North Atlantic Ocean OSSE System: Evaluation of Operational Ocean Observing System Components and Supplemental Seasonal Observations for Potentially Improving Tropical Cyclone Prediction in Coupled Systems

George R. Halliwell, Jr. (corresponding author): NOAA/AOML/PhOD, 4301 Rickenbacker Cswy, Miami, FL; 1-305-361-4346; george.halliwell@noaa.gov

Michael F. Mehari: CIMAS, University of Miami and NOAA/AOML/PhOD, 4301 Rickenbacker Cswy, Miami, FL; 1-305-361-4323; michael.f.mehari@noaa.gov

Matthieu Le Hénaff: CIMAS, University of Miami and NOAA/AOML/PhOD, 4301 Rickenbacker Cswy, Miami, FL; 1-305-361-4323; matthieu.lehenaff@noaa.gov

Villy H. Kourafalou: RSMAS, University of Miami, 4600 Rickenbacker Cswy, Miami, FL; 1-305-421-4905; vkourafalou@rsmas.miami.edu

Ioannis S. Androulidakis: RSMAS, University of Miami, 4600 Rickenbacker Cswy, Miami, FL; 1-305-421-4905; iandroulidakis@rsmas.miami.edu

Hee-Sook Kang: RSMAS, University of Miami, 4600 Rickenbacker Cswy, Miami, FL; 1-305-421-4648; hkang@rsmas.miami.edu

Robert M. Atlas: NOAA/AOML, 4301 Rickenbacker Cswy, Miami. FL; 1-305-361-4300; robert.atlas@noaa.gov

Journal of Operational Oceanography

(revised 11 April 2017)

Abstract. Observing System Simulated Experiments (OSSEs) performed during the 2014 North Atlantic hurricane season quantify ocean observing system impacts with respect to improving ocean model initialization in coupled tropical cyclone (TC) prediction systems. The suitability of the OSSE system Forecast Model (FM) with respect to the previously-validated Nature Run (NR) is demonstrated first. Analyses are then performed to determine the calibration required to obtain credible OSSE impact assessments. Impacts on errors and biases in fields important to TC prediction are first quantified for three major components of the existing operational ocean observing system. Satellite altimetry provides the greatest positive impact, followed by Argo floats and sea surface temperature (SST) measurements from both satellite and in-situ systems. The OSSE system is then used to investigate observing system enhancements, specifically regional underwater glider deployments during the 2014 hurricane season. These deployments resulted in modest positive impacts on ocean analyses that were limited by (1) errors in the horizontal structure of the increment field imposed by individual gliders and (2) memory loss in the spreading of these corrections by nonlinear model dynamics. The high-resolution, three-dimensional representation of the truth available in OSSE systems allows these issues to be studied without high-density ocean observations.

Key words: OSSE, ocean data assimilation, tropical cyclones

1. Introduction

To improve tropical cyclone (TC) prediction, the next-generation Hurricane Weather Research and Forecasting (HWRF) model presently being tested by the U. S. National Oceanic and Atmospheric Administration Environmental Modelling Center (NOAA/EMC) will use a data-assimilative global ocean analysis product for initialization. The specific product used is the NOAA Real-Time Ocean Forecast System (RTOFS; Mehra and Rivin, 2010) which is presently initialized from the U. S. Navy Global Ocean Forecast System (GOFS; Chassignet et al., 2007; Cummings and Smedstad, 2013). Furthermore, the U.S.

Navy Coupled Ocean-Atmosphere Prediction System for TC forecasting (COAMPS-TC; Doyle et al., 2014) model recently implemented ocean coupling, with the ocean model initialized directly by the GOFS product. These developments demonstrate that global and regional ocean prediction systems (Tonani et al., 2015) that assimilate the operational ocean observing system (Legler et al., 2015; Schiller et al., 2015) will play an important future role in TC forecasting. The present study specifically addresses strengths and limitations of major components of the current operational ocean observing system, and begins to address future ocean observing requirements for improved TC prediction by coupled models.

Accurate forecasts of TC intensity evolution by coupled prediction systems require accurate prediction of the enthalpy flux that provides energy from the ocean to the storm. Halliwell et al. (2008; 2011) demonstrated that errors in ocean model initialization make a leading-order contribution to errors in predicted sea surface temperature (SST) cooling that then degrades predicted enthalpy flux. One contributing factor is that inaccurate initial upper-ocean stratification degrades the performance of model vertical mixing parameterizations that control entrainment of colder water into the ocean mixed layer. Another contributing factor is inaccurate representation of pre-storm upper-ocean heat content. Strong TCs can force maximum SST cooling of up to several degrees Celsius over regions of low heat content where the seasonal thermocline is located close to the surface and cold water can be rapidly entrained into the mixed layer. By contrast, these storms force cooling of <1ºC in regions with high heat content associated with deep seasonal thermoclines (e.g. Shay and Black, 2000). Energetic ocean eddies and boundary currents are often associated with large horizontal differences in heat content while dynamical processes associated with these features significantly alter the forced SST cooling pattern (Jacob and Shay, 2003; Jaimes and Shay, 2009; 2010; Jaimes et al., 2011). Initialization of Hurricane Gonzalo (2014) HWRF forecasts by data-assimilative ocean analyses substantially improved predicted intensity

compared to initialization by an unconstrained ocean analysis (Dong et al., 2017) For all of these reasons, the ocean component of coupled prediction systems must be initialized as accurately as possible to produce credible surface enthalpy flux and storm intensity forecasts.

Although it is known that the existing operational ocean observing system significantly reduces errors in operational ocean analysis products (e.g. Cummings and Smedstad, 2014; Oke et al., 2015), a quantitative assessment of the importance of these observations specifically with respect to the TC prediction problem does not exist. Furthermore, prior assessments have often relied on Observing System Experiments (OSEs) which are based on performing twin experiments where one assimilates all observations and the second denies the observing system being evaluated. Because OSEs evaluate real observing systems, some observations must always be withheld for evaluation. Alternatively, linearized adjoint methods have also been used to evaluate the impact of actual observations (e.g. Cummings and Smedstad, 2014) using the analysis to represent the truth. By contrast, Observing System Simulation Experiments (OSSEs) perform twin experiments using synthetic observations simulated from a free-running, high resolution Nature Run (NR). The NR is an unconstrained model run that has been validated as a representation of the "true" ocean through rigorous evaluation in comparison to observations. The high-resolution, three-dimensional "truth" provided by the NR permits more detailed evaluations to be conducted than are possible with OSEs. Consequently, a rigorously validated and calibrated OSSE system can be used not only to evaluate the impact of future observing systems or alternate deployment strategies for existing systems, but also to provide a more thorough evaluation of existing observing systems.

The present analysis uses a new ocean OSSE system recently expanded to cover the North Atlantic hurricane region (Figure 1) to specifically assess observing system impacts with respect to improving ocean model initialization for TC prediction. Evaluation and

validation of the system NR with respect to representing the "truth" has already been conducted (Kourafalou et al., 2016; Androulidakis et al., 2016). Results from the two additional steps required for OSSE system validation and calibration are presented herein prior to describing the impact assessments.

The ocean OSSE system design and observing system evaluation methodology are described first (Section 2) followed by results from the two additional validation/calibration steps (Section 3). Three primary components of the existing operational ocean observing system (satellite altimetry, Argo floats, and both satellite and in-situ SST) are then evaluated in Section 4. Strategies are then examined in Section 5 for deploying additional instrumentation to augment the operational ocean observation system during hurricane season, focusing on the regional deployment of underwater gliders that collect repeat profiles of temperature and salinity. In Section 6, factors limiting the ability of individual underwater gliders (and other ocean profilers) to correct the model state are identified using the high-resolution, three-dimensional representation of the truth available from the NR. Conclusions are presented in Section 7. Although impacts with respect to the TC prediction problem are emphasized in the present analysis, aspects of this analysis are applicable to other short-term ocean prediction applications.

2. The Ocean OSSE System

Overview

The OSSE system developed by the joint AOML/CIMAS/RSMAS Ocean Modelling and OSSE Center (OMOC; http://cimas.rsmas.miami.edu/omoc.html) incorporates system components and rigorous validation methodologies required to ensure credible observing system assessments that have long been in use in the atmosphere (e.g. Arnold and Dey, 1986; Atlas, 1997; Hoffman and Atlas, 2016), but not in the ocean. A prototype system was initially validated in the Gulf of Mexico (Halliwell et al., 2014), and then used to evaluate the impact

of airborne ocean profile surveys conducted during the 2010 Deepwater Horizon oil spill for improving ocean forecast model initialization (Halliwell et al., 2015).

For the present TC-related effort, this prototype system was expanded into a larger Atlantic Ocean domain (Figure 1) so that the entire North Atlantic hurricane region was situated far from open-ocean boundaries. The NR and the Forecast Model (FM) of the fraternal-twin OSSE system use substantially different configurations, including horizontal and vertical resolution (Table 1), of the Hybrid Coordinate Ocean Model (HYCOM; Bleck, 2002; Chassignet et al., 2003; Halliwell, 2004). The FM is used to assimilate real observing systems for OSEs (which are performed herein as part of the OSSE system validation procedure) and synthetic observing systems simulated from the NR for OSSEs. The final component of the OSSE system is the toolbox that simulates ocean observing systems from the NR and adds realistic errors, including representation errors (Table 2). These errors must substantially represent the errors contained in the actual observations.

The altered configuration of the FM with respect to the NR (Table 1) was designed to degrade the performance of the FM so that ocean climatology and variability simulated by unconstrained runs of the two models are neither too similar nor too dissimilar. This is one of the basic requirements of a robust OSSE system as described in Atlas (1997) and Hoffman and Atlas (2016; see the supplemental checklist). Errors in the FM with respect to the NR must have similar magnitudes and properties as errors between the best available ocean models and the actual ocean; otherwise, biased impact assessments will be obtained. In particular, model parameters were adjusted so that the FM was more diffusive and thus behaved like an older, lower-resolution ocean model. A major configuration change from the initial Gulf of Mexico implementation is that both models employ the standard hybrid vertical coordinate structure (isopycnal, sigma and Cartesian coordinates) instead of one employing only fixed level coordinates. To make this change work, the FM uses a

substantially different vertical discretization from the NR with lower resolution and layer target densities referenced to the surface instead of 2000 dbar as used by the NR. This aggressive effort to alter the model configuration was necessary to mitigate the fact that both the NR and FM are forced by the Navy NAVGEM atmospheric model. Although three-hourly fields are used to force the NR and six-hourly fields are used to force the FM, differences are too small to produce sufficient error growth rates between the models.

The data assimilation scheme is described in Halliwell et al. (2014). State variables assimilated are layer temperature, layer salinity, and layer thickness. The Cooper and Haines (1996) method is used to assimilate satellite altimetry by correcting model vertical profiles of layer thickness so that water mass properties are preserved. All temperature-salinity profiles are assimilated after first re-mapping these profiles into the ocean model hybrid coordinate vertical discretization so that assimilation is performed on the native vertical grid. For profiles that measure only temperature, corresponding salinity profiles based on climatological correlations are generated prior to the vertical re-mapping. SST is assimilated alone within the upper two model layers, which always remain within the mixed layer.

The background error covariance for the DA system is obtained from a decade-long unconstrained simulation by the FM. Covariances are calculated separately for each month as described in Halliwell et al. (2014) to primarily resolve the mesoscale structure and not the larger horizontal scales of the seasonal cycle present in upper-ocean fields. Localization radii based on the Gaspari and Cohn (1999) taper function are applied with maximum radii of 300 km for altimetry, 100 km for SST, and 280 km for profile data. The combined influence of the background covariance field tapered by the localization function governs the structure of the correction field induced by each observation as described in Section 7.

All data-assimilative experiments are performed as described in Halliwell et al. (2014) using a daily update cycle. Consequently, no attempt is made to evaluate the impact of

observations collected within storms where a daily update cycle is inadequate due to the rapid ocean response to storm forcing. Impact assessments presented herein are therefore only applicable to error reduction over regions located ahead of approaching storms.

OSSE System Evaluation

Rigorous validation of an OSSE system must first demonstrate that the NR can be used to represent the true ocean based on realistic reproduction of mean climatology and variability. Model evaluation and scientific analysis of the NR was performed by Kourafalou et al (2016) and Androulidakis et al (2016). The NR is not expected to be accurate in all aspects; for example, it is not likely to accurately reproduce deep ocean flow. However, the evaluation demonstrated that it accurately reproduces the climatology and variability of upper-ocean fields in the North Atlantic hurricane region that are important to the TC prediction problem. In particular, SST and sea surface salinity (SSS) comparisons with satellite, ARGO and NDBC data, especially during periods and over areas of TC-ocean interaction, showed significantly high skill scores (Androulidakis et al., 2016; Kourafalou et al., 2016). The simulated surface circulation was deemed to be realistic in comparison to maps generated from surface drifters from the Global Drifter Program and also from AVISO altimetry data, while eddy kinetic eddy levels also were realistic (Kourafalou et al., 2016). The simulated upper 100 m temperature, salinity and stratification frequency, which are important for TC prediction, also showed good agreement in comparison to ARGO observations (Kourafalou et al., 2016). The NR simulation was also used to investigate specific TC-related upper-ocean applications such as the Amazon plume interaction with passing hurricanes (Androulidakis et al., 2016) and the impact of hurricanes on Gulf Stream transport over the N. Atlantic region (Kourafalou et al., 2016).

The two additional steps required for comprehensive OSSE system validation are presented herein. The first one is to demonstrate that the NR and FM models are neither too

similar nor too dissimilar. The second one is to directly evaluate the full OSSE system and quantitatively determine if calibration is necessary to realize credible observing system impact assessments (e.g. Hoffman et al., 1990; Atlas, 1997). This system evaluation is achieved by comparing OSEs and OSSEs that assimilate the identical sets of real and synthetic observations, respectively, to determine if similar impact assessments are obtained. Results of that evaluation are then used to calibrate the skill scores that are used to assess observing system impacts.

