster, D., Dias, P.S., Swinbank, R., Takeuchi, Y., Tennant, W., Wilson, L., Worley, S.,
472 2010. The thorpex interactive grand global ensemble. *Bull Am Meteorol Soc* 91.
473 <https://doi.org/10.1175/2010BAMS2853.1>
- 474 Bouttier, F., and Courtier, P., 2002. Data assimilation concepts and methods March 1999. Training.
- 475 Bouttier, F., and Courtier, P., 1999. Data assimilation concepts and methods, in: ECMWF
476 *Meteorological Training Course Lecture Series*.

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

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

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

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

492 Edwards, C.A., Moore, A.M., Hoteit, I., and Cornuelle, B.D. (2015). Regional ocean data
493 assimilation. *Annual Review of Marine Science*, 7, 21–42, doi:10.1146/annurev-marine-
494 010814-015821.

495 EUMETSAT/OSI-SAF, GHRSSST level 3P global subskin sea surface temperature from the
496 Advanced Very High Resolution Radiometer (AVHRR) on the MetOp-A satellite, 2008,
497 <http://dx.doi.org/10.5067/GHGMT-3PE01>, Ver. 1. PO.DAAC, CA, USA. Dataset accessed
498 2018.10.01.

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

501 Fletcher, S. J., M. Zupanski, M. R. Goodliff, A. J. Kliwer, A. S. Jones, J. M. Forsythe, T.-C. Wu,
502 M. J. Hossen and S. Van Loon, 2023. Lognormal and Mixed Gaussian-lognormal based
503 Kalman Filters. *Monthly Weather Review*, 151, 761-774. [https://doi.org/10.1175/MWR-D-](https://doi.org/10.1175/MWR-D-22-0072.1)
504 [22-0072.1](https://doi.org/10.1175/MWR-D-22-0072.1)

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

554 Jacques, D., Zawadzki, I., 2015. The impacts of representing the correlation of errors in radar data
555 assimilation. Part II: Model output as background estimates. *Mon Weather Rev* 143, 2637–
556 2656. <https://doi.org/10.1175/MWR-D-14-00243.1>

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

560 Kondo, K., Miyoshi, T., 2016. Impact of removing covariance localization in an ensemble Kalman
561 Filter: Experiments with 10 240 members using an intermediate AGCM. *Mon Weather Rev*
562 144, 4849–4865. <https://doi.org/10.1175/MWR-D-15-0388.1>

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

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

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

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

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

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

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

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

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

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

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

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

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

602 Necker, T., Geiss, S., Weissmann, M., Ruiz, J., Miyoshi, T., Lien, G.Y., 2020a. A convective-scale
603 1,000-member ensemble simulation and potential applications. *Quarterly Journal of the*
604 *Royal Meteorological Society* 146, 1423–1442. <https://doi.org/10.1002/qj.3744>

605 Necker, T., Weissmann, M., Ruckstuhl, Y., Anderson, J., Miyoshi, T., 2020b. Sampling error
606 correction evaluated using a convective-scale 1000-member ensemble. *Mon Weather Rev*
607 148, 1229–1249. <https://doi.org/10.1175/MWR-D-19-0154.1>

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

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

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

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

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

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

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

627 Sanikommu, S., Banerjee, D.S., Baduru, B., Paul, B., Paul, A., Chakraborty, K., Hoteit, I., 2019.
628 Impact of dynamical representational errors on an Indian Ocean ensemble data assimilation

629 system. Quarterly Journal of the Royal Meteorological Society 145.
630 <https://doi.org/10.1002/qj.3649>

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

635 Sanikommu, S., Paul, A., Sluka, T., Ravichandran, Muthalugu, Kalnay, E., 2017. The pre-Argo
636 ocean reanalyses may be seriously affected by the spatial coverage of moored buoys. Sci Rep
637 7, 46685. <https://doi.org/10.1038/srep46685>

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

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

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

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

- 651 Sura, P., Sardeshmukh, P.D., 2008. A global view of non-Gaussian SST variability. *J Phys Oceanogr*
652 38. <https://doi.org/10.1175/2007JPO3761.1>
- 653 Toye, H., Kortas, S., Zhan, P., Hoteit, I., 2018. A fault-tolerant HPC scheduler extension for large
654 and operational ensemble data assimilation: Application to the Red Sea. *J Comput Sci* 27,
655 46–56. <https://doi.org/10.1016/j.jocs.2018.04.018>
- 656 Toye, H., Sanikommu, S., Raboudi, N.F., Hoteit, I., 2020. A hybrid ensemble adjustment Kalman
657 filter based high-resolution data assimilation system for the Red Sea: Implementation and
658 evaluation. *Quarterly Journal of the Royal Meteorological Society* 146.
659 <https://doi.org/10.1002/qj.3894>
- 660 Toye, H., Zhan, P., Gopalakrishnan, G., Kartadikaria, A.R., Huang, H., Knio, O., Hoteit, I., 2017.
661 Ensemble data assimilation in the Red Sea: sensitivity to ensemble selection and atmospheric
662 forcing. *Ocean Dyn* 67, 915–933. <https://doi.org/10.1007/s10236-017-1064-1>
- 663 Toye, H., Zhan, P., Sana, F., Sanikommu, S., Raboudi, N., Hoteit, I., 2021. Adaptive ensemble
664 optimal interpolation for efficient data assimilation in the red sea. *J Comput Sci* 51.
665 <https://doi.org/10.1016/j.jocs.2021.101317>
- 666 Vandenbulcke, L., Barth, A., 2015. A stochastic operational forecasting system of the Black Sea:
667 Technique and validation. *Ocean Model* (Oxf) 93.
668 <https://doi.org/10.1016/j.ocemod.2015.07.010>
- 669 Wang, X., Lei, T., 2014. GSI-based four-dimensional ensemble-variational (4DEnsVar) data
670 assimilation: Formulation and single-resolution experiments with real data for NCEP global
671 forecast system. *Mon Weather Rev* 142. <https://doi.org/10.1175/MWR-D-13-00303.1>

