https://doi.org/10.5194/os-2019-85
Preprint. Discussion started: 6 August 2019

Ocean Science
Discussions
Open Access
EGU

Evaluation of fraternal versus identical twin approaches for observation impact assessments: An EnKF-based ocean assimilation application for the Gulf of Mexico

Liuqian Yu[1,2], Katja Fennel[1], Bin Wang[1], Arnaud Laurent[1], Keith R. Thompson[1] and Lynn K. Shay[3]

[1]Department of Oceanography, Dalhousie University, Halifax, Nova Scotia, Canada

[2]Department of Mathematics, The Hong Kong University of Science and Technology, Hong Kong

[3]Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, Florida, USA

E-mail: liuqianyu@ust.hk

https://doi.org/10.5194/os-2019-85
Preprint. Discussion started: 6 August 2019

Ocean Science
Discussions
Open Access
EGU

Abstract

Assessments of ocean data assimilation (DA) systems and observing system design experiments typically rely on identical or fraternal twin experiments. The identical twin approach has been recognized as yielding biased impact assessments in atmospheric predictions but these shortcomings are not sufficiently appreciated for oceanic DA applications. Here we present the first direct comparison of the fraternal and identical twin approach in an ocean DA application. We assess the assimilation impact for both approaches in a DA system for the Gulf of Mexico that uses the Ensemble Kalman Filter. Our comparisons show that, despite a reasonable error growth rate in both approaches, the identical twin produces a biased skill assessment overestimating the improvement from assimilating sea surface height and sea surface temperature observations while underestimating the value of assimilating temperature and salinity profiles. Such biases can lead to an undervaluation of some observing assets (in this case profilers) and thus misguided distribution of observing system investments.

https://doi.org/10.5194/os-2019-85
Preprint. Discussion started: 6 August 2019

1. Introduction

Ocean data assimilation (DA), i.e. the incorporation of observations into ocean models to obtain the best possible estimate of the ocean state, has become standard practice for improving the accuracy of model predictions and reanalyses. Benefiting from the rapid expansion of ocean observing platforms and advances in computing power, various ocean DA applications at both regional and global scales have been developed in support of ocean hindcasts, nowcasts and forecasts (e.g., see recent reviews in Moore et al. 2019 and Fennel et al. 2019). Necessarily the credibility of a DA system demands rigorous validation. It is straightforward to assess the assimilation impact (i.e. the differences between ocean state estimates from a model run with and without assimilation), where a better fit of the model state to observations following assimilation might be considered as positive. But in practice, the value of such an assessment is limited because it does not consider independent observations (i.e., observations that have not been assimilated into the system).

An alternative assessment approach is to conduct twin experiments (e.g., Anderson et al., 1996; Halliwell et al., 2014). The essential steps of a twin experiment are to 1) predefine a simulation as the "truth", 2) sample synthetic observations from this "truth", 3) assimilate these observations into a different simulation referred to as the forecast run, and 4) assess the skill of this assimilative run against a non-assimilative ("free") run using independent observations sampled from the "truth". Conventionally, if the chosen "truth" and forecast runs are from two different model types or significantly different configurations of the same model type (e.g., using different physical parameterizations and/or spatial resolution), the method is referred to as the 'fraternal twin' approach; if the same model but with perturbed initial, forcing or boundary conditions is used, the method

https://doi.org/10.5194/os-2019-85
Preprint. Discussion started: 6 August 2019

48 is referred to as 'identical twin' approach (Halliwell et al., 2014). In addition to validating
49 DA systems, twin experiments are used for Observing System Simulation Experiments
50 (OSSEs) that evaluate the impact of different ocean observing system designs on predictive
51 skill (e.g., Oke and O'Kane 2011; Halliwell et al. 2015, 2017). Ideally, the "truth" and
52 forecast simulations in the twin system used for the OSSE should be from two different
53 models.

54 The identical twin approach has been more commonly used in oceanic DA
55 applications (e.g., Friedrichs 2001, Counillon and Bertino, 2009b; Simon and Bertino, 2009;
56 Srinivasan et al., 2011; Song et al, 2016; Yu et al., 2018) although it is well known from
57 atmospheric OSSEs that this approach provides biased impact assessments when the error
58 growth rate between the "truth" and forecast runs is insufficient (e.g., Arnold and Dey 1986;
59 Atlas 1997; Hoffman and Atlas 2016). This fact is not yet sufficiently recognized in
60 applications of ocean OSSEs and skill assessments of oceanic DA systems (Halliwell et al.,
61 2014). To avoid the potential bias in impact assesments associated with identical twin
62 experiments, Halliwell et al. (2014) proposed to apply a criterion that has long been used
63 in realistic atmospheric OSSEs. They suggested that the model for the forecast run should
64 be configured differently enough from that for the "truth" run so that the rate of error growth
65 between them has the same magnitude as that between state-of-the-art ocean models and
66 the true ocean. In practice this is difficult to assess.

67 Alternatively, Halliwell et al. (2014) suggested to compare the assimilation impact
68 in the fraternal twin framework with that in a realistic configuration; if a similar impact is
69 obtained in both twin and realistic configurations, the fraternal twin DA framework can be
70 considered appropriate for assessing assimilation impact and conducting OSSEs (Halliwell

https://doi.org/10.5194/os-2019-85
Preprint. Discussion started: 6 August 2019

et al. 2014). Such fraternal OSSEs have proven instructive for evaluating the assimilation impact of different observing platforms in the Gulf of Mexico (Halliwell et al., 2015) and North Atlantic (Halliwell et al., 2017). However, a direct comparison of fraternal and identical twin approaches has not yet been conducted for an ocean application, to the best of our knowledge. Motivated by this, we use an ocean DA system for the Gulf of Mexico (GOM) to compare and contrast the fraternal and identical twin approaches in an assimilation impact assessment. We use 'Fraternal twin' to refer to the case where two different models are used. The rationale for choosing the GOM as our testbed is that the non-deterministic aspects of the circulation in the GOM, including the northward penetration of Loop Current (LC) intrusions and the associated eddy shedding, require DA for accurately hindcasting/forecasting the circulation. The need for accurate nowcasts and predictions was particularily acute during the 2010 Deepwater Horizon (DwH) oil spill. Previous data assimilation applications in the GOM have focussed primarily on improvements of the surface current fields observeable from satellite or drifters but did not examine the assimilation impact on subsurface flow fields. As the DwH oil spill has shown, knowledge of model skill in simulating the subsurface circulation is also important. Utilizing twin experiments, we aim to examine the assimilation impact on the subsurface circulation.

Toward this objective we implement an advanced ensemble DA technique, the Ensemble Kalman Filter (EnKF), for a high-resolution (horizontal resolution of 5 km) model covering the entire GOM. The EnKF utilizes flow-dependent background error covariances in contrast to the time-invariant covariance in optimal interpolation (OI-) or variational-based DA systems that have previously been used in GOM (e.g., Counillon and

https://doi.org/10.5194/os-2019-85
Preprint. Discussion started: 6 August 2019

94 Bertino 2009a, 2009b; Jacobs et al. 2014). By rigorously assessing the skill of the EnKF-
95 based assimilative model (with an emphasis on the subsurface fields) through fraternal and
96 identical twin experiments and OSSEs, we demonstrate how the identical twin approach
97 yields misleading conclusions in this practical application. We also address whether an
98 improved skill in reproducing the surface dynamics of the LC and associated eddies
99 translates into improved skill in simulating the subsurface circulation.
100
101 **2. Model description and experimental setup**
102 **2.1 The physical model**
103 The model is configured using the Regional Ocean Modelling System (Haidvogel
104 et al., 2008; ROMS, http://myroms.org) for the GOM (Fig. 1a). It has a horizontal resolution
105 of 5 km and 36 terrain-following vertical layers with higher resolution near the surface and
106 bottom. Vertical turbulent mixing is parameterized using the Mellor and Yamada (1982)
107 Level 2.5 closure scheme, and bottom friction is specified using a quadratic drag
108 formulation. The model utilizes a third-order accurate, non-oscillatory advection scheme
109 for tracers (HSIMT, Wu and Zhu, 2010), which is mass-conservative and positive-definite
110 with low dissipation and no overshooting, and is forced with the atmospheric forcing fields
111 from the European Centre for Medium-Range Weather Forecasts (ECMWF)
112 (http://apps.ecmwf.int/datasets/). River input is prescribed as in Xue et al. (2013), with
113 daily runoff from US Geological Survey for rivers inside the US and long-term
114 climatological estimates for rivers in Mexico and Cuba. The model is one-way nested inside
115 the 1/12° data-assimilative global Hybrid Coordinate Ocean Model (HYCOM) (Chassignet
116 et al., 2009). Tidal forcing is neglected because tides are small in the GOM.

https://doi.org/10.5194/os-2019-85
Preprint. Discussion started: 6 August 2019

Previous studies have highlighted two important aspects for model skill in the GOM, a sufficiently high horizontal resolution for representing the mesoscale dynamics (e.g., Chassignet et al., 2005) and an accurate representation of the LC inflow through the Yucatan Strait (e.g., Oey, 2003). Our model meets the two requirements. The 5-km horizontal resolution is sufficient to resolve mesoscale processes (the baroclinic Rossby radius is 30 to 40 km in the central GOM, see, Oey et al., 2005). And our ROMS model is nested in a data-assimilative HYCOM model which simulates an accurate structure of the LC and its eddies. Initial model-data comparisons showed that the model has skill in statistically simulating the main features of the LC intrusion with a slight tendency to overestimating its northward penetration.