Observing System Evaluation Procedures for TC Applications

Because observing system impact assessments are made with respect to specific applications, impacts on nine ocean fields important to TC prediction are emphasized herein. To assess impacts on ocean dynamics, sea surface height (SSH), dynamic height at the surface relative to 1000 m ($D_{0\text{-}1000}$) and the depth of the 20ºC isotherm (H_{20}) are considered. In TC regions, H_{20} maps are often used as a proxy for representing the structure of upper-ocean boundary currents and eddies (e.g. Meyers et al., 2014). To assess impacts on ocean thermodynamics, two surface fields [SST and sea surface salinity (SSS)] and four subsurface fields [tropical cyclone heat potential (TCHP), depth of the 26ºC isotherm (H_{26}), mean temperature over the upper 100 m ($\overline{T}_{0-100}$), and temperature difference between the surface and 100 m (ΔT_{0-100}, included as a crude measure of stratification)] are considered.

TCHP, also referred to as Ocean Heat Content (Leipper,1972), is calculated by:

$$\mathrm{TCHP} = c_p \int_0^{H_{26}} \rho \left[T(z) - 26 \right] dz, \tag{1}$$

where c_p is specific heat of seawater at constant pressure. TCHP derived from satellite altimetry and SST (Meyers et al., 2014) is used in the Statistical Hurricane Intensity Prediction Scheme (SHIPS) at the U. S National Hurricane Center (DeMaria *et al*. 2005,

Mainelli *et al.*, 2008). TCHP serves as an index of the ocean thermal energy potentially available to maintain or intensify storms (e.g. Lin et al., 2013). The H_{26} field represents the thickness of the upper-ocean layer that is sufficiently warm to support storms.

Observing system impact assessments are based on both RMS variability errors and mean biases present in ocean analyses fields. The mean square difference and mean bias between two fields are given by

$$\begin{aligned} MSD &= \frac{1}{n}\sum_{i=1}^{n}\left(x_i - y_i\right)^2 \\ B &= \langle x \rangle - \langle y \rangle \end{aligned} . \tag{2}$$

Defining *MSD´* as the mean square difference between the two fields after removing mean values:

$$MSD' = \frac{1}{n}\sum_{i=1}^{n}\left[\left(x_i - \langle x \rangle\right) - \left(y_i - \langle y \rangle\right)\right]^2, \tag{3}$$

it can be shown that

$$MSD = MSD' + B^2. \tag{4}$$

If the analysed fields are horizontal maps, *MSD´* represents the contribution of errors in field structure while B^2 represents the contribution of bias to total *MSD*. For individual experiments, *MSD*, *MSD´*, and *B* are all calculated with respect to the truth. For the OSE-OSSE comparison experiments used for system evaluation, truth is represented by actual observations (OSE) or the identical set of synthetic observations simulated from the NR (OSSE). For OSSEs, truth is represented by fields extracted or calculated from the NR.

The impact of assimilating a particular set of observations on field variability errors is quantified in terms of improvement over a reference experiment where observations being evaluated are denied. A normalized mean square difference is calculated using

$$MSD'_{norm} = \frac{MSD'_{EXPT}}{MSD'_{REF}}, \quad (5)$$

where EXPT denotes the experiment being evaluated and REF denotes the reference experiment. An associated skill score is then calculated:

$$S' = 1 - MSD'_{norm} \quad (6)$$

By definition, *S'* is zero for REF while values exceeding 0 represent improvement in EXPT over REF. In most cases herein, REF is chosen to be the unconstrained FM.

3. OSSE System Evaluation and Calibration

Evaluation of NR and FM Model Configurations

We first stipulate that differences between unconstrained simulations by the two models represent errors between the FM and the truth represented by the NR. The different model configurations must insure that errors between the models have magnitudes and properties similar to errors that exist between the best contemporary ocean models and the actual ocean. This criterion is first tested by analysing long-term time series fields of SSH anomaly (SSHA) available from the Archiving, Validation, and Interpretation of Satellite Oceanographic (AVISO) data center (http://www.aviso.oceanobs.com), and of both TCHP and H_{26} from the AOML upper-ocean heat content product (Goni and Knaff, 2009; Goni et al., 2009). The criterion is further tested by comparing actual Argo profiles to synthetic profiles simulated from both the NR and FM at the same times and locations.

Weekly AVISO SSHA fields and weekly-averaged SSHA fields from both the NR and unconstrained FM are first compared over the six-year runs of both models. The magnitude and pattern of $RMSE' = MSD'^{1/2}$ between NR and observed SSHA are similar to the magnitude and pattern between NR and FM SSHA (Figures 2a, 2b). For upper-ocean thermodynamical variability important to TCs, daily fields of TCHP and H_{26} for six North

Atlantic hurricane seasons (July-October, 2009-2014) obtained from the AOML TCHP product (Goni and Knaff, 2009; Goni et al., 2009) are compared to daily fields from the NR and FM. For both fields, the magnitudes and patterns of *RMSE'* between the NR and observations (Figures 2c, 2e) are again similar to the magnitudes and patterns between the NR and FM (Figures 2d, 2f). However, the agreement is not as tight for TCHP because error magnitudes between the two models are somewhat smaller than those between the NR and the actual ocean.

RMS errors between unconstrained simulations by the two models for eight fields (SSH cannot be calculated from Argo profiles) are the same order of magnitude as errors between the NR and the actual ocean. Errors are larger for the two dynamical fields and smaller for the six thermodynamical fields (Table 3). Averaged over all eight fields, errors between the models are about 11% smaller than errors between the NR and the actual ocean. The criterion being evaluated is therefore substantially, but not perfectly, satisfied. Although the lack of perfection may act to degrade impact assessments, calibration derived from the OSE-OSSE comparisons described in the next section provides a means to correct for this and any other imperfections present in OSSE system design.

System Calibration

To determine if calibration is necessary, OSE-OSSE comparison experiments are analysed. These include: the unconstrained FM, the Control, three experiments that deny 1, 2, 3 altimeters respectively (but assimilate all SST and XBT), and one experiment that denies all SST (but assimilates all altimeters). The altimeters used are: Haiyang-2a, Saral-Altika and Jason-2. Each experiment is run twice, one assimilating actual observations (OSE) and the other assimilating the identical set of synthetic observations (OSSE). The unconstrained FM is run from October 2008 to December 2014 while all other experiments are run from March to October 2014. Actual Argo float profiles and synthetic float profiles simulated from the

NR at the same locations and times (with realistic errors added) are withheld from assimilation to represent the truth. Fields calculated from these actual and synthetic Argo profiles are each compared to fields calculated from corresponding Argo profiles extracted from the unconstrained FM experiment [which represents REF in (5)].

MSD'_{norm} from (5) and S' from (6) are both used in the calibration analysis because they are non-dimensional, which allows several model fields to be compared to increase the statistical significance of results. Statistics are calculated over the time interval 1 July through 31 October 2014. MSD'_{norm} is first compared between OSE-OSSE experiment pairs for several different variables and for the multiple combinations of EXPT and REF that are listed in Figure 3. The high correlation in the resulting scatter plot (Figure 3a) is encouraging, but there is a tendency for OSSEs to underestimate MSD'_{norm}; i.e. overestimate observing system impacts. Such discrepancies are expected in the OSSE framework and can arise from design flaws in the OSSE system, such as the FM and NR not exactly satisfying OSSE system requirements or from incorrect error modelling in the simulation of synthetic observations from the NR. Fortunately, the present overestimate is small and correctable by appropriate calibration (e.g. Hoffman et al, 1990). Re-design of the OSSE system is not warranted.

The skill score S' from (6) will subsequently be used to determine observing system impacts on field variability, but only after being corrected for the OSSE tendency to overestimate impacts. A corrected MSD'_{norm} is first calculated based on the linear regression fit in Figure 3a, and then used in (6) to calculate calibrated S' (S'_C). The resulting scatterplot of S'_C in Figure 3b demonstrates that the OSSE tendency to overestimate impacts has been effectively eliminated. The RMS amplitude of the residual of the linear S'_C fit shown if Figure 3a is 0.076, which is a measure of the uncertainty in any individual realization of this statistic. Because of this uncertainty, multiple realizations of this statistic should be calculated using different model variables to provide robust impact assessments.

4. Impact of Operational Ocean Observing System Components

Procedure

The experiments designed to evaluate impacts of three primary observing system components (altimetry, Argo, and SST), and also to evaluate impacts of subcomponents of the altimetry and SST systems, are listed in Table 4. The unconstrained FM simulation is the reference experiment against which error reduction resulting from data assimilation is measured. The control experiment assimilates operational ocean observing system components. This experiment is initialized on 1 March 2014 from the long unconstrained FM run to allow sufficient time for adjustment by the start of hurricane season on 1 June, and then is continued through 31 October 2014. Data distribution during July through October 2014 for the assimilated observing system components, with the exception of satellite-derived SST (obtained from the Navy MCSST product), is illustrated in Figure 4. Although the global XBT dataset is also assimilated by the control experiment, it is not evaluated herein because the sparse coverage is not designed to improve ocean analyses for short-term forecast applications. Only five XBT transects were run between July and October 2014 (Figure 4b), none in the Caribbean Sea and Gulf of Mexico (Figure 4c).

Two types of OSSE experiments are performed (Table 4). For “data denial” experiments, individual components of the ocean observing system assimilated in the control experiment are denied for evaluation, which determines their impact when added to other system components. For “individual observing system” experiments, each system is assimilated alone to measure impact in isolation. Data denial experiments determine impact based on S'_C decrease and bias increase with respect to the control experiment whereas individual observing system experiments measure impact based on positive S'_C and bias decrease with respect to the unconstrained FM. All OSSE experiments were initialized on

1 June 2014, data denial experiments from the control experiment and individual observing system experiments from the FM, and then run through 31 October 2014.

Quantitative assessments are performed over the two large subdomains shown in Figure 1, specifically the Intra-Americas Sea (IAS) and the open Atlantic Ocean outside of the IAS, to test if regional differences exist. MSD'_{norm} and B are both calculated over the two subdomains from daily model archives, and then temporally averaged over the core of the hurricane season from 1 August through 15 October 2014. The delayed start time allows two months from the start of each experiment for solutions to equilibrate. The averaged MSD'_{norm} is then corrected based on the linear fit in Figure 4a to calculate the calibrated skill score S'_C. The statistics for the data denial experiments (Figure 5) are calculated for those experiments listed in Table 4 that deny altimetry, Argo, and SST. The same statistics from the individual observing system experiments (Figure 6) are calculated for those experiments listed in Table 4 that individually assimilate the same three observing systems. In both Figures 5 and 6, S'_C and B from both the control experiment and the unconstrained FM are also included. As expected, S'_C is largest for the control experiment and zero by definition for the unconstrained FM, while B is usually largest for the unconstrained FM.

Altimetry Data Impact

Satellite altimetry denial demonstrates that it has a positive skill score impact on all five variables based on S'_C reduction (Figure 5), particularly for TCHP, ΔT_{0-100}, and SSS. Altimetry impact tends to be greater in the IAS where stronger mesoscale features are associated with larger horizontal upper-ocean heat content differences compared to the open Atlantic (e.g. Jaimes and Shay, 2009). When assimilated alone, altimetry generally has positive impact toward correcting the horizontal structure of all fields in both analysis domains based on positive S'_C values (Figure 6), with the smallest correction occurring for SSS. The correction of the H_{20} field structure is substantially larger in the IAS, again due to

the stronger mesoscale features. Altimetry is effective at correcting field structure, because horizontal scales across the mesoscale range are well resolved along track, while cross-track mesoscale structure fills in over several days due to the presence of four altimeters. Although the Cooper-Haines algorithm only corrects temperature and salinity profiles beneath the mixed layer, positive impact is still realized for surface fields SST and SSS, presumably due to the correction of horizontal mesoscale structure. In both the data denial and individual observing system experiments, altimetry assimilation does not consistently reduce biases in model fields (Figures 5 and 6). Positive bias impact is not possible because mean model SSH is arbitrary and a reference mean is not available.

The impact of the number of altimeters assimilated is presented in Figure 7. Similar results are obtained in the open Atlantic and IAS regions. As expected, assimilation of the first altimeter produces the largest impact. Denial of one altimeter, and even of two altimeters, produces very small additional skill score reduction. A minimum of two altimeters is necessary to achieve skill scores >90% of the skill present in the control experiment, while assimilating the third altimeter achieves >95% of this skill. The number of altimeters has little influence on bias reduction, which is primarily provided by other components of the ocean observing system. The primary value of the fourth altimeter, and to a substantial extent the third altimeter, is to provide the necessary redundancy to avoid significant negative impacts if one altimeter is lost.

Argo Data Impact

Denial of Argo profiles demonstrates that this system has a positive impact on correcting the structure of all fields except SST (Figure 5). Argo impact is largest for SSS which is not directly measured by the other observing system components. The assimilation of Argo profiles alone demonstrates positive impact on the horizontal structure of all fields except TCHP in both regions and SST in the IAS (Figure 6). Argo also contributes to bias reduction

in most fields in both domains. Overall, Argo has a positive impact on reducing ocean model initialization errors with respect to the TC prediction problem, but to a substantially smaller extent than altimetry due to the comparatively coarse space-time coverage that aliases the mesoscale. The inability to substantially correct SST and TCHP may result because both fields are strongly controlled by surface fluxes and upper-ocean mixing and the coarse space-time sampling by Argo is not adequate to provide substantial correction.

SST Data Impact

Denial of all satellite and in-situ SST observations demonstrates that these measurements have positive impact on the horizontal structure of only two fields: SST and TCHP (Figure 5). Large positive bias impact is evident in the same two fields, especially over the IAS region. Assimilation of SST observing systems alone produced positive impacts on the horizontal structure of only the SST and TCHP fields (Figure 6). Very large bias reduction is also realized for SST and TCHP.

Impacts of denying components of the SST observing system on SST and TCHP are presented in Figure 8. The satellite-derived MCSST product is the only one that has a detectable positive impact on SST and TCHP horizontal field structure in both the open Atlantic and IAS regions. A similar result is obtained from the corresponding individual observing system experiments (Figure 8) except that small positive impact on field structure also exists for SST over the open Atlantic due to the assimilation of ship intake SST and surface drifter SST. Satellite SST is the only component of the SST observing system that substantially corrects horizontal structure through the mesoscale range due to the high-resolution sampling, at least in cloud-free areas. Space-time coverage by the three in-situ components is inadequate for this purpose.

