Rigorous Evaluation of a Fraternal Twin Ocean OSSE System for the Open Gulf of Mexico

G. R. HALLIWELL JR.,* A. SRINIVASAN,[+] V. KOURAFALOU,[#] H. YANG,[@] D. WILLEY,[@]
M. LE HÉNAFF,[&] AND R. ATLAS**

NOAA/AOML/Physical Oceanography Division, Miami, Florida
[+] *Meteorology and Physical Oceanography Division, Rosenstiel School of Marine and Atmospheric Science,
University of Miami, and Tendral, LLC, Miami, Florida*
[#] *Meteorology and Physical Oceanography Division, Rosenstiel School of Marine and Atmospheric Science,
University of Miami, Miami, Florida*
[@] *Cooperative Institute for Marine and Atmospheric Studies, University of Miami, and NOAA/AOML/Physical
Oceanography Division, Miami, Florida*
[&] *Cooperative Institute for Marine and Atmospheric Studies, University of Miami, Miami, Florida*
***NOAA/AOML, Miami, Florida*

ABSTRACT

A new fraternal twin ocean observing system simulation experiment (OSSE) system is validated in a Gulf of Mexico domain. It is the first ocean system that takes full advantage of design criteria and rigorous evaluation procedures developed to validate atmosphere OSSE systems that have not been fully implemented for the ocean. These procedures are necessary to determine a priori that the OSSE system does not overestimate or underestimate observing system impacts. The new system consists of 1) a nature run (NR) stipulated to represent the true ocean, 2) a data assimilation system consisting of a second ocean model (the "forecast model") coupled to a new ocean data assimilation system, and 3) software to simulate observations from the NR and to add realistic errors. The system design is described to illustrate the requirements of a validated OSSE system. The chosen NR reproduces the climatology and variability of ocean phenomena with sufficient realism. Although the same ocean model type is used (the "fraternal twin" approach), the forecast model is configured differently so that it approximately satisfies the requirement that differences (errors) with respect to the NR grow at the same rate as errors that develop between state-of-the-art ocean models and the true ocean. Rigorous evaluation procedures developed for atmospheric OSSEs are then applied by first performing observing system experiments (OSEs) to evaluate one or more existing observing systems. OSSEs are then performed that are identical except for the assimilation of synthetic observations simulated from the NR. Very similar impact assessments were realized between each OSE–OSSE pair, thus validating the system without the need for calibration.

1. Introduction

Observing system simulation experiments (OSSEs) provide a rigorous and cost-effective approach to evaluate the impact of new atmospheric and oceanic observing systems prior to deployment. OSSEs are essentially an extension of observing system experiments (OSEs), also referred to as data denial experiments. OSEs determine the impact of existing observing systems using data denial experiments, with one assimilating all observations and the other denying the observing system of interest. Impact is determined by the increase in analysis and forecast errors resulting from denial of that system. OSSEs extend this procedure to the evaluation of new observing systems, or alternate deployment strategies for existing systems. Data denial experiments are performed that assimilate synthetic observations sampled from a realistic high-resolution nature run (NR) stipulated to represent the "true" atmosphere or ocean. One experiment assimilates all synthetic observations including the new observing system, while the other denies the new system. An OSSE system thus consists of 1) the atmospheric or oceanic model used to perform the NR, 2) a data assimilation

Corresponding author address: George Halliwell, NOAA/AOML/
Physical Oceanography Division, 4301 Rickenbacker Causeway,
Miami, FL 33149.
E-mail: george.halliwell@noaa.gov

system (DAS) consisting of a different atmospheric or oceanic model (typically referred to as the "forecast model") coupled to a data assimilation (DA) procedure, and 3) a toolbox to simulate realistic synthetic observations from the NR. Although conceptually straightforward, it is necessary to validate the system through rigorous evaluation and to determine if results must be calibrated to ensure that realistic impact assessments are obtained.

OSSEs have been in longer use and are more advanced for the atmosphere compared to the ocean. Early atmospheric OSSEs were generally performed using systems with design flaws and/or without prior validation, which often led to overestimates or underestimates of observing system impact that were not discovered until after system deployment. Over time, design criteria and rigorous evaluation procedures, which include calibration when necessary, were developed to ensure that realistic impact assessments are produced (e.g., Atlas et al. 1985a,b; Atlas 1997). Over the prior three decades, realistic atmospheric OSSEs have been conducted for many purposes—for example, to evaluate the potential for future observing systems, to improve numerical weather prediction, to plan for the Global Weather Experiment, and to plan for the Earth Observing System (e.g., Atlas et al. 1985a, 1999, 2001; Arnold and Dey 1986; Hoffman et al. 1990). The use of OSEs and OSSEs to test the impact of scatterometer winds on numerical weather prediction was reviewed by Atlas et al. (2001). The use of OSSEs to document the impact of lidar winds on numerical weather prediction was reviewed by Atlas and Emmitt (2008). Atmospheric OSSEs have also evaluated trade-offs in the design of observing systems and observing networks (Atlas and Emmitt 1991; Rohaly and Krishnamurti 1993), and tested new methodologies for data assimilation (Atlas and Bloom 1989; Daley 2001).

By contrast, ocean OSSEs performed to date have not followed the complete set of design strategies and rigorous validation techniques developed for the atmosphere. For ocean OSSEs to be credible, they must pass specific tests for realism developed for atmospheric OSSEs that are applicable to the ocean as well. This paper presents the first demonstration of this approach for the ocean with the intent of setting a new standard for future ocean OSSEs. Specifically, it describes the new ocean OSSE system developed at the Ocean Modeling and OSSE Center (OMOC; http://cimas.rsmas.miami.edu/omoc. html), a joint center involving the National Oceanic and Atmospheric Administration (NOAA) Atlantic Oceanographic and Meteorological Laboratory (AOML), the Rosenstiel School of Marine and Atmospheric Science (RSMAS) of the University of Miami, and the Cooperative Institute for Marine and Atmospheric Studies (CIMAS).

Section 2 describes the full set of OSSE procedures developed for the atmosphere and the rationale supporting them, and then discusses how these procedures should be extended to ocean OSSE systems. Section 3 describes the design of the ocean OSSE system components and its implementation in a Gulf of Mexico domain. Section 4 evaluates the suitability of the two ocean model configurations chosen for the NR and the forecast model. Section 5 describes the rigorous validation of the OSSE system, while the overall results are summarized in section 6. A list of acronyms is provided in appendix A, and a description of the DA procedure used in the OSSE system is presented in appendix B.

2. OSSE methodology

a. Rationale for OSSE system design

The established procedures to design and perform OSSEs documented in the atmospheric OSSE literature are summarized by Atlas (1997). The NR is a long unconstrained simulation performed at high resolution using a state-of-the-art general circulation model. For OSSEs to be credible, it is essential that the NR provide the most accurate possible representation of the true atmosphere or ocean, that is, possess a model climatology and variability with statistical properties that agree with observations to within specified limits. Given the present state of ocean models, an NR need not be adequate in all respects as long as the evaluation demonstrates that key phenomena being measured by the observing systems in question are reproduced with sufficient accuracy. Phenomena that are inaccurately reproduced cannot be considered in the OSSE system evaluation.

Criteria governing the selection of the forecast model are presented in Atlas et al. 1985a,b and Atlas (1997). Model errors arise because of 1) errors in the initial state; 2) numerical truncation errors due to insufficient resolution; 3) errors in the representation of physical processes, both resolved and parameterized; and 4) errors in both surface and lateral boundary conditions. If the identical model with the same resolution and boundary conditions is used for both the NR and the forecast model, then initialization errors are solely responsible for error growth between them. By contrast, all four factors contribute to the error growth rate between ocean model simulations and the true ocean. The resulting insufficient error growth rate between the forecast and NR models can lead to biased OSSE impact assessments, typically an overestimation of impact when sparse data are assimilated and an underestimation when dense (e.g., satellite) data are assimilated (Atlas et al. 1985b). This situation is referred to as the "identical twin" problem.

Factors other than errors in the initial state must therefore contribute significantly to error growth rate between the models. Ideally, differences (errors) between the two models should grow to the same magnitude as, and have properties similar to, errors that presently exist between state-of-the-art general circulation models and the true ocean. At the same time, errors in the forecast model cannot become so large that they produce unrealistic representations of climatology and variability. These requirements can be substantially realized by using two different model types and running the forecast model at lower resolution to introduce additional truncation errors. Alternatively, the chosen forecast model can be a different configuration of the same model type used for the NR as long as different physical parameterizations, truncation errors, and boundary condition errors are appropriately introduced. This latter method is referred to as the "fraternal twin" approach, and it is used for the ocean OSSE system presented herein.

An OSSE system must also include software that realistically simulates both existing and planned observing systems from the NR (Atlas 1997). In particular, all errors present in the actual observations must be added to these synthetic observations. In addition to random instrument errors, horizontal and vertically correlated errors, representation errors, and bias must be introduced appropriately. Failure to realistically add all errors will lead to inaccurate impact assessments.

Practically speaking, it is not possible to choose NR and forecast models that perfectly satisfy all of the criteria described above. To reduce the chances of overestimating or underestimating observing system impacts, the two model choices should be evaluated prior to performing OSSEs to make sure that these criteria are at least substantially satisfied. However, thorough validation of the OSSE system must be achieved through a rigorous evaluation procedure that compares OSSEs to reference OSEs (Atlas 1997). A set of reference OSEs are first performed to evaluate components of the present-day observing system. A set of OSSEs are then performed that are identical to the reference OSEs except for the assimilation of synthetic observations simulated from the NR. The OSSE system is validated if impact assessments produced by the OSSEs and reference OSEs are consistently the same. If consistent overestimates or underestimates of impact are obtained from the OSSEs, then the system can potentially be validated after adding a calibration step (e.g., Hoffman et al. 1990) that applies a correction factor. However, if large overestimates or underestimates are consistently realized, or if large random assessment errors are encountered, then the OSSE system design will need to be revisited. Calibration by this method must focus on revisiting the suitability of the NR; adjusting differences in physics, truncation, and boundary condition errors between the two model choices; and determining whether errors added to the synthetic observations are realistic.

b. Status of ocean OSSE development

Although numerous ocean OSEs have been successfully performed to evaluate existing ocean observing systems such as satellite altimetry, satellite SST, and Argo floats (e.g., Oke and Schiller 2007b), we are not aware of any ocean OSSEs published in the literature that used a system that follows all established atmospheric procedures. In this sense, ocean OSSEs are presently at a stage analogous to the early years of atmospheric OSSE development. Various approaches have been used to perform ocean OSSEs. One method uses a single model coupled to an ensemble data assimilation system where observing system impacts are evaluated by the reduction in ensemble error statistics (e.g., Mourre et al. 2006; Le Hénaff et al. 2008). Most published ocean OSSE studies did not use a full-fledged DAS for assimilation. Instead, field reconstruction techniques were often used, specifically generating two- or three-dimensional maps from synthetic observations using procedures such as multivariate optimum interpolation, Kalman filter interpolation, or projection onto dominant empirical orthogonal functions (e.g., Oke and Schillera 2007a; Vecchi and Harrison, 2007; Ballabrera-Poy et al. 2007; Sakov and Oke 2008; Kamenkovich et al. 2009). Studies by Guinehut et al. (2002, 2004) used the field reconstruction approach, but they did not specifically refer to their approach as an OSSE.

Ocean OSSE studies that assimilated synthetic observations into a full-fledged DAS have been performed, but none to our knowledge followed all of the design criteria and rigorous validation procedures established for the atmosphere. Morss and Battisti (2004a,b) used the same coupled ENSO forecast system for both the NR and forecast models, so that results were potentially influenced by the identical twin problem. A similar approach was followed by Taillandier et al. (2006), who evaluated the impact of assimilating Argo floats in the Mediterranean. The evaluation of the planned *Aquarius* sea surface salinity (SSS) satellite by Tranchant et al. (2008) did employ different NR and forecast models, but it did not evaluate the OSSE system in comparison to reference OSEs. In that study, the OSSE was performed by first assimilating actual observations into DAS and then performing a second experiment where synthetic SSS sampled from an NR were added to the actual observations assimilated in the first run instead of assimilating all synthetic observations.

TABLE 1. The two HYCOM model configurations used in the fraternal twin OSSE system.

Model Attribute	Standard model configuration (selected as the forecast model)	Alternate model configuration (selected as the NR model)
Vertical coordinate	Hybrid	Fixed σ–z
Vertical mixing	KPP	Mellor–Yamada level 2.5
Thickness diffusion velocity (m s^{-1})	0.010	0.011
Scalar diffusion velocity (m s^{-1})	0.005	0.007
Viscosity, Laplacian diffusion velocity (m s^{-1})	0.002 86	0.003 16
Viscosity, biharmonic diffusion velocity (m s^{-1})	0.020	0.022
Coefficient of quadratic bottom friction	0.0022	0.0023
Vertical grid remapping	WENO	PPM

Ocean OSSE studies to date have provided valuable information on observing system impacts. However, given the large expense of altering or extending existing ocean observing systems and introducing new observing systems, the time has come to develop validated ocean OSSE systems employing all of the methods long accepted by the atmospheric science community that enable a priori determination of the credibility of impact assessments produced by the system.

3. Ocean OSSE system design

To address the need for a rigorous ocean OSSE system, a comprehensive methodology has been developed, following the established procedure of atmospheric OSSEs. A state-of-the art community model is used and the methodology is demonstrated in a dynamically complex open sea environment (within the Gulf of Mexico), over an 8-yr period (2004–10). In the following, and unless specified otherwise, "OSSE" implies ocean OSSE.

a. Configuration of the NR and forecast models

The fraternal twin OSSE system uses two different realizations of the same ocean model type configured to produce substantially different physics and truncation errors. The Hybrid Coordinate Ocean Model (HYCOM) is a primitive equation ocean model with Lagrangian vertical coordinates that quasi-optimally resolve vertical structure throughout the ocean. The standard vertical coordinate configuration is isopycnic in the stratified ocean interior, but it dynamically transitions to level coordinates near the surface to provide resolution in the surface mixed layer. In the coastal ocean, the vertical coordinate system dynamically transitions to user-specified level or terrain-following (σ) coordinates. At the end of each baroclinic time step, a vertical grid generator attempts to restore isopycnic target densities in each layer by moving layer interfaces vertically and

then remapping model layer variables. This restoration is not possible near the surface, where specified minimum layer thicknesses are maintained. Model equations are presented in Bleck (2002), while subsequent evolution and evaluation of the model is summarized in Chassignet et al. (2003) and Halliwell (2004). HYCOM allows flexible choices of vertical coordinate type (Halliwell et al. 2009) and also contains multiple choices of numerical algorithms and subgrid-scale parameterizations, all suitable for introducing different physics and truncation errors.