- 672 Whitaker, J.S., Hamill, T.M., 2002. Ensemble data assimilation without perturbed observations. *Mon*
673 *Weather Rev* 130. [https://doi.org/10.1175/1520-0493\(2002\)130<1913:EDAWPO>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2)
- 674 Whitaker, J. S., & Hamill, T. M. (2012). Evaluating methods to account for system errors in ensemble
675 data assimilation. *Monthly Weather Review*, 140(9). <https://doi.org/10.1175/MWR-D-11->
676 00276.1
- 677 Yao, F., Hoteit, I., Pratt, L.J., Bower, A.S., Köhl, A., Gopalakrishnan, G., Rivas, D., 2014a. Seasonal
678 overturning circulation in the Red Sea: 2. Winter circulation. *J Geophys Res Oceans* 119,
679 2263–2289. <https://doi.org/10.1002/2013JC009331>
- 680 Yao, F., Hoteit, I., Pratt, L.J., Bower, A.S., Zhai, P., Köhl, A., Gopalakrishnan, G., 2014b. Seasonal
681 overturning circulation in the Red Sea: 1. Model validation and summer circulation. *J*
682 *Geophys Res Oceans* 119, 2238–2262. <https://doi.org/10.1002/2013JC009004>
- 683 Zhai, P., Bower, A.S., Smethie, W.M., Pratt, L.J., 2015. Formation and spreading of Red Sea Outflow
684 Water in the Red Sea. *J Geophys Res Oceans* 120. <https://doi.org/10.1002/2015JC010751>
- 685 Zhan, P., Gopalakrishnan, G., Subramanian, A.C., Guo, D., Hoteit, I., 2018. Sensitivity Studies of
686 the Red Sea Eddies Using Adjoint Method. *J Geophys Res Oceans* 123, 8329–8345.
687 <https://doi.org/10.1029/2018JC014531>
- 688 Zhan, P., Guo, D., Krokos, G., Dong, J., Duran, R., Hoteit, I., 2022. Submesoscale Processes in the
689 Upper Red Sea. *J Geophys Res Oceans* 127. <https://doi.org/10.1029/2021JC018015>
- 690 Zhan, P., Krokos, G., Guo, D., Hoteit, I., 2019. Three-Dimensional Signature of the Red Sea Eddies
691 and Eddy-Induced Transport. *Geophys Res Lett* 46. <https://doi.org/10.1029/2018GL081387>

692

693

694 **Tables**

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

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

699

700

701

702

703

704

705

706

707

708

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

Assimilation Experiment	SSH	SST	T&S profiles
<i>SSHrun</i>	Yes; whole domain	No	No
<i>SSTrun</i>	No	Yes; whole domain	No
<i>SSTnrsrcrun</i>	No	Yes; only in the NRS	No
<i>SSTsrsrcrun</i>	No	Yes; only in the SRS	No
<i>SSTgoarun</i>	No	Yes; only in the GoA	No

713

714

715

716

717

718

719

720

721

722

723

724

725

726

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

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

730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749

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

	SST	SSS	SSH
<i>Init_A500exp</i>	26	14	4
<i>Atm_A500exp</i>	27	38	3
<i>A500exp</i>	59	86	4

