



A note on NCOM temperature forecast error calibration using the ensemble transform

Emanuel Coelho ^{a,b,*}, Germana Peggion ^{a,c}, Clark Rowley ^a, Gregg Jacobs ^a, Richard Allard ^a, Elaina Rodriguez ^c

^a Naval Research Laboratory, Stennis Space Center, MS, USA

^b University of Southern Mississippi, MS, USA

^c University of New Orleans, Louisiana, LA, USA

ARTICLE INFO

Article history:

Received 12 June 2008

Received in revised form 24 July 2008

Accepted 22 January 2009

Available online 28 February 2009

Keywords:

Ocean ensembles

Forecasting

Ocean models

Forecasting errors

Environmental assessment

(43–45N, 9–10.5E)

ABSTRACT

During the MREA07 trial, off the NW coast of Italy in the late spring and summer of 2007, Navy Coastal Ocean Modeling (NCOM) multiple nests free-run ensembles were made available in real-time for the LASIE07 and BPO7 events and a fairly complete set of observations were collected inside the inner nests domains. This note addresses the problem of predicting NCOM local unbiased 0–24 h forecast errors by perturbing a limited number of possible error sources through Monte-Carlo simulations, without local data assimilation. It discusses preliminary results using the Ensemble Transform [Bishop, C.H., and Toth, Z., 1999: Ensemble transformation and adaptive observations. *Journal of the Atmospheric Sciences*, 56, 1748–1765] to calibrate the ensemble spread by adjusting its characteristics (spread–skill relationship and magnitude) to an observed or pre-estimated error field. A small (10 members) ensemble of free runs was used for water column temperature forecast Root Mean Square (RMS) error prediction. After being post-processed they were compared with observed errors and those estimated using time variability as an error proxy. The ensemble runs were generated through atmospheric forcing perturbations using the space–time deformation method as proposed by [Hong, H.X., Bishop, C., 2007. Ensemble and probabilistic forecasting. IUGG XXIV General Assembly 2007, Perugia, Italy, 2–13 July], keeping independent initial conditions. Because at the starting time all runs shared the same IC, the ensemble was run for roughly two weeks for spinning up and then used during the following 10 days for data comparisons, during which the ensemble spread did not diverge and was consistent with the observed dynamics. Comparisons of ensemble spread of temperature profiles with local observed errors and time variability (assumed as an error proxy) showed that they were consistent through this 10 day analysis period, with performances above the non-calibrated ensemble estimates and time-variability used as error proxy.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

When considering numerical prediction of ocean dynamic states using nested domains, several sources of error can contribute to cascading uncertainty into state variable estimation (Coelho and Rixen, 2008). These sources of error include the errors of the initial and lateral boundary conditions, local forcing, bathymetry errors, numerical approximations and filtering, errors due to approximations when assimilating observations, errors in the forcing terms and unresolved scales (sub-grid variability). To address this problem, local unbiased (correlation) and persistent errors (bias) of the Navy Coastal Ocean Modeling (NCOM) System nested in global ocean domains, are typically reduced and monitored by assimilating dynamical balanced analysis fields of state variables, derived from observation networks, using the Navy Coupled Ocean Data Assimilation (NCODA) system (e.g., Cummings, 2005). This system also provides an error estimate of these analysis fields at an analysis time.

In recent implementations (Coelho and Rixen, 2008; Fabre et al., 2008), ensemble based stochastic methods have been used to track these NCOM analysis multi-scale ocean errors by running the model several times using different forcing and starting from different initial conditions. The resultant ensemble spread was constrained at each new analysis time by the new estimate of the analysis errors using a technique named Ensemble Transform (ET) (Bishop and Toth, 1999). In order to be accurate, the perturbed ensemble members should be taken from a fairly large number of independent runs to resolve state variables error covariances and should include all significant sources of error and uncertainty (Judd et al., 2007; Lermusiaux et al., 2006). Since this is not easy to obtain in operational timeframes, and once a smaller number of runs are selected, one can expect the ensemble to perform differently inside the simulation domain and through time depending on the number of the dominant error modes. This limitation motivates on-going work in developing dedicated metrics to diagnose and prognoses ensemble performances through the overall domains and forecasting lead times.

In any case, it is anticipated that a small number of runs may still provide useful information under certain conditions (e.g. when there are no strong non-linearity and bias errors are on the same order of

* Corresponding author. Naval Research Laboratory, Stennis Space Center, MS39529, USA. Tel.: +1 228 688 5710.

E-mail address: emanuel.coelho.ctr.po@nrlssc.navy.mil (E. Coelho).

magnitude of the correlations errors). Furthermore, if the ensemble estimates define a domain that contain the most relevant features and scales of the physical system, then they can be improved in their consistency through calibration and post-processing by adjusting their spread and bias to some training sequence. These methods have been successfully used for meteorological ensemble calibration (e.g. [Doblas-Reyes et al., 2005](#); [Hamill and Whitaker, 2007](#)) and for multi-model ocean ensembles applications (e.g. [Rixen et al., 2008](#); [Coelho, 2008](#)).

It should be noted that with a small number of independent runs we should not expect to resolve the full ocean state covariances with the original model grid resolution, but one can expect that a small number of runs between 10 and 15 may still be adequate to track single variable forecast errors on a re-sampled spatial domain as long as the number of independent variables can be kept within the order of $O(10^3)$, following the estimates of [Judd et al. \(2007\)](#). This note will discuss the limitations of a small ensemble size used during the MREA07 trial and proposes a method to improve forecast error prediction consistency for specific target variables, applicable also for non-state variables estimates when there are not many observations or prior to use observations into the assimilation process.

Several methods have been used to perturb the initial conditions fields based on the observed errors. In particular [Bishop and Toth \(1999\)](#) proposed a technique named Ensemble Transform that allows computing dynamically balanced initial conditions perturbations that are consistent with a best estimate of the error covariance. On the other hand, ensemble calibration can also be sought through post-processing using Bayesian methods (e.g. [Gneiting et al., 2004](#), [Coelho et al., 2005](#) and [Rixen and Coelho, 2006](#)), within the limits of the known cross-correlations among the observed and modeled variables. This work combine both techniques as a post-processing method, applied to local single variable ensemble spread calibration. The methodology uses the perturbed model statistics re-scaled through an estimate of the error variance, to obtain short-term estimates of posterior normal probability distributions envelopes of a selected ensemble variable.