The two configurations chosen for the fraternal twin OSSE system employ the standard hybrid vertical coordinate system and an alternate fixed σ–z vertical coordinate system (Table 1). In the alternate configuration, σ coordinates are restricted to the inner continental shelf regions to limit regions where large pressure gradient errors may occur (Halliwell et al. 2009). The standard configuration uses the K-profile parameterization (KPP) vertical mixing scheme (Large et al. 1994), while the alternate configuration uses the Mellor–Yamada level 2.5 turbulence closure (Mellor and Yamada 1982). Horizontal mixing and diffusion coefficients are varied between the two configurations (Table 1). The vertical remapping algorithm in the hybrid grid generator is also varied, with the standard configuration using the weighted essentially nonoscillatory (WENO) algorithm and the alternate configuration using the piecewise polynomial mapper (PPM) algorithm. Vertical remapping is also necessary for the fixed σ–z configuration because the model Lagrangian vertical coordinates rely on the hybrid grid generator to restore fixed-layer thicknesses. Taken together, these choices introduce substantially different physics and truncation errors between the two model configurations. Errors in surface forcing and boundary conditions are not explicitly introduced, although small boundary condition errors may result by interpolation from the original 0.08° (the resolution of the forecast model used in the DAS) to the 0.04° resolution of the NR. It is demonstrated later that failure to explicitly

TABLE 2. Summary of the experiments including the three long (2004–10), unconstrained simulations along with the reference OSEs and corresponding OSSEs run with the forecast model. Experiment HIRES2 using the alternate model configuration was chosen as the NR, while the standard configuration used for experiment LORES was chosen as the forecast model.

Experiment	Time interval	Characteristics
HIRES1	1 Jan 2004–31 Dec 2010	0.04° grid using standard HYCOM configuration
HIRES2 (NR model)	1 Jan 2004–31 Dec 2010	0.04° grid using alternate HYCOM configuration
LORES (forecast model)	1 Jan 2004–31 Dec 2010	0.08° grid using standard HYCOM configuration
DAFREE	1 Jan 2010–20 Dec 2010	Extracted from experiment LORES
OSE1	1 Jan 2010–20 Dec 2010	Assimilate all observations listed in Table 3
OSE2	1 Jan 2010–20 Dec 2010	Deny WP-3D profiles
OSE3	1 Jan 2010–20 Dec 2010	Deny WP-3D profiles, *Jason-2* and *Envisat* altimeters
OSSE1	1 Jan 2010–20 Dec 2010	Assimilate all synthetic observations listed in Table 3
OSSE2	1 Jan 2010–20 Dec 2010	Deny WP-3D profiles
OSSE3	1 Jan 2010–20 Dec 2010	Deny WP-3D profiles, *Jason-2* and *Envisat* altimeters

introduce these errors did not compromise the validation of the OSSE system.

b. Selection of the NR and forecast models

In an OSSE system, the NR model is chosen to perform the most statistically realistic simulation of the ocean as possible. To determine which model configuration should preferentially be chosen to perform the NR, two unconstrained high-resolution experiments (HIRES1 and HIRES2; Table 2) were run using the "standard" and "alternate" configurations listed in Table 1. Both experiments were run from 2004 through 2010 on a 0.04° Mercator mesh with 32 hybrid vertical layers. They were forced by fields obtained from a regional mesoscale atmospheric model, specifically the data-assimilative U.S. Navy Coupled Ocean–Atmosphere Mesoscale Prediction System (COAMPS) run with a horizontal resolution of 27 km. Both were also nested within a model-generated high-resolution Atlantic Ocean climate simulation performed by the Naval Research Laboratory. The initial field for 1 January 2004 used to initialize both models was obtained from a data-assimilative Gulf of Mexico (GOM) HYCOM analysis that was employed in several studies of GOM circulation (Prasad and Hogan 2007; Halliwell et al. 2009; Kourafalou et al. 2009). Model archives were saved at 6-h intervals to resolve higher-frequency variability such as near-inertial oscillations in the ocean velocity fields. Because of the identical initialization, a comparison of these model runs is conducted over the 2005–10 time interval to allow differences (errors) between the simulations to grow and equilibrate over the first year of integration.

Snapshot maps of SSH and SST along with zonal cross sections are presented to illustrate the performance of these two model configurations (Fig. 1). Important features of the circulation possess a realistic structure in both experiments, and the cross sections reveal the different vertical coordinate structure. Analysis of these

and other fields demonstrated that insignificant differences exist in the realism with which each model represents the ocean variability. The Loop Current (LC) always displayed realistic pathways and also shed warm eddies in a realistic manner, although the 2005–10 time interval is too short to precisely evaluate errors in the frequency of shedding events. Eddies in the interior Gulf displayed realistic structure and variability, particularly the relatively small cyclones adjacent to the LC that often contribute to the eddy shedding process. Le Hénaff et al. (2012) evaluated another experiment that used HYCOM with a configuration similar to HIRES1, and demonstrated satisfactory thermal structure and northward extension of the LC.

Taken together, these assessments demonstrate that neither model configuration can be preferentially accepted or rejected as being suitable for performing the NR. Motivated in part by the desire to evaluate the new DAS with the standard vertical coordinate configuration, the alternate model configuration experiment HIRES2 is chosen to be the NR, while the standard configuration of HYCOM is used as the forecast model component of the DAS. With the NR performed at 0.04° resolution, the forecast model is configured on a 0.08° Mercator mesh that consists of every other grid point of the NR mesh to introduce additional truncation errors.

c. Data assimilation system

The DAS consists of the forecast model coupled to the Tendral Statistical Interpolation System (T-SIS). T-SIS is a statistical interpolation package for use with ocean circulation models in analysis, forecasting, and system evaluation applications. The prediction/background error covariance needed in the estimation procedure can be flexibly specified from several common approximations and parameterizations of the full error covariance matrix. The package also provides a comprehensive set of support routines for handling most of the common
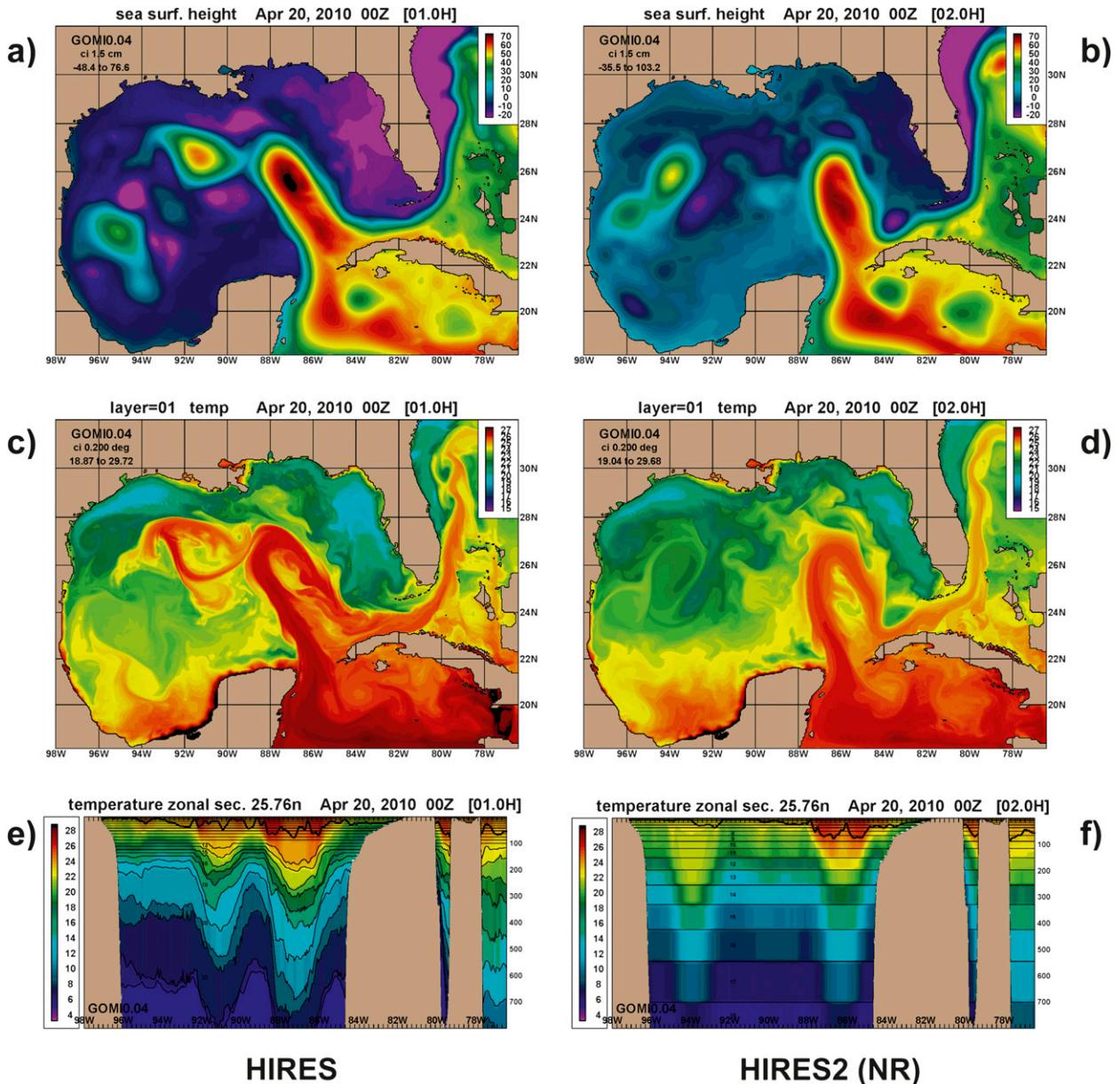
FIG. 1. Model fields (snapshots on 10 Apr 2010) from the two parallel long free runs (2004–10) of the 0.04° GOM HYCOM using the two fraternal twin configurations (Table 1). Shown are (top) SSH, (middle) SST, and (bottom) zonal cross sections at 25.76°N for the (left) standard hybrid configuration (experiment HIRES1) and the (right) σ–z (predominantly z) fixed-coordinate configuration (experiment HIRES2, which was chosen as the NR). Note that although the ocean features are the same, they are represented differently in the two models. The differences are related in part to different numerical truncation errors associated with the two vertical coordinate choices.

observations types and Python support codes that per-form statistical pre- and postprocessing, visualization, and quality control. T-SIS can be used with all ocean model types, and the version used herein is optimized for assimilation into the Lagrangian vertical coordinate layers of HYCOM. Technical details of the T-SIS and its implementation in the present study are contained in appendix B. Because this is a new system, an evaluation

of T-SIS performance is presented in section 5c prior to evaluating the OSSE system.

d. Ocean observations

The assimilated datasets along with key parameters used for their assimilation are listed in Table 3. Because of the enhanced ocean observation effort in the eastern Gulf of Mexico during 2010 spurred by the *Deepwater*

TABLE 3. Summary of observation types and observing systems assimilated along with key input parameters required by the DAS. The local observation radius defines the radius from each model grid point outside of which observations are excluded. The minimum assimilation depth is the isobath within which observations are not assimilated.

Observation type	Observing system	Observation error	Local observation radius (km)	Minimum assimilation depth
Satellite SSH	*Jason-1* altimeter	0.04 m	150	300
	Jason-2 altimeter	0.04 m	150	300
	Envisat altimeter	0.04 m	150	300
Satellite SST	MCSST	0.5°C	40	30
In situ SST	Surface buoy	0.3°C	40	30
	Surface drifter	0.3°C	40	30
	Ship intake	0.4°C	40	30
XBT *T* profiles	Ship	Variable from GDEM3	150	30
	Airborne (WP-3D)	Variable from GDEM3	150	30
XCTD *T*, *S* profiles	Airborne (WP-3D)	Variable from GDEM3	150	30
XCP *T* profiles	Airborne (WP-3D)	Variable from GDEM3	150	30

Horizon oil spill (Liu et al. 2011), the OSSE system evaluation is conducted during this time interval. Datasets include along-track measurements of SSH anomaly (SSHA) from three altimeters: *Jason-1*, *Jason-2*, and the *Environmental Satellite* (*Envisat*), along with the mean dynamic topography field removed from these data, all obtained from the Archiving, Validation, and Interpretation of Satellite Oceanographic data (AVISO) center (http://www.aviso.oceanobs.com). SST from the satellite-derived multichannel sea surface temperature (MCSST) product, in situ measurements collected by ship and surface buoys, and in situ measurements collected by surface drifters were all obtained from the U.S. Global Ocean Data Assimilation Experiment (USGODAE) server (http://usgodae.org) and used for assimilation.

Subsurface measurements include XBT profiles collected from ships that were also obtained from the USGODAE server. Profiles collected by Argo floats were not available in the GOM during the oil spill. Subsurface measurements also include profiles of airborne expendable BT (AXBT), CTD (AXCTD), and current profilers (AXCP) collected by the NOAA WP-3D hurricane research aircraft on nine flight days between 8 May and 9 July 2010 (Shay et al. 2011). The aircraft sampled profiles across the eastern Gulf of Mexico in quasi-synoptic lawnmower patterns with sufficient two-dimensional resolution to resolve the path of the LC and the associated cyclones and anticyclones. Survey maps for all flight days are presented in Shay et al. (2011). Most of the probes that were dropped on each flight day were AXBTs, with most sampling to nearly 400 m and others sampling to 800 m. On most flight days, the AXBTs were supplemented by a small number of AXCPs that sampled temperature and velocity profiles to depths up to 1800 m and AXCTDs that sampled temperature and salinity to at least 1000 m. All velocity

profiles were used solely for evaluation. Lagrangian trajectories obtained from the AOML surface drifter dataset were also obtained for evaluation.

e. Synthetic observations

A set of synthetic observations identical to the actual observations listed in Table 3 were sampled from the NR. For each observation type, realistic errors are added (Table 4). Instrument and other local random errors are added to each individual observation using a random number generator that assumes a Gaussian probability density distribution. In addition to uncorrelated local instrument errors, representation errors that may have horizontal or depth correlation scales large compared to model resolution must be accounted for. For example, actual altimetry measurements resolve ocean eddy and frontal variability in the along-track direction that are unresolved or poorly resolved in the NR. Inspection of NR SSH fields demonstrates that eddies with diameters <40 km (corresponding to wavelengths <80 km) are not adequately simulated. SSH variability associated with smaller submesoscale eddies must therefore be added to the synthetic altimetry data (Table 4).

