Progress in Oceanography 130 (2015) 32–46

Contents lists available at ScienceDirect

Progress in Oceanography

journal homepage: www.elsevier.com/locate/pocean

OSSE impact analysis of airborne ocean surveys for improving upper-ocean dynamical and thermodynamical forecasts in the Gulf of Mexico

George R. Halliwell Jr [a,*], Vassiliki Kourafalou [b], Matthieu Le Hénaff [c], Lynn K. Shay [b], Robert Atlas [d]

[a] NOAA/AOML/PhOD, 4301 Rickenbacker Causeway, Miami, FL 33149, USA
[b] MPO/RSMAS, University of Miami, 4600 Rickenbacker Causeway, Miami, FL 33149, USA
[c] CIMAS, University of Miami, 4600 Rickenbacker Causeway, Miami, FL 33149, USA
[d] NOAA/AOML, 4301 Rickenbacker Causeway, Miami, FL 33149, USA

ARTICLE INFO

Article history:
Received 6 January 2014
Received in revised form 9 September 2014
Accepted 14 September 2014
Available online 22 September 2014

ABSTRACT

A prototype, rigorously validated ocean Observing System Simulation Experiment (OSSE) system is used to evaluate the impact of different sampling strategies for rapid-response airborne ocean profile surveys in the eastern interior Gulf of Mexico. Impacts are assessed with respect to improving ocean analyses, and forecasts initialized from those analyses, for two applications: improving oil spill forecasts and improving the ocean model response to tropical cyclone (TC) forcing. Rapid model error growth in this region requires that repeat surveys be conducted frequently in time, with separation of less than 4 days required to approach maximum error reduction in model analyses. Substantial additional error reduction in model dynamical fields is achieved by deploying deep (1000 m) AXCTDs instead of shallow (400 m) AXBTs. Shallow AXBTs constrain the ocean thermal field over the upper 400 m nearly as well as deep AXCTDs. However, in addition to constraining ocean fields over a greater depth range, AXCTDs also measure salinity profiles and more accurately constrain upper-ocean density than AXBTs, leading to a more accurate representation of upper ocean pressure and velocity fields. Sampling AXCTD profiles over a one-half degree array compared to one degree leads to substantial additional error reduction by constraining variability with horizontal scales too small to be corrected by satellite altimetry assimilation. A 2-day lag in availability of airborne profiles does not increase errors in dynamical ocean fields, but it does increase errors in upper-ocean thermal field including Tropical Cyclone Heat Potential (TCHP), demonstrating that these profiles must be rapidly made available for assimilation to improve TC forecasts. The additional error reduction in ocean analyses achieved by assimilation of airborne surveys translates into significantly improved forecasts persisting over time intervals ranging between 1 and 2 weeks for most model variables but several weeks for TCHP. In particular, upper-ocean temperature forecasts can be significantly improved for an extended interval of time by conducting airborne profile surveys.

Published by Elsevier Ltd.

Introduction

Several rapid-response airborne surveys profiled the upper ocean over the interior eastern Gulf of Mexico (GoM) during the spring and summer of 2010. This field campaign was motivated by the need to improve initialization of ocean forecast models used by the NOAA Office of Response and Restoration to forecast oil transport and dispersion from the Deepwater Horizon (DWH) spill (e.g., Liu et al., 2011a, 2011b). A combination of Airborne eXpendable BathyThermographs (AXBTs), Conductivity-Temperature-Depth profilers (AXCTDs), and Current Profilers (AXCPs) were deployed 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 in quasi-synoptic lawnmower patterns (e.g. Fig. 1a) with sufficient horizontal resolution to resolve the path of the Loop Current (LC) plus associated cyclones and anticyclones. These oceanic features are associated with strong currents and fronts, making this a challenging region to monitor and forecast. The rapid implementation of the repeat airborne survey program during the spill was motivated by this challenge.

An Observing System Experiment (OSE) performed using the Navy regional GoM nowcast–forecast system (based on the Hybrid Coordinate Ocean Model, HYCOM) demonstrated that assimilation

* Corresponding author. Tel.: +1 305 361 4346.
E-mail address: george.halliwell@noaa.gov (G.R. Halliwell Jr).

http://dx.doi.org/10.1016/j.pocean.2014.09.004
0079-6611/Published by Elsevier Ltd.

of the airborne profiles reduced upper-ocean temperature errors by >30% and reduced biases to near zero (Shay et al., 2011). Although the mesoscale airborne surveys had a positive impact on forecasts, it is not clear whether they were conducted in an optimal manner. Individual surveys were organized on short notice due to availability of aircraft, relied on available profilers, and were conducted irregularly in time with repeat intervals ranging from 3 days to 2 weeks due to availability of aircraft. Pre-planning to optimize impacts in terms of horizontal resolution, temporal survey frequency, probe type, profile depths, and minimizing data latency for assimilation was necessarily limited.

Observing System Simulation Experiments (OSSEs) provide a means to test the impact of new ocean observing systems and of alternate deployment strategies for existing systems. These impacts are typically assessed based on the ability to monitor and forecast specific oceanographic phenomena of interest. In the present study, OSSE methodology is employed specifically to determine the impact of rapid-response airborne surveys on ocean analyses and forecasts, focusing on several questions related to the design and execution of the surveys. The overarching goal is to establish a robust methodology that can be used to optimize the observational design and quantify the improvement of initial fields provided to ocean forecast models and, moreover, the improvement of forecasts initialized by those fields. Experiments are performed herein specifically to determine the impact of (1) delayed availability of ocean profiles for assimilation; (2) instrument type (shallow AXBTs versus deep AXCTDs); (3) horizontal resolution of profile surveys; and (4) temporal interval between surveys. Impacts are assessed by calculating error metrics for several model fields produced by both analysis and forecast experiments. Given the initial motivation for conducting the rapid-response GoM surveys to improve DWH oil spill prediction, the impact of the airborne surveys on the analysis and forecast of ocean dynamical fields (sea surface height and velocity) is determined. Rapid-response airborne surveys have also been used to measure the ocean ahead of hurricanes to improve ocean model initialization in coupled hurricane forecast models (e.g. Shay and Uhlhorn, 2008; Uhlhorn and Shay, 2012, 2013; Sanabia et al., 2013). For this reason, the impact of airborne surveys on initializing and forecasting upper-ocean thermodynamical fields important for hurricane forecasting is also assessed.

A recently-developed fraternal-twin ocean OSSE system validated for this region (Halliwell et al., 2014) is employed herein to address these questions. This is a prototype OSSE system that has been developed by the joint NOAA–AOML and University of Miami/RSMAS Ocean Modeling and OSSE Center (OMOC; http://cimas.rsmas.miami.edu/omoc.html). System validation was achieved using an evaluation procedure that compared OSSEs to reference Observing System Experiments (OSEs). Each OSE–OSSE experiment pair was identical except for assimilating real (OSE) and synthetic (OSSE) observations, respectively. The key step was to demonstrate that similar impact assessments were consistently obtained between all corresponding OSE–OSSE pairs based on multiple error metrics calculated for several model variables. This design and validation exercise followed long-established procedures used in atmospheric OSSEs (e.g. Atlas et al., 1985a, 1985b; Arnold and Dey, 1986; Atlas, 1997).

Section 'The ocean OSSE system' presents a brief overview of the OSSE system. Section 'OSSE analysis procedures' describes how the different observations are analyzed and assimilated. Section 'OSSE impact assessments based on ocean analyses' describes OSSE impact assessments based on ocean analyses, while Section 'OSSE impact assessments based on ocean forecasts' describes impact assessments based on ocean forecasts. Section 'Summary' summarizes the results.

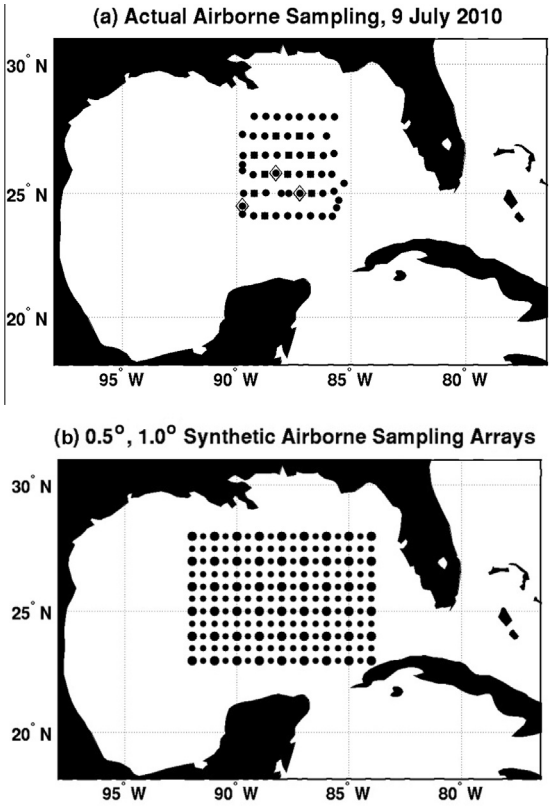

Fig. 1. (a) Airborne profile array conducted by a NOAA WP-3D hurricane research aircraft during one of the 9 flights performed during the DWH oil spill (9 July 2010; Shay et al., 2011) showing profiles collected using AXBTs (dots), AXCTDs (open diamonds showing the three released at AXBT profile locations), and AXCPs (squares). (b) Synthetic airborne profile arrays sampled for the OSSEs, where all dots delineate the 0.5° array and the large dots denote the 1° array.

The ocean OSSE system

The OSSE system configuration is described in detail in Halliwell et al. (2014) and is briefly summarized here. This "fraternal twin" system employs two different realizations of HYCOM, each configured to produce substantially different physics and truncation errors. One is used to perform the Nature Run (NR), while the other is used as the data-assimilative forecast model (FM). 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, ranging from sigma or z-coordinate in shelf regions to isopycnic in deep waters (Halliwell et al., 2009; Kourafalou et al., 2009; Schiller et al., 2011). The model also contains multiple choices of numerical algorithms and subgrid-scale parameterizations which can be varied to introduce the different physics and truncation errors required of a credible OSSE system. To add additional truncation errors, the FM is run at lower horizontal resolution than the NR. The present study uses the same system configuration that was validated in the interior Gulf of Mexico. A fixed σ–z vertical coordinate system with 26 layers on a 0.04° Mercator mesh is used for the NR while a hybrid z–σ-isopycnic vertical coordinate system with 26 layers on a 0.08° Mercator mesh is used for the FM. The lower-resolution mesh consists of every other point of the NR mesh. Details of the two model configurations are described in Halliwell et al. (2014).