The situation is very different for bias reduction. Denial of all SST observing system components (Figure 8) leads to SST bias magnitudes comparable to the unconstrained FM,

demonstrating that other ocean observing system components do not significantly correct this bias. By contrast, denial of any individual component of the SST observing system does not lead to significantly larger bias magnitudes over the control experiment (Figure 8). Each individual component of the SST observing system contributes substantially to bias reduction. This result is confirmed by the individual observing system experiments (Figure 8), with satellite SST having the largest impact.

5. Underwater Glider Enhancements to the Operational Ocean Observing System

Synthetic underwater glider deployments measuring repeat profiles of temperature and salinity are conducted to demonstrate how ocean OSSEs can be used to evaluate seasonal enhancements to the existing ocean observing system. Three cases are studied which consider a single glider deployment, a regional deployment of six gliders, and a more-extensive deployment of fourteen gliders that provides extensive coverage over a large region off the Gulf of Mexico and southeastern coasts of the U. S. For each of the three glider deployment strategies investigated, two experiments are performed: one adds the glider to the control experiment and the second assimilates only the glider data (Table 5). This set of six experiments is run from 1 June to 31 October 2014, all using the same 280 km localization radius for profile assimilation as all previous experiments. The other single glider experiments in Table 5 will be discussed later.

Single Underwater Glider Deployment

The single underwater glider is deployed north of Puerto Rico, a choice motivated by the deployment of an actual glider in this general region during the 2014 hurricane season (Domingues et al. 2015; Dong et al., 2017). This and all other synthetic gliders deployed in subsequent experiments follow a simple pre-determined three-rung ladder-shaped track that covers a region of 1º longitude by 2.25º latitude (Figure 9). In all cases, gliders are deployed

at the southwestern corner of the ladder and initially track to the east, then north, then west along the middle track, then north to the top rung before zigzagging back to the initial deployment location. The synthetic gliders move at 0.25 m s^{-1} and sample eight profiles per day. The resulting profiles were located very close together compared to the ~7 km model grid point spacing. The high observation correlation among these profiles tends to degrade the least-squares problem solved during each analysis cycle. Given the daily assimilation cycle, tests determined that it was adequate to assimilate only the single daily profile closest to the analysis time.

The impact assessments in Figure 9 focus on four fields relevant to TC applications: H_{20}, SST, SSS, and TCHP. Statistics are calculated over the small box shown in the top panel of Figure 9 that contains the three-rung ladder-shaped glider track and extends three grid points outside the track in each direction. Adding a single glider to the control experiment produces a small improvement in field structure for SSS, but little improvement in the other three fields. Adding the glider also produces modest bias reduction in all four fields over the control experiment. When the single glider is assimilated alone, improvement in field structure is again realized only for SSS (Figure 9) while the glider produces large bias reduction for SST, SSS, and TCHP and modest bias reduction for H_{20}. While the bias reduction in both experiments demonstrates substantial positive impact in the vicinity of the glider, little positive impact is realized toward correcting horizontal field structure as represented by the skill score, a result that is explored in Section 6.

Regional Multiple Underwater Glider Deployment

The potential benefits of multiple underwater glider deployments are explored by releasing six synthetic gliders over a broad region north of Puerto Rico on 1 June 2014 (Figure 10). Overall, there is a slight improvement over the single glider case in the correction of H_{20} and SSS field structure, with S'_C increasing by 10-15%. When the six gliders are assimilated

alone, positive skill values of ~0.25 are realized for these same two fields. Little or no improvement is realised for SST and TCHP. Bias reduction is equal to or larger than the reduction achieved in the single glider case. Overall, multiple gliders have larger overall positive impact over the larger sampled area as expected, but still provide limited correction of ocean field structure because the 250-300 km instrument separation does not resolve the mesoscale. In the OSSE analysis of airborne ocean temperature-salinity profiles in the Gulf of Mexico by Halliwell et al. (2015), it was demonstrated that profiles should be separated by < 1 degree in both latitude and longitude to substantially resolve mesoscale structure.

A Coastal Underwater Glider "Shield"

The final case evaluates the potential impact of deploying an underwater glider "shield" to correct ocean model initialization over large regions offshore of the Gulf of Mexico and the eastern U. S. coastlines where it is critically important to improve TC intensity forecasts prior to possible landfall. A total of 8 synthetic gliders are released in the open Atlantic while six are released in the Gulf of Mexico and northwest Caribbean Sea (Figure 11). The two analysis regions in this case (Atlantic and IAS) are outlined in the left panel of Figure 11.

When added to the control experiment, the gliders have minor impact on horizontal field structure in both the Atlantic and IAS regions (Figure 11). A slight (≤10%) increase in S'_C is realized for SST and TCHP only. By contrast, a large positive impact on bias correction is realized in both regions. When assimilated alone, the gliders have a detectable positive impact on field structure except for SST and SSS in the IAS, with the largest positive impact realized for H_{20} in both regions and TCHP in the IAS. Large bias reduction is realized for all fields in both regions except for SSS in the IAS. Field structure correction is again limited by glider separation distances that do not resolve the mesoscale. However, spreading gliders out over larger areas as done here still permits modest correction of larger-scale field

structure presumably associated with gyre-scale structure, and also provides large bias correction of most model fields.

6. Factors Limiting Positive Impacts of Ocean Profile Assimilation

Overview

Factors that limit the positive impact of assimilating underwater glider data (and data from other types of ocean profilers) are now examined, taking advantage of the validated, high-resolution, three-dimensional representation of the truth available from the NR. When a single observation or profile is assimilated, the structure of the increment field that is added to the first guess to produce the analysis is governed by how innovation is spread by the background error covariance function multiplied by the Gaspari and Cohn (1999) localization radius function. "Cumulative correction" of a model field is then defined as the sum of all daily increment fields plus the spreading produced by the nonlinear model operator.

Evolution of the ocean state vector in the DA experiment is written as:

$$\mathbf{x}_{n+1} - \mathbf{x}_n = \mathbf{L}\mathbf{x}_n + \mathbf{P_b}\mathbf{H}^T \left(\mathbf{H}\mathbf{P_b}\mathbf{H}^T + \mathbf{R}\right)^{-1} \left[\mathbf{y} - \mathbf{H}\left(\mathbf{x}_n + \mathbf{L}\mathbf{x}_n\right)\right], \quad (7)$$

where $\mathbf{x}_n$ and x_{n+1} are the model state vector before and after assimilation cycle n, $\mathbf{L}$ is the nonlinear model operator, $\mathbf{P_b}$ is the background error covariance matrix, $\mathbf{y}$ is the glider observation, and $\mathbf{H}$ is the observation operator (interpolation to observation locations). Each daily update cycle is executed as follows: A one-day forecast is run where the model operator updates the state vector $\mathbf{x}_n$ to produce a first guess for the analysis. Observations are then assimilated to generate the increment vector given by the rightmost term of (7) that is added to the first guess to produce the analysis $\mathbf{x}_{n+1}$ that is subsequently used to initialize the next cycle. The cumulative correction due to glider assimilation at time n=1 exactly equals the increment field from the first assimilation cycle. With increasing time, the cumulative correction between each DA experiment and the unconstrained FM gradually diverges from

the sum of the daily increments as the nonlinear model operator **L** spreads the influence beyond the local correction region.

Three issues are examined in this context: (1) the impact of the prescribed form of the background error covariance matrix on the increment field; (2) the consequences of reducing the Gaspari and Cohn (1999) localization radius on the increment field; and (3) the spreading of cumulative correction by the nonlinear model operator. Four additional single-glider experiments are run for 30 days to address these issues (Table 5). These experiments are initialized on 10 September 2014 and run through 10 October to provide results representative of peak hurricane season. Two experiments (G30D1 and G30D2) assimilate the single glider alone using the 280 km localization radius, with the first assimilating profiles over all 30 days and the second assimilating profiles only on the first two days. One other experiment (G30D3) repeats G30D1, but uses a smaller 90 km localization radius. The final experiment G30D4 assimilates the single glider over 30 days, but adds it to the observing systems assimilated by the control experiment. Experiments G30D1 through G30D3 were initialized from the unconstrained FM, while experiment G30D4 was initialized from the control experiment.

Background Error Covariance Matrix

During each analysis cycle, the structure of the increment field is controlled by the background covariance matrix $\mathbf{P_b}$ in Equation (7) which maps innovations at observation locations into model space. This mapping tends to smooth smaller scales associated with energetic submesoscale and frontal-scale variability typically present in the ocean. The realistic, high-resolution representation of the truth available from the NR enables us to directly analyse the consequences of this smoothing along with the impact of spreading by the model operator. Without a validated OSSE system, this type of analysis can only be performed where and when very-high resolution observational coverage is available, an

uncommon situation. A rare example is Jacobs et al. (2014), who were able to test assumptions concerning $\mathbf{P_b}$ structure in the Gulf of Mexico given the deployment of densely-spaced ocean drifters by the Grand Lagrangian Deployment (GLAD) program (Coelho et al., 2014).

Figure 12a illustrates the difference in TCHP between the single glider experiment G30D1 and the unconstrained FM after two days of assimilation. Figure 12b illustrates the negative of the error field (NR minus FM) present at time zero. After two days, the cumulative correction approximately equals the sum of the first two increment fields. The horizontal structure of TCHP correction is governed by the structure of $\mathbf{P_b}$ tapered by the 280 km localization radius function applied during the layer-by-layer assimilation of the upper-ocean temperature profile used to calculate TCHP. Corrections are confined within this radius, but multiplication by the localization radius function reduces the magnitude of $\mathbf{P_b}$ by a factor of e^{-1} at a radius of ~120 km, and to insignificant values at radii exceeding ~220 km, the latter defining an effective radius of influence. The resulting localized correction structure (Figure 12a) differs from the errors present between the FM and the NR (Figure 12b), particularly with respect to small-scale structure and frontal-scale boundaries, as expected.

Given that altimetry assimilation significantly corrects the structure of the mesoscale eddy field, this same analysis is repeated using experiment G30D4 which adds the single glider to the operational ocean observing system and is initialized from the control experiment. After two days, the cumulative correction again approximately equals the sum of the first two increment fields (Figure 12c). In this case, the correction is smaller in magnitude compared to G30D1 where the glider is assimilated alone. However, small-scale structure is still not accurately corrected (Figure 12d).

Sensitivity to Localization Radius

Evaluation of the impact of localization radius is performed by running two other versions of the single glider experiment described in Figure 9 using alternate localization radii of 160 and 90 km. Reduction of localization radius did not improve the ability of the single glider to correct horizontal field structures within the small analysis box shown in Figure 9, but did reduce the bias corrections for the four model fields listed in Table 6. The optimum choice of localization radius presumably involves a trade-off between maximizing the area over which ocean fields are corrected to maximize bias reduction while minimizing the introduction of spurious structure in the eddy field.

Nonlinear Model Operator

The spreading of cumulative corrections by the nonlinear model operator is illustrated in Figure 14 for the 30-day experiments G30D1, G30D2, and G30D3 listed in Table 5. Whether the glider profiles were assimilated for all 30 days or for only 2 days, and whether a localization radius of 280 km or 90 km was used, the model operator acts to spread out the influence of the glider over an approximately 6 x 6 degree box after 30 days, spreading more to the west than to the east. By the end of 30 days, the TCHP change patterns are no longer significantly correlated between the 280 km and 90 km radius experiments. Although the glider profile assimilation eventually influences a large area, the changes outside the local correction region are not correlated with the truth as represented by the NR due to the chaotic nature of mesoscale variability.

7. Conclusions

The present analysis demonstrates how a rigorously validated and calibrated ocean OSSE system can be used to evaluate observations for a specific purpose, in this case emphasizing the improvement of ocean model initialization in coupled TC prediction systems over the

North Atlantic hurricane region. A fraternal-twin OSSE system approach was employed that was previously validated in the Gulf of Mexico (Halliwell et al., 2014; 2015). Three primary components of the operational ocean observing system, and strategies for deploying additional instrumentation during hurricane season, were evaluated for this purpose. The NR used by the OSSE system has already been thoroughly evaluated and vetted as a realistic representation of the "true" ocean with respect to short-term ocean forecast applications, emphasizing the coupled TC prediction problem and the correct representation of associated ocean-atmosphere processes (Kourafalou et al., 2016; Androulidakis et al., 2016). The remaining evaluation steps required to fully validate and calibrate the OSSE system presented herein demonstrate that the present fraternal twin OSSE system is capable of providing credible impact assessments with only modest calibration.

Satellite altimetry assimilation provides the greatest overall positive impact by substantially correcting mesoscale structure in ocean model dynamical and thermodynamical (e.g. TCHP) fields important to TC prediction. Two of the four available altimeters reduce errors in ocean mesoscale structure by over 90 percent of the total improvement achieved by assimilating all four. By contrast, altimetry assimilation has little impact on bias reduction. Argo profile assimilation results in small error reduction in the horizontal structure of some, but not all, model fields due to the limited (~3 degrees, 10 days) space-time profile coverage. However, Argo profile assimilation does produce substantial bias reduction is most fields. SST measurements are primarily effective only at reducing upper-ocean thermal errors. Satellite SST measurements from the Navy MCSST product provide substantial correction of mesoscale structure due to dense horizontal sampling in cloud-free regions. The three in-situ subcomponents of the SST observing system do not have sufficient space-time resolution to significantly correct mesoscale structure. However, all individual satellite and in-situ components of the SST observing system contribute to large bias reduction in upper-ocean

thermal fields, which is important to the TC prediction problem. Collectively, these three operational observing system components provide substantial error and bias reduction in ocean model initialization errors. Although the emphasis herein is on TC applications, these positive impacts will benefit other ocean prediction applications.