2.2 Experimental framework

The deterministic formulation of the EnKF (DEnKF), first introduced by Sakov and Oke (2008), was implemented in the GOM model. The DEnKF has been successfully used in previous ocean assimilation applications (e.g., Simon et al., 2015; Jones et al., 2016; Yu et al., 2018). The algorithm consists of sequential forecast and analysis steps, where the model ensemble is propagated forward in time during the forecast step and updated with available observations using the Kalman Filter analysis equation during the analysis step. The analysis equation is given as:

$$\boldsymbol{x}^a = \boldsymbol{x}^f + \boldsymbol{K}(\boldsymbol{d} - \boldsymbol{H}\boldsymbol{x}^f), \quad (2)$$

where $\boldsymbol{x}$ is the $n \times 1$ model state estimate vector (n is the number of model state variables at all grid points), the superscripts a and f represent the analysis and the forecast estimates, respectively, $\boldsymbol{d}$ is the $m \times 1$ vector of observations (m is the number of available

https://doi.org/10.5194/os-2019-85
Preprint. Discussion started: 6 August 2019

observations), $\boldsymbol{H}$ is the linear $m \times n$ measurement operator mapping the model state onto the observations, and $\boldsymbol{K}$ is the $n \times m$ Kalman gain matrix, given as

$$\boldsymbol{K} = \boldsymbol{P}^f \boldsymbol{H}^T (\boldsymbol{H}\boldsymbol{P}^f \boldsymbol{H}^T + \boldsymbol{R})^{-1}, \quad (3)$$

where $\boldsymbol{P}^f$ is the $n \times n$ forecast error covariance matrix (approximated by the forecast ensemble), $\boldsymbol{R}$ is the $m \times m$ observation error covariance, and T denotes the matrix transpose. Details on the DEnKF derivation and implementation can be found in Sakov and Oke (2008).

2.2.1 Fraternal twin experiments

In fraternal twin experiments, the "truth" is generated by interpolating the daily outputs of the 1/12° data-assimilative global HYCOM (Chassignet et al., 2009) onto the ROMS model grid. Synthetic observations are sampled from the "truth", including SSH, SST, and temperature and salinity profiles. Typical Gaussian observation errors of $N(0, 2$ cm) for SSH, $N(0, 0.3$ °C) for temperature (both SST and temperature profiles), and $N(0,$ 0.01) for salinity are added to the sampled data. SSH and SST are sampled weekly at every fifth horizontal grid point to yield a spatial resolution of ~1/4°. SSH in regions shallower than 300 m is not used for assimilation because dynamics in shelf areas where wind and buoyancy forcing dominate could substantially deviate from the geostrophic state weakening the correlation between SSH and subsurface temperature and salinity fields. For SST, only those in regions shallower than 10 m are excluded. Importantly when preparing the synthetic SSH observations, the mean dynamic topography (MDT) of the HYCOM "truth" run had to be removed from the sampled SSH data and the MDT of the ROMS model had to be added. The MDTs of the HYCOM and ROMS models were obtained by averaging their respective daily SSH outputs from 2010 to 2016.

https://doi.org/10.5194/os-2019-85
Preprint. Discussion started: 6 August 2019

Temperature and salinity profiles were sampled with two different sampling schemes (see locations in Fig. 1a, b). The first scheme adopts the sampling dates and locations used in the survey described in Shay et al. (2011). The key features of this scheme are that the sampling is centered on the LC region, the majority (363 out of 472) of temperature profiles are limited to the upper 400 m, and very few (34) salinity profiles were collected. In the second scheme coverage was extended such that temperature and salinity profiles are sampled simultaneously over the entire central GOM down to 1000 m depth on 23 instead of 9 dates.

A non-assimilative run, subsequently referred to as the free run, is initialized on 1 April 2010 from the global HYCOM and compared with the data-assimilative runs to evaluate the impact of the assimilation.

In the DA experiments, 20-member ensembles are started from different initial conditions and forced by perturbed boundary conditions and wind fields. The initial conditions were created by using three-dimensional (3D) fields from daily HYCOM outputs within a 20-day window centered on the initialization date of 1 April 2010. The boundary conditions were generated by applying a time lag of up to +/-10 days to the boundary condition (i.e., the first member's boundary conditions are 10 days ahead) following Counillon and Bertino (2009b). The perturbed wind fields were created by first conducting an empirical orthogonal function (EOF) decomposition of the wind field and then adding perturbations from the mixture of the first 4 EOF modes to the wind field, where the four perturbation modes were multiplied with zero-mean unit-variance random numbers and a scale factor of 0.5 similar to Thacker et al. (2012) and Li et al. (2016).

https://doi.org/10.5194/os-2019-85
Preprint. Discussion started: 6 August 2019

We used an ensemble of 20 as it was the largest size feasible given the computing resources available to us and found this to work well in our application. The same ensemble size has also been used in previous studies (e.g., Hu et al., 2012; Mattern et al., 2013). Distance-based localization with an influence radius of 50 km was applied as described in Evensen (2003) to prevent the potential negative effects of spurious correlations between distant grid points. An inflation factor of 1.05 was applied to the ensemble anomalies inflating the ensemble spread around its mean at every assimilation step as introduced by Anderson and Anderson (1999). This accounts for the potential underestimation of the forecast error covariance due to the small ensemble size. The choice of localization radius and inflation factor are based on initial tests and takes into account that the baroclinic Rossby radius in the central GOM is 30 to 40 km (Oey et al., 2005) to avoid choosing too small localization radius value.

Observations are assimilated weekly from 2 April to 3 September 2010 updating the 3D temperature and salinity fields. On each assimilation date, the observations (regardless of observation types) are assimilated simultaneously in one single step. After the last assimilation step on 3 September 2010, the ensemble is run without any data assimilation for 4 more weeks. Three assimilation experiments (referred to as F1, F2 and F3) are conducted. F1 assimilates weekly SSH and SST, F2 and F3 assimilate the temperature and salinity profiles following the two sampling schemes described earlier (Fig. 1a, b) in addition to SSH and SST. Model-data misfit is quantified by computing the Mean Absolute Deviations (MAD) of model simulations from the "truth" for the open Gulf (defined as regions deeper than 300 m). For ensemble assimilation runs, the forecast ensemble mean at assimilation steps is used for calculating the MAD.

https://doi.org/10.5194/os-2019-85
Preprint. Discussion started: 6 August 2019

207 ***2.2.2 Identical twin experiments***

208 The identical twin experiments have a similar setup as the fraternal twin
209 experiments except that the "truth" is not taken from HYCOM but generated from a ROMS
210 simulation that differs from the free run only in its initial and boundary conditions and wind
211 forcing. The "truth" run is started on 1 April 2010 from an initial state from an earlier
212 ROMS simulation, and is forced with boundary conditions that are lagging behind those of
213 the free run by 14 days and wind fields reconstructed from the first 10 EOFs of the realistic
214 ECMWF wind. Since the same model architecture is used in free and reference runs for the
215 identical twin, there is no need to correct MDT when sampling SSH observations.

216 Similar to the fraternal twin setup, three assimilation experiments are conducted in
217 the identical twin framework (I1, I2 and I3) that assimilate the same combinations of
218 observations as in F1, F2 and F3.

