Deep-Sea Research II 129 (2016) 374–393

Contents lists available at ScienceDirect

Deep-Sea Research II

journal homepage: www.elsevier.com/locate/dsr2

The performance of the US Navy's RELO ensemble, NCOM, HYCOM during the period of GLAD at-sea experiment in the Gulf of Mexico

Mozheng Wei [a,*], Gregg Jacobs [a], Clark Rowley [a], Charlie N. Barron [a], Pat Hogan [a], Peter Spence [b], Ole Martin Smedstad [b], Paul Martin [a], Philip Muscarella [c], Emanuel Coelho [d]

[a] Naval Research Laboratory, Stennis Space Center, MS 39529, USA
[b] QinetiQ-North America, Stennis Space Center, MS 39529, USA
[c] ASEE Post doc at NRL, Stennis Space Center, MS 39529, USA
[d] University of New Orleans at NRL, Stennis Space Center, MS 39529, USA

ARTICLE INFO

Available online 12 September 2013

Keywords:
Ocean ensemble prediction
Ocean data assimilation
Ensemble spread and reliability
Forecast accuracy and skill
Lagrangian coherent structure

ABSTRACT

A suite of real-time ocean model forecasts was carried out successfully at NRL to provide modeling support and guidance to the CARTHE GLAD at-sea experiment during summer 2012. The forecast systems include two RELO ensembles and three single models using NCOM and HYCOM with different resolutions. All of these forecast outputs are archived and made available on web servers for the CARTHE scientists. The detailed descriptions of these forecast systems and the products presented in this paper provide a much-needed background to the scientists in CARTHE and others who will use our forecasts and GLAD drifter observations to do further research after the future public release of the CARTHE GLAD data.

A calibrated ensemble system with enhanced spread and reliability is proposed in this project. It is found that this calibrated ensemble outperforms the un-calibrated ensemble in terms of quantitative forecasting accuracy, skill and reliability for all the variables and observation spaces we have evaluated. The metrics used include RMS error, anomaly correlation, spread-reliability and Talagrand rank histogram. Both ensembles are compared with three single-model forecasts with NCOM and HYCOM with different resolutions. The advantages of ensembles are demonstrated.

RELO ensembles have been applied to Lagrangian trajectory prediction, and it is demonstrated that either ensemble can provide valuable uncertainty information in addition to predicting the particle trajectory with highest probability in comparison with a single ocean model forecast. The calibrated ensemble with more reliability is able to capture some trajectories in different, even opposite directions which are missed by the un-calibrated ensemble. When the ensembles are applied to computing the LCS (Lagrangian Coherent Structure), the uncertainties of the LCSs, which cannot be estimated from a single model forecast, are identified. Another finding is that the LCS depends on the model resolution. The model with highest resolution produces the finest small-scale LCS structures, while the model with lowest resolution generates only large scale LCSs. The work on using ocean ensembles in Lagrangian ocean dynamics presented in this paper represents our initial attempt in this field. It is expected that this work will lead to more extensive new research in this area in the near future.

Published by Elsevier Ltd.

1. Introduction

The Grand Lagrangian Deployment (GLAD) is an at-sea experiment that was supported by the Consortium for Advanced Research on Transport of Hydrocarbon in the Environment (CARTHE, http://www.carthe.org/, and other papers in this special issue) and the Gulf of Mexico Research Initiative (GoMRI, http://gulfresearchinitiative.org/). GoMRI is funded by BP following the Deep Water Horizon (DWH) drilling rig explosion approximately 60 km off the coast of Louisiana on April 20, 2010. CARTHE is one of the consortia of GoMRI and it comprises 26 principal investigators from 12 universities and research institutions including the Naval Research Laboratory (NRL) distributed across four Gulf of Mexico states and four other states. The GLAD experiment was conducted in the northern Gulf of Mexico (GOM) from 17 July to 3 August 2012 by the members of the GoMRI CARTHE consortium. During the experiment, 317 near-surface drifters were released to directly measure transport and dispersion processes down to spatial scales as small as 100 m. More details about the GLAD experiment can be found in the articles of this special issue.

* Corresponding author.
E-mail address: Mozheng.Wei@nrlssc.navy.mil (M. Wei).

0967-0645/$ - see front matter Published by Elsevier Ltd.
http://dx.doi.org/10.1016/j.dsr2.2013.09.002

As the modeling team for CARTHE, here at NRL we are focusing on the numerical modeling, data assimilation (DA) and forecasting to support and provide numerical guidance to the GLAD experiment. To accomplish this mission, we have successfully run two RELO (Relocatable Circulation Prediction System) ensembles, each with 32 members, and three single-model deterministic forecasts using the Navy Coastal Ocean Model (NCOM Martin, 2000; Barron et al., 2006) and the Hybrid Coordinate Ocean Model (HYCOM, http://www.hycom.org, Bleck (2002), Chassignet et al. (2003), Halliwell (2004)) with different resolutions. All of these five ocean forecast systems were run in real-time, assimilating routine in situ and satellite observations processed at the U.S. Naval Oceanographic Office (NAVOCEANO), located at Stennis Space Center, MS.

To prepare for these important numerical and in situ experiments, all forecast experiments started from May 16, 2012 to initialize the ocean models and test the support software needed to distribute real-time forecasts. The forecast outputs are post-processed, archived and made available on NRL web servers for the scientists and students in CARTHE. Raw forecast data are also available on request by the CARTHE participants. Every day, real-time model outputs were first post-processed and evaluated by the scientists at NRL. Any important information and findings were provided to the CARTHE scientists in other organizations via email or regular tele-conferences. The implementation and operation of these forecast systems were conducted smoothly without significant delays in delivery. These products provided real-time guidance to the GLAD drifter deployment. One of the goals of this paper is to describe the details of these numerical forecast systems and the corresponding products including the RELO ensembles, and particularly the calibrated ensemble which is proposed for this mission. After the CARTHE GLAD data are released to the public in the future, it is expected that other scientists will use the GLAD drifter data set and our real-time forecasts to do further research. The material presented in this paper will be valuable for scientists from both inside and outside the CARTHE project.

In recent years, advances have been made in ensemble forecasting in both numerical weather prediction (NWP) and ocean prediction. In fact, ensemble products have become essential components of the daily operational products at all major NWP centers. The ensemble method uses a sample of numerical forecasts that represents our best knowledge about the possible evolution of a dynamical system. Ideally an ensemble forecast system (EFS) should include forecast uncertainties related to both the initial analysis values and the numerical model. The first generation of EFSs was implemented at the major meteorological centers about 20 years ago, and the details have been described in many publications, e.g. Toth and Kalnay (1993), Houtekamer et al. (1996), Molteni et al. (1996). These EFSs have been improved regularly over the past years in areas such as initial perturbation generation methods, methods for representing model related uncertainties, and computational efficiency. The performance of different ensemble methods or systems has been studied and compared in the literature, e.g. Wei and Toth (2003), Buizza et al. (2005), Bowler (2006), Wei et al. (2006). A more recent summary and discussion of newer ensemble methods can be found in Wei et al. (2008). The basic properties, advantages and disadvantages of different methods are summarized in Tables 1 and 2 of Wei et al. (2008).

Ensemble prediction method has also been frequently applied to ocean modeling (Evensen, 1994). A real-time realistic ensemble ocean prediction with data assimilation and adaptive sampling was completed with CMRE in 1996 (Lermusiaux, 1999). Further work using the ensemble technique for ocean forecast and uncertainty estimation can be found in Lermusiaux et al. (2006a, 2006b). Yin and Oey (2007) employed 20 members using the breeding method to study an eddy shedding event in the Gulf of Mexico. The authors successfully estimated the locations and strengths of the Loop Current and ring for July to September 2005. Counillon and Bertino (2009) studied eddy shedding and meso scale dynamics in the GOM by using a 10-member ensemble based on HYCOM with 5 km resolution. They used different values of a parameter in the optimal interpolation DA to generate the initial perturbations, while the atmospheric and lateral boundary conditions are perturbed randomly. The Loop Current and eddy fronts from observations were successfully predicted by their ensemble forecast, although the ensemble spread is two to three times smaller than forecast error. An ocean ensemble prediction system using breeding method to perturb all the model variables, and using the observations from operational OceanMAPS forecasting system has been developed by O'Kane et al. (2011) at the Australian Bureau of Meteorology.

At NRL Stennis Space Center, the RELO ensemble forecast system has been developed to provide a capability for a rapidly relocatable ocean ensemble forecast and data assimilation system for use in operational forecast support for the U.S. Navy's missions (Rowley, 2008, 2010; Rowley et al., 2012; Wei et al., 2013). A schematic showing the configuration of the RELO system with 32 ensemble members as used in this paper is presented in Fig. 1 of Wei et al. (2013). The forecast component in RELO ensemble is NCOM, (Martin, 2000; Barron et al., 2006). NCOM is a primitive-equation ocean model developed at NRL for local, regional, and global forecasting of temperature, salinity, sound speed, and currents. The data assimilation component is the Navy Coupled Ocean Data Assimilation System (NCODA; Cummings, 2005), which is based on a 3D-Var formulation. The observational data used for assimilation include satellite sea surface temperature (SST), satellite sea surface height anomaly (SSHA, or altimetry), satellite microwave-derived sea ice concentration, and in situ surface and profile data from ships, drifters, fixed buoys, profiling floats, XBTs, CTDs, and gliders. Both NCOM and NCODA are used operationally at two U.S. Navy operational centers, namely the Fleet Numerical Meteorology and Oceanography Center (FNMOC), located in Monterey, CA, and the U.S. Naval Oceanographic Office (NAVOCEANO), located at Stennis Space Center, MS.

It is known that an ideal EFS should have a spread that has amplitude comparable to the ensemble mean error and grow at a similar rate (Buizza et al., 2005; Wei et al., 2006, 2008). Since the method presently used in NCODA to estimate model forecast error underestimates the analysis error used to generate the initial perturbations through the Ensemble Transform (ET) process, the initial perturbations are small and cannot match the real analysis error variance. As a result, the RELO ensemble is under-dispersive, and the spread is smaller than the ensemble mean error. Although estimating analysis error in a 3D-Var based DA system such as NCODA is challenging (Lermusiaux et al., 2000; Lermusiaux, 2002), the Lanczos method with proper calibration can be used to produce reasonably good analysis error variance with extra computational cost (Wei et al., 2012). Another simpler poor-man's method is to use multi-analysis data from different DA systems or operational centers as demonstrated in Wei et al. (2010).

Another logical step to enhance the ensemble spread is to account for the model related uncertainties in the RELO ensemble. To achieve this, Wei et al. (2013) proposed and examined three different schemes for perturbing the horizontal and vertical mixing parameters. It is found that the RELO ensemble spread is enhanced to a limited extent. The results show that the scheme with perturbing both horizontal and vertical mixing parameters based on Gaussian distribution produces the largest spread increment. This scheme will be used in this paper. However, results also show that the RELO ensemble is still under-dispersive, even with this parameter perturbation scheme. To further improve the RELO ensemble for the CARTHE GLAD experiment, we propose

a calibration to enhance the initial spread inside the ET based on previous estimates of the ensemble spread and the forecast error. The superiority of the calibrated ensemble will be explored and demonstrated. Some other methods, such as stochastic forcing (Lermusiaux, 2006), can potentially enhance the spread and will be explored in the future.

We will compare the two RELO ensembles with three single-model deterministic forecasts with different resolutions. The impact of model resolution will be examined. In addition, we will apply the RELO ensembles to Lagrangian dynamics and prediction, including particle trajectory prediction and the Lagrangian Coherent Structure (LCS). The advantages of using the ensemble, especially the calibrated, more reliable ensemble, will be demonstrated in all of these cases.

Section 2 provides brief descriptions of the ET formulation for initial perturbations, the time-deformation technique to generate surface forcing perturbations from an atmospheric model for the RELO ensembles, the methodology for perturbing the mixing parameters, the configurations for the RELO ensembles, NCOM, and HYCOM, and the experimental setup. The major results from two RELO ensembles and three single-model forecasts with different resolutions in terms of various metrics for accuracy, reliability and forecast skill are presented in Section 3. Also shown in Section 3 are the results from applications of the ensembles to Lagrangian dynamics and particle trajectory prediction. A discussion and conclusions are presented in Section 4.

2. Methodologies, RELO ensembles, NCOM and HYCOM, and experimental setup

Before we provide our main experimental results in Section 3, we briefly describe the methodologies used in our RELO ensemble and the calibrated ensemble. Brief descriptions of NCOM and HYCOM with different resolutions for generating single deterministic forecasts are given. The RELO configuration and the experiment design will also be described.

2.1. Initial perturbations for RELO ensemble

The NRL RELO ensemble prediction system uses the Ensemble Transform (ET) method, which transfers forecast perturbations from the previous cycle into new perturbations that have the estimated initial analysis error variance. It is then followed by a rescaling using the same initial analysis error variance information. The details and properties of the method are described in Wei et al. (2005, 2008, 2013) and McLay et al. (2007). Only a very brief description is provided here. Let

$$\mathbf{Z}^f = \frac{1}{\sqrt{k-1}}[\mathbf{z}_1^f, \mathbf{z}_2^f, \ldots, \mathbf{z}_k^f], \quad \mathbf{Z}^a = \frac{1}{\sqrt{k-1}}[\mathbf{z}_1^a, \mathbf{z}_2^a, \ldots, \mathbf{z}_k^a], \tag{1}$$

where the n dimensional state vectors $\mathbf{z}_i^f = \mathbf{x}_i^f - \mathbf{x}^f$ and $\mathbf{z}_i^a = \mathbf{x}_i^a - \mathbf{x}^a$ $(i = 1, 2, \ldots k)$ are k ensemble forecast and analysis perturbations for all model variables, respectively. Here $\mathbf{x}^f$ is the mean of k ensemble forecasts from NCOM, and $\mathbf{x}^a$ is the analysis from the independent NCODA DA system. Unless stated otherwise, the lower and upper case bold letters indicate vectors and matrices, respectively. In the ensemble representation, the $n \times n$ forecast and analysis covariance matrices are approximated, respectively, as

$$\mathbf{P}^f = \mathbf{Z}^f \mathbf{Z}^{fT} \text{ and } \mathbf{P}^a = \mathbf{Z}^a \mathbf{Z}^{aT}, \tag{2}$$

where superscript T indicates the matrix transpose. For a given set of forecast perturbations $\mathbf{Z}^f$ at time t, the analysis perturbations $\mathbf{Z}^a$ are obtained through an ensemble transformation $\mathbf{T}$ such that

$$\mathbf{Z}^a = \mathbf{Z}^f \mathbf{T} \tag{3}$$

In RELO, the best estimate of the analysis error variance is derived from NCODA. Suppose the diagonal matrix $\mathbf{P}^a$ is composed of the analysis error variances obtained from the operational NCODA system. The ET transformation matrix $\mathbf{T}$ can be constructed as follows. For an ensemble forecast system, the forecast perturbations $\mathbf{Z}^f$ can be generated by Eq. (1). One can solve the following eigenvalue problem.

$$\mathbf{Z}^{fT} \mathbf{P}^{a-1} \mathbf{Z}^f = C \mathbf{\Gamma} \mathbf{C}^{-1}, \tag{4}$$

where $\mathbf{C}$ contains column orthonormal eigenvectors $(\mathbf{c}_i)$ of $\mathbf{Z}^{fT} \mathbf{P}^{a-1} \mathbf{Z}^f$ (also the singular vectors of $\mathbf{P}^{a-1/2} \mathbf{Z}^f$), and $\mathbf{\Gamma}$ is a diagonal matrix containing the associated eigenvalues (λ_i) with magnitude in decreasing order. The ET transformation matrix can be constructed as $\mathbf{T} = \mathbf{C}\mathbf{G}^{-1/2}$, where $\mathbf{G} = diag(\lambda_1, \lambda_2, \ldots, \lambda_{k-1}, \alpha)$ and α is a non-zero constant. According to Eq. (3), the analysis perturbations are given by $\mathbf{Z}_p^a = \mathbf{Z}^f \mathbf{C}\mathbf{G}^{-1/2}$. It can be shown that these perturbations are not centered. The final new analysis perturbations with simplex transformation imposed can be constructed through transformation

$$\mathbf{Z}^a = \mathbf{Z}_p^a \mathbf{C}^T = \mathbf{Z}^f \mathbf{C}\mathbf{G}^{-1/2} \mathbf{C}^T \tag{5}$$

It can be shown that the new analysis perturbations in Eq. (5) are centered. In addition, this method has the advantage that the ensemble perturbations span a subspace that has a maximum number of degrees of freedom. Wei et al. (2008) also showed that the orthogonality of the initial perturbations will increase as the number of ensemble members increases. If the number of ensemble members approaches infinity, then the transformed perturbations will be orthogonal under the inverse of the analysis error variance norm. In addition to the flow dependent spatial structure, the covariance constructed from the initial perturbations is approximately consistent with the analysis covariance from the DA, if the number of ensemble members is large.