Seasonal enhancements to the operational ocean observing system were evaluated, specifically regional underwater glider deployments. Impacts were assessed for a single glider and a group of six gliders deployed in the open Atlantic north of Puerto Rico, and also for a glider "shield" with 14 instruments deployed offshore of the Gulf of Mexico and southeast U. S. coasts. Given the high-resolution representation of the truth provided by the NR, individual glider profiles (and presumably ocean profiles collected by other instruments) were found to have limited ability to correct field structure in their vicinity. The background error covariance that maps innovations into model space smooths the structure of the increment field (e.g. Jacobs et al., 2014) so that energetic model errors with scales smaller than the mesoscale range are not accurately corrected. Also, horizontal separation of instruments in the multiple glider experiments was not adequate to correct mesoscale structure. Despite these limitations, our results suggest that spreading gliders over larger areas does produce a modest correction in larger-scale field structure, presumably because each individual glider corrects bias in its vicinity. This same factor probably underpins the ability of the Argo float array to modestly correct horizontal structure in model fields despite coarse space-time sampling that does not resolve the mesoscale.

Results from OSSE impact assessments are strictly valid only for the particular design of the OSSE system that was used, which includes the choice of NR and FM models and the DA methodology. The present results therefore represent an initial assessment of observing system impacts with respect to accurate initialization of coupled prediction models. Follow-on OSSE evaluations should be conducted with different DA methodologies such as the

Ensemble Kalman Filter (Evensen, 1994) that update background error covariance estimates during analysis runs, or multi-scale DA procedures (e.g. Haley and Lermusaiux, 2010), that can potentially improve correction of mesoscale structure.

Acknowledgments: The authors acknowledge support from the NOAA Hurricane Sandy Disaster Relief Act (NA15OAR4320064) and from the NOAA Quantitative Observing System Assessment Program (QOSAP, P8R2W02PQF). A. Srinivasan of Tendral LLC developed the architecture of the initial DA system that has been adapted for the OMOC OSSE system. G. Halliwell, M. Mehari, and M. Le Hénaff acknowledge internal support from the NOAA/AOML Physical Oceanography Division. The authors acknowledge the NOAA Research and Development High Performance Computing Program (http://rdhpcs.noaa.gov) for providing computing and storage resources that have contributed to the research results presented in this paper.

References

Androulidakis Y, Kourafalou VK, Halliwell Jr GR, Le Hénaff M, Kang H-S, Mehari M, Atlas R. 2016. Hurricane interaction with the upper-ocean in the Amazon-Orinoco plume region. Ocean Dyn., 66(12): 1559-1588.

Arnold CP, and Dey CH. 1986. Observing system simulation experiments: past, present, and future. Bull. Amer. Meteor. Soc., 67:687-695.

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

Bleck R. 2002. An oceanic general circulation framed in hybrid isopycnic-Cartesian coordinates. Ocean Modelling, 4: 55-88.

Chassignet EP, Smith LT, Halliwell, Jr GR, and Bleck R. 2003. North Atlantic simulation with the HYbrid Coordinate Ocean Model (HYCOM): Impact of the vertical coordinate choice, reference density, and thermobaricity. J. Phys. Oceanogr. 33:2504-2526.

Chassignet EP, Hurlburt HE, Smedstad OM, Halliwell, Jr GR, Hogan PJ, Wallcraft AJ, and Bleck R. 2007. Ocean Prediction with the HYbrid Coordinate Ocean Model (HYCOM). J. Marine Systems, 65:60-83.

Coelho EF, Hogan P, Jacobs G, Thoppil P, Huntley H, Haus B, Lipphardt, Jr. B, Kirwan, Jr. AD, Ryan E, Olascoaga J. et al. 2014. Ocean current estimation using a multi-model ensemble Kalman filter during the grand lagrangian deployment experiment (GLAD). Ocean Modelling, 87:86-106.

Cooper M, and Haines K. 1996. Altimetric assimilation with water property conservation. J. Geophys. Res., 101:1059-1078.

Cummings JA, and Smedstad OM. 2014. Ocean data impacts in global HYCOM. J. Atmos. Oceanic. Technol., 31:1771-1791.

DeMaria M, Mainelli M, Shay LK, Knaff JA, and Kaplan J. 2005. Further improvements to the statistical hurricane intensity prediction scheme (SHIPS). Wea. and Forecasting, 20:531-543.

Domingues R, Goni G, Bringas F, Lee S-K,Kim H-S, Halliwell G, Dong J, Morell J, and Pomales L. 2015. Upper ocean response to Hurricane Gonzalo (2014): Salinity effects revealed by targeted and sustained underwater glider observations. Geophys. Res. Lett. 42:7131-7138.

Dong, J., Domingues R, Goni, G, Halliwell, G, Kim H-S, Lee S-K, Mehari M, Bringas F, Morell J, and Pomales L. 2017. Impact of assimilating underwater glider data on Hurricane Gonzalo (2014) forecast. Wea. and Forecasting, in press.

Doyle JD, Hodur RM, Chen S, Jin Y, Moskatis JR, Wang S, Hendricks EA, Jin H, and Smith TA. 2014. Tropical cyclone prediction using COAMPS-TC. Oceanography, 27:104-115.

Evensen, G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J. Geophys. Res., 99(C5):10143–10162.

Gaspari G, and Cohn SE. 1999. Construction of correlation functions in two and three dimensions. Quart. J. Roy. Metereol. Soc., 125:723–757.

Goni GJ and Knaff J. 2009. Tropical cyclone heat potential. Bull. Amer. Meteorol. Soc., 90:S54-S56.

Goni GJ, DeMaria M, Knaff J, Sampson C, Ginis I, Bringas F, Mavume C, Lauer C, Lin I-I, Ali MM, et al. 2009. Applications of satellite-derived ocean measurements to tropical cyclone intensity forecasting. Oceanogr., 22:176-183.

Haley PJ and Lermusiaux PFJ. 2010. Multiscale two-way embedding schemes for free-surface primitive equations in the “Multidisciplinary Simulation, Estimation, and Assimilation System”. Ocean Dyn. 60:1497-1537.

Halliwell Jr. GR. 2004. Evaluation of Vertical Coordinate and Vertical Mixing Algorithms in the Hybrid-Coordinate Ocean Model (HYCOM). Ocean Modelling, 7:285-322.

Halliwell Jr. GR, Shay LK, Jacob SD, Smedstad OM, and Uhlhorn EW. 2008. Improving ocean model initialization for coupled tropical cyclone forecast models using GODAE nowcasts. Mon. Wea. Rev., 136:2576-2591.

Halliwell Jr. GR, Shay LK, Brewster J, and Teague WJ. 2011. Evaluation and sensitivity analysis of an ocean model response to hurricane Ivan. Mon. Wea. Rev., 139:921-945.

Halliwell Jr. GR, Srinivasan A, Kourafalou V, Yang H, Willey D, Le Henaff M, and Atlas R. 2014. Rigorous evaluation of a fraternal twin ocean OSSE system in the open Gulf of Mexico. J. Atmos. Ocean. Technol., 31:105-130, doi:10.1175/JTECH-D-13-00011.1.

Hoffman RN and Atlas R. 2016, Future Observing System Simulation Experiments. Bull. Amer. Meteorol. Soc., 96:1601-1616.

Halliwell Jr. GR, Kourafalou V, Le Henaff M, Atlas R, and Shay LK, 2015: OSSE impact analysis of airborne ocean surveys for improving upper-ocean dynamical and thermodynamical forecasts in the Gulf of Mexico. Prog. in Oceanogr., 130:32-46, doi:10.1016/j.pocean.2014.09.004.

Hoffman RN, Grassotti C, Isaacs RG, Louis J-F, and Nehrkorn T. 1990. Assessment of the impact of simulated satellite lidar wind and retrieved 183-GHz water vapor observations on a global data assimilation system. Mon. Wea. Rev., 118:2513-1542.

Jacob SD and Shay LK. 2003. The role of oceanic mesoscale features on the tropical cyclone induced mixed layer response. J. Phys. Oceanogr. 33:649-676.

Jacobs GA, Bartels B, Bogucki DJ, Beron-Vera F, Chen SS, Coelho EF, Curcic M, Griffa A, Gough M, Haus BK, et al. 2014. Data assimilation considerations for improved ocean predictability during the Gulf of Mexico Grand Lagrangian Deployment (GLAD). Ocean Modelling, 83:98-117.

Jaimes B, and Shay LK. 2009. Mixed layer cooling in mesoscale ocean eddies during hurricanes Katrina and Rita. Mon. Wea. Rev., 137:4188-4207.

Jaimes B, Shay LK, 2010. Near-inertial wave wake of hurricanes Katrina and Rita over mesoscale oceanic eddies. J. Phys. Oceanogr., 40:1320-1337.

Jaimes, B, Shay LK, and Halliwell GR. 2011. The response of quasigeostrophic oceanic vortices to tropical cyclone forcing. J. Phys. Oceanogr., 41:1965-1985.

Kourafalou V, Androulikakis Y, Halliwell Jr GR, Kang H-S, Mehari M, Le Hénaff M, Atlas R, and Lumpkin R. 2016. North Atlantic ocean OSSE system development: Nature Run evaluation and application to hurricane interaction within the Gulf Stream. Progress in Oceanogr., 148:1-25.

Legler DM, Freeland HJ, Lumpkin R, Ball G, McPhaden MJ, North S, Crowley R, Goni GJ, Send U, and Merrifield MA. 2015. The current status of the real-time in situ global ocean observing system for operational oceanography. J. Oper. Oceanogr., 8:s189-s200.

Leipper DF, and Volgenau D. 1972. Hurricane Heat Potential of the Gulf of Mexico, J. Phys. Oceanogr., 2:218-224.

Lin I-I, Goni GJ, Knaff JA, Forbes C, and Ali MM. 2013. Ocean heat content for tropical cyclone intensity forecasting and its impact on storm surge. Nat. Hazards, 66:1481-1500.

Mainelli M, DeMaria M, Shay LK, and Goni G. 2008. Application of oceanic heat content estimation to operational forecasting of recent category 5 hurricanes. Wea. And Forecast., 23:3-16.

Mehra A, and Rivin I. 2010. A real time ocean forecast system for the North Atlantic Ocean. Terr Atmos. Ocean Sci, 21:211-228.

Meyers P, Shay LK, and Brewster JK. 2014. Development and analysis of the Systematically Merged Atlantic Regional Temperature and Salinity Climatology for Oceanic Heat Content Estimates. J. Atmos. Oceanic Tech., 31:131-149.

Oke P, Larnicol G, Jones EM, Kourafalou V, Sperrevik AK, Carse F, Tanajura CAS, Mourre B, Tonani M, Brassington GB, et al. 2015. Assessing the impact of observations on ocean forecasts and reanalyses: Part 2, regional applications. J. Operational Oceanogr. 8:s63-s79.

Schiller A, Bell M, Brassington G, Brasseur P, Barcela R, De Mey P, Dombrowsky E, Gehlen M, Hernandez F, Kourafalou V, et al. 2015. Synthesis of new scientific challenges for GODAE OceanView. J. Oper. Oceanogr., 8:s259-s271.

Shay LK., Goni GJ, and Black PG. 2000. Effects of a warm oceanic feature on hurricane Opal. Mon. Wea. Rev., 128:1366-1383.

Tonani, M, Balmaseda M, Bertion L, Blockley E, Brassington G, Davidson F, Drillet Y, Hogan P, Kuragano T, Lee T, et al. 2015. Status and future of global and regional ocean prediction systems. J. Oper. Oceanogr., 8:s201-s220.

Table 1. The NR and FM model configurations chosen for the fraternal twin Atlantic OSSE system.

Model Attribute	**Nature Run Model (HYCOM)**	**OSSE System Forecast Model (HYCOM)**
Horizontal resolution	0.04º Mercator (1951 x 1387)	0.08º Mercator (976 x 694)
Vertical discretization	Hybrid, 35 layers (2000 m reference pressure)	Hybrid, 26 layers (0 m reference pressure)
Time steps (baroclinic/barotropic)	180/6 s	480/16 s
Bathymetry	New, 0.04º product from NRL	0.08º bathymetry from HYCOM Atlantic climatological run
Atmospheric forcing	Navy NOGAPS model (every 3 hours)	Navy NOGAPS model (every 6 hours)
Initial and boundary conditions	Global HYCOM (interpolated to higher-resolution NR mesh)	HYCOM Atlantic climatological run
Thermobaric pressure gradient correction	yes	no
KPP critical bulk Ri	0.45	0.30
KPP double diffusion	on	off
KPP nonlocal b.l. mixing	on	off
KPP background IW viscosity	1 x 10-4	2 x 10-4
KPP background IW diffusivity	1 x 10-5	2 x 10-5
KPP max shear inst. Viscosity	5 x 10-3	7 x 10-3
KPP max shear inst. Diffusivity	5 x 10-3	7 x 10-3
Minimum mixed layer thickness	10 m	12 m
Quadratic bottom friction coefficient	0.0023	0.0032
Diffusion velocity for Laplacian viscosity	0.00286	0.0044
Diffusion velocity for biharmonic viscosity	0.02	0.03
Diffusion velocity for biharmonic thickness diffusion	0.01	0.017
Diffusion velocity for Laplacian scalar diffusion	0.0050	0.0087
Hybrid grid generator vertical remapping algorithm	WENO-like	Piecewise Linear Mapping (PLM)
Hybrid grid generator inverse relaxation coefficient	1 baroclinic time step	4 baroclinic time steps

Table 2. The major components of the operational ocean observing system that are assimilated, including subcomponents of the altimetry and SST observing systems. Errors added to the synthetic observations are summarized.

