Ocean Modelling 87 (2015) 86–106

Contents lists available at ScienceDirect

Ocean Modelling

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

Virtual Special Issue
Gulf of Mexico Modelling – Lessons from the spill

Ocean current estimation using a Multi-Model Ensemble Kalman Filter during the Grand Lagrangian Deployment experiment (GLAD)

Emanuel F. Coelho [a,b,*], P. Hogan [b], G. Jacobs [b], P. Thoppil [b], H.S. Huntley [c], B.K. Haus [d], B.L. Lipphardt Jr. [c], A.D. Kirwan Jr. [c], E.H. Ryan [d], J. Olascoaga [d], F. Beron-Vera [d], A.C. Poje [e], A. Griffa [f], T.M. Özgökmen [d], A.J. Mariano [d], G. Novelli [d], A.C. Haza [d], D. Bogucki [g], S.S. Chen [d], M. Curcic [d], M. Iskandarani [d], F. Judt [d], N. Laxague [d], A.J.H.M. Reniers [d], A. Valle-Levinson [h], M. Wei [b]

[a] University of New Orleans, Department of Physics, 2000 Lakeshore Drive, New Orleans, LA 70148, USA
[b] Naval Research Laboratory, Stennis Space Center, MS 39529, USA
[c] University of Delaware, College of Earth, Ocean and Environment, Newark, DE 19716, USA
[d] University of Miami, Rosenstiel School of Marine & Atmospheric Science, 4600 Rickenbacker Causeway, Miami, FL 33149, USA
[e] City University of New York, College of Staten Island, Staten Island, NY 10314, USA
[f] Consiglio Nazionale delle Ricerche – Istituto di Scienze Marine U.O.S. di Puzzuolo di Lerici, Forte S.Teresa 19032, Puzzuolo di Lerici (SP), Italy
[g] Texas A&M University-Corpus Christi, Department of Physical & Environmental Sciences, 6300 Ocean Dr., Corpus Christi, TX 78412, USA
[h] University of Florida, Civil and Coastal Engineering Department, Gainesville, FL 32611, USA

ARTICLE INFO

Article history:
Received 2 December 2013
Received in revised form 20 October 2014
Accepted 4 November 2014
Available online 27 December 2014

Keywords:
Ocean modeling
Data assimilation
Ensemble forecasting
Ensemble Kalman Filter
Lagrangian observations
Ocean currents

ABSTRACT

In the summer and fall of 2012, during the GLAD experiment in the Gulf of Mexico, the Consortium for Advanced Research on Transport of Hydrocarbon in the Environment (CARTHE) used several ocean models to assist the deployment of more than 300 surface drifters. The Navy Coastal Ocean Model (NCOM) at 1 km and 3 km resolutions, the US Navy operational NCOM at 3 km resolution (AMSEAS), and two versions of the Hybrid Coordinates Ocean Model (HYCOM) set at 4 km were running daily and delivering 72-h range forecasts. They all assimilated remote sensing and local profile data but they were not assimilating the drifter's observations. This work presents a non-intrusive methodology named Multi-Model Ensemble Kalman Filter that allows assimilating the local drifter data into such a set of models, to produce improved ocean currents forecasts. The filter is to be used when several modeling systems or ensembles are available and/or observations are not entirely handled by the operational data assimilation process. It allows using generic in situ measurements over short time windows to improve the predictability of local ocean dynamics and associated high-resolution parameters of interest for which a forward model exists (e.g. oil spill plumes). Results can be used for operational applications or to derive enhanced background fields for other data assimilation systems, thus providing an expedite method to non-intrusively assimilate local observations of variables with complex operators. Results for the GLAD experiment show the method can improve water velocity predictions along the observed drifter trajectories, hence enhancing the skills of the models to predict individual trajectories.

1. Introduction

The skills of local and regional high-resolution (i.e. order 1 km resolution) ocean model forecasts running in real-time can be improved by assimilating local and remote observations. The intrusive assimilation process typically consists on the generation of an analysis with corrected model states that are then used to initialize the next forecast cycles (e.g. Lermusiaux and Robinson, 1999; Cummings, 2005; Ngodock et al., 2007; Lunde and Coelho, 2009). The residual errors of these improved analysis along with other sources of uncertainty in boundary conditions, forcing, and model parameters can be used to construct ensembles with multiple runs

* Corresponding author at: Centre for Maritime Research and Experimentation (CMRE), La Spezia, Italy.
E-mail address: emanuel.coelho@cmre.nato.int (E.F. Coelho).

http://dx.doi.org/10.1016/j.ocemod.2014.11.001
1463-5003/

that will define an error-subspace domain with the possible realizations for the ocean states (e.g. Coelho et al., 2009; Lunde and Coelho, 2009; Wei et al., in press).

Generic regional implementations can include several forecasting systems, some in ensemble mode as described in the previous paragraph, or just running with different options, resolutions, forcing, and boundary conditions. All together they will span a larger set that defines a domain containing the best guesses for the ocean state. In the past, several approaches have been used to aggregate and track the best solutions for specific applications within these multi-model ensemble domains (e.g. Coelho et al., 2005; Logutov and Robinson, 2005; Rixen and Coelho, 2007; Leslie et al., 2008; Lenartz et al., 2010, among others). The motivation for this work is the observation that these multi-model-based forecasts generally exhibit improved predictive skills relative to any of the individual runs.

The Consortium for Advanced Research on Transport of Hydrocarbon in the Environment (CARTHE[1]) conducted the Grand Lagrangian Deployment (GLAD) experiment, sponsored by the Gulf of Mexico Research Initiative. Several ocean models were used to support this deployment in the northern Gulf of Mexico (GoM) of more than 300 surface drifters, in late July 2012. These drifters were similar to the CODE-ARGOS systems (Poulain, 1999), drogued at a depth around 1 m to reduce direct windage. Their positions were tracked through the SPOT-Globalstar system every 5 min, and the majority of drifters persisted past October 2012, limited primarily by battery life.

Real-time ocean modeling was conducted during the period from July to November 2012 to support the deployment and tracking of the drifter network. The set of available models included the Navy Research Laboratory (NRL) Navy Coastal Ocean Model running at 1 km (NCOM 1 km) and 3 km (NCOM 3 km) resolutions, the US Navy operational Intra-Americas Seas NCOM (AMSEAS) at 3 km resolution – all with tidal forcing, and two versions of the NRL Hybrid Coordinates Ocean Model at 4 km, one with tidal forcing (HYC-T) and one without tidal forcing (HYCOM). All these models run with time steps shorter than 300 s for numerical stability, but results were only saved every 3 h for the NCOM grids and every 6 h for the HYCOM grids. For the present work all model data were taken with a 6 h temporal resolution and interpolated to a common 4 km grid. All of these systems assimilated satellite altimetry, sea surface temperature, and routine ocean temperature and salinity profile observations, although each system used different options to construct their analysis. None of the systems assimilated data from the drifters.

The GLAD models had limited success tracking the drifter network and capturing accurately the dominant features (see Tables 1–3). Carrier et al. (2014), showed the intrusive assimilation of the GLAD drifter data into the NCOM at 3 km improved the model estimates. In this paper we present a different approach using a non-intrusive cycling methodology named Multi-Model Ensemble Kalman Filter (MEKF) that combines the available models with the drifter data to further improve the short-range forecasting skill over that of any of the individual models or their direct aggregation (un-weighted mean).

The method is designed to be used on-scene or to complement routine intrusive data assimilation systems and has the advantage of running in post-processing, hence allowing finer tailoring to specific locations and/or applications. It combines the in situ measurements of surface currents estimated from surface drift trajectories and the multiple estimates of surface velocities as delivered by the several models to derive consistent optimal forecasts of the ocean states. Although this work focuses on surface velocities, the method is also applicable to multi-variate analysis or to correct variables of any kind that are directly correlated with the ocean state and for which a forward model exists (e.g. oil spill plumes).

The MEKF assumes that a prior system is running with several forecast models with potentially different resolutions, parameter choices, forcing and initialization, either through multiple independent systems and/or by single model ensemble simulations. The filter derives an optimal linear combination of these estimates that minimize deviations from a prior ensemble mean and that still provide the best fit to the data during a short time window at the beginning of the forecast. These deviations from the prior aggregation are taken within the bounds of the several models' covariances and observations representation errors, together with their

[1] http://carthe.org.

Table 1
Performance metrics for model velocity forecasts for 0–24 h. The models NCOM 1 km, NCOM 3 km, AMSEAS, HYCOM and HYC T are described in the text and correspond to those used by the MEKF. The row "persistency" shows the results using the first measured velocity within the forecast cycle, for each track, taken as constant for the full forecast range. The row "mean" corresponds to the 5 models un-weighted mean. The row "LSF" shows the un-weighted mean of the models after a least-square fit to the data using the data within the first 24 h and the row "consensus" shows the MEKF results based on the global fits during the first 24 h. As such both the "LSF" and "consensus" should be interpreted as an analysis during the period 0–24 h. The "RMS" column corresponds to the root mean square of the model-observations mismatches in m/s. "Mean Err." shows the mean model-observation mismatch in m/s. "SNR" corresponds to the signal to noise ratio estimated as the product between the observed velocity standard deviation and the RMS, such that for values above 1 we could expect added value from the forecast. "Corr. Mag." shows the magnitude of the correlation between predicted and observed velocity taken as a complex variables. The column "Corr. Angle" shows the angle in degrees of the correlation predictions. Note that the small angle by the HYC T might not be relevant since the magnitude of the velocity correlation is very small. The two best values in each column are highlighted in bold. The worst value in each column is marked with italics.

Model	RMS	Mean Err.	SNR	Corr. Mag.	Corr. Angle
NCOM 1 km	0.40	0.34	1.03	0.47	−6
NCOM 3 km	0.38	0.33	1.07	0.54	−9
AMSEAS	0.41	0.35	1.01	0.54	*−13*
HYCOM	0.40	0.35	1.02	0.46	−8
HYC T	*0.47*	*0.40*	*0.87*	*0.32*	**0**
Persistency	0.41	0.32	1.01	0.52	−8
Mean	0.34	0.30	1.20	0.59	−8
LSF	0.34	0.30	1.23	0.60	−11
Consensus	**0.32**	**0.27**	**1.29**	**0.64**	**−3**

Table 2
Same as Table 1 but for the forecast period 24–48 h.

Model	RMS	Mean Err.	SNR	Corr. Mag.	Corr. Angle
NCOM 1 km	0.41	0.36	1.00	0.43	−8
NCOM 3 km	0.39	0.34	1.04	0.51	−12
AMSEAS	0.42	0.36	0.98	0.51	−14
HYCOM	0.42	0.37	0.98	0.42	−9
HYC T	*0.48*	*0.41*	0.86	*0.30*	**0**
Persistency	*0.48*	0.40	*0.85*	0.34	*−23*
Mean	0.36	0.31	1.15	0.56	−9
LSF	0.37	0.31	1.15	0.55	−14
Consensus	**0.35**	**0.29**	**1.19**	**0.57**	**−6**

Table 3
Same as Table 1 but for the forecast period 48–72 h.

Model	RMS	Mean Err.	SNR	Corr. Mag.	Corr. Angle
NCOM 1 km	0.41	0.36	0.99	0.43	−6
NCOM 3 km	0.39	0.34	1.04	0.51	−10
AMSEAS	0.43	0.36	0.97	0.49	−14
HYCOM	0.42	0.37	0.97	0.40	−10
HYC T	0.47	0.40	0.87	0.30	**1**
Persistency	*0.52*	*0.43*	*0.79*	*0.23*	*−37*
Mean	0.36	0.31	1.14	**0.54**	−9
LSF	0.37	0.32	1.13	0.53	−13
Consensus	**0.35**	**0.30**	**1.16**	**0.54**	**−4**

individual least-squares projections onto the observations. The solution constitutes the optimal posterior consensus forecast for the variables of interest, given the set of most recent observations, and is valid through the remaining portion of a forecast range until the next scheduled routine data assimilation cycle occurs. These improved estimates can be used directly for operational applications or as new background fields for the routine assimilation systems used by the individual ocean models.

The errors of the MEKF velocity estimates were computed along the different GLAD Lagrangian trajectories and used to benchmark the performance of the filter in comparison to each of the individual real-time model runs. The results demonstrate that this non-intrusive single variable approach can improve significantly the skill to predict ocean currents along the dynamical trajectories. Ongoing work will expand the approach by integrating this method with other incremental and variational data assimilation methods and will implement a fully multivariate analysis combining profiles of ocean measurements with the surface velocity data collected by the drifters.

Section 2 presents an overview of the GLAD experiment and discusses some examples of Lagrangian trajectory analysis. The Multi-Model Ensemble Kalman Filter (MEKF) is presented in Section 3. The implementation results are examined in Section 4, and conclusions are provided in Section 5.

2. The Grand Lagrangian Deployment experiment (GLAD)

CARTHE conducted the large-scale GLAD experiment in the Gulf of Mexico during the period July 20–31, 2012, more than 300 low-cost custom-made drifters were deployed by the R/V Walton Smith near the Deepwater Horizon site off the Louisiana coast. It constituted an essential first step to study the complex and elusive surface ocean currents that transport pollutants and to characterize the dispersion and the complex multi-scale interactions among mesoscale and sub-mesoscale oceanic flows (e.g. Olascoaga et al., 2013; Poje et al., in press).

Using SPOT GPS units, the drifters were similar in design to the CODE-ARGOS surface drifters (e.g. Poulain, 1999). They were drogued at a depth of approximately 1 m to decouple their motion from direct wind forcing (estimated 1–3% windage, Davis, 1985) and damp wave induced motions without introducing a wave-phase related bias (Davis, 1985). Positions were reported at 5-min intervals, although significant data gaps occurred mostly due to adverse weather conditions, which required careful data processing to produce accurate trajectory and velocity estimates. Data processing consisted of the removal of bad data points defined as outliers in position and/or velocity magnitude, sharp turns and those with inconsistent short-term position sequences near large reception gaps. Data gaps were then filled using a

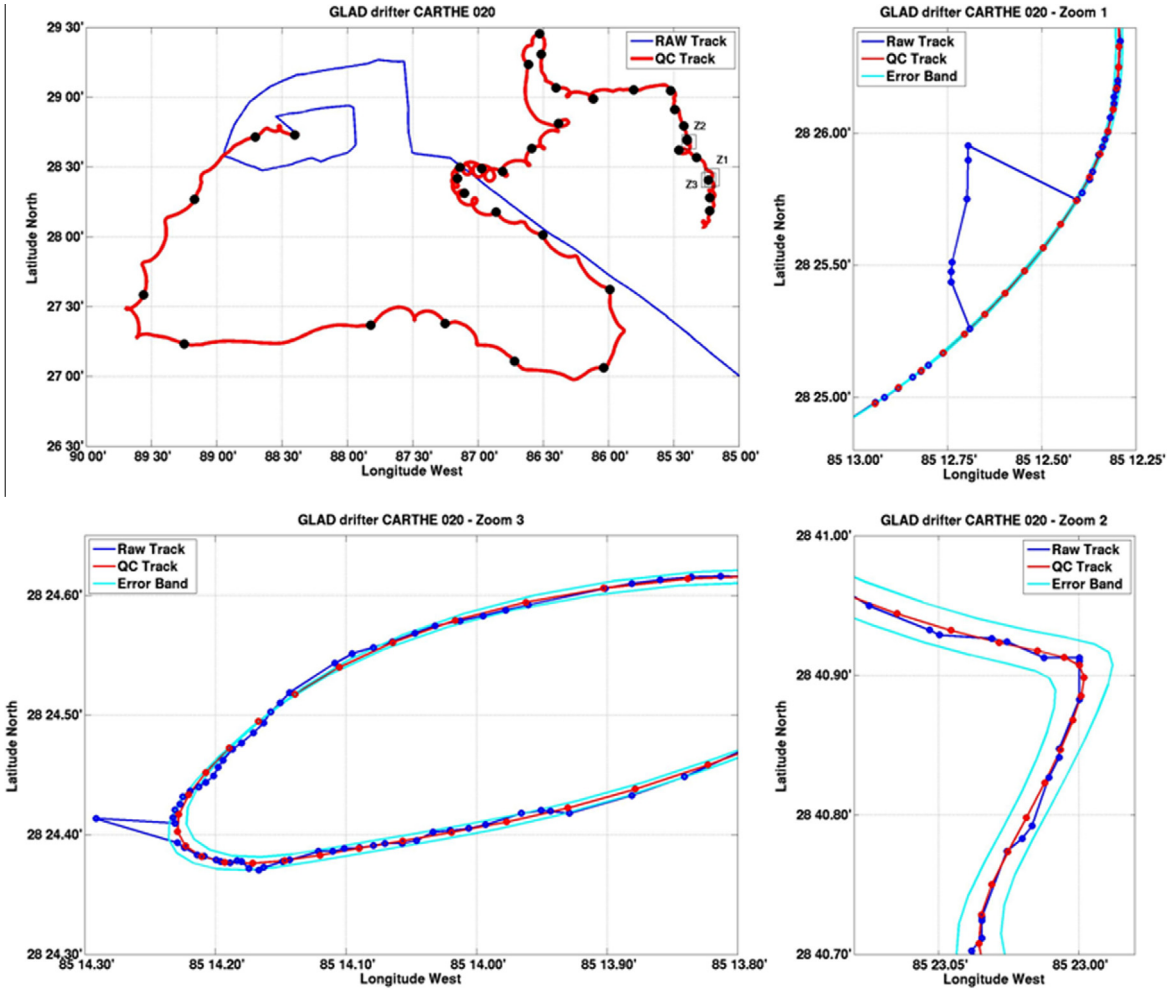

Fig. 1. Drifter CARTHE-020 quality control and data processing details. This drifter reported valid data between July 21, 2012 at 12:15 UTC and September 29 at 23:45 UTC. The blue dots and lines in all panels correspond to raw observations as directly measured through the SPOT units with a 5-min instrumental sampling rate. The red dots and lines correspond to the final processed 15-min sampled positions. The black dots in the upper left panel show 2-day intervals along the track. The cyan lines show the error bands along the tracks based on the gaps in the raw data after the removal of bad positions. The panels with Zoom 1, 2 and 3 show three examples of data processing, where outliers were removed from the raw data (Zoom 1 and 3) and the impact of the filter smoothing prior to 15-min decimation for the final data set (Zoom 2).

non-causal spline interpolation to produce regularly spaced positions at 5 min intervals. These data sets were then used to compute the drifters's local velocities. Finally, the position and velocity data were low-pass filtered with a 1-h period cut-off and sampled at uniform 15-min intervals over a uniform time grid, equal for all drifters. Fig. 1 shows some examples of data processing results for the drifter CARTHE-020.