The NR was forced by fields obtained from a regional mesoscale atmospheric model, specifically the data-assimilative Navy

Coupled Ocean–Atmosphere Mesoscale Prediction System (COAMPS) run with a horizontal resolution of 27 km. It was nested within a model-generated high-resolution Atlantic Ocean climate simulation performed by the Naval Research Laboratory. The FM is coupled to a data assimilation (DA) procedure that allows the required prediction/background error covariance to be flexibly specified from several common approximations and parameterizations of the full error covariance matrix. A variant of the Cooper and Haines (1996) algorithm is used to correct subsurface temperature and salinity through assimilation of altimeter Sea Surface Height (SSH) anomalies. Technical details of the DA system and its implementation are contained in Halliwell et al. (2014).

The OMOC OSSE system is optimized for assimilation into the Lagrangian vertical coordinate layers of HYCOM. All profiles must be layerized prior to assimilation by re-mapping onto the model hybrid isopycnic-sigma-z vertical coordinate system used by the FM. Layerization requires that both temperature and salinity are available for all profiles. For profiles of temperature alone, synthetic salinity profiles are generated based on climatological T–S relationships (Thacker et al., 2004). Consequently, the difference between assimilating AXCTDs and AXBTs over their common depth range is that the former involves assimilating measured temperature and salinity profiles while the latter involves assimilating measured temperature and estimated salinity profiles. The analysis procedure updates each model layer separately in a vertically decoupled manner.

OSSE analysis procedures

Ocean analysis windows

Two ocean analysis approaches are used in this study. Retrospective ocean analyses are performed using "reanalysis" mode, where all observations are assumed to be available for assimilation on the analysis day. These analyses are performed daily using a 1-day update cycle. Reanalysis mode is used herein to study the impact of changing the temporal interval between repeat airborne surveys, such as those conducted in response to the DWH oil spill. Near-real-time (NRT) analyses must follow a different approach because the availability of observations for assimilation can be delayed by up to a few days, particularly for altimetry. NRT-mode analyses therefore use an extended update cycle that starts with an initial guess field several days prior to the analysis time when nearly all delayed observations are available. Each analysis is then performed over this several-day interval using 1-day update cycles that assimilate all observations available on each day. In the present study, NRT mode is used retrospectively to study the impact of individual rapid-response airborne surveys, and also to evaluate the impact of latency in the availability of airborne profiles and other observations for assimilation.

Observations for assimilation

All ocean observing systems other than airborne profiles that were available for assimilation during the 2010 analysis interval are listed in Table 1. To perform OSSEs, these actual observations are replaced by observations simulated from the NR. For NRT-mode experiments, realistic temporal delays in the availability of these observations must be introduced. The chosen delays approximate the actual delays encountered by operational ocean analysis-forecast centers and are listed in Table 1. Synthetic data sets include along-track measurements of SSH anomaly (SSHA) from three altimeters: Jason-1, Jason-2, and Envisat, along with the mean dynamic topography field that must be added to SSHA. The 2005–2010 mean SSH field from the NR is used for mean dynamic topography. Synthetic SST data is sampled to represent the satellite-derived Multi-Channel Sea Surface Temperature (MCSST) product while *in-situ* SST is sampled to represent data collected by ships, surface meteorological buoys, and surface drifters. Subsurface measurements include XBT profiles collected from ships. Velocity observations are not currently assimilated by the OMOC OSSE system.

Airborne survey maps for all flight days are presented in Shay et al. (2011), and one example (9 July 2010) is presented in Fig. 1a. Most probes that were deployed on each flight day were AXBTs, with some sampling to nearly 400 m depth and others sampling to nearly 800 m depth. On most flight days, the AXBTs were supplemented by a smaller number of AXCPs that sampled temperature and velocity profiles to depths up to 1800 m and AXCTDs that sampled temperature and salinity to depths up to 1000 m (Shay and Uhlhorn, 2008). To perform OSSEs, temperature and salinity profiles to 1000 m are simulated at the grid points in Fig. 1b for AXCTDs while temperature alone to 400 m is simulated for AXBTs. Corresponding salinity profiles for AXBTs are generated based on climatological relationships (Thacker et al., 2004). When synthetic profiles are sampled from the layerized NR model, vertical interpolation is performed to generate values at the same z levels sampled by the actual profiles. All temperature-salinity profiles are then layerized prior to assimilation by re-mapping onto the vertical layer structure used by the FM.

Realistic errors are added to the synthetic observations from each observing system as described in Halliwell et al. (2014). Instrument and other local random errors are added to each individual observation using a random number generator that assumes a Gaussian probability density distribution. Representation errors resulting from ocean features with space and time scales resolved by actual observations, but not by the NR, are also added. Spatially correlated errors present in along-track satellite altimetry measurements are added as described in Halliwell et al. (2014).

Rationale for OSSE impact evaluation

The impact of ocean observing systems must be evaluated in the context of specific applications. For the present study, observations are evaluated in terms of improving ocean analyses and forecasts for two applications: (1) predicting the transport and dispersion of oil spills; and (2) predicting the upper-ocean thermal response to tropical cyclones (TCs). Both applications demand accurate initialization and prediction of mesoscale ocean features and their associated currents while the TC application also demands highly-accurate initialization and prediction of upper-ocean thermal and density structure within these ocean features. It is important to initialize density (stratification) profiles accurately for the TC application to enable ocean model vertical mixing

Table 1
Time lags for which observations other than airborne profiles were assimilated using the NRT analysis procedure. 'X' indicates that observations are available at the given lag time while blanks indicate that they are not available.

Lag Day	Lag Hour	Altim. Jason-1	Altim. Envisat	Altim. Jason-2	MCSST	Surface drifter SST	Met. Buoy SST	Ship SST	Ship XBT profiles
0	0–24				X	X	X		
−1	24–48			X	X	X	X	X	X
−2	48–72		X	X	X	X	X	X	X
−3	72–96	X	X	X	X	X	X	X	X

parameterizations to work properly. Oceanic structure in the eastern interior GoM is dominated by the Loop Current (LC) and associated cyclones and anticyclones. These features affect TC forecasts because they are associated with large horizontal changes in the thermal energy available to maintain storm intensity, and because storm-forced SST cooling patterns are distorted by the near-surface background velocity field and its associated vorticity (Jaimes et al., 2011).

The LC and associated anticyclones are associated with a thick surface warm layer and large available thermal energy for maintaining TC intensity. By contrast, other regions in the interior Gulf are dominated by Gulf Common Water associated with a thin surface warm layer and less available thermal energy. The smallest available thermal energy is typically encountered in cold-core cyclones associated with the LC. The Tropical Cyclone Heat Potential (TCHP), also known as Ocean Heat Content (OHC), is an index of thermal energy availability relative to 26 °C (e.g. Leipper and Volgenau, 1972; Shay et al., 2000; Mainelli et al., 2008; Goni et al., 2009; Lin et al., 2013). Other indices have recently been developed, including heat content plus stratification effects (Shay and Brewster, 2010; Meyers et al., 2014), vertically averaged temperature (Price, 2009), and potential energy based (Vincent et al., 2012). TCHP is used herein because it is the parameter presently used in operational statistical–dynamical TC forecast models (DeMaria et al., 2005; Mainelli et al., 2008).

The error analyses used to determine observing system impact involve the calculation of multiple error metrics for several different model fields. Improved oil spill forecasts require that dynamically balanced ocean currents be initialized as accurately as possible. Error analyses are therefore performed for dynamical variables sea surface height (SSH), sea surface zonal velocity (SSU), and sea surface meridional velocity (SSV). Because the TC application requires accurate initialization of the upper-ocean thermodynamical structure, error analyses are also performed for fields of TCHP, SST, and sea surface salinity (SSS).

Oceanic variability during the analysis interval

To provide context for interpreting the OSSE impact assessments, four SSH maps from the NR during the study interval are presented in Fig. 2. The NR has been evaluated (Halliwell et al., 2014, who extended results from Le Hénaff et al., 2012) for the mesoscale processes of interest, and found acceptable as a satisfactory representation of the "true" ocean in the interior eastern GoM. The eastern and central parts of the domain sampled by the synthetic airborne profiles are dominated by LC variability as expected. During May, the LC path follows a short northward extension to about 26.5°N. Over the next 3 months, the northward extension gradually increases to about 27.5°N. By 1 October, a new anticyclonic eddy has almost completely detached from the LC. Later in the month, however, the ring begins to reattach (not shown). The NR reproduces realistic cold-core eddy variability along the LC front (e.g. Le Hénaff et al., 2012; Halliwell et al., 2014). For example, during the near-detachment, a small cold-core eddy is centered near 85°W, 25°N just to the east of the narrow neck of slightly higher SSH that tenuously connects the nascent eddy to the LC. However, the cold eddy is unable to extend westward to complete the detachment after 1 October. Oceanic variability in the western one-third of the analysis domain is much weaker than in the immediate LC region.

OSSE impact assessments based on ocean analyses

Experimental design and evaluation procedure

OSSE experiments and their configurations are listed in Table 2. Experiment FREE is an unconstrained run performed by the FM, which develops large errors with respect to the NR (also unconstrained) as expected and thus serves as a reference to determine total error reduction resulting from data assimilation. Experiment NOP3A is conducted in reanalysis mode and assimilates with no data delay all synthetic observations with the exception of airborne profiles. Experiment NOP3B is the same as NOP3A except for being run in NRT mode, assimilating observations with the delays shown in Table 1. Both NOP3A and NOP3B serve as reference experiments to determine the further error reduction achieved in the reanalysis-mode and NRT-mode experiments, respectively, by assimilating airborne profiles.

All other experiments assimilate synthetic airborne profiles. The NRT experiments are designed to determine the impact of data

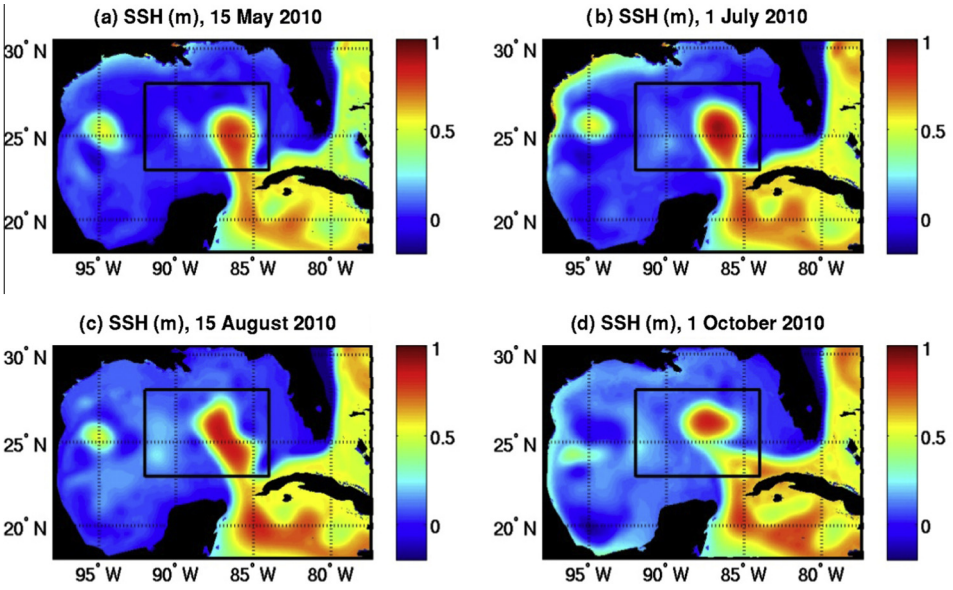

Fig. 2. SSH maps from the Nature Run at four times during the 2010 analysis interval: (a) 15 May, (b) 1 July, (c) 15 August, and (d) 1 October illustrating the Loop Current and eddy variability that occurred within the synthetic airborne sampling region outlined by the black boxes.