To model this error, a random number generator first calculates a Gaussian-distributed error value with an RMS amplitude of 0.08 m. This error value is then added to all n sampling points within an along-track window of length 100 km after multiplication by a factor of $n^{-3/2}$. This window is then advanced by one along-track sampling point and the procedure is repeated. The value of the multiplication factor is chosen so that the resulting RMS error magnitude at each point equals the intended value of 0.08 m. An example of the along-track errors added along one track segment is presented in Fig. 2. The integral along-track correlation length scale of this example is estimated from

TABLE 4. Summary of the errors added to the synthetic observation types and observing systems sampled from the NR. Although velocity profiles were not assimilated, errors were added to AXCP velocity profiles since they were used for evaluation.

Observation type	Observing system	Random instrument error	Representation error	Other errors
Satellite SSH	*Jason-1* altimeter	0.03 m	Random 0.08 m Length scale 100 km	Internal tides Random 0.03 m Length scale 5 km
	Jason-2 altimeter	0.03 m	Random 0.08 m Length scale 100 km	Internal tides Random 0.03 m Length scale 5 km
	Envisat altimeter	0.03 m	Random 0.08 m Length scale 100 km	Internal tides Random 0.03 m Length scale 5 km
Satellite SST	MCSST	0.12°C	Random 0.05°C	
In situ SST	Surface buoy	0.05°C	Random 0.05°C	
	Surface drifter	0.05°C	Random 0.05°C	
	Ship intake	0.12°C	Random 0.05°C	
XBT T profiles	Ship	0.02°C	Random 0.02°C	Depth error RMS amplitude 2 m at 100-m depth (see text)
	Airborne (WP-3D)	0.02°C	Random 0.02°C	Depth error RMS amplitude 2 m at 100-m depth (see text)
XCTD T, S profiles	Airborne (WP-3D)	T 0.02°C S 0.05 psu	Random T 0.02°C S 0.05 psu	Depth error RMS amplitude 2 m at 100-m depth (see text)
XCP T, u, and v profiles	Airborne (WP-3D)	T 0.02°C u 0.02 m s^{-1} v 0.02 m s^{-1}	Random T 0.02°C u 0.02 m s^{-1} v 0.02 m s^{-1}	Depth error RMS amplitude 2 m at 100-m depth (see text)

$$L = 2 \int_0^T r(\tau)\, d\tau, \qquad (1)$$

where the autocorrelation function r of distance lag τ is integrated out to lag T set by the first zero crossing. The resulting value of 46 km demonstrates that the representation error adds variability over space scales not adequately resolved by the NR.

Fall-rate errors must be accounted for in ocean profiles. This error is modeled by using a random number generator to first calculate a depth error value for each individual profile assuming a Gaussian distribution with an RMS amplitude of 2 m. The resulting depth error for each profile instrument was applied equally throughout its vertical extent below 100 m but tapered above that depth. Each resulting depth error profile was then translated into error profiles for all measured variables. Random instrument measurement errors for AXBTs, AXCTDs, and AXCPs used herein are presented in Shay et al. (2011). Another random representation error is added to synthetic SST from both satellite and in situ sources to account for the inability of the model to resolve very small-scale structure in the surface SST field (Table 4).

4. Evaluation of the two model configurations

a. Climatology

To determine whether the configurations of the NR and forecast models substantially satisfy basic requirements as outlined in section 2a, experiment HIRES (the NR) is compared to an unconstrained low-resolution
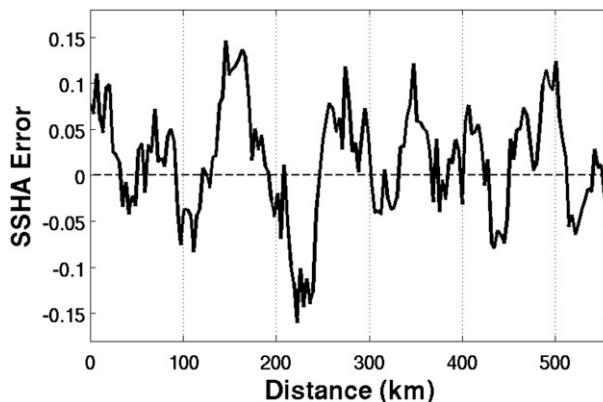

FIG. 2. An example of the SSHA spatially correlated representation error added to the synthetic altimetry along-track data.
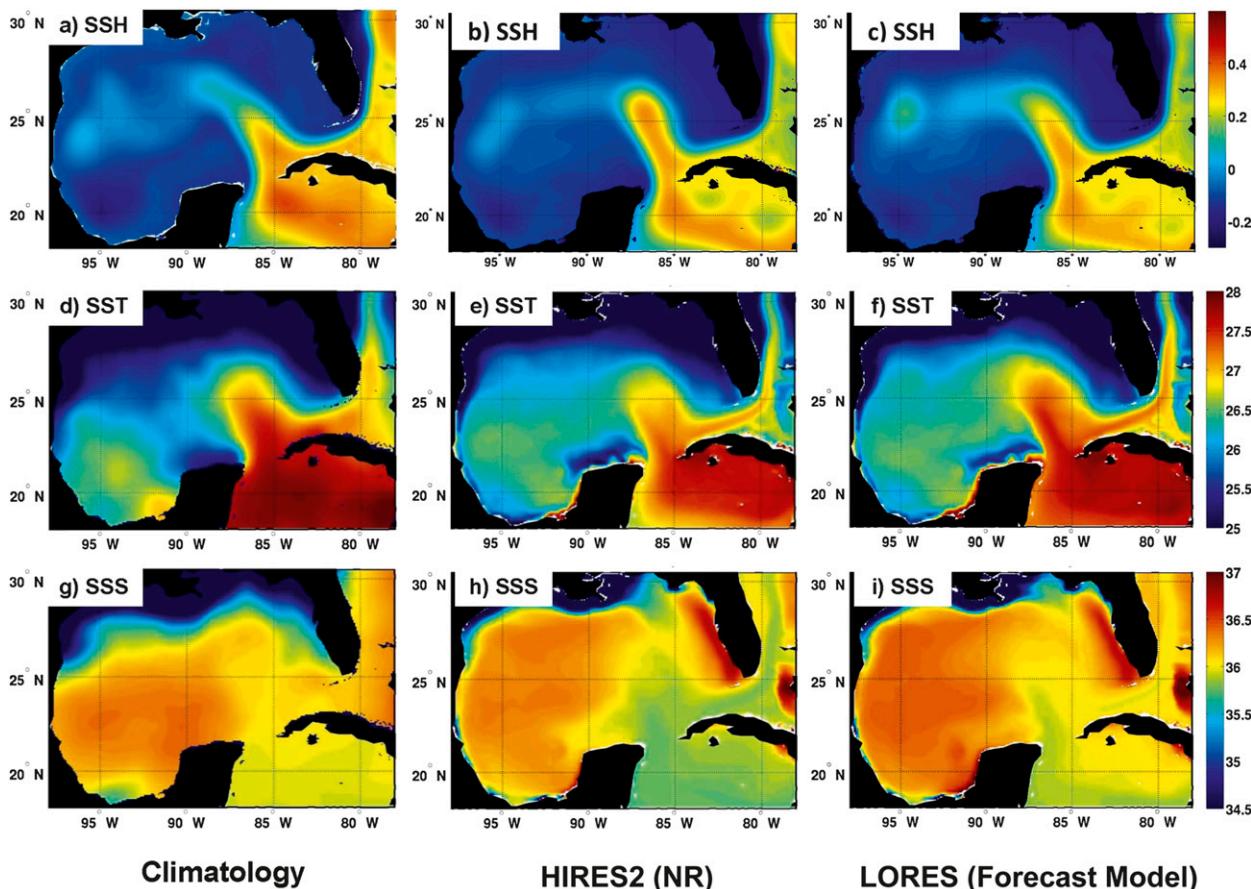
FIG. 3. Mean climatology of (a)–(c) SSH (m), (d)–(f) SST (°C), and (g)–(i) SSS (psu). (a) Mean SSH is from the CNES mean dynamic topography derived from altimetry and obtained from AVISO. Climatological mean (d) SST and (g) SSS are obtained from the U.S. Navy GDEM3 climatology. Other mean fields are derived from the (b),(e),(h) NR model (experiment HIRES2) and the (c),(f),(i) forecast model (experiment LORES).

experiment (LORES; Table 2) that was run using the forecast model as configured in the DAS. LORES is therefore identical to HIRES1, except for being run at lower resolution (0.08°). The 6-yr (2005–10) mean fields of SSH, SST, and SSS from experiments HIRES2 and LORES are compared to each other and to mean fields of SSH obtained from the Centre National d'Etudes Spatiales (CNES) mean dynamic topography derived from altimetry and mean fields of temperature and salinity from the U.S. Navy Generalized Digital Environment Model version 3 (GDEM3) ocean climatology (Carnes 2009). The mean patterns of all variables produced by the two experiments are similar to each other and to the climatological mean patterns (Fig. 3). This similarity demonstrates that both the NR and forecast models produce statistically realistic climatological structure as required for a valid OSSE system. The SSH patterns outline the impact of the mean penetration of the LC and also display an east–west ridge of SSH extending

westward across the central latitudes of the Gulf that denotes the pathway of westward-propagating anticyclones that detach from the LC. The mean SST patterns from both models are dominated by the protrusion of warm water associated with the mean LC superimposed on a general northward decrease. The mean SSS patterns from both models show high SSS in the western interior GOM and lower SSS in the eastern interior Gulf.

Close inspection of the model mean fields reveals differences in structural detail, demonstrating the impact from the different physics and truncation errors as required for an OSSE system. For the most part, differences between the models are no larger than the differences between each model and climatology, and thus satisfy the requirement that both models reproduce statistically realistic climatology. The mean extension of the LC is sharper in the models, which results in part from higher model resolution and shorter model temporal averaging interval compared to climatology. An
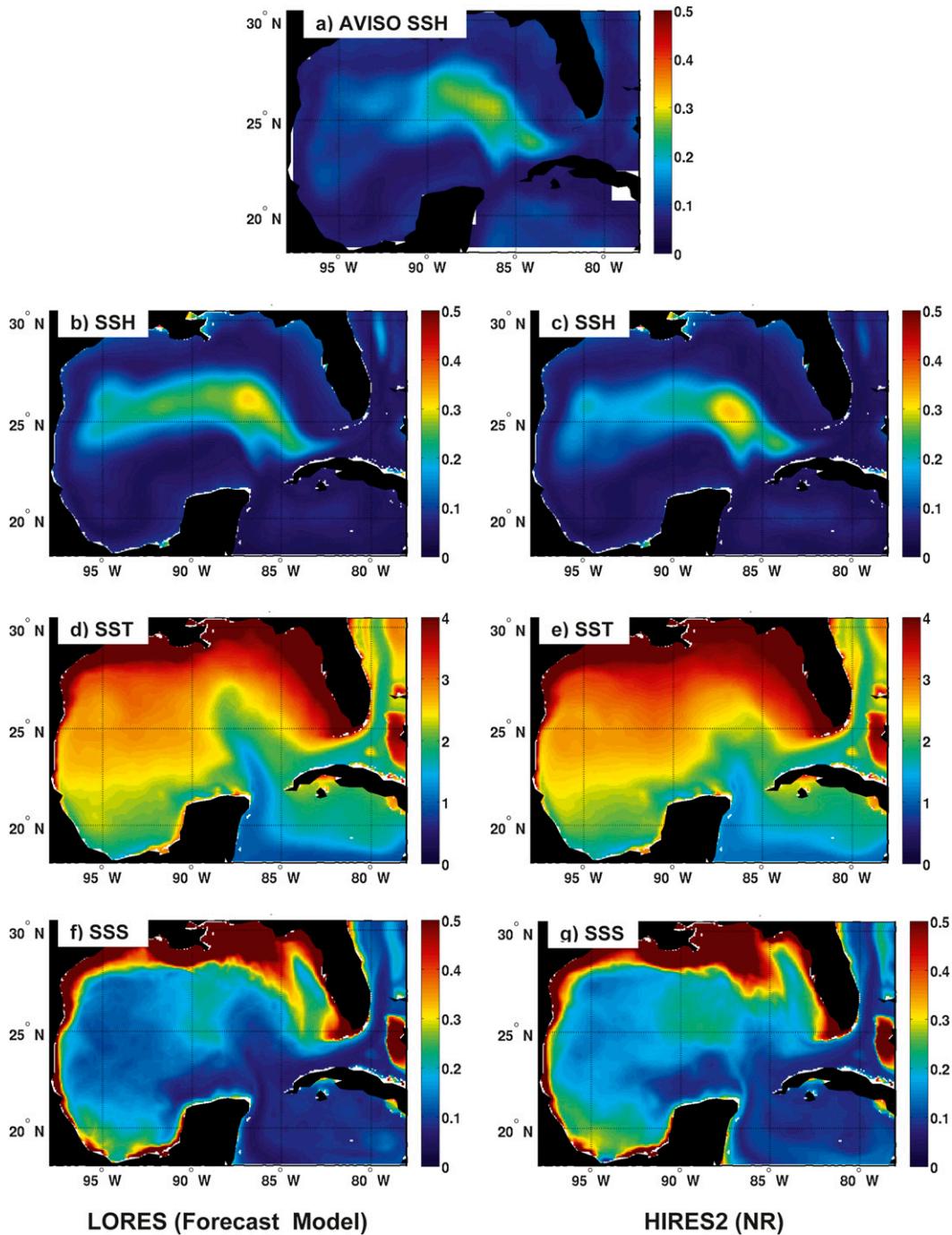
FIG. 4. (a) 2005–10 RMS amplitude field of SSH (m) from weekly AVISO maps, RMS amplitude fields from the unconstrained 0.08° hybrid coordinate forecast model run (experiment LORES) for (b) SSH (m), (d) SST (°C), and (f) SSS (psu). RMS amplitude fields from the unconstrained 0.04° fixed coordinate NR (experiment HIRES2) for (c) SSH (m), (e) SST (°C), and (g) SSS (psu).

obvious exception to statistically realistic climatology exists for mean SSS over the west Florida shelf, where both model salinities are much larger than climatology, possibly due to the use of climatological river runoff in the models. Although the NR salinity is not valid in this region, this does not impact the present study, which focuses on the open Gulf.