2.2. Surface forcing perturbations for RELO ensembles using time-deformation technique

The surface forcing perturbations for the RELO ensemble are generated from real-time meteorological model forecast fields obtained from FNMOC, which produces operational forecast fields using the Navy Operational Global Prediction Center System (NOGAPS, for global) and the Coupled Ocean Atmosphere Mesoscale Prediction System (COAMPS, for regional) forecast systems. The atmospheric model fields are used to produce surface forcing fields for the RELO and NCOM. The ocean surface forcing fields include wind stress, surface pressure, shortwave and longwave radiation, air temperature, and specific humidity. Throughout our experiments, COAMPS atmospheric data fields, which are available at a 3 h interval and updated every 12 h, are used to produce surface forcing for single and ensemble forecasts of the ocean.

For the RELO ensemble, perturbed surface forcing fields for different ensemble members are drawn from the single-model prepared forcing using a time-deformation with random shifting technique. A number of the completely independent random fields are generated every 24 h with a desired de-correlation length. For each ensemble member, forcing is prepared at the same 3 h interval, but with the values computed at shifted times, by linear interpolation of the forcing originally prepared for the single model forecast. The time shifts are defined using a set of independent random fields generated every 24 h with a defined spatial de-correlation, so that any interpolated field is not correlated with any other interpolated field 24 h away, and the atmospheric

forcing for each ensemble member will be independent. More detailed mathematical formulae are given in Wei et al. (2013).

2.3. Mixing parameter perturbations in RELO ensemble

The impacts of perturbing model parameters on the RELO ensemble spread, reliability, accuracy, and forecasting skill were investigated in Wei et al. (2013). In that study, two parameters that play critical roles in describing the horizontal and vertical mixing in NCOM (Martin, 2000; Barron et al., 2006) were chosen. The parameterization formulae in atmospheric and ocean models are in fact approximate representations of unresolved ocean mixing processes in terms of model variables at the resolved scales. The different perturbation schemes described in Wei et al. (2013) are just attempts to describe phenomena at scales smaller than those resolved by the model.

By default, the NCOM used in RELO ensemble uses the Smagorinsky formulation for horizontal mixing parameterization (Smagorinsky, 1963); the control run scaling uses the parameter $smag=0.1$ as the default value. The Smagorinsky scheme scales the rate of mixing according to the horizontal velocity shear and is considered more physically based than other available options. The eddy coefficients are isotropic and are independent of coordinate rotation. For vertical mixing parameterization, NCOM also has multiple options. The default choice is the Mellor-Yamada Level 2 turbulence closure scheme (MYL2, Mellor and Yamada, 1974; Mellor and Durbin, 1975). MYL2 is a simplified scheme that assumes there is an approximate local balance between shear production, buoyancy production, and dissipation in the turbulent kinetic energy (TKE) equation, allowing the TKE to be calculated algebraically from the mean vertical density and velocity gradients, and the turbulence length scale to be estimated by an empirical formula. Ocean forecasts produced at the Navy operational centers typically use the MYL2 formulation due to the overriding importance of efficiently using computational resources. The RELO experiments reported here use the default MYL2 for vertical mixing parameterization, with the ensemble experiments examining variations in the parameter $b1\_myl2$ that scales the TKE in the MYL2 scheme and affects the predicted depth of mixing. The default operational value $b1\_myl2=15.0$ is used in the control run. Advantages and disadvantages of MYL2 in comparison with the other options were discussed in Martin (2000).

Three different parameter perturbation schemes were tested in Wei et al. (2013). In this study, we use the scheme that produced the largest spread increment for the RELO ensemble, i.e., the scheme in which these two critical parameters in the horizontal and vertical mixing turbulence parameterization, *smag* and $b1\_myl2$, are perturbed using a Gaussian distribution. The mean and standard deviation of *smag* are chosen as 0.125 and 0.01875, while for $b1\_myl2$, the values are 17.5 and 0.625 respectively. With these selected values, 99.99% of the random values generated by Gaussian distributions for *smag* and $b1\_myl2$ will be in the range of

$$smag\_range = mean \pm 4std = [0.05, 0.2],$$

$$b1\_myl2\_range = mean \pm 4std = [15.0, 20.0].$$

Under these distributions, values of these two randomly generated parameters are expected to fall within reasonable ranges and allow NCOM to run smoothly. The RELO ensemble based on this choice of parameter perturbation is denoted by gom32r (or *r* in figures).

Because the technique employed in the NCODA DA system underestimates the initial analysis error, the initial perturbations generated by the ET method in RELO ensemble are small. With the introduction of parameter perturbations in gom32r to account for the model uncertainties from mixing parameterizations, the spread has been improved to some degrees. However, the initial ensemble spread in scheme gom32r is still smaller than the ensemble mean error as shown in Wei et al. (2013). In an effort to adjust the initial spread to be more representative of the ensemble mean error, we take an ad hoc approach to calibrate the initial spread magnitude based on the difference between the RMS error of ensemble mean and the spread. This kind of ad hoc approach has proven to be effective in operational ensemble systems at major NWP centers (Houtekamer et al., 1996; Buizza et al., 2005; Bowler et al., 2009; Wei et al., 2008). In this study, we configure another new RELO ensemble with a calibrated ensemble spread, in which the magnitudes of initial perturbations generated in the ET in gom32r are increased by 50% immediately. This system is denoted as gom32q (or *q* in figures). The results using different various verification metrics in later sections will demonstrate that this calibration process has enhanced the initial spread, forecast accuracy, skill and reliability in comparison with gom32r. One of the advantages of this ad hoc calibration is that the spatial structure of the initial perturbations is not altered.

2.4. RELO configuration, NCOM, HYCOM and experimental design

To prepare for the GLAD at-sea experiment, we have carried out a series of real-time ocean prediction experiments since May 16, 2012. These include the two RELO ensembles, gom32r and gom32q described in above section, each with 32 ensemble members, two single NCOMs at 3 km and 1 km resolutions, and HYCOM with 4 km resolution. At the time of writing, we have completed the real-time forecasts starting from September 17, 2012. Therefore, for validation we choose the forecast series from June 1 to September 17, 2012, totaling 109 days. The forecast length during the experiment is 72 h, with output every 6 h. The configuration of the RELO experiments, consisting of 32-member ensembles using the NCOM and NCODA DA system, is depicted in Fig. 1 of Wei et al. (2013). The future developments include a stochastic physics parameterization to account for more sources of model error, and using the ensemble to provide the background error covariance information for the NCODA analysis.

Both RELO ensembles *r* and *q*, and the 3 km NCOM single forecast have a horizontal domain that covers the GOM from 98 to 79°W and 18 to 31°N with model grid spacing 3 by 3 km. The grid dimensions are 640 and 481 in the longitude and latitude directions respectively. This single NCOM forecast is denoted as ncom3km (or 3k in figures). The number of vertical levels is 49, with 34 sigma levels in the upper ocean and *z*-levels starting from level 35 to the bottom of the sea. The advantages of this kind of hybrid sigma-*z* coordinate have been discussed in Martin (2000) and Barron et al. (2006). The vertical grid extends down to 5500 m.

Another single forecast with NCOM has the horizontal resolution of 1 by 1 km (denoted by ncom1km, or 1k in figures) covering the GOM from 97.95 to 80.25°W and 18.05 to 30.79°N, with grid dimensions 1800 and 1420 in the longitude and latitude directions respectively. The vertical coordinate and resolution are the same as ncom3km. The only difference between ncom1km and ncom3km is in the horizontal resolution. Tidal forcing (barotropic tidal height and transports at the open boundary, and tidal potential in the interior) is turned on for gom32r, gom32r, ncom3km and ncom1km.

The final single forecast is generated with the Gulf of Mexico HYbrid Coordinate Ocean Model (HYCOM). HYCOM is on a Mercator projection covering the region from 18°N to 32°N, and from 98 to 76.4°W. The horizontal grid resolution is 1/25°, ~4 km

Table 1
RELO ensembles and single ocean model descriptions.

	gom32r (r)	gom32q (q)	ncom3km (3k)	ncom1km (1km)	hycom4km (4k)
Model	NCOM(32-member)	NCOM(32-member)	NCOM	NCOM	HYCOM
Resolution	3 km 49 Hybrid σ-z levels	3 km 49 Hybrid σ-z levels	3 km 49 Hybrid σ-z levels	1 km 49 Hybrid σ-z levels	4 km 20 Hybrid levels
Tidal forcing	On	On	On	On	Off
Perturbations	*Analysis*: NCODA 3D-Var. *Initial perturbations*: ET, Surface forcing perturbs from COAMPS atmospheric fields based on time-deformation technique. *Model error*: Perturbing vertical and horizontal turbulence mixing parameters with Gaussian distribution.	All the components from gom32r+additional initial perturbation calibration	N/A	N/A	N/A

resolution (indicated by hycom4km, or 4k in figures). The model employs 20 hybrid vertical coordinate surfaces. The vertical coordinates can be isopycnals (density tracking, which is the best in the deep stratified ocean), levels of equal pressure (nearly fixed depths, which are the best used in the mixed layers) and unstratified ocean and sigma-levels (terrain-following, are the best choice in shallow water). HYCOM combines all three approaches by choosing the optimal distribution at every time step. The model makes a dynamically smooth transition between coordinate types by using the layered continuity equation. The model is nested in a climatology generated from a multi-year, climatologically-forced, 0.08° HYCOM Atlantic Ocean simulation. There is no tidal forcing turned on during this run. All single forecast models and two RELO ensembles together with their configurations are summarized in Table 1.

3. Results from the RELO ensembles, NCOM and HYCOM

3.1. RELO ensemble spread with the calibrated initial perturbations

It is known that ensemble mean and ensemble spread are the very basic attributes of an ensemble prediction system. In general, the ensemble mean outperforms the single deterministic forecast in terms of the root mean square (RMS) error and the absolute error. The ensemble spread is closely related to the range, reliability, and sharpness or resolution of the EFS (Wei and Toth, 2003; Wei et al., 2006, 2008, 2012, 2013). Before the RELO ensemble is compared with the single forecasts from NCOM and HYCOM, we concentrate on the comparisons between the ensemble r and the calibrated ensemble q in this section.

Since the GLAD at-sea experiment started from July 17, 2012, and lasted until August 3, 2012, we choose to show the different snapshots at 00UTC on July 17, 2012 to demonstrate the different ensemble properties. Fig. 1 shows the ensemble plumes at 00UTC July 17, 2012 originating on the surface from a location of 60 km off the Louisiana coast at (88.39 W, 26.74 N) which is the location of Deepwater Horizon (DWH) oil spill incident beginning April 20, 2010. From the top to bottom, the ensemble plumes for temperature, salinity, and the zonal and meridional velocity components u and v are shown respectively for ensembles r and q. The black dashed lines follow individual ensemble members, and thick dash-dotted curves indicate the bounds one standard deviation above and below the mean. The comparisons between the left and right columns show clearly the impact of the calibration on the spread. The spread has been increased for all the variables due to the calibration. The ensemble mean and median of the two ensembles are very similar in most cases except for salinity. It is a common practice that the ensemble mean is regularly used to predict forecast events, as it has been shown that the mean from a reliable ensemble performs better than a single deterministic forecast (Toth and Kalnay, 1993). The ensemble plumes at 1500 m also showed similar impact from the calibration, except that the ensemble spread is generally smaller than at the surface (not shown).

Little if any growth as a function of forecast lead time is evident in the ensemble spreads for temperature and meridional velocity at the surface, while the spreads for salinity and zonal velocity show clear growth. This result can be observed more clearly in Fig. 2, which shows the ensemble spread started from 00UTC July 17, 2012 as a function of the forecast lead time and depth for ensembles r and q. Because the assimilation data stream has very few observations in the deep water, NCODA has little new information to produce large visible changes to the analysis. Since the analysis error variance estimated from NCODA is small, the initial ensemble spread at these depths is small as well. There is little variation below 200 m for temperature and velocity, so we restrict the plots to the upper 200 m where spread variations are better resolved by the color range. For salinity, there is little variation below 60 m. For temperature, the largest spread is not at the surface but in a 50-m thick depth range centered at 50 m.

The velocity spreads are largest near the surface due to the atmospheric forcing perturbations introduced through the time-deformation technique. The atmospheric forcing perturbations on the surface propagate downward very slowly as the forecast lead time increases for most variables, though the spread for u tends to propagate deeper as a function of time than the other variables. The spread for u propagates to about 50 m from the surface in 48 h. The spread for salinity is also larger from the surface to 10 m, due to fresh water mixing near the surface.

To get a better picture of the horizontal distribution of ensemble spread and a direct comparison of these two ensembles, we show the initial ensemble spread distributions at 00UTC July 17, 2012 for T, S, u and v for both ensembles r and q in Fig. 3. As expected, the calibrated ensemble spreads q are larger than the spreads from ensemble r for all the variables. For temperature, the largest spread reflects ocean state variability around the Yucatan Current, the Florida Current, and advection around the Loop Current eddy. The ensemble also shows relatively high uncertainty at the surface south of the Mississippi River Delta, which is near the DWH site. There is a large uncertainty in the surface salinity near the Mississippi and Atchafalaya river outflows. The large fresh water inputs lead to low salinity values near the coasts of Louisiana and Mississippi; this area of mixing is where the largest surface salinity variations are located. The ensemble spread reflects large variations in surface velocity within 200 km of the Louisiana coast, and in regions near the Loop Current and Yucatan Current.