The MREA07 (BP07 and LASIE trials), took place off La Spezia, Italy in the spring and summer of 2007 (e.g., [LeGac and Hermand, 2007](#)). During the trial, mesoscale relocatable NCOM implementations using the RELO system were made available in real-time without performing local data assimilation, though remote sensing and global data was assimilated on the outer nests used for boundary conditions and initialization. In standard implementations the RELO system runs together with the Navy Coupled Ocean Data Assimilation (NCODA) system that performs observations quality control and produce local analysis for assimilation that in the present version are based on a Multi-Variate Optimum Interpolation technique (e.g. [Cummings, 2005](#)). NCODA also provides the analysis error fields that are used to re-set the ensemble spread of the initial fields in operational ensemble runs using the same ET technique (e.g. [Fabre et al., 2008](#)). This present solution does not provide reliable analysis error covariances but it is planned that the NCODA system will evolve in the near future into using hybrid Monte-Carlo ensembles (e.g. [Lermusiaux et al., 2006](#)) and Variational analysis (e.g. [Ngodock et al., 2007](#)). This will improve error covariance estimates and produce analysis fields consistent with the boundary conditions and other forcing fields. For this specific implementation, the NCODA assisted assimilation process in the inner nests was turned off to allow a fully independent analysis of the model results and observations, simulating a scenario where no local data would be available in useful timeframes.

During this trial the free-run error fields of the RELO system were estimated using an ensemble of 10 independent runs with independent initial conditions starting from a common field far back in time and perturbed through atmospheric forcing using space-time deformation of the surface forcing fields ([Hong and Bishop, 2007](#)). The ensemble spread of the free runs was then re-scaled in post-processing through an Ensemble Transform ([Bishop and Toth, 1999](#)) using the temporal variability as an error proxy. These preliminary

error estimates were then used for model benchmarking and aiming specific ocean-acoustic applications (e.g. [Carriere et al., 2009](#)) and to estimate the relative impact of different observational strategies ([Coelho et al., 2007](#)).

2. RELO-NCOM setup

The Relocatable Navy Coastal Ocean Model (RELO-NCOM) is a scalable, portable, and user-friendly system for nowcasting and short-term (2–3 day) forecasting simulations. There are two major components: 1) NCOM ([Martin, 2000](#)) and 2) the Navy Coupled Ocean Data Assimilation (NCODA) ([Cummings, 2005](#)) for data analysis and model initialization. For a rapid configuration, the system relies on a set of data and products available on a global scale (bathymetry, winds, analysis of the remote sensing data). These products are generally on a low resolution and it is possible to substitute them with local and high-resolution databases. RELO-NCOM meets the naval requirements to generate real-time description of the environmental variables and it is operational at the US Naval Oceanographic Office (NAVO).

There is a fundamental difference between assessing an ocean model configuration in a research and an operational mode. Both need to be designed, calibrated, and evaluated to encompass the dominant dynamics of a given region. The goal is to provide the best possible representation of the dynamical features of a specific area. However, a predictive system that supports operational applications must be rapidly relocatable anywhere in the ocean (oil-spill response and naval operations are the most relevant applications), and easily reconfigured. The principal goal is to provide good representations everywhere with the available data (i.e., in spite of the absence of complete sets of observations), motivating the need to associate with the system a reliable error diagnostics and prediction tool, to allow tracking consistently the error dynamics.

For the MREA07 trial the RELO-NCOM was deliberately set on its default mode as for a generic application with little or no tuning of the physical and numerical parameters. Furthermore, no MREA07 or other data were assimilated into the inner nests. The goal of this implementation was to test the modeling skills of these free runs and estimate the relevance of the atmospheric forcing as a single source of error.

The daily predictive cycle during MREA07 is described as follows:

- NCOM is started from the previous day's nowcast (–24 h) and forced by the available operational winds. Open Boundary Conditions (OBC) are extracted from the simulation of the parent domain. The OBC for the outer most nest are extracted from NCOM configured on a global scale at a $1/8^\circ$ resolution (NCOM-GL) which is operational at the Naval Oceanographic Office (NAVO) (http://www7320.nrlssc.navy.mil/global_ncom/index.html) ([Barron et al., 2006](#)). However, this procedure is not restricted to NCOM–NCOM nesting; any nest could be coupled with several other dynamical model formulations.
- During the nowcast, temperature (T) and salinity (S) fields are nudged to the nowcast fields of the parent simulations. The nudging during the hindcast phase has been suggested to provide a minimum connection with real-time data since NCOM-GL assimilates sea surface temperatures (SST) and Modular Ocean Data Assimilation System (MODAS) synthetics (with the surface height derived from the Naval Layer Ocean Model (NLOM) (http://www7320.nrlssc.navy.mil/global_nlom/)). No data are nudged after the nowcast (0 h).
- A short-term (2-day) forecast is provided. The 48-hour interval has been chosen because this is the typical period in which meteorological mesoscale forecasts are available and reliable.
- The nested domains run then in sequence using boundary conditions from the outer nests (i.e., one way nesting). Although NCOM provides a tile nesting approach, the default procedure

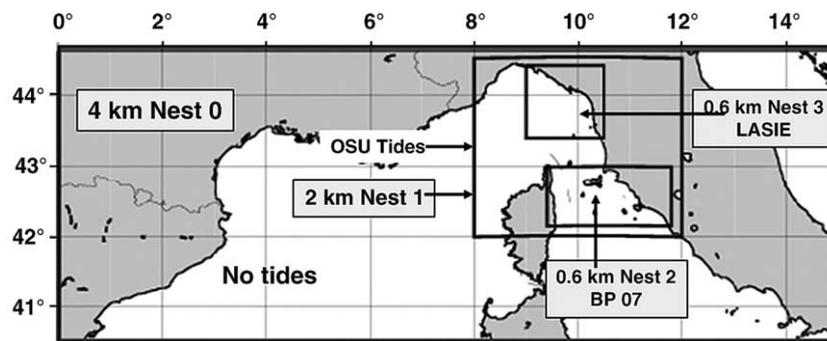


Fig. 1. The triple nest configuration for MREA_07.

allows an easy and rapid configuration and assessment of each domain, and more importantly, a possible different choice of the vertical coordinate between nests. Fig. 1. illustrates the triple nested configuration for the MREA07 exercise.