752

753

754

755

756

757

758

759

760

761

762

763

764

765

766

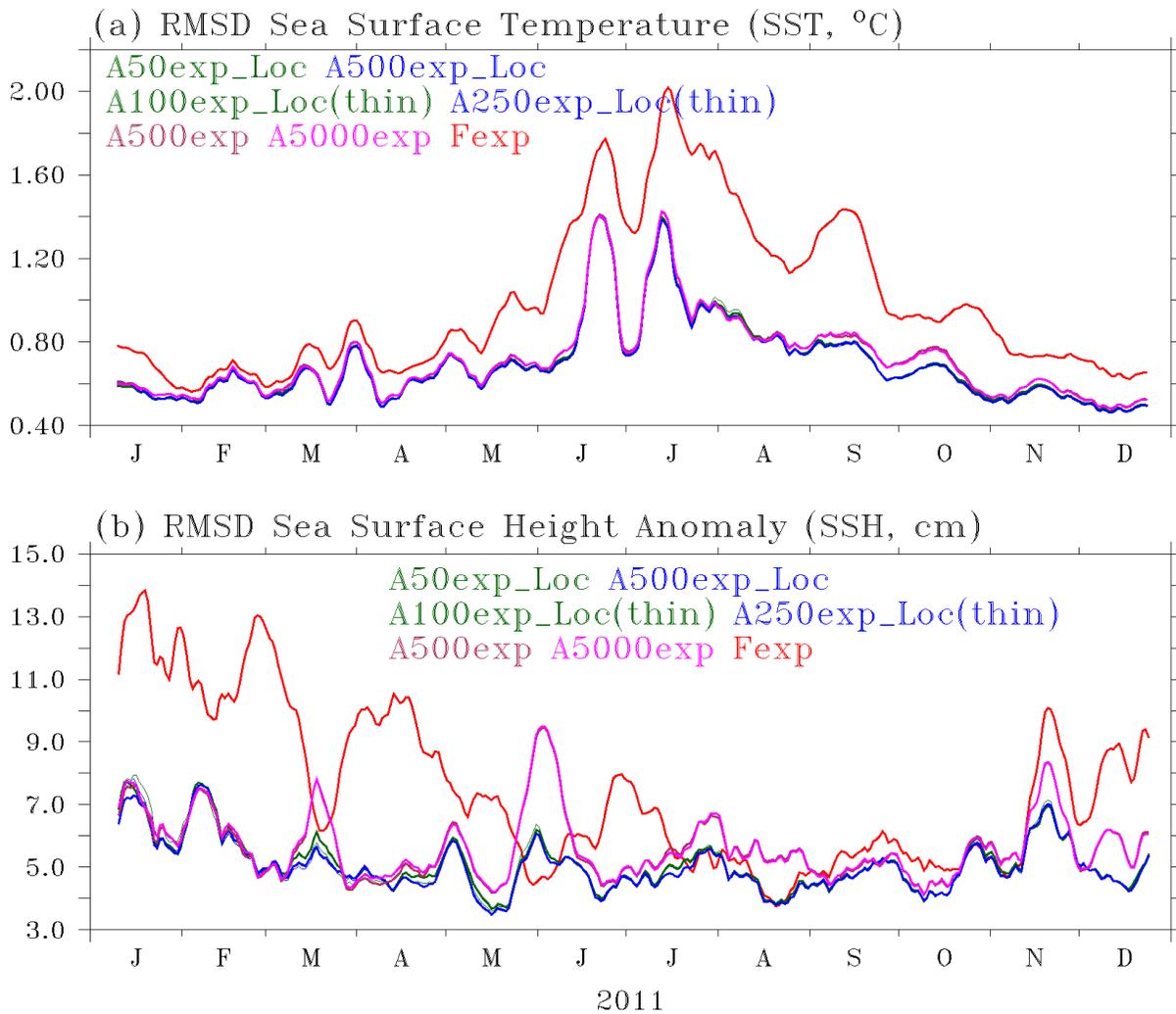
767

768

769

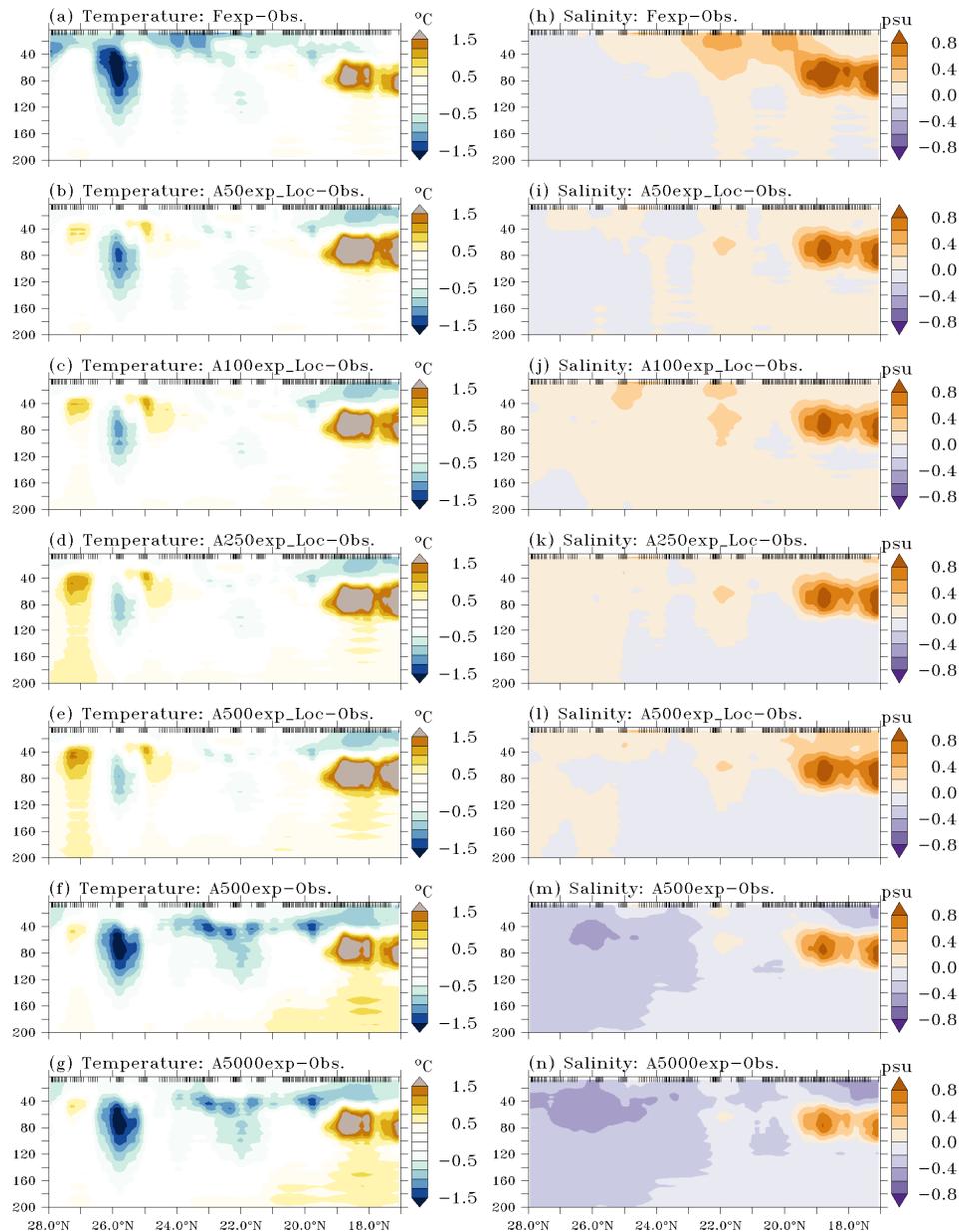
770

771 **Figures**



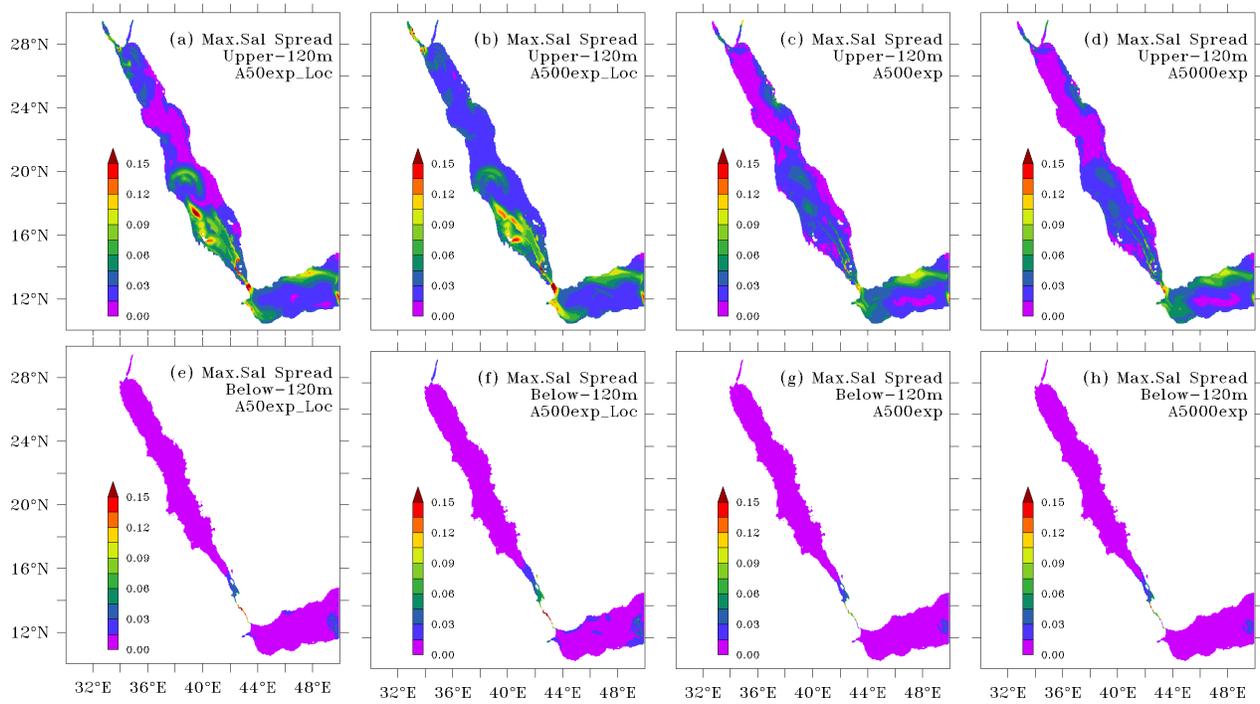
772

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



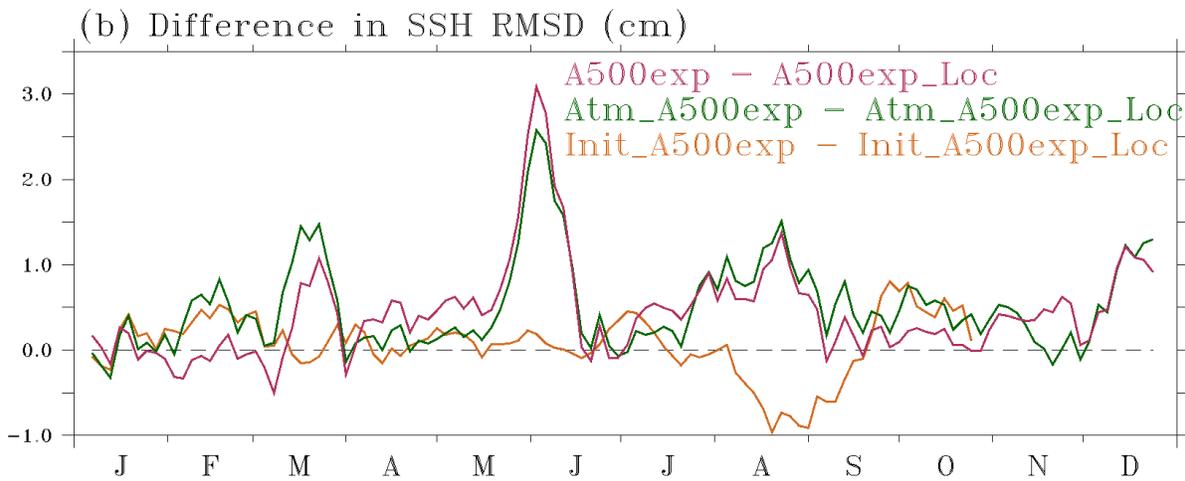