The spread comparisons between ensembles r and q shown in Figs. 1–3 are snapshots of the two ensembles at 00UTC July 17, 2012, from one particular location or vertical level. To obtain more

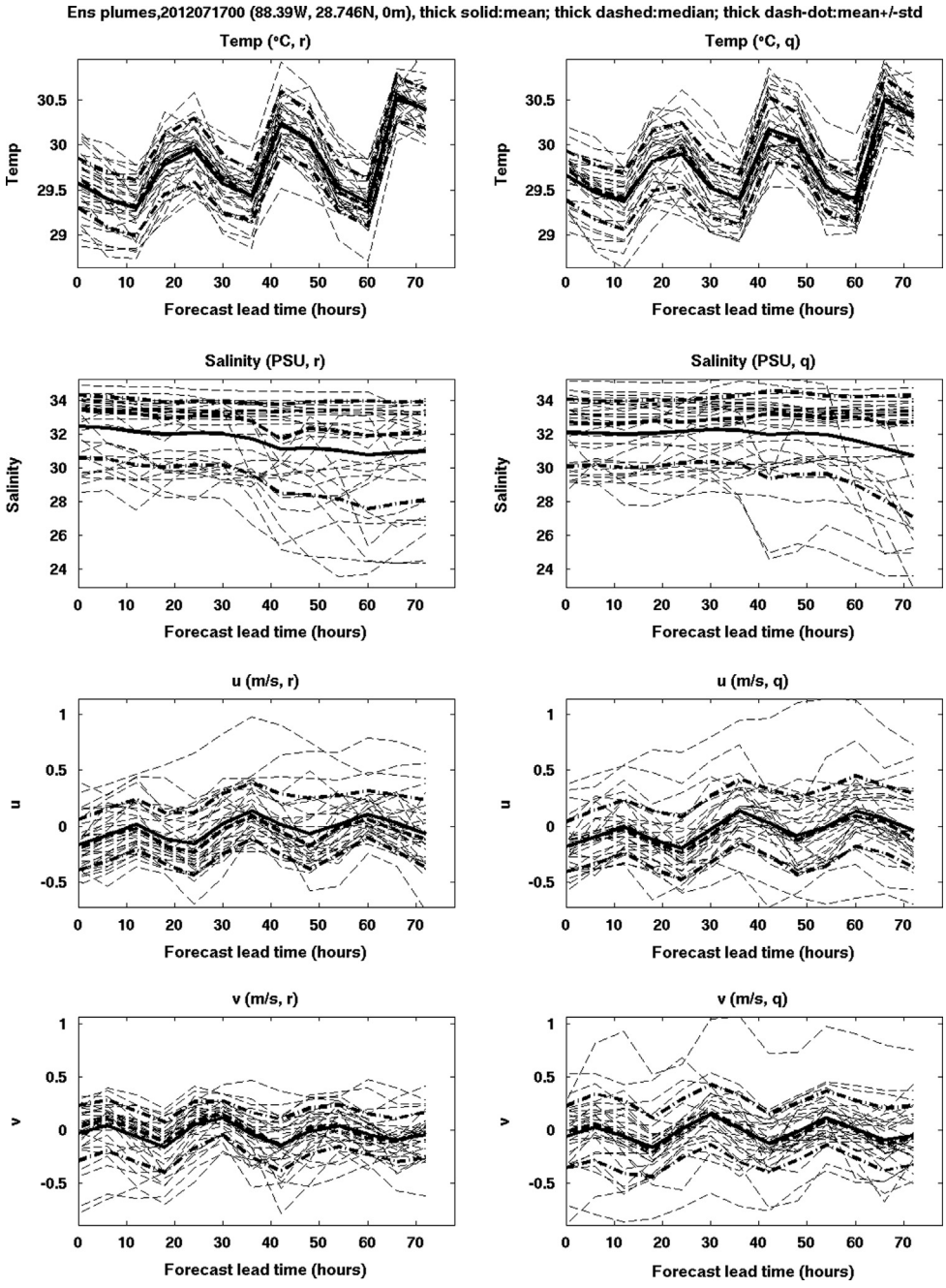

Fig. 1. Ensemble plumes from 00UTC July 17, 2012 at (88.39°W, 28.74°N, 0 m) for ensembles *r* (left) and *q* (right). Plots from top to bottom are temperature, salinity, *u*, and *v*. Dashed lines are the 32 individual ensemble members, the ensemble mean and median are indicated by thick and thick dashed lines. The thick dash-dotted lines indicate ensemble mean +/− one standard deviation.

statistically meaningful comparisons, in Fig. 4 we plot the ensemble spreads at 00UTC on each day during our experimental period from June 1 to September 17, 2012, totaling 109 cycles. The spreads of temperature and salinity for ensembles *r* and *q* are shown for forecast lead times of 24, 48 and 72 h respectively from top to bottom. In our computation of various verification metrics, we have interpolated all ensemble forecasts onto the observation locations. The routine in situ and remote sensing observations from the NAVOCEANO operational NCODA DA system are used, so numbers and locations of observations on each day vary. To have the best statistical meaning, all spread values are averaged over the full observation space. It is clear that the spreads for the calibrated ensemble *q* are consistently larger than those of ensemble *r* for both temperature and salinity at 24, 48 and 72 h forecast lead times. The results in this figure also show a large spread variability on different dates during this period for both variables, particularly salinity.

To look at the spread comparison from another perspective, we plot the temperature spreads for ensembles *r* and *q* as a function of forecast lead time in Fig. 5. All the spread values are averaged over the 109 days from June 1 to September 17, 2012. In addition, they are averaged over the full observation space (A), and for the layer from 0 to 100 m (B). In this study, our evaluations are carried out for different observation spaces, including the full observation space, near surface (upper 1 m) and the ocean interior over a range from 0 to 100 m. Two factors motivate the evaluations over

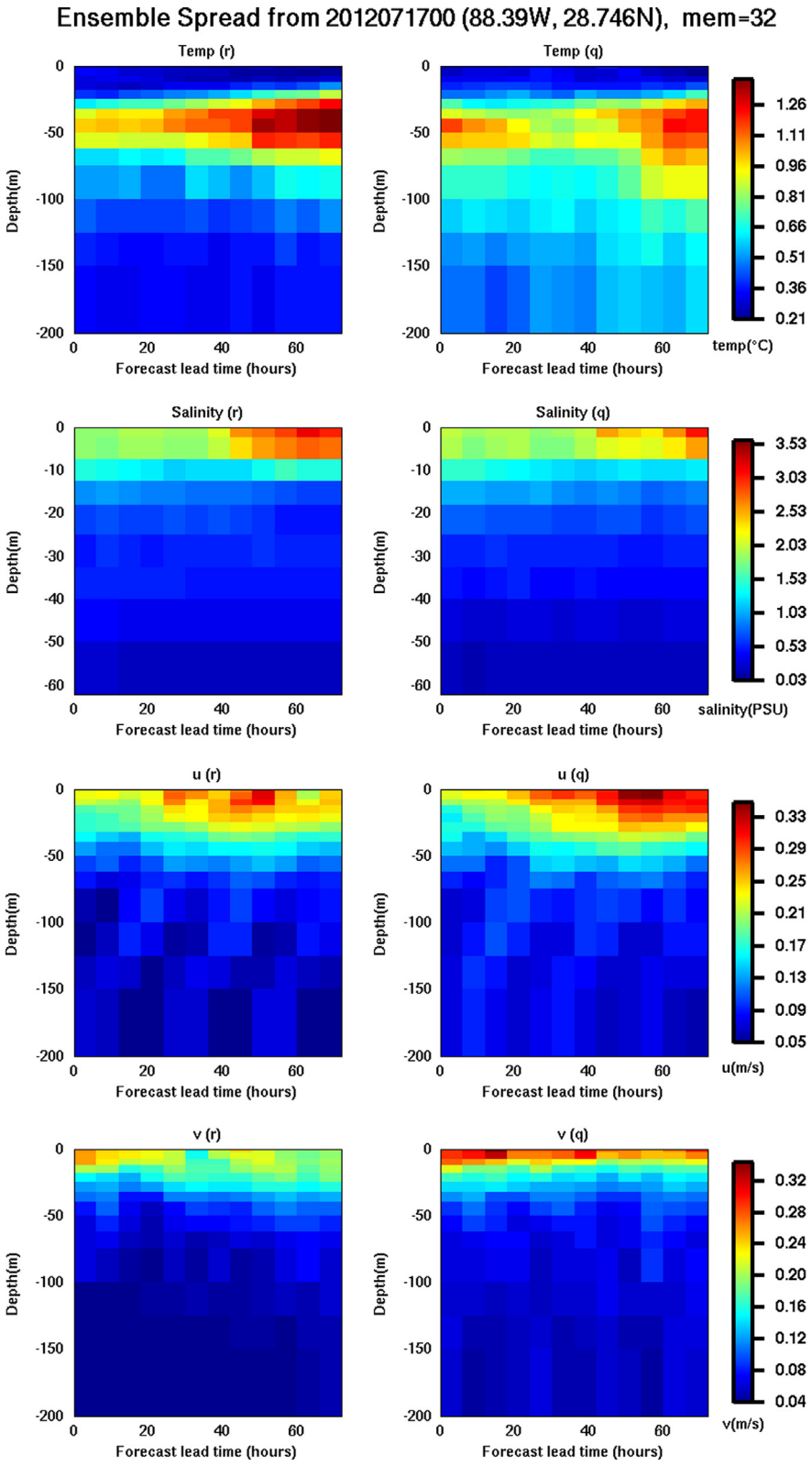

Fig. 2. Ensemble spread as a function of forecast lead time and water depth from 00UTC July 17, 2012 at (88.39°W, 28.74°N) for ensembles *r* (left) and *q* (right). From top to bottom the plots are for temperature, salinity, *u*, and *v*.

different observation spaces. First, these layers are dynamically distinct domains, with the surface dominated by air-sea interactions and highly variable wind-driven currents, and the interior controlled by mesoscale dynamics and internal mixing processes. Second, the density of observations is much larger near the surface than in the interior, so an un-weighted average over the entire domain is strongly skewed toward the near-surface. In both observation spaces, ensemble spreads from both systems grow slightly within 72 h, but the enhancement of spread from the calibrated ensemble *q* is evident.

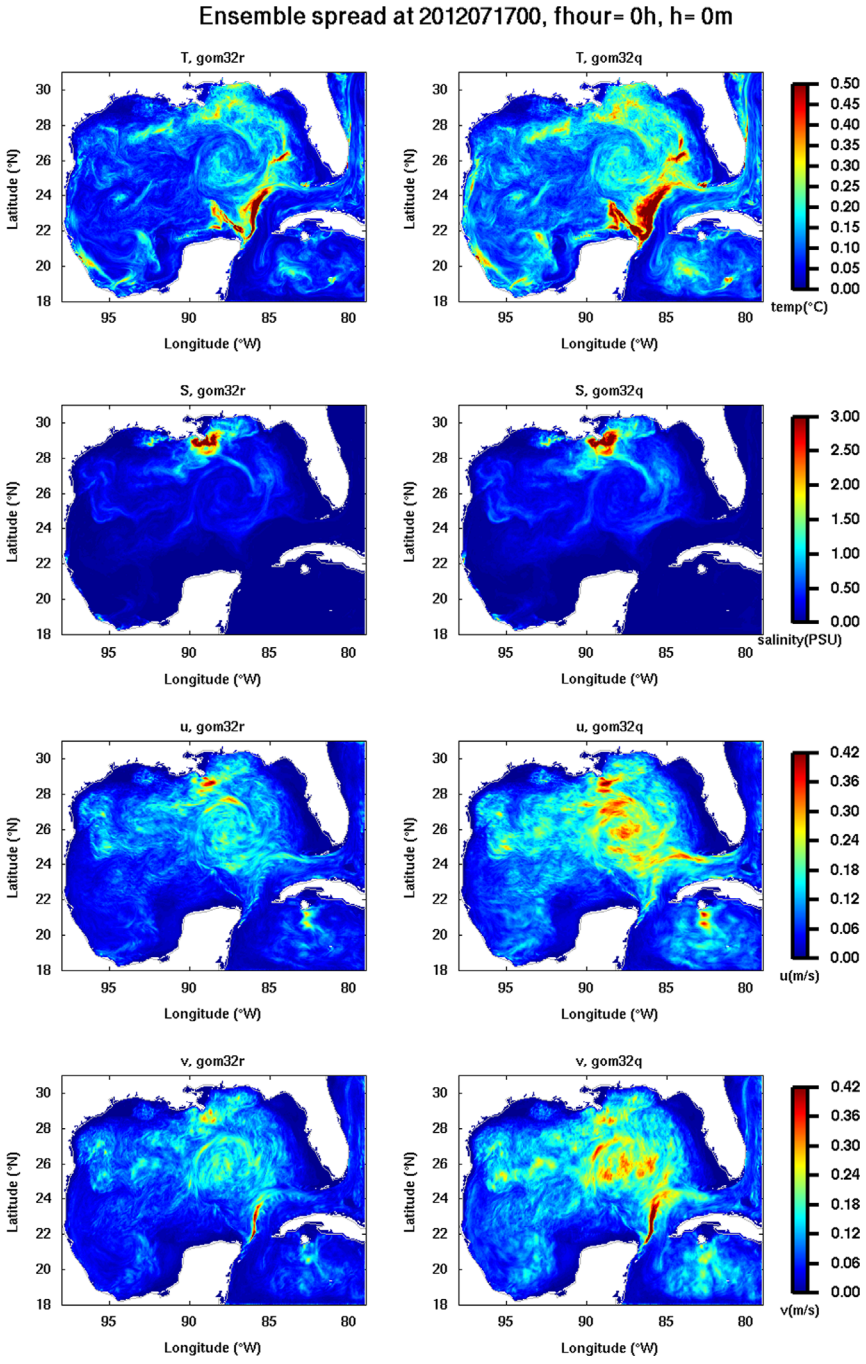

Fig. 3. Initial ensemble spread for r (left panel) and q (right panel) at 00UTC July 17, 2012 at the surface for temperature, salinity, u, and v (from top to bottom).

3.2. *Forecast accuracy and skill*

The RELO ensemble is developed to predict the ocean state with the ensemble mean and to predict the forecast uncertainty with the ensemble spread presented in the previous section. One of the most commonly used metrics to quantify the forecast error is the RMS difference between the ensemble mean and subsequent observations corresponding to the forecast time. It is expected that forecast accuracy generally decreases as the forecast lead time increases. This change in accuracy is represented as a growth rate in the RMS forecast error. In general the estimated uncertainty of a forecast is proportional to the ensemble spread, and it can be shown that an ideal ensemble should have an ensemble spread that has a similar magnitude and growth rate to the ensemble RMS error (Wei and Toth, 2003; Buizza et al., 2005; Wei et al., 2006, 2008). One of the main reasons for perturbing the mixing parameters in RELO is to account for model-related uncertainties and their contributions to ensemble spread growth (Wei et al., 2013). Without representations of model-related uncertainty, the ensemble will normally be under-dispersive and underestimate the true forecast uncertainty. If some important sources of uncertainties are neglected, the reliability of the forecast will be reduced.