In this model configuration, all domains are forced with the Coupled Ocean Atmosphere Mesoscale Prediction System (COAMPS^{®1}) Europe-2 winds (27 km) (Hodur, 1997) and heat fluxes from 0.5° Navy Operational Global Atmospheric Prediction System (NOGAPS, Rosmond et al., 2002). Monthly river discharges are extracted from the global river data set of NCOM-GL (Barron and Smedstad, 2002), with the Arno, Magra, and Serchio transports provided by the Istituto Idrografico Italiano. The vertical resolution of each domain has 38 σ - and 7 z -levels (45 levels). The outer nest (nest0) is at 4 km horizontal resolution with the primary purpose of serving as a buffer zone between NCOM-GL's NOGAPS forcing and the higher resolution wind data set. Nest 1 (2 km resolution) includes tides. Tides are specified at the boundaries from the Oregon State University tide model (Egbert and Erofeeva, 2002). Nest2 and nest3 are at about 0.6 km resolution and configured for the BP07 (Elba) and LASIE07 (LaSpezia) domains, respectively. An ensemble of 10 independent runs of the inner nests was also made available in real-time, using similar set-ups but with perturbed atmospheric forcing using the space–time deformations method (Hong and Bishop, 2007).

One of the most pressing issues of real-time operational forecasting is to provide the information in a timely manner. Ocean forecasts are usually one of the final components of a long string of products developed at several different centers: a delay in acquiring one of the input data (e.g., winds, boundary conditions), the classic computer breakdowns (just to mention a few issues) may create a domino effect and ultimately a late delivery of the forecast. In order to avoid delays in the queue submission which are often occurring at the supercomputer sites, the full forecast cycle is performed at the Naval Research Laboratory – Stennis Space Centre (NRLSSC) on dual processor Optron-based LINUX platforms. The latest NOGAPS and COAMPS analyses and forecasts are usually available at NRLSSC before 1000GMT, but NCOM-GL daily hindcasts and forecasts arrive at about 1130GMT. Therefore, to speed up the delivery of the results, the OBC for nest0 are extracted from the NCOM-GL 72 h forecast of the previous day. This makes it possible to start the simulations at about 1000GMT and complete the forecast cycle usually before NCOM-GL latest files are available at NRLSSC. Unfortunately, only a partial COAMPS data set is archived at NRLSSC, so the price for this procedure is the use of NOGAPS-0.5 heat fluxes.

The model results are written to NetCDF files at user specified z -levels and time increments. It is important that the z -levels be consistent with the NCOM vertical grid. A coarse vertical resolution in the NetCDF files may remove features reproduced by the model; a too fine vertical resolution increases the computational cost and memory requirement

without increasing the physical accuracy of the solutions. For this real-time exercise, the NCOM fields were provided on 47-levels and at a 1 h increment. To reduce the amount of transferred data, only the 48 h forecast (i.e., no hindcast) of the model and only a few upper vertical levels for the ensemble spread were posted on the MREA07 ftp server, generally at about 1230GMT and 1500GMT, respectively.

3. RELO-NCOM control analysis and data comparison

This note will focus on the analysis and discussion for the period June 10 to 25, 2007 and for the nest 3 area only. In this region, dynamics were mostly dominated by a persistent cyclonic gyre centered roughly at 43°40'N and 9°20'W, modulated by smaller re-circulation cells north and east, closer to the coast. The shapes and temperature distributions of these smaller cells were strongly perturbed by the wind forcing. During the “sirocco” south-easterly winds (e.g. 06/19 06:00 snapshot displayed in Fig. 2, left panel) the average surface temperatures were higher, with warmer waters trapped closer to the eastern coast. During the “libeccio” south-westerly winds ((e.g. 06/23 12:00 snapshot displayed in Fig. 2, right panel), the cold eddy signature becomes more noticeable and different re-circulation patterns can be found between the eddy and the coastline.

The Sea Surface Temperature (SST) images obtained from NOAA AVHRR displayed in Fig. 3, although with different resolutions, concur with the analysis of the previous paragraph.

The water column was strongly stratified during the whole period. Model temperature hindcast and forecast estimates were compared with 160 CTD profiles collected during the trial in the period June 4–26, 2007 by three ships in the area (RV Planet, RV Leonardo and NI Galatea). The daily CTDs' covered both deep and shallow water throughout most of the surveying time. For this work only profiles inside the nest 3 domain were used. For each CTD, the nearest (in space and time) hourly model profile was extracted. No horizontal or temporal interpolation is performed on the model or data. Since observations are on a higher vertical resolution relative to model estimates, the model temperature at a specific z -level should be compared with the mean value of the observed values between the intermediate levels up and below (i.e. for the model estimate T_i at level Z_i , observations should be averaged between the levels $(Z_{i-1} + Z_i)/2$ and $(Z_i + Z_{i+1})/2$). The model-data comparisons displayed in Fig. 4 show that temperature errors were more noticeable on average at the bottom of the well mixed layer (at roughly 50 m depth), with the surface waters typically cooler than observations and warmer waters below. Temperature errors were very small below the 200 m depth. It is also noticeable that these error characteristics did not change significantly during the analysis period, though significant changes occur in the forcing and dynamic responses as mentioned above.

From these comparisons one can assume that the prediction skills of the model were limited, not significantly above model persistency, such that these free-run RELO-NCOM fields could be considered as an analysis tool capable of providing reasonable spatial distributions of the

¹ COAMPS is a registered trademark of the Naval Research Laboratory.