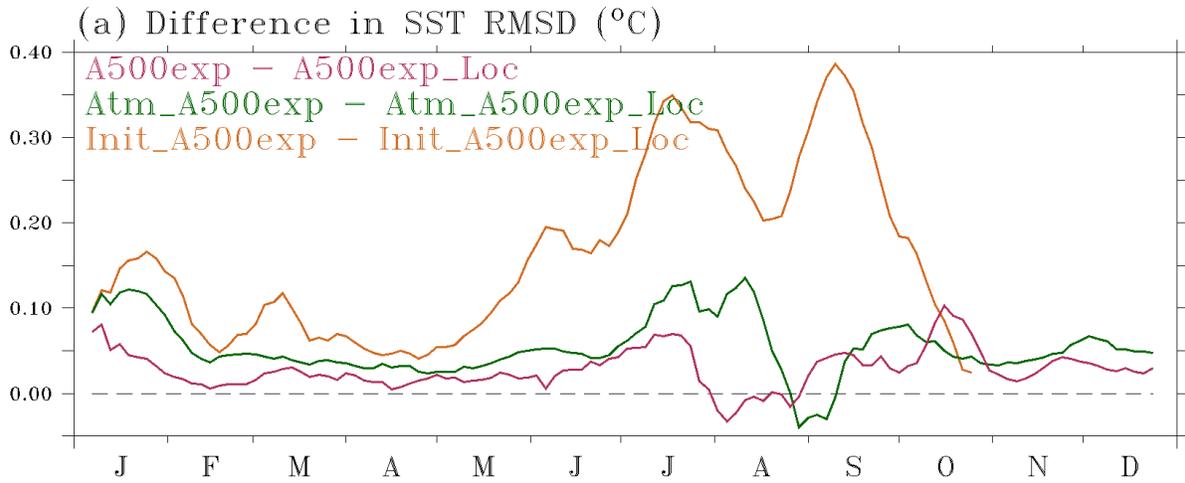
778

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



786

787 Figure 3. Maximum ensemble salinity spread in the (a-d) upper 120m and (e-h) below 120m. Results
 788 are shown for (a,e) *A50exp_loc*, (b,f) *A500exp_loc*, (c,g) *A500exp*, (d,h) *A5000exp* on 1st October,
 789 2011.



2011

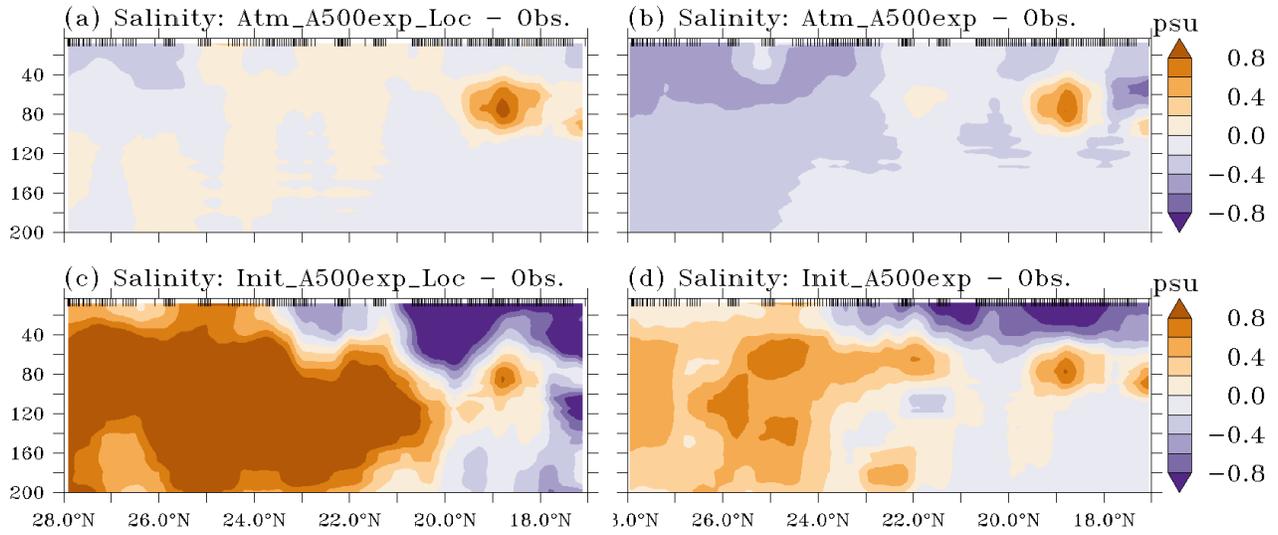
790

791 Figure 4. (a) Difference in SST RMSD ($^{\circ}\text{C}$) between non-localization and localization experiments.

792 Results from six different experiments are represented; $Init_A500exp - InitA500exp_loc$ (orange),

793 $Atm_A500exp - Atm_A500exp_loc$ (green), $A500exp - A500exp$ (maroon). Panel b is same as that of

794 a except for SSH (cm).