Table 2
List of experiments performed by the forecast model and their key characteristics.

Experiment	Analysis Method	Airborne profiles assimilated	Profile Horizontal Resolution	Timing of Airborne Missions
FREE	No assimilation	None	N/A	N/A
NOP3A	Reanalysis mode	None	N/A	N/A
NOP3B	NRT mode	None	N/A	N/A
AXBTHR	NRT mode	400 m AXBT	0.5°	Lag days 0 and −1
AXCTDHR	NRT mode	1000 m AXCTD	0.5°	Lag days 0 and −1
AXCTDLR	NRT mode	1000 m AXCTD	1.0°	Lag days 0 and −1
DELAY0	NRT mode	1000 m AXCTD	0.5°	Lag day 0 only
DELAY2	NRT mode	1000 m AXCTD	0.5°	Lag day −2 only
DT1	Reanalysis mode	1000 m AXCTD	0.5°	Every day
DT2	Reanalysis mode	1000 m AXCTD	0.5°	Every 2 days
DT4	Reanalysis mode	1000 m AXCTD	0.5°	Every 4 days
DT8	Reanalysis mode	1000 m AXCTD	0.5°	Every 8 days
DT16	Reanalysis mode	1000 m AXCTD	0.5°	Every 16 days

latency, probe type (including profile depth), and horizontal resolution, while reanalysis-mode experiments are designed to determine the impact of temporal survey frequency. To execute each NRT-mode experiment, twenty-six 4-day analysis cycles are run, with each cycle separated by a 7 day interval. The first cycle is initialized on 1 May 2010 and the final one is initialized on 29 October 2010. Each cycle is initialized by ocean fields from NRT-mode experiment NOP3A so that each resulting analysis is influenced only by the airborne profiles collected during that particular 4-day cycle. The resulting product consists of ocean analyses separated by 7 day time intervals beginning on 5 May 2010 and ending on 2 November 2010. The motivation for running multiple analysis cycles is to calculate robust statistics for the impact assessments. By contrast, reanalysis mode experiments are run sequentially from 1 May through 1 November using a daily update cycle to enable sequential airborne surveys separated by a prespecified number of days to be assimilated. For both NRT and reanalysis mode experiments, model archives are saved 6 h after each analysis time, which provides time for the barotropic adjustment of model fields after analysis. Details on each experiment are introduced in subsequent sections.

The error metrics analyzed herein are mean bias, RMS errors, and combinations of metrics represented by the Murphy (1988) skill score and Taylor (2001) diagrams. These metrics were all used in the OSSE system evaluation of Halliwell et al. (2014), where they are described in detail. For each experiment, error metrics are calculated in comparison to fields from the NR over all common model grid points located within the region where synthetic profiles were simulated (Fig. 1b). The selected domain in Fig. 1b is too large for a single aircraft to sample in a single day, but it was chosen to facilitate the calculation of robust statistics. Results presented herein remain valid for single aircraft surveys, but expected error reductions would obviously be confined to a smaller area.

Impact of altimetry assimilation

Before evaluating the impact of assimilating airborne survey profiles, experiments NOP3A and NOP3B are analyzed to establish baseline reference points for subsequent reanalysis-mode and NRT-mode experiments, respectively. Due to poor coverage by real-time *in situ* observing systems, altimetry is primarily responsible for constraining three-dimensional ocean fields in the interior GoM in operational ocean analysis products. Error reduction due primarily to altimetry assimilation is evaluated by calculating RMS errors and Murphy skill scores (Table 3) for the three dynamical and three thermodynamical model fields. Comparing reanalysis-mode experiment NOP3A to the unconstrained experiment FREE over all common model grid points within the region sampled by airborne surveys (Fig. 1a), the percentage reduction of RMS error due primarily to altimetry assimilation was 38% for TCHP, 37% for SST, 11% for SSS, 58% for SSH, 39% for SSU, and 44% for SSV. For experiment FREE, the Murphy skill score is significant for SST (0.48), barely significant for TCHP (0.07), and insignificant for all other fields. (A score of 1.0 represents a perfect comparison while values <0 are insignificant.) The skill scores for experiment NOP3A are much larger for all fields except SSS, which retains an insignificant score of −0.25. Altimetry assimilation has the largest impact on upper ocean dynamical variables as expected, with skill scores increasing from −0.51 to 0.80 for SSH, from −0.27 to 0.52 for SSU, and from −0.27 to 0.60 for SSV. It has a smaller positive impact on TCHP and SST, which is presumably achieved by constraining the three-dimensional structure of dynamical ocean features associated with horizontal temperature variability. It has no discernible impact on SSS.

The impact of data latency for altimetry is determined by comparing RMS errors and skill scores between reanalysis-mode experiment NOP3A and NRT-mode experiment NOP3B (Table 3). RMS errors are either the same or insignificantly smaller for NOP3B due to the data delays. Consequently, typical delays in altimetry data availability have minimal impact on constraining ocean fields in the GoM interior.

Impact of data latency of airborne profiles

The impact of delays in airborne profile availability is assessed by evaluating and comparing NRT-mode experiment NOP3B to NRT-mode experiments DELAY0, DELAY2, and AXCTDHR (Table 2), all of which assimilated AXCTD profiles on the 0.5° survey grid. Experiment DELAY0 assimilates profiles only on the final day of each 4-day analysis cycle while DELAY2 represents a 2-day delay by assimilating profiles only on the second day of each cycle. Experiment AXCTDHR assimilates profiles on both the third and fourth days of each cycle to represent the impact of two back-to-back daily airborne surveys. RMS errors with respect to the NR from these four experiments, calculated over all common model grid points within the region sampled by synthetic profiles (Fig. 1) for all analysis times from June through October 2010, are plotted sequentially as time series (Fig. 3). Mean RMS errors and Murphy skill scores for these experiments calculated over all weekly analysis times between 1 June and 30 September are listed in Table 4. The tabulated statistics are restricted to a shorter time interval than the time series plots to represent summer conditions and avoid the impact of autumn cooling in the upper ocean.

As expected, experiment AXCTDHR produces the largest additional reduction in RMS error compared to NOP3B for the three thermodynamical model fields (47% for TCHP and 35% for both SST and SSS), and also for the three dynamical model fields (46% for SSH, 32% for SSU, and 35% for SSV). Skill scores for this

Table 3
RMS errors and Murphy skill scores among three forecast model experiments that illustrate the impact of altimetry assimilation and the NR for six model fields (variables are defined in the text). The errors and scores were calculated over all common model grid points within the domain sampled by airborne surveys from 1 June through 30 September 2010 using daily analyses for experiment FREE and reanalysis-mode experiment NOP3A, and all weekly analyses within this time interval for NRT-mode experiment NOP3B. Percentage reduction of RMS error in comparison to experiment FREE for all variables is shown in parentheses for experiments NOP3A and NOP3B.

Metric	Experiment	TCHP	SST	SSS	SSH	SSU	SSV
RMS Error	FREE	17.6	0.70	0.38	0.31	0.49	0.59
	NOP3A	10.9 (−38%)	0.44 (−37%)	0.34 (−11%)	0.13 (−58%)	0.30 (−39%)	0.33 (−44%)
	NOP3B	11.2 (−36%)	0.46 (−34%)	0.34 (−11%)	0.13 (−58%)	0.31 (−37%)	0.34 (−42%)
Murphy Skill Score	FREE	0.07	0.48	−0.32	−0.51	−0.27	−0.20
	NOP3A	0.50	0.75	−0.25	0.80	0.52	0.62
	NOP3B	0.49	0.73	−0.24	0.78	0.49	0.61

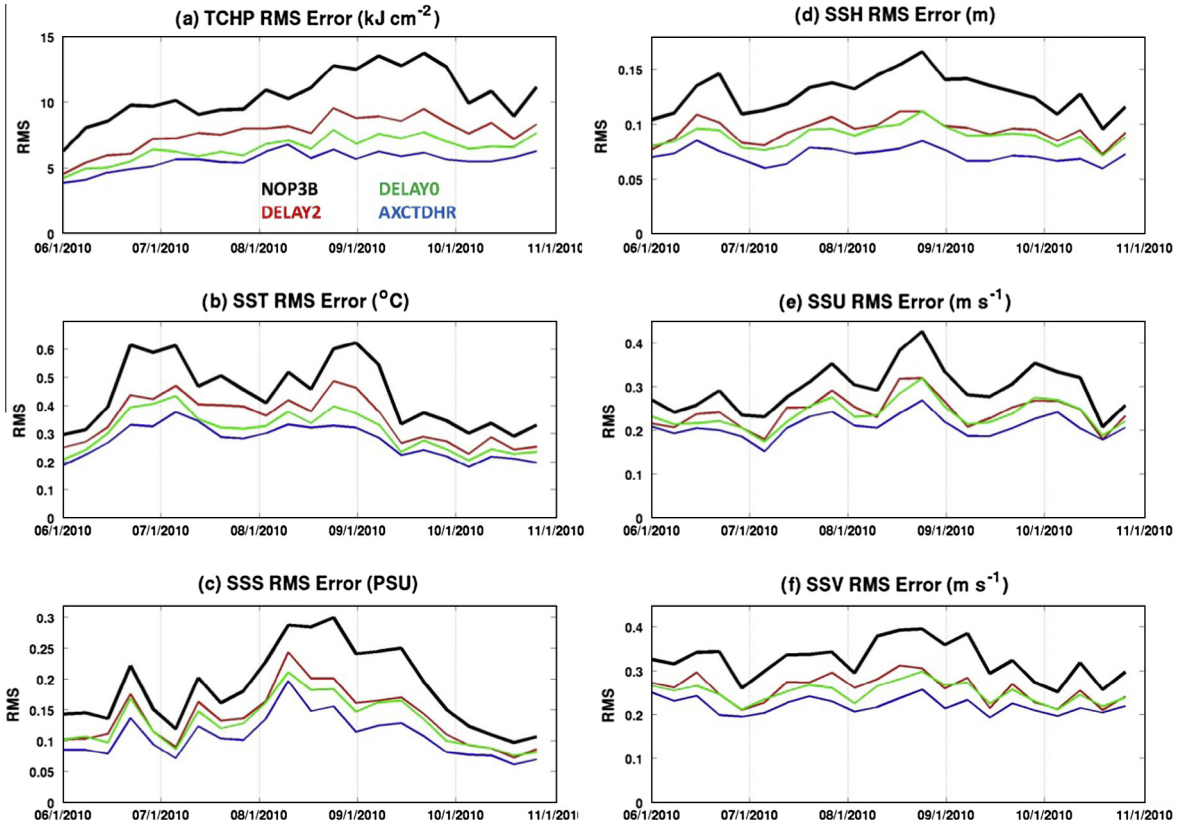

Fig. 3. Time series of RMS error between the Nature Run and the 7-day analysis products from OSSE experiments NOP3, DELAY0, DELAY2, and AXCTDHR for (a) TCHP, (b) SST, (c) SSS, (d) SSH, (e) SSU, and (f) SSV. Line colors are given by the legend in panel (a).