To compare the forecast accuracy of the RELO ensembles with and without calibration, we compute the RMS error of the ensemble means from both ensembles r and q for temperature and salinity. In addition, the RMS errors of the single deterministic forecasts are also calculated for the same period of time and against the same operational observations as the ensembles, in order to have a fair comparison with both ensembles. In this study,

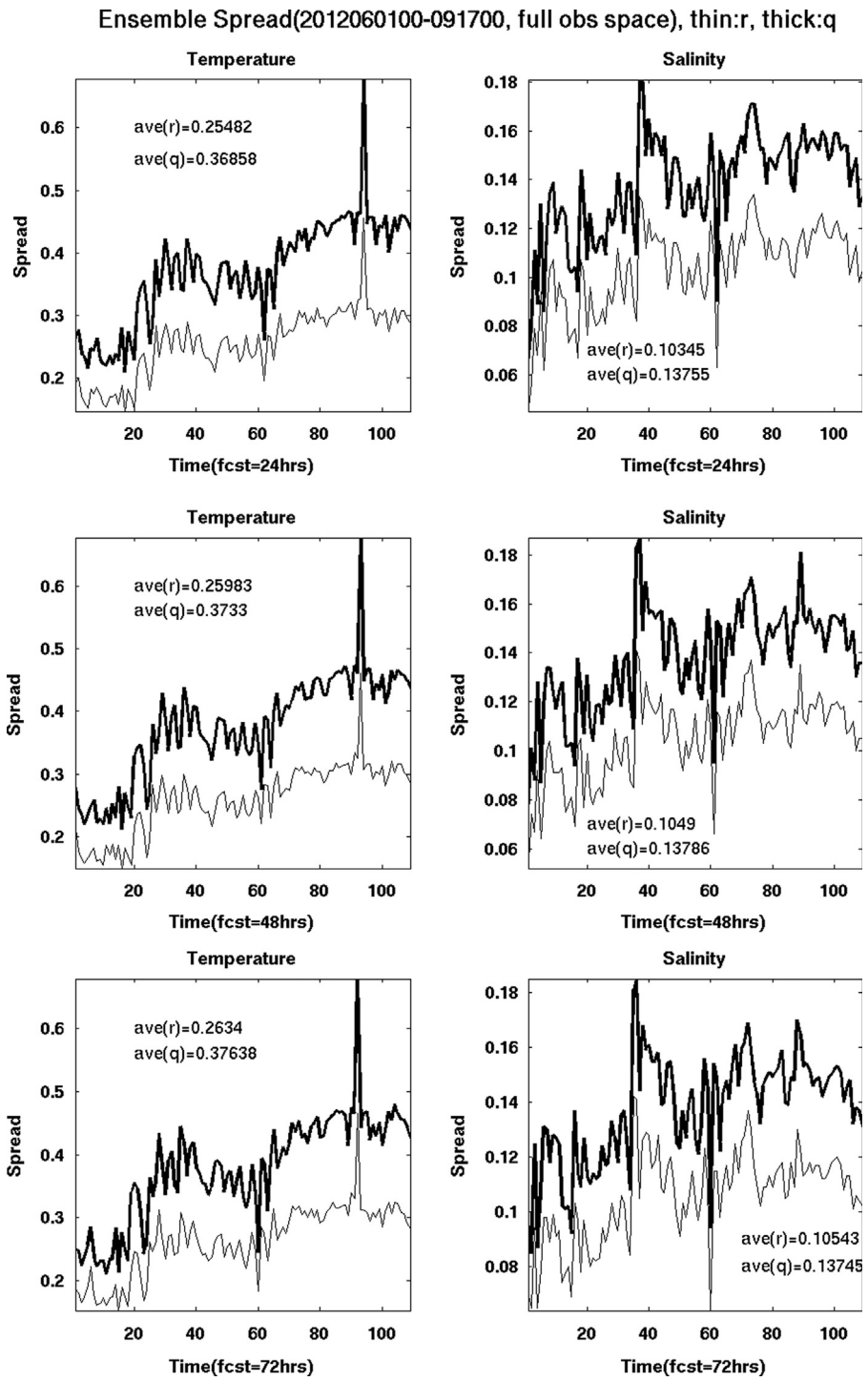

Fig. 4. Ensemble spreads for r (thin) and q (thick) as functions of day during the experimental period from June 1 to September 17, 2012. The spreads of T (left panel) and salinity (right panel) are shown for forecast lead times of 24, 48 and 72 h respectively from top to bottom panels. The spread values are averaged over the full observation space.

we have run three separate single deterministic forecasts for the same period of experiments as described in Section 2, namely ncom3km, ncom1km and hycom4km.

Fig. 6 shows the RMS errors of the two ensemble means and two single forecasts for temperature and salinity as a function of lead time. The RMS error is the difference between the model forecast and the truth represented by unassimilated observations valid during the forecast interval; thus it is a direct measure of forecast accuracy. The RMS values for ensembles r, q, and single forecasts ncom3km and ncom1km are indicated in different line styles respectively. All the RMS values are averaged over the 109 days from June 1 to September 17, 2012. The RMS values are averaged over the full observation space (top), and the layer between 0 and 100 m (bottom).

For temperature, ncom1km has the lowest RMS values. This is consistent with our expectations, as a higher resolution model often produces more accurate forecasts than lower resolution models. In the full observation space, the ensemble with

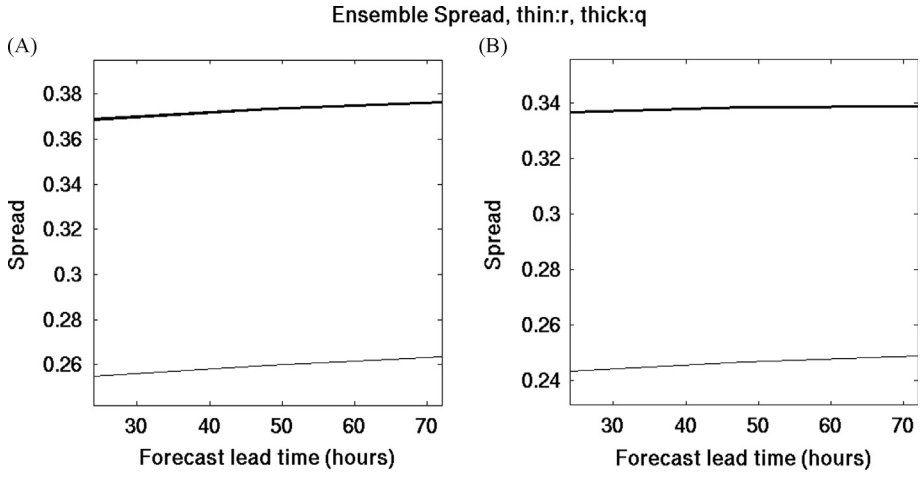

Fig. 5. Ensemble spreads of temperature for r (thin) and q (thick) as functions of forecast lead time. The spread values are averaged over the 109 days from June 1 to September 17, 2012, and over the full observation space (A), the layer between 0 and 100 m (B).

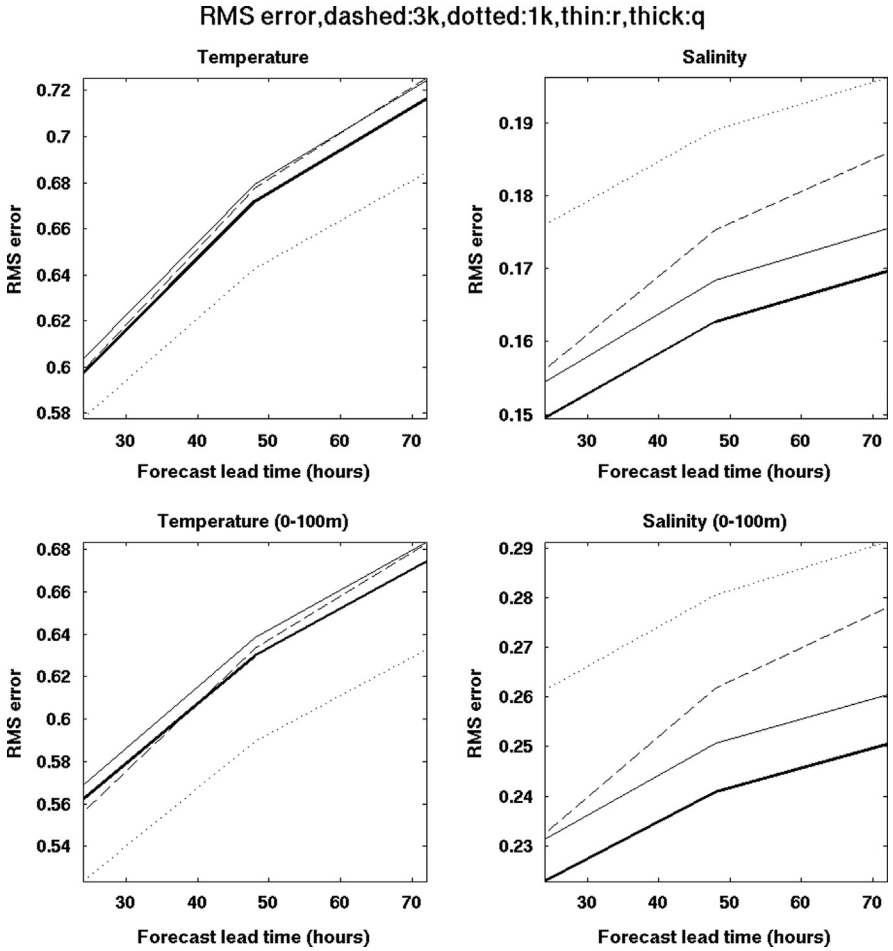

Fig. 6. The RMS errors of temperature (left panel) and salinity (right panel) for ensembles r (thin solid), q (thick solid), ncom3km (dashed), ncom1km (dotted) as functions of lead time. All the RMS values are averaged over the 109 days from June 1 to September 17, 2012, and averaged over the full observation space (top panel), and the layer between 0 and 100 m (bottom panel).

calibration (q) has lower RMS values than ensemble r, which is similar to ncom3km. In the observation space of 0–100 m, ensemble q also has lower RMS values than ensemble r. However, ncom3km has a slightly lower value at the 24 h forecast time. For salinity in both observation spaces, the calibrated ensemble mean (q) is the most accurate, followed by the un-calibrated ensemble (r), ncom3km and ncom1km. It is surprising to note that ncom3km has lower RMS values than ncom1km for salinity which will be explained later in this section. It is clear that the calibration introduced in the initial perturbations has improved the accuracy of ensemble mean for both temperature and salinity in both observation spaces verified. RMS errors for hycom4km

were also computed, the values are larger than all of these forecasts for all the forecast lead times in both observation spaces (not shown). This is expected as hycom4km has lowest resolution. The second factor is the tidal forcing that was not applied in the hycom4km experiment. The third factor may come from the interpolation in the observation space, since for verification purpose, all the forecasts are interpolated onto the observation space which is based on ncom3km experiment.

Similar to most operational ensemble systems in both meteoroand oceanography, the ensemble spreads are generally under-dispersive and grow slower than the ensemble mean errors, particularly in the ensembles without accounting for model-related errors (Wei and Toth, 2003; Buizza et al., 2005; Wei et al., 2006, 2008; Bowler et al., 2009; McLay et al., 2007; Reynolds et al., 2011). If we compare the RMS errors of temperature in both observation spaces (left) with the ensemble spreads in the same spaces in Fig. 5, even the enhanced ensemble spreads with calibration (q, thick) are still smaller than their corresponding RMS errors. The initially small ensemble spread is a consequence of the underestimation of the analysis error variance computed from the 3D-Var NCODA system. This results in a smaller initial ensemble spread during the initial perturbation generation process. At NRL, an effort is being made to make a better estimate of the analysis error from NCODA; we expect this will improve our future RELO ensembles. It is also clear that the ensemble spreads do not grow as fast as the ensemble mean RMS error. This indicates that just perturbing the two mixing parameters is not sufficient to account for most of the model-related uncertainties. Our plan is to introduce stochastic forcing at all the model grid points for all the model variables. We expect that when the model uncertainties are more completely accounted for, the ensemble spread will grow at a rate similar to the ensemble mean RMS error, and raise the RELO ensemble reliability to an even higher level.

Next, we assess the forecast skill of these two ensembles and compare them with the three single forecasts. To quantify the skill, we compute the anomaly correlation (AC) of the ensemble mean and the single forecasts. Again, the observations are used as the truth. The AC is preferred to a simple correlation coefficient (CC), which is defined as the correlation between the forecast and the observed values. The CC does not take forecast bias into account. It is possible for a forecast with large error to have a high CC value. It is well established practice to use the AC with climatology as a reference to account for seasonal variation (Wilks, 2006). The AC for any forecast variable f at a particular forecast lead time is defined as the correlation between the forecast and observation anomalies with respect to climatology, i.e.,

$$AC = \frac{\overline{(f-c)(y-c)}}{\sqrt{\overline{(f-c)^2}}\sqrt{\overline{(y-c)^2}}},$$

where c,y are the climate data and observation fields at the same verifying locations as the forecast, and the over-bar indicates the geographical mean over the verifying space. Therefore, the AC measures similarities in the pattern of departure (or anomalies) from the climatology field; it is a pattern correlation and regarded as a skill score relative to climatology. It is arguably the most commonly used metric in NWP centers (Buizza et al., 2005). The

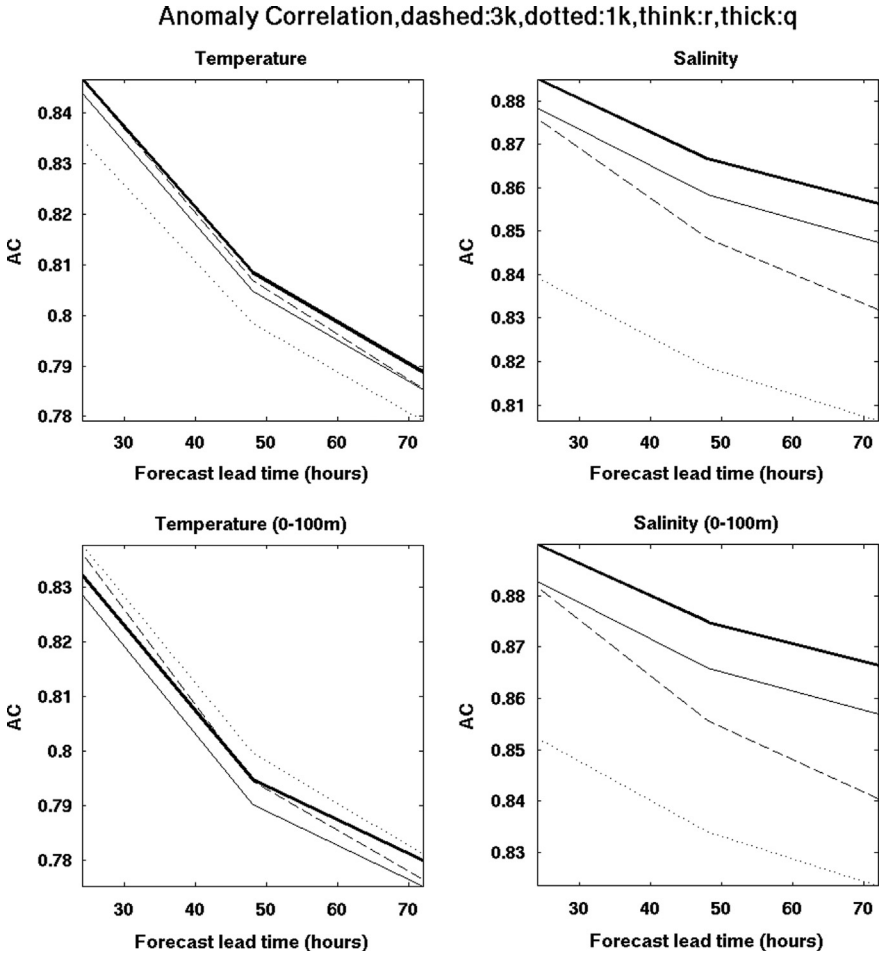

Fig. 7. The anomaly correlation of temperature (left panel) and salinity (right panel) for ensembles r (thin solid), q (thick solid), ncom3km (dashed), ncom1km (dotted) as functions of lead time. All AC values are averaged over the 109 days from June 1 to September 17, 2012, and averaged over the full observation space (top panel), and the layer between 0 and 100 m (bottom panel).

climatological data we used were obtained from the Navy's ocean operational center NAVOCEANO.