219

220 **3. Results**

221 **3.1. Assessment of the fraternal and identical twin experiment setup**

222 We first examine the credibility of the fraternal and identical twin setups by
223 comparing the error growth rates in SSH between the free run and the "truth" for both twins
224 (Fig. 2). The fraternal twin has a slightly higher error growth rate (0.048 cm/day) than the
225 identical twin (0.040 cm/day), but both are of similar magnitude to that between the free
226 run and real observations (0.042 cm/day). This meets the requirement suggested by
227 Halliwell et al. (2014) that the errors between the free run and the "truth" should grow at a
228 similar rate as errors that develop between the state-of-the-art ocean models and the true
229 ocean. The comparison in Fig. 3 also shows that differences between the "truth" and free

https://doi.org/10.5194/os-2019-85
Preprint. Discussion started: 6 August 2019

runs in SSH and subsurface salinity fields are obvious and qualitatively comparable between the fraternal and identical twin experiments. This satisfies the other requirements suggested in Halliwell et al. (2014), namely that the free run is able to reproduce the main features of the simulated phenomenon (i.e. the LC intrusion) with some realism, and that there are sufficient differences between the free and "truth" runs for the assimilation method to correct.

3.2. Impact of assimilation in fraternal twin experiments

Temporally and spatially averaged MADs between the fraternal twin assimilation runs and the free run are summarized in Table 1 (temporal evolution is shown in Fig. S1 in the supplement). Assimilating SSH and SST in F1 significantly reduces the MADs of SSH (by 51%), temperature (by 29%) and velocity fields (by 25%), and slightly reduces MADs in salinity (by 11%) (Table 1). After the last assimilation step, MADs remain low for at least 4 weeks (Fig. S1). Assimilating additional temperature and salinity profiles (in F2 and F3) further benefits temperature and especially salinity fields, in particular in F3, where the salinity MAD are reduced by 23%, but has almost no effect on SSH and velocity MAD.

In F1 the MAD in SSH, temperature, and velocity components is reduced for almost the entire domain, with the most significant reductions in the LC region (Fig. 4). The reduction in salinity MAD is relatively small in F1 but larger in F3 where additional temperature and salinity profiles are assimilated (Fig. S2 in the supplement). In contrast to SSH, temperature, and velocity, the biggest impact of assimilation on the salinity field is on the shelf where salinity is more variable than in the open Gulf because of river inputs.

Vertically, the reductions of spatially and temporally averaged MAD extend to nearly 900 m depth for temperature and velocity, and 500 m for salinity (Fig. 5). The

https://doi.org/10.5194/os-2019-85
Preprint. Discussion started: 6 August 2019

253 maximum reductions in MAD amount to 0.6 ℃ for temperature at 200 m, 0.12 for surface
254 salinity, and 0.07 m/s for surface velocity (Fig. 5). Assimilating temperature and salinity
255 profiles in F3 leads to greater reductions of temperature and salinity MAD primarily in the
256 upper 300 m compared to F1.

257 Next, we assess the impact of assimilation on subsurface temperature (Fig. S3 in the
258 supplement) and salinity fields (Fig. 6). The "true" spatial distribution of mean temperature
259 and salinity at 400 m depth in August shows only a weak northward intrusion of warm and
260 salty LC water and a detached anticyclonic eddy. Compared to the "truth", the free run
261 overestimates the northward extension of the LC (depicted by the 12 ℃ isotherm and 35.5
262 isohaline), and the detached eddy is misaligned. Assimilation corrects the extension and
263 angle of the LC and the position of the eddy, significantly reducing the averaged MAD
264 error by 47% (52%) and 31% (46%) for temperature and salinity, respectively in the F1 (F3)
265 run.

266 Lastly, we examine the assimilation impact on subsurface circulation in a
267 comparison of August mean circulation at 400 m depth of the fraternal twin runs (Fig. 7).
268 The "truth" shows a limited northeastward extension of the LC with two eddies shedding
269 (Fig. 7d). As mentioned already above, the free run overestimates the northward extension
270 and simulates a more energetic detached anticyclonic eddy that has propagated further west
271 (Fig. 7e). Assimilation in F1 brings the simulated shape, strength and location of the LC
272 and LC eddies closer to the "truth" with an overall MAD reduction of ~45% compared to
273 the free run (Fig. 7f). A closer look at the LC intrusion region (Fig. 7g, h, i) and the western
274 (Fig. 7a, b, c) and northern shelf breaks (Fig. 7j, k, l) shows that the greatest improvement
275 in subsurface circulation is in the open Gulf and LC region where mesoscale processes

https://doi.org/10.5194/os-2019-85
Preprint. Discussion started: 6 August 2019

dominate (MAD reduction of ~57%), whereas the improvement in circulation is weaker along the shelf regions where submesoscale processes are important and influences of the open ocean, bathymetry and local wind and river forcing coexist (MAD reductions of ~25% and ~42% on the western and northern shelf, respectively). Specifically, the small-scale currents surrounding the spill site observed in the "truth" (i.e., the strong anticyclonic eddy to the east of the spill site and cyclonic eddy to its southwest) are not satisfactorily represented in either the free run or F1. The results of F2 and F3 are very similar to F1.

3.3. Assimilation impact in identical versus fraternal twins

Assimilating SSH and SST in identical twin I1 leads to even larger error reductions than in the fraternal twin F1 with domain-averaged MAD reductions in temperature of 45% (versus 29% in the fraternal twin), salinity of 21% (versus 11%) and velocity fields of 46% (versus 25%) (Table 1). However, the benefit of assimilating additional temperature and salinity profiles in I2 and I3 on temperature and salinity fields in the identical twin framework is much smaller than in the fraternal twin (Table 1).

With respect to the simulated subsurface circulation, the improvement by assimilating SSH and SST is also much greater in identical twin I1 (Fig. S4 in the supplement) than in fraternal twin F1 with a MAD reduction of ~67% versus ~45%. In addition, a remarkable improvement in subsurface circulation following assimilation in I1 is observed not only in the LC intrusion region (MAD reduction of ~69%) but also on the shelves (~55% and ~63%, respectively, on the western and northern shelves), including the region near the DwH spill site (Fig. S4).

https://doi.org/10.5194/os-2019-85
Preprint. Discussion started: 6 August 2019

299 **4. Discussion**

300 We implemented the EnKF technique in a high-resolution regional model for the
301 GOM. The skill of this data-assimilative system was assessed through a series of fraternal
302 and identical twin experiments assimilating data from different observing system
303 configurations. The differences between the two approaches have important implications
304 for observing system design studies.

305 Consistent with previous assimilation studies in the GOM (e.g., Wang et al., 2003;
306 Counillon and Bertino 2009b; Hoteit et al., 2013), our fraternal and identical twin
307 experiments both show that assimilating altimetry data can constrain a range of large-scale
308 to mesoscale features such as the LC and associated eddies. The warmer and more saline
309 LC and its eddies have a temperature and salinity signature that is distinct from the so-
310 called Gulf Common Water and have a clear signal of elevated SSH. Assimilation of SSH
311 using the multivariate EnKF therefore can adjust temperature and salinity profiles based on
312 the SSH information. Assimilation of SSH and SST substantially corrects the subsurface
313 temperature, salinity and velocity fields from the surface to depths of up to 900 m, with
314 clear improvements in location and intensity of the LC and LC eddies.

315 The fraternal twin experiments show that salinity is less constrained than
316 temperature when assimilating only SSH and SST. Assimilation of additional temperature
317 profiles (experiment F2) only slightly improves salinity; inclusion of salinity profiles
318 (experiment F3) is more effective in improving salinity. This highlights the value of
319 assimilating salinity profiles to constrain model salinity fields. The importance of salinity
320 measurements has also been reported in the realistic DA configuration by Halliwell et al.
321 (2015). However, such additional benefits of assimilating temperature and salinity profiles

https://doi.org/10.5194/os-2019-85
Preprint. Discussion started: 6 August 2019

on model-simulated temperature and salinity fields are not observed in the identical twin experiments, which already yield much greater improvements when assimilating SSH and SST alone. It follows that, the additional information content in the subsurface observations (i.e., profiles) within the identical twin system is much smaller than that for the fraternal twin.