Observing System Component	Instrument	Instrument Measurement RMS Error	RMS Representation Errors	Other RMS Errors
Satellite Altimetry	Cryosat	0.02 m	0.02 m; correlation length scale 40 km	Internal tides 0.01 m (length scale 5 km)
	Jason-2	0.02 m	0.02 m; correlation length scale 40 km	Internal tides 0.01 m (length scale 5 km)
	Envisat	0.02 m	0.02 m; correlation length scale 40 km	Internal tides 0.01 m (length scale 5 km)
	Haiyang-2a	0.02 m	0.02 m; correlation length scale 40 km	Internal tides 0.01 m (length scald 5 km)
SST	Satellite MCSST	0.3ºC	0.2 ºC	
	In-situ fixed surface buoy	0.1ºC	0.2 ºC	
	In-situ surface drifter	0.1ºC	0.2 ºC	
	In-situ ship intake	0.2ºC	0.2 ºC	
Argo	Argo profiling floats	T: 0.005ºC S: 0.005 PSU	T: 0.15ºC S: 0.08 PSU (taper to zero from surface to 200 m)	Depth error – 2 m (taper to zero above 100 m)
XBT	Primarily ship transects	0.05ºC	0.2ºC (taper to zero from surface to 200 m)	Depth error – 1.5% of depth fall rate

Table 3. *RMSE'* statistics calculated using real Argo profiles and synthetic Argo profiles simulated from the NR and FM at the same times and locations. Statistics between the NR and unconstrained FM are compared to statistics between the NR and the actual ocean. The right column presents the percentage by which the error between the NR and FM exceeds the error between the NR and the actual ocean. A value of 100% means that the error magnitude between the two models equals the error magnitude between the NR and the true ocean.

Ocean Field	*RMSE*, NR vs. FM	*RMSE*, NR vs. Real Ocean	NR vs. FM Error Percentage
$D_{0\text{-}1000}$ (dyn m)	0.26	0.20	130%
H_{20} (m)	60	58	103%
H_{26} (m)	19	28	68%
TCHP (kJ cm^{-2})	19	25	76%
$\bar{T}_{0–100}$ (ºC)	1.13	1.38	82%
$\Delta T_{0–100}$ (ºC)	1.44	2.09	69%
SST (ºC)	0.62	0.66	94%
SSS (PSU)	0.23	0.26	88%
Average Error Percentage, NR vs. FM			89%

Table 4. Experiments run to evaluate three primary components of the operational ocean observing system: altimetry, Argo, and SST, along with subcomponents of the altimetry and SST systems.

Data Denial Experiments			**Individual Observing System Experiments**		
Experiments	Run Dates	Initialization	Experiments	Run Dates	Initialization
Control (assimilate existing observing systems)	03/01/14-10/31/14	FM	Unconstrained FM run	09/19/09-10/31/14	Navy GOFS analysis
Deny 1, 2, 3, and 4 altimeters	06/01/14-10/31/14	Control	Add 1, 2, 3, and 4 altimeters	06/01/14-10/31/14	FM
Deny Argo	06/01/14-10/31/14	Control	Add Argo	06/01/14-10/31/14	FM
Deny all SST	06/01/14-10/31/14	Control	Add all SST	06/01/14-10/31/14	FM
Deny satellite SST	06/01/14-10/31/14	Control	Add satellite SST	06/01/14-10/31/14	FM
Deny ship SST	06/01/14-10/31/14	Control	Add ship SST	06/01/14-10/31/14	FM
Deny drifter SST	06/01/14-10/31/14	Control	Add drifter SST	06/01/14-10/31/14	FM
Deny buoy SST	06/01/14-10/31/14	Control	Add buoy SST	06/01/14-10/31/14	FM

Table 5. Experiments to evaluate 2014 hurricane season enhancements to the operational ocean observing system by deploying underwater gliders. The right column gives the imposed Gaspari and Cohn (1999) localization radius.

Experiment	Deployment region	Run dates	Localization radius (km)
Control plus 1 glider	North of Puerto Rico	06/01/14 to 10/31/14	280
1 glider only	North of Puerto Rico	06/01/14 to 10/31/14	280
Control plus 6 gliders	North of Puerto Rico	06/01/14 to 10/31/14	280
6 gliders only	North of Puerto Rico	06/01/14 to 10/31/14	280
Control plus 14 gliders	6 in Gulf of Mexico and NW Caribbean; 8 off SE U.S. coast	06/01/14 to 10/31/14	280
14 gliders only	6 in Gulf of Mexico and NW Caribbean; 8 off SE U.S. coast	06/01/14 to 10/31/14	280
G30D1 (1 glider only)	North of Puerto Rico	09/10/14 to 10/10/14	280
G30D2 (1 glider only, assimilated for only two days)	North of Puerto Rico	09/10/14 to 10/10/14	280
G30D3 (1 glider only)	North of Puerto Rico	09/10/14 to 10/10/14	90
G30D4 (Control experiment plus 1 glider only)	North of Puerto Rico	09/10/14 to 10/10/14	280

Table 6. Mean biases in four fields comparing the unconstrained FM to three single glider experiments that used three different choices of the localization radius R. Biases are calculated over the small box shown in the top panel of Figure 9.

Analysis Region	H_{20} (km)	SST (ºC)	SSS (PSU)	TCHP (kJ cm^{-2})
FM R=280 km	36.9	-0.29	-0.14	-5.76
Single glider R=280 km	23.3	-0.14	-0.04	-1.09
Single glider R=160 km	24.2	-0.18	-0.07	-2.28
Single glider R=90 km	27.4	-0.23	-0.10	-3.94

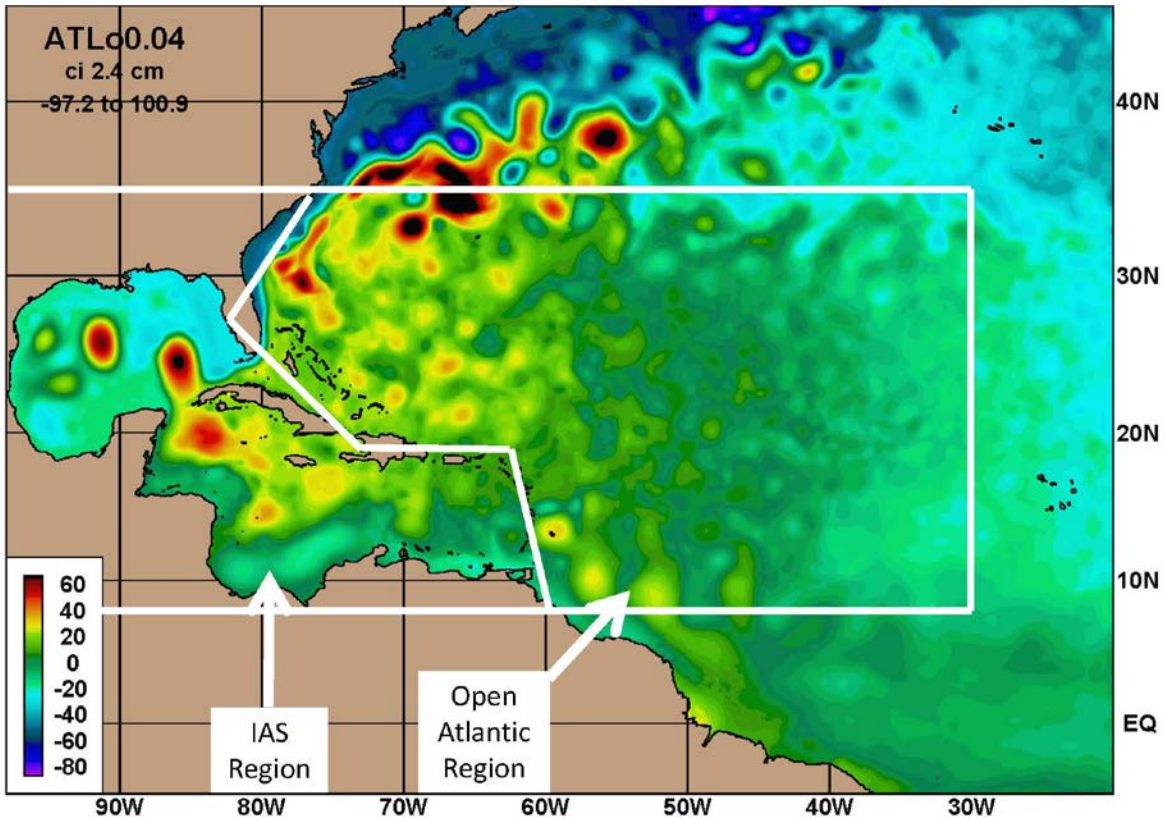
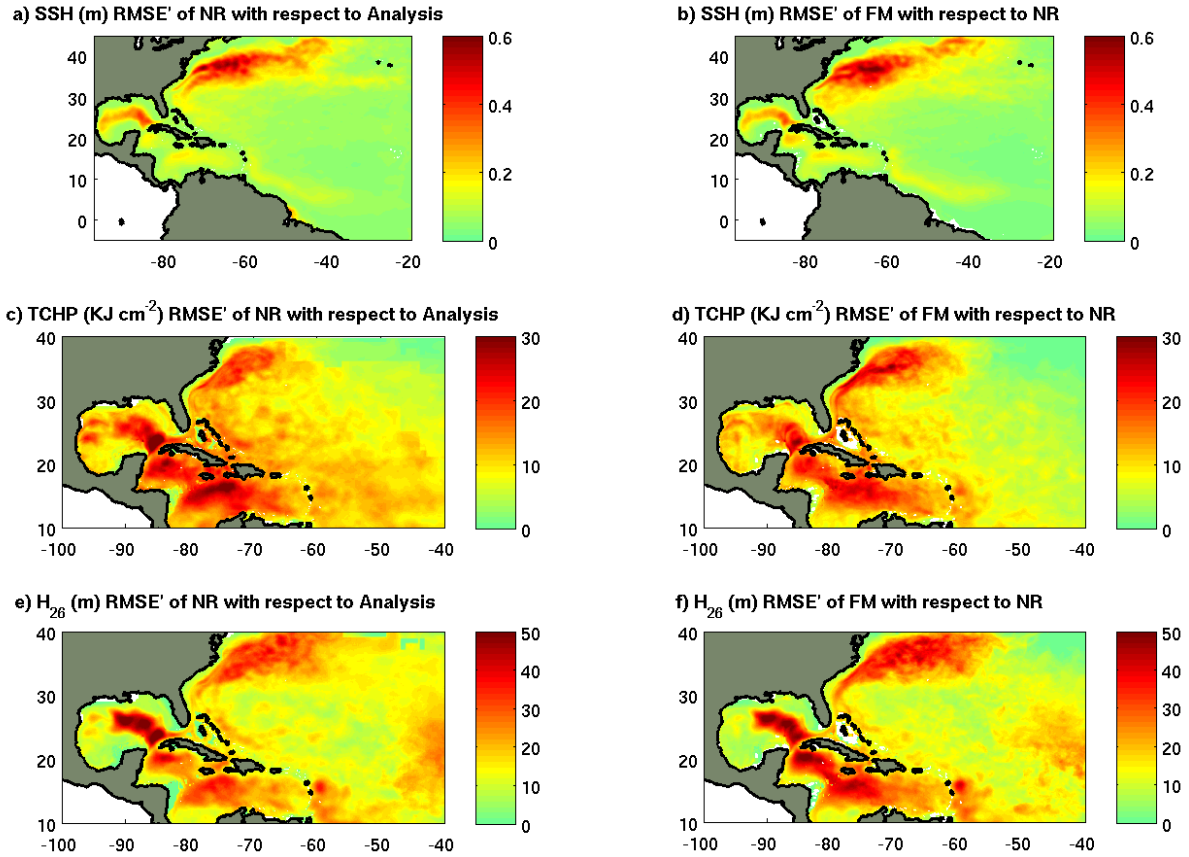


Figure 1. Example model field (SSH, cm) from the NR with white lines illustrating the two primary evaluation regions within the North Atlantic Hurricane domain.



Fi**gure 2.** RMS differences in SSH between (a) the NR and the Aviso analysis and (b) the FM and the NR. RMS differences in TCHP between (c) the NR and the AOML analysis product and (d) the FM and the NR. RMS differences in H_{26} between (e) the NR and the AOML analysis product and (f) the FM and the NR.

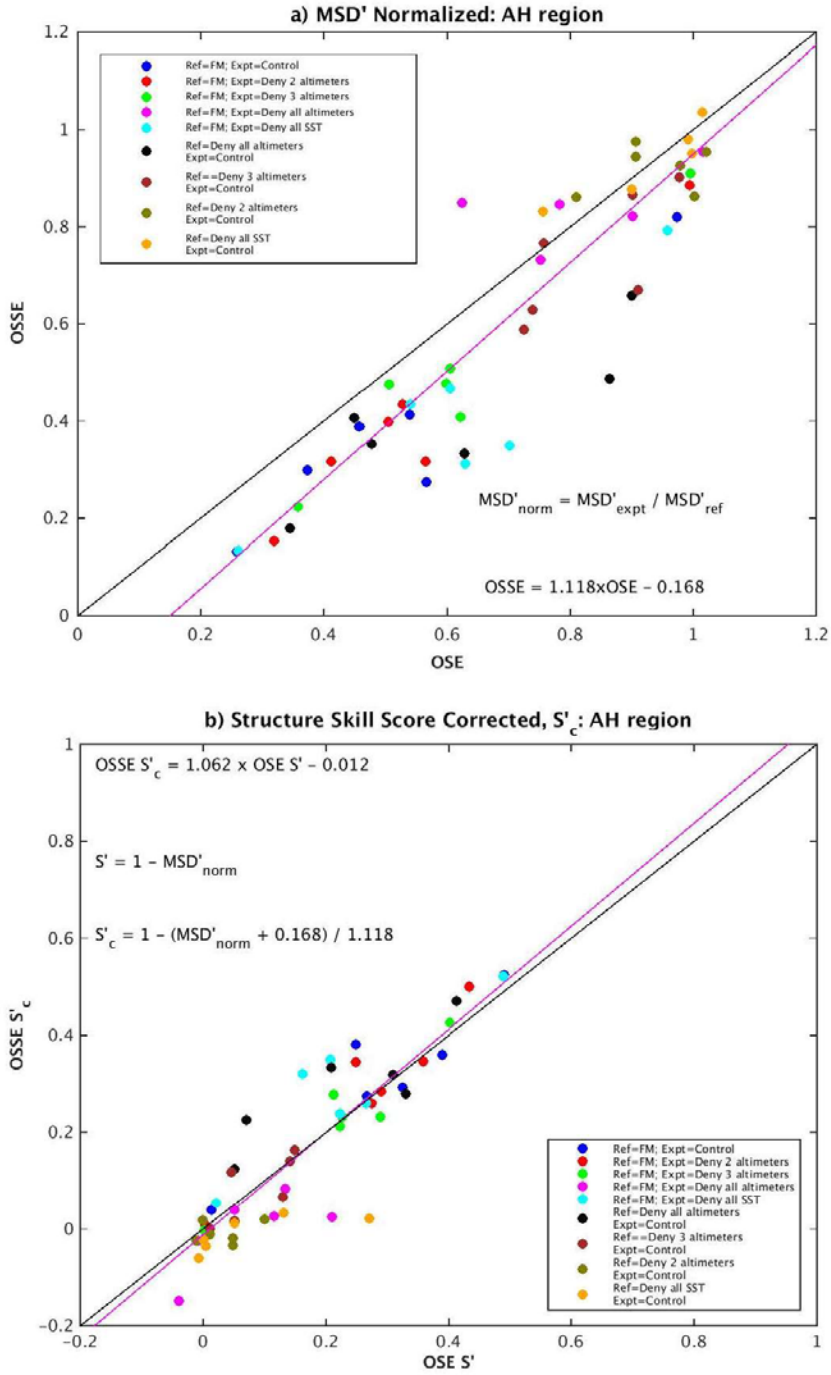


Figure 3. Scatterplots of MSD'_{norm} (top) and calibrated skill score S'_C (bottom). The legends denote the experiment pairs used to calculate MSD'_{norm} from (5). The linear regression equations are presented in both panels. Actual fits are denoted by purple lines while perfect fits are indicated by black lines. The equation used to calculate S'_C is shown in (b).