During the experiment, multiple ocean models were running with resolutions from 1 to 4 km, delivering each day ocean currents forecasts for a period equal to or larger than 72 h. These runs were independently updated every day, by assimilating the most recent in situ and satellite observations. They were used primarily to fine-tune drifter launch positions and identify features that could dominate local ocean dispersion.

The drifters deployment scheme is summarized in Fig. 2. It consisted of a large-scale survey (LSS) with 20 drifters launched across the northern GoM on July 20–21, in an inward spiral toward the region of concentrated drifter deployments. This was followed by four tight groups of dozens of drifters each, spaced from 100 to 15 km apart. An additional four drifters were released on the shelf. The deployments following the LSS were designed to sample sub-mesoscale dispersion characteristics in different flow regimes.

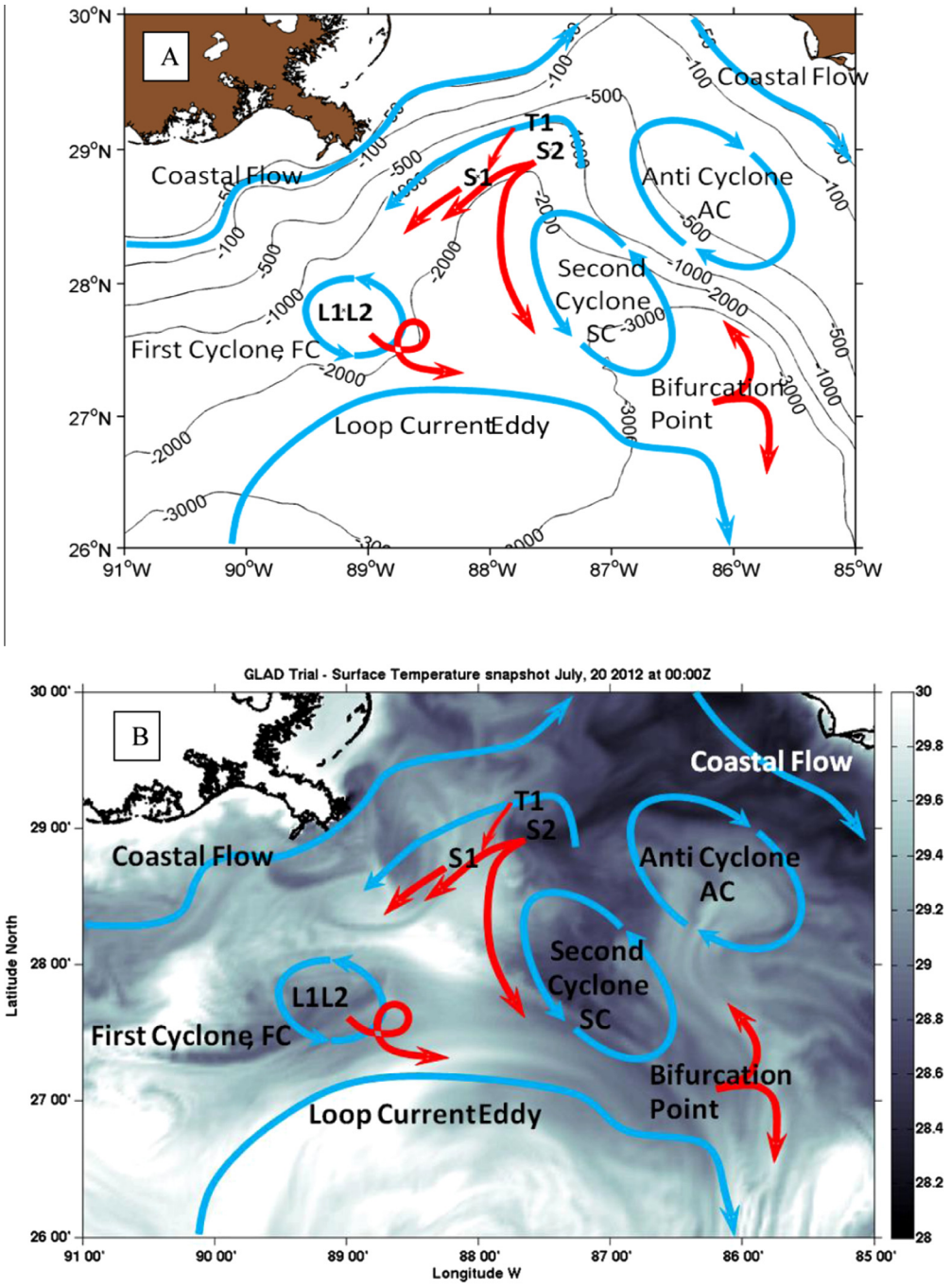

Fig. 2. Ocean circulation features during GLAD. Panel A shows the deployment positions of the drifter arrays at the locations S1, S2, T1, and L1L2 overlaid on the bathymetry. The initial flow directions of the drifter arrays are indicated by the red arrows. The blue arrows show the general circulation and mesoscale features during this time as inferred from independent sea surface height satellite observations. Panel B shows the same information overlaid on the surface temperature estimated by the RELO-NCOM 3 km model on July 20, 2012 at 00:00 UTC showing a reasonable agreement between the initial drifter movements and the alignment of the predicted frontal systems.

In the first intensive deployment (S1) on July 22, centered at 28.8°N 88.1°W, 90 drifters were launched in an S-shaped formation within a 12 km by 9 km area. The second intensive deployment of 90 drifters on July 26 (S2) targeted the eastern edge of the Mississippi River outflow at 29.2°N 87.6°W, crossing the salinity front observed from the ship. A third deployment of 27 drifters (T1) on July 29, occurred at the head of the DeSoto Canyon at 29.0°N 87.5°W. Meanwhile, a small cyclone was observed in satellite SST and some of the model simulations at 27.8°N 89.2°W. This region was then selected for the deployment of the remaining 63 drifters (L1L2) on July 30–31 in an L-shaped configuration.

The Loop Current during July 2012 had extended far northwest into the GoM, and a Loop Current Eddy (LCE) had detached from it just prior to the GLAD drifter deployments. This LCE constrained the southern boundary of the drifter trajectories (about 27°N) throughout the experiment. The general flow north of the LCE consisted of a coastal jet flowing eastward inshore of the 50 m isobath. The jet extended from the Louisiana shelf, past the Mississippi River outflow, eastward past the Mississippi and Alabama coasts, and southeast along the Florida coast. This coastal jet persisted throughout late October. The sea surface salinity from the model runs also suggested a significant advection of Mississippi River fresh water to the east from the Atchafalaya basin and from the Mississippi delta.

Until early September, the drifter trajectories evolved along these complex coastal dynamics influenced by the two coastal cyclonic eddies strongly interacting with each other and with the LCE and by a northeast anti-cyclonic eddy that evolved in shape. By late August some of the drifters were caught in a feature current that developed between the LCE and an enhanced secondary cyclone to the northeast.

The different models displayed somewhat different structures as seen in Fig. 3, even though they all assimilated the same observations. To illustrate the complexity of the observed coastal dynamics and the challenges that remain in predicting Lagrangian trajectories when multiple systems are available, Fig. 3 shows the trajectory of the drifter CARTHE 001 for 10 days prior to August 19, 2012 overlaid on the mean surface temperature fields and mean surface velocities as predicted by each of the model runs. Overall the models seem to capture a well-defined frontal system along or near the observed trajectory. As a result we can see a good general alignment of the temperature fronts and velocity directions with the observed drifter progression during most of the track. The drifter trajectory was however occasionally crossing temperature gradients as predicted by the models and showing misalignments relative to the velocity estimates, making difficult the determination of what model or models could be delivering the best forecasts of the surface velocity.

In late August, Tropical Storm Isaac crossed the experiment area, bringing winds above 25 m/s and producing very high measured drifter velocities, above 2 m/s. Some of the drifters were stranded onshore, and those in deeper water were dispersed offshore as can be seen in Figs. 4A–4C.

Many of the drifters lasted more than 80 days until late October, and only a few were entrained by the Florida Current and exited the Gulf. The vast majority of the drifters remained within the central and northeastern Gulf.

3. The Multi-Model Ensemble Kalman Filter (MEKF)

When multiple forecasting systems are available with different set-up options, resolutions, initialization, forcing and boundary conditions, one can use them to define a probability space as the span of the detected realizations of the ocean state. Typically these domains will contain only a small number of realizations when compared to the total number of state variables being predicted. Also, some of these model estimates might be strongly correlated especially during the first forecast hours if they are assimilating similar sets of observations. The MEKF is designed to track an optimal solution within such a linear space that will correspond to the smallest departure from a prior guess but that also provides the best fit to a given sequence of observations over a short time window. The major assumptions underlying this technique are that the prior domain will include the true ocean states and that the systems are unbiased through iterations with independent intrusive data assimilation systems.

Solutions to track best guesses within these multi-model domains have been addressed in previous work and can be

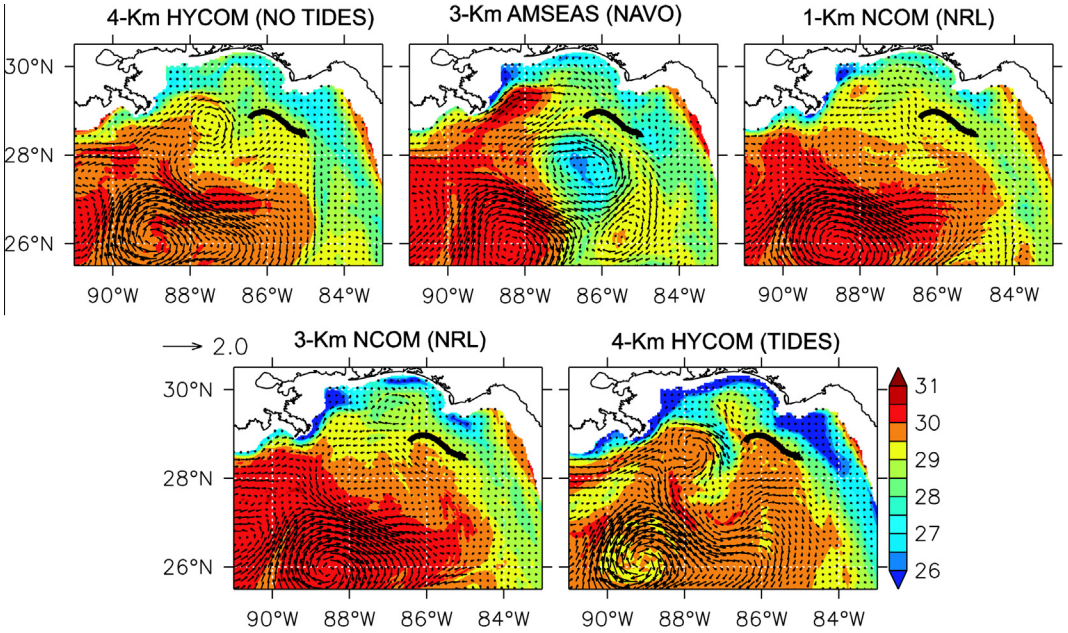

Fig. 3. Surface temperature and currents averaged between August 10 and August 19, 2012 00:00 UTC for the five models run in real-time during the GLAD experiment. The black line overlay on the plots shows the CARTHE-001 drifter trajectory for the same period.

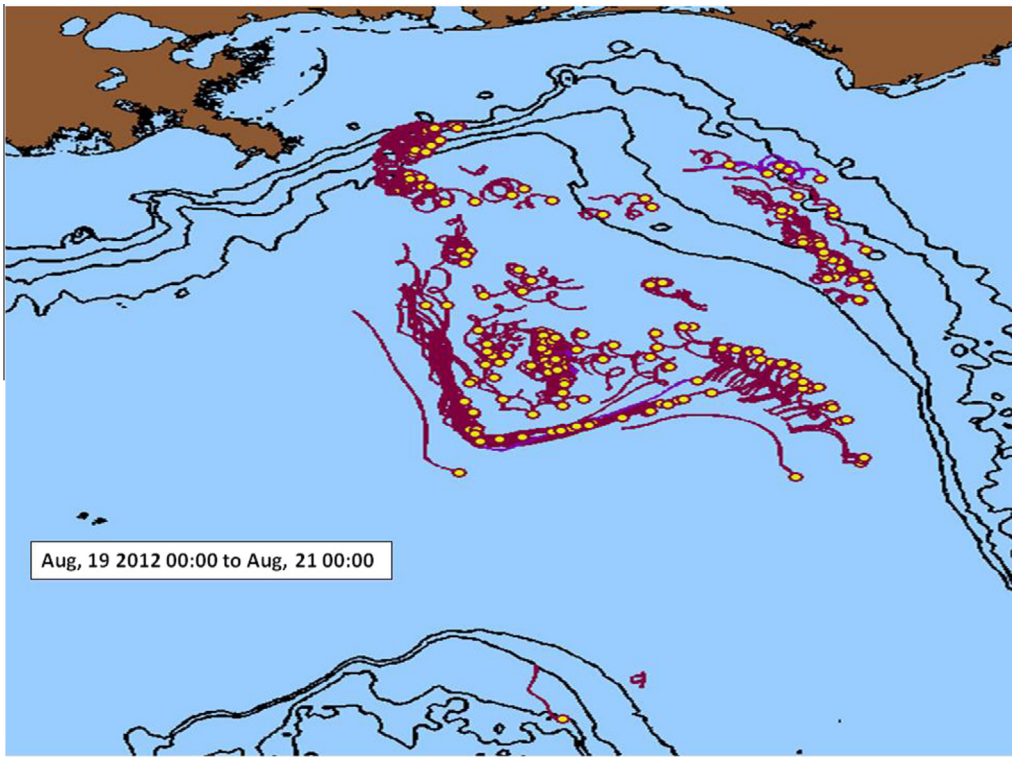

Fig. 4A. GLAD drifter network (extracted from the real time display run by the CARTHE team at the University of Delaware) during the period August 19–21, 2012 (days 30–32 in the trial) before the arrival of the Tropical Storm Isaac to the region. The lines show the last 48 h of the tracks, and the large dots correspond to the reported positions at the end of the snapshot time. Times are UTC.

summarized by three different categories: the local (Lagrangian) approach; the consensus approach; and the Eulerian approach. In the Local approach (e.g. Coelho et al., 2005) aggregation solutions were identified as the best linear combination of the models for each independent realization or application, independently of each other. As such, the optimal (weighted average) forecasts of drifter paths and velocities were computed independently for each drifter, based only on the local fits to the data collected by the same platform over a recent past. This platform-centric method maximizes the individual fit to observations but does not provide a robust extrapolation for positions away from their locations or to other platforms since it does not take into account the data and/or model space–time correlations. In the Consensus approach (Logutov and Robinson, 2005; Rixen and Coelho, 2007; Leslie et al., 2008), the solutions were found by choosing weights for the model runs based on global fits to all available observations. This approach is computationally efficient but when the available models have poor global skills it may not produce improved results. Also remote model errors may strongly degrade local accuracy if there is no regularization step. Finally, in the Eulerian approach (Lenartz et al., 2010), weights were computed at each point on a common grid, constrained by a high-dimensional error covariance matrix, computed from recent model climatology. This solution has the advantage of allowing observed innovations to produce locally consistent fits, thus handling a large number of degrees of freedom. However, since the number of models is small and they might be strongly correlated it requires the use of regularization schemes and crude approximations of the prior error covariance. Furthermore, the smooth estimates of this error covariance matrix may not be valid when individual models are undergoing independent data assimilation, resulting in discrepancies between consecutive forecasting cycles.