Concerning the representation of ocean variability, RMS amplitude maps of SSH, SST, and SSS fluctuations from both HIRES2 and LORES all have similar spatial
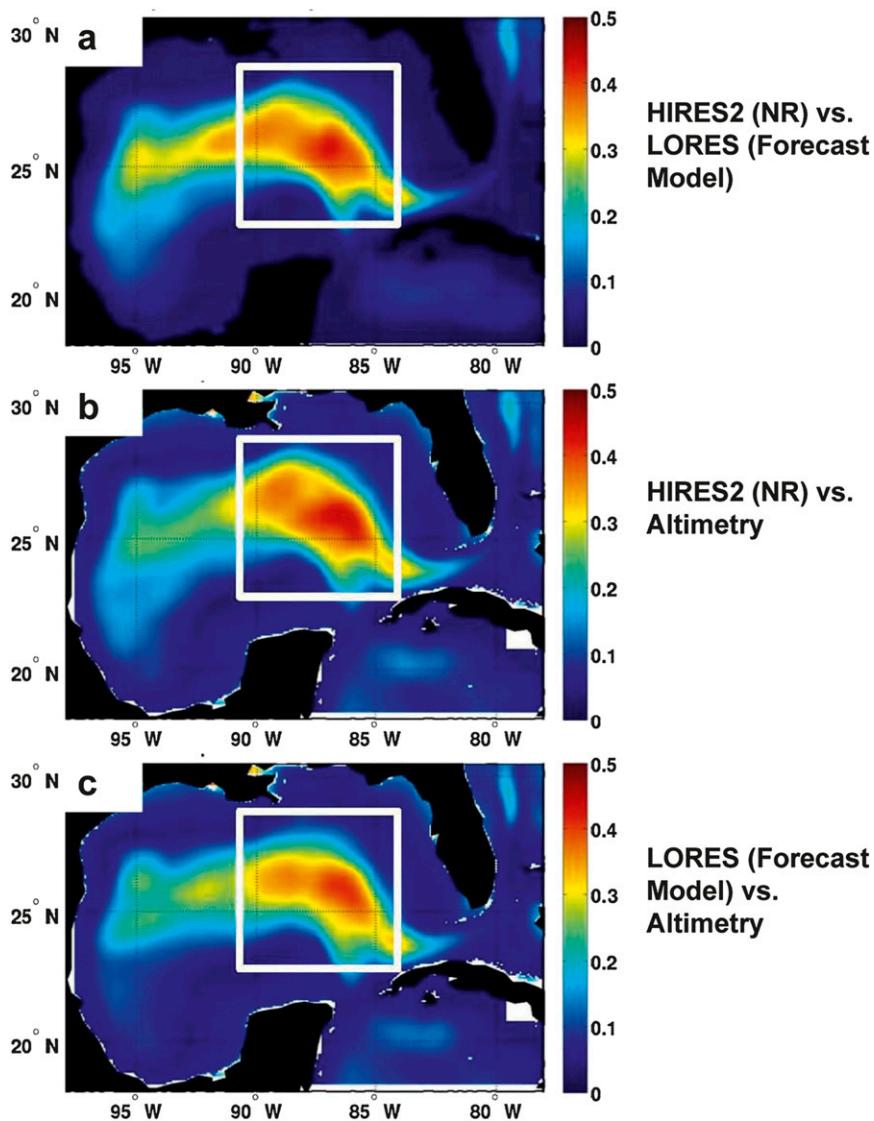
FIG. 5. RMS difference of SSH (m) between (a) the NR (HIRES2) and forecast (LORES) models, (b) between the NR model and AVISO altimetry, and (c) between the forecast model and AVISO altimetry. The boxes outline the interior GOM region where a detailed evaluation of the OSSE system and observing system impact was conducted.

structure (Fig. 4). The SSH amplitude is largest near 26°N, 87°W, where the LC variability is large and eddy shedding frequently occurs. A ridge of large RMS variability extends westward along the preferred pathway of detached anticyclonic eddies. Both SST and SSS variability (Fig. 4) tend to increase toward the north and are large in northern coastal regions, where the response to seasonal and synoptic atmospheric forcing and to river runoff is largest. Variability tends to be smallest within the LC and Florida Current waters. Comparing the RMS amplitude of model SSH to altimetry-derived SSH (Fig. 4), the broad structures are similar but larger model peak amplitudes exist in the eddy shedding

region, while the westward extension of large SSH amplitude along the detached eddy pathway is less pronounced in climatology. These model-climatology amplitude differences may substantially result from higher model resolution and shorter model temporal averaging interval compared to climatology. Consequently, the models appear to represent the distribution of variability with sufficient realism for use in a valid OSSE system. Furthermore, differences in the detailed structure of the amplitude patterns between the two models again reveal impacts from differences in physics and truncation errors required by the OSSE system.

b. Model error analysis

Since the NR cannot be initialized with a perfect representation of the true ocean, subsequent error growth rates cannot be properly determined. As a result, evaluation of the two models must out of necessity rely on comparisons of the magnitude and distribution of RMS errors between the two models and compare these to the same error statistics between the two models and the true ocean. In the present analysis, weekly AVISO gridded SSH maps from 2005 to 2010 are interpolated to the model grids and these error statistics are calculated (Fig. 5). Consideration was given to comparing satellite-derived SST maps, but SST is dominated by the annual cycle and model SST variability tends to follow the imposed surface air temperature, limiting the usefulness of this comparison. Because a more stringent evaluation of error growth between the models is not possible, the OSSE system validation critically relies on the OSE–OSSE comparisons described in section 5.

The magnitude and pattern of RMS errors are very similar between the two models, and also between each model and observed fields (Fig. 5). In all cases, large values exist in the LC eddy shedding region and extend to the west along the eddy pathway. This substantial similarity among RMS error statistics demonstrates that the two model configurations are reasonable choices for the OSSE system. The OSSE validation effort will be conducted within the box shown in the panels of Fig. 5, which encloses the domain within which the P-3 aircraft surveys were conducted. Within this subregion, time series of spatial RMS error between experiments HIRES2 and LORES, and also between each experiment and weekly AVISO maps, are plotted (Fig. 6). In each case, the magnitude of this error oscillates over time as the SSH patterns by chance either more closely or less closely resemble each other. However, the mean RMS error magnitudes and the ranges over which they oscillate are very close among the three cases. The evidence does not support rejecting these model configurations.

5. OSSE system validation

a. Experiments

The previous determination that two chosen model configurations substantially meet the requirements of a valid fraternal twin OSSE system does not guarantee valid impact assessments will be produced. Additional evaluation is required for system validation, which involves comparing OSSEs to reference OSEs. To perform this evaluation, seven additional experiments are analyzed (Table 2). Experiment OSE1 assimilates all real observations, while OSE2 denies the WP-3D profiles and
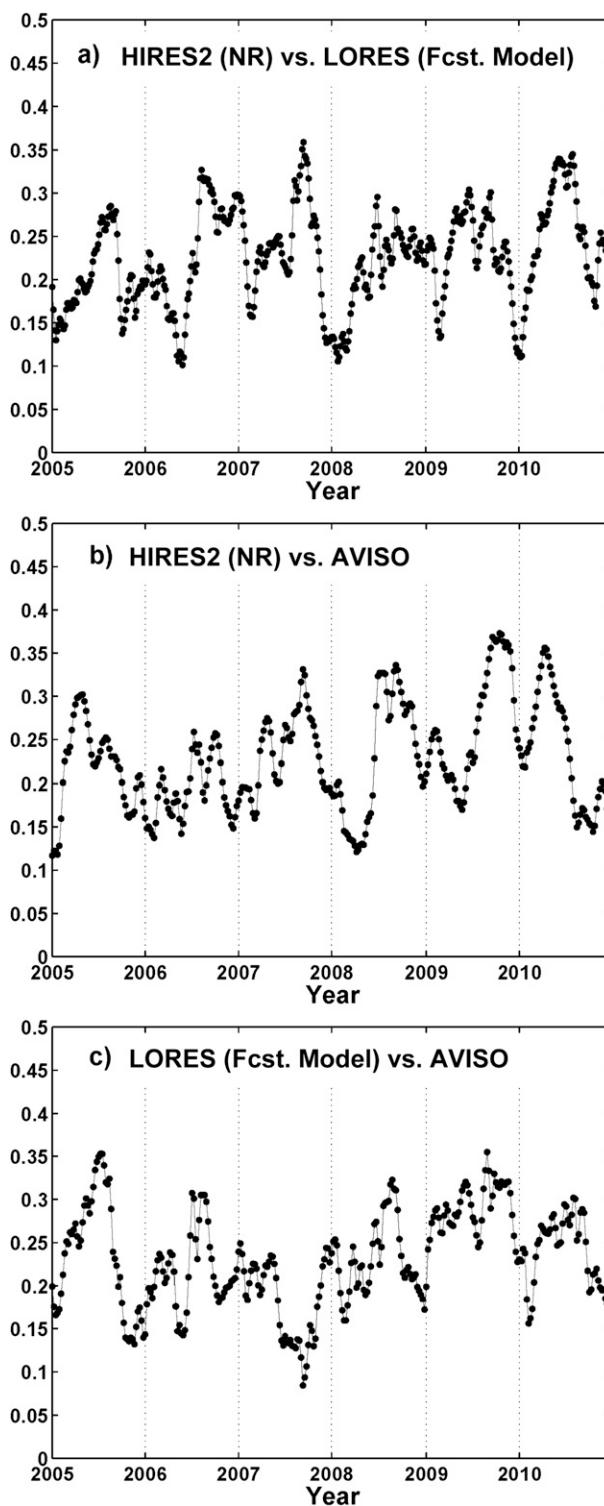

FIG. 6. Time series of SSH RMS amplitude (m) calculated over the boxes outlined in Fig. 5 for (a) the NR (HIRES2) vs the forecast (LORES) model, (b) the NR model vs AVISO altimetry, and (c) the forecast model vs AVISO altimetry.

TABLE 5. Estimates of integral correlation scales over time and in the zonal and meridional directions.

Field	Integral time scale (days)	Integral zonal scale (km)	Integral meridional scale (km)
H_{20}	55	172	153
TCHP	57	164	155
u (15 m)	33	146	97
v (15 m)	27	95	120

OSE3 further denies two of the three altimeters (*Jason-2* and *Envisat*). Experiments OSSE1, OSSE2, and OSSE3 are performed that are identical except for assimilating synthetic observations. All of the OSE and OSSE runs are initialized on 1 January 2010 using fields from experiment LORES, and use the same atmospheric forcing and ocean boundary conditions as LORES. For comparison to unconstrained simulation results, an experiment without data assimilation (DAFREE) is used, which consists of fields extracted from LORES over the same time interval as the OSE–OSSE pairs.

b. Evaluation methodology

The OSE experiments are evaluated by comparing actual airborne profiles to the same profiles extracted from the experiments. The OSSE experiments are evaluated by comparing synthetic airborne profiles simulated from the NR with realistic errors added to the same synthetic profiles extracted from the experiments. Evaluation of OSEs is generally impeded by the limited availability of unassimilated observations. This is particularly true for determining the impact of withholding airborne profiles by comparing experiments OSE1 to OSE2 because temperature and salinity profiles were assimilated by OSE1. This is not an issue in comparing OSE2 to OSE3 or OSE3 to DAFREE because no airborne profiles were assimilated. Two sets of unassimilated observations are available to evaluate the impact of withholding airborne profiles: velocity profiles sampled by AXCPs and trajectories of surface drifters. To be consistent with the surface drifters, which are drogued to a depth of 15 m, velocity components at a depth of 15 m are extracted from the AXCP profiles for evaluation.

The temperature profiles provide three fields for evaluation. (Salinity profiles from AXCTDs are too sparse to provide robust statistics.) First, observed profiles spanning the upper 250 m of the water column are compared to synthetic profiles at the same locations extracted from model fields. To evaluate the impact of data denial on the horizontal structure of ocean features, maps of 20°C isotherm depth (H_{20}) are calculated from the profiles. Finally, tropical cyclone heat potential (TCHP; Leipper and Volgenau 1972; Mainelli et al. 2008), also referred to as ocean heat content (OHC), is calculated. TCHP is the thermal energy required to heat all near-surface water above 26° from 26°C to the observed temperature. One of the initial planned applications of the OSSE system will be to evaluate observing strategies for improving ocean model initialization for coupled hurricane forecast models. Accurate initialization of TCHP is important for the ocean model to provide accurate SST forecasts.

Several types of statistical comparisons are performed. RMS error and mean bias are calculated on individual flight days between modeled and observed fields and presented as time series. Other bulk statistical analyses are performed using all airborne observations collected over the nine flight days, which is especially important for velocity field evaluation because of the limited number of velocity profiles collected on individual flight days in comparison to temperature. These bulk analyses include Taylor (2001) diagrams, which simultaneously illustrate three related error metrics between two fields: correlation coefficient, RMS amplitude, and RMS error. They also include the Murphy (1988) skill score Σ, defined as

$$\Sigma = r^2 - [r - (\sigma_Y/\sigma_X)]^2 - [(\overline{Y} - \overline{X})/\sigma_X]^2, \quad (2)$$

where r is the correlation coefficient and $\overline{X}$, $\overline{Y}$, σ_X, and σ_Y are the means and standard deviations of the two fields. It equals the squared correlation coefficient reduced by errors in both the mean values and RMS amplitudes, which can decrease Σ to the point where it becomes negative. The skill is considered to be significant if $\Sigma > 0$. The Taylor diagrams and Σ analyses partly complement each other by having two metrics in common: correlation coefficient and RMS amplitude. However, Taylor diagrams also display the impact of RMS differences but do not include the impact of mean bias, which is included in Σ.