Table 4
RMS errors and Murphy skill scores among four forecast model experiments that illustrate the impact of data latency and the NR for six model fields. The errors and scores were calculated over all common model grid points within the domain sampled by airborne surveys comparing the weekly analyses from these NRT-mode experiments within the time interval 1 June through 30 September 2010. Percentage reduction of RMS error from experiment NOP3B for all variables is shown in parentheses for experiments AXCTDHR, DELAY0, and DELAY2.

	Experiment	TCHP	SST	SSS	SSH	SSU	SSV
RMS Error	NOP3B	11.2	0.46	0.34	0.13	0.31	0.34
	AXCTDHR	5.99 (−47%)	0.30 (−35%)	0.22 (−35%)	0.07 (−46%)	0.21 (−32%)	0.22 (−35%)
	DELAY0	6.99 (−38%)	0.34 (−26%)	0.23 (−32%)	0.09 (−31%)	0.24 (−23%)	0.26 (−24%)
	DELAY2	8.50 (−24%)	0.36 (−22%)	0.24 (−29%)	0.10 (−23%)	0.25 (−19%)	0.27 (−21%)
Murphy Skill Score	NOP3B	0.49	0.73	−0.24	0.78	0.49	0.61
	AXCTDHR	0.89	0.91	0.68	0.92	0.76	0.82
	DELAY0	0.85	0.88	0.46	0.88	0.69	0.77
	DELAY2	0.78	0.86	0.42	0.87	0.66	0.75

experiment are all highly significant (e.g. 0.89 for TCHP and 0.92 for SSH). Although the lowest skill score from AXCTDHR is 0.68 for SSS, the largest improvement over NOP3B (from −0.24) occurred for this variable, demonstrating the high value of *in-situ* salinity profiles. Ocean analysis products typically have large salinity errors due to limited observations, so AXCTD profiles are particularly important for accurately constraining this field. This can be particularly important for TC forecasting because it improves the

accuracy of vertical density profiles in initial ocean fields that in turn will improve the performance of ocean model vertical mixing parameterizations. The single-survey experiment DELAY0 produces somewhat less RMS error reduction in all six model variables, which demonstrates a positive impact from conducting surveys over two sequential days. Experiment DELAY2 produces only slightly less error reduction and slightly lower skill scores in variables SSS, SSH, SSU, and SSV. However, the impact of data delay is larger for TCHP and SST, demonstrating that delays should be minimized for the purpose of initializing the ocean component of coupled TC forecast models.

The impact of airborne profile assimilation on bias reduction for the three thermodynamical fields is substantial for all three experiments (Fig. 4). Experiment AXCTDHR provides the largest bias reduction, with values near zero for TCHP and SST but remaining positive for SSS (average of ≈0.06 PSU). Airborne profile assimilation is therefore highly effective at reducing biases in model thermodynamical fields. Two sequential daily surveys are most effective for bias reduction. Somewhat smaller bias reduction is achieved in experiments DELAY0 and DELAY2, again demonstrating the advantage of conducting back-to-back daily surveys.

Impacts of airborne instrument type and horizontal resolution

The mix of probe types that were used and profile depths that were measured during the DWH oil spill (Fig. 1a), during which only a limited number of deep AXCTDs and AXCPs were deployed compared to AXBTs (Fig. 1a), raises the question of the impact of probe type on ocean analysis errors. This impact is assessed by comparing the error reduction and skill score increases achieved by NRT experiment AXBTHR, which assimilates 400 m AXBT profiles (Table 2) to both AXCTDHR and NOP3B. Inspection of RMS error time series (Fig. 5), mean RMS errors and skill scores (Table 5) demonstrates that shallow AXBTs constrain TCHP and SST as effectively as deeper AXCTDs. This result was expected (but now is quantified) because both probe types sample temperature over the upper 400 m. However, shallow AXBTs produce substantially less error reduction for SSS compared to deep AXCTDs (21% versus 32%). The use of climatological *T*–*S* relationships to generate AXBT salinity profiles (Thacker et al., 2004) does reduce SSS error, but not as much as the assimilation of actual salinity profiles measured by AXCTDs. The reduced profile depths and the less-accurate derived salinity profiles assimilated by AXBTHR both act to limit error reduction in SSH analyses (31% versus 46%). The additional percentage RMS error reduction in AXCTDHR with respect to AXBTHR is smaller in the noisier SSU and SSV fields. The use of deep AXCTD fields versus shallow AXBTs also has little impact on TCHP and SST bias reduction, but has substantial impact on SSS bias reduction (Fig. 6). Overall, exclusive use of deep AXCTDs provides little additional constraint on upper-ocean temperature and heat content over shallow AXBTs, but it does provide a modest additional constraint on upper-ocean dynamical fields and a more accurate representation of salinity.

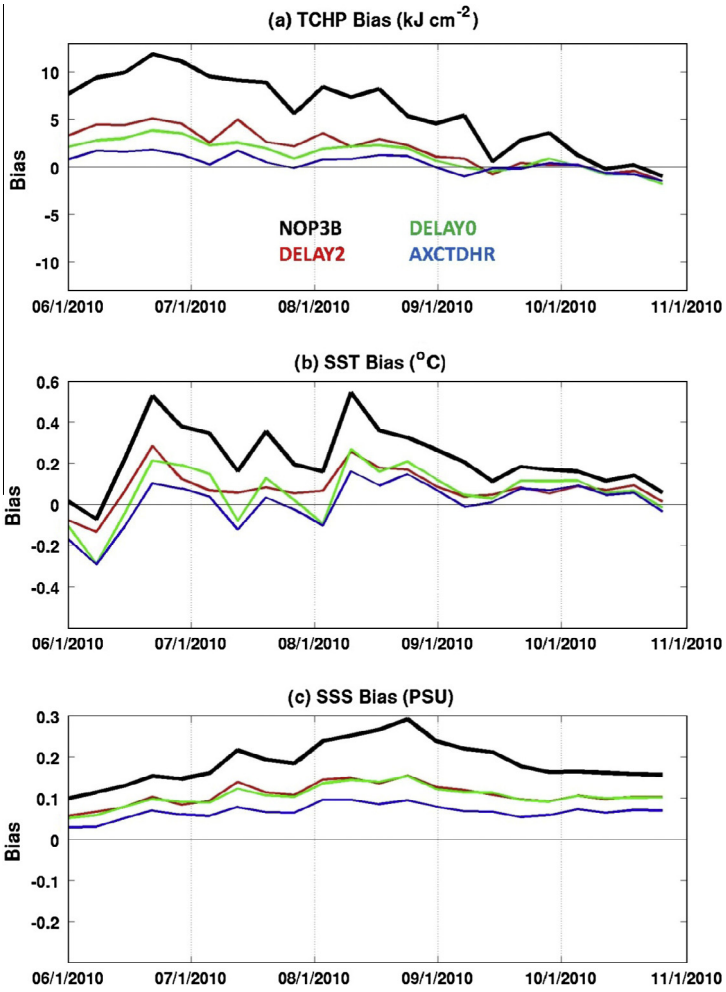

Fig. 4. Time series of model mean bias between the Nature Run and the 7-day analysis products from OSSE experiments NOP3, DELAY0, DELAY2, and AXCTDHR for (a) TCHP, (b) SST, and (c) SSS, Line colors are given by the legend in panel (a).

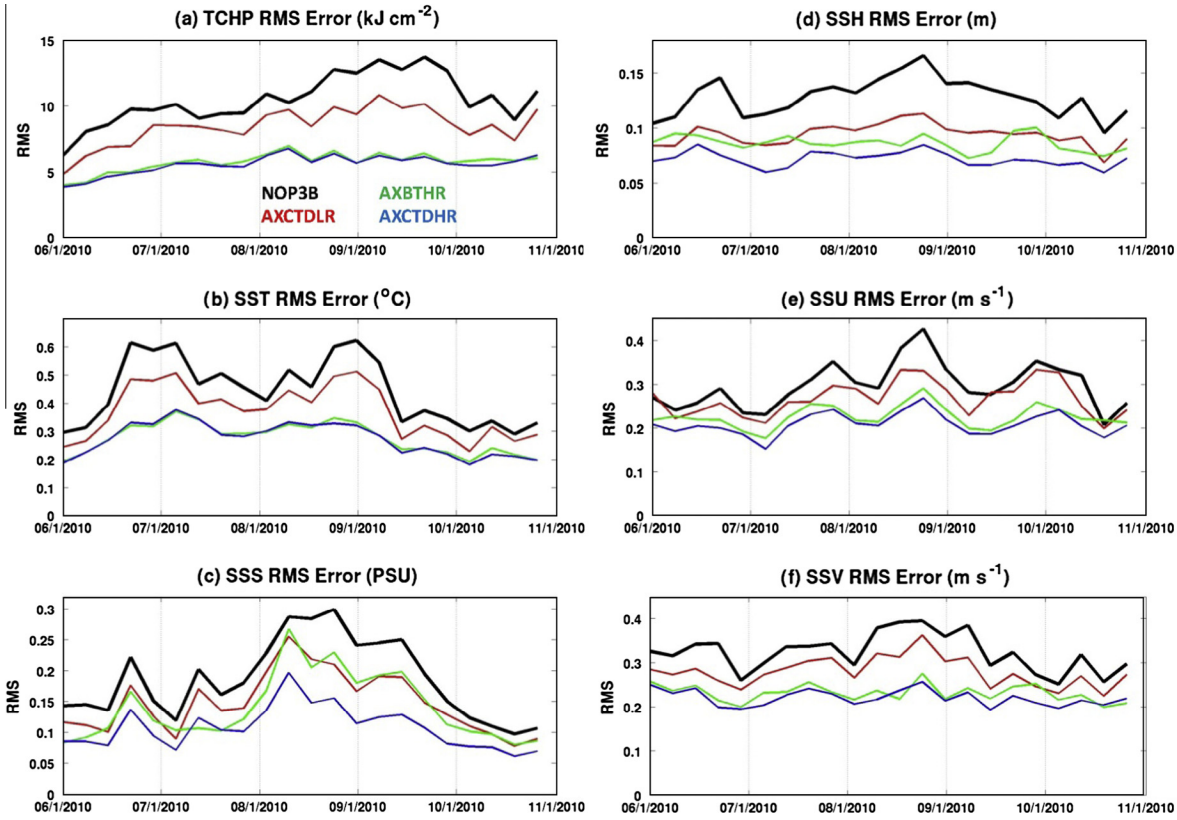

Fig. 5. Time series of RMS error between the Nature Run and the 7-day analysis products from OSSE experiments NOP3, AXCTDLR, AXBTHR, and AXCTDHR for (a) TCHP, (b) SST, (c) SSS, (d) SSH, (e) SSU, and (f) SSV. Line colors are given by the legend in panel (a).