The MEKF can produce either local or consensus analysis, by tracking individual or subsets of Lagrangian platforms or by computing domain-wide consensus estimates. The skill resulting from the MEKF will be strongly dependent on the skills of the prior models and on the available observations. In fact, a major assumption is that each individual model is unbiased (i.e. that on average their estimates do not have persistent errors). For this purpose it is then assumed the bulk of the model state corrections resulting from local and remote observations are being performed through routine intrusive data assimilation running on each model separately. As a result, one should expect the MEKF to produce only small deviations relative to a prior ensemble mean. The MEKF problem is formulated next, following as close as possible the standard notation for Kalman filters described in Ide et al. (1997).

Assume there is an n by d matrix $\mathbf{M} = \{\mathbf{m}_i^T,\ i = 1,\ldots,n\}$ with n unbiased model estimates (i.e. with a zero domain-wide mean error) of d pre-normalized state variables included in each vector $\mathbf{m_i}$. These variables are non-dimensional and already mapped onto a common grid. As such, this matrix will include the available forecasts within the range $[0, t_f]$ for all variables over the full set of possible locations in the grid.

The MEKF will then find and track a set of n weights $\mathbf{w} = \{w_i,\ i = 1, n\}$, such that the linear combination $\sum_{i=1}^{n} w_i \mathbf{m}_i$ will provide an optimal forecast, given a set of s observations $\mathbf{y}^o$ during a time window $[0, t_a]$ included inside $[0, t_f]$. As a major constraint we want the weights vector to be consistent with the corresponding models, such that if two models are strongly correlated their weights should be similar, while the models that produce a better least-square fit will have larger weights. This analysis can be viewed within the framework of particle filters applied to data assimilation problems (e.g. van Leeuwen, 2009), by assuming the model realizations are random draws of the possible ocean states and for which one can identify a different weight or relative likelihood. The main difference for the present work is that we assume these models (or particles) to be strongly correlated and that we only have access to a mean estimate of each model representing a class of particles,

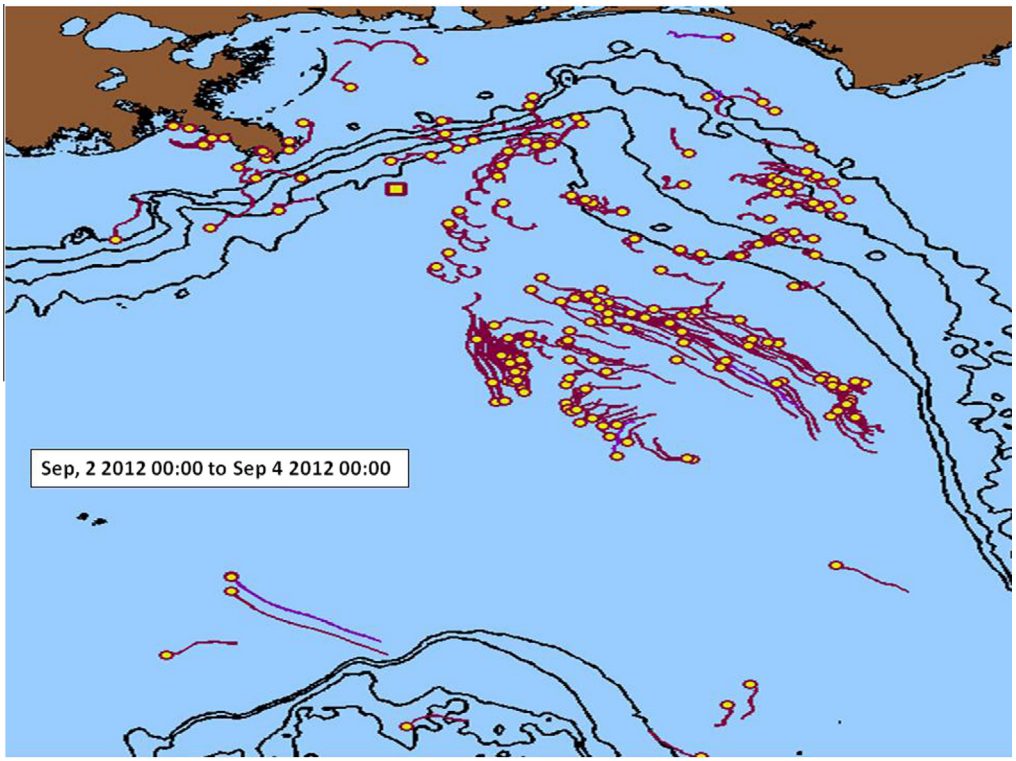

Fig. 4B. Same as Fig. 4A but for the period September 2–4, 2012 (days 43–45 in the trial) just after Tropical Storm Isaac made landfall (August 28, 2012).

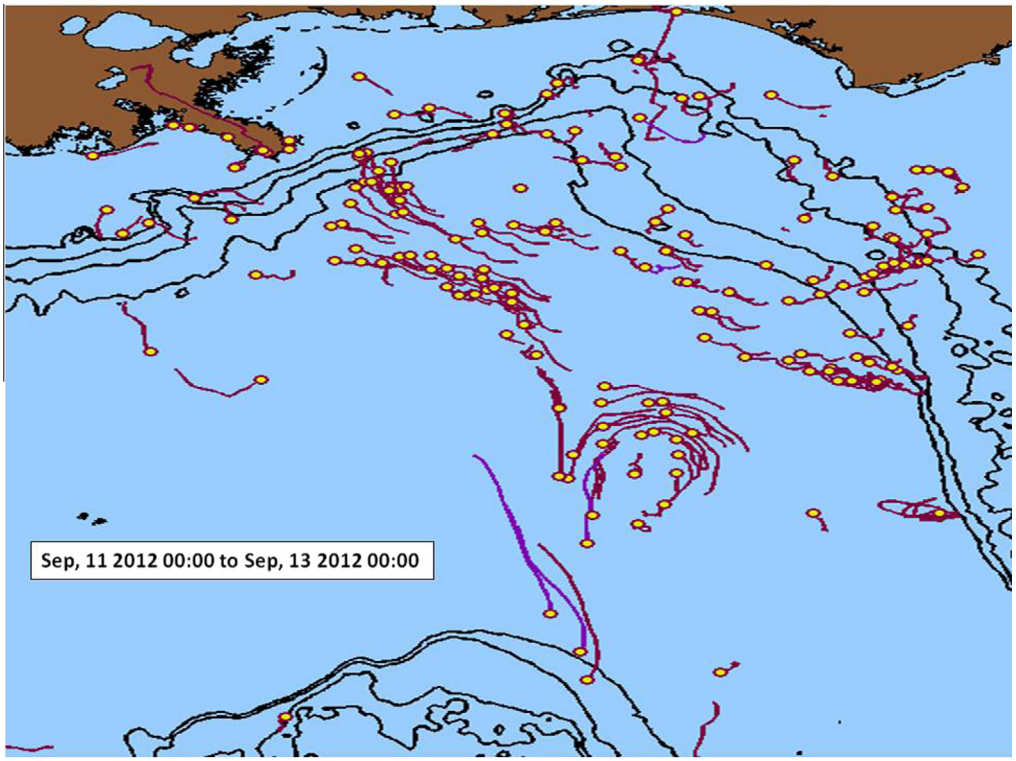

Fig. 4C. Same as Fig. 4A but referring to the period September 11–13, 2012 (days 52–54 in the trial) after the passage of the Tropical Storm Isaac.

and thus the global data fit and updates need to be addressed accordingly. On the other hand, this constrained aggregation approach can also be viewed in the perspective of more traditional data assimilation work (e.g. Lorenc, 1995) where solutions are found through the minimizations of cost functions that include the relevant criteria and constraints.

A challenge one needs to consider when searching for an optimal model aggregation is that the set of weights computed using

observations made during the short time window $[0\,t_a]$ might not be optimal through the full forecast range t_f. On the other hand if through localization we develop different sets of weights per region and/or variable, the linear combination of the model states using these different weights will require further constraint by the model physics and dynamics. These more complex localization and multivariate aspects of the filter will be addressed in future work. By taking into account these considerations, the present analysis will only set up the basic framework for single variable global weights estimation and test the hypothesis that these weights will persist through a full forecast range within a single variable tracking problem.

The next sections show that the optimal weights computed by the MEKF using a small portion of drifters's dynamical trajectories during the GLAD experiment did in fact persist during the subsequent forecast periods. The MEKF forecasts were performing above any of the models individually, their mean and the results obtained by direct linear regression. Since this analysis with the GLAD data was made for single variable (water velocity) the MEKF estimates did not be change the balances and relations between the other state variables. Future work will study whether this consensus based on GLAD drifters only can improve the full model state.

The set of observations used in the MEKF framework undergo through the same pre-normalization as the model states to produce the set of data innovations $\mathbf{y_i}$ for each of the model estimates $\mathbf{m_i}$:

$$\mathbf{y_i} = \mathbf{y}^{o} - H(\mathbf{m_i}) \tag{1}$$

Here the observation operator H maps the model estimates $\mathbf{m}_i$ with dimension d to the observation space with dimension s, so that $H(\mathbf{m_i})$ is the vector with the ith model estimate of the observed quantities and $H(\mathbf{M})$ is the n by s matrix of all model estimates thereof.

We can start by considering the expected value or un-weighted mean of the available models as a first (or prior) guess for the optimal aggregation. As will be discussed below, this corresponds to the assumption that the true state is included in the model set and that the models are all equally likely. One should note this solution is independent of the data and corresponding innovations.

As a second option we can compute a weight, independently for each model, that would provide a least-squares regression to the global data. These regression coefficients will provide a data-fit of each model and can be related to individual model likelihoods and importance sampling in classical particle filter solutions that have been applied within several geophysical contexts (van Leeuwen, 2009). Here, each draw is considered independent of the others, and the expected value of the model states, considering these likelihoods as weights, produces an estimate that is closer to the observations over $[0\,t_a]$ in a least-squares sense. However, this approach does not take into account the correlations among the several model estimates and among the various observed variables, and as such when applied for model aggregation it might amplify or attenuate significantly the model estimates away from the observations themselves. Since it does not account for consistencies among the several model estimates and along the dynamical trajectories predicted by each model, it may also produce an over-fitting to the data and may be less likely to persist through the full forecast range. These considerations motivated the use of constrained regressions to the observations and models correlations in particle filter formulations and provide further motivation for the MEKF as described next.

The MEKF will search for a solution that preserves the multimodel set information included in the model–model covariance and in the joint regression to observations. For this purpose, we start by defining a vector of random entities $\mathbf{x}(j) = \{x_i(j), i = 1,\ldots,n\}$ corresponding to the state vector that includes all possible versions that could be given by the available models taken as stochastic entities at a location and time denoted as j. There will be d occurrences of this random vector at different locations and times, and the expected value or consensus given by $x(j) = \frac{1}{n}\sum_{i=1}^{n}\bar{x}_i(j) = n\sum_{i=1}^{n} w_i m_i(j)$ provides the best global aggregation (for all values of j), given a sequence of observations and the prior model runs. One should note all values of $x_i(j)$ at different locations and times will be delivered by the same ith model entity that has an expected value $\bar{x}_i(j)$ given by the weighted model draw $m_i(j)$. As such, there will be only n parameters w_i that we want to find and track.

As we can see from the expression below for a generic estimate $\bar{x}_i$, the weight w_i is the conditional probability of model i, relative to the other models, and assesses its relative skill in predicting the local state:

$$\begin{aligned}\bar{x}_i &= \int_{x_{i=1,\ldots,n}} x_i p(x_{i=1,\ldots,n})\mathrm{d}x_{i=1,\ldots,n} \\ &= \int_{x_{j\neq i}} p(x_{j\neq i} \setminus x_i)\mathrm{d}x_{j\neq i} \int_{x_i} x_i p(x_i)\mathrm{d}x_i = w_i m_i\end{aligned} \tag{2}$$

In here $p(x_i = 1,\ldots,n)$ is the joint probability distribution of all stochastic model entities, $p(x_{j\neq i}\backslash x_i)$ is the conditional distribution given the model i and $p(x_i)$ is the marginal probability distribution for the random entity represented by the model i. The prior model conditional distributions without the insight given by the observations or other criteria should all be equal. However, once there are observations or other criteria that allow assessing relative performances, one can expect these conditional distributions to be different for each model. As a result better performing models will have larger weights relative to the poorly performing models. The methodology below details how these weights (or conditional probabilities) can be computed once we have a Least Square Fit of each model run, relative to a given set of observations.

This definition is different from the traditional Kalman filter and particle filter formulations used for ocean data assimilation in that although it uses the model state as the random variable of interest and the draws of available model runs, it assumes that the random components are the models themselves as global entities and not the individual state variables. As such, each available model run would be associated with a weight or likelihood determined by the joint skill in representing all the observations. These are to be determined through a joint-probability distribution of the model states since they can be strongly correlated. One should note the number of the MEKF weights is much smaller than the number of degrees of freedom of traditional data assimilation formulations, producing faster results with modest computer resource requirements. However, local corrections will be determined by a model–model covariance and by each model's space–time scales, so they will include much less detail and will not be limited by or preserving the correlations of the ocean state with the observed innovations.

The weights are determined in the MEKF for any generic model variable at a location and time identified by the index j through a forecast step described by:

$$\begin{aligned}\mathbf{w}^f &= \mathbf{w} \\ x^f(j) &= \mathbf{w}^{fT}\mathbf{m}(j) \\ \mathbf{P}^f(j) &= \frac{n^2}{d-1}\mathbf{W}^f\tilde{\mathbf{M}}\tilde{\mathbf{M}}^{\mathbf{T}}\mathbf{W}^{f\mathbf{T}}\end{aligned} \tag{3}$$

followed by an analysis step that updates the weights prior and covariance, described by:

$$\begin{aligned}\mathbf{w} &= \mathbf{w}^f + \mathbf{K}\Delta\dot{\mathbf{w}} \\ x(j) &= w^T\mathbf{m}(j) \\ \mathbf{P}(j) &= n^2\mathbf{W}\mathbf{P}(j)^f\mathbf{W}^T + \mathbf{Q}(j)\end{aligned} \tag{4}$$

In these equations the superscript f stands for the forecast cycle. The vector $\mathbf{m}(j)$ is the n-vector with the several model draws for the specific variable at location and time j and in the notation henceforth simplified. The vectors $\mathbf{w}^f$ and $\mathbf{w}$ are the model weights forecast and analysis, respectively. The n by n matrix $\mathbf{P^f}$ is the forecast error covariance of the n-dimensional random variable $\mathbf{x}$. The model–model analysis residual error covariance matrix is denoted by $\mathbf{Q}$ and will be defined below. The estimates $\mathbf{P^f}$, $\mathbf{P}$ and $\mathbf{Q}$ will be the same for all variables and locations j for the purposes of the present work, though they will become different if we apply a localization step. $\mathbf{W}$ is an n by n diagonal matrix with the weights $\mathbf{w}$ along the main diagonal and zeros elsewhere and $\tilde{\mathbf{M}}$ is the n by d zero mean model estimates matrix (i.e. each model represented in a row of $\tilde{\mathbf{M}}$ had its mean subtracted). The matrix $\mathbf{K}$ denoting the Kalman gain and the innovations vector $\Delta\dot{\mathbf{w}}$ will be defined below.