To analyze errors in trajectories between real surface drifters and synthetic surface drifters released at the same locations as the real drifters, and then advected by model velocity fields, the trajectory skill score proposed by Liu and Weisberg (2011) is used:

$$S = \begin{cases} 1 - s/n & s \leq n \\ 0 & s > n \end{cases}, \quad (3)$$

where the index $s = \sum_{i=1}^{M} d_i / \sum_{i=1}^{M} l_{oi}$ is the sum of the separation distances (d_i) at times t_i of individual position fixes up to time t_M of the final position fix divided by the total length of the observed trajectory (l_{oi}) at time t_M. Parameter n (which is set to 1) is a user-defined tolerance threshold representing the critical value of no skill ($S = 0$).
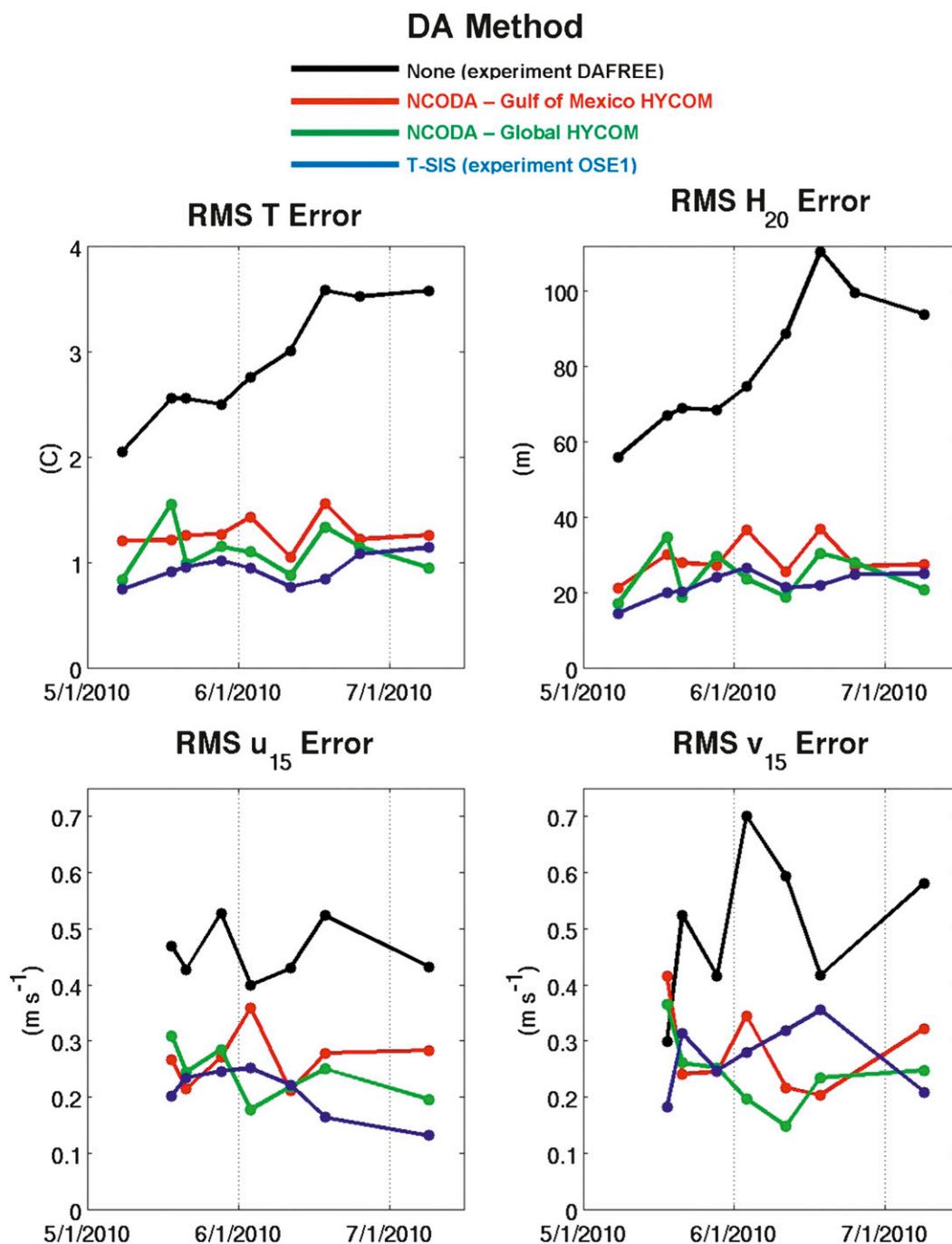
DA Method

FIG. 7. RMS error between model profiles and observed airborne profiles for (top left) all temperature values measured between 0 and 250 m, (top right) H_{20} derived from T profiles, (bottom left) u_{15}, and (bottom right) v_{15}. The legend at the top summarizes the DA method used in the experiments. All observations including airborne T and S profiles (but not velocity) were assimilated in these experiments.

Given the long time scales of LC variability, concern about the number of degrees of freedom in these statistical analyses is justified. To explore this issue, integral temporal, zonal, and meridional correlation scales are calculated using Eq. (1) from daily maps of four variables from the unconstrained experiment DAFREE from May through December 2010 (Table 5). Integral time scales for velocity components are about one-half
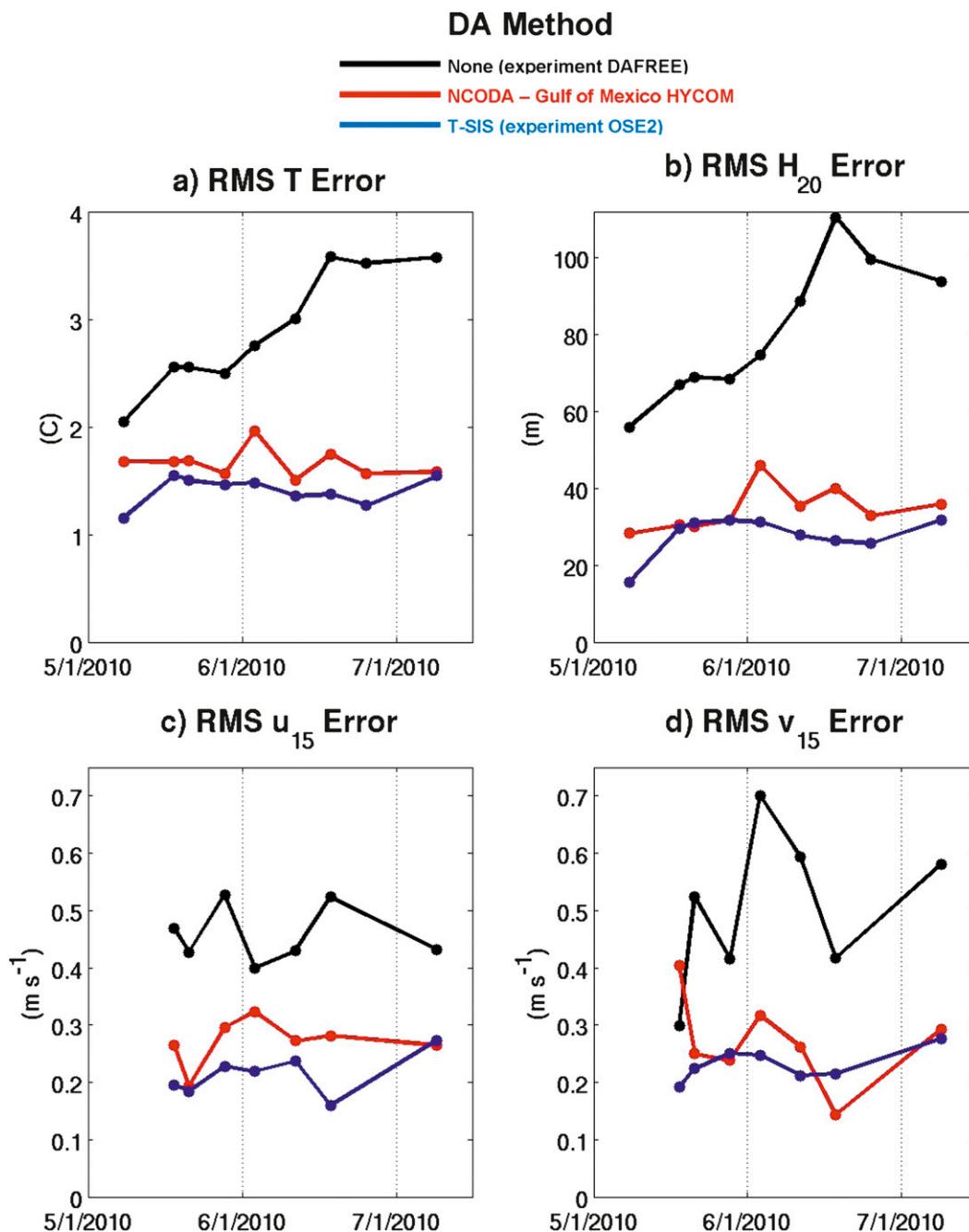
DA Method

— None (experiment DAFREE)
— NCODA – Gulf of Mexico HYCOM
— T-SIS (experiment OSE2)

FIG. 8. RMS error between model profiles and observed airborne profiles for (a) all temperature values measured between 0 and 250 m, (b) H_{20} derived from T profiles, (c) u_{15}, and (d) v_{15}. The legend at the top summarizes the DA method used in the experiments. In these experiments, airborne profiles were not assimilated and are used solely for evaluation.

of the scales for H_{20} and TCHP. The 2-month span of the airborne surveys equals approximately two integral time scales for the velocity components and one integral time scale for the other fields. Fortunately, additional degrees of freedom are obtained from the spatial extent of the sampling. Integral zonal and meridional scales range

from slightly less than 100 to 172 km, with slightly smaller values on average for the velocity components compared to the other fields. The spatial coverage of the airborne surveys typically spans 3–4 integral zonal and meridional scales for velocity components and ≤3 integral zonal and meridional scales for the other fields.
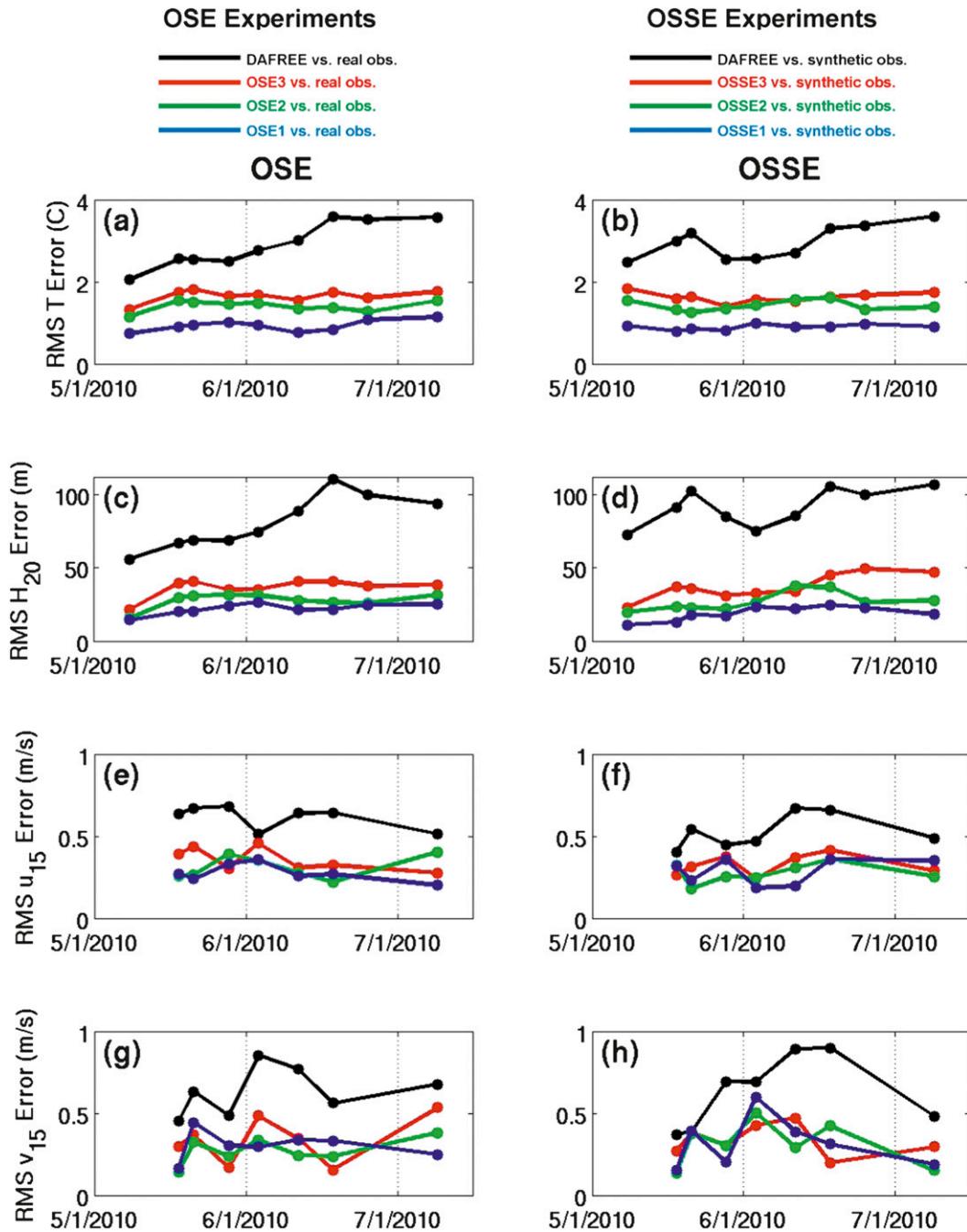
FIG. 9. RMS error between model profiles and observed airborne profiles of four variables: all temperature values measured between 0 and 250 m for (a) OSE and (b) OSSE experiments, H_{20} for (c) OSE and (d) OSSE experiments, u_{15} for (e) OSE and (f) OSSE experiments, and v_5 for (g) OSE and (h) OSSE experiments. Specific experiments are listed in the legend at top.

Thus, the two-dimensional fields span about 25 independent space–time integral correlations scales for velocity components and about 8 independent space–time integral correlation scales for the other fields. In evaluating model velocity errors versus AXCP profiles,

the apparent advantage of a larger number of space–time correlation scales is diminished because relatively few velocity profiles were sampled compared to temperature. Overall, the number of quasi-independent samples in the observational datasets is less than ideal,

TABLE 6. The mean RMS errors for all of the time series graphed in Fig. 9 along with percentage increases in this error due to denying different components of the observing system for four different fields. Results are shown for both the OSE and OSSE experiments. The bold percentage increases shown in parentheses resulting from the denial of airborne observations for both velocity components were calculated from experiments performed by the U.S. Navy using the GOM HYCOM analysis system and reported in Shay et al. (2011).