Shown in Fig. 7 are the AC values of the temperature and salinity averaged over the full observation space, the space between 0 and 100 m for the two ensembles and single forecasts from ncom3km and ncom1km. They are also averaged over the same 109-day period. In the full observation space for temperature, the calibrated ensemble *q* has highest skill score, while ensemble *r* and ncom3km have similar values. Higher resolution model ncom1km has lower AC value than ensemble *r* and ncom3km. In the observation space of 0–100 m, ncom1km has higher AC values than ensemble *q* and ncom3km, while the un-calibrated ensemble *r* has even lower AC value.

For salinity, the calibrated ensemble *q* is most skillful, followed by ensemble *r*, ncom3km and ncom1km, in both observation spaces. As in the accuracy comparisons displayed in Fig. 6, it is very clear that the calibrated ensemble *q* is more skillful than the un-calibrated ensemble *r* for both observation spaces and both variables. For both temperature and salinity in either observation space, hycom4km has the lowest AC values (not shown). This is understandable for the reasons we have explained in the paragraph about Fig. 6. A surprising result from Figs. 6 and 7 is that ncom1km performs poorer than ncom3km in terms of RMS error and AC for salinity. This may be due to errors introduced during the interpolation of forecasts onto ncom3km observation space for verifications. Since salinity observations are generally sparser than temperature observations, larger interpolation errors could be incurred from a much higher resolution ncom1km forecasts for salinity.

3.3. Ensemble reliability

As shown in the previous sections, the ensemble spread is an important attribute of an ensemble forecast system. The spread of a reliable ensemble forecast varies in space and time, and should capture the forecast errors as a function of the forecast lead time. Therefore, a reliable ensemble spread should have similar magnitude and growth rate to the ensemble mean error. If an ensemble spread is too small, it will miss important dynamic events, especially extreme ones. If an ensemble spread is too large, it will make the ensemble less sharp and less reliable with lower resolution. The results from the previous sections clearly show that the calibrated ensemble *q* is more accurate and skillful in all the observation spaces and for both variables we have evaluated. In this section, we compare the two ensemble spreads by using two other metrics which are especially designed for ensemble systems. The first one is the spread-reliability diagram. It is computed with 20 bins based on our 32-member ensembles, using observations as the truth. The exact steps for our RELO ensembles are outlined in Appendix A of Wei et al. (2013).

The top panel in Fig. 8 shows ensemble spread-reliability diagrams for salinity using observations as the truth for lead times of 24, 48, and 72 h in the full observation space. The same is shown for the observation space between 0 and 100 m (bottom). To have maximum statistical significance, all the values are averaged over a large number of samples within the respective observation spaces from 1 June to 19 September 2012. Since ensemble spread is expected to represent the forecast uncertainty, the spread-reliability curve over such a large sample should coincide with the diagonal line denoting equality between ensemble spreads and ensemble errors. The spread-reliability for salinity in Fig. 8 shows that the ensemble spread of *r* is small or under-dispersive for all the ranges for all forecast lead times on both observation spaces. Ensemble *r* is over-confident, and under-predicts the forecast error variance, which is consistent with the results in Figs. 5 and 6.

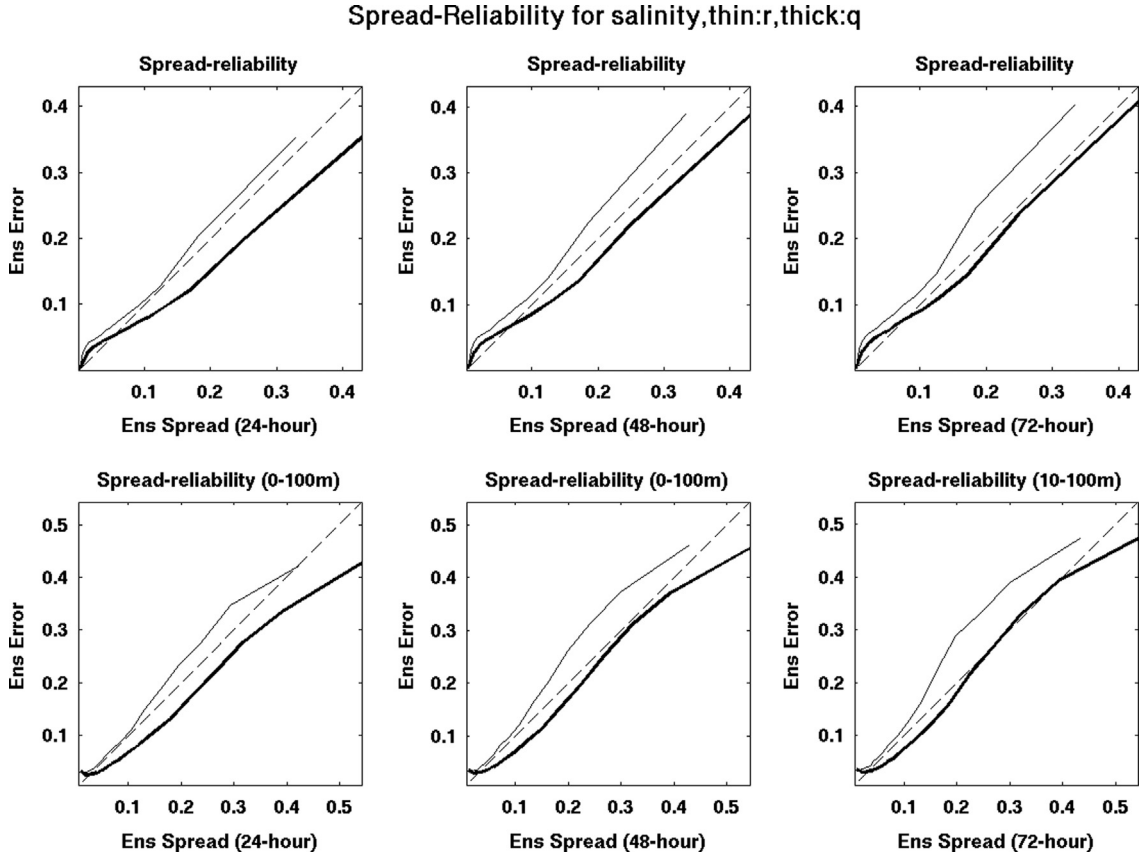

Fig. 8. Ensemble forecast spread-reliability diagrams for salinity (*r*: thin, *q*: thick) using observation as truth for lead times of 24, 48, and 72 h (from left to right). Values are averaged over the 109 days period from June 1 to September 17, 2012, and over the full observation space (top panel) and the layer between 0 and 100 m (bottom panel).

The spread-reliability curve for the calibrated ensemble q is closer to the diagonal line for all forecast lead times and both observation spaces, except for the 24-h forecast in the full observation space. In this case, ensemble q is slightly over-dispersive, and over-predicts the forecast error variance. The spread-reliability curves for temperature at the same three forecast lead times have been computed for the full observation space, for 0–100 m, and the near-surface. The results show the calibrated ensemble q is closer to the diagonal line than ensemble r for all three forecast lead times and in all three observation spaces (not shown).

Another popular metric to diagnose the ensemble spread reliability and its consistency is the rank histogram or Talagrand histogram (Talagrand et al., 1997; Wilks, 2006). The procedures for computing this histogram for our RELO ensembles using observations as truth are described in detail in Appendix B of Wei et al. (2013). Here, we compute the rank histograms for both temperature and salinity at three forecast lead times, namely 24, 48, and 72 h, and in three domains, including the full observation space, the space between 0 and 100 m, and, for temperature only, the surface. Shown in Fig. 9 are the temperature rank histograms for ensembles r (white) and q (black) at lead times of 24, 48, and 72 h in the full observation space, the observation space between 0 and 100 m, and the surface.

One notices immediately that the temperature rank histograms for ensemble q are much flatter than for ensemble r for all three forecast lead times and for all three observation spaces. This is also reflected in the values of consistency index of the rank histogram of both ensembles. The consistency index of ensemble q is about 20–30% lower in comparison with ensemble r in each of these cases. The same is computed for salinity for the full

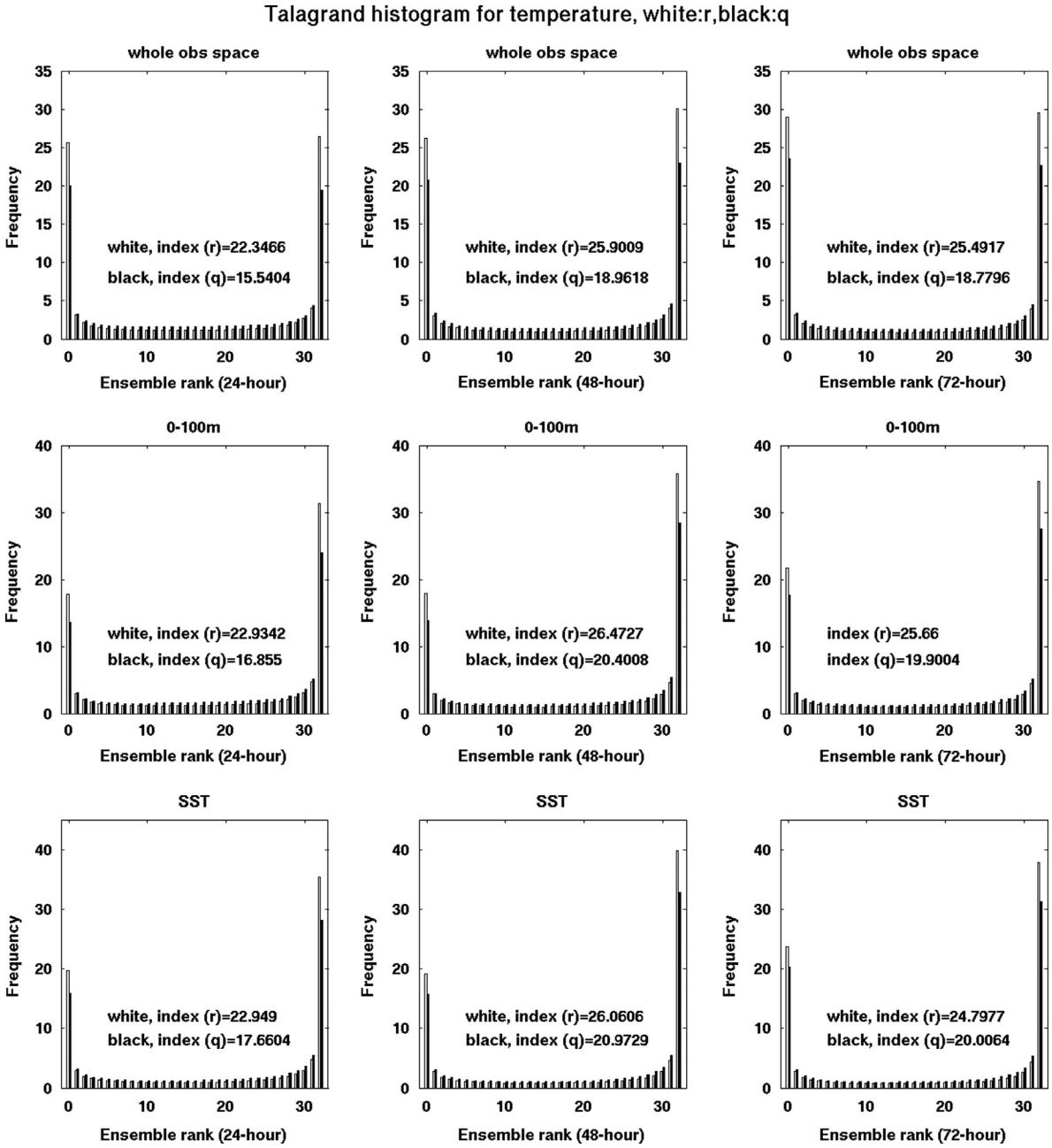

Fig. 9. Talagrand rank histograms for temperature (r: white, q: black) using observation as truth for lead times of 24, 48, and 72 h (from left to right). All the values are averaged over the 109 days period from June 1 to September 17, 2012, and over the full observation space (top panel), the layer between 0 and 100 m (middle panel), and the surface (bottom panel). Consistency index is also indicated in each case for both ensembles.

observation space and the observation space between 0 and 100 m (not shown). Again, the rank histograms for ensemble q are much flatter than ensemble r in all cases. The consistency index for ensemble q is about 50% lower than that of ensemble r, and is much closer to 1, which is the value for an ideal ensemble system, compared with those for temperature. In summary, the calibrated ensemble q is much more consistent than ensemble r.

3.4. Ensembles in Lagrangian prediction

During the CARTHE GLAD experiment, the NRL real-time numerical model runs including ncom3km, ncom1km, hycom4km and both RELO ensembles q and r provided valuable numerical guidance for the drifter deployment from July 17, 2012. So far, the results and findings discussed in all previous sections are based on the Eulerian formulation, in which the ocean states are described at fixed grid points. In this section, we describe the results from the application of RELO ensembles to the Lagrangian dynamics and prediction, i.e., the results from the ensembles will be based on following particle trajectories.