Let $\eta(\mathbf{x};\mathbf{x^f},\mathbf{P^f})$ be the prior n-dimensional multi-variate distribution of the stochastic model entities in the vector $\mathbf{x}$. It is assumed to be normal with expected value $\mathbf{x^f}$ and a covariance given by the n by n matrix $\mathbf{P^f}$. The generic expression for a normal distribution is given by

$$\eta_N\left(\mathbf{x};\mathbf{x^f},\mathbf{P^f}\right) = (2\pi)^{-n/2}\left|\mathbf{P^f}\right|^{-1/2}\exp\left[-1/2\left(\mathbf{x}-\mathbf{x^f}\right)^{\mathbf{T}}\mathbf{P^{f^{-1}}}\left(\mathbf{x}-\mathbf{x^f}\right)\right] \tag{5}$$

where $|\mathbf{P^f}|$ is the determinant of the model error covariance matrix $\mathbf{P^f}$. This matrix has the variances of each of the stochastic models in $\mathbf{x}$ along its diagonal, while the off-diagonal terms represent the cross-models covariances.

When there is no other prior information the background expectation (and mode under the normal distribution assumption) of $\mathbf{x}$ can be taken as the vector of the given model states with equal skills in representing the truth (i.e. equal likelihoods). Assuming they will include the true state, the sum of all conditional probabilities will have to add up to 1. This corresponds to giving a weight $w_i^f = 1/n$ to each model entity. This solution denotes the maximum uncertainty solution, since it assumes a uniform likelihood for the model estimates. The background state vector at the index j, is then given by

$$\mathbf{x^f}(j) = \{\bar{x}_i(j), i = 1,\ldots,n\} = \left\{w_i^f m_i(j), i = 1,\ldots,n\right\} \tag{6}$$

To compute the model–model covariance we would need access to an ensemble of realizations of each stochastic model entity at each location, but we only have their mean. However, if we assume the statistics of each entity to be stationary and ergodic in the domain of interest, then we can compute the model variances and covariances by averaging across the d realizations of $\mathbf{x}$ instead of averaging across the ensemble dimension. This assumption might not always hold but it can be mitigated by running the MEKF in localized space–time sub-domains. Using this assumption the covariance of $\mathbf{x}$ will then be given by the n by n matrix $\mathbf{P^f}$ as:

$$\mathbf{P^f} = \frac{n^2}{d-1}\mathbf{W^f}\tilde{\mathbf{M}}\tilde{\mathbf{M}}^{\mathbf{T}}\mathbf{W^{f^T}} \tag{7}$$

where $\mathbf{W^f}$ is a matrix with the prior or background weights (taken as $1/n$) along its diagonal and zeros elsewhere.

Now considering the set of observations $\mathbf{y^o}$ and known observations operator H, each of the individual models will have a least-squares regression to the observed data given by the weights $\dot{\mathbf{w}} = \{\dot{w}_i, i = 1,\ldots,n\}$ found by minimizing the square increments:

$$z_i^2 = \sum_{j=1}^{s}\left(n\dot{w}_i H_j(\mathbf{m}_i) - y_j^o\right)^2 \tag{8}$$

For a linear H, the condition $\frac{\partial(z_i^2)}{\partial(\dot{w}_i)} = 0$ produces the set of weights for the a best fit solution hereafter called Least Squares Fit (LSF):

$$\dot{\mathbf{w}} = \frac{1}{n(s-1)}\left[\mathbf{y^o}^T H(\mathbf{M})^T\mathbf{N}^{-1}\right]^T \tag{9}$$

where $\mathbf{N}$ is a n by n normalization diagonal matrix with the elements corresponding to the variances of the projected model estimates summed over all observations:

$$N_{ii} = 1/(s-1)H(\mathbf{m}_i)^T H(\mathbf{m}_i) \tag{10}$$

These LSF model observations have a n by n residual error covariance $\mathbf{R}$ given by:

$$\mathbf{R} = \frac{1}{s-1}\dot{\mathbf{D}}\dot{\mathbf{D}}^{\mathbf{T}} \tag{11}$$

Here $\dot{\mathbf{D}} = \{nH(\dot{\mathbf{w}}_i\mathbf{m}_i) - \mathbf{y}^o, i = 1,\ldots,n\}$ is an n by s matrix with the residual errors of the LSF, for all models. Thus the diagonal terms in $\mathbf{R}$ represent the LSF residual error variances for each model and the off-diagonal terms the residual error covariances among the different models.

We can also assume the LSF solution to be represented by a multi-variate normal distribution describing the state estimates of $\mathbf{x^{LSF}}\backslash\mathbf{x}$ denoted as $\eta(\mathbf{x};\mathbf{x^{LSF}},\mathbf{R})$. As such, the LSF weights will correspond to the conditional probabilities for each model mean to reproduce the observations. Note that though this LSF solution produces likelihoods or weights that can be conceptually equivalent to traditional particle filter formulations, they differ in the way the residual error matrix is formulated since it preserves the information of the global cross-correlation of model-observations misfits under ergodic approximations. Another note of caution is that as the available draws (or means of the stochastic entities) approach the data, the sum of these weights (or likelihoods) will converge to unity, but when all models are uncorrelated with the data their sum will converge to zero. Therefore the norm $\bar{\dot{w}} = \sum_{i=1}^{n}\dot{w}_i$ can be used as a metric benchmarking the overall likelihood of the available models to reproduce the observations i.e. when $\dot{w}$ is very small most of the true ocean state will likely reside on the null-space of the domain span by the available models draws and vice versa. As such, when this norm is small one can expect the overall forecast performance to be poor.

The objective of the MEKF is then to combine the above distributions into a single estimate that could provide an optimal consensus. Using Bayes' theorem and following a procedure similar to the one described in e.g. Lorenc (1995) and Arulampalam et al. (2002), within the context of variational analysis and Bayesian filters, we can find a posterior probability distribution function for the generic model local state vector, given the models' LSF, as:

$$p(\mathbf{x}|\mathbf{x^{LSF}}) = \frac{p(\mathbf{x})p(\mathbf{x^{LSF}}|\mathbf{x})}{p(\mathbf{x^{LSF}})} = A\,\eta\left(\mathbf{x};\mathbf{x^f},\mathbf{P^f}\right)\eta(\mathbf{x};\mathbf{x^{LSF}},\mathbf{R}) \tag{12}$$

where A is a normalization since

$$p(\mathbf{x^{LSF}}) = \int_{j=1}^{m} p(\mathbf{x^{LSF}}|\mathbf{x})p(\mathbf{x})\mathrm{d}\mathbf{x} = 1/A \tag{13}$$

is the integration through all possible values of $\mathbf{x}$.

For a generic variable at location and time j we can now define the ensemble domain with the possible states for the vector $\mathbf{x}$ as the linear space of the available model draws (i.e. the possible model realization in the basis will be proportional to the available draw of the respective model). Defining the vectors $\mathbf{x}$, $\mathbf{x^f}$ and $\mathbf{x^{LSF}}$ by introducing the diagonal weights matrices $\mathbf{W}$, $\mathbf{W^f}$ and $\dot{\mathbf{W}}$ as

$$\mathbf{x} = \mathbf{W}\mathbf{m}(j) = \begin{bmatrix} w_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_n \end{bmatrix} \mathbf{m}(j),$$
$$\mathbf{x}^{\mathbf{f}} = \mathbf{W}^{\mathbf{f}}\mathbf{m}(j) = \begin{bmatrix} w_1^f & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_n^f \end{bmatrix} \mathbf{m}(j) \quad (14)$$
$$\mathbf{x}^{\mathbf{LSF}} = \dot{\mathbf{W}}\mathbf{m}(j) = \begin{bmatrix} \dot{w}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \dot{w}_n \end{bmatrix} \mathbf{m}(j)$$

where w_i are the variable ensemble weights, w_i^f are the weights corresponding to the expected value of the prior solution and $\dot{w}_i$, are the weights of the expected value of the LSF solution. After some simple algebra, using the generic expression for the normal distribution above in Eq. (5), we can then expand equation (12) into:

$$\begin{aligned} P(\mathbf{x}|\mathbf{x}^{\mathrm{LSF}}) &= B\exp\left[-1/2(\mathbf{x}-\mathbf{x}^f)^T\mathbf{P}^{f^{-1}}(\mathbf{x}-\mathbf{x}^f) - 1/2(\mathbf{x}^{\mathrm{LSF}}-\mathbf{x})^T\mathbf{R}^{-1}(\mathbf{x}^{\mathrm{LSF}}-\mathbf{x})\right] \\ &= B\exp\left[-1/2\mathbf{m}(j)^T\left(\mathbf{W}-\mathbf{W}^f\right)^T\mathbf{P}^{f^{-1}}\left(\mathbf{W}-\mathbf{W}^f\right)\mathbf{m}(j)\right. \\ &\quad \left. -1/2\mathbf{m}(j)^T\left(\dot{\mathbf{W}}-\mathbf{W}\right)^T\mathbf{R}^{-1}(\dot{\mathbf{W}}-\mathbf{W})\mathbf{m}(j)\right] \end{aligned} \quad (15)$$

where $B(A,\mathbf{P}^{\mathbf{f}},\boldsymbol{R})$ can be treated as a constant since it does not depend on $\mathbf{x}$.

In order to obtain the posterior weights for the final solution and taking into account the result expressed in (2) one can now compute the components of the expected vector $\bar{\mathbf{x}}(j)$ through:

$$\bar{x}_i(j) = \int_{\text{ensemble}} x_i(j)p(\mathbf{x}\setminus\mathbf{x}^{\mathbf{LSF}})\mathrm{d}\mathbf{x} = w_i m_i(j) \quad (16)$$

A direct generic solution is rather complex and in order to solve Eq. (16) numerically we would need to have access to an ensemble of possible weights following some unknown probability distribution. However, noting that for distributions of the normal class the expected vector and mode vector will be the same, one can more simply find the expected values for $\mathbf{x}$ by finding the set of parameters that maximize the posterior distribution i.e. by finding the set of weights $\mathbf{W}$ that maximize $p(\mathbf{x}|\mathbf{x}^{\mathbf{LSF}})$.

Denoting the exponential characteristics of normal distributions, for the purpose of finding a maximum we can compute:

$$\begin{aligned} -\ln\left(p(\mathbf{x}\setminus\mathbf{x}^{\mathbf{LSF}})\right) &= 1/2\mathbf{m}(j)^T\left[\Delta\mathbf{W}^T\mathbf{P}^{\mathbf{f}^{-1}}\Delta\mathbf{W}\right. \\ &\quad \left.+\left(\Delta\dot{\mathbf{W}}-\Delta\mathbf{W}\right)^T\mathbf{R}^{-1}\left(\Delta\dot{\mathbf{W}}-\Delta\mathbf{W}\right)\right]\mathbf{m}(j) - \ln(B) \end{aligned} \quad (17)$$

where the correction ($\Delta\mathbf{W}$) and increment ($\Delta\dot{\mathbf{W}}$) matrices are defined as

$$\Delta\mathbf{W} = \mathbf{W} - \mathbf{W}^{\mathbf{f}} \quad \text{and} \quad \Delta\dot{\mathbf{W}} = \dot{\mathbf{W}} - \mathbf{W}^{\mathbf{f}} \quad (18)$$

Noting that $\ln(B)$ and the estimates $\mathbf{m}(j)$ are independent of the weight selection, then the maximization of $p(\mathbf{x}|\mathbf{x}^{\mathbf{LSF}})$ corresponds to minimizing the cost function J defined as the expression between brackets above:

$$\begin{aligned} P(\mathbf{x}|\mathbf{x}^{\mathrm{LSF}})_{MAX} &\Rightarrow \boldsymbol{J}(\Delta\mathbf{W})_{MIN} \\ &= \left[\Delta\mathbf{W}\mathbf{P}^{\mathbf{f}^{-1}}\Delta\mathbf{W}^T + (\Delta\mathbf{W}-\Delta\dot{\mathbf{W}})\mathbf{R}^{-1}(\Delta\mathbf{W}-\Delta\dot{\mathbf{W}})^T\right]_{MIN} \end{aligned} \quad (19)$$

Now, by setting:

$$\frac{\partial\boldsymbol{J}(\Delta\mathbf{W})}{\partial\Delta\mathbf{W}^{439}} = 0 \quad (20)$$

we can find

$$\Delta\mathbf{W}\left(\mathbf{P}^{\mathbf{f}^{-1}} + \mathbf{R}^{-1}\right) = \dot{\mathbf{W}}\mathbf{R}^{-1} \quad (21)$$

Multiplying both sides of this expression on the right by $\mathbf{R}$ and on the left by $\mathbf{P}^{\mathbf{f}}\Delta\mathbf{W}^{-1}$ we can find the final form for the optimal set of weights and define the Kalman gain matrix $\mathbf{K}$ used by the filter as:

$$\Delta\mathbf{W} = \Delta\dot{\mathbf{W}}(\mathbf{P}^f + \mathbf{R})^{-1}\mathbf{P}^f = \Delta\dot{\mathbf{W}}\mathbf{K} \quad (22)$$

Finally, taking the values along the main diagonal of the weights matrices we can produce the weights vectors used in the MEKF as:

$$\mathbf{w} = \mathbf{w}^f + \Delta\dot{\mathbf{w}}\mathbf{K} = \mathbf{w}^f + \mathbf{K}(\dot{\mathbf{w}} - \mathbf{w}^f) \quad (23)$$

Since these weights correspond to the conditional probabilities for each given model, the sum of these posterior expected weights can also be used as a proxy of the joint likelihood of the final consensus to preserve both the prior information and observations: Very small values will correspond to large model spread and large data misfits by all models while values closer to unity will correspond to small spread in the model estimates that are accurately capturing the observations.

If the hypothesis formulated in the previous chapter regarding the persistence of the weights is true and these posterior weights are optimal through the full forecast range $[0,t_f]$, then the errors estimated during the analysis can be used to predict the forecast errors. The model–model residual error matrix $\mathbf{Q}$ can be computed after estimating the MEKF weights as $\mathbf{Q} = 1/(s-1)\mathbf{D}\mathbf{D}^T$ from $\mathbf{D} = \{nw_iH(\mathbf{m}_i) - \mathbf{y}^o, i = 1,\ldots,n\}$. Under the same stationarity and ergodic assumptions mentioned above, the summations for $\mathbf{Q}$ are taken along all observations and as such the diagonal terms will represent the residual error variances of each model estimate after applying the optimal weight and the off-diagonal terms the residual error covariances among the different models. This estimate allows us then to compute the generic analysis error covariance $\mathbf{P}$ as defined in the expressions (5) for the Kalman filter.

Since our final variable of interest is the local estimate of a state parameter that is computed by the mean of the components in the vector $\mathbf{x}$, we can also introduce a local error estimate for the generic variable $\mathbf{x}(j)$. We start by identifying the variance of the vector $\mathbf{x}$ as $v(j) = \frac{1}{n-1}\sum_{i=1}^{n}(nw_im(j)_i - x_i(j))^2$ and defining the n by n matrix $\boldsymbol{\Upsilon}(\mathbf{j})$ with the value of v along its diagonal and zeros elsewhere. Next we assume the local model error covariances will be stationary over the entire grid and for all times in the forecast, such that a normalized model–model correlation matrix $\mathbf{C}$ exists and is the same for all possible j, as given by:

$$\mathbf{C} = \frac{n^2}{d-1}\mathbf{W}\tilde{\mathbf{M}}_{\mathbf{n}}\tilde{\mathbf{M}}_{\mathbf{n}}^T\mathbf{W}^T \quad (24)$$

where $\tilde{\mathbf{M}}_{\mathbf{n}}$ corresponds here to the normalized model estimates matrix but where each column is divided by each model standard deviation. As such the diagonal terms of the matrix $\mathbf{C}$ will correspond to the squares of each model optimal weight and the off-diagonal terms will only show the relative model–model normalized covariances. We can then define a local forecast error covariance for the estimates $\mathbf{x}(j)$ as:

$$\mathbf{P}_x(j) = \boldsymbol{\Upsilon}(j)\mathbf{C} + \mathbf{Q} \quad (25)$$

These error estimates are not used in the present work but they will permit the identification of local sensitivities of the consensus errors and to address targeted observations and adaptive sampling problems as developed in Coelho et al. (2009) among others. These

aspects of error forecast estimates will be expanded in future work and are introduced here only to allow a more comprehensive description of the methodology.

Though the MEKF formulation as detailed here can be generalized and applied to variables of any kind the following sections will discuss the results of implementing this methodology only for a single variable problem, tracking the velocities measured by the drifters deployed during the GLAD experiment.

4. Ocean currents predictions using the MEKF during the GLAD experiment

During the GLAD experiment in the summer-fall 2012 the data from more than 300 drifters were used together with remote sensing imagery and other in situ measurements to observe and understand the motions in the upper ocean of the north-central GoM. These observations were integrated with the estimates delivered by several model runs to better capture the dynamics of the observed dispersion. This section discusses how the MEKF can be used to improve the ocean current estimates by using a non-intrusive four-dimensional analysis that outperforms any of the individual model estimates.