Another major difference between the fraternal and identical twin approaches lies in the assimilation impact on subsurface circulation. In the fraternal twin experiments, assimilating satellite altimetry effectively constrains the large to mesoscale structures on the order of 100 km that dominate the deep GOM. The improved circulation in deep GOM has a positive but relatively limited impact on the circulation near the DwH spill site, which is located in the transition zone between the open Gulf (where the circulation is dominated by the mesoscale LC and its eddies) and the shelf (where currents are largely driven by wind and density forcing). The assimilation of SSH, SST and additional temperature and salinity profiles (at a resolution of ~70 km) in our fraternal twin experiments provides limited constraints on the small-scale circulation features in this region. This is consistent with Wang et al. (2003) who found that assimilating SSH and SST could not accurately resolve smaller-scale eddies in the DeSoto Canyon region near the DwH site. It has been suggested previously that higher-resolution localized observations (Lin et al., 2007; Jacobs et al., 2014; Carrier et al., 2014; Berta et al., 2015; Muscarella et al., 2015) and even finer model resolution (< 5 km, Ledwell et al., 2016) are needed to better constrain these submesoscale features. In contrast to the fraternal twin, the identical twin I1, which assimilates only SSH and SST, yields remarkable improvements not only in the mesoscale

https://doi.org/10.5194/os-2019-85
Preprint. Discussion started: 6 August 2019

344 circulation dominating the open GOM but also the smaller-scale processes prevailing along
345 the shelf breaks, including the DeSoto Canyon region where the spill site is located.
346 These results provide two examples of how the identical twin approach yields
347 misleading impact assessments: 1) the improvement in subsurface fields resulting from
348 assimilating SSH and SST is overestimated, and 2) the value of additional profiles is
349 underestimated. Undervaluing the information provided by a class of observational assets
350 is particularly troublesome in the context of OSSEs. While this issue is well known in the
351 context of atmospheric OSSEs (e.g., Arnold and Dey 1986; Atlas 1997; Hoffman and Atlas
352 2016), it is not yet sufficiently recognized for ocean OSSEs and skill assessments of oceanic
353 DA systems. Halliwell et al. (2014)'s set of design criteria and evaluation procedures for
354 ocean OSSEs serves as guidance for designing twin experiments for a data-assimilative
355 system. Their main criterion is that the rate of error growth between simulated and observed
356 states must be similar between the twin framework and reality. However, we found a similar
357 rate of error growth in SSH in both twin experiments and in reality, yet the identical twin
358 proved problematic. Thus, assessing error growth in just one ocean property appears to have
359 been insufficient. In all, our results clearly support the use of the fraternal over the identical
360 twin approach, but they also hint that other criteria in addition to assessing the rate of error
361 growth between the forecast run and "truth" are needed to obtain more credible impact
362 assessment from fraternal twin.
363
364 **5. Conclusions**
365 We presented a direct comparison of fraternal and identical twin approaches for
366 assessing data assimilation impact in an EnKF-based ocean DA system for Gulf of Mexico.

https://doi.org/10.5194/os-2019-85
Preprint. Discussion started: 6 August 2019

To the best of our knowledge, this is the first direct comparison of fraternal and identical twin approaches for an oceanic DA system and first demonstration of how the identical twin approach can yield misleading assessments in practice. Our comparisons show that the identical twin approach overestimates the improvement in model skill resulting from assimilating SSH and SST, including for the subsurface circulation, while underestimating the value of additional information from temperature and salinity profiles. In the context of observing system design, such biased assessments are problematic and can lead to misguided decisions on balancing investments between different observing assets. We conclude that skill assessments and OSSEs from identical twin experiments should be avoided or, at least, regarded with caution. While the fraternal twin approach is more robust, questions remain about how to best choose a credible framework. In our case, the rate of error growth in SSH appears to have been an insufficient criterion.

https://doi.org/10.5194/os-2019-85
Preprint. Discussion started: 6 August 2019

379 **Code and data availability.** The ROMS model code can be accessed at
380 http://www.myroms.com (last access: 16 June 2016). ROMS data assimilation model
381 outputs are publicly available through the Gulf of Mexico Research Initiative Information
382 & Data Cooperative (GRIIDC) at https://data.gulfresearchinitiative.org/
383 data/R5.x275.000:0009. HYCOM data can be downloaded at
384 http://tds.hycom.org/thredds/catalog.html (last access: 9 July 2019).
385
386 **Author contributions.** LY and KF conceived the study. LY carried out the model
387 simulations and analysis. BW assisted in preparing the HYCOM data and validating the
388 free model run. AL, KT and LS provided inputs to the model setup and data assimilation
389 techniques. LY and KF discussed the results and wrote the manuscript with contributions
390 from all co-authors.
391
392 **Competing interests.** The authors declare that they have no conflict of interest.
393
394 **Acknowledgements.** This research was made possible in part by a grant from The Gulf of
395 Mexico Research Initiative (GoMRI-V-487). Data are publicly available through the Gulf
396 of Mexico Research Initiative Information & Data Cooperative (GRIIDC) at
397 https://data.gulfresearchinitiative.org/ data/R5.x275.000:0009. This work used the Extreme
398 Science and Engineering Discovery Environment (XSede) Comet at the San Diego
399 Supercomputer Center through allocation (TG-OCE170001). LY also acknowledges
400 support from the Nova Scotia Graduate Fellowship program and thanks Thyng et al. 2016
401 for making the “cmocean” color schemes available.

https://doi.org/10.5194/os-2019-85
Preprint. Discussion started: 6 August 2019

References:

Anderson, D. L. T., Sheinbaum, J., and Haines, K.: Data assimilation in ocean models, Rep. Prog. Phys., 59, 1209–1266, 1996.

Anderson, J. L., and Anderson, S. L.: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts, Mon. Weather Rev., 127(12), 2741–2758, https://doi.org/10.1175/1520-0493(1999)127<2741:AMCIOT>2.0.CO;2, 1999.

Arnold, C.P., Dey, C.H.: Observing-Systems Simulation Experiments : Past, Present, and Future, Bull. Am. Meteorol. Soc., 67, 687–695, 1986.

Atlas, R.: Atmospheric observations and experiments to assess their usefulness in data assimilation, J. Meteorol. Soc. Japan, 75, 111–130, 1997.

Berta, M., Griffa, A., Magaldi, M. G., Ozgokmen, T. M., Poje, A. C., Haza, A. C., and Josefina Olascoaga, M.: Improved surface velocity and trajectory estimates in the Gulf of Mexico from blended satellite altimetry and drifter data, J. Atmos. Ocean Tech., 32(10), 1880–1901. https://doi.org/10.1175/JTECH-D-14-00226.1, 2015.

Carrier, M. J., Ngodock, H., Smith, S., and Jacobs, G.: Impact of assimilating ocean velocity observations inferred from Lagrangian drifter data using the NCOM-4DVAR, Mon. Weather Rev., 142(4), 1509–1524. https://doi.org/10.1175/MWR-D-13-00236.1, 2014.

Chassignet, E. P., Hurlburt, H. E., Metzger, E. J., Smedstad, O., Cummings, J., Halliwell, G., Bleck, R., Baraille, R., Wallcraft, A. J., Lozano, C., Tolman, H.L., Srinivasan, A., Hankin, S., Cornillon, P., Weisberg, R., Barth, A., He, R., Werner, F., Wilkin, J.: US GODAE: Global ocean prediction with the HYbrid Coordinate Ocean Model

https://doi.org/10.5194/os-2019-85
Preprint. Discussion started: 6 August 2019

425 (HYCOM), Oceanography, 22(2), 64–75. https://doi.org/10.5670/oceanog.2009.39,
426 2009.
427 Chassignet, E. P., Hurlburt, H. E., Smedstad, O. M., Barron, C. N., Ko, D. S., Rhodes, R.
428 C., Shriver, J. F., Wallcraft, A. J., Arnone, R.: Assessment of data assimilative ocean
429 models in the Gulf of Mexico using ocean color, in: Circulation in the Gulf of Mexico:
430 Observations and Models, edited by: Sturges, W., and Lugo-Fernández, A.,
431 Geophysical Monograph Series (Vol. 161, pp. 87–100). Washington, DC: American
432 Geophysical Union, 2005.
433 Counillon, F., and Bertino, L.: Ensemble Optimal Interpolation: Multivariate properties in
434 the Gulf of Mexico, Tellus, Series A: Dynamic Meteorology and Oceanography, 61(2),
435 296–308. https://doi.org/10.1111/j.1600-0870.2008.00383.x, 2009a.
436 Counillon, F., and Bertino, L.: High-resolution ensemble forecasting for the Gulf of Mexico
437 eddies and fronts, Ocean Dynam., 59(1), 83–95. https://doi.org/10.1007/s10236-008-
438 0167-0, 2009b.
439 Evensen, G.: The Ensemble Kalman Filter: Theoretical formulation and practical
440 implementation, Ocean Dynam., 53(4), 343–367. https://doi.org/10.1007/s10236-003-
441 0036-9, 2003.
442 Fennel, K., Gehlen, M., Brasseur, P., Brown, C.W., Ciavatta, S., Cossarini, G., Crise, A.,
443 Edwards, C.A., Ford, D., Friedrichs, M.A.M., Gregoire, M., Jones, E., Kim, H.-C.,
444 Lamouroux, J., Murtugudde, R., Perruche, C.: Advancing Marine Biogeochemical and
445 Ecosystem Reanalyses and Forecasts as Tools for Monitoring and Managing
446 Ecosystem Health, Front. Mar. Sci., 6, 1–9. https://doi.org/10.3389/fmars.2019.00089,
447 2019