Field	Observations assimilated	Mean RMS error		Percentage increase of mean RMS error	
		OSE	OSSE	OSE	OSSE
T from surface to 250 m	All	0.93	0.90		
	Deny airborne observations	1.41	1.42	51	58
	Further deny two of three altimeters	1.66	1.62	17	14
	None (DAFREE)	2.90	2.97	75	83
H_{20}	All	20.6	18.9		
	Deny airborne observations	27.8	27.0	35	43
	Further deny two of three altimeters	36.5	37.2	31	38
	None (DAFREE)	82.7	91.5	121	146
u_{15}	All	0.34	0.29		
	Deny airborne observations	0.31	0.28	−11 (**−7**)	−3
	Further deny two of three altimeters	0.36	0.32	15	18
	None (DAFREE)	0.61	0.53	75	62
v_{15}	All	0.30	0.32		
	Deny airborne observations	0.27	0.31	−10 (**−1**)	−3
	Further deny two of three altimeters	0.34	0.34	24	8
	None (DAFREE)	0.63	0.63	88	88

but results of the subsequent evaluation demonstrate that robust statistical results are obtained, more so for temperature and fields derived from temperature compared to velocity components.

We confine the evaluation of the OSSE system to the open Gulf of Mexico because of the lack of altimetry assimilation within the 300-m isobath, the use of climatological river runoff in the model, and the large SSS errors over the west Florida shelf. In the future, the OSSE system evaluated herein will be ported to a new 0.02° HYCOM-based Gulf of Mexico nowcast–forecast system and to higher-resolution coastal models nested within it that will all employ realistic high-frequency river runoff (Schiller et al. 2011) for the purpose of evaluating coastal ocean observing systems.

c. Evaluation of the T-SIS DA methodology

Before conducting the OSSE system evaluation, the performance of the new T-SIS DA methodology is analyzed in comparison to two operational HYCOM Navy Coupled Ocean Data Assimilation (NCODA) ocean analysis products produced by the U.S. Navy using the operational HYCOM nowcast–forecast system (e.g., Chassignet et al. 2007). Because these products assimilated all available observations including the airborne profiles, they are compared to experiment OSE1. The quality of the analysis products is measured by the error reduction resulting from assimilation in comparison to the unconstrained experiment DAFREE (Fig. 7). For temperature between the surface and 250 m on each of

the nine flight days, the largest RMS errors from DAFREE range between 2° and 4°C (Fig. 7a). All three analysis products produced substantial error reduction to values averaging close to 1°C. The T-SIS product produced the largest error reduction on most days, ranging between 0.7° and 1.1°C. Large error reduction is also achieved for H_{20} by all products, with T-SIS again producing slightly smaller errors than the others (Fig. 7b). All three analysis products produced error reduction in the velocity components, although the fractional reduction is less than for the other fields (Figs. 7c,d). Although the velocity component statistics are noisy because of relatively sparse velocity sampling limited to seven of the nine flight days, error reduction by T-SIS is still comparable to error reduction by the other products.

Because temperature profiles were assimilated and H_{20} maps were calculated from these profiles, the larger error reduction produced by the T-SIS is encouraging but essentially illustrates an improved goodness-of-fit rather than improved performance. A second evaluation is therefore performed by comparing error reduction due to altimetry assimilation between the T-SIS experiment OSE2 and an experiment run with the HYCOM Gulf of Mexico nowcast–forecast system that also assimilated all observations except the airborne profiles (Shay et al. 2011). These experiments are then evaluated against the airborne profiles that, this time, were not assimilated (Fig. 8). Essentially the same results are obtained, validating the decision to use the new T-SIS methodology for the OSSE system.
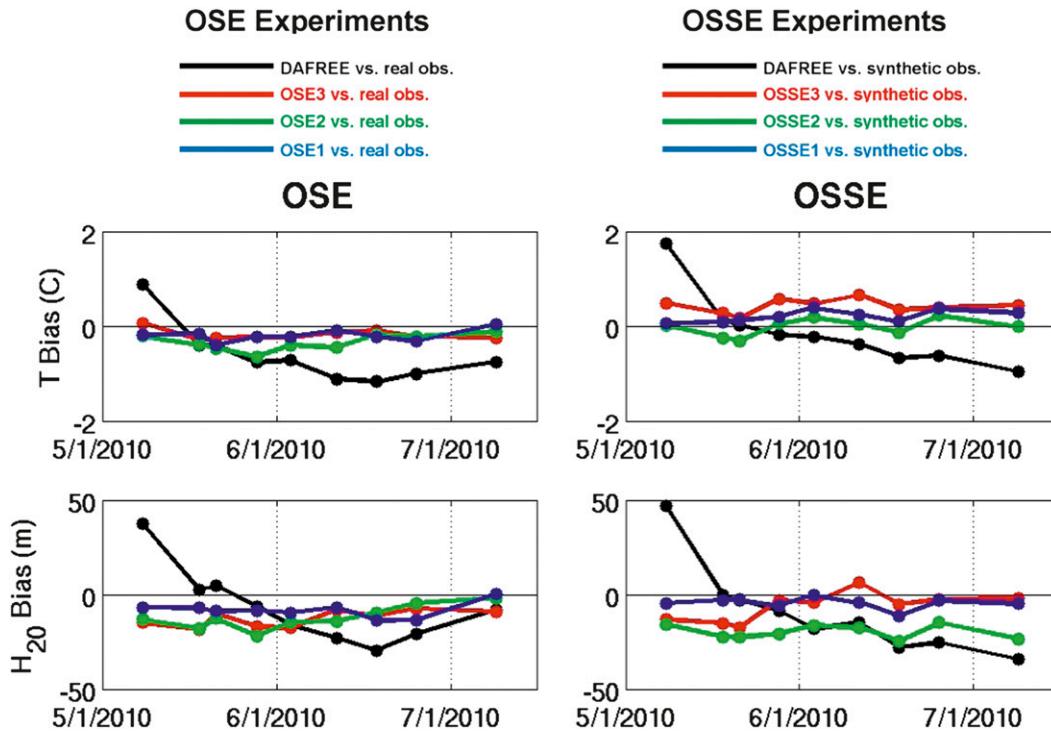
Fig. 10. As in Fig. 9, but for mean bias of two of the fields: temperature between 0 and 250 m for (top left) OSE and (top right) OSSE experiments, and H_{20} for (bottom left) OSE and (bottom right) OSSE experiments.

d. OSSE system evaluation using airborne observations

Comparing temperature profiles between the surface and 250 m, OSE1 produces the smallest RMS error, ranging from 0.7° to 1.1°C (Fig. 9). RMS errors of all time series graphed in Fig. 9 averaged over all flight days, along with the percentage increase in RMS errors resulting from denial of observations, are listed in Table 6. The mean RMS error for OSE1 is 0.93°C. Results for OSSE1 are very similar, with a mean RMS error of 0.90°C. Denial of airborne observations in OSE2 (OSSE2) increased errors by an average of 51% (58%) while further denial of two of the three altimeters in OSE3 (OSSE3) only resulted in an additional error increase of 17% (14%). Denial of all remaining observations in DAFREE, which primarily reveals the impact of denying the single altimeter assimilated in OSE3 (OSSE3), further increased errors by an average of 75% (83%). Altimetry observations and airborne profiles both have a large impact on reducing upper-ocean temperature errors. The close corresponding between the OSE and OSSE results is encouraging although some differences are evident on individual flight days. In addition to statistical uncertainty of error estimates on individual flight days, the nonuniform sampling patterns conducted on the nine flight days (Shay et al. 2011) contributes to day-to-day error differences.

The smallest RMS errors in H_{20} were produced by OSE1 (OSSE1) with mean values of 20.6 m (18.9 m). Denial of airborne observations in OSE2 (OSSE2) increased errors by an average of 35% (43%), while further denial of two of the three altimeters in OSE3 (OSSE3) resulted in an additional error increase of 31% (38%). Denial of all remaining observations in DAFREE further increased errors by an average of 121% (146%). Close correspondence between OSE and OSSE results are again realized, but one notable difference in observing system impact based on H_{20} versus temperature error reduction is the comparatively larger impact of denying two altimeters in both OSE3 and OSSE3. As expected, altimeters play the dominant role in constraining the horizontal structure of oceanic boundary currents and mesoscale eddies as outlined by H_{20} maps. By contrast, the impact of airborne profiles on correcting H_{20} is somewhat smaller than on correcting upper-ocean temperature.

RMS errors in u_{15} and v_{15} extracted from airborne velocity profiles (Fig. 9; Table 6) demonstrate that assimilation of a single altimeter has the largest impact on reducing velocity errors, while assimilation of additional altimeters and airborne profiles has little additional
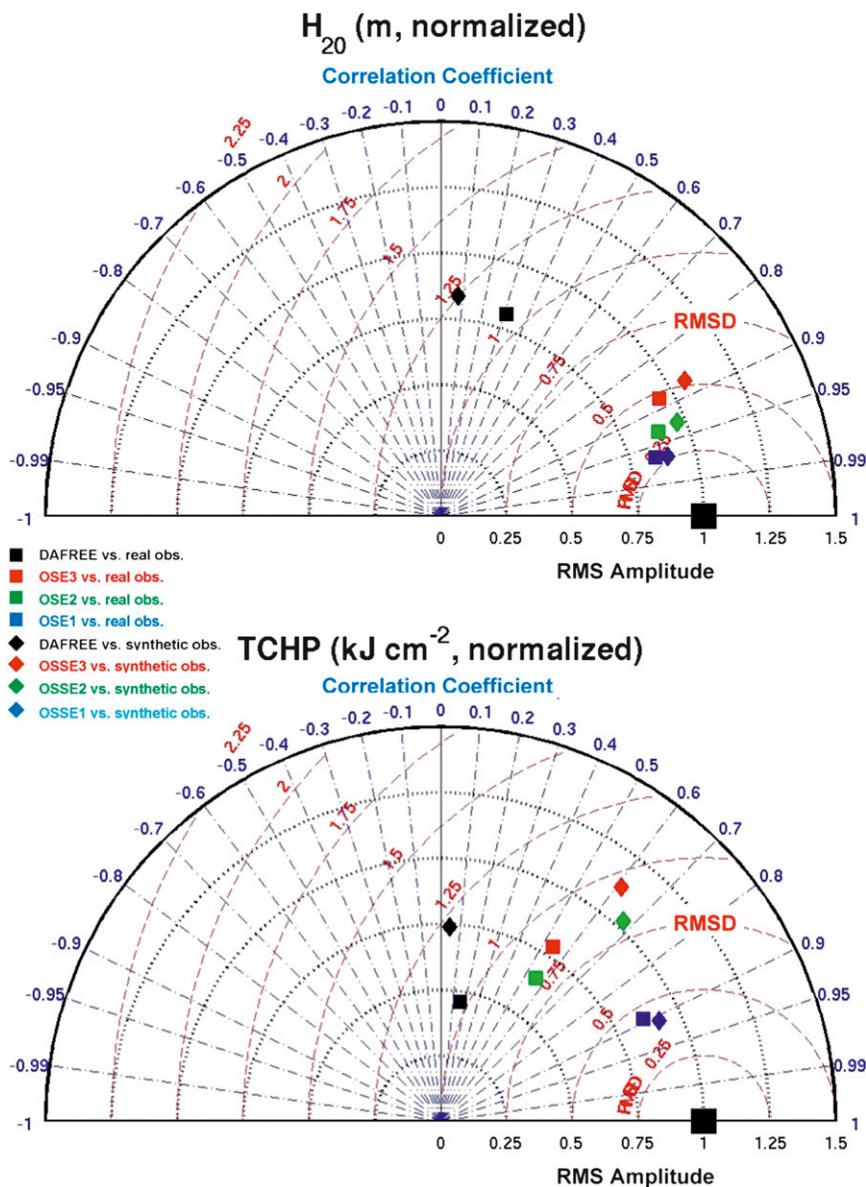
FIG. 11. Taylor diagrams comparing H_{20} and TCHP values at the locations of all AXBT, AXCTD, and AXCP drops (which all recorded temperature) over the nine WP-3D flight days. Mean values for each flight day were removed before calculating the statistical metrics. The large black square denotes a perfect comparison between the two fields. The RMS error axis is represented by red dashed curves, the RMS amplitude axis is represented by black dotted curves, and the correlation axis is represented by blue dashed–dotted lines. A large black square marks the point of perfect comparison. The legend at left identifies points from the individual experiments.

impact. The OSE and OSSE experiments again produce similar results. The relatively sparse velocity sampling results in noisy statistics that contribute to the difficulty in detecting impacts of assimilating other observations beyond one altimeter. One interesting result is that denial of airborne temperature and salinity profiles actually results in decreased velocity errors (Table 6). This

surprising result motivated additional effort to determine if this indicated problems with the T-SIS DA algorithm. The OSE conducted using the U.S. Navy HYCOM Gulf of Mexico nowcast–forecast system to evaluate the impact of airborne profile assimilation during the nine flight days and reported in Shay et al. (2011) was revisited to analyze the impact on velocity

TABLE 7. Murphy skill scores (Σ) from Eq. (2) for four different fields and for both the OSE and OSSE experiments.

Field	Observations assimilated	Σ skill score	
		OSE	OSSE
H_{20}	All	0.92	0.93
	Deny airborne observations	0.87	0.86
	Further deny two of three altimeters	0.77	0.73
	None (DAFREE)	−0.14	−0.57
TCHP	All	0.79	0.82
	Deny airborne observations	0.30	0.33
	Further deny two of three altimeters	0.23	0.11
	None (DAFREE)	−0.06	−0.48
u_{15}	All	0.63	0.43
	Deny airborne observations	0.60	0.50
	Further deny two of three altimeters	0.56	0.30
	None (DAFREE)	−0.31	−0.82
v_{15}	All	0.42	0.50
	Deny airborne observations	0.54	0.53
	Further deny two of three altimeters	0.26	0.40
	None (DAFREE)	−1.39	−0.88

errors. Results from the U.S. Navy experiment that assimilated all observations were used in Fig. 7 to evaluate the performance of the T-SIS system. Comparison to the experiment that denied the airborne profiles revealed that u_{15} and v_{15} errors decreased by 7% and 1%, respectively, because of denial compared to 11% and 10%, respectively, produced by the T-SIS OSE (Table 6). Although the decreases were smaller, the results were similar and the T-SIS DA methodology cannot be singled out as being flawed. Further research is warranted to understand why this happens and to devise strategies to improve the quality of ocean DA systems.