4.1. Velocity tracking using the numerical models

The models running during the GLAD experiment (introduced in Section 2) assimilated standard data (satellite altimetry, sea surface temperature, and temperature and salinity profiles). They all produced daily updates of 72-h or longer range forecasts, using atmospheric forcing from a COAMPS™ atmospheric model run at 25 km resolution and made available by the US Navy Fleet Numerical Meteorology and Oceanography Centre (FNMOC). Lateral boundary conditions were taken from the global HYCOM (e.g. Metzger et al., 2006) or global NCOM (e.g. Barron et al., 2007). Note the MEKF is not aimed to benchmark or assess individual model skills; instead the MEKF will combine the information from all simulations into a best guess estimate, based on their consistency with local observations.

The GLAD drifters were quality controlled and filtered as discussed in Section 2, resulting in a database with positions and velocity estimates on a regular time grid, sampled at 15 min along the drifter's trajectories. Fig. 5 shows a short time series of the observed velocities for the drifter CARTHE-001 corresponding to the period August 19–29, 2012. Although the following paragraphs will be focused on this drifter, similar conclusions can be drawn for other drifters.

The plots in Figs. 5A and 5B show the model velocity components parallel to the drifter track (along-track), interpolated for the same time and position along the trajectory. The NRL NCOM at 1 km estimates are shown in Fig. 5A and the estimates from NRL NCOM 3 km in Fig. 5B. They both use the 24–48 h forecast period. The plots also show the residual velocity component perpendicular to the drifter track (cross-track), which represents model misfits (the drifters measure zero cross-track velocities by definition). All the data shown in these figures were filtered with a cut-off frequency of 12 h and decimated at 0.5 h. When the cross-track velocity estimate are larger than a representation error threshold taken at 0.1 m/s or when the along track velocities are negative one can conclude the models were not interpreting correctly the drifter track as a dynamical trajectory. When the cross-track velocities are negligible and the along track components positive we can assume the models are correctly interpreting the track as a dynamical trajectory. However, if the along-track velocities are either too large or too small then they were either over-estimating or under-estimating its kinetic energy content.

To allow interpreting the causes of bad skills in capturing the dynamical trajectories the plots in Fig. 5 also include a quasi-geostrophic velocity proxy based on the density gradients taken from the model fields along the trajectory and that could diagnose long range misalignments of the observed tracks relative to the predicted density surfaces. They also show the estimates from the model free-surface slopes (mostly dominated by tidal and inertial signals) that diagnose free-surface lead velocity estimates

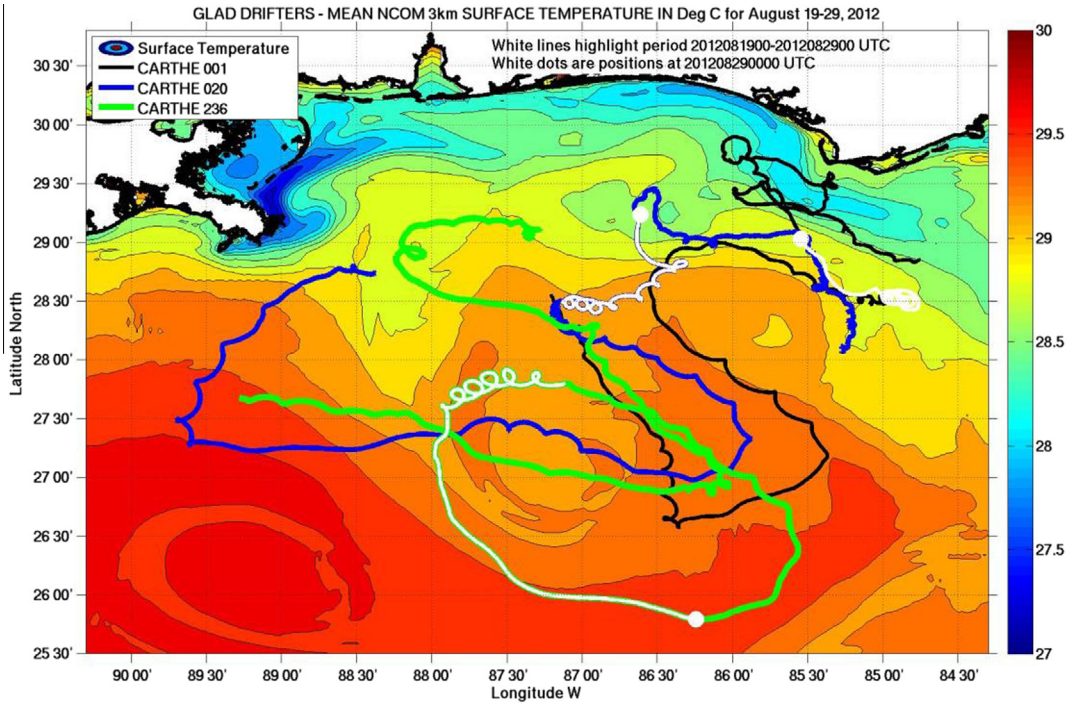

Fig. 4D. GLAD drifter network detail showing drifters CARTHE-001, CARTHE-020 and CARTHE-236 overlay on the NCOM 3 km mean surface temperature for the period August 19, 2012 00:00 UTC–August 29, 2012 00:00 UTC. This same period is highlighted as white along each track.

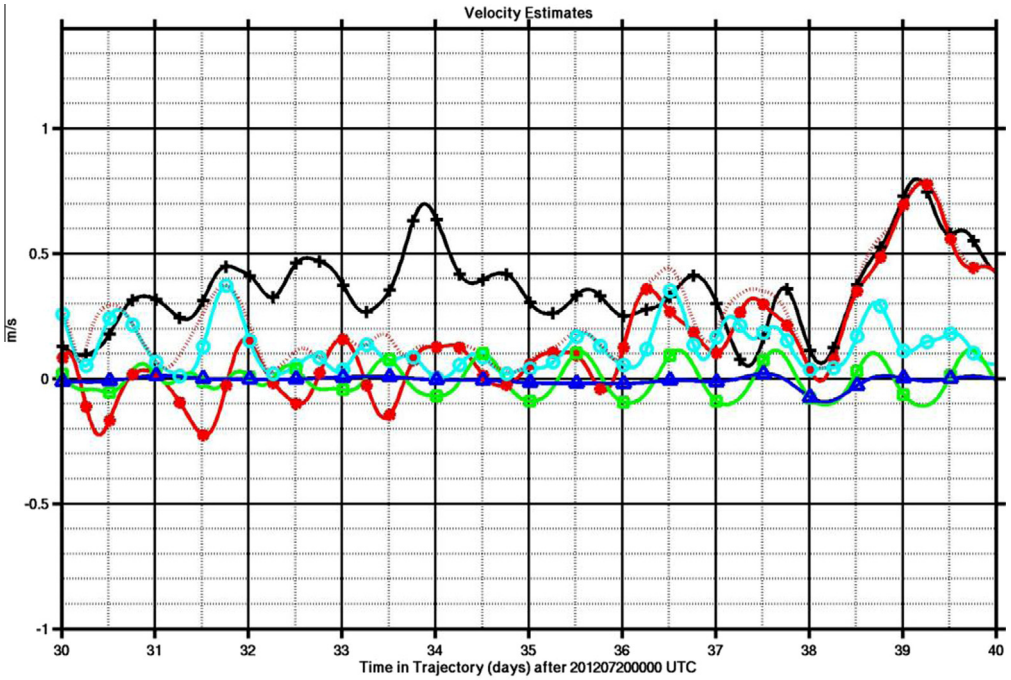

Fig. 5A. Velocities along the drifter track from the model NCOM 1 km and corresponding along the track velocities measured by the drifter CARTHE-001 between August 19, 2012 (day 30) and August 29, 2012 (day 40). The black line (+) corresponds to the drifter observed velocities. The model velocity estimates parallel to the drifter track are marked in red (●) and the residual (perpendicular to the track) estimates in cyan (○). The quasi-geostrophic velocity proxy discussed in the text is shown in blue (Δ), and the velocity proxies based on the along-track surface elevation are displayed in green (□). The dashed red line corresponds to the total magnitude of the model current along the trajectory.

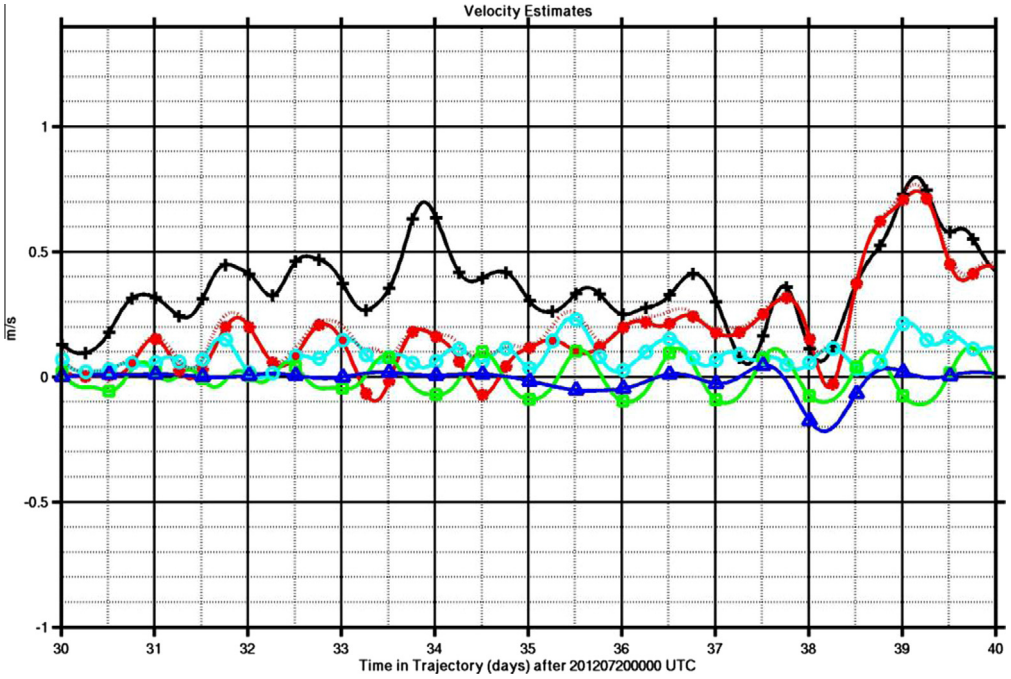

Fig. 5B. Same as Fig. 5A using the NCOM 3 km model estimates.

(i.e. resulting from the available potential energy along the trajectory). The quasi-geostrophic proxy is computed as $V_{qg} \sim \Delta\rho/\rho\ g/f_c\ \Delta\boldsymbol{\varepsilon}^{\mathbf{t}}/\Delta l$, where ρ is the water density; g the gravity acceleration; ε^t the de-tided free surface perturbation; Δl the drifter travel distance; f_c the Coriolis frequency. The free-surface elevation proxy is defined as $V_e \sim sqrt(2g\Delta\varepsilon\ \Delta\rho/\rho)\ f_c\ \Delta t$, where ε is the full free surface perturbation and Δt the time interval between samples. If the models were correct, the cross-track velocities should be negligible, the quasi-geostrophic (QG) proxy should be small along the track, except when along track surface slopes were present and that could force the drifter to cross the density fronts by converting kinetic energy to potential energy and vice versa. The free-surface

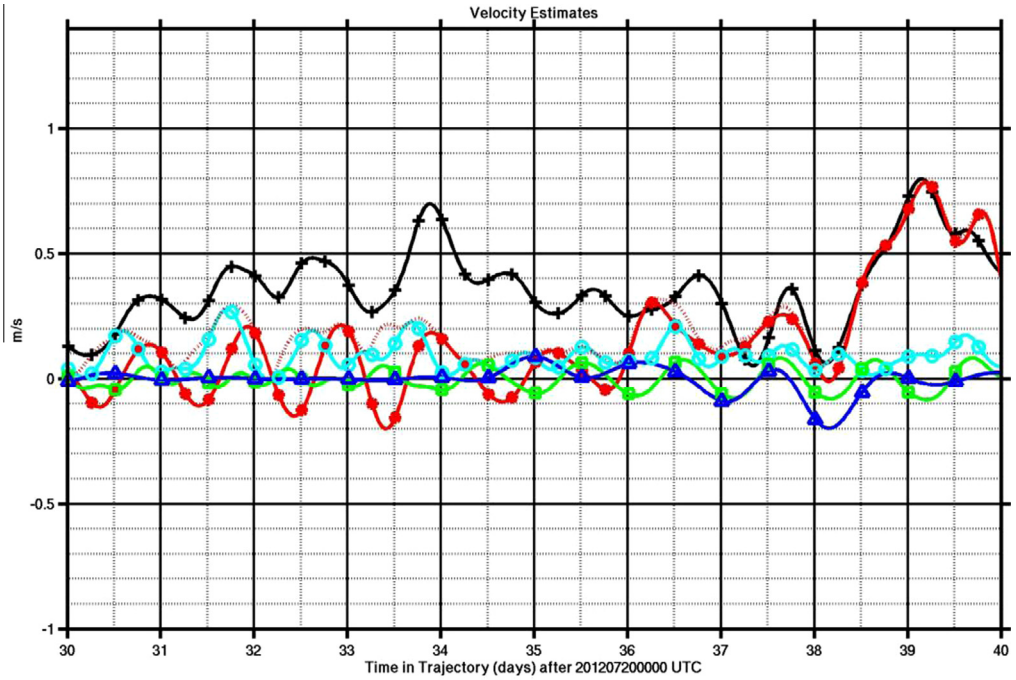

Fig. 5C. Same a Fig. 5A using the MEKF consensus estimates.

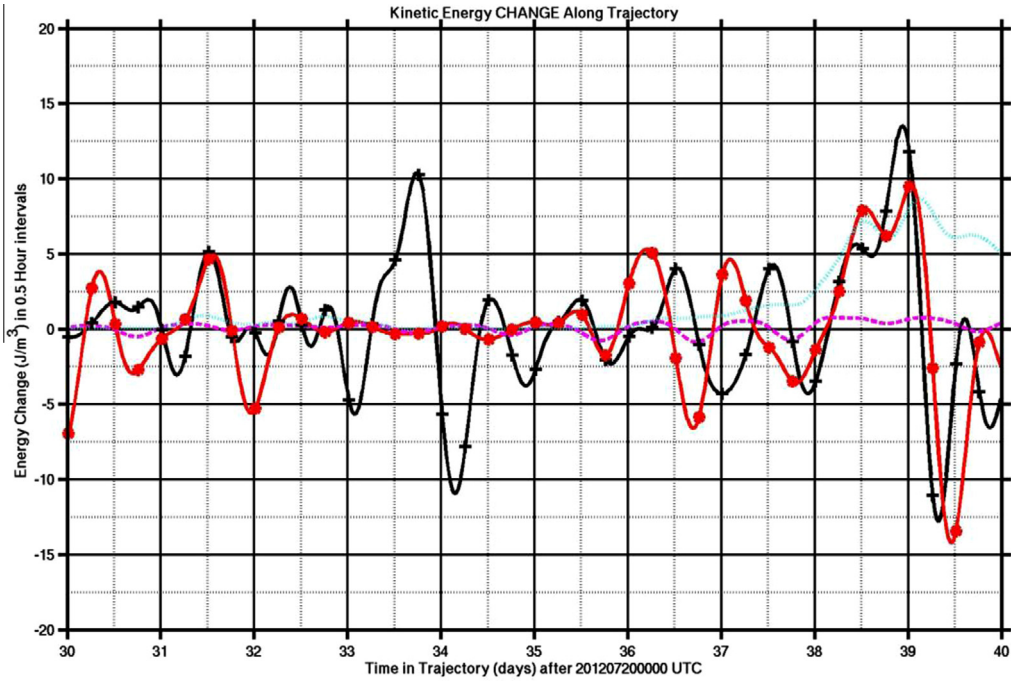

Fig. 6A. Kinetic energy change (work done) estimated by the NCOM 1 km model and corresponding estimates along the track as measured by drifter CARTHE-001, between August 19, 2012 (day 30) and August 29, 2012 (day 40). The black line (+) shows the drifter observations taken at 30 min intervals. The red line (●) shows the estimates from the model velocities. The wind stress energy inputs per unit volume, integrated during the same 30 min intervals are marked in cyan (–) and the net heat fluxes in magenta (..).

lead velocities proxy allows diagnosing along the trajectory the occurrence of these conversions of the potential energy into kinetic energy and vice versa.