https://doi.org/10.5194/os-2019-85
Preprint. Discussion started: 6 August 2019

448 Friedrichs, M.A.M.: A data assimilative marine ecosystem model of the central equatorial
449 Pacific: Numerical twin experiments, J. Mar. Res. 59, 859–894.
450 https://doi.org/10.1357/00222400160497544, 2001.
451 Haidvogel, D. B., Arango, H., Budgell, W. P., Cornuelle, B. D., Cur- chitser, E., Di Lorenzo,
452 E., Fennel, K., Geyer, W. R., Hermann, A. J., Lanerolle, L., Levin, J., McWilliams, J.
453 C., Miller, A. J., Moore, A. M., Powell, T. M., Shchepetkin, A. F., Sherwood, C. R.,
454 Signell, R. P., Warner, J. C., and Wilkin, J.: Ocean forecasting in terrain-following
455 coordinates: formulation and skill assessment of the regional ocean modeling system,
456 J. Comput. Phys., 227, 3595–3624, 2008.
457 Halliwell, G. R., Kourafalou, V., Le Hénaff, M., Shay, L. K., and Atlas, R.: OSSE impact
458 analysis of airborne ocean surveys for improving upper-ocean dynamical and
459 thermodynamical forecasts in the Gulf of Mexico, Prog. Oceanogr., 130, 32–46.
460 https://doi.org/10.1016/j.pocean.2014.09.004, 2015.
461 Halliwell, G. R., Srinivasan, A., Kourafalou, V., Yang, H., Willey, D., Le Hénaff, M., and
462 Atlas, R.: Rigorous evaluation of a fraternal twin ocean OSSE system for the open
463 Gulf of Mexico, J. Atmos. Ocean Tech., 31(1), 105–130.
464 https://doi.org/10.1175/JTECH-D-13-00011.1, 2014.
465 Halliwell, G.R., Mehari, M.F., Le Hénaff, M., Kourafalou, V.H., Androulidakis, I.S., Kang,
466 H.S., Atlas, R.: North Atlantic Ocean OSSE system: Evaluation of operational ocean
467 observing system components and supplemental seasonal observations for potentially
468 improving tropical cyclone prediction in coupled systems, J. Oper. Oceanogr., 10,
469 154–175. https://doi.org/10.1080/1755876X.2017.1322770, 2017.
470 Hoffman, R.N., Atlas, R.: Future observing system simulation experiments. Bull. Am.

https://doi.org/10.5194/os-2019-85
Preprint. Discussion started: 6 August 2019

471 Meteorol. Soc. 97, 1601–1616. https://doi.org/10.1175/BAMS-D-15-00200.1, 2016.
472 Hoteit, I., Hoar, T., Gopalakrishnan, G., Collins, N., Anderson, J., Cornuelle, B., Kohl, A.,
473 Heimbach, P.: A MITgcm/DART ensemble analysis and prediction system with
474 application to the Gulf of Mexico, Dynam. Atmos. Oceans, 63, 1–23.
475 https://doi.org/10.1016/j.dynatmoce.2013.03.002, 2013.
476 Hu, J., Fennel, K., Mattern, J. P., and Wilkin, J.: Data assimilation with a local Ensemble
477 Kalman Filter applied to a three-dimensional biological model of the Middle Atlantic
478 Bight. J. Marine Syst., 94, 145–156. https://doi.org/10.1016/j.jmarsys.2011.11.016,
479 2012.
480 Jacobs, G. A., Bartels, B. P., Bogucki, D. J., Beron-Vera, F. J., Chen, S. S., Coelho, E. F.,
481 Curcic, M., Griffa, A., Gough, M., Haus, B. K., Haza, A. C., Helber, R. W., Hogan, P.
482 J., Huntley, H. S., Iskandarani, M., Judt, F., Kirwan, A. D., Laxague, N., Valle-
483 Levinson, A., Lipphardt, B. L., J. Mariano, A., Ngodock, H. E., Novelli, G., Olascoaga,
484 M. J., Özgökmen, T. M., Poje, A. C., Reniers, A. J.H.M., Rowley, C. D., Ryan, E. H.,
485 Smith, S. R., Spence, P. L., Thoppil, P. G., Wei, M.: Data assimilation considerations
486 for improved ocean predictability during the Gulf of Mexico Grand Lagrangian
487 Deployment (GLAD), Ocean Model., 83, 98–117.
488 https://doi.org/10.1016/j.ocemod.2014.09.003, 2014.
489 Jones, E. M., Baird, M. E., Mongin, M., Parslow, J., Skerratt, J., Margvelashvili, N., Matear,
490 R. J., Wild-Allen, K., Robson, B., Rizwi, F., Oke, P., King, E., Schroeder, T., Steven,
491 A., Taylor, J.: Use of remote-sensing reflectance to constrain a data assimilating
492 marine biogeochemical model of the Great Barrier Reef, Biogeosciences, 13, 6441–
493 6469. https:// doi.org/10.5194, 2016.

https://doi.org/10.5194/os-2019-85
Preprint. Discussion started: 6 August 2019

494 Ledwell, J. R., He, R., Xue, Z., DiMarco, S. F., Spencer, L. J., and Chapman, P.: Dispersion
495 of a tracer in the deep Gulf of Mexico, J. Geophys. Res.-Oceans, 121, 1110–
496 1132. doi:10.1002/2015JC011405, 2016.
497 Li, G., Iskandarani, M., Hénaff, M. Le, Winokur, J., Le Maître, O. P., and Knio, O. M.:
498 Quantifying initial and wind forcing uncertainties in the Gulf of Mexico, Computat.
499 Geosci., 1133–1153. https://doi.org/10.1007/s10596-016-9581-4, 2016.
500 Lin, X. H., Oey, L. Y., and Wang, D. P.: Altimetry and drifter data assimilations of loop
501 current and eddies, J. Geophys. Res.-Oceans, 112(5), 1–24.
502 https://doi.org/10.1029/2006JC003779, 2007.
503 Mattern, J. P., Dowd, M., and Fennel, K.: Particle filter-based data assimilation for a three-
504 dimensional biological ocean model and satellite observations, J. Geophys. Res.-
505 Oceans, 118, 2746–2760. https://doi.org/10.1002/jgrc.20213, 2013.
506 Mellor, G. L., and Ezer, T.: A gulf stream model and an altimery assimilation scheme. J.
507 Geophys. Res.-Oceans, 96, 8779–8795, 1991.
508 Muscarella, P., Carrier, M. J., Ngodock, H., Smith, S., Lipphardt, B. L., Kirwan, A. D., and
509 Huntley, H. S.: Do assimilated drifter velocities improve Lagrangian predictability in
510 an operational ocean model? Mon. Weather Rev., 143(5), 1822–1832.
511 https://doi.org/10.1175/MWR-D-14-00164.1, 2015.
512 Oey, L.-Y., and Lee, H.-C.: Effects of winds and Caribbean eddies on the frequency of
513 Loop Current eddy shedding: A numerical model study, J. Geophys. Res.-Oceans,
514 108(C10), 3324. https://doi.org/10.1029/2002JC001698, 2003.
515 Oey, L.-Y., Ezer, T., Lee, H.-C.: Loop Current, rings and related circulation in the Gulf of
516 Mexico: a review of numerical models and future challenges, in: Circulation in the