A basic application of numerical model output to the Lagrangian dynamics is to use the forecast velocity fields to predict the particle or drifter trajectories on the water surface. Earlier work on the prediction and predictability of drifter trajectories in ocean simulations includes Özgökmen et al. (2000, 2001). This is a significant contribution ocean models make to the disaster response community, for search and rescue, and for contaminant monitoring and mitigation, such as that in the aftermath of DWH oil spill. Many examples of using the numerical model forecast velocities to predict particle trajectories are available in the literature. The value of trajectory predictions by ocean models have been demonstrated particularly well after the 2010 DWH incident (Maltrud et al., 2010; Huntley et al., 2011b; Mariano et al., 2011). A special collection of papers about Lagrangian trajectory predictions with state-of-the-art ocean models entirely devoted to the 2010 DWH oil spill has been published in Liu et al. (2011). Most of these studies are based on single model forecasts; only a few use with ensembles. However, those ensemble results are actually composed of different forecasts from different models or organizations, and they are not produced from true ensemble prediction systems based on dynamics. As a result, the uncertainties of those predicted trajectories are hard to assess. In this paper we will concentrate on the application of RELO ensemble forecasts to Lagrangian prediction. It is well known that an ensemble forecast system provides not only more accurate ensemble mean to describe the ocean state, but also valuable uncertainty information which is not available from traditional single deterministic forecasts. For simplicity, in the following we advect particles using a fourth-order Runge–Kutta integration with 2-dimensional interpolated forecast surface velocity on the ocean surface in the GOM. There is no attempt to account for diffusive processes or subgrid-scale uncertainties.

One example of the extra information that can be provided only by an ensemble system is shown in Fig. 10. We choose seven particles to represent drifters or surface oil patches from the DWH disaster indicated by A, B, C, D, E, F, G on the surface of GOM. From the water currents generated from any ocean prediction model such as HYCOM or NCOM, one can forecast the particle trajectories. To demonstrate the importance of uncertainty information provided by the calibrated ensemble, particles F and G are placed near hyperbolic locations. For each of these particles, we integrate the velocity fields provided by each ensemble member to obtain a particle trajectory. Thus, there are 32 different possible trajectories from each location from one ensemble forecast. The trajectories of these particles from ensemble r are shown in the top panel of Fig. 10. The corresponding trajectories generated by ensemble q are displayed in the bottom panel of Fig. 10. Also plotted in Fig. 10 are the sea surface height (SSH) in colored contours, and the surface current velocity indicated by arrows. The correlation between SSH and velocity is clear. In addition, the particle trajectories tend to follow the directions of velocity as expected.

We also plot the predicted trajectory for each particle based on the ensemble mean in thick red curve in this figure. If ensemble spread distribution is close to Gaussian, then ensemble mean and mode will be similar. Fig. 1 shows that this is the case for both u and v. Therefore, the trajectory based on ensemble mean is almost the same as the one based on mode with the highest probability, i.e. the most likely scenario for the particle to follow. Traditionally, the trajectories of these particles are generated by using a single model such as ncom3km, ncom1km or hycom4km. Each model will generate just one trajectory (one possible outcome) for each particle. Due to the uncertainties of initial conditions, and the chaotic nature of ocean dynamical system that is highly nonlinear, there should be a range of possibilities of trajectories that each particle might follow. In contrast to a single trajectory produced by single model for a particle, either ensemble r or ensemble q can provide 32 possible trajectories representing different possibilities. It is not difficult to see that the information provided by ensemble is more complete with both the most likely trajectory and uncertainties. This extra information will help users or decision makers to make more scientifically sound decisions.

As in previous sections, we are also very interested in the impact of the calibrated ensemble on the particle trajectories in

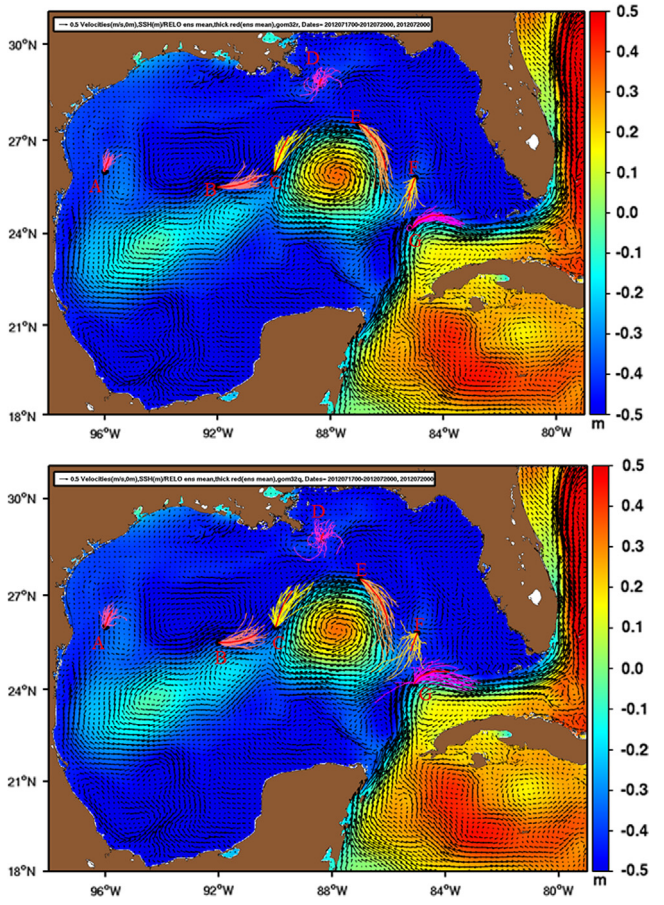

Fig. 10. The trajectories of seven particles (A–G) predicted by ensemble members from r (top panel) and q (bottom panel). The predicted trajectories by ensemble means are denoted by thick red curves. Particle D is chosen to be at the location of DWH accident. Superimposed are the SSH in color contour and surface water velocity indicated by arrows.

comparison with the un-calibrated ensemble. One clear difference between the trajectories predicted by these two ensembles is the spread of ensemble trajectories, which can be similarly defined as the RMS distance between these trajectories predicted by different ensemble members. For each of these seven particles, the spread of trajectories in ensemble q (bottom) is obviously larger. This means that ensemble q can capture a wider range of possibilities of future movement of the particle than ensemble r (top). In another word, when both ensembles are used to predict a particle trajectory, ensemble q could capture some possible different moving directions that might be missed by ensemble r.

To display this more clearly, we zoom in a smaller domain containing particles F and G. This is shown in Fig. 11. All 32 members of ensemble r predict particle F moving southward. However, among the 32 members of ensemble q, there is one member predicting that particle F will move toward the northwest, and another member predicting southeast movement. That is, ensemble q predicts 1/32 probability for particle F to move either northwest or southeast, but these probabilities are missed by the less reliable ensemble r. For particle G, all the members of ensemble r predict that the particle will follow the Loop Current toward the Florida Strait. However, ensemble q predicts a non-zero likelihood for particle G to move westward (2/32) or northward (2/32). These uncertainties are not picked up by ensemble r due to its lesser reliability as we discussed in previous sections. During our real-time ensemble runs for the CARTHE GLAD at-sea experiment, we have noticed other examples on other dates when the calibrated ensemble q captures greater uncertainty than ensemble r near those sensitive, hyperbolic locations.

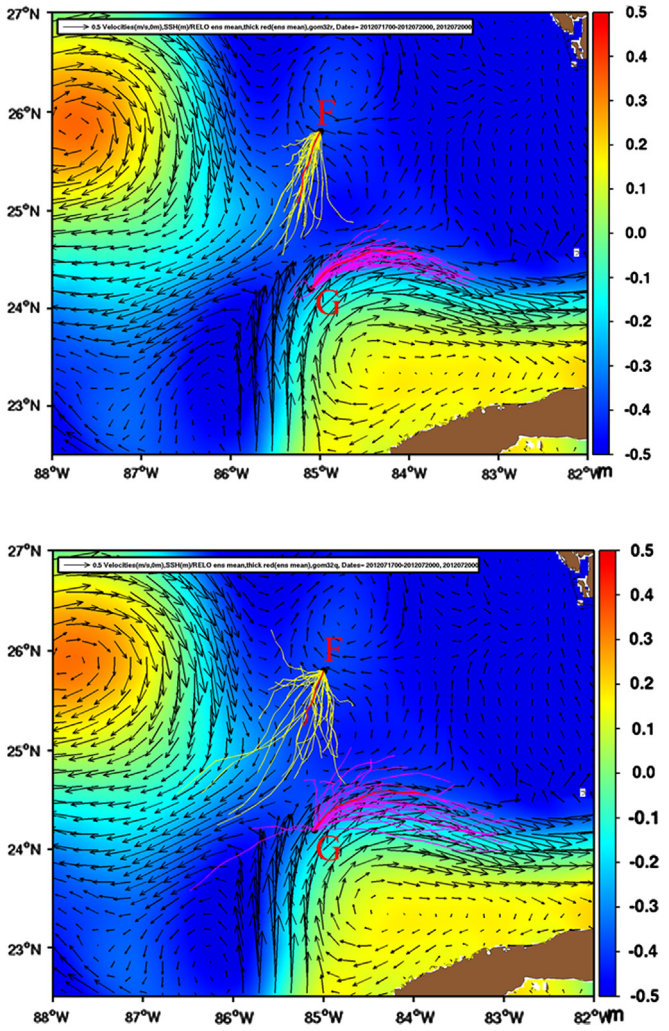

Fig. 11. As in Fig. 10, but for the smaller domain around particles F and G.

3.5. Ensembles in Lagrangian coherent structure

Another important area of ensemble application is the Lagrangian Coherent Structure (LCS). The ensemble can provide not only the most likely LCS, but also its associated uncertainties. There have been many applications of the LCS in computational fluid dynamics (Haller and Yuan, 2000; Shadden et al., 2005; Haller and Sapsis, 2011). This technique has been adopted in the ocean modeling community to study the tracer distribution and prediction (Lekien et al., 2005; Lermusiaux et al., 2006a, 2006b; Coulliette et al., 2007; Olascoaga et al., 2008; Beron-Vera et al., 2008; Shadden et al., 2009; Olascoaga, 2010; Huntley et al., 2011a, 2011b; Olascoaga and Haller, 2012). Conceptually, LCSs are the locally most strongly attracting or repelling material surfaces in the flow. They move with the flow and provide core surfaces organizing the advection of tracers. A common practice to identify the LCS is to use the finite-time Lyapunov exponent (FTLE); some slightly varying definitions can be found in the literature. In this paper, we adopt the FTLE to define the LCS following Shadden et al. (2005) and Haller and Sapsis (2011), in which a robust mathematical description and definition is given. The FTLE has proven to be an effective tool to identify LCSs, and is a measure of the finite-time averaged maximum separation rate of two initially close fluid particles. In this study, we concentrate on 2-dimensional surface flow of ocean. Suppose the ocean velocity field generated by NCOM or HYCOM is $\mathbf{v}(x,y,t)=(u(x,y,t),v(x,y,t))$. The dynamical equation is given by

$$\frac{d\mathbf{x}}{dt}=\mathbf{v}(x,y,t).$$

If we follow a particle at time t_0 to a later time t, the integration of the above equation will provide a flow map $F(t_0,t)$ which maps the particle at the initial position to the current position at time t, i.e. $\mathbf{x}(t)=F(t_0,t)\mathbf{x}(t_0)$. A matrix can be formed by using the gradients of the flow map as

$$C=\left(\frac{dF}{d\mathbf{x}}\right)^T\left(\frac{dF}{d\mathbf{x}}\right),$$

with the superscript T indicating matrix transformation. This symmetric matrix is called the right Cauchy–Green deformation tensor, and is the function of t_0, $\mathbf{x}_0$, t, and $\mathbf{x}$. The largest FTLE associated with this trajectory over the time interval $t-t_0$ is defined as

$$\sigma(\mathbf{x}_0,t_0,\mathbf{x},t)=\frac{1}{|t-t_0|}\log\sqrt{\lambda_{\max}(C)},$$

where $\lambda_{\max}(C)$ denotes the largest eigenvalue of C. Therefore, the FTLE is the time-averaged maximum exponential stretching about the trajectory from time t_0 to t. The ridges of the largest FTLE indicate the LCSs. There are two types of LCSs. The first kind is the repelling LCS, which is the material surface formed by the trajectories of the dynamical system that repel other trajectories at the locally highest rate for the time interval $t-t_0$. The second is the attracting LCS, which is the material surface that attracts nearby trajectories at the locally highest rate for the time interval $t-t_0$. A common way of computing the repelling LCS at time t_0 is to integrate a set of trajectories forward starting from an array of initial conditions up to a time t. The largest FTLE can be used to identify the repelling LCS, which is associated with the stable manifold. Another separate backward integration from time t to t_0 is needed to locate the attracting LCS, which is associated with the unstable manifold. More detail can be found out in Shadden et al. (2005) and Haller and Sapsis (2011).

The attracting LCS has been used in tracer prediction and pollutant dispersal modeling in Olascoaga et al. (2008), Olascoaga (2010) and Olascoaga and Haller (2012). Beron-Vera et al. (2008) computed attracting LCSs to identify mesoscale eddies. Repelling LCSs are material lines that act as moving barriers to transport. They have important impacts on the movements of particles and pollutants on ocean surface. Lekien et al. (2005) and Lermusiaux et al. (2006a) and Coulliette et al. (2007) have used repelling LCSs for optimizing pollution management and release in coastal oceans in California and Florida. The repelling LCS was also used by Shadden et al. (2009) to help drifter release in Monterey Bay. However, all of these studies on LCSs from numerical ocean models are based on the single deterministic forecasts. While identifying the LCSs in ocean is an important contribution, it is also important to understand the uncertainties related to these LCSs. With a single-model deterministic forecast, the uncertainties of the LCSs cannot be estimated.