From the analysis of Figs. 5A and 5B one can see for both the NCOM 1 km and NCOM 3 km models that the variability of the along-track velocity estimates shows a persistent near-inertial forcing, consistently with the estimates of potential–kinetic energy transfer suggested by the free-surface velocity proxy. However, phase lags and smaller magnitudes of the model estimates develop with respect to the velocity of the drifter. The NCOM 1 km model shows more frequent negative along-track velocities estimates and larger cross-track estimates than the NCOM 3 km model suggesting the latter is doing a better job in capturing the general aspects of this dynamical trajectory. However, the quasi-geostrophic proxy for the

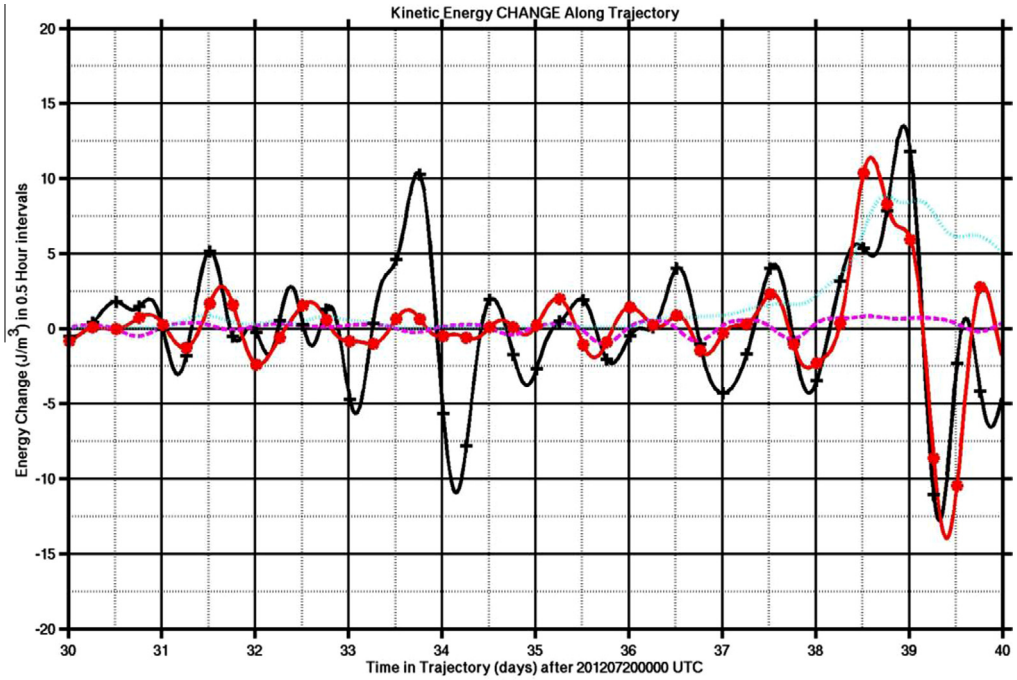

Fig. 6B. Same a Fig. 6A using the NCOM 3 km model estimates.

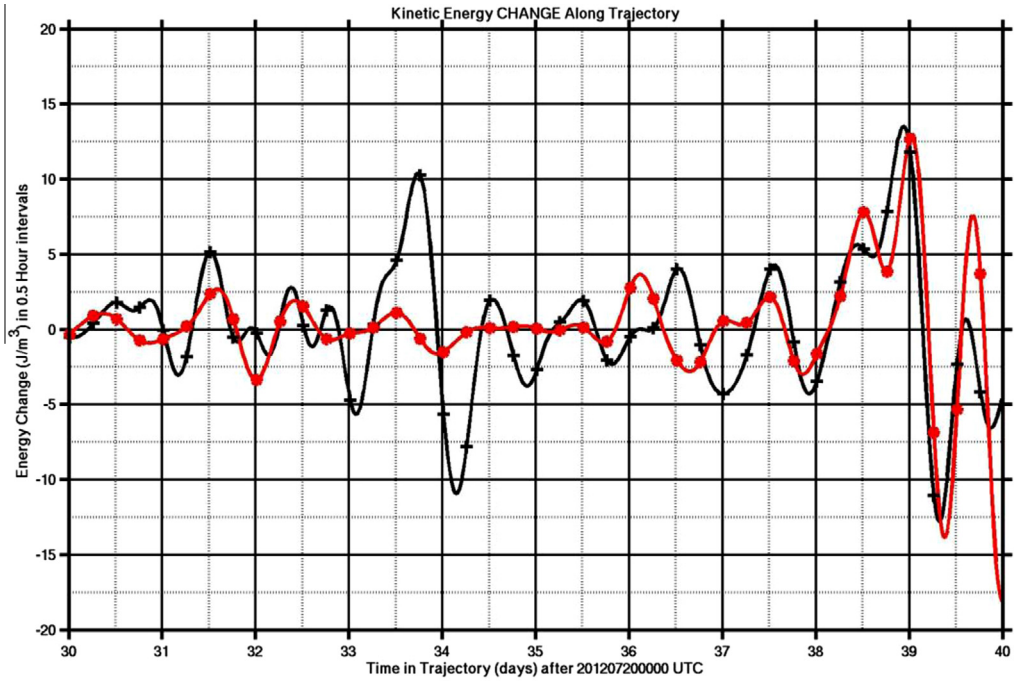

Fig. 6C. Same a Fig. 6A using the MEKF consensus estimates.

NCOM 3 km showed larger values compared to the NCOM 1 km, suggesting some misalignment of the density fronts as predicted in the model relative to the observed trajectory.

Figs. 6A and 6B displays the corresponding changes in kinetic energy for the same cases displayed in Figs. 5A and 5B, as measured by the drifters (black line) and those estimated by the NCOM 1 km and 3 km model runs (using the same 24–48 h segments). The changes in kinetic energy can be either due to conversion of potential energy to kinetic energy and vice versa (displayed by the good correlation between the free-surface velocity proxy and these curves), due to energy flux divergence or convergence or due to external forcing (directly through momentum fluxes or indirectly through net heat fluxes). The energy flux divergence and convergence depends on remote fields outside the trajectory and should be slowly evolving. These can justify the mean differences in energy content over the period of several days. For

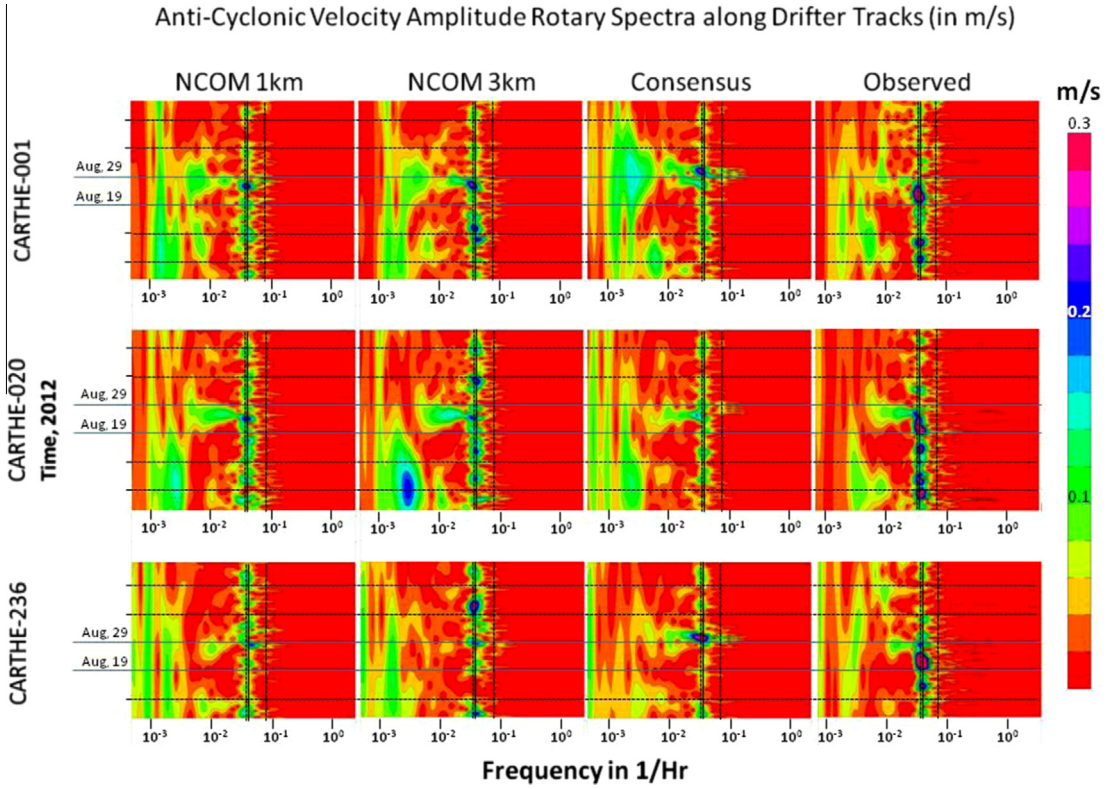

Fig. 7A. Anti-cyclonic velocity rotary spectra for drifters CARTHE-001, CARTHE-020 and CARTHE-236. The vertical axes correspond to the time from July 20, 2012 to October 30, 2012. The period August 19–29 is shown between the two horizontal lines. From left to right, the columns show estimates from NCOM 1 km, NCOM 3 km, the consensus, and the observations. The vertical lines on all the plots show the semi-diurnal frequency on the right, the inertial frequency in the middle, and the diurnal frequency on the left. The horizontal axis corresponds to the frequency logarithmic scale and is shown below the plots. The range of frequencies are from 5e−4 h^{-1}–1/(83 days) to 2 h^{-1} (1/(0.5 h).

interpretation purposes the plots also show the external forcing terms used by each model along the trajectory, namely the energy inputs from the wind stress and by the net heat fluxes.

In the time series displayed in Fig. 6 we can see examples when the impulses were consistent with the observations (e.g. during days 38–39 corresponding to August 27–28, due to the strong winds produced by Tropical Storm Isaac). They also show examples when the models were not capturing well the observed changes in kinetic energy either by lagging and/or underestimating the impulses (e.g. in days 31–33 corresponding to August 20–22). The large impulse on day 34 (August 23) produced a sharp increase in the drifter speed. This corresponds to a strong inertial response as can be seen in Fig. 7A. This impulse was not seen at all by any model and was likely a response to a sharp wind event that was not well reproduced by the smooth 25 km resolution COAMPS winds that was used as the atmospheric forcing an all the models. Figs. 6 and 7 shows that overall models had performances changing within the several forecasts cycles and sometimes not consistently among each other. They also suggest that although NCOM 3 km may be better capturing phases of the impulses, the NCOM 1 km was better in reproducing their magnitudes.

The trajectory of drifter CARTHE-001 was complex as displayed in Fig. 4D. It included periods of fast, straight motion with constant kinetic energy alternating with very intensive high-frequency anti-cyclonic rotation, superimposed with larger scale rotation both cyclonic and anti-cyclonic. Some of these kinetic energy changes are likely associated with local forcing (e.g. August 20, 23 and 27). For events like these, the model estimates were consistent according to the atmospheric forcing being used. However, some other impulse-like events as those on August 25–26 cannot be directly linked to external forcing. The phase lags and amplitude underestimations suggest the energy characteristics in the models were not well aligned with the trajectory and/or that the model energy fluxes due to the local dynamics contained significant errors.

During the 10-day analysis period displayed in Figs. 5 and 6, drifter CARTHE-001 was near the northeast GoM shelf break. However, these results are consistent with the analysis of other drifters like CARTHE-020 and CARTHE-236, in different on-shore and off-shore regions as displayed in Figs. 4D and 7. Because both the NCOM 1 km and NCOM 3 km models underestimated along-track velocities, showed phase lags and cross-track components, neither model could be judged to be better than the other for the regions sampled.

Fig. 7 shows the time evolution of the velocity amplitude rotary spectra for the drifters CARTHE-001, CARTHE-020 and CARTHE-236. By direct inspection one can see the spectral properties are similar for all three drifters. The cyclonic spectral energy in Fig. 7B was lower and mostly concentrated at lower frequencies, while the anti-cyclonic spectra in Fig. 7A shows the most energetic events at the near-inertial ranges.

Fig. 7A shows that significant anti-cyclonic inertial energy appears after approximately August 19 in response to a passing frontal system for all three drifters. Also noticeable is the inertial response to the impulse of kinetic energy in August 23, seen in all drifters and not detected by any of the models, as discussed above for the drifter CARTHE-001. Another energetic event occurs in response to the passage of Hurricane/Tropical Storm Isaac in early September, characterized by large velocities and by a broad-band transient rotary spectral response, mostly anti-cyclonic and

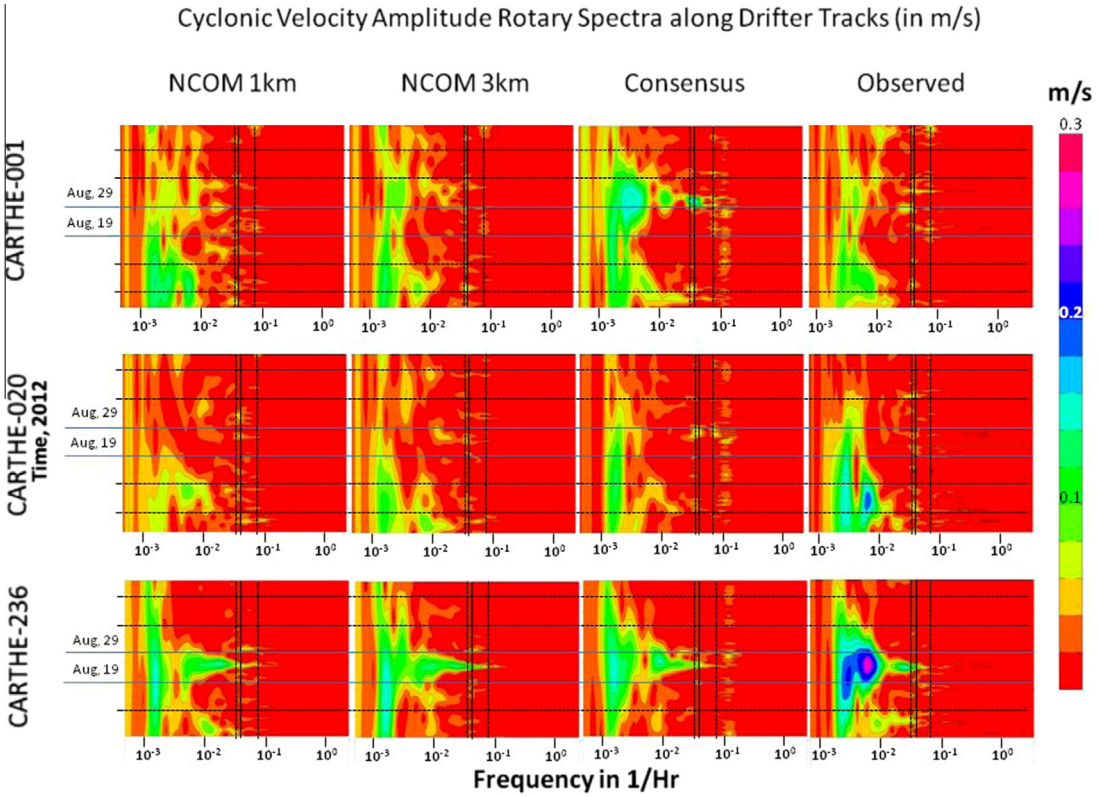

Fig. 7B. As Fig. 7A, but for the cyclonic velocity rotary spectra components.

well captured by the models. This however had a negligible inertial response, suggesting the impact of the storm over the surface to have been like a time impulse with the energy propagating fast across scales and eventually being lost to the sub-surface layers. Also, the cyclonic response of drifter CARTHE-236 caught in an accelerated loop current gyre in the latter half of August was not well captured by the models.

Overall, these examples suggest that models were capturing most energetic responses, although attenuated and with some lags and amplitude mismatches. Models do track dominant kinematic features in the region. The results also demonstrate the degradation in model along-track velocity forecast skill once local dynamics became non-stationary and have large near-inertial responses. Since each individual model carries different assumptions that result in different forecasts and consequently in different errors characteristics in space and time.

4.2. Velocity tracking using the MEKF

Daily outputs of the five models used during the GLAD experiment were interpolated to a common 4 km resolution grid, with 6-h time sampling and covering a 72-h forecast period. All together they defined the domain of uncertainty used by the MEKF and contained the best guesses of the true ocean states at each running cycle. The domain size was however too small to address all the details of a multivariate problem with a large number of degrees of freedom such that for this application we cannot expect the filter to correct large model errors or to produce estimates significantly different from the best individual runs of each cycle.