Table 5
RMS errors and Murphy skill scores among four forecast model experiments that illustrate the impact of instrument type and the NR for six model fields. The errors and scores were calculated over all common model grid points within the domain sampled by airborne surveys comparing the weekly analyses from these NRT-mode experiments within the time interval 1 June through 30 September 2010. Percentage reduction of RMS error from experiment NOP3B for all variables is shown in parentheses for experiments AXCTDHR, AXBTHR, and AXCTDLR.

	Experiment	TCHP	SST	SSS	SSH	SSU	SSV
RMS Error	NOP3B	11.2	0.46	0.34	0.13	0.31	0.34
	AXCTDHR	5.99 (−47%)	0.30 (−35%)	0.22 (−35%)	0.07 (−46%)	0.21 (−32%)	0.22 (−35%)
	AXBTHR	6.15 (−45%)	0.30 (−35%)	0.27 (−21%)	0.09 (−31%)	0.23 (−23%)	0.24 (−24%)
	AXCTDLR	9.40 (−24%)	0.39 (−22%)	0.26 (−29%)	0.10 (−23%)	0.28 (−19%)	0.29 (−21%)
Murphy Skill Score	NOP3B	0.49	0.73	−0.24	0.78	0.49	0.61
	AXCTDHR	0.89	0.91	0.68	0.92	0.76	0.82
	AXBTHR	0.89	0.91	0.47	0.83	0.72	0.80
	AXCTDLR	0.73	0.84	0.32	0.88	0.59	0.71

The impact of horizontal resolution is assessed using NRT-mode experiment AXCTDLR (Fig. 1b, Table 2) which is identical to AXCTDHR except for assimilating profiles sampled on the 1.0° grid instead of the 0.5° grid. Inspection of RMS error time series (Fig. 5) along with mean RMS errors and skill scores (Table 5) with respect to the NR demonstrates that reduced horizontal resolution results in a substantial RMS error increase for all six model fields. For example, RMS error in TCHP produced by experiment NOP3B is further reduced by 47% and 24% in experiments AXCTDHR and AXCTDLR, respectively (Table 5). The corresponding additional error reductions in SSH over NOP3B are 46% and 23%. Skill scores for AXCTDLR are smaller than those for AXCTDHR (Table 5). This skill score reduction is greatest for SSS (0.68 for high resolution to 0.32 for low resolution), again emphasizing the importance of *in-situ* salinity profiles. Lower horizontal resolution has a more modest negative impact on bias reduction for TCHP and SST than does replacing AXCTDs with AXBTs (Fig. 6). Overall, high-resolution airborne surveys constrain ocean features with finer-scale horizontal resolution than is possible with satellite altimetry alone because of low cross-track and temporal resolution. This is particularly important in energetic ocean regimes containing strong frontal boundaries and current jets, such as the interior eastern GoM.

RMS error maps for experiment AXCTDHR calculated within the airborne sampling domain for TCHP (Fig. 7a) and SSH (Fig. 8a) enable the spatial distribution of these errors to be analyzed. Within this domain, the largest errors for both variables exist in the eastern part of the domain where the LC and associated eddies dominate the variability. Because this experiment produced the largest error reduction, it is presented as the reference experiment against which the impacts of not assimilating airborne profiles, decreasing the horizontal resolution of these profiles, and deploying shallow AXBTs instead of deep AXCTDs can be assessed. Figs. 7b and 8b present the increases in RMS error in experiment NOP3B over those produced by AXCTDHR (Figs. 7a and 8a) for TCHP and

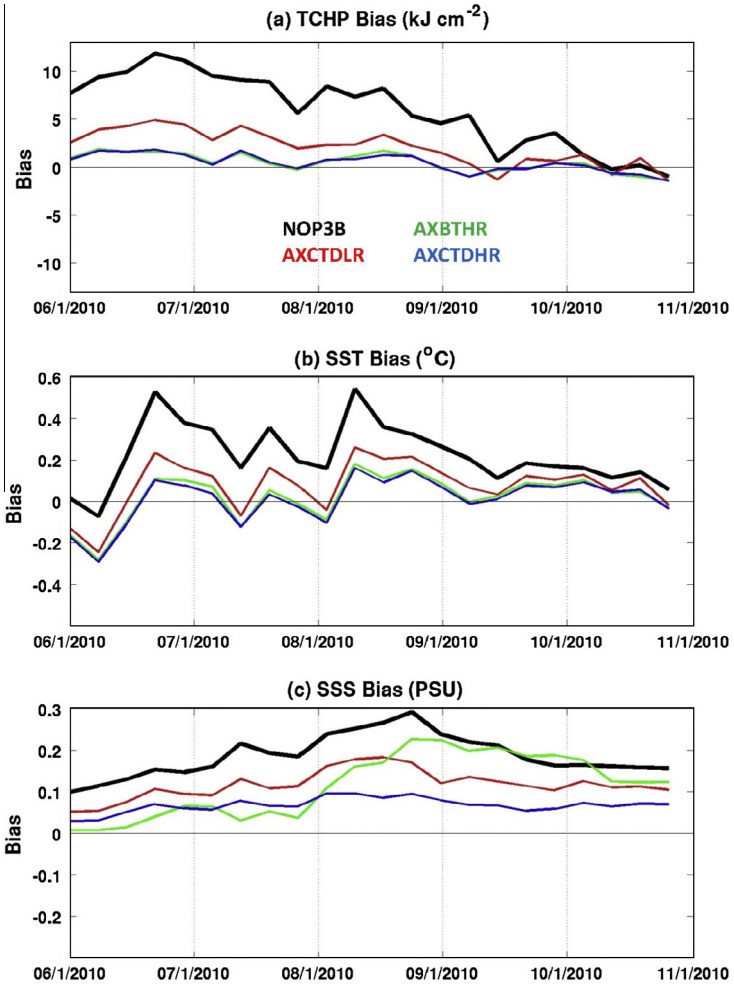

Fig. 6. Time series of model mean bias between the Nature Run and the 7-day analysis products from OSSE experiments NOP3, AXCTDLR, AXBTHR, and AXCTDHR for (a) TCHP, (b) SST, and (c) SSS, Line colors are given by the legend in panel (a).

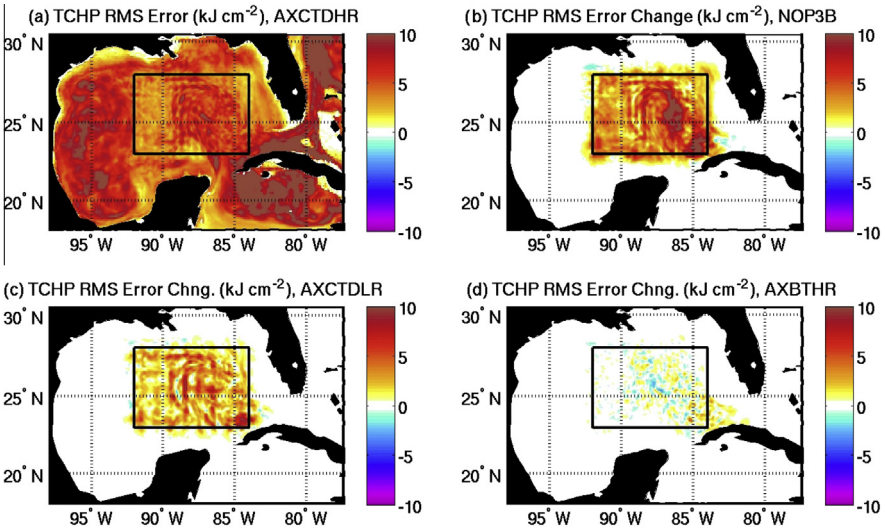

Fig. 7. (a) RMS error map for TCHP between the Nature Run and OSSE experiment AXCTDHR; change (increase) in RMS error (with respect to AXCTDHR) in experiment: (b) NOP3B, when all airborne profiles are denied; (c) AXCTDLR, showing the impact of lower profile resolution; and (d) AXBTHR, showing the impact of assimilating shallow AXBTs versus deep AXCTDs.

SSH, respectively, revealing the error increase patterns resulting from denial of airborne observations. Large error increases are observed for both fields, with the largest values found in the eastern part of the survey domain in the LC extension and eddy formation region where the strongest ocean features and frontal boundaries are present (Fig. 2).

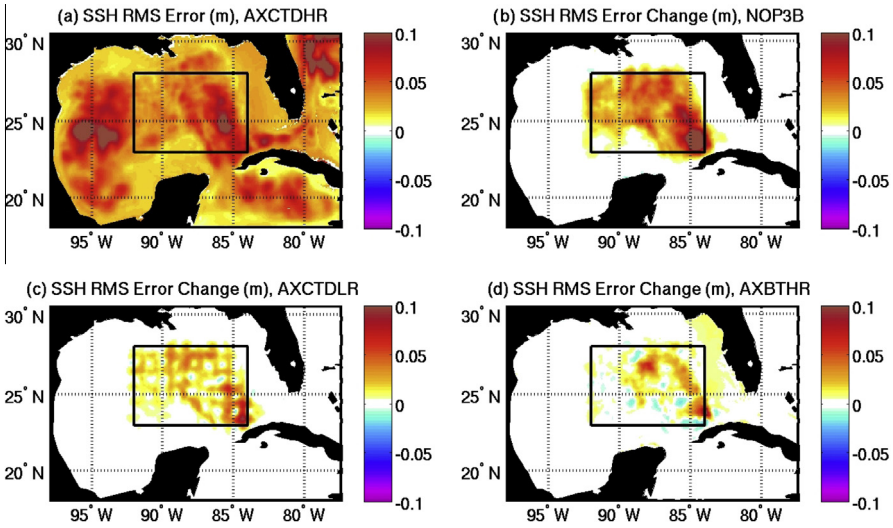

Fig. 8. Same as Fig. 7 for SSH.