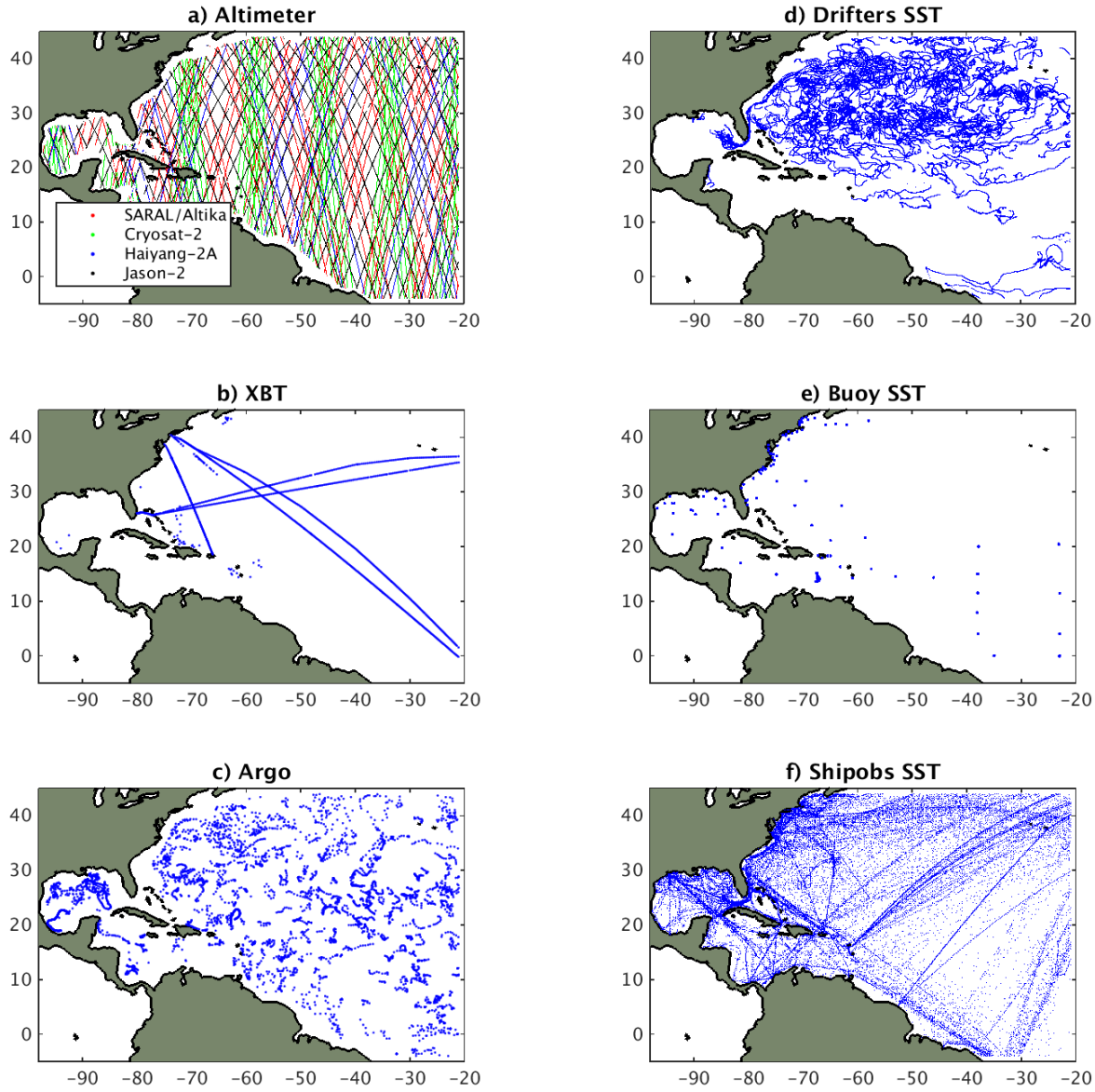


Figure 4. Components of the operational ocean observing system that are assimilated. (a) Track distribution for the four available altimeters (upper left) are illustrated for one ten-day interval only. All in-situ measurements from (b) XBTs, (c) Argo floats, and three SST systems [(d) surface drifters, (e) fixed surface buoys, and (f) ship intake] are shown during the assessment period from 1 July through 31 October 2014. The distribution of satellite-derived SST from the Navy MCSST product is not shown.

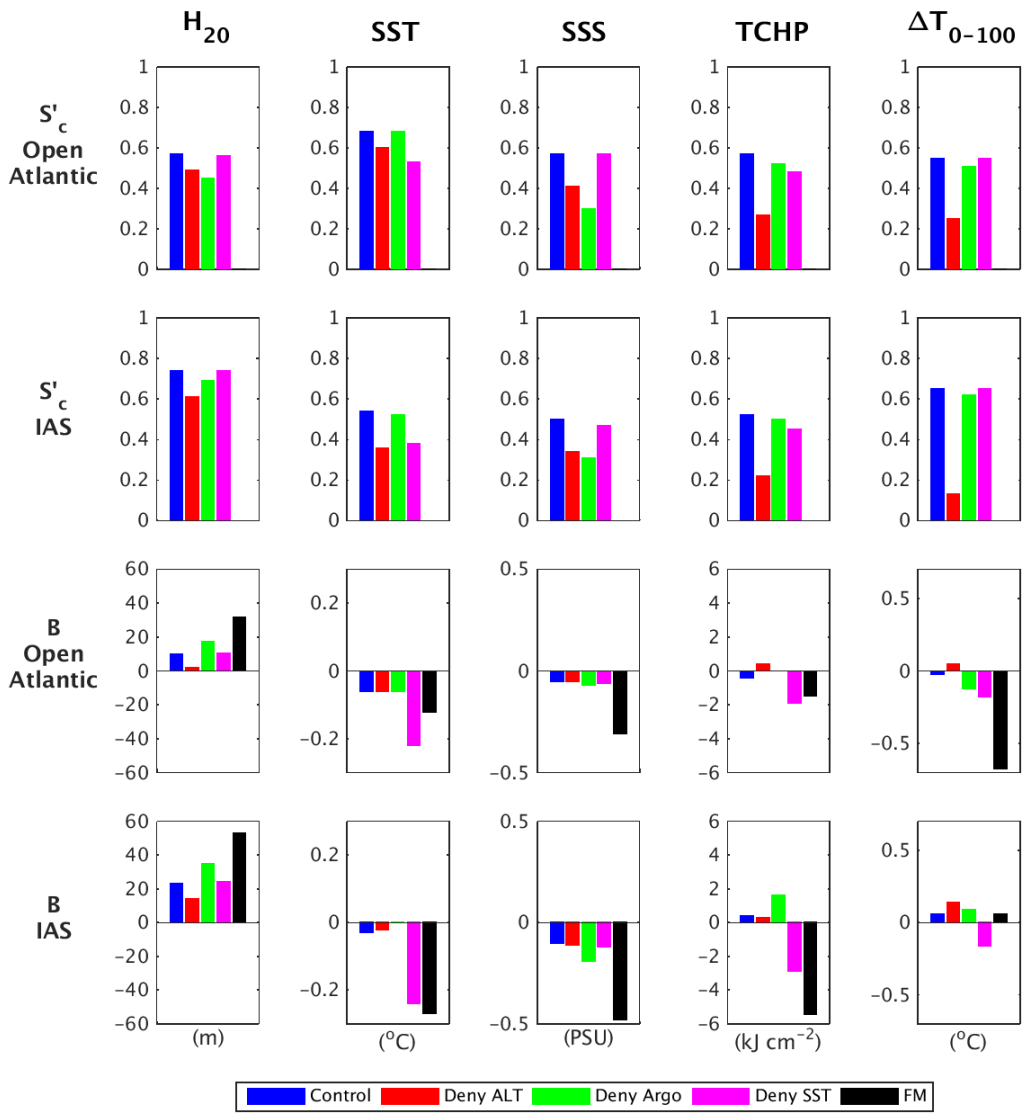


Figure 5. Calibrated skill score S'_C and bias B from the data denial experiments evaluating the three primary observing systems for five model variables calculated over both the open Atlantic and IAS regions (Figure 1). The experiments are listed in the legend at the bottom. The FM is used as the reference experiment REF, so S'_C measures improvement over the unconstrained FM run. By definition, $S'_C = 0$ for the FM experiment.

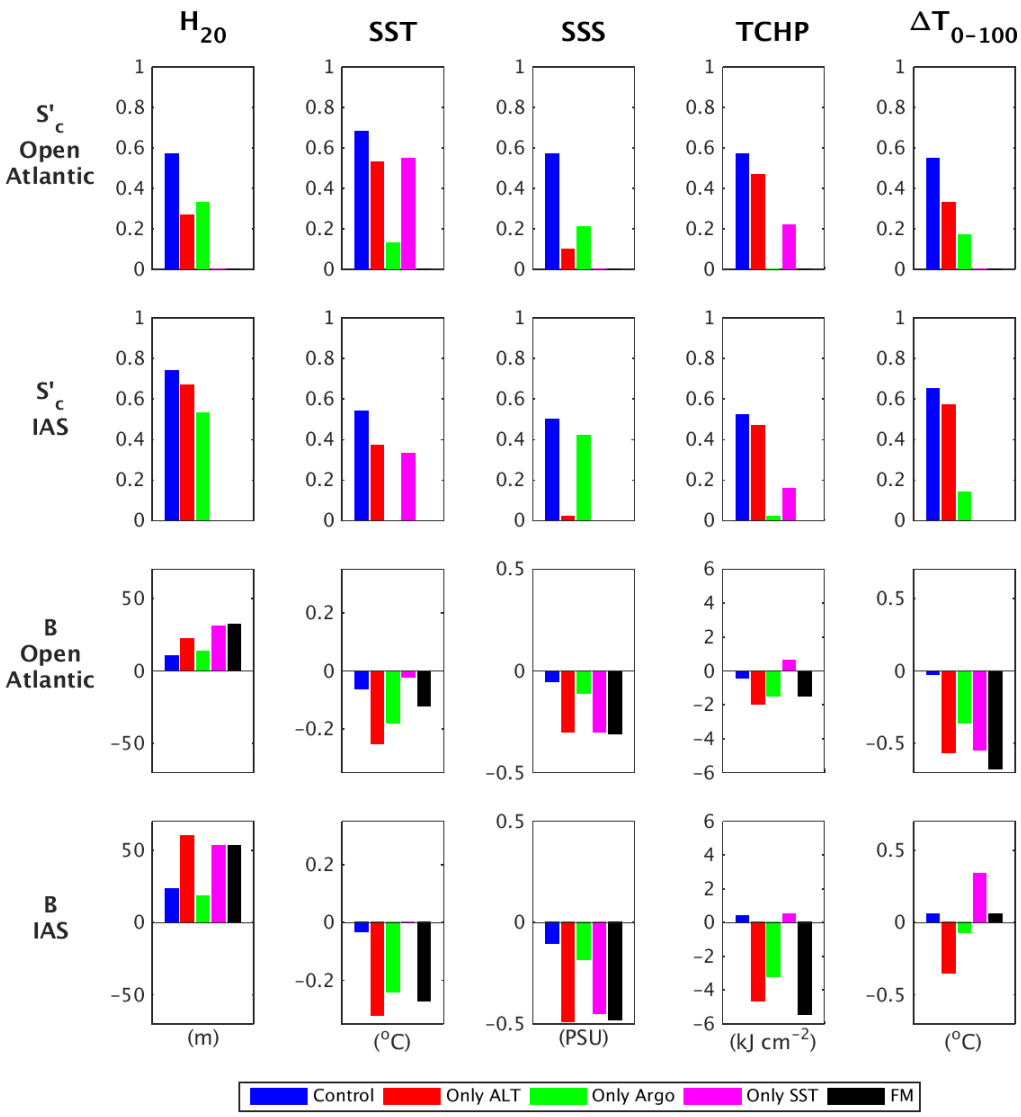


Figure 6. Same as Figure 5 except from the individual observing system experiments.