The MEKF was cycled for each day of the trial using the procedures detailed in Section 3. The least-squares fit coefficients were computed using a time window corresponding to the first 24 h of the forecast cycles (hindcast period) and used to estimate the optimal posterior weights that will produce the consensus analysis for the remaining re-gridded forecast, covering the period 24–72 h (forecast period). The observations made during the hindcast time window were used in a single step, such that the filter background was always the corresponding cycle multi-model mean (i.e. $\mathbf{w}_f = \{1/n; i = 1, \ldots, n\}$). The resulting ocean surface velocities were then compared to the observed drifter velocities to benchmark the filter performance during both the hindcast and forecast phases, assessing if the filter was improving alignment with the observed dynamical trajectories represented by the drifter tracks.

The data observed during the forecast periods was not used to compute the posterior estimates but was collected along the same dynamical trajectories of the data used for the analysis, hence potentially showing the maximum impact of the innovations. During the hindcast period, by design, the optimal posterior estimates should on average outperform any of the models (i.e. have a least-squares error, averaged over the full hindcast period, smaller than any of the single model runs). The hypothesis to test is that the MEKF produces a consensus estimate that provides on average a solution better aligned with the observed trajectories.

Assuming these trajectories will be invariant over the full 72 h run periods, the optimal fit from the first 24 h should persist during the remainder of the forecast period. This will mean the MEKF consensus estimate will also remain valid through the full 72 h period. If such is the case the filter estimates would also outperform on average any of the single models during the forecast period.

The MEKF weights computed for each run cycle are shown in Fig. 8. They change throughout the experiment suggesting that no single model consistently outperformed the others in terms of predictive skills. When the model produces large mismatches with observation, the overall skill of a run will be low and the magnitudes of the corresponding LSF weight small. If a run is highly correlated with the observations the mismatches will be smaller

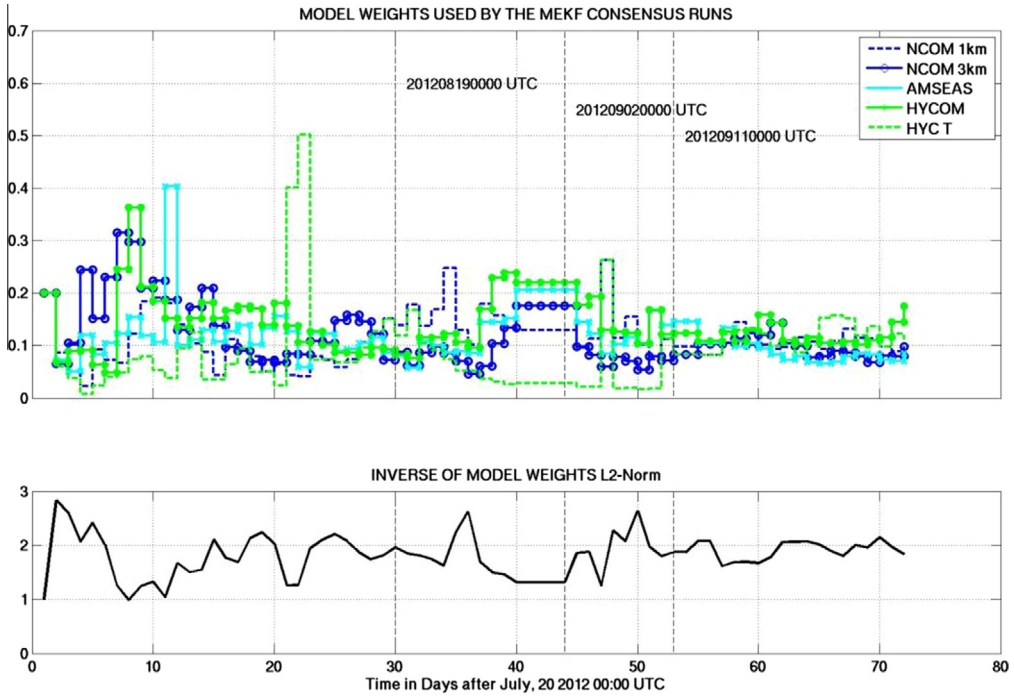

Fig. 8. The upper panel of the figure shows the MEKF weights for the different models through the GLAD experiment. During the days 40–43 the weights were not updated. The lower panel shows the inverse of the sum of the weights magnitude and that was used to re-scale the final MEKF consensus estimates.

and the LSF weights will approach $1/n$. When there is a background model variance (diagonal of the matrix **P**) that is much larger than the LSF residual errors (diagonal of **R**), the final posterior weights will approach the LSF weights (i.e. the weights will converge to the maximizing the likelihood function). On the other hand, if all the LSF residual errors become much larger than the background model variances the final weights will approach the background weights ($1/n$ for all models in this case). In this second situation the observations will be producing a very small or no impact in the final solution. As a note of care, when all models have equally low skills and the background model variance is large, the posterior weights can be very small and damp the magnitudes of the consensus estimates, which will converge to a trivial solution. To avoid this behavior the weights are therefore normalized by the inverse of the sum of the weights magnitude, as shown in Fig. 8 (lower panel). This operation guaranteed the final velocities and kinetic energy did not change just due to the changes in the data and/or models skill. The resulting velocity and kinetic energy changes estimates along the trajectory of drifter CARTHE 001 are shown in Figs. 5C and 6C.

Fig. 5C compares the measured along-track velocities with those computed from the MEKF consensus for the drifter CARTHE-001. For the period shown, the NCOM 1 km and the AMSEAS models were the runs with larger weights (Fig. 8). Also note that all models had small weights on on August 24 and 25 (Fig. 8; days 35–36 in the trial) suggesting degraded consensus forecast skill during this two-day period. From the analysis of this figure we can infer the consensus captures the variability of most events (learning from both the run NCOM 1 km and NCOM 3 km). However, the consensus still shows significant cross-track velocities, and negative along-track values suggesting limited skill in forecasting the trajectories. The MEKF consensus also shows significant quasi-geostrophic velocity proxy.

Fig. 6C displays the equivalent changes in kinetic energy along the trajectory as predicted by the MEKF consensus. Though still with some inconsistencies we can see there are some improvements relative to the MCOM 1 km and NCOM 3 km in reproducing some of the impulses along the trajectory, particularly in matching the response to the front passage during August 20–21 and 25–26 and a better fit in the response to the Tropical Storm Isaac during August 27–28.

This analysis suggests the MEKF consensus is improving the velocity estimates by attenuating the features in the models that are not consistent with observations, while maintaining those that are better correlated with the data. The consensus spectra estimates shown in Fig. 7 further suggests this property either by distributing the energy or by better reproduction both cyclonic and anti-cyclonic events, though still with significant mismatches.

Fig. 9 shows the square errors of the MEKF (consensus) forecasts relative to the observed velocities using a 24–48-h forecast window through the full life of drifter CARTHE-001. The figure also shows the equivalent results using the estimates from the NCOM 1 km and 3 km as reference. For exemplification the MEKF was also run using the Local option for this drifter (i.e. using only the LSF with the data from the drifter CARTHE-001). The Local approach (identified as the Buoy Tracking Analysis in the figure) provides smaller RMS error for some cases but does not outperform the consensus overall. The consensus estimate has residuals that are close to the Local solution and generally performs better than the NCOM 1 km and 3 km estimates. The large error displayed by the models and the consensus on August 22–23 (days 33–34) correspond to the response to the sharp impulse discussed in Fig. 6 that was not seen in any of the models and triggered the energetic near-inertial response seen in Fig. 7 for all drifters.

The performance of the consensus using the full drifter network is shown in Figs. 10A–10D and in Tables 1–3. The figures show the root-mean-square (RMS) of the mismatches between the model estimates and the measured velocities, along 6 h bins over the full 72-h forecast cycles, for the several individual models and their mean (colored lines). The figures also show the final MEKF consensus (black line) that outperforms all other single model estimates during the full forecast period and relative to their mean (red line)

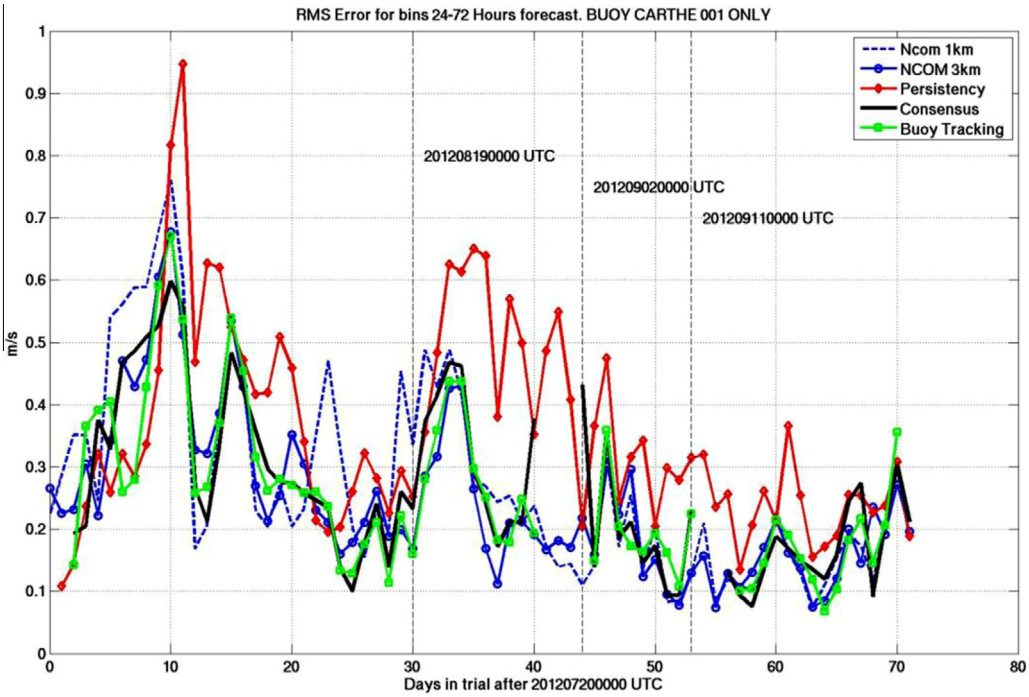

Fig. 9. Root-mean-square (RMS) errors of velocities along drifter CARTHE-001 trajectory. RMS errors are shown for the NCOM 1 km, NCOM 3 km, Velocity persistency, consensus using the full drifter network, and the analysis using only the drifter CARTHE-001 data for the posterior optimal estimation. The RMS errors were computed using the model-observation mismatches during the forecast periods 24–72 h of each daily cycle, such that for example the RMS for day 20 shows the RMS errors averaged for the forecasts covering days 21–23.

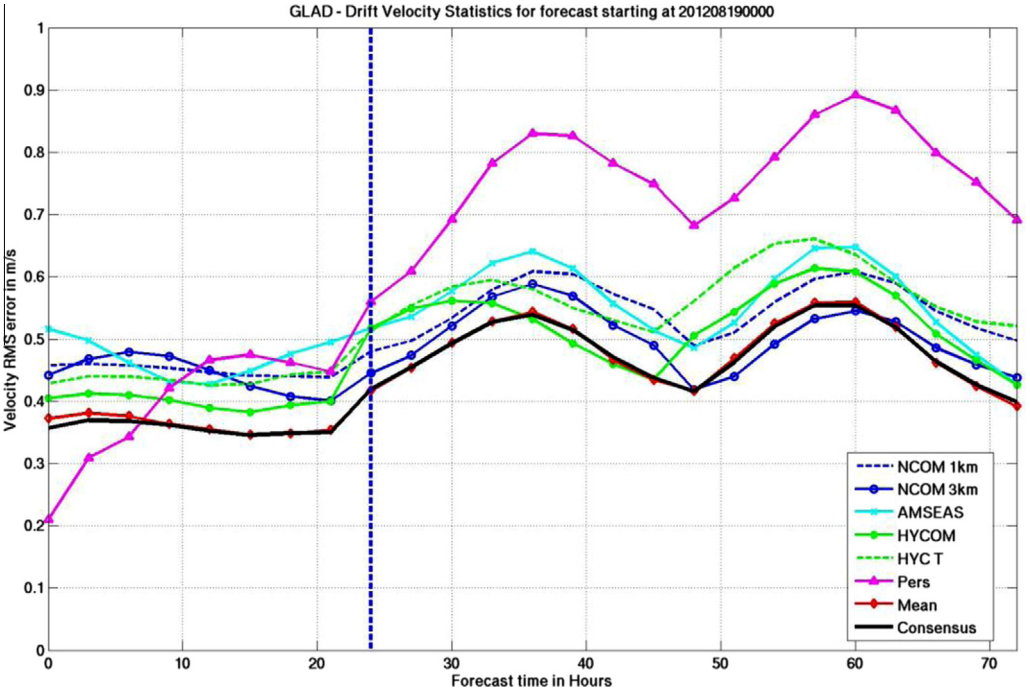

Fig. 10A. Root-mean-square surface velocity error using the full drifter network and 6 h time bins for the forecast run on August 19, 2012 at 00:00 UTC (day 30) before the Tropical Storm Isaac crossed the region. This cycle used the several runs from August 19 00:00 UTC as the model backgrounds and the drifter position during the period August 19–20 to compute the optimal posterior weights using the MEKF. As such the first 24 h (to the left of the blue vertical dotted line) the thick black line shows the MEKF residual analysis errors while the values to the right are true MEKF forecast errors. For reference, the velocity persistency error assuming the velocity observed by each platform at the analysis time remained constant is shown as the red dotted line. The other curves correspond to the different models used by the filter.

used by the MEKF as the background. As reference, to compare results relative to a trivial solution, the figures and tables also display the performance relative to a persistency estimate defined as if the first measured velocity by each drifter and assumed to persist throughout the full 72 h range. The tables also show the overall results when considering the LSF solution only. These are not

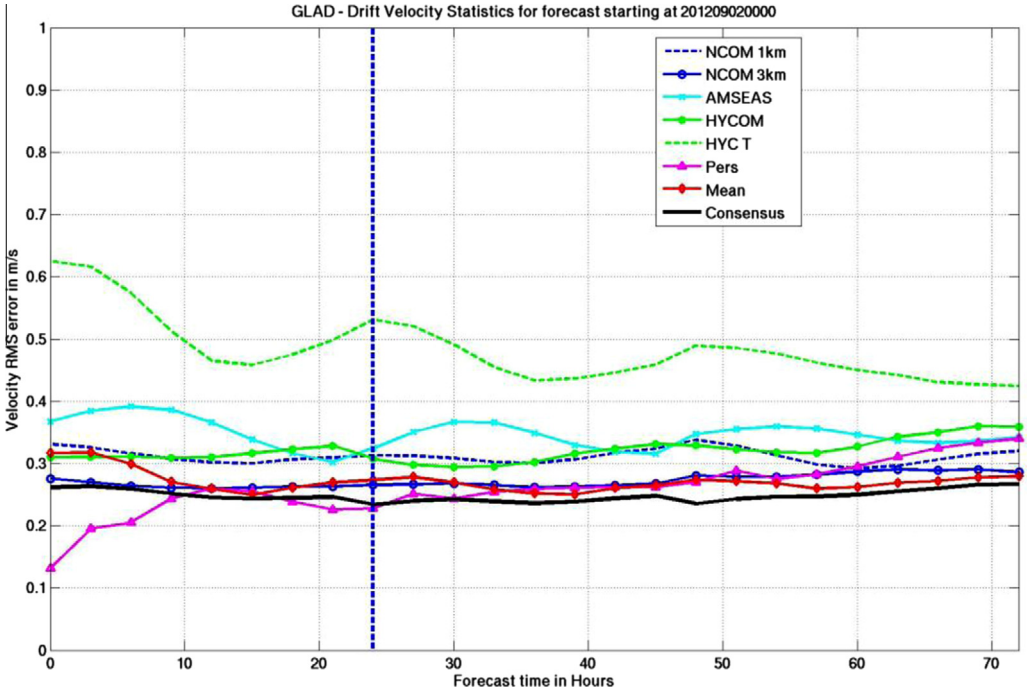

Fig. 10B. Same as Fig. 10A for the run on September 2, 2012 (day 45), a few days after the landfall of Tropical Storm Isaac in August 28–29 2012.