https://doi.org/10.5194/os-2019-85
Preprint. Discussion started: 6 August 2019

517 Gulf of Mexico: Observations and Models, edited by: Sturges, W., and Lugo-
518 Fernández, A., Geophysical Monograph Series (Vol. 161, pp. 87–100). Washington,
519 DC: American Geophysical Union, 2005.
520 Oke, P. R., O'Kane, T.J. (Eds.): Observing system design and assessment. Operational
521 Oceanography in the 21st Century. Springer, Netherlands, pp. 123–151, 2011.
522 Sakov, P., and Oke, P. R.: A deterministic formulation of the ensemble Kalman filter: an
523 alternative to ensemble square root filters, Tellus A, 60(2), 361–371.
524 https://doi.org/10.1111/j.1600-0870.2007.00299.x, 2008.
525 Shay, L. K., Jaimes, B., Brewster, J. K., Meyers, P., McCaskill, E. C., Uhlhorn, E., Marks,
526 F., Halliwell Jr., G. R., Smedstad, O. M., and Hogan, P.: Airborne ocean surveys of
527 the Loop Current complex from NOAA WP-3D in support of the Deepwater Horizon
528 oil spill, in: Monitoring and Modeling the Deepwater Horizon Oil Spill: A Record-
529 Breaking Enterprise, edited by: Liu, Y., Macfadyen, A., Ji, Z.-G., and Weisberg, R.
530 H., Geophysical Monograph Series (Vol. 195, pp. 131-152). Washington, DC:
531 American Geophysical Union. doi:10.1029/2011GM001101, 2011.
532 Simon, E., and Bertino, L.: Application of the Gaussian anamorphosis to assimilation in a
533 3-D coupled physical-ecosystem model of the North Atlantic with the EnKF: a twin
534 experiment, Ocean Sci., 5, 495-510, https://doi.org/10.5194/os-5-495-2009, 2009.
535 Simon, E., Samuelsen, A., Bertino, L., and Mouysset, S.: Experiences in multiyear
536 combined state-parameter estimation with an ecosystem model of the North Atlantic
537 and Arctic Oceans using the Ensemble Kalman Filter. J. of Marine Syst., 152, 1–17.
538 https://doi.org/10.1016/j.jmarsys.2015.07.004, 2015.
539 Song, H., Edwards, C. A., Moore, A. M., and Fiechter, J.: Data assimilation in a coupled

https://doi.org/10.5194/os-2019-85
Preprint. Discussion started: 6 August 2019

physical-biogeochemical model of the California Current System using an incremental lognormal 4-dimensional variational approach: Part 2-Joint physical and biological data ssimilation twin experiments. Ocean Model., 106, 146–158. https://doi.org/10.1016/j.ocemod.2016.04.001, 2016.

Srinivasan, A., Chassignet, E. P., Bertino, L., Brankart, J.-M., Brasseur, P., Chin, T. M., Counillon, F., Cummings, J. A., Mariano, A. J., Smedstad, O. M., and Thacker, W. C.: A comparison of sequential assimilation schemes for ocean prediction with the HYbrid Coordinate Ocean Model (HYCOM): Twin experiments with static forecast error covariances. Ocean Model., 37(3–4), 85–111. https://doi.org/10.1016/j.ocemod.2011.01.006, 2011.

Thacker, W. C., Srinivasan, A., Iskandarani, M., Knio, O. M., and Hénaff, M. Le.: Propagating boundary uncertainties using polynomial expansions, Ocean Model., 43–44, 52–63. https://doi.org/10.1016/j.ocemod.2011.11.011, 2012.

Thyng, K. M., Greene, C. A., Hetland, R. D., Zimmerle, H. M., and DiMarco, S. F.: True colors of oceanography: Guidelines for effective and accurate colormap selection. Oceanography, 29(3), 9013. https://doi.org/10.5670/oceanog.2016.66, 2016

Wang, D.-P., Oey, L.-Y., Ezer, T., and Hamilton, P.: Near-surface currents in DeSoto Canyon (1997–99): comparison of current meters, satellite observation, and model simulation, J. Phys. Oceanogr., 33(1), 313–326. https://doi.org/10.1175/1520-0485(2003)033<0313:NSCIDC>2.0.CO;2, 2003.

Wu, H., and Zhu, J.: Advection scheme with 3rd high-order spatial interpolation at the middle temporal level and its application to saltwater intrusion in the Changjiang Estuary. Ocean Model., 33(1–2), 33–51.

https://doi.org/10.5194/os-2019-85
Preprint. Discussion started: 6 August 2019

563 https://doi.org/10.1016/j.ocemod.2009.12.001, 2010.

564 Yu, L., Fennel, K., Bertino, L., El, M., and Thompson, K. R.: Insights on multivariate
565 updates of physical and biogeochemical ocean variables using an Ensemble Kalman
566 Filter and an idealized model of upwelling. Ocean Model., 126, 13–28.
567 https://doi.org/10.1016/j.ocemod.2018.04.005, 2018.

568 Xue, Z., He, R., Fennel, K., Cai, W. J., Lohrenz, S., and Hopkinson, C.: Modeling ocean
569 circulation and biogeochemical variability in the Gulf of Mexico. Biogeosciences,
570 10(11), 7219–7234. https://doi.org/10.5194/bg-10-7219-2013, 2013.

https://doi.org/10.5194/os-2019-85
Preprint. Discussion started: 6 August 2019

571 **Table 1.** Mean Absolute Deviation (MAD) from the "truth" of physical variables for free
572 and data assimilation runs in fraternal twin and identical experiments. The MAD were
573 averaged over all grid cells excluding the shelves (defined by water depths < 300 m) and
574 daily snapshots from 1 April to 1 October 2010. At assimilation steps the forecast ensemble
575 mean was used for the calculation. The percentage change relative to the free run is
576 presented in parentheses.

577

	SSH (cm)	T (°C)	S	U (m/s)
Fraternal twin				
Free	11	0.72	0.15	0.21
F1 (satellite only)	5.3 (-51%)	0.51 (-29%)	0.13 (-11%)	0.16 (-25%)
F2 (satellite and scheme 1)	5.3 (-52%)	0.50 (-30%)	0.13 (-13%)	0.16 (-25%)
F3 (satellite and scheme 2)	5.4 (-51%)	0.48 (-33%)	0.11 (-23%)	0.16 (-26%)
Identical twin				
Free	10	0.58	0.093	0.20
I1 (satellite only)	4.2 (-59%)	0.32 (-45%)	0.073 (-21%)	0.11 (-46%)
I2 (satellite and scheme 1)	4.1 (-60%)	0.31 (-47%)	0.072 (-23%)	0.11 (-47%)
I3 (satellite and scheme 2)	4.4 (-57%)	0.29 (-50%)	0.068 (-27%)	0.11 (-46%)

578

https://doi.org/10.5194/os-2019-85
Preprint. Discussion started: 6 August 2019

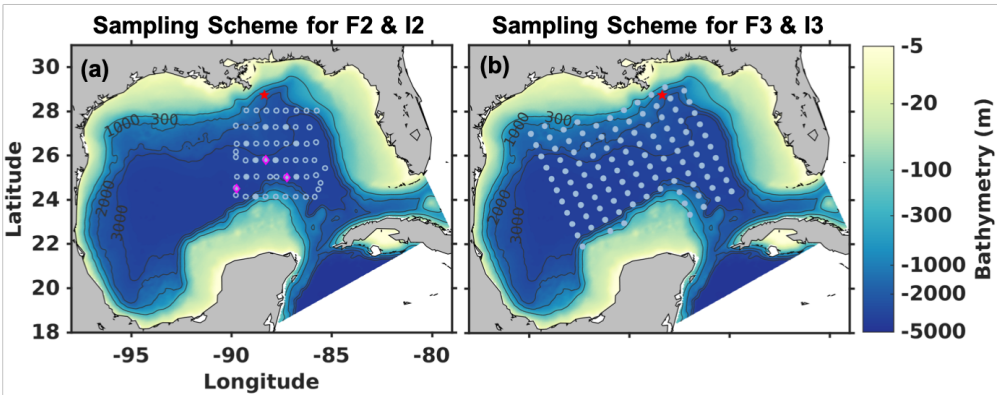
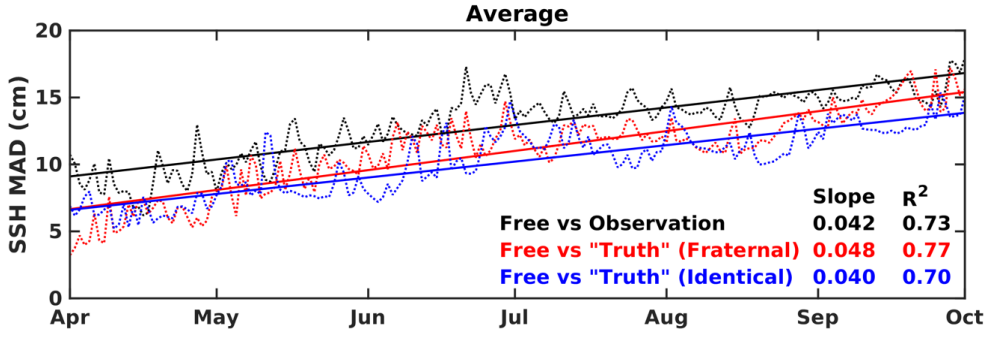