Impacts of assimilation and data denial on bias reduction are also investigated on the nine flight days for upper-ocean temperature and H_{20} (Fig. 10). In general, similar reduction of bias due to assimilation is evident in both fields for both the OSE and OSSE experiments. Most of the bias reduction is achieved through the assimilation of a single altimeter by OSE3 and OSSE3. Additional bias reduction is not clearly evident when the remaining altimeters and then the airborne profiles are assimilated.

Because of the limited statistical significance of error analysis on individual flight days, particularly for velocity components, statistical error analysis is also conducted over the combined nine flight days. In calculating RMS errors, mean values of all analyzed fields are calculated and removed separately on each flight day. Errors in both H_{20} and TCHP are evaluated using Taylor

diagrams (Fig. 11). Prior to calculating the statistics for these diagrams, the RMS amplitudes of all fields are normalized by the RMS amplitude of the observed field. For reference, all Taylor diagrams presented herein contain a large black square plotted at the location indicating a perfect comparison between two fields—that is, a correlation coefficient of 1.0, identical normalized RMS amplitudes of 1.0, and zero RMS error. The largest error reduction for H_{20} is achieved by the assimilation of a single altimeter in both OSE3 and OSSE3, while smaller improvements are achieved by the assimilation of all three altimeters in OSE2 and OSSE2, and the airborne profiles in OSE1 and OSSE1. Given that H_{20} is a proxy for the horizontal structure of the LC and adjacent eddies, the dominant importance of assimilating at least one altimeter in correctly locating ocean features is evident. A small additional error reduction is achieved by assimilating the airborne profiles, in agreement with the time series in Fig. 9.

For TCHP, assimilation of altimetry produces only a small improvement in accuracy in both the OSE and OSSE experiments. Instead, the largest error reduction in both sets of experiments is achieved by the assimilation of the airborne profiles. These conclusions are supported by the skill scores Σ for H_{20} and TCHP (Table 7). The scores for DAFREE are negative and insignificant for both fields and are large and significant for both OSE1 and OSSE1. Denial of airborne observations produces slightly smaller skill scores for H_{20} but much smaller (although still significant) scores for TCHP in both OSE2 and OSSE2. Overall, the same impact assessments are realized from the OSE and OSSE experiments.

The Taylor diagrams for u_{15} and v_{15} (Fig. 12) clearly demonstrate that the assimilation of a single altimeter has the greatest impact on error reduction in both the OSE and OSSE experiments. Assimilation of additional altimeters and the airborne profiles does not produce significant further error reduction in both the OSEs and OSSEs. These assessments are supported by the skill score analysis (Table 7). The skill scores again reveal the tendency for errors to decrease when airborne profiles are denied for u_{15} (OSSE1 vs OSSE2) and v_{15} (OSE1 vs OSE2 and OSSE1 vs OSSE2).

e. OSSE system evaluation using surface drifter trajectories

Additional evaluation of observing system impact on current velocity analysis is performed against the set of Gulf of Mexico surface drifters drogued at 15 m that was released by the NOAA AOML starting in early June 2010 in response to the *Deepwater Horizon* oil spill (Fig. 12; Lumpkin and Elipot 2010). For comparison,
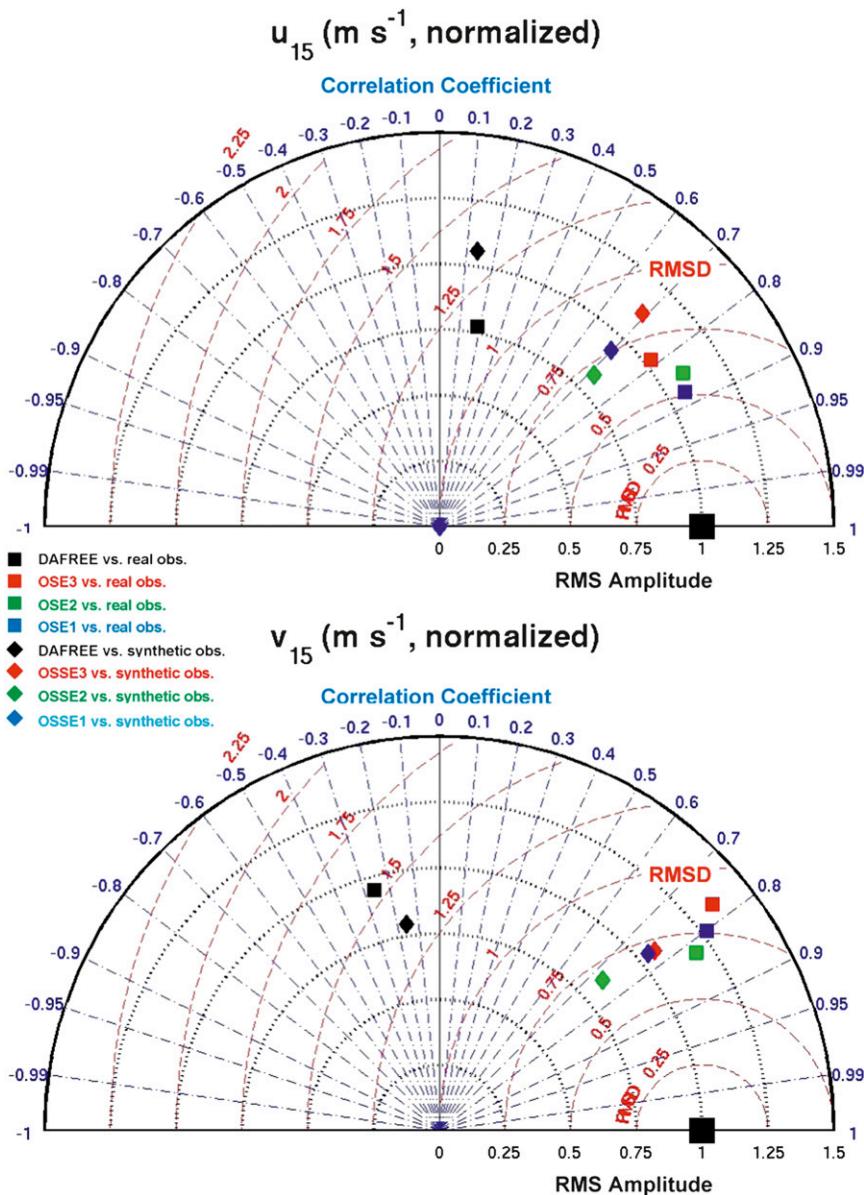
FIG. 12. As in Fig. 11, but comparing u_{15} and v_{15} from OSE experiments (squares) and OSSE experiments (diamonds) over all nine WP-3D flight days. Vertical means were removed from each profile before calculating the statistical metrics.

synthetic drifters are released in archived model fields at the locations of the actual drifters every 2 days from 12 June to 10 December, and then advected by a trajectory model using 15-m velocity fields from the experiments. Synthetic particle releases are confined to the subdomain bounded by 22°–29°N, 91°–84°W, as illustrated by the white boxes in Fig. 5. Particle advection is performed using fourth-order Runge–Kutta horizontal interpolation of the velocity field (Garraffo et al. 2001). A Lagrangian stochastic model is used to add unresolved

small-scale velocity fluctuations. Further details of this trajectory model are presented in Mariano et al. (2011).

The trajectory skill score analysis is only applied to evaluate the impact of denying two of three altimeters and further denying all observations because airborne profiles were only collected during a small fraction of the June–December time interval of this analysis. In calculating S from Eq. (3), 6-hourly position fixes were used out to a maximum fix time t_M of 40 h. For both the OSE

TABLE 8. Trajectory skill scores for several experiments calculated using the method of Liu and Weisberg (2011) [Eq. (3) with $n = 1$].

Experiment	Trajectory skill score
OSE2	0.35
OSE3	0.26
DAFREE (vs observations)	0.09
OSSE2	0.31
OSSE3	0.24
DAFREE (vs NR)	0.10

and OSSE experiments, the largest increase in S resulted from assimilating a single altimeter, as was true in the previous analyses in comparison to airborne measurements (Table 8). The trajectory skill score increased when the additional two altimeters were assimilated. This improvement was not noted in the Taylor diagram analysis in Fig. 11 comparing model and observed velocity component profiles. Roughly similar results are obtained from the OSE and OSSE experiments.

f. OSSE system validation

The overall similarity of impact assessments obtained from the OSSEs and reference OSEs demonstrate that calibration will not be necessary and that the OSSE system as configured for the interior Gulf of Mexico can be declared valid for evaluating new observing systems. Although the OSE and OSSE results are very similar on average, significant uncertainty can exist in any single impact assessment based on error analysis for a single field that is not extensively sampled. To limit this uncertainty, it is necessary to evaluate future OSSE results against a sufficiently large set of observations to realize statistically valid assessments. Because individual observing systems may have a larger impact on the accuracy of some analyzed and forecast fields and a smaller impact on others, it is important to evaluate the impacts on multiple ocean fields. It is also good practice to evaluate impacts using multiple error metrics to be sure that assessments are not sensitive to the choice of metrics. Finally, the system has been declared valid with respect to the chosen model configurations and the choice of T-SIS as the DA methodology. Sensitivity to DA methodology is an important issue that will be addressed in future research.

6. Summary

A new prototype fraternal twin ocean OSSE system designed with two substantially different configurations of the Hybrid Coordinate Ocean Model (HYCOM) as the NR and forecast models has been developed for initial application to observing system evaluation in the open Gulf of Mexico. The DAS consists of the forecast model and the new T-SIS statistical interpolation methodology optimized for assimilation into the hybrid vertical coordinate system of the ocean model. The novel aspect of this work is the development of an ocean OSSE system that incorporates design criteria and rigorous evaluation procedures long established for atmospheric OSSE systems that enable credible observing system impact assessments to be obtained.

It is demonstrated that 1) the chosen NR model reproduces both the climatology and variability of ocean phenomena with sufficient accuracy to represent the "true" ocean, and 2) the chosen forecast model configured with different physics and truncation errors substantially satisfies the requirement that error growth rate between the two models be very similar to error growth between individual state-of-the-art models and the true ocean. A rigorous evaluation comparing OSSEs to reference OSEs then demonstrated that the OSSE system produces valid impact assessments without requiring calibration. Impact assessments from future OSSE studies conducted with this system should be made based on multiple error metrics, and in particular on error analyses conducted for multiple ocean fields because observing systems often have larger impacts on constraining some fields and smaller impacts on others.

The ocean OSSE system evaluated herein will be initially used to assess the use of targeted airborne and in situ ocean observations to improve ocean model initialization for coupled hurricane forecasting. Issues to be evaluated for airborne surveys include determining the impact of varying the horizontal resolution of profiles, the time interval between surveys, the depth range over which profiles are taken, and instrument type (e.g., AXBTs sampling temperature only) compared to AXCTDs sampling both temperature and salinity. The OSSE system can be extended to other ocean domains and will be eventually used to evaluate observing system impacts (both existing and new systems) for a broad range of oceanographic applications including observing strategies for basin-scale to global ocean climate variability.

APPENDIX A

List of Acronyms

AVISO	Archiving, Validation, and Interpretation of Satellite Oceanographic data (satellite data distribution center)
CNES	Centre National d'Etudes Spatiales (France)
DA	Data assimilation
DAS	DA system consisting of the forecast model plus DA methodology
GDEM3	U.S. Navy Generalized Digital Environmental Model version 3
HYCOM	Hybrid Coordinate Ocean Model
MCSST	U.S. Navy multichannel sea surface temperature product
NR	Nature run
NR model	Atmospheric or oceanic model used to perform the NR
OMOC	Joint AOML–RSMAS–CIMAS Ocean Modeling and OSSE Center
OSE	Observing System Experiment
OSSE	Observing system simulation experiment
PPM	Piecewise polynomial mapping technique to vertically remap ocean profiles
SSH	Sea surface height
SST	Sea surface temperature
TCHP	Tropical cyclone heat potential
WENO	WENO technique to vertically remap ocean profiles
WP-3D	Model of Lockheed Orion aircraft used for hurricane research

APPENDIX B

T-SIS Ocean DAS

The T-SIS DA scheme is based on multivariate linear statistical estimation wherein the best linear unbiased estimate of the state of the ocean (the analysis $\mathbf{x}^a$) is obtained by updating the previous model forecast $\mathbf{x}^f$ using

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}^f), \qquad (B1)$$

where $\mathbf{y}$ represents the observations to be assimilated, $\mathbf{H}$ is the observation operator, $\mathbf{K}$ is a matrix of optimization parameters often called the gain matrix. The Gauss–Markov formula prescribes a gain matrix that is optimal in a least squares sense (e.g., Bennet 1992; Wunsch 1996):

$$\mathbf{K} = \mathbf{P}^f \mathbf{H}^{\mathrm{T}} (\mathbf{H}\mathbf{P}^f \mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}, \qquad (B2)$$

where $\mathbf{P}^f$ is the forecast error covariance matrix, $\mathbf{R}$ is the observation error covariance matrix, and superscript T denotes matrix transpose. Formally, $\mathbf{P}^f = E(\mathbf{e}^f \mathbf{e}^f)$ is the covariance matrix of the forecast error given $\mathbf{e}^f = \mathbf{x}^f - \mathbf{x}^{\mathrm{true}}$, where E represents an ensemble average and $\mathbf{x}^{\mathrm{true}}$ is the true state of the ocean. The forecast is assumed to be statistically unbiased with $E(\mathbf{e}^f) = 0$.

Because of the lack of complete and accurate information on the true oceanic state, $\mathbf{P}^f$ is a difficult quantity to determine. Numerous approximations of $\mathbf{P}^f$ have been used to represent the multivariate and spatial correlations as accurately as possible in a numerically efficient fashion. T-SIS includes five parameterizations of $\mathbf{P}^f$ discussed in Srinivasan et al. (2011) and Thacker et al. (2012). For the present OSSE application, the error covariance is prescribed using an ensemble of model states sampled at different times:

$$\mathbf{P}^f = (M-1)^{-1} \sum_{m-1}^{M} (\mathbf{x}_m^f - \overline{\mathbf{x}^f})(\mathbf{x}_m^f - \overline{\mathbf{x}^f})^{\mathrm{T}}, \quad (B3)$$

where $\mathbf{x}_m^f$ is the mth sample of the forecast ensemble, $\overline{\mathbf{x}^f}$ is the ensemble mean, and M is the number of samples. The underlying assumption is that the time variability can be related to error covariance. However, the magnitude of the true error covariance is likely to be smaller than the time variability. Therefore, the prescribed error covariance typically requires scaling to realistic levels. Measurements are assumed to be uncorrelated and a diagonal observation error covariance matrix is used.