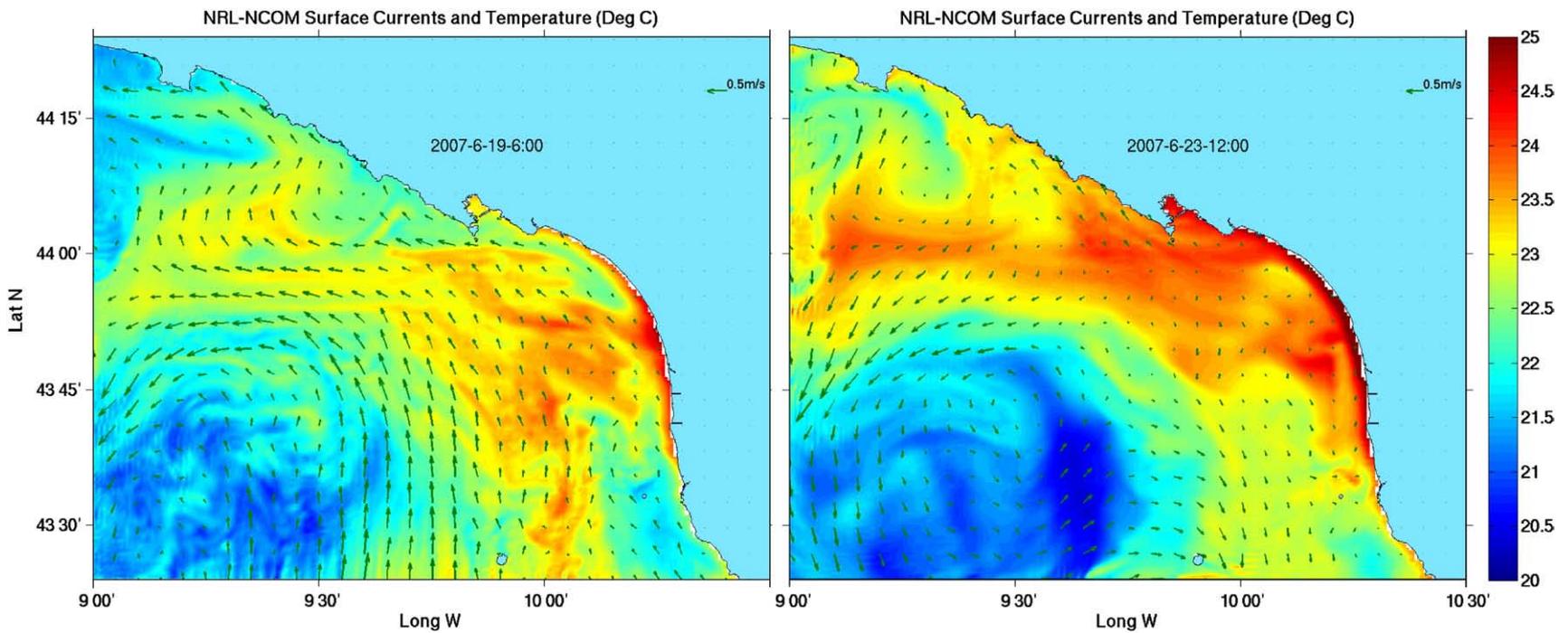


Fig. 2. RELO-NCOM upper layer temperature snapshots for the days 06/19 (left panel) and 06/23 (right panel). The snapshots hours, displayed in the images, correspond to the wind maximum stress for each day. During the 19th winds were predominantly south-easterly (“Sirocco”) and during the 23rd they were predominantly south-westerly (“Libeccio”). Both panels display how flow patterns changes around the persistent gyre in the South-West corner, with warmer waters intruding northward during the “Libeccio” event.

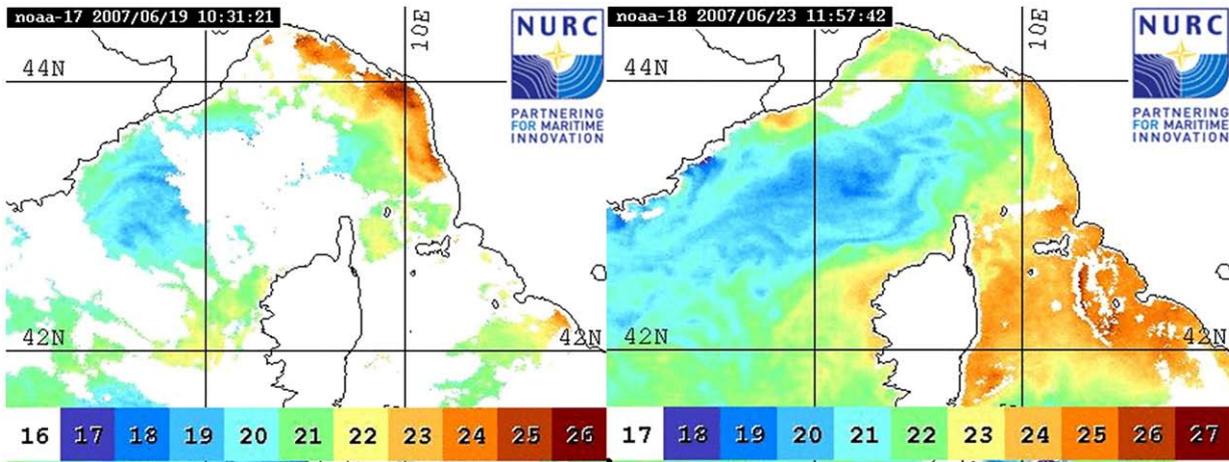


Fig. 3. NOAA AVHRR Sea Surface Temperature estimates for 06/19 (left panel) and 06/23 (right panel). During the 19th winds were predominantly from the south-east (“Sirocco”) and during the 23rd from the south-west (“Libeccio”). Images were produced by automatic processing using NURC TERASCAN software.

temperature fields, up to at least 48 h. This is mostly due to the persistent nature of the dominant local dynamics that did not change significantly during the analysis period. In other more dynamic areas one could expect these free-run errors to increase significantly after a few hours and differences between forecast lead times also to become more noticeable.

Since there were no significant differences between these errors, the discussion below regarding error prediction will use the 0–24 h and 24–48 h temperature forecasts as equivalent estimates.

4. Ensemble re-scaling using the ensemble transform

The ocean is driven by surface fluxes that are determined by the atmospheric state and are one major source of uncertainty. Predicted atmospheric fields often contain the forecast feature of interest, but

they can be misplaced in space and time (e.g. Hoffman et al., 1995). This characteristics motivated the attempts to represent forecast errors in terms of a shift of a forecast in space and time similar to the pseudo-random fields method described by Evensen (2003) and applied in ocean ensemble generation problems (e.g. Demirov et al., 2003). For the present work, the atmospheric forcing perturbations used to force the ocean ensemble members were produced using the method developed by Hong and Bishop (2007). It uses only time shifts of the forecast, with a choice of parameters to provide a good precision in the atmospheric perturbations, though accuracy may not be guaranteed over the whole simulation period.