Figs. 7c and 8c present the increases in RMS error in experiment AXCTDLR over the RMS error produced by experiment AXCTDHR, revealing the error increase pattern resulting from reduced horizontal resolution. Substantial error increases are observed for both TCHP (Fig. 7c) and SSH (Fig. 8c), being more confined to the southeastern region of the survey domain for SSH where the largest variability exists. Small-scale structure is observed in both error increase patterns, with the largest increases tending to occur at the 0.5° airborne sampling grid points that were denied in experiment AXCTDLR (the small dots in Fig. 1b). This structure exists despite the relatively large horizontal localization radius (140 km) specified for assimilating the airborne profiles by the DA system. It demonstrates that the higher-resolution sampling reduces RMS errors primarily by constraining smaller-scale ocean variability in this energetic frontal region that is not well constrained by altimetry. Figs. 7d and 8d present the increases in RMS error in experiment AXBTHR over the RMS error produced by experiment AXCTDHR, revealing the error increase pattern resulting from replacing deep AXCTDs with shallow AXBTs. This change of probes and profile depths has insignificant impact on TCHP error (Fig. 7d), as noted earlier. Error increase is evident for SSH only over a limited part of the survey domain (Fig. 8d) which forms an arc that tends to follow the boundary of the extended LC path present during most of the study interval (Fig. 2). This region begins north of the Yucatan peninsula at 24°N, extends farther north to 27°N, and turns eastward and then southward toward the entrance to the Florida Straits. The reduced profile depths of the AXBTs along with the reduced accuracy of derived salinity profiles apparently reduce the ability to constrain boundaries separating distinct ocean features along the LC path.

Finally, reanalysis-mode experiments were also conducted to assess the impact of changing probe type and horizontal resolution. Very similar results were obtained compared to the NRT-mode experiments, so they are not presented herein.

Impact of temporal sampling intervals for repeat airborne surveys

The mesoscale airborne survey program conducted in response to the DWH oil spill was designed to repeatedly sample the ocean over an extended interval of time. Surveys were separated by time intervals ranging from 3 days to 2 weeks. For future programs of this type, it will be beneficial to pre-determine quasi-optimum temporal sampling intervals based on the resulting error reduction in model analysis products used to initialize forecast models. The impact of temporal survey frequency is assessed herein by comparing error reduction in reanalysis-mode experiments DTN, where $N = 1, 2, 4, 8$, and 16 gives the number of days between airborne surveys (Table 2), to error reduction in reanalysis-mode experiment NOP3A.

A convenient visualization of error reduction as a function of time interval between airborne surveys is provided by plotting three related error metrics (RMS amplitude, RMS error, and correlation coefficient) on Taylor (2001) diagrams (Fig. 9). Points from all of the DTN experiments on these diagrams are compared to the reference point from experiment NOP3A, and also to the reference point that represents a perfect comparison. As expected, RMS error decreases and correlation increases monotonically for all fields, as the temporal sampling interval decreases. The largest improvement over experiment NOP3A is realized for SSH, which exceeds 50% for both experiments DT1 and DT2. The second largest improvement is for SSS while the least improvement is for SST. For all six model fields, only modest additional improvements are achieved by DT1 compared to DT2. Except for SSS, little error decrease is realized for sampling intervals $\geqslant$8 days and it is only for sampling intervals $\leqslant$4 days that large error reduction is achieved. All experiments DT1 through DT16 produce progressively smaller RMS errors and larger skill scores in comparison to reanalysis-mode experiment NOP3A (Table 6). Improvement in these metrics is substantial with decreasing time intervals down to 2 days with only small additional improvement achieved with daily surveys.

Although detectable error reduction is achieved on average for survey intervals up to 16 days, the large error reduction achieved by individual surveys is short-lived. To illustrate this point, time series of RMS error for individual model fields from the DTN experiments are inspected (not shown). The typical pattern is for large error reduction to be achieved on the analysis day followed by relatively rapid error growth until the error is again reduced on the next analysis day, essentially producing a "sawtooth" pattern in the time series. To illustrate this in a compact manner for experiments DT2, DT4, and DT8, all error correction-rebound cycles within the time interval June through October 2010 are averaged together to produce a mean error cycle as a function of time from analysis day. These mean cycles for DT2, DT4, and DT8 are plotted in Fig. 10 for all six model fields out to day 8 from the analysis time, which requires the correction-rebound cycles for DT2 and DT4 to be repeated over this full time interval. The RMS error from experiment NOP3A averaged in the identical manner as DT8 is also

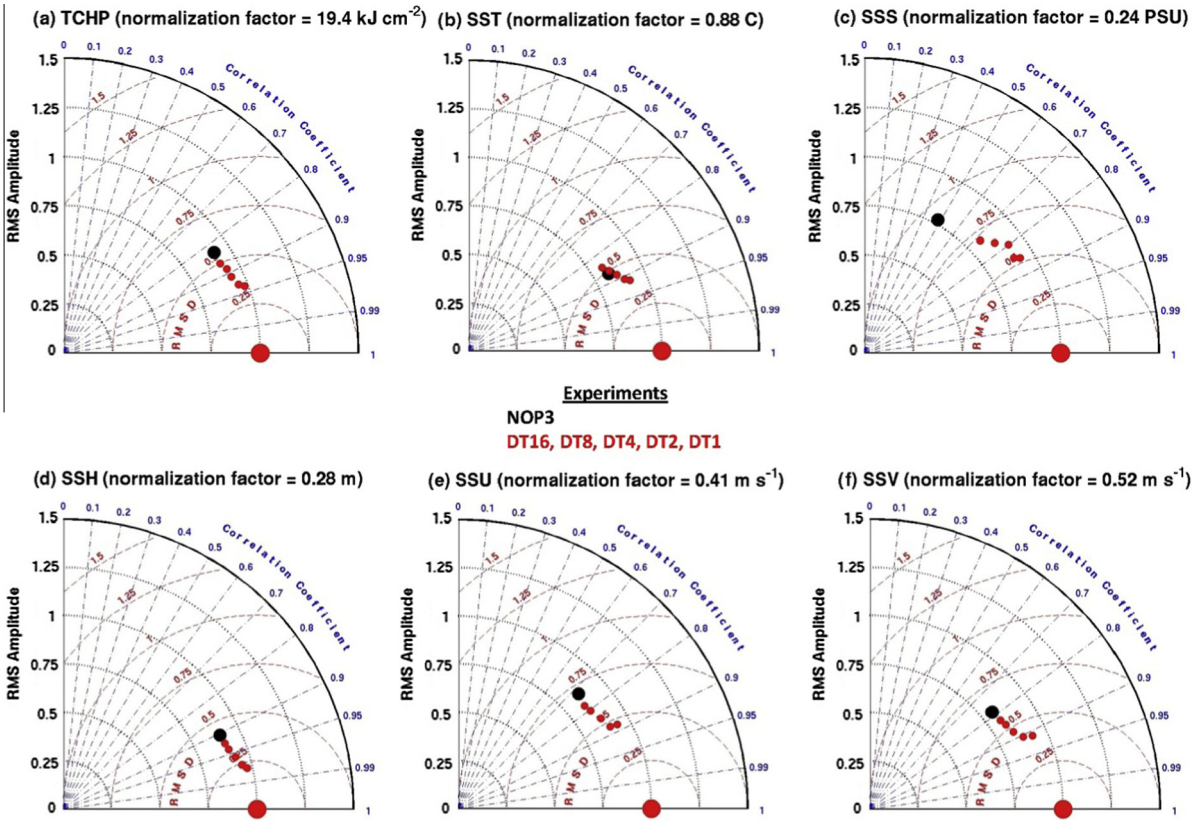

Fig. 9. Taylor diagrams comparing experiment NOP3A (large black dots) to experiment DT1, DT2, DT4, DT8, and DT16 (small red dots, ordered in sequence from bottom right toward top left) for (a) TCHP, (b) SST, (c) SSS, (d) SSH, (e) SSU, and (6) SSV. The large red dot signifies a perfect comparison.

Table 6
RMS errors and Murphy skill scores among six forecast model experiments that illustrate the impact of temporal separation between airborne surveys and the NR for six model fields. The errors and scores were calculated over all common model grid points within the domain sampled by airborne surveys from 1 June through 30 September 2010 using daily analyses from these reanalysis-mode experiments. Percentage reduction of RMS error from experiment NOP3B for all variables is shown in parentheses for experiments DT1, DT2, DT4, DT8, and DT16.

	Experiment	TCHP	SST	SSS	SSH	SSU	SSV
RMS Error	NOP3A	10.9	0.44	0.34	0.13	0.30	0.33
	DT1	6.36 (−42%)	0.28 (−36%)	0.15 (−56%)	0.07 (−46%)	0.20 (−35%)	0.22 (−33%)
	DT2	6.57 (−40%)	0.30 (−32%)	0.16 (−53%)	0.08 (−38%)	0.20 (−33%)	0.23 (−30%)
	DT4	7.53 (−31%)	0.35 (−20%)	0.19 (−44%)	0.09 (−31%)	0.23 (−23%)	0.25 (−24%)
	DT8	8.44 (−23%)	0.40 (−9%)	0.22 (−35%)	0.10 (−23%)	0.25 (−17%)	0.28 (−15%)
	DT16	9.37 (−14%)	0.44 (0%)	0.25 (−26%)	0.11 (−15%)	0.27 (−10%)	0.30 (−9%)
Murphy Skill Score	NOP3A	0.50	0.75	−0.25	0.80	0.52	0.62
	DT1	0.88	0.92	0.82	0.94	0.79	0.84
	DT2	0.87	0.91	0.79	0.93	0.79	0.83
	DT4	0.82	0.87	0.68	0.90	0.72	0.79
	DT8	0.77	0.83	0.58	0.87	0.66	0.74
	DT16	0.69	0.77	0.44	0.84	0.61	0.70

included in Fig. 10 for reference. These patterns demonstrate that because of relatively rapid error growth, it is necessary to conduct frequent airborne surveys (at least every 4 days and preferably every 2 days) to maintain most of the error reduction achieved by individual surveys.