Fig. 1. Model domain and bathymetry. The red star denotes the location of the DwH oil rig. (a) Sampling scheme for twin experiments F2 and I2. The symbols represent stations where temperature (circles) and salinity (magenta diamonds) profiles were collected by Shay et al. (2011), with deep temperature or salinity profiles (down to 1000 m) marked as filled circles or magenta diamonds and shallow temperature profiles (down to 400 m) as open circles. (b) Sampling scheme for F3 and I3. The dots represent stations where temperature and salinity profiles extending to 1000 m depth were sampled from the "truth" run.

https://doi.org/10.5194/os-2019-85
Preprint. Discussion started: 6 August 2019

Ocean Science
Discussions
Open Access
EGU



587

588 **Fig. 2.** Time series of MAD error (cm) averaged over the open Gulf (excluding shelf regions
589 shallower than 300 m) for free run's SSH in relative to the SSH from the satellite
590 observation (black dashed line), the "truth" in the fraternal (red) and identical (blue) twin
591 experiments, respectively. The corresponding colored solid lines are linear regressions of
592 the time series, where the slope values represent the respective MAD error growth rate in
593 unit of cm/day.

https://doi.org/10.5194/os-2019-85
Preprint. Discussion started: 6 August 2019

Ocean Science
Discussions
Open Access
EGU

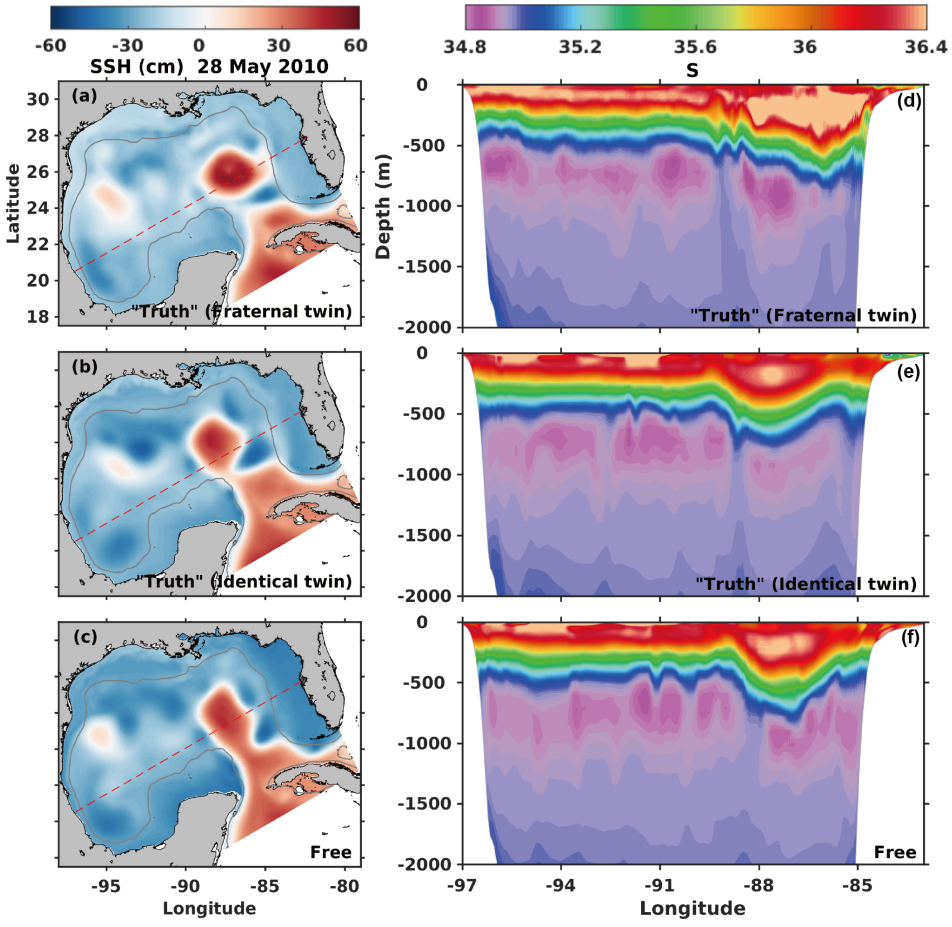
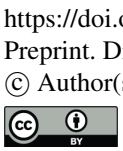
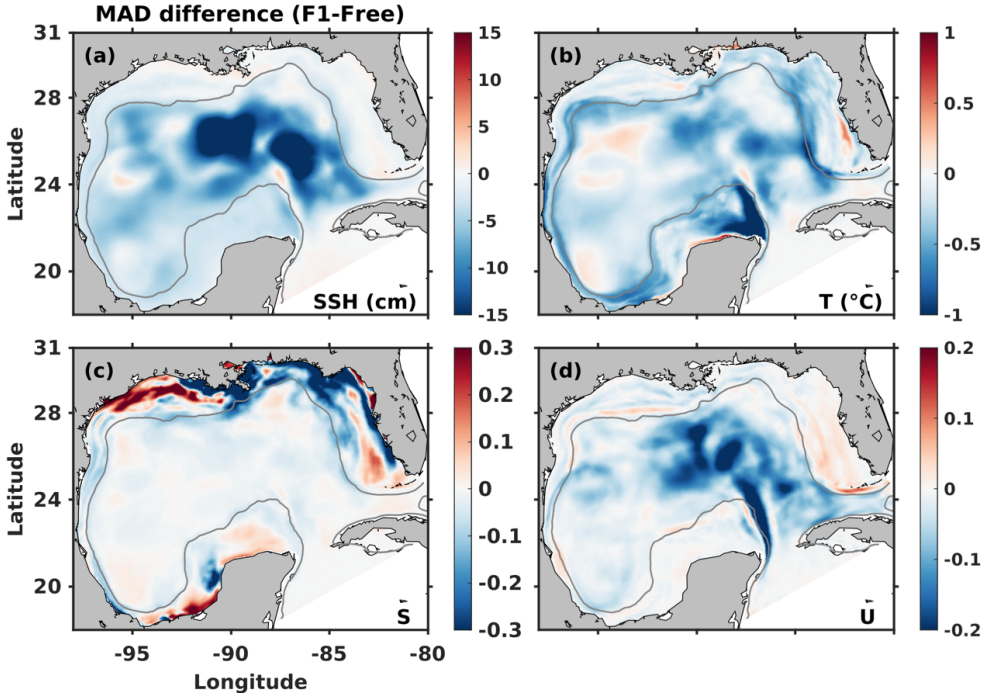


Fig. 3. Sea surface height (SSH, cm) and transect of salinity (S) on 28 May 2010. Panels (a) and (d) are from HYCOM and used as the "truth" in the fraternal twin experiments. Panels (b) and (e) are from ROMS and used as "truth" in identical twin experiments. Panels (c) and (f) are from the free ROMS run. The gray contour in the SSH maps marks the bathymetric depth of 300 m, and the red dashed line shows the position of the transect in panels (d-f).

https://doi.org/10.5194/os-2019-85
Preprint. Discussion started: 6 August 2019



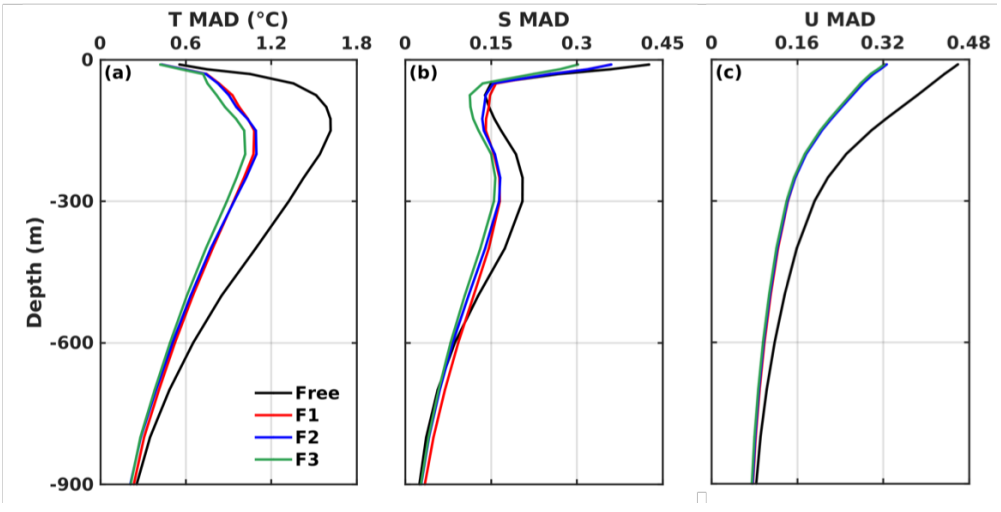


601

602 **Fig. 4.** The difference of physical variables' time-averaged (daily snapshots from 1 April
603 to 1 October) MAD between fraternal twin F1 and the free run. MAD of temperature and
604 velocity were averaged over the entire water column. Negative values (cold colors)
605 correspond to a decrease in MAD compared to free run, whereas positive values (warm
606 colors) correspond to an increase. The gray contour marks the bathymetric depth of 300 m.