If we neglect bathymetry, error induced by numerical approximations and other sources of possible model bias, the ensemble transform (ET) method of generating initial perturbations applied in atmospheric

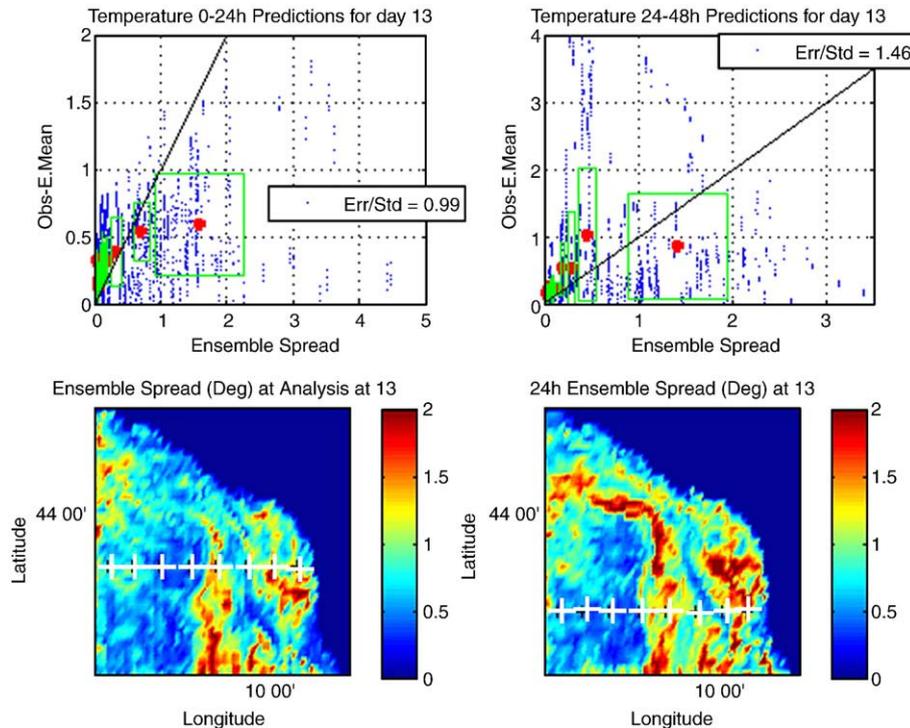


Fig. 4. RELO-NCOM water temperature bias and RMS error estimates. The four panels in the left show the RMS errors along each simulation day (24 h period), using different model estimates compared with the observations. The color plot named “A04” in the upper left uses hindcast atmospheric forcing fields, the plot named “Pers” uses model persistence (hour 0 snapshot) and the plots below named “F24” and “F48” use 24 and 48 h lead forecasts respectively. The four panels in the right show the error bias (24 h mean errors) using the same model estimates.

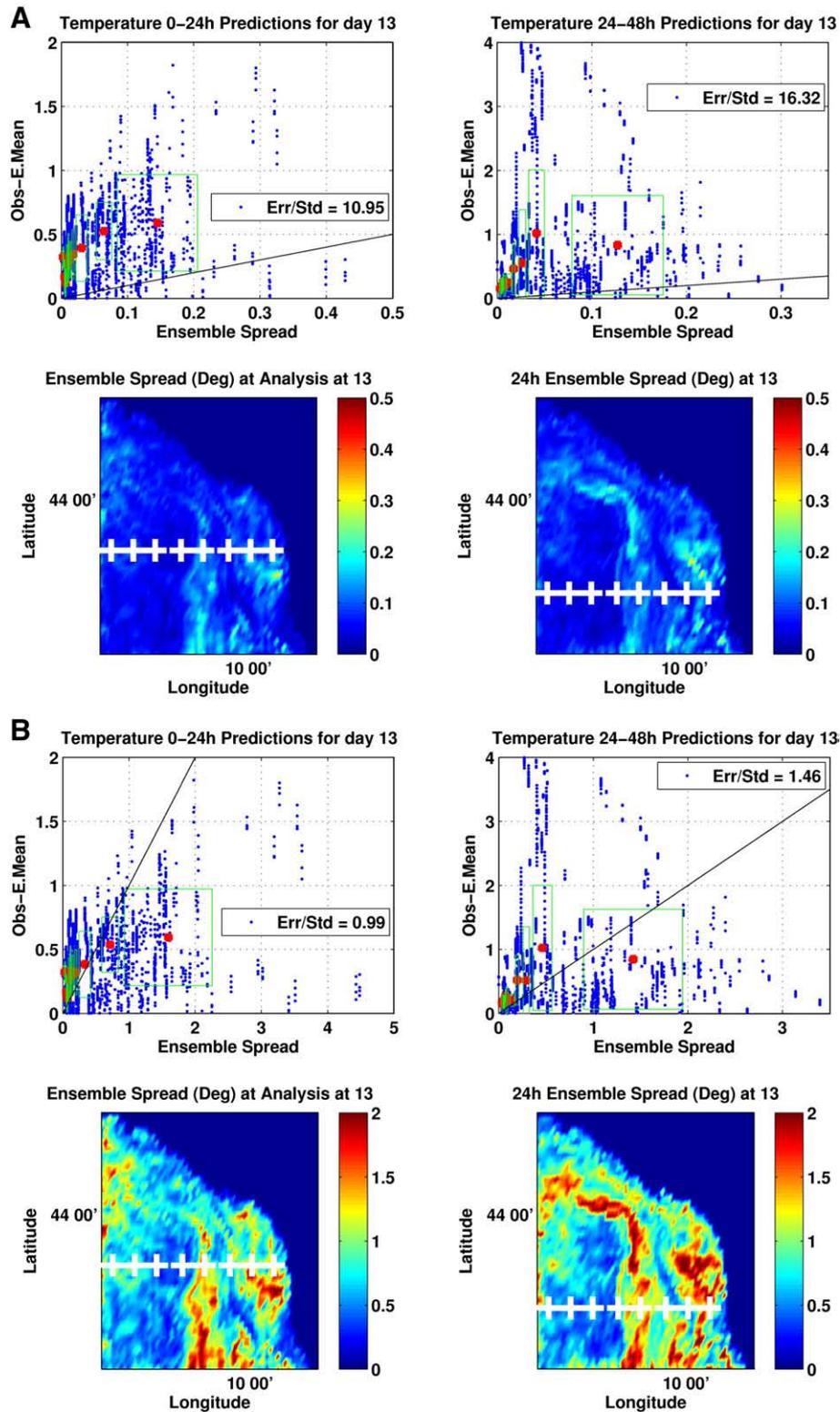


Fig. 5. Error scatter plots computed using the run of June 13. The upper scatter diagrams show the ensemble spread vs. observed forecast error before re-scaling (A) and after re-scaling (B). The forecast errors were computed using the 0–24 h forecasts (panels in the left) and using the 24–48 h model forecasts (panels in the right). The color plot below each scatter diagrams shows the surface temperature error estimate (ensemble standard deviation) at hour 00:00 (left) and 24:00 (right) relative to the simulation day and the white crosses depict the locations used for model-data comparison.

ensemble forecasts (Bishop and Toth, 1999) can be used to re-balance and re-shape the IC fields of the ensemble subset. Besides assuring that all detected error growing modes will be equally represented, the advantage of this technique is such that: it respects hydrodynamic balances by ensuring that initial perturbations are a linear sum of forecast perturba-

tions from the preceding forecast; and ensures that the initial perturbations are equally likely and orthogonal under a measure of the probability of initial condition error based on the best available estimate of initial condition error variance. This technique does not provide though an initial set of background perturbations that need to be introduced using

complementary methods, such as forcing from an ensemble of atmospheric forecasts as mentioned in the previous paragraph.