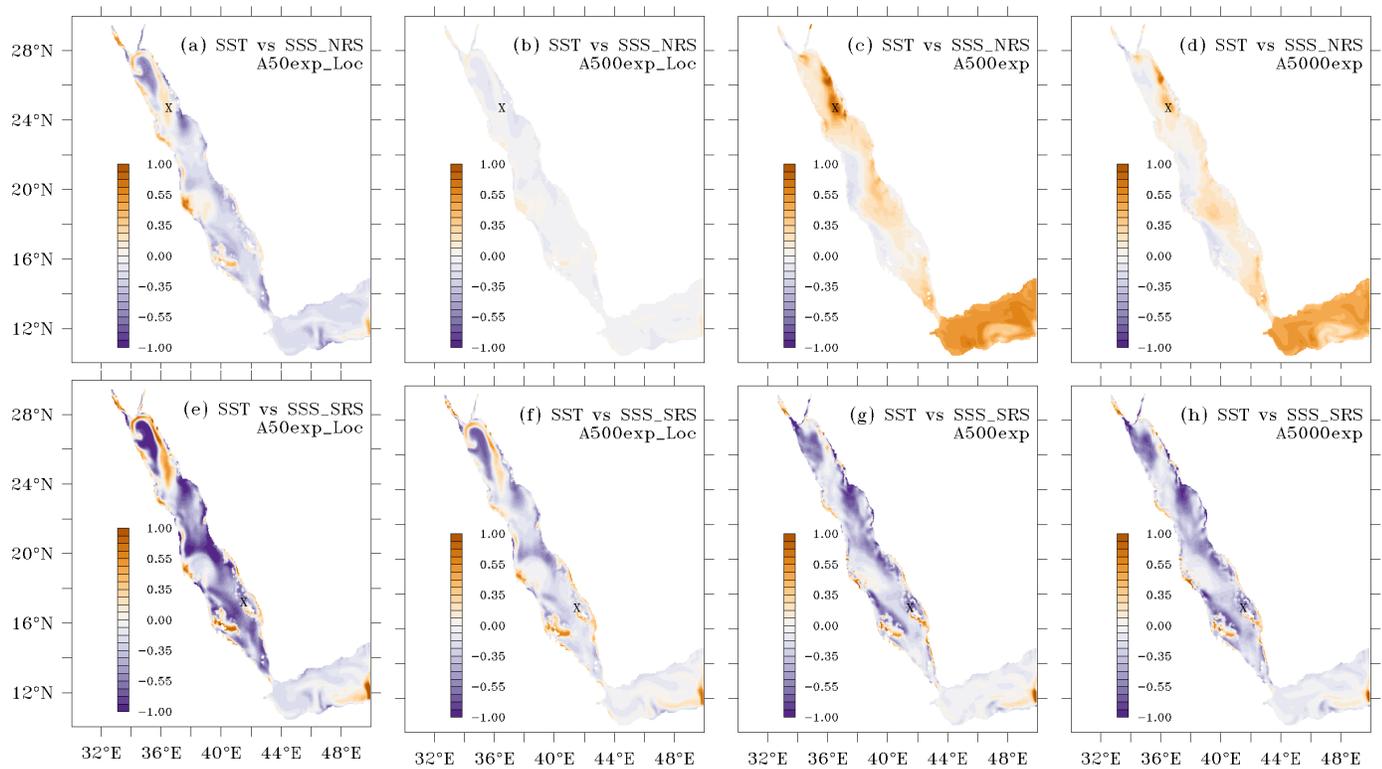
795

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

799

800

801



802

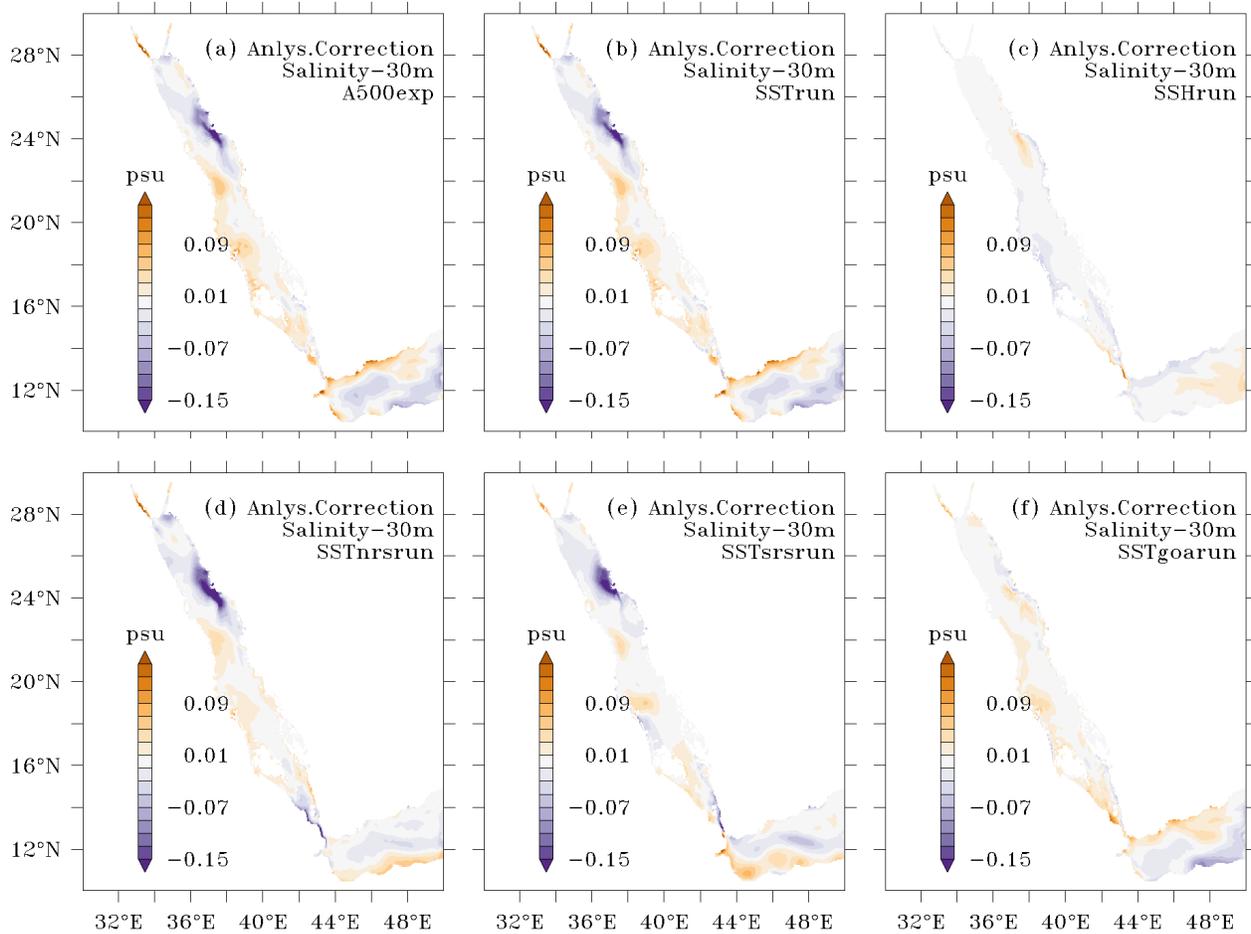
803 Figure 6: Ensemble cross-correlations, between SST and SSS at a point location, in (a,e) *A50exp_loc*,