607

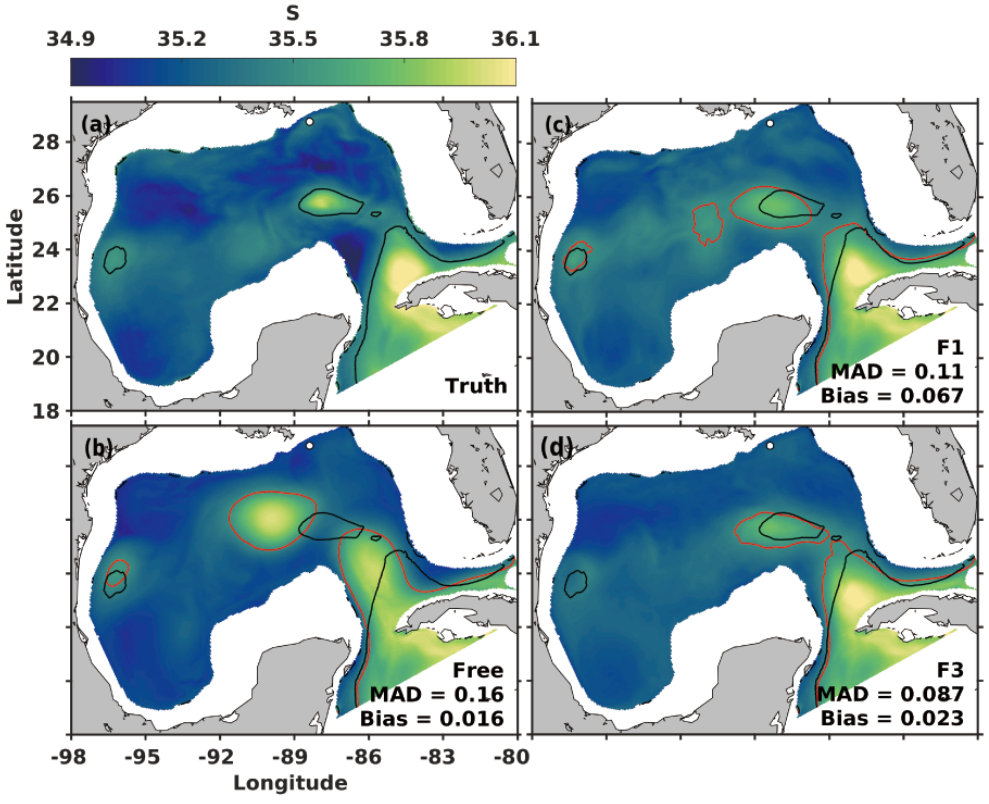
https://doi.org/10.5194/os-2019-85
Preprint. Discussion started: 6 August 2019



608

609 **Fig. 5.** Profiles of MAD averaged over the open Gulf (excluding shelf regions shallower
610 than 300 m) and daily snapshots from 1 April to 1 October 2010 for (a) temperature (T, ℃),
611 (b) salinity (S), and (c) velocity (U, m/s) from the free run and the fraternal twin runs.

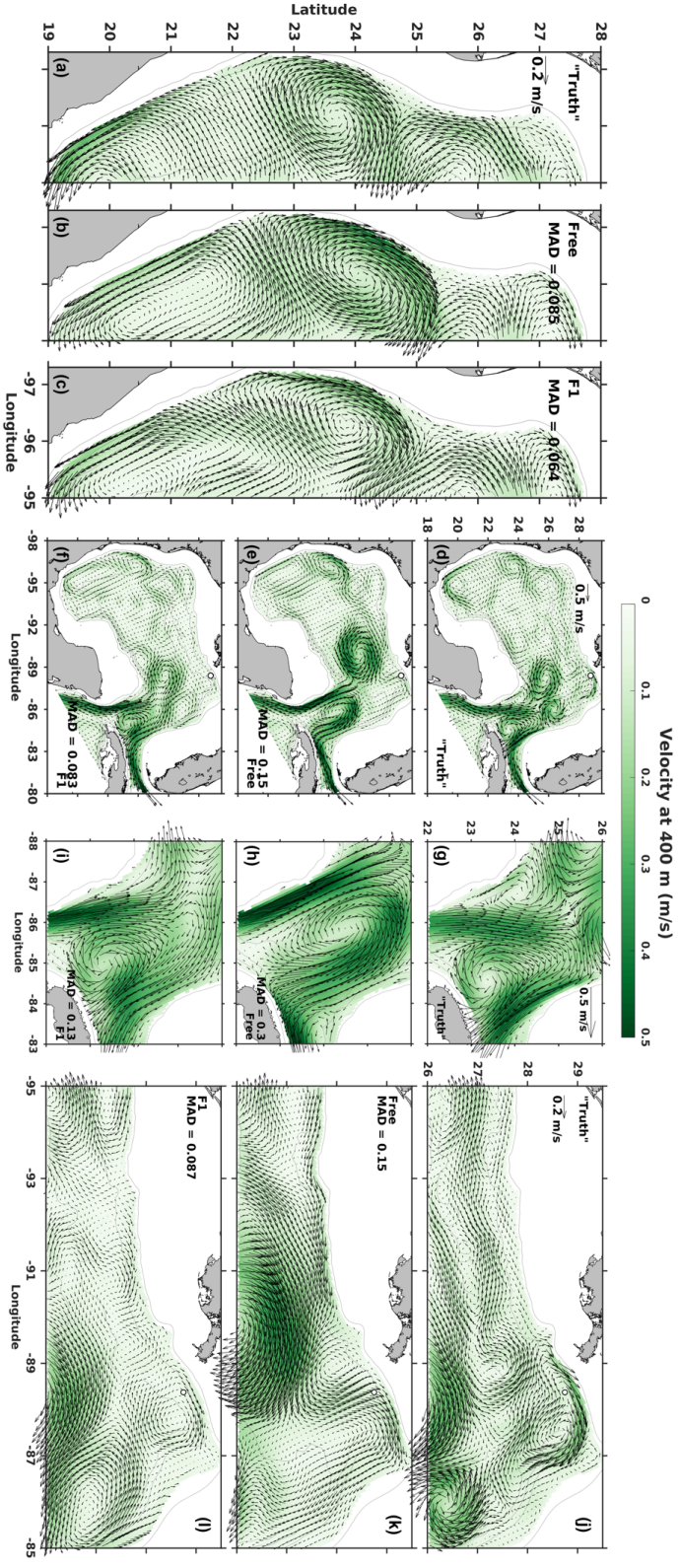
https://doi.org/10.5194/os-2019-85
Preprint. Discussion started: 6 August 2019



612

613 **Fig. 6.** August-mean salinity (S) at 400 m from the (a) "Truth", (b) Free, (c) F1 and (d) F3
614 run in fraternal twin experiments. The white dot denotes the location of the Deepwater
615 Horizon oil rig. The contours mark the 35.5 isohaline, where the black contours denote the
616 isotherm or isohaline for the "truth" while red contours denote those for the actual
617 simulation in each panel. The horizontal domain averaged MAD and Bias values at 400 m
618 for each experiment in relative to the "truth" are also presented in respective panel.

https://doi.org/10.5194/os-2019-85
Preprint. Discussion started: 6 August 2019



https://doi.org/10.5194/os-2019-85
Preprint. Discussion started: 6 August 2019

620 **Fig. 7.** August-mean velocity at 400 m in the (a, d, g, j) "truth", (b, e, h, k) free and (c, f, I,
621 l) F1 run in fraternal twin experiments. Panels in the 1st, 3rd and 4th columns are zoomed into
622 the western shelf, central Gulf, and norther shelf, respectively. The white dot denotes the
623 location of the DwH oil rig, and gray contours mark the bathymetric depths of 300, 1000,
624 2000 and 3000 m, respectively.