As detailed in Bishop and Toth (1999), through the ET ensemble generation technique, K forecast perturbations of N state variables $\mathbf{X}^o(N \times K)$, can be transformed into a set of perturbations \mathbf{X}^r that are consistent with the background error analysis covariance \mathbf{P}_g^a , using

$$\mathbf{X}^r = \mathbf{X}^o \mathbf{T}$$

where \mathbf{T} is a transformation matrix determined by the eigenvectors and eigenvalues of the projections of the magnitude of the predicted analysis perturbations on the inverse of the error analysis covariance matrix. If the number of ensemble members equals the number of state variables, this projection guarantees the perturbations covariance to be equal to the error covariance.

Through this transform we can then obtain a set of perturbed fields that are consistent with an independent estimate of the error covariance. In operational implementations these initial fields are used as new initial conditions for the K independent ensemble runs, providing a method to assimilate the observed errors into the ensemble forecasts. For the present application and to use this method in post-processing a persistency assumption during the 48 h forecast cycles is taken, regarding the projection of the ensemble covariances into the observed errors.

5. MREA07 error predictions

For the present application since no data are to be used the ET is computed using the temperature 48 h forecast time variances, as estimated by the RELO-NCOM free runs, producing a diagonal error covariance matrix \mathbf{P}_g^a . Besides allowing for a faster transform, this approach allows keeping the shapes of the off-diagonal terms (spatial cross-correlations) as estimated by the ensemble, while consistently re-scaling the analysis errors, without introducing further analytical or numerical approximations.

The temperature estimates ensemble spatial correlations are then updated only by the RELO-NCOM independent runs. This method allows keeping error covariance updates, without the cost of computing and inverting very large matrices. Furthermore, since only a limited number of ensemble members are available, this method limits the growth of spurious cross-correlations. The same transform matrix \mathbf{T} is applied to all time steps of the ensemble estimates.

The resulting ensemble spread (standard deviations) for each temperature estimate is then compared against the absolute value of the RELO-NCOM vs. data mismatches and displayed in scatter diagrams as those shown in Fig. 5 for days Jun 13 and 14, before and after applying the ET. The statistical significance of each of these individual estimates (small blue dots) is negligible, such that they are grouped in equally populated bins with 1000 elements, defined along the ensemble spread axis. These bins displayed inside the scatter diagrams as large red dots will have similar likelihoods and will be statistically relevant. For the ensemble to be accurate, bins should be aligned along the main diagonal, highlighted as a black line on the plots. The green rectangles around the bins show the standard deviations of each bin along each axis (error and ensemble spread). Other relevant statistic is the mean ratio between measured error vs. ensemble spread, (Err/Std in the figures) that should be close to 1 for the ensemble to be accurate.

The graphics in Fig. 5 left of the black line show the scatter diagrams for days 13 (left upper plot) and day 14 (right upper plot) computed from the ensemble before post-processing. From the bin distribution we can see the ensemble to have a positive spread-skill relationship, through all ranges of the observed errors, such that estimates of smaller ensemble spread are well correlated with smaller errors and estimates of larger error are well correlated with the larger errors, through all ranges of observed errors. However, we can see that the ensemble was grossly under-predicting the magnitudes of the observed errors in roughly one order of magnitude. This is most likely due to the fact the initial fields and other major sources of error besides atmospheric forcing were not being properly perturbed.

The data of June 13 were used as the initial day to start the procedure and adjust the ensemble spread to the observed error. For

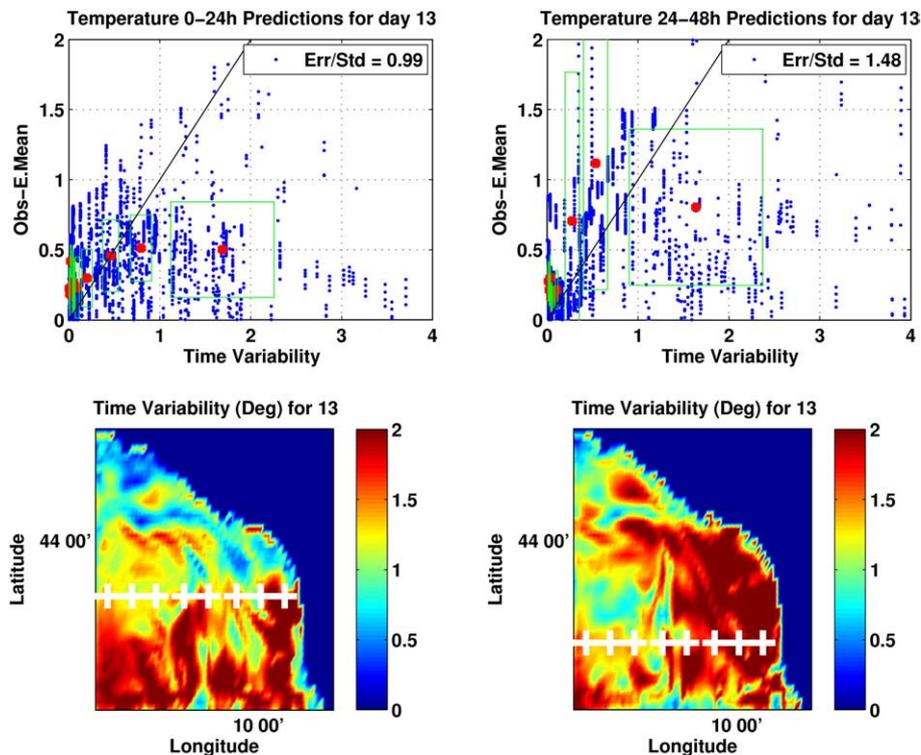


Fig. 6. Same as in Fig. 5-B, but using the time variability as an error proxy instead of the ensemble spread as an error estimate.