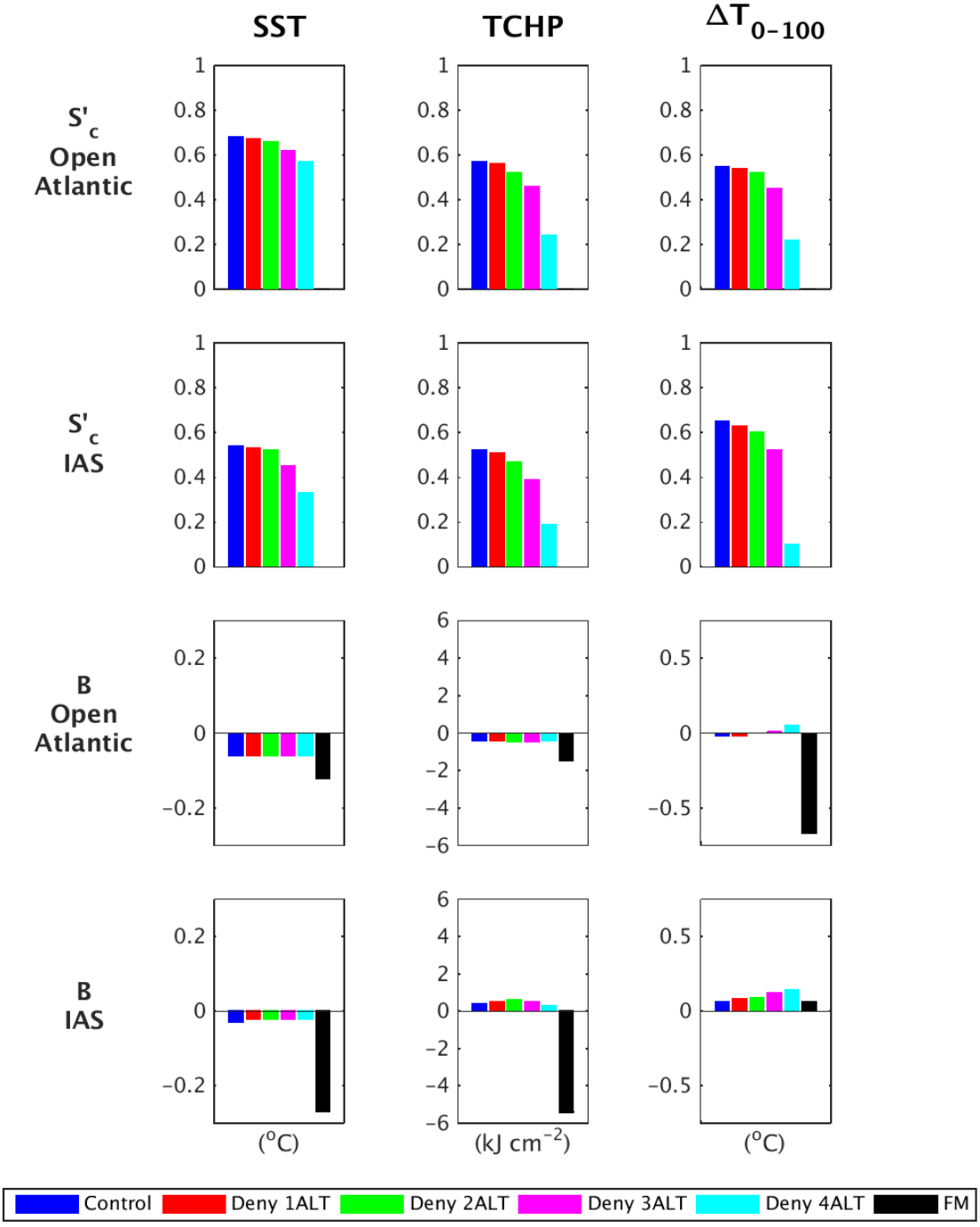


Figure 7. Calibrated skill score S'_C and bias B from the data denial experiments that deny 1, 2, 3, and 4 altimeters in comparison to the control experiment and unconstrained FM run for three model variables calculated over both the open Atlantic and IAS regions (Figure 1). The experiments are listed in the legend at the bottom. The FM is used as the reference experiment REF, so S'_C measures improvement over the unconstrained FM run. By definition, $S'_C = 0$ for the FM experiment.

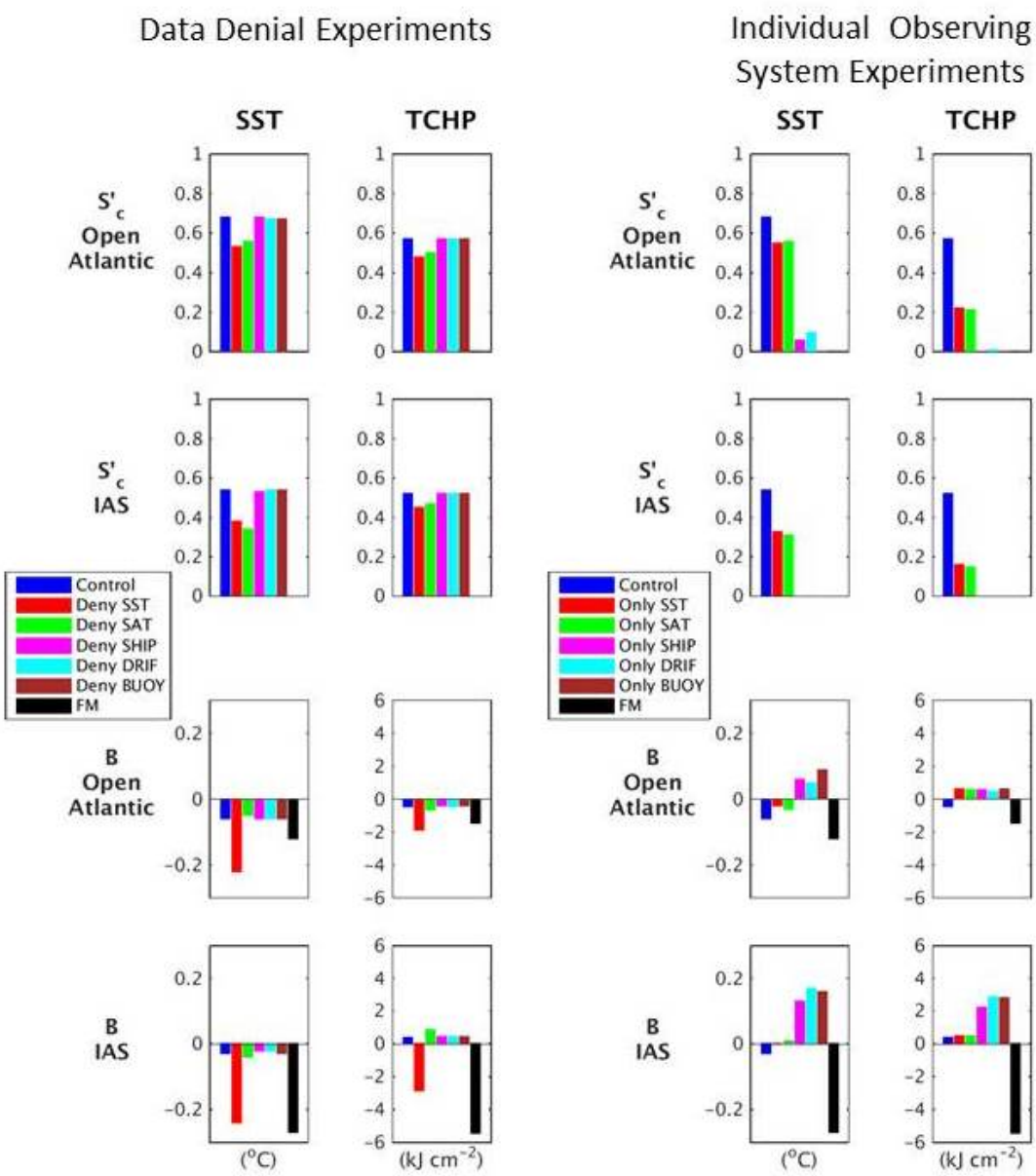


Figure 8. Calibrated skill score S'_C and bias B from the data denial experiments (left columns) that deny the entire SST observing system, and that deny individual components of this system (satellite MCSST, ship, drifter, and buoy), calculated over both the open Atlantic and IAS regions (Figure 1). The same statistics for individual observing system experiments are shown in the right columns. Results from the unconstrained FM run are included for reference. Statistics are shown only for the two model variables strongly influenced by SST assimilation. The experiments are listed in the legends. The FM is used as the reference experiment REF, so S'_C measures improvement over the unconstrained FM run. By definition, $S'_C = 0$ for the FM experiment.

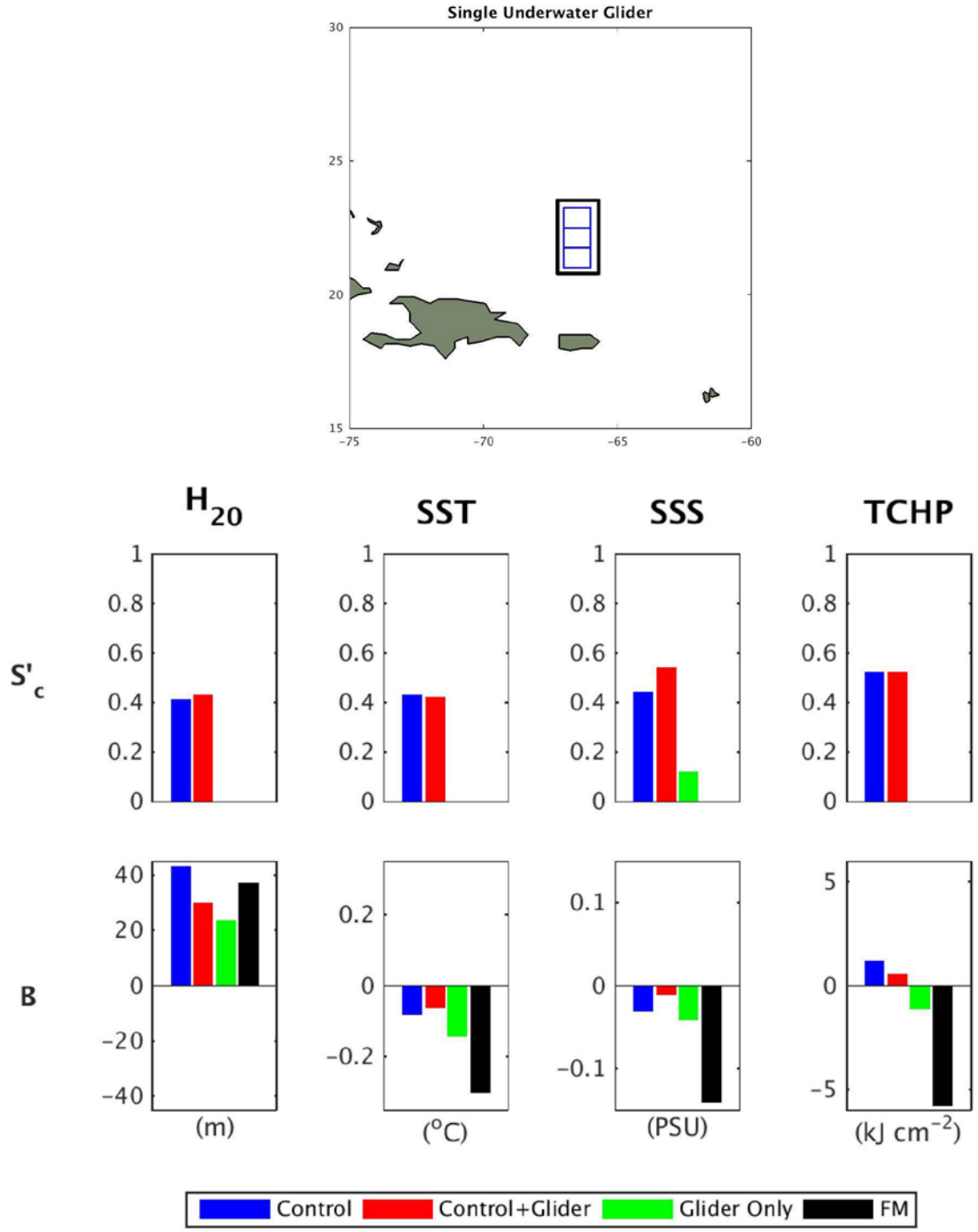


Figure 9. Calibrated skill score S'_C and bias B of four variables from two single-glider experiments, one adding the glider to the control experiment and the other adding the glider to the FM (glider only). Comparison is made to statistics from the control experiment and the unconstrained FM. Statistics are calculated over the black box surrounding the blue ladder-shaped glider track in the top panel.

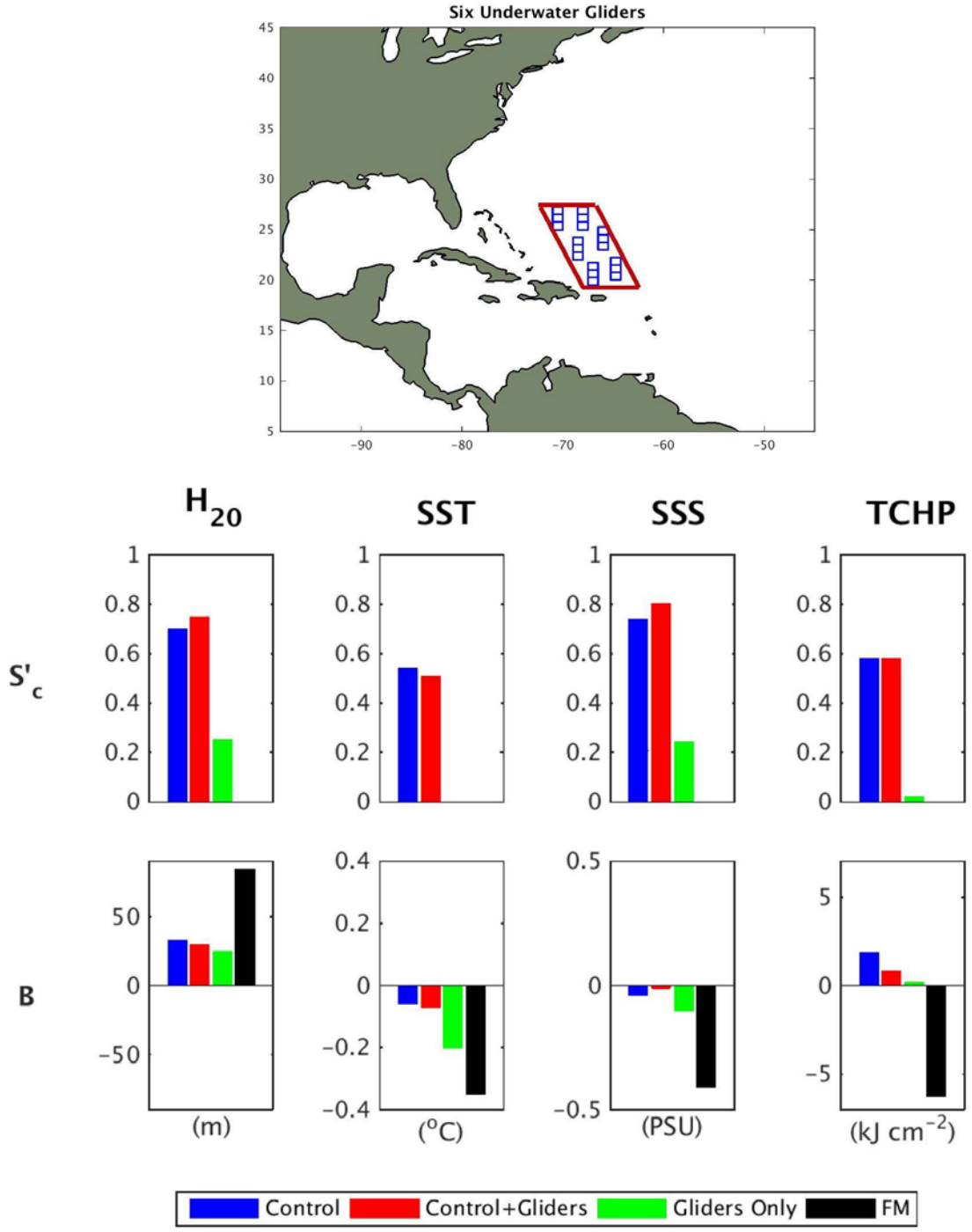


Figure 10. Calibrated skill score S'_C and bias B for four variables from two six-glider experiments, one adding the gliders to the control experiment and the other adding the gliders to the FM (glider only). Comparison is made to statistics from the control experiment and the unconstrained FM. Statistics are calculated over the parallelogram-shaped region shown in the top panel containing the glider paths.