OSSE impact assessments based on ocean forecasts

Each NRT experiment provides ocean analysis fields at 7-day time intervals that can be used to initialize forecasts. To assess the impact of airborne profile assimilation on ocean forecasts, the set of fifteen analyses produced between 26 May and 1 September 2010 separated by 7-day time intervals were used to initialize 60-day forecasts. The first three analyses produced during May were excluded because TCHP is either small or zero over much of the interior eastern GoM. Analyses produced after 1 September could not be used because the 60-day forecasts extended beyond the termination time for all experiments.

RMS error time series [RMS(t)] for all fifteen individual forecasts initialized by experiment AXCTDHR are presented in Fig. 11 for the six model fields, and these are referenced to mean RMS error determined from experiment FREE. For all model fields, there is considerable scatter in RMS error evolution among individual forecast experiments. For thermodynamical variables, RMS(t) tends to grow relatively rapidly for TCHP and SST over the initial 10 forecast days,

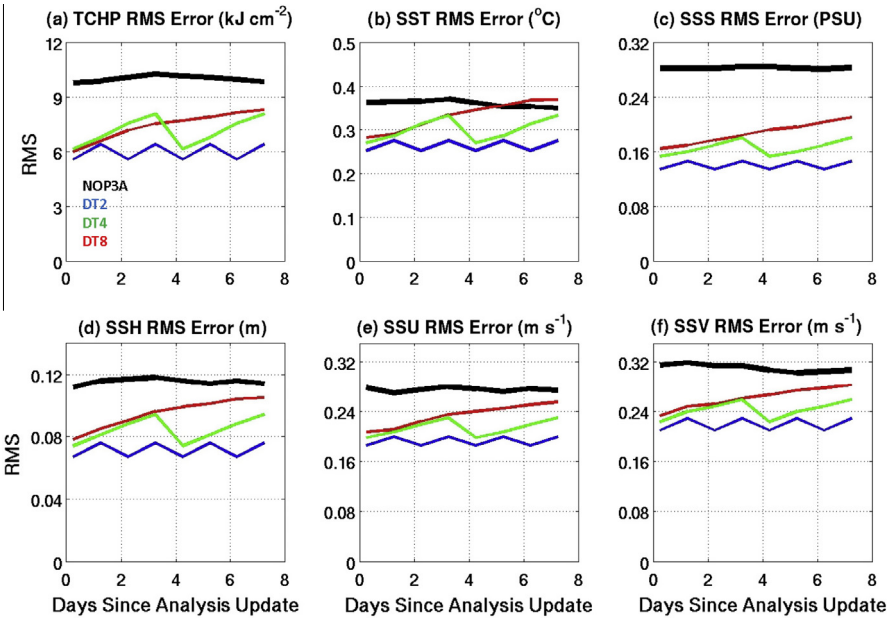

Fig. 10. RMS error between the NR and OSSE experiments NOP3A (thick black line), DT2 (blue line), DT4 (green line), and DT8 (red line) for (a) SSH, (b) SST, (c) SSS, (d) SSH, (e) SSU, and (f) SSV), all calculated over the domain sampled by synthetic airborne surveys (Fig. 1b). These curves illustrate the RMS error growth in DT2, DT4, and DT8 as a function of time since the analysis update. Mean RMS error at each time is averaged over the 18 7-day analysis times from June through October 2010.

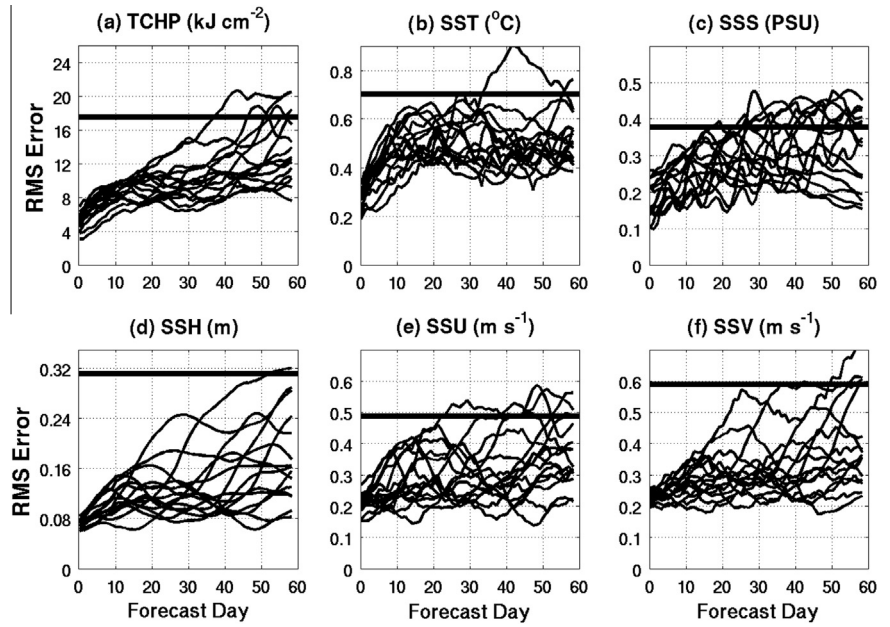

Fig. 11. RMS forecast errors for six model fields from fifteen 60-day forecasts initialized from experiment AXCTDHR, each initialized from a 4-day analysis separated by 7-day time intervals between 26 May and 1 September 2010. The horizontal black lines represent the mean RMS error during this time interval from experiment FREE. For each forecast, RMS error is calculated daily between the experiment and the NR over all common grid points in the domain sampled by the airborne surveys (Fig. 1b).

after which it tends to grow slowly for TCHP and level off for SST. Errors for TCHP remain substantially smaller than the mean error from experiment FREE for the majority of individual forecasts. The rapid leveling off for SST may result from the tendency for model SST to follow surface atmospheric temperature imposed by the surface forcing. Forecasts for SSS tend to grow monotonically, but with considerable scatter and short-period fluctuations in amplitude. For dynamical fields, RMS error for SSH grows substantially for a small number of forecasts. For other forecasts, however, the average error after day 10 tends to level off. As expected, the predictability of the LC system is sensitive to the initial state of the LC and associated eddies. The RMS error time series for SSU and SSV are noisier, but the same pattern seen for SSH remains evident at least for SSV. Overall, substantial forecast error reduction is achieved by assimilating AXCTDs on two back-to-back days. Depending on the initial state of the ocean, some forecasts retain most of this improvement even after 60 forecast days, while other forecasts produce much more rapid error growth with time.

To facilitate visualization of forecast error growth, individual time series of RMS(t)–RMS(0) are calculated over all 15 forecasts

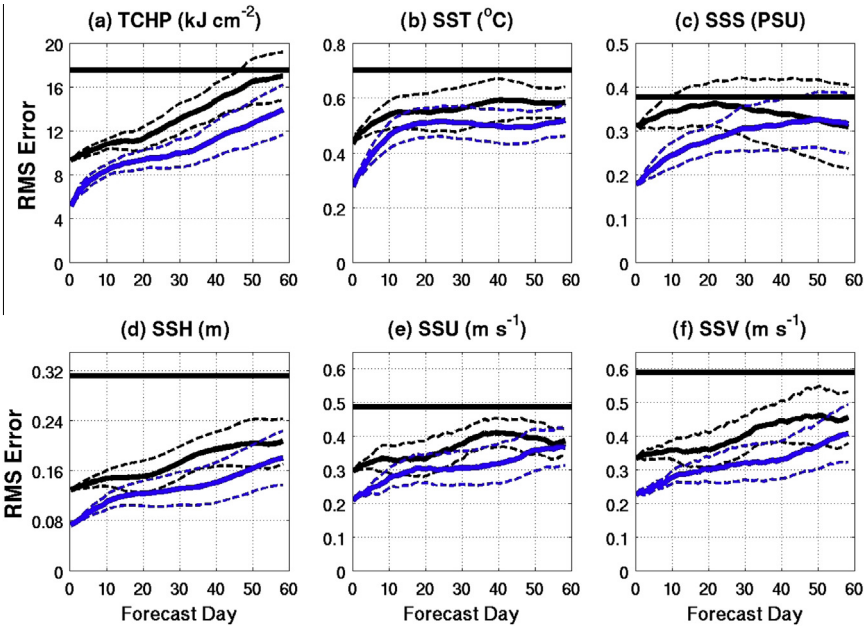

Fig. 12. RMS forecast errors for six model fields. The horizontal black lines represent the mean RMS error during this time interval from experiment FREE. The other black lines are forecasts initialized from experiment NOP3B while the blue lines are forecasts initialized from experiment AXCTDHR. Errors from a total of fifteen forecasts are averaged on every day, each initialized from a 7-day analysis between 26 May and 1 September 2010 and calculated over the domain sampled by synthetic airborne surveys (Fig. 1b). For AXCTDHR, this is the average of the fifteen forecasts shown in Fig. 11. The dashed lines delineate approximate 95% confidence limits.

for each model field, averaged together, and then plotted in Fig. 12. The evolution of mean error in forecasts initialized by experiment AXCTDHR is compared to forecasts initialized by experiment NOP3B to visualize the improvement resulting from assimilation of airborne profiles. Confidence limits (95%) for mean RMS(t)–RMS(0) are also plotted in Fig. 12, although these limits are approximate given the non-normal distribution of forecast error at each time. In all model fields, the additional error reduction at the beginning of the AXCTDHR forecasts with respect to the NOP3B forecasts shrinks relatively rapidly over the first 10–15 days, more so in the thermodynamical fields than in the dynamical fields. However, mean forecast error from AXCTDHR remains less that the mean error from NOP3B over all forecast times and for all fields, except near the end of the 60-day forecast interval for both SSS and SSU.

Confidence intervals increase substantially at longer forecast times, so a much larger number of forecasts sampling a broader range of initial states of the LC system should be conducted to accurately determine predictability improvements resulting from airborne profile assimilation. A rough measure of predictability time scale is estimated from the present analysis as the forecast day where the confidence intervals first overlap. This time scale equals 47 days for TCHP (although they nearly overlap at day 18), 8 days for SST, 21 days for SSS, 12 days for SSH, 8 days for SSU, and 12 days for SSV. The longer time lag before TCHP confidence limits permanently overlap suggests that the airborne surveys are particularly effective at improving extended upper-ocean temperature forecasts.

Summary

Although assimilation of satellite altimetry substantially constrains the three-dimensional structure of ocean boundary currents and mesoscale eddies in ocean analysis products, significant additional error reduction is achievable by conducting rapid-response airborne surveys that sample upper-ocean profiles in a quasi-synoptic pattern. The present study employed a prototype ocean OSSE system to quantify the degree of improvement that can be achieved as a function of several aspects of airborne survey design. The design and validation of this OSSE system has followed procedures developed for atmospheric OSSE systems to insure that credible observing system impact assessments are realized.