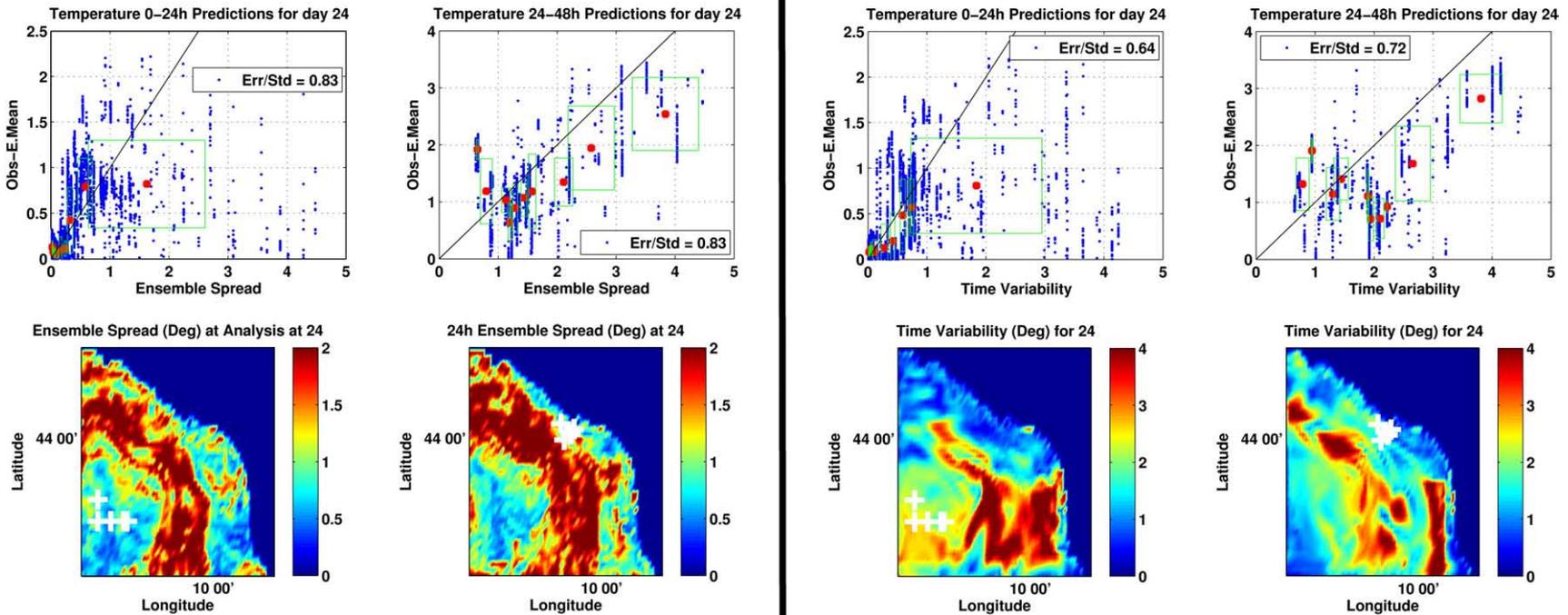


Fig. 7. Same results as described for Fig. 5-B (on the left) and Fig. 6 (on the right) but for the model run of June 24. The panels left of the vertical line show the results using the calibrated ensemble. Panels in the right show the same results but using the time variability as an error proxy.

this purpose, a multiplication factor of 4 was estimated from the data and applied to the temporal standard deviations used to compute the ET throughout the simulation period. This value was estimated iteratively in order to bring the ratio Err/Std from a value of 11 before the transform to 1. As a result, the red bins also became closer to the main diagonal as we can see on the scatter diagrams right of the vertical black line in Fig. 5. For the following day represented by the 24–48 h forecast this ratio increased slightly to 1.5, though the bins remained close to the main diagonal.

Other relevant result from Fig. 5 is the spatial distribution of the error estimates. In the lower color maps one can see the ensemble spread at the surface for days 13 and 14. The black crosses show the points where data was collected during those days respectively. One can see that the spatial patterns were not strongly changed by the transform and the areas with larger estimated errors are shaped along the boundaries of the persistent cyclonic eddy in the SW portion of the domain as one could expect. The sampling locations during these two days included several runs across the boundaries of this cyclonic gyre.

Since the ET was using the temporal standard deviation to re-scale the ensemble spread one could argue that the information contained in the ensemble would be erased and time variability would be the dominant error proxy. In order to evaluate this hypothesis the same scatter diagrams were computed using the temporal standard deviation instead of ensemble spread, as displayed in Fig. 6. To keep an equivalent accuracy a multiplication factor of 7.8 was also applied to set the ratio Err/Std to 1 for the day 13 data. From the scatter diagrams one can see that this error proxy keeps similar positive spread–skill relations, though the spatial distribution of errors is significantly different from those estimates by the ensemble and not so well correlated with the dominant dynamics.

Using the tuning parameters estimated for day 13, one can estimate the ensemble spread and the time-variability error proxy for the following forecast days. Since observations were made until June 25, Fig. 7 displays the same diagrams for the last two days of June 24 (0/24 h in the labels) and 25 (24/48 h in the labels) when model-data comparisons were possible. The four plots panel in the left shows the results using the transformed ensemble and the panel in the right shows the same results using the time-variability proxy. One can see that the ensemble spread was kept consistent with the dynamics and the performance of both the transformed ensemble and time variability as error proxy seems close in performance. However, looking to the spatial distribution of the predicted surface temperature errors as displayed in

the lower color maps for days June 24 and 25 one can see that the ensemble responded consistently with the “Sirocco” and “Libeccio” wind events, spreading the areas of larger uncertainty around the cyclonic eddy, not so well represented by the time-variability proxy.

In order to obtain more objective performance estimates, daily performance statistics were computed as displayed in Table 1. These include the ratio Err/Std as an estimate of the error estimate accuracy, the bins correlation coefficient (C) as an estimate of the spread–skill and the bin deviation from the main diagonal (Bin Bias – BB) as an estimate of the error estimates bias.

Overall, during the period June, 13 to 25 the positive spread–skill was kept for all estimates (ensemble with and without transform and time variability), with the ensemble performing slightly better showing a 0.8 correlation coefficient among the bins while the time proxy had a 0.7 coefficient. The ratio Err/Std was also kept consistently through this period such that on average through this period the ensemble value was 13.4, the ET was kept as 1 and the time proxy as 1.1.

The mean differences between bin coordinates (i.e. deviations from the main diagonal) can also be used as an error bias estimate. Through this 12 day period (June 13 to 25) the ensemble estimates after the transform remained unbiased while the original ensemble had a value of 0.4 and the time-variability proxy showed also a negligible negative bias of 0.03.