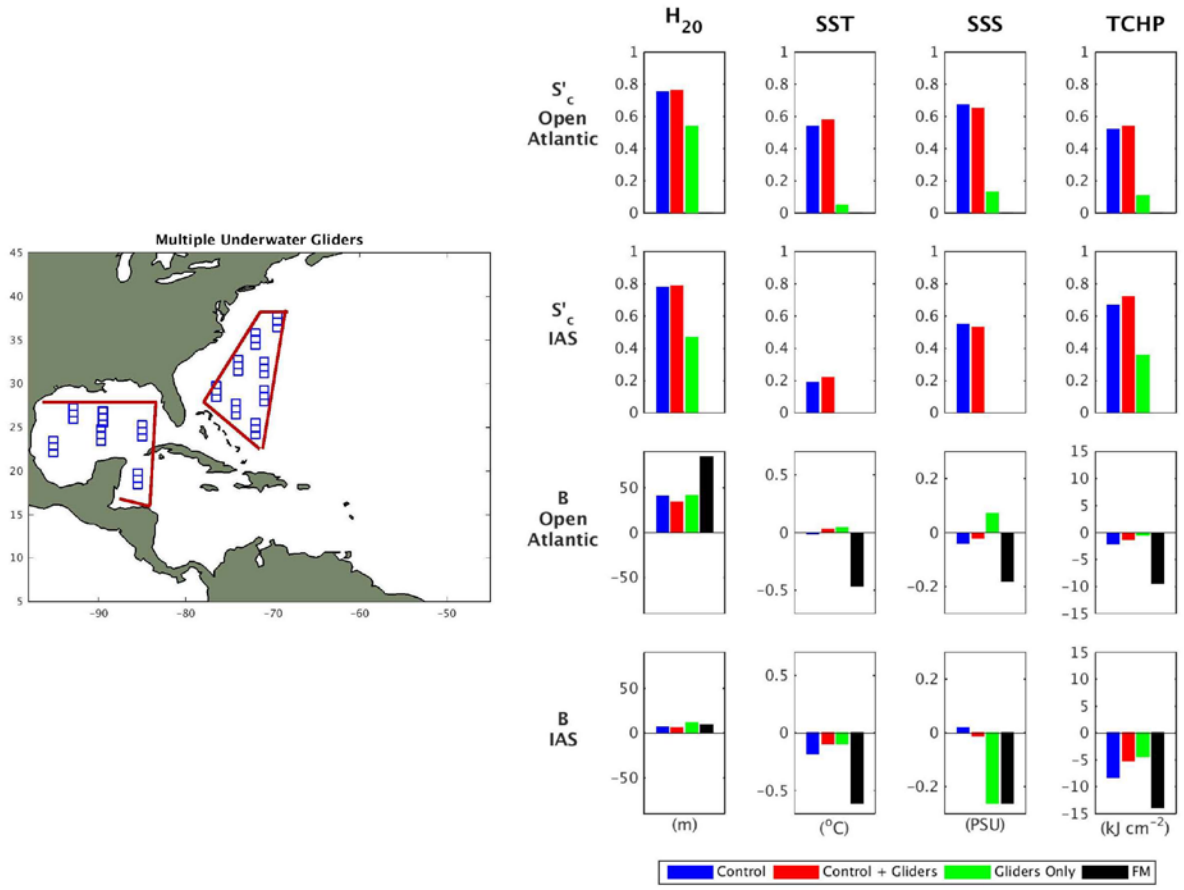


Figure 11. Calibrated skill score S'_C and bias B for four variables from the experiment that adds 14 gliders to the control experiment, with the synthetic gliders deployed as shown (left panel) in deep water off the U. S. Gulf of Mexico and southeast Atlantic coasts. Comparison is made to statistics from the control experiment and the unconstrained FM. Statistics are calculated separately for open Atlantic and IAS regions containing the gliders as outlined in the left panel.

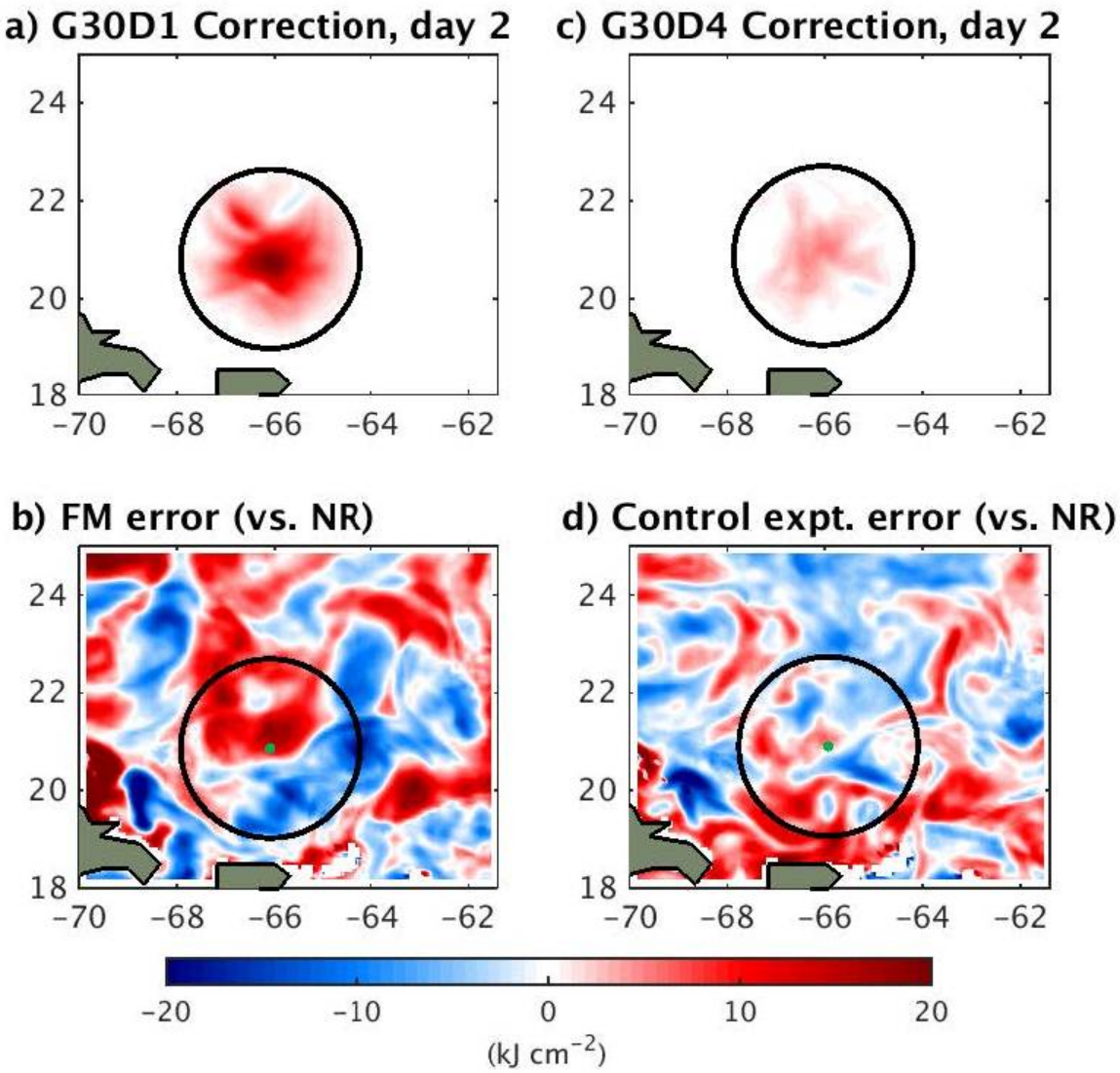


Figure 12. (a) Increment field for TCHP resulting from two consecutive days of glider profile assimilation from experiment G30D1 (Control plus glider assimilation) in Table 5; (b) TCHP difference on day zero, NR minus FM, which represents the negative of the model error present at the initial time; (c) same as (a) from experiment G30D4 (glider assimilation only) in Table 6; and (d) TCHP difference on day zero, NR minus control experiment, which represents the negative of the model error present at the initial time.

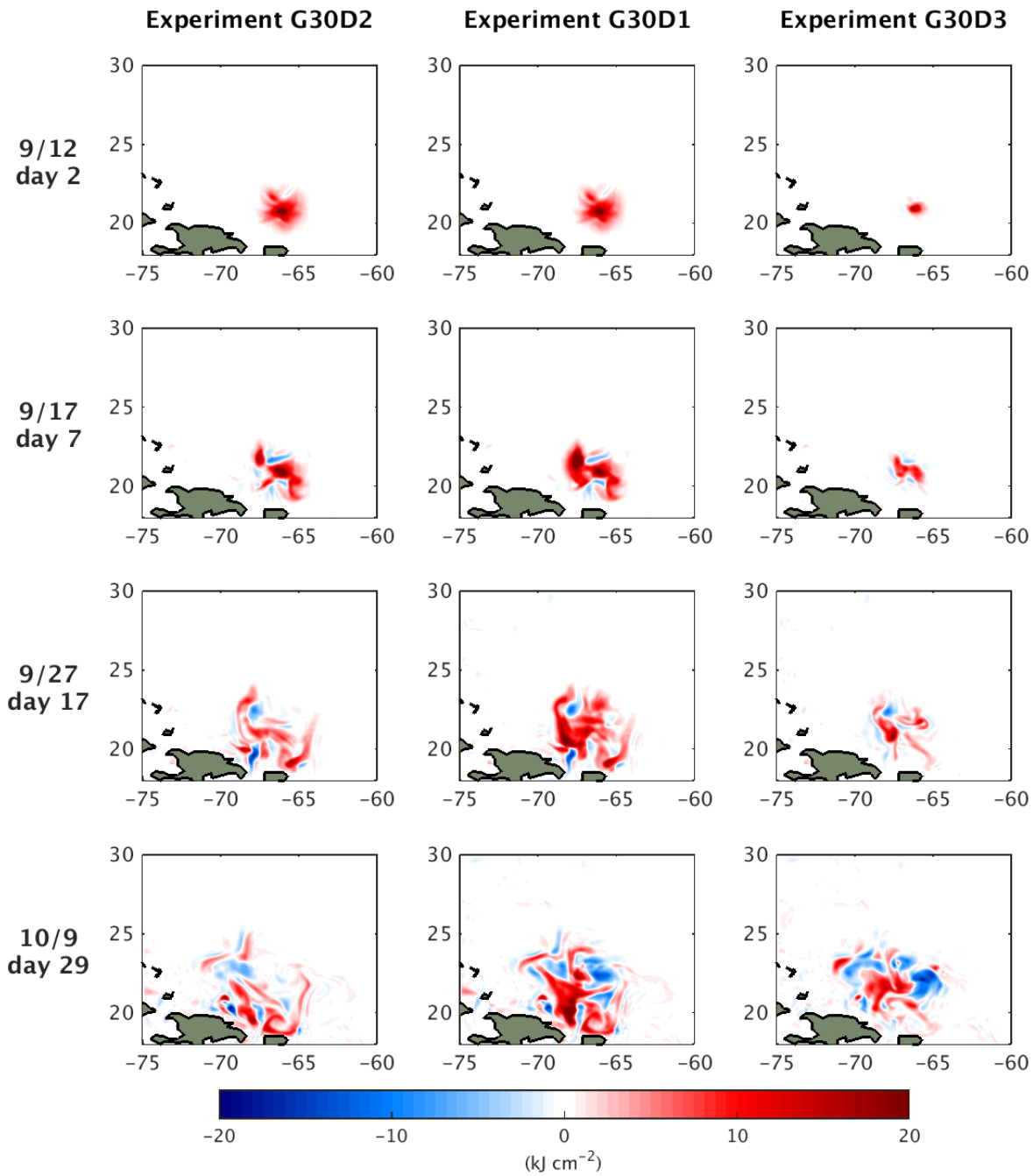


Figure 13. Correction fields for TCHP from three thirty-day experiments (G30D1, G30D2, and G30D3, Table 5) that assimilate a single glider beginning 10 September 2014. The left column is for the experiment that assimilated the glider profiles only during the first two days while the middle column is for the experiment that assimilated glider profiles over all 30 days. The right column is for the experiment that assimilates glider profiles over all 30 days, but uses a localization radius of 90 km instead of 280 km.