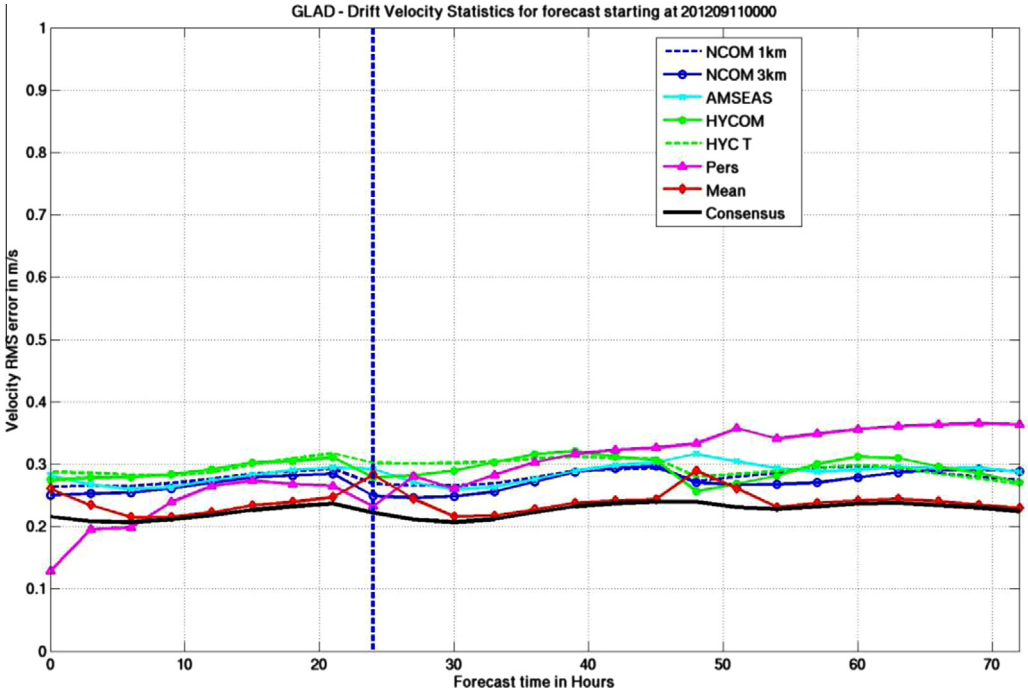

Fig. 10C. Same as Fig. 10A but for the MEKF run using the model estimates from September 11, 2012 (day 54), well after the Tropical Storm Isaac had left the region.

includes in the figures but the corresponding curves were mostly between the MEKF consensus and the mean. Note that the first 24 h in each cycle corresponds to an analysis for the MEKF consensus estimate, as such one should expect it to perform better than any of the other estimates that were not using the observations during this period. Figs. 10A–10C refer to single daily runs corresponding to the dates used in Figs. 4A–4C. They represent the performance of the MEKF during single days of the experiment and highlight the three different dynamical regimes explained above with the drifters covering different regions and with different overall spread.

Fig. 10A shows the comparisons of the model velocities with the corresponding observed drifter velocities using the forecast run in August 19 (thus covering the period August 19 00:00 UTC–August 22 00:00 UTC 2012). This period immediately followed the passage of an atmospheric frontal system across the northeastern Gulf. The

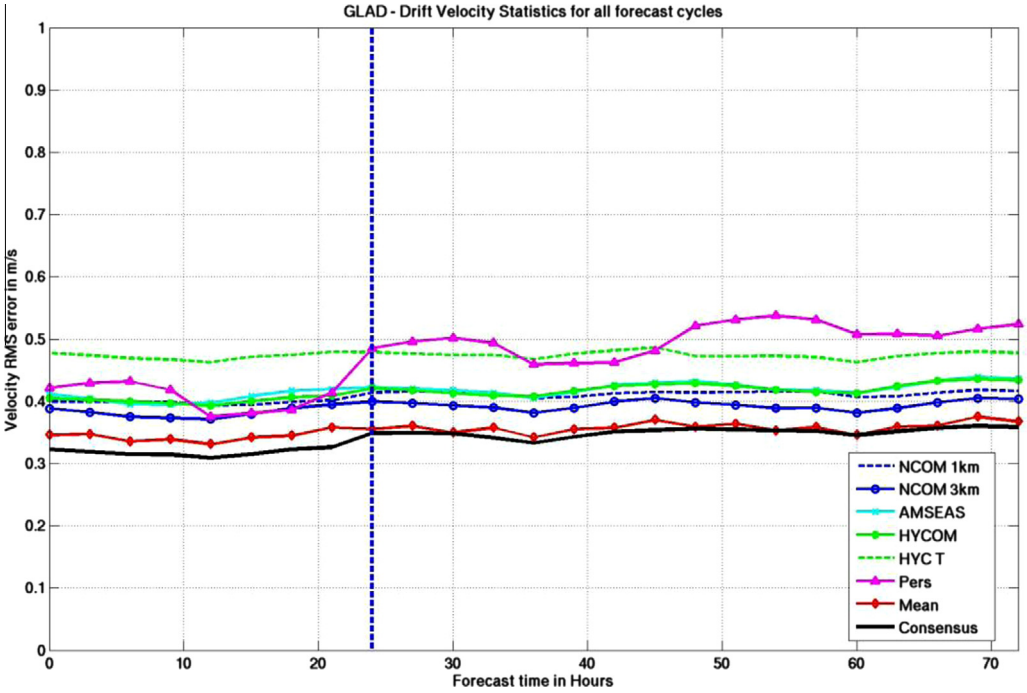

Fig. 10D. Same as Fig. 10A, using the full range of forecast cycles from July, 20 to September, 30 2012. The RMS errors for this figure were computed over 6 h bins that include the full set of measured vs. modeled velocities along the drifters trajectories in all cycles for the same forecast bin period.

atmospheric front triggered inertial residual motions in many drifters along the shelf and coastal areas, as can also be seen in Figs. 3, 5 and 7. During this period most of the drifters were still grouped in their deployment clusters and followed the dynamical features where they were launched as described in Section 2. The results show a highly variable response in time and that the MEKF consensus was only slightly outperforming the ensemble mean. During this period all models were showing similar fits to the observations. The persistency performance shows significant oscillations suggesting this period was mainly dominated by a strong inertial motion.

Fig. 10B shows the equivalent comparisons using the forecasts run in September, 2. Hurricane Isaac passed through the region shortly before this period. It made landfall in the central Gulf August 28. The storm increased drifter velocities (which reached values above 2 m/s), and caused several drifters to ground in shallow waters along the Louisiana, Mississippi, and Alabama coasts. The winds of this storm also wiped out most of the residual inertial motion that were set earlier around August 20–22 and before as we can see in the examples in Figs. 4B and 7. In this example the model HYCOM with tidal forcing (HYC-T) showed an RMS error that was above the others likely due to the different handling of the data assimilation on this particular cycle. This situation when one or more models show degraded skills is when the MEKF consensus produces a significantly more accurate solution compared to the ensemble mean as visible in the figure.

Finally, Fig. 10C shows the comparisons for the forecast beginning September 11, well after Hurricane Isaac. During this period the drifter network was widely dispersed but strongly constrained along what seem to be well developed frontal systems through the central and eastern Gulf. Overall during this period all models have much smaller and stable RMS errors than the previous dates. Nevertheless, the MEKF consensus provides a more accurate solution during the full forecast period.

The overall RMS errors using the full set of forecast runs are shown in Fig. 10D and in Tables 1–3. They concur with the single days analysis discussed in the previous paragraphs, confirming that the MEKF produces a consensus analysis that outperforms on average the entire model set on both the hindcast and forecast periods as well as their mean and the individual Least Square Fits (LSF row in Tables 1–3). The smaller residual RMS errors of the MEKF consensus, estimated during the first 24 h, persist during the remainder of the forecast cycles, giving an indication of the robustness of the method, even during extreme events. These results show that the posterior weights computed by the MEKF using the observation made during the first 24 h were still producing better results during the true forecast periods (24–72 h). This result concurs with the hypothesis stated in Section 3 (i.e. the optimal weights computed by the MEKF during a small portion of the dynamical trajectories persist during the following forecast period).

Fig. 10D and Tables 1–3 also allows us to compare consecutive forecast cycles. The forecast delivered by the MEKF for the ranges 24–48 h or 48–72 h, based on 24- or 48-h old simulations and performing the non-intrusive analysis of the drifter data, were better on average than the estimates given by the updated runs over the range 0–24 h after the intrusive assimilation of the routine data. These more recent runs were assimilating new observations (profile and remote sensing data). None of the models was assimilating the drifter observations but Carrier et al. (2014) discusses tests on a version of the NCOM 3 km assimilating the drifter velocity data that showed improved model skill.

5. Summary and conclusions

The Multi-Model Ensemble Kalman Filter (MEKF) produces optimal aggregation estimates that combine multiple operational models and local observations into an improved forecast. These estimates can be used for on-scene forecasting or as background states for additional intrusive data assimilation Results have shown the approach to improve the consistency of surface velocity estimates with turn-around times compatible with operational applications.

The MEKF method was used to track the velocity data from more than 300 surface drifters deployed during the GLAD experiment in the GoM, using an ensemble of five models with different setups. Results showed the MEKF produced a robust consensus analysis that on average outperformed any of the individual model runs, their mean or model linear regressions to observations for the entire forecast cycles, even during extreme events. However, results showed the method still had limitations in reproducing drifter tracks as dynamical trajectories, producing residual cross drifter track velocity components comparable to those estimated by individual runs. Comparison of consecutive run cycles show the MEKF forecasts from 24- or 48-h old simulations performed better than newly updated runs using the standard data assimilation suggesting the optimal weights computed during a 24 h time window persisted throughout the full range of the available forecasts.

Although the tests were focused on the velocity estimates, the method can also be used for different variables or include other parameters that could be directly correlated with the ocean state and for which a forward model exists (e.g. oil spill images).

The results discussed in this paper can be viewed as part of a framework that allows multi-scale multi-model assimilation of observations into ocean models. The approach starts by correcting the large-scale and low-frequency features in lower resolution outer nests. These corrected estimates then force inner higher resolution nests through the boundaries. The smaller scales reproduced by these inner nests can be further corrected internally through higher resolution analysis and so forth. When we get to a limit when the available observations are not sufficient to constrain the higher resolution or whenever there will be variables with operators that cannot be directly be used by the operational data assimilations systems, a final local zoom-in into the area and variables of interest can then be performed using the non-intrusive MEKF approach.

The MEKF method can expand the use of high resolution local observations, combining several assimilation methods and scales, to improve the local accuracy in the areas and variables of interest by aligning the model runs with the detected dynamical trajectories. This "telescopic" forecast and assimilation approach allows dynamical features or instabilities as detected by the data to be sequentially projected onto the scales reproduced by each domain resolution. These can then be evolved and projected onto the smaller grids to correct new significant dynamical modes consistently with the observed scales, while keeping coherence at the boundaries through the analysis of the same data sets.

Future work will use and discuss this methodology for the multivariate problem processing state and non-state variables to evaluate the impact of other measured variables like temperature and salinity profiles or satellite imagery on the prediction of surface velocities.

Acknowledgments

This research was made possible in part by a Grant from BP/The Gulf of Mexico Research Initiative, and in part by the US Office of Naval Research grant N0003913WX02913. This paper is contribution NRL/JA/7320-13-1783 and has been approved for public release.

References

Arulampalam, M.S., Maskell, S., Gordon, N., Clapp, T., 2002. A tutorial on particle filters for online nonlinear/non-gaussian Bayesian tracking. IEEE Trans. Signal Process. 50 (2), 174–188.

Barron, C.N., Kara, A.B., Rhodes, R.C., Rowley, C., Smedstad, L.F., 2007. Validation Test Report for the 1/8 deg Global Navy Coastal Ocean Model Nowcast/Forecast System. NRL Tech Report NRL/MR/7320-07-9019.

Carrier, M.J., Ngodock, H.E., Smith, S., Jacobs, G., Muscarella, P., Özgökmen, T., Haus, B., Lipphardt, B., 2014. Impact of ocean velocity observations inferred from Lagrangian drifter data using the NCOM-4DVAR. Mon. Weather Rev. 142 (4), 1509–1524.

Coelho, E.F., Rixen, M., Signell, R., 2005. NATO Tactical Ocean Modeling System: Concept Applicability. NATO Undersea Research Centre-Italy Serial Report, SR411.

Coelho, E.F., Peggion, G., Rowley, C., Jacobs, G., Allard, R., Rodriguez, E., 2009a. A note on NCOM temperature error calibration using the ensemble transform. J. Mar. Syst. 78, S272–S281, Rapid Environmental Assessment special issue.

Coelho, E.F., Rowley, C., Jacobs, G., 2009. Ocean Data Assimilation Guidance Using Uncertainty Forecasts, MTS/IEEE Oceans'09, 090611-014.

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

Davis, R.E., 1985. Drifter observations of coastal surface currents during CODE: the method and descriptive view. J. Geophys. Res. 90 (C3), 4741–4755.

Ide, K., Courtier, P., Ghil, M., Lorenc, A., 1997. Unified notation for data assimilation: operational, sequential and variational. J. Meteor. Soc. Jpn. 75, 181–189.

Lenartz, F., Mourre, B., Barth, A., Beckers, J.M., Vandenbulcke, L., Rixen, M., 2010. Enhanced ocean temperature forecast skills through 3-D super-ensemble multi-model fusion. Geophys. Res. Lett. 37, L19606.

Lermusiaux, P.F.J., Robinson, A.R., 1999. Data assimilation via error subspace statistical estimation. Part I: Theory and schemes. Mon. Weather Rev. 127 (7), 1385–1407.

Leslie, W.G., Robinson, A.R., Haley, P.J., Logotouv, O., Moreno, P.A., Lermusiaux, P.F.J., Coelho, E.F., 2008. Verification and training of real-time forecasting of multi-scale ocean dynamics for maritime rapid environmental assessment. J. Mar. Syst. 69 (1), 3–16.

Logutov, O.G., Robinson, A.R., 2005. Multi-model fusion and error parameter estimation. Q. J. R. Meteorol. Soc. 131 (613), 3397–3408. http://dx.doi.org/10.1256/qj.05.99.

Lorenc, A.C., 1995. Atmospheric data assimilation. Met Office, Scientific Paper 34, 17 pp.

Lunde, B.N., Coelho, E.F., 2009. Implementations of the Navy Coupled Ocean Data Assimilation System at the Naval Oceanographic Office. MTS/IEEE Oceans'09, 090602–015.

Metzger, E., Hurlburt, H.E., Wallcraft, A.J., Chassignet, E.P., Cummings, J.A., Smedstad, O.M., 2006. Global ocean prediction using HYCOM. HPCMP Users Group Conference HPCMP-UGC'06 proceedings, 0-7695-2797-3/06, IEEE.

Ngodock, H.E., Smith, S.R., Jacobs, G.A., 2007. Cycling the representer algorithm for variational data assimilation with a nonlinear reduced gravity ocean model. Ocean Model. 19 (3), 101–111.

Olascoaga, M.J., Beron-Vera, F.J., Haller, G., Trinanes, J., Iskandarani, M., Coelho, E.F., Haus, B., Huntley, H.S., Jacobs, G., Kirwan Jr., A.D., Lipphardt Jr., B.L., Özgökmen, T., Reniers, A.J.H.M., Valle-Levinson, A., 2013. Drifter motion in the Gulf of Mexico constrained by altimetric Lagrangian coherent structures. Geophys. Res. Lett. 40 (23), 6171–6175.

Poje, A.C., Özgökmen, T.M., Lipphart Jr, B., Haus, B., Ryan, E.H., Haza, A.C., Jacobs, G., Reniers, A.J.H.M., Olascoaga, J., Novelli, G., Griffa, A., Beron-Vera, F.J., Chen, S., Hogan, P., Coelho, E., Kirwan Jr, A.D, Huntley, H., Mariano, A.J., 2014. Submesoscale dispersion in the vicinity of the deepwater Horizon spill. Proc. Nat. Acad. Sci. 111 (35), 12693–12698.

Poulain, P.-M., 1999. Drifter observations of surface circulation in the Adriatic Sea between December 1994 and March 1996. J. Mar. Syst. 20, 231–253.

Rixen, M., Coelho, E.F., 2007. Operational surface drift prediction using linear and non-linear hyper-ensemble statistics on atmospheric and ocean models. J. Mar. Syst. 65, 105–121.

van Leeuwen, P.J., 2009. Particle filtering in geophysical systems. Mon. Weather Rev. 137, 4089–4114.

Wei, M., Jacobs, G., Rowley, C., Barron, C.N., Hogan, P., Spence, P., Smedstad, O.M., Martin, P., Muscarella, P., Coelho, E.F., in press. The performance of the US Navy's RELO ensemble, NCOM, HYCOM during the period of GLAD at-sea experiment in the Gulf of Mexico. Deep Sea Res. II http://dx.doi.org/10.1016/j.dsr2.2013.09.002.