804 (b,f) *A500exp_loc*, (c,g) *A500exp*, (d,h) *A5000exp* on 1st October, 2011. The cross-correlations are

805 shown for two different point locations (indicated as a solid black dot), one in the NRS (upper panels)

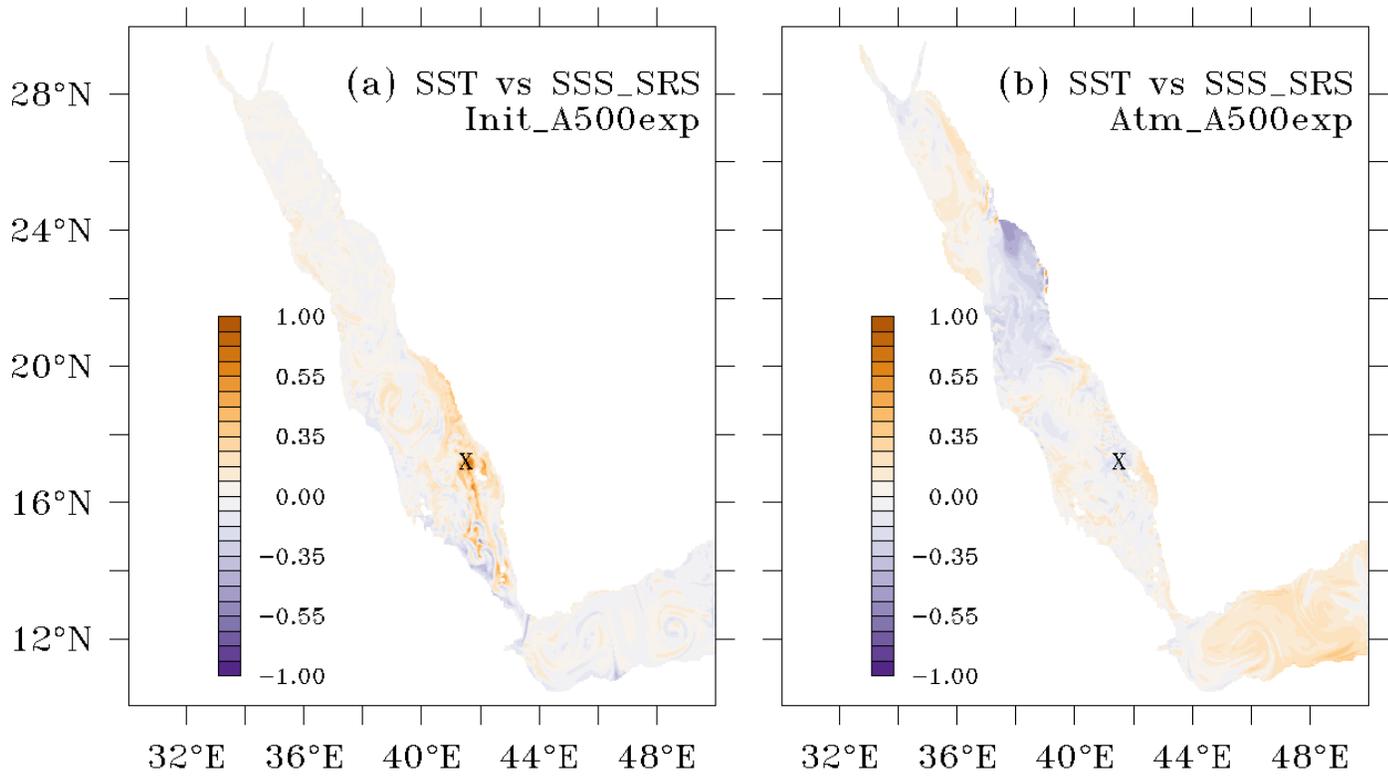
806 and another in the SRS (bottom panels).