In general, the model state vector used in estimation procedures contains all of the prognostic variables. However, the state vector used for T-SIS is a subset of the HYCOM prognostic variables, specifically layer thickness, layer temperature, layer salinity, layer density, and the diagnosed SSH anomaly. In addition, the state vector is further subdivided into three subvectors, one consisting of SSH anomaly, another of layer thickness, and another consisting of layer temperature, salinity, and density. Each subvector is assumed to be uncorrelated with the others, making $\mathbf{P}^f$ block diagonal.

In the present study, this matrix is estimated from a long unconstrained run of the forecast model (experiment LORES in Table 2). Ensemble states are stratified by month so that annual cycles in model variables do not affect the statistics. For example, the January matrix is calculated using six model archives during each of the years 2005–10 separated by five days (days 3, 8, 13, 18, 23, and 28).

The SSH anomaly field is not directly assimilated because in HYCOM, it is diagnosed from the prognostic bottom pressure and internal density fields. Instead, a layerized version of the Cooper and Haines (1996) procedure is used to adjust model layer thicknesses in the isopycnic-coordinate interior in response to SSH anomaly innovations. Prior to calculating SSH innovations, the mean dynamic topography (MDT) is added back into the altimetry observations. Because HYCOM has arbitrary mean SSH within the Gulf of Mexico domain, the difference in domain mean SSH between model and observations is calculated and the value is added to all observed altimetry data. Altimetry is not assimilated where water depth is less than 300 m. These adjustments constitute the first step of the update cycle prior to assimilating other fields in the state vector.

To optimize system performance for the HYCOM Lagrangian vertical coordinate system (essentially a stack of shallow water layers), subsurface profile observations are first layerized (remapped onto the model hybrid isopycnic σ–z vertical coordinate system) prior to assimilation. The analysis procedure then updates each layer separately in a vertically decoupled manner. For temperature profiles that do not have corresponding salinity profiles, synthetic salinity profiles are generated from climatological temperature minus salinity ($T - S$) relationships (Thacker et al. 2004) to permit layerization. The situation with velocity components is more complicated given that they are decomposed into barotropic and baroclinic components in the model. The barotropic velocity components are not included in the estimation procedure because the observation types, sampling frequencies, and assimilation time window are not appropriate to constrain barotropic velocity. In this initial version of T-SIS, baroclinic velocity components are also excluded from the estimation procedure since the cross correlations required to update them are typically not robust enough. Instead, a geostrophic velocity update increment is calculated from layer pressure increments as a postprocessing step after all other fields have been updated and is then used to adjust both the barotropic and baroclinic velocity components.

The above-mentioned modifications do make the estimation procedure less than fully multivariate, but it remains effective in the absence of robust cross correlations

between the subvectors (see section 5c; Figs. 7 and 8). The analysis is performed in the model grid space using a simple, nonadaptive, distance-based localization using observations within the localization window around a particular grid point. A quasi-Gaussian, isotropic, distance-dependent localization function (Gaspari and Cohn 1999) is used to impose a smooth localization of the error covariance and the innovations to yield a spatially continuous analysis. The localization radius, beyond which the ensemble-based covariance between two points is artificially reduced to zero, is uniform in space and set to 300 km. This corresponds to an e-folding radius of about 90 km. If instances of negative layer thickness occur after performing an analysis cycle, then they are corrected as a postprocessing step. The next cycle is then restarted from the analysis in a straightforward manner without using incremental updating or nudging. A daily update cycle is used for this study.

REFERENCES

Arnold, C. P., Jr., and C. H. Dey, 1986: Observing-systems 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.

——, and S. C. Bloom, 1989: Global surface wind vectors resulting from the assimilation of satellite wind speed data in atmospheric general circulation models. *OCEANS '89 Proceedings: An International Conference Addressing Methods for Understanding the Global Ocean,* Vol. 1, IEEE, 260–265.

——, and G. D. Emmitt, 1991: Impact of several orbit inclinations for the impact of laws on global climate studies. Preprints, *Second Symp. on Global Change Studies,* New Orleans, LA, Amer. Meteor. Soc., 28–32.

——, and ——, 2008: Review of observing system simulation experiments to evaluate the potential impact of lidar winds on numerical weather prediction. *Reviewed and Revised Papers Presented at the 24th International Laser Radar Conference,* International Laser Radar Conference, 726–729.

——, E. Kalnay, W. E. Baker, J. Susskind, D. Reuter, and M. Halem, 1985a: Simulation studies of the impact of future observing systems on weather prediction. Preprints, *Seventh Conf. on Numerical Weather Prediction,* Montreal, QC, Canada, Amer. Meteor. Soc., 145–151.

——, ——, and M. Halem, 1985b: Impact of satellite temperature soundings and wind data on numerical weather prediction. *Opt. Eng.,* **24,** 341–346.

——, S. C. Bloom, R. N. Hoffman, E. Brin, J. Ardizzone, J. Terry, D. Bungato, and J. C. Jusem, 1999: Geophysical validation of NSCAT winds using atmospheric data and analyses. *J. Geophys. Res.,* **104** (C5), 11 405–11 424.

——, and Coauthors, 2001: The effects of marine winds from scatterometer data on weather analysis and forecasting. *Bull. Amer. Meteor. Soc.,* **82,** 1965–1990.

Ballabrera-Poy, J., E. Hackett, R. Murtugudde, and A. J. Busalacchi, 2007: An observing system simulation experiment for an

optimal moored instrument array in the tropical Indian Ocean. *J. Climate,* **20,** 3284–3299.

Bennet, A. F., 1992: *Inverse Methods in Physical Oceanography.* Cambridge University Press, 346 pp.

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

Carnes, M., 2009: Description and evaluation of GDEM-V3.0. Naval Research Laboratory Tech. Rep. NRL/MR/7330-09-9165, 21 pp.

Chassignet, E. P., L. T. Smith, G. R. Halliwell, and R. Bleck, 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.

——, H. E. Hurlburt, O. M. Smedstad, G. R. Halliwell, P. J. Hogan, A. J. Wallcraft, and R. Bleck, 2007: Ocean prediction with the Hybrid Coordinate Ocean Model (HYCOM). *J. Mar. Syst.,* **65,** 60–83.

Cooper, M., and K. Haines, 1996: Altimetric assimilation with water property conservation. *J. Geophys. Res.,* **101** (C1), 1059–1077.

Daley, R., 2001: *Atmospheric Data Analysis.* Cambridge University Press, 457 pp.

Garraffo, Z. D., A. J. Mariano, A. Griffa, C. Veneziani, and E. P. Chassignet, 2001: Lagrangian data in a high-resolution numerical simulation of the North Atlantic—I. Comparison with in situ drifter data. *J. Mar. Syst.,* **29,** 157–176.

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

Guinehut, S., G. Larnicol, and P. Y. Le Traon, 2002: Design of an array of profiling floats in the North Atlantic from model simulations. *J. Mar. Syst.,* **35,** 1–9.

——, P. Y. Le Traon, G. Larnicol, and S. Phillips, 2004: Combining Argo and remote-sensing data to estimate the ocean three-dimensional temperature fields—A first approach based on simulated observations. *J. Mar. Syst.,* **46,** 85–98.

Halliwell, G. R., Jr., 2004: Evaluation of vertical coordinate and vertical mixing algorithms in the Hybrid-Coordinate Ocean Model (HYCOM). *Ocean Modell.,* **7,** 285–322.

——, A. Barth, R. H. Weisberg, P. Hogan, O. M. Smedstad, and J. Cummings, 2009: Impact of GODAE products on nested HYCOM simulations of the west Florida shelf. *Ocean Dyn.,* **59,** 139–155.

Hoffman, R. N., C. Grassotti, R. G. Isaacs, J.-F. Louis, and T. Nehrkorn, 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.

Kamenkovich, I., W. Cheng, E. S. Sarachik, and D. E. Harrison, 2009: Simulation of the Argo observing system in an ocean general circulation model. *J. Geophys. Res.,* **114,** C09021, doi:10.1029/2008JC005184.

Kourafalou, V. H., G. Peng, H. Kang, P. J. Hogan, O.-M. Smedstad, and R. H. Weisberg, 2009: Evaluation of Global Ocean Data Assimilation Experiment products on south Florida nested simulations with the Hybrid Coordinate Ocean Model. *Ocean Dyn.,* **59,** 47–66, doi:10.1007/s10236-008-0160-7.

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

Le Hénaff, M., P. De Mey, B. Mourre, and P.-Y. Le Traon, 2008: Contribution of a wide-swath altimeter in a shelf seas

assimilation system: Impact of the satellite roll errors. *J. Atmos. Oceanic Technol.,* **25,** 2133–2144.

——, V. H. Kourafalou, Y. Morel, and A. Srinivasan, 2012: Simulating the dynamics and intensification of cyclonic Loop Current frontal eddies in the Gulf of Mexico. *J. Geophys. Res.,* **117,** C02034, doi:10.1029/2011JC007279.

Leipper, D. F., and D. Volgenau, 1972: Hurricane heat potential of the Gulf of Mexico. *J. Phys. Oceanogr.,* **2,** 218–224.

Liu, Y., and R. H. Weisberg, 2011: Evaluation of trajectory modeling in different dynamic regions using normalized cumulative Lagrangian separation. *J. Geophys. Res.,* **116,** C09013, doi:10.1029/2010JC006837.

——, ——, C. Hu, and L. Zheng, 2011: Tracking the Deepwater Horizon oil spill: A modeling perspective. *Eos, Trans. Amer. Geophys. Union,* **92,** 45–46, doi:10.1029/2011EO060001.

Lumpkin, R., and S. Elipot, 2010: Surface drifter pair spreading in the North Atlantic. *J. Geophys. Res.,* **115,** C12017, doi:10.1029/2010JC006338.

Mainelli, M., M. DeMaria, L. K. Shay, and G. Goni, 2008: Application of oceanic heat content estimation to operational forecasting of recent category 5 hurricanes. *Wea. Forecasting,* **23,** 3–16.

Mariano, A. J., V. H. Kourafalou, A. Srinivasan, H. Kang, G. R. Halliwell, E. H. Ryan, and M. Roffer, 2011: On the modeling of the 2010 Gulf of Mexico oil spill. *Dyn. Atmos. Oceans,* **52,** 322–340.

Mellor, G. L., and T. Yamada, 1982: Development of a turbulence closure model for geophysical fluid problems. *Rev. Geophys.,* **20,** 851–875.

Morss, R. E., and D. S. Battisti, 2004a: Designing efficient observing networks for ENSO prediction. *J. Climate,* **17,** 3074–3089.

——, and ——, 2004b: Evaluating observing requirements for ENSO prediction: Experiments with an intermediate coupled model. *J. Climate,* **17,** 3057–3073.

Mourre, B., P. De Mey, Y. Menard, F. Lyard, and C. Le Provost, 2006: Relative performance of future altimeter systems and tide gauges in constraining a model of North Sea high-frequency barotropic dynamics. *Ocean Dyn.,* **56,** 473–486.

Murphy, A. H., 1988: Skill scores based on the mean square error and their relationships to the correlation coefficient. *Mon. Wea. Rev.,* **116,** 2417–2424.

Oke, P. R., and A. Schiller, 2007a: A model-based assessment and design of a tropical Indian Ocean mooring array. *J. Climate,* **20,** 3269–3283.

——, and ——, 2007b: Impact of Argo, SST, and altimeter data on an eddy-resolving ocean analysis. *Geophys. Res. Lett.,* **34,** L19601, doi:10.1029/2007GL031549.

Prasad, T. G., and P. J. Hogan, 2007: Upper-ocean response to Hurricane Ivan in a 1/25° nested Gulf of Mexico HYCOM. *J. Geophys. Res.,* **112,** C04013, doi:10.129/2006JC003695.

Rohaly, G. D., and T. N. Krishnamurti, 1993: An observing system simulation experiment for the Laser Atmospheric Wind Sounder (LAWS). *J. Appl. Meteor.,* **32,** 1453–1471.

Sakov, R., and P. R. Oke, 2008: Objective array design: Application to the tropical Indian Ocean. *J. Atmos. Oceanic Technol.,* **25,** 794–807.

Schiller, R. V., V. H. Kourafalou, P. J. Hogan, and N. D. Walker, 2011: The dynamics of the Mississippi River plume: Impact of topography, wind and offshore forcing on the fate of plume waters. *J. Geophys. Res.,* **116,** C06029, doi:10.1029/2010JC006883.

Shay, L. K., and Coauthors, 2011: Airborne ocean surveys of the Loop Current complex from NOAA WP-3D in support of

the Deep Water Horizon oil spill. *Monitoring and Modeling the* Deepwater Horizon *Oil Spill: A Record-Breaking Enterprise, Geophys. Monogr.,* Vol. 195, Amer. Geophys. Union, 131–151.

Srinivasan, A., and Coauthors, 2011: Comparison of sequential data assimilation schemes for ocean prediction with HYCOM: Twin experiments with static forecast error covariances. *Ocean Modell.,* **37,** 85–111.

Taillandier, V., A. Griffa, P.-M. Poulain, and K. Béranger, 2006: Assimilation of Argo float positions in the north western Mediterranean Sea and impact on ocean circulation simulations. *Geophys. Res. Lett.,* **33,** L11604, doi:10.1029/2005GL025552.

Taylor, K. E., 2001: Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.,* **106** (D7), 7183–7192.

Thacker, W. C., S.-K. Lee, and G. R. Halliwell Jr., 2004: Assimilating 20 years of Atlantic XBT data into HYCOM: A first look. *Ocean Modell.,* **7,** 183–210.

——, A. Srinivasan, M. Iskandarani, O. M. Knio, and M. Le Henaff, 2012: Propagating boundary uncertainties using polynomial expansions. *Ocean Modell.,* **43–44,** 52–63.

Tranchant, B., C.-E. Testut, L. Renault, N. Ferry, F. Birol, and P. Brasseur, 2008: Expected impact of the future SMOS and Aquarius ocean surface salinity missions in the Mercator Ocean operational systems: New perspectives to monitor ocean circulation. *Remote Sens. Environ.,* **112,** 1476–1487.

Vecchi, G. A., and M. J. Harrison, 2007: An observing system simulation experiment for the Indian Ocean. *J. Climate,* **20,** 3300–3343.

Wunsch, C., 1996: *The Ocean Circulation Inverse Problem.* Cambridge University Press, 442 pp.