We will use our real-time ensemble data generated for the CARTHE GLAD drafter deployment to identify the LCSs and their uncertainties. The LCSs from ensemble will be compared with

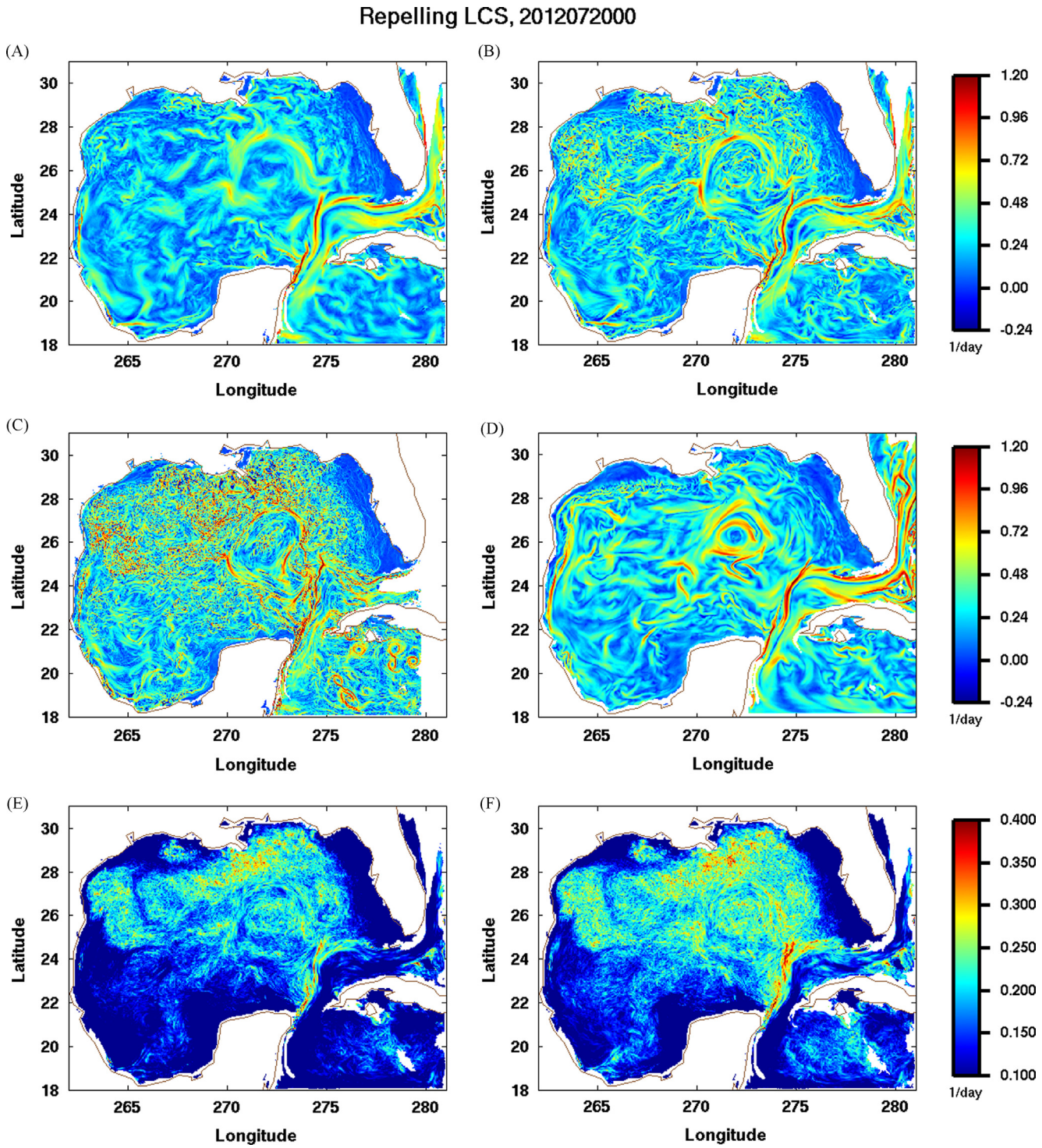

Fig. 12. The repelling LCSs (1/day) on the ocean surface over the GOM at 00UTC, July 20, 2012, generated by ensemble mean of gom32q (A), ncom3km (B), ncom1km (C), and hycom4km (D). The STD of LCSs from r and q are displayed in (E) and (F).

those from single ocean models. This also provides an opportunity to study the sensitivity of the LCS to the model and its resolution. As an initial step to apply ensembles to LCSs in ocean simulation, we consider only repelling LCSs in this study. Application to attracting LCSs will be explored in the near future.

Fig. 12 shows the repelling LCS on the ocean surface at 00UTC, July 20, 2012 which was the first date of deployment of GLAD Large Scale Spiral (LSS) drifters. There were total 20 drifters deployed to study the large scale dynamics of the GOM. The repelling LCSs are computed based on the 3-day forecasts of ensemble mean of gom32q (A), ncom3km (B), ncom1km (C), and hycom4km (D). Since there is little difference between the LCSs from the ensemble means of gom32r and gom32q in terms of magnitude and scale, only the LCS from the ensemble mean of gom32q is shown.

However, the advantage of using an ensemble lies in the fact that it can provide valuable uncertainty information about the LCS. The LCS of each ensemble member is computed for both ensembles *q* and *r*. Each LCS of a different ensemble member represents a possible realization of structure within this ensemble. The standard deviation (STD) of 32 LCSs based on the 32 individual members is a good estimate of the uncertainty of LCSs described by the ensemble. The STDs for gom32r and gom32q are shown in Fig. 12(E) and (F) respectively

One clear difference among these repelling LCSs from different models and the ensemble mean is the spatial scale due to different resolutions of the models being used. The LCS generated from ncom3km shows smaller scale structures than hycom4km which has lower resolution. Although gom32q uses NCOM with 3 by 3 km resolution, the LCS from the ensemble mean displays larger spatial structures than that from ncom3km. This is probably due to the filtering effect of ensemble mean, which removes some smaller scale features. The LCS from ncom1km in Fig. 12(C) reveals the smallest scale features of repelling LCS. This will make ncom1km an ideal candidate to study the sub-mesoscale eddy structures in the GOM. The large scale repelling LCS around the Loop Current is identified by gom32q, ncom3km and ncom1km, but not by hycom4km, which instead shows a smaller circular repelling LCS near the north of the Loop Current. All the models show a long, robust large-scale repelling LCS starting from the Yucatan Current, connecting the Loop, Florida, and the Gulf Stream Currents. Along this large-scale repelling LCS, there also exist large uncertainties as shown by the STD of LCSs (Fig. 12E and F). Overall, the STD of LCSs from either ensemble shows similar, relatively large uncertainties in similar regions, such as the northern GOM near the DWH location, and the areas around the Loop Current. However, the calibrated ensemble *q* produces more pronounced uncertainty structures at more locations than ensemble *r*.

Starting from July 20, 2012, 20 GLAD LSS drifters were deployed around the DWH location. It is interesting to see the LCSs and their related uncertainties near these drifters. In order to look at more detailed structures of LCSs in the areas of our interests and experiments at the same time, a much smaller domain is needed. Fig. 13 shows the same as Fig. 12, but with a smaller domain. Now, much more detailed LCS structures and uncertainties can be seen clearly from all four forecasts, particularly ncom1km, which displays well organized small scale repelling LCSs. Again, the detail structures of LCSs identified using numerical model output depend on the model resolutions.

Most of the drifters were more likely deployed in the troughs of LCSs; very few were on the ridges of LCSs. Since the repelling LCS acts as a transport barrier, if the drifter is on the ridge of the LCS, it could fall to either side of the LCS. Once the drifter is on one side of the LCS, it will be trapped on that side, as it is almost impossible for a drifter to cross the LCS barriers. Therefore, the movement of a drifter depends greatly on how repelling LCSs change with time, and the drifter movement is generally constrained by the repelling LCSs. Accurately identified repelling LCSs will provide helpful guidance to drifter deployment, and to forecasting the trajectories of drifters (Lekien et al., 2005; Lermusiaux et al., 2006a; Coulliette et al., 2007; Shadden et al., 2009).

As we discussed, this is only the initial step in applying the ensembles to repelling LCS studies. Our next step will be to compare the predicted drifter trajectories, the identified LCSs and their associated uncertainties against the observed drifter trajectories from the CARTHE GLAD data set. It is expected that the full advantages of the ensembles will be demonstrated with this valuable source of drifter data from the CARTHE GLAD. We will report those results when they are available.

4. Discussion and conclusions

As the designated modeling team within CARTHE to support and provide guidance to the GLAD at-sea experiment in the summer of 2012, we have run several real-time ocean model forecasts, starting on May 16, 2012, well before the GLAD drifter deployment. These include two ensembles (gom32r and gom32q), ncom3km, ncom1km and hycom4km. All of these forecast outputs are archived and made available on web servers for all the CARTHE scientists and students involved in this project. The raw forecast data are also available if requested. The real-time forecast results were evaluated every cycle by the local scientists at NRL, and important information and findings were provided to CARTHE scientists in other organizations via emails or regular teleconferences. The implementation and operation of these forecast systems were conducted successfully without a glitch and provided great value and real-time guidance to the drifter deployment. In this paper, we describe the details of these numerical forecast systems and the corresponding products including the RELO ensembles, particularly the calibrated ensemble. The introduction and description of these forecast systems will provide background to scientists inside and outside the CARTHE project.

In addition, we examine the performance and efficiency of these forecast systems based on our comprehensive evaluations. We choose to verify these forecast products against Navy's operational observations used in NCODA for 109 days from 00UTC June 1 to 00UTC September 19, 2012. Detailed comparisons of these forecast systems are carried out based on the most commonly used verification metrics. The advantages and disadvantages of different systems or models are studied and summarized. We demonstrate the differences between the ensemble and single forecasts, and in particular we propose a calibrated ensemble (gom32q) with enhanced initial spread to overcome a difficulty in analysis error estimation in the present 3D-Var-based NCODA DA system. Because the calculation used in NCODA underestimates the analysis error, the initial ensemble perturbations generated through the ET cannot match the real analysis error variance. As a result, our ensemble spread is smaller than the ensemble mean error, and the reliability of the ensemble is compromised. Another separate effort has been underway to improve the analysis error directly in NCODA, but this will take time to develop and evaluate. The mixing parameter perturbation scheme introduced is also part of these efforts to improve the RELO ensemble spread. The proposed calibration scheme in this paper is based on the RELO ensemble with the mixing parameter perturbations, and has been proven to be an efficient and effective method to further improve the ensemble spread.

To understand the direct impacts of this spread calibration on the overall performance, these two ensemble spreads are compared directly from different perspectives. These include ensemble plumes of a single location, vertical and horizontal distributions,

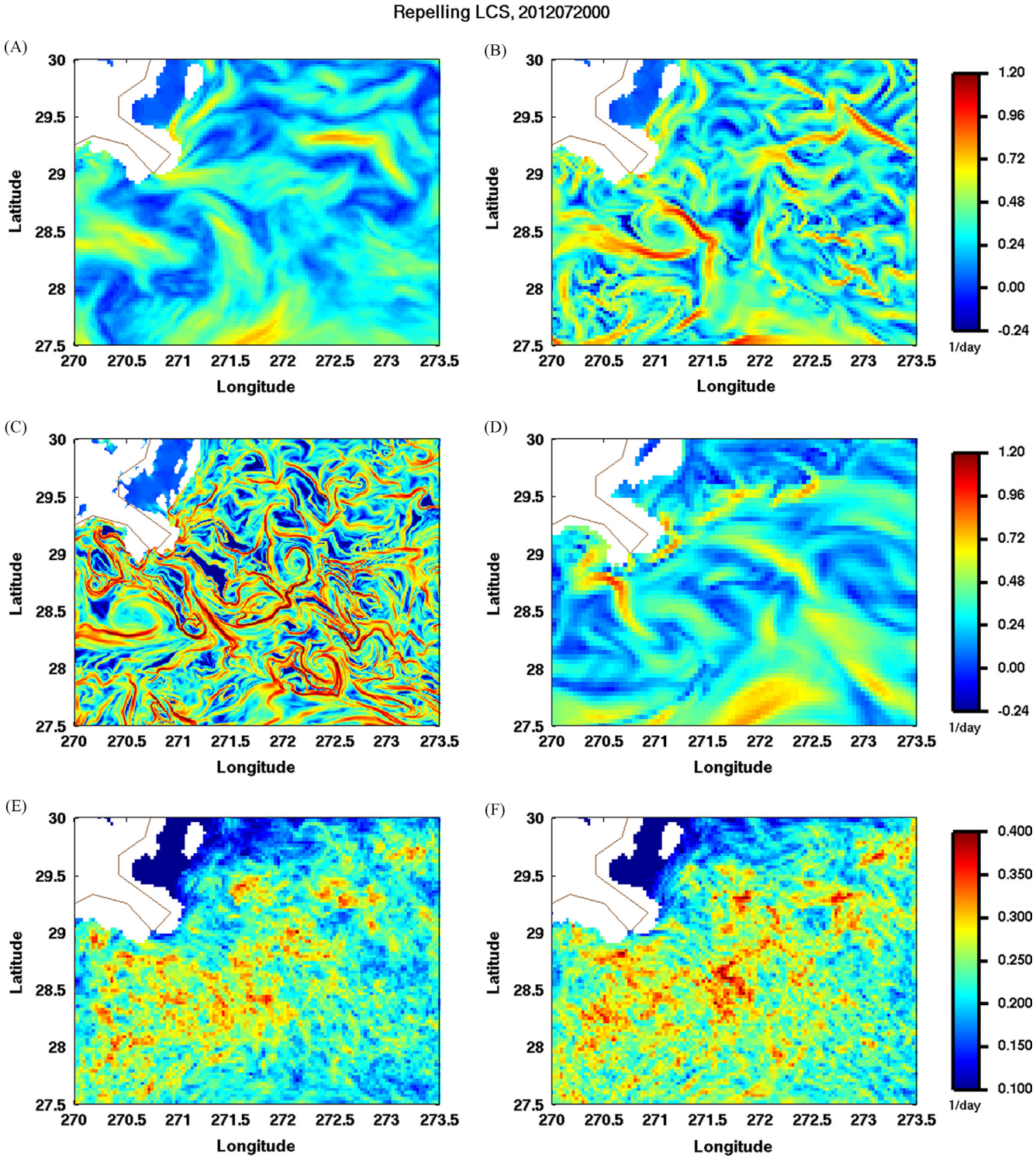

Fig. 13. As in Fig. 12, but for the smaller domain around the GLAD drifters area.

long-time averages over the whole period of the experiment, and averages over different observation spaces with different dynamics. The results show that ensemble spread is clearly enhanced from the calibration. It is found that this calibrated ensemble gom32q is superior to the un-calibrated ensemble gom32r in terms of quantitative forecasting accuracy, skill, and reliability. The metrics we evaluated include RMS error, anomaly correlation, spread-reliability, and Talagrand rank histogram.

For temperature in the full observation space and averaged over 109 days, ncom1km has lowest RMS value due to its much higher resolution, while the calibrated ensemble q has lower RMS value than ensemble r. For salinity the calibrated ensemble mean q is the most accurate, followed by the un-calibrated ensemble r, ncom3km and ncom1km in both full observation space and upper 0–100 m. To quantify the forecast skills of different systems, we compute AC using observations as truth, and climatology as a reference to account for seasonal variation. For temperature, ensemble q has the highest skill score, followed by ensemble r. For salinity, again the calibrated ensemble q is most skillful, followed by ensemble r, ncom3km and ncom1km in both observation

spaces. One surprising result is that ncom1km performs poorer than ncom3km for salinity. This is due to the larger interpolation errors incurred during the interpolations of very high resolution forecasts onto the very sparse observation space for verifications.

To compare the reliability and consistency of two ensembles, we use spread-reliability diagram and Talagrand rank histogram. The spread-reliability curves at three forecast lead times (24, 48, and 72 h) for temperature are computed for full observation space, 0–100 m and surface. The results indicate that the calibrated ensemble q is closer to the diagonal line (more reliable) than ensemble r for all three forecast lead times and in all three observation spaces. For salinity, ensemble q is closer to diagonal line for all forecast lead times and both observation spaces, except for the 24-h forecast in the full observation space. In this case, both ensembles are relatively close to the diagonal line, but ensemble q is slightly over-dispersive, and over-predicts the forecast error variance.

The rank histograms, which describe both reliability and consistency of an ensemble, are computed and compared. For temperature, ensemble q has a much flatter distribution than ensemble r for all three forecast lead times and for all three observation spaces. The associated consistency index of ensemble q is about 20–30% lower than ensemble r in each of these cases. For salinity in the full observation space and the observation space between 0 and 100 m, ensemble q has much flatter rank histograms than ensemble r in all cases. The consistency index for ensemble q is about 50% lower than that of ensemble r. The index values are much closer to 1, which is the value for an ideal ensemble system.