807



808
 809 Figure 7. Analysis corrections of salinity (psu) at 30m depth on 1st October, 2011, as resulted from (a)
 810 *A500exp*, (b) *SSTrun* (c) *SSHrun*, (d) *SSTnrsrun*, (e) *SSTsrsrun*, and (f) *SSTgoarun*.

811
 812
 813
 814
 815



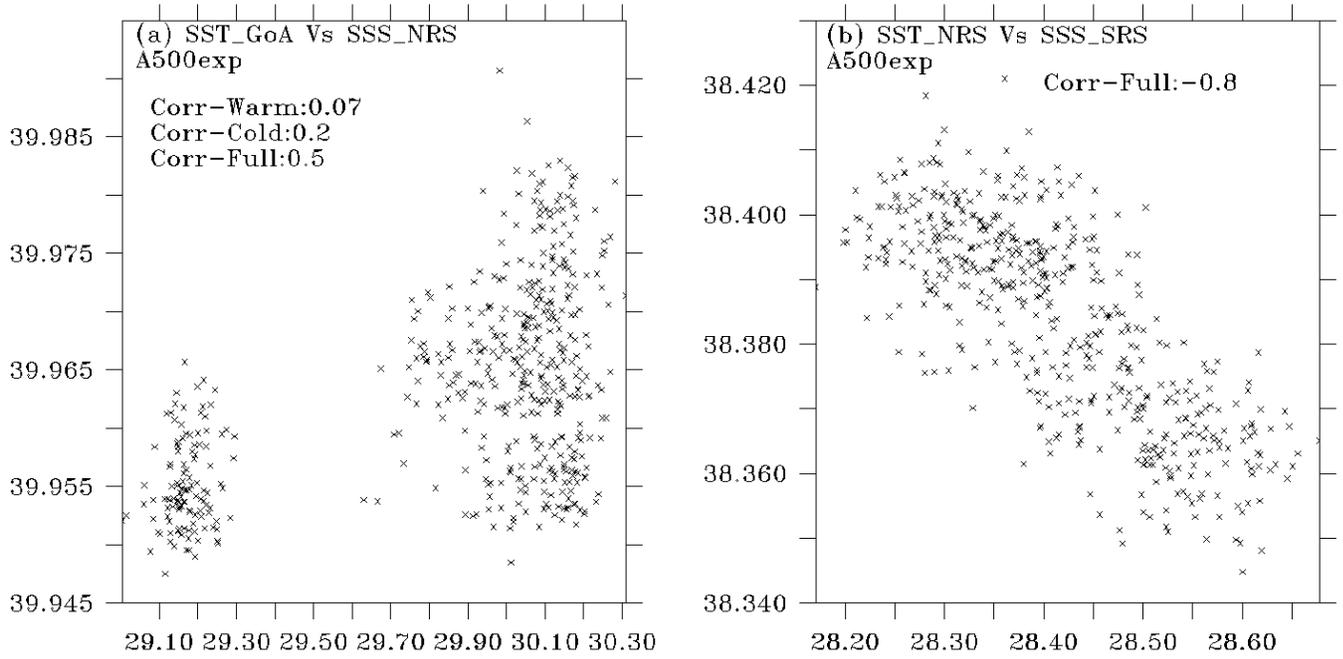
816

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

818 *Init_A500exp* and (b) *Atm_A500exp*.

819

820



821

822 Figure 9. Scatter between SST and SSS in the *A500exp* ensemble corresponds to 1st October 2011.

823 Scatter is between (a) SST in the GoA and SSS in the NRS, and (b) SST in the NRS and SSS in the

824 SRS.

825

826

827

828

829

830

831

832

833

834

835