6. Concluding remarks

The work presented above showed that some level of predictability of stochastic environmental variables through numerical modeling could be achieved using Monte-Carlo methods, producing ensemble based error estimates along with the predicted state variables, even using a limited number of ensemble runs. However, the system performance will be space and time dependent requiring an accurate metrics system to produce both diagnostics and prognostics of the precision and accuracy of the outputs.

The Ensemble Transform (ET) approach was successfully applied for free-run ocean Mesoscale error prediction calibration, by re-scaling RELO-NCOM ensembles produced through atmospheric perturbations. Independent data was used for this analysis where the model runs were not assimilating any local data. Results show that the ensemble spread did not diverge and was consistent with the observed dynamics throughout the simulation period. The ensemble showed a positive spread–skill through all ranges of the observed errors.

Comparisons of ensemble spread of the temperature profiles with local observed errors and time variability (assumed as an error proxy) showed that they were consistent through a 12 day analysis period. The ET calibrated ensemble had slightly better performance statistics than the time-variability error proxy, most likely due to the fact that the ensemble predicted errors were better correlated with the local observed dynamics.

Results show that the ensemble spread did not diverge and was consistent with the observed dynamics throughout the simulation period. Furthermore, comparisons of ensemble spread of the temperature profiles with local observed errors and time variability (assumed as an error proxy) showed that they were consistent through the 12 day analysis period, with performances above the non-calibrated ensemble estimates and time-variability used as error proxy. Overall error estimates became unbiased and the system was able to accurately separate large errors from smaller errors with a positive spread–skill relationship, through all ranges of the observed errors.

Acknowledgments

This work was supported by the Office of Naval Research under the grant N00014-08-2-1146. The authors thank the MREA07 technical and science teams and the ships’ crews that carried the in-situ data collection.

Table 1

This table shows the daily mean values of the ration between individual observed error magnitudes vs the correspondent ensemble standard deviation (Err/Std), the correlation coefficient or linear regression slope of the 1000 point bin averages (Corr. Coef) and the difference between the bins ensemble standard deviation and bin errors in °C (Bin Bias BB).

Day	Err/Std			Corr.Coef.			Bin BIAS (BB)		
	Ens	ET	Time	Ens	ET	Time	Ens	ET	Time
06/13	11.0	1.0	1.0	0.84	0.84	0.75	0.3	0.0	0.0
06/14	16.3	1.5	1.5	0.74	0.73	0.67	0.4	0.1	0.1
06/17	20.3	1.6	1.8	0.79	0.80	0.55	0.3	0.1	0.2
06/18	10.0	0.8	1.0	0.67	0.67	0.89	0.5	−0.2	0.0
06/20	17.4	1.4	1.8	0.48	0.53	0.38	0.4	0.1	0.2
06/21	12.1	0.9	1.0	0.89	0.90	0.85	0.3	0.0	0.0
06/22	10.3	0.8	0.9	0.85	0.86	0.76	0.3	−0.1	−0.1
06/23	13.0	1.0	0.9	0.90	0.91	0.90	0.3	0.0	0.0
06/24	11.9	0.8	0.6	0.85	0.85	0.94	0.2	−0.1	−0.2
06/25	11.9	0.8	0.7	0.70	0.69	0.46	1.3	−0.3	−0.5
Mean	13.4	1.0	1.1	0.8	0.8	0.7	0.43	0.00	−0.03

Each one of these estimates was computed for the ensemble without post-processing (Ens), with the ET post-processing (ET) and for the post-processed time variability used as an error proxy (time). The row at the bottom shows the overall averages during the experiment.

References

- Barron, C.N., Smedstad, L.F., 2002. Global river inflow within the navy coastal ocean model. *Proceeding of MTS/IEEE Oceans 2002 Conference*, Biloxi, MS, 29 October 2002, pp. 1472–1479.
- Barron, C.N., et al., 2006. Formulation, implementation and examination of vertical coordinate choices in the global Navy Coastal Ocean Model (NCOM). *Ocean Modelling* 11, 347–375.
- Bishop, C.H., Toth, Z., 1999. Ensemble transformation and adaptive observations. *Journal of the Atmospheric Sciences* 56, 1748–1765.
- Carriere, O., Hermand, J.-P., LeGac, J.C., Rixen, M., 2009. Full-field tomography and Kalman tracking of the range-dependent sound speed field in a coastal water environment. *Journal of Marine Systems* 78, S282–S392.
- Coelho, E.F., 2008. Operational oceanography. Zoom-in modeling for local applications. *Proceedings of the International Conference on Mathematics and Continuum Mechanics*. Pub. Centro Internacional de Matematica (CIM). ISBN: 978-989-95011-2-6, pp. 207–217. Porto-Portugal.
- Coelho, E.F., Rixen, M., 2008. Maritime rapid environmental assessment. *New trends in operational oceanography*. *Journal of Marine Systems* 69 (1–2).
- Coelho, E.F., Rixen, M., Signell, R., 2005. NATO tactical ocean modeling system: concept applicability. *NATO Undersea Research Centre Serial Report*, p. SR411.
- Coelho, E.F., Peggion, G., Rowley, C., Bishop, C., Xiaodong, X., Jacobs, G., 2007. Error modeling and sampling guidance using the ensemble transform and ensemble Kalman filter during the BP07 trial. *Rapid Environmental Assessment Conference. Coastal Processes: Challenges for monitoring and prediction*, NATO Undersea Research Centre, September.
- Cummings, J., 2005. Operational multivariate ocean data assimilation. *Quarterly Journal of the Royal Meteorological Society* 131, 3583–3604.
- Demirov, E., Pinardi, N., Fratianni, C., Tonani, M., Giacomelli, L., De Mey, P., 2003. Assimilation scheme of the Mediterranean Forecasting System: operational implementation. *Annales Geophysicae* 21, 189–204.
- Doblas-Reyes, F.J., Hagerdorn, R., Palmer, T.N., 2005. The rationale behind the success of multi-model ensembles in seasonal forecasting — II. Calibration and combination. *Tellus* 57A, 234–252.
- Egbert, G.D., Erofeeva, S.Y., 2002. Efficient inverse modeling of barotropic ocean tides. *Journal of Atmospheric and Oceanic Technology* 19, 183–204.
- Evensen, G., 2003. The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dynamics* 53, 343–367.
- Fabre, J.P., Rowley, C., Jacobs, G., Coelho, E.F., Bishop, C., Hong, X., Cummings, J., 2008. Environmental acoustic variability