Another contribution from the ensemble forecasts discussed in this paper is the application of ensembles to Lagrangian trajectory prediction. It is demonstrated that the ensemble can generate important uncertainty information in addition to predicting the particle trajectory with the highest probability, in contrast to a single ocean model forecast. In addition, we show the impact of ensemble spread on the trajectory prediction. The calibrated ensemble q with more reliability can pick up completely different trajectory directions which are missed by the less reliable ensemble r.

Finally, we apply the ensemble to computing LCSs for the GOM. The repelling LCSs identified by ensembles q and r are compared with those generated by the single models with different resolutions (ncom3km, ncom1km, hycom4km). As expected, the LCSs based on ensemble means of both ensembles q and r are similar. Each ensemble is composed of 32 ncom3km model runs with the initial perturbations generated based on the analysis error from NCODA. It is interesting to note that the LCSs identified by the ensemble means have larger spatial scales than those produced by ncom3km. This is due to the filtering effect of ensemble mean which removes some small scale features. This can be an advantage in situations where only larger scales of transport barriers are needed, such as tracer prediction in longer time scales.

Our results also show that the repelling LCSs are sensitive to model resolution. The LCSs produced by hycom4km have the largest scales, while ncom1km, which has highest resolution in our experiments, is able to produce the finest small-scale LCS structures that cannot be generated by using lower resolution models such as ncom3km, hycom4km or the ensemble mean. However, the uncertainties of LCSs generated by all these single models cannot be estimated. The real advantage of the ensemble in this application is the capability for estimating these uncertainties (Lermusiaux et al., 2006a); the uncertainties of LCSs provided by these two ensembles are different due to different ensemble spreads.

We plan to continue to explore the application of ensembles to Lagrangian trajectory and LCS prediction. It is well known that ensemble techniques have been used at major NWP centers for about 20 years. Almost all the major operational meteorological centers have adopted ensemble forecast systems, and ensemble products have already become essential components in those centers. The benefits over single forecast have been widely recognized and accepted by the public, not just researchers. The application of ensemble approaches in the Lagrangian framework of ocean prediction is still largely unexplored. The work presented in this paper is our first step in this direction. Our next immediate tasks include improving the efficiency of computing the dominant FTLE values for ensemble systems, and improving the computational efficiency of attracting LCSs in ensembles.

Acknowledgments

This research was made possible in part by a grant from BP/The Gulf of Mexico Research Initiative (GoMRI) through the Consortium for Advanced Research on Transport of Hydrocarbon in the Environment (CARTHE). It was also partly funded through the PRACTICE 6.2 project at NRL and supported by the Office of Naval Research (Program Element 0602435N). We thank our colleagues at NRL at Stennis Space Center for their assistance, particularly Germana Peggion, Jan Dastugue, David Sitton, Michael Phelps and the scientists from other organizations in CARTHE.

References

Barron, C.N., Kara, A., Martin, P., Rhodes, R., Smedstad, L., 2006. Formulation, implementation and examination of vertical coordinate choices in the Global Navy Coastal Ocean Model (NCOM). Ocean Model. 11, 347–375.

Beron-Vera, F.J., Olascoaga, M.J., Goni, G.J., 2008. Oceanic mesoscale eddies as revealed by Lagrangian coherent structures. Geophys. Res. Lett. 35, L12603, http://dx.doi.org/10.1029/2008GL033957.

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

Bowler, N.E., 2006. Comparison of error breeding, singular vectors, random perturbations and ensemble Kalman filter perturbation strategies on a simple model. Tellus 58A, 538–548.

Bowler, N.E., Arribas, A., Beare, S., Mylne, K., Shutts, G., 2009. The local ETKF and SKEB:upgrade to the MOGREPS short-range ensemble prediction system. Quart. J. R. Meteorol. Soc. 135, 767–776.

Buizza, R., Houtekamer, P., Toth, Z., Pellerin, P., Wei, M., Zhu, Y., 2005. A comparison of the ECMWF MSC and NCEP global ensemble prediction systems. Mon. Weather Rev. 133, 1076–1097.

Chassignet, E.P., Smith, L.T., Halliwell, G.R., Bleck, R., 2003. North Atlantic simulations with the HYbrid Coordinate Ocean Model (HYCOM): impact of the vertical coordinate choice, reference pressure, and thermobaricity. J. Phys. Oceanogr. 33 (12), 2504–2526.

Coulliette, C., Lekien, F., Paduano, J., Haller, G., Marsden, J., 2007. Optimal pollution mitigation in Monterey Bay based on coastal radar data and nonlinear dynamics. Environ. Sci. Technol. 41, 6562–6572.

Counillon, F, Bertino, L., 2009. High-resolution ensemble forecast for the Gulf of Mexico eddies and fronts. Ocean Dyn. 59, 83–95.

Cummings, J., 2005. Operational multivariate ocean data assimilation. Quart. J. R. Meteorol. Soc. 131, 3583–3604.

Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrohpic model using Monte Carlo methods to forecast error statistics. No. C5. J. Geophys. Res. 99, 10143–10162.

Haller, G., Yuan, G., 2000. Lagrangian coherent structures and mixing in two-dimensional Turbulence. Physica D 147, 352–370.

Haller, G., Sapsis, T., 2011. Lagrangian coherent structures and the smallest finite-time Lyapunov exponent. Chaos 21, 1–5.

Halliwell, G.R., 2004. Evaluation of vertical coordinate and vertical mixing algorithms in the HYbrid Coordinate Ocean Model (HYCOM). Ocean Model. 7 (3–4), 285–322.

Houtekamer, P., Lefaivrem, L., Derome, J., Ritchie, H., Mitchell, H., 1996. A system simulation approach to ensemble prediction. Mon. Weather Rev. 124, 1225–1242.

Huntley, H.S., Lipphardt, B.L., Kirwan, A.D., 2011a. Lagrangian predictability assessed in the East China Sea. Ocean Model. 36, 163–178.

Huntley, H.S., Lipphardt, B.L., Kirwan, A.D., 2011b. Surface drift predictions of the deepwater horizon spill: the Lagrangian perspective. In: Liu, Y., MacFadyen, A., Ji, Z.-G., Weisberg, R.H. (Eds.), Monitoring and Modeling the Deepwater Horizon Oil Spill: A Record-Breaking Enterprise. Geophysical Monograph Series, vol. 195. AGU, Washington, D.C., pp. 179–195.

Lekien, F., Coulliette, C., Mariano, A.J., Ryan, E., Shay, L.K., Haller, G., Marsden, J., 2005. Pollution release tied to invariant manifolds: a case study for the coast of Florida. Physica D 210, 1–20.
Lermusiaux, P.F.J., 1999. Estimation and study of mesoscale variability in the Strait of Sicily. Dyn. Atmos. Oceans 29, 255–303.
Lermusiaux, P.F.J., Anderson, D.G.M., Lozano, C.J., 2000. On the mapping of multivariate geophysical fields: error and variability subspace estimates. Quart. J. R. Meteorol. Soc. B 126, 1387–1430.
Lermusiaux, P.F.J., 2002. On the mapping of multivariate geophysical fields: sensitivity to size, scales and dynamics. J. Atmos. Oceanic Technol. 19, 1602–1637.
Lermusiaux, P.F.J., 2006. Uncertainty estimation and prediction for interdisciplinary ocean dynamics. special issue on “uncertainty quantification”. J. Comput. Phys. 217, 176–199.
Lermusiaux, P.F.J., Chiu, C.-S., Gawarkiewicz, G.G., Abbot, P., Robinson, A.R., Miller, R.N., Haley, P.J., Leslie, W.G., Majumdar, S.J., Pang, A., Lekien, F., 2006a. Quantifying Uncertainties in Ocean Predictions. In: Paluszkiewicz, T., Harper, S. (Eds.), Special Issue on Advances in Computational Oceanography. Office of Naval Research, vol. 19 (1), pp. 92–105. http://dx.doi.org/10.5670/oceanog.2006.93.
Lermusiaux, P.F.J., Malanotte-Rizzoli, P., Stammer, D., Carton, J., Cummings, J., Moore, A.M., 2006b. Progress and Prospects of U.S. Data Assimilation in Ocean Research. Oceanography, In: Paluszkiewicz, T., Harper, S., (Eds.), Special Issue on Advances in Computational Oceanography. vol. 19(1), pp. 172–183.
Liu, Y., MacFadyen, A., Ji, Z.-G., Weisberg, R.H. (Eds.), 2011. Monitoring and Modeling the Deepwater Horizon Oil Spill: A Record-Breaking Enterprise. Geophysical Monograph Series, vol. 195. AGU, Washington D.C., p. 271.
Mariano, A.J., Kourafalou, V.H., Srinivasan, A., Kang, H., Halliwell, G.R., Ryan, E., Roffer, M., 2011. On the modeling of the 2010 Gulf of Mexico oil spill. Dyn. Atmos. Oceans 52, 322–340.
Martin, P.J., 2000. A description of the Navy Coastal Ocean Model Version 1. Naval Research Laboratory, Stennis Space Center, MS p. 42. (Technical Report NRL/FR/7322-00-9962).
Maltrud, M., Peacock, S., Visbeck, M., 2010. On the possible long-term fate of oil released in the Deepwater Horizon incident, estimated using ensembles of dye release simulations. Environ. Res. Lett. 5, 1–7.
McLay, J., Bishop, C., Reynolds, C.A., 2007. The ensemble-transform scheme adapted for the generation of stochastic forecast perturbations. Quart. J. R. Meteorol. Soc. 133, 1257–1266.
Mellor, G.L., Yamada, T., 1974. A hierarchy of turbulence closure models for planetary boundary layers. J. Atmos. Sci. 31, 1791–1806.
Mellor, G.L., Durbin, P., 1975. The structure and dynamics of the ocean surface mixed layer. J. Phys. Oceanogr. 5, 718–728.
Molteni, F., Buizza, R., Palmer, T., Petroliagis, T., 1996. The ECMWF ensemble prediction system: methodology and validation. Quart. J. R. Meteorol. Soc. 122, 73–119.
O'Kane, T., Oke, P., Sandery, P., 2011. Prediction the East Australian Current. Ocean Model. 38, 251–266.
Olascoaga, M.J., Beron-Vera, F.J., Brand, L.E., Kocak, H., 2008. Tracing the early development of harmful algal blooms with the aid of lagrangian coherent structure. J. Geophys. Res. 113, C12014,1–10, http://dx.doi.org/10.1029/2007JC004533.
Olascoaga, M.J., 2010. Isolation on the West Florida Shelf with implication for red tides and pollutant dispersal in the Gulf of Mexico. Nonlinear Process. Geophys. 17, 685–696.
Olascoaga, M.J., Haller, G., 2012. Forecasting sudden changes in environmental contamination patterns. Proc. Natl. Acad. Sci. 109, 4738–4743.
Özgökmen, T.M., Griffa, A., Mariano, A.J., Piterbarg, L.I., 2000. On the predictability of Lagrangian trajectories in the ocean. J. Atmos. Oceanic Technol. 17, 366–383.
Özgökmen, T.M., Piterbarg, L.I., Mariano, A.J., Ryan, E., 2001. Predictability of drifter trajectories in the tropical Pacific Ocean. J. Phys. Oceanogr. 31, 2691–2720.
Reynolds, C.A., Ridout, J., McLay, J., 2011. Examination of parameter variations in the US Navy global ensemble. Tellus 63A, 841–857.
Rowley, C., 2008. RELO SYSTEM USER GUIDE. Oceanography Division Naval Research Laboratory. Stennis Space Center, MS USA p. 59.
Rowley, C., 2010. Validation Test Report for the RELO System. Naval Research Laboratory, Stennis Space Center, MS p. 69. (Oceanography Division, NRL Report NRL/MR/7320–10-9216).
Rowley, C., Richman, J., Emanuel, C., 2012. Boundary Condition Uncertainty in the NRL Relocatable Ocean Ensemble Forecast System. AGU Ocean Science Meeting, Salt Lake City, UT 20–25 Feburary 2012.
Shadden, S.C., Lekien, F., Marsden, J.E., 2005. Definition and properties of Lagrangian coherent structures from finite-time Lyapunov exponents in two-dimensional aperiodic flows. Physica D 212, 271–304.
Shadden, S.C., Lekien, F., Paduan, J.D., Chavez, F., Marsden, J.E., 2009. The correlation between surface drifters and coherent structures based on HF radar in Monterey Bay. Deep-Sea Res. Part II: Top. Stud. Oceanogr. 56, 161–172.
Smagorinsky, J., 1963. General circulation experiments with the primitive equations. I: The basic experiment. Mon. Weather Rev. 91, 99–164.
Talagrand, O., Vautard, R., Strauss, B., 1997. Evaluation of Probabilistic System. In: Proceedings of the Workshop on Predictability, ECMWF. Reading, UK, pp. 1–25.
Toth, Z., Kalnay, E., 1993. Ensemble forecasting at NMC: the generation of perturbations. Bull. Am. Meterol. Soc. 174, 2317–2330.
Wei, M., Toth, Z., 2003. A new measure of ensemble performance: perturbations versus error correlation analysis (PECA). Mon. Weather Rev. 131, 1549–1565.
Wei, M., Toth, Z., Wobus, R., Zhu, Y., Bishop, C., 2005. Initial perturbations for NCEP ensemble forecast system. In: Thorpex Symposium Proceedings for the First THORPEX Internal Science Symposium, 6–10 December 2004, Montreal, Canada. The Symposium Proceedings in a WMO Publication 2005, WMO TD no. 1237, WWRP THORPEX no. 6, 2005, pp. 227–230.
Wei, M., Toth, Z., Wobus, R., Zhu, Y., Bishop, C., Wang, X., 2006. Ensemble transform Kalman Filter-based ensemble perturbations in an operational global prediction system at NCEP. Tellus 58A, 28–44.
Wei, M., Toth, Z., Wobus, R., Zhu, Y., 2008. Initial perturbations based on the ensemble transform (ET) technique in the NCEP global ensemble forecast system. Tellus 60A, 62–79.
Wei, M., Toth, Z., Zhu, Y., 2010. Analysis differences and error variance estimates from multi-center analysis data. Aust. Meteorol. Oceanogr. J. 59, 25–34.
Wei, M., Pondeca, M., Toth, Z., Parrish, D., 2012. Estimation and calibration of observation impact signals using the Lanczos method in NOAA/NCEP data assimilation system. Nonlinear Process. Geophys. 19, 541–557.
Wei, M., Rowley, C., Martin, P., Barron, C., Jacobs, G., 2013. The U.S. Navy's RELO ensemble prediction system and its performance in the Gulf of Mexico. Quart. J. R. Meteorol. Soc. 139A, http://dx.doi.org/10.1002/qj.2199.
Wilks, D.S., 2006. Statistical Methods in the Atmospheric Sciences. Cambridge Press„ San Diego, USA p. 627.
Yin, X-Q., Oey, L-Y., 2007. Bred ensemble forecast of loop current and rings. Ocean Model. 17, 300–326.