The impacts of different airborne sampling strategies were assessed with two specific applications in mind. The first application is to improve the accuracy of ocean analyses and forecasts for the purpose of oil spill forecasting, which requires accurate representation of ocean dynamical fields. Airborne survey programs such as the repeat survey program conducted in response to the Deepwater Horizon oil spill can more-optimally constrain model analyses and forecasts if several strategies are followed. In the interior eastern GoM, model errors grow rapidly which requires that repeat surveys be conducted frequently in time, separated by no more than 4 days, to retain most of the analysis error reduction between surveys. The actual rapid-response surveys conducted for the oil spill were separated by up to 2 weeks. During such long gaps, the large majority of the error reduction achieved by the previous analysis was probably lost by the time of the next survey.

Substantial additional error reduction in dynamical fields is achieved by deploying deep (1000 m) AXCTDs instead of shallow (400 m) AXBTs, and also by increasing horizontal profile resolution to less than one degree. In addition to constraining ocean fields over a greater depth range, assimilation of measured salinity profiles from AXCTDs more accurately constrains the upper-ocean density field than the derived salinity profiles assimilated along with AXBT measured temperature profiles. The increased resolution constrains the structure of ocean fields with horizontal scales too small to be adequately constrained solely by satellite altimetry assimilation.

The second application of the present analysis is to reduce ocean model initialization errors in coupled TC forecast models with the overarching goal of improving intensity forecasts. This application typically involves conducting a one-time rapid-response ocean survey or series of surveys ahead of individual storms such as Hurricanes Isidore and Lili in 2004 (Shay and

Uhlhorn, 2008; Uhlhorn and Shay, 2012, 2013) as opposed to an extended repeat-survey program. The impacts of probe type and depth along with horizontal resolution on ocean dynamical fields summarized above is also relevant to this application because background vorticity associated with mesoscale ocean features distorts the SST response to TC forcing (Jaimes et al., 2011). However, additional stringent requirements exist for this application, specifically that the upper-ocean thermal and density fields be accurately constrained by the airborne profiles within mesoscale ocean features. In this context, shallow AXBTs constrain the ocean thermal field over the upper 400 m nearly as well as deep AXCTDs since both probes sample temperature over this depth range. However, the additional improvement in salinity (and hence density) achieved by assimilating AXCTD profiles will reduce density profile errors that then contribute to improved SSH analyses. These profiles must be rapidly made available on the GTS or directly to operational forecast centers for assimilation. Although a 2-day lag does not significantly increase errors in dynamical ocean fields, errors in the upper-ocean thermal field including TCHP do increase, demonstrating importance of data latency for the TC application. Substantial additional error reduction is achieved by conducting two rapid-response airborne surveys on back-to-back days compared to a single pre-storm survey.

The additional error reduction in ocean analyses achieved by assimilation of airborne surveys translates into significantly improved forecasts persisting over time intervals between 1 and 2 weeks for SSH, SSU, SSV, and SST, about 3 weeks for SSS, and 8 weeks for TCHP. Relatively short time intervals exist for the three dynamical fields due to rapid error growth in this energetic ocean region. Short time intervals also exist for SST in part because this field is already well constrained due to good observational coverage by satellite and *in situ* instruments, and also because rapid adjustment of SST toward surface atmospheric temperature occurs during the forecasts. A longer time interval exists for SSS because it is very poorly constrained without airborne profile assimilation, but is well constrained with airborne profile assimilation, particularly when deep AXCTDs are used. The longest time interval exists for TCHP, demonstrating that subsurface upper-ocean temperature forecasts are significantly improved for an extended interval of time.

Airborne survey impacts documented herein are valid for the interior GoM, particularly for the region dominated by the LC system. This is the specific region within which the OSSE system was first validated. Impact assessments may differ in other oceanographic regions, particularly less-energetic regions with weaker currents and frontal boundaries where error growth may be slower and predictability time scales longer. Our plans are to extend this OSSE system to other oceanographic regions where these issues can be studied.

Acknowledgments

Support is acknowledged from the USWRP Hurricane Forecast Improvement Project, from the NOAA Office of Weather and Air Quality through the OSSE testbed, and from a NOAA Science Box Grant. GRH acknowledges support from Grant OAR-M8R2WHSP01 and from the Physical Oceanography Division of NOAA/AOML. VHK and MLH acknowledge support from NOAA NA10OAR4320143 and NA13OAR4830224. LKS acknowledges support from NOAA and BOEM in the acquisition and analysis of the oceanic profiles during DWH. LKS also acknowledges the pilots, technicians and engineers at the NOAA Aircraft Operations Center who made it possible to acquire oceanographic profiles from the research aircraft during DWH.

References

Arnold, C.P., Dey, C.H., 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.

Atlas, R., Kalnay, E., Halem, M., 1985. Impact of satellite temperature soundings and wind data on numerical weather prediction. Opt. Eng. 24, 341–346.

Atlas, R., Kalnay, E., Baker, W.E., Susskind, J., Reuter, D., Halem, M., 1985b. Simulation studies of the impact of future observing systems on weather prediction. Proceedings, 7th AMS Conference on Numerical Weather Prediction, pp. 145–151.

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

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

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

DeMaria, M., Mainelli, M., Shay, L.K., Knaff, J.A., Kaplan, J., 2005. Further improvements to the statistical hurricane intensity prediction scheme (SHIPS). Weather Forecast. 20, 531–543.

Goni, G., DeMaria, M., Knaff, J., Sampson, C., Ginis, I., Bringas, F., Mavume, A., Laner, C., Lin, I.-I., Ali, M.M., Sandery, P., Ramos-Buarque, S., Kang, K., Mehra, A., Chassigner, E., Halliwell, G., 2009. Applications of satellite-derived ocean measurements to tropical cyclone intensity forecasting. Oceanography 22, 176–183.

Halliwell Jr., G.R., 2004. Evaluation of vertical coordinate and vertical mixing algorithms in the hybrid-coordinate ocean model (HYCOM). Ocean Model. 7, 285–322.

Halliwell Jr., G.R., Barth, A., Weisberg, R.H., Hogan, P., Smedstad, O.M., Cummings, J., 2009. Impact of GODAE products on nested HYCOM simulations of the West Florida Shelf. Ocean Dyn. 59, 139–155.

Halliwell Jr., G.R., Srinivasan, A., Kourafalou, V., Yang, H., Willey, D., Le Hénaff, M., Atlas, R., 2014. Rigorous evaluation of a fraternal twin ocean OSSE system in the open Gulf of Mexico. J. Atmos. Ocean. Technol. 31 (1), 105–130. http://dx.doi.org/10.1175/JTECH-D-13-00011.1.

Jaimes, B., Shay, L.K., Halliwell, G.R., 2011. The response of quasigeostrophic vortices to tropical cyclone forcing. J. Phys. Oceanogr. 41, 1965–1985.

Kourafalou, V.H., Peng, G., Kang, H., Hogan, P.J., Smedstad, O.M., Weisberg, R.H., 2009. Evaluation of global ocean data assimilation experiment products on South Florida nested simulations with the Hybrid Coordinate Ocean Model. Ocean Dyn. 59 (1), 47–66. http://dx.doi.org/10.1007/s10236-008-0160-7.

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

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

Lin, I.-I., Goni, G.A., Knaff, J.A., Forbes, C., Ali, M.M., 2013. Ocean heat content for tropical cyclone intensity forecasting and its impact on storm surge. Nat. Hazards 66, 1481–1500.

Liu, Y., MacFadyen, A., Ji, Z.-G., Weisberg, R.H. (Eds.), 2011a. Monitoring and modeling the deepwater horizon oil spill: a record-breaking enterprise. Geophys. Monogr. Ser. vol. 195, 271 pp., AGU, Washington, D.C. doi:10.1029/GM195.

Liu, Y., Weisberg, R.H., Hu, C., Zheng, L., 2011b. Tracking the Deepwater Horizon oil spill: a modeling perspective. EOS, Trans. Amer. Geophys. Union 92 (6), 45–46.

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

Meyers, P.C., Shay, L.K., Brewster, J.K., 2014. The development of the systematically merged Atlantic regional temperature and salinity climatology for hurricane intensity forecasting. J. Atmos. Oceanic Tech. 31, 131–149.

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

Price, J.F., 2009. Metrics of hurricane-ocean interaction: vertically integrated or vertically averaged ocean temperature. Ocean Sci. 5, 351–368.

Sanabia, E.R., Barrett, B.S., Black, P.G., Chen, S., Cummings, J.A., 2013. Real-time upper-ocean temperature observations from aircraft during operational hurricane reconnaissance missions: AXBT demonstration project year one results. Weather Forecast. 28, 1404–1422.

Schiller, R.V., Kourafalou, V.H., Hogan, P.J., Walker, N.D., 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. http://dx.doi.org/10.1029/2010JC006883.

Shay, L.K., Uhlhorn, E., 2008. Loop Current response to hurricanes Isidore and Lili. Mon. Weather Rev. 137, 3248–3274.

Shay, L.K., Brewster, J.K., 2010. Oceanic heat content variability in the eastern Pacific Ocean for hurricane intensity forecasting. Mon. Weather Rev. 138, 2110–2131.

Shay, L.K., Goni, G.J., Black, P.G., 2000. Effects of a warm oceanic feature on Hurricane Opal. Mon. Weather Rev. 125 (5), 1366–1383.

Shay, L.K., Jaimes, B., Brewster, J.K., Meyers, P., McCaskill, E.C., Uhlhorn, E., Marks, F., Halliwell, Jr. G.R., Smedstad, O.-M., Hogan, P., 2011. Airborne ocean surveys of

the Loop Current complex from NOAA WP-3D in support of the Deep Water Horizon oil spill. In: Liu et al. (Eds.), Monitoring and Modeling the Deepwater Horizon Oil Spill: A Record-Breaking Enterprise. AGU Monograph Series, pp. 131–151.

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

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

Uhlhorn, E., Shay, L.K., 2012. Loop Current mixed layer response to hurricane Lili (2002) Part I: Observations. J. Phys. Oceanogr. 42, 400–419.

Uhlhorn, E., Shay, L.K., 2013. Loop Current mixed layer response to hurricane Lili (2002) Part II: Modeling results. J. Phys. Oceanogr. 43, 1173–1192.

Vincent, E.M., Lengaigne, M., Vialard, J., Madec, G., Jourdain, N.C., Masson, S., 2012. Assessing the oceanic control on the amplitude of sea surface cooling induced by tropical cyclones. J. Geophys. Res. 117, C05023. http://dx.doi.org/10.1029/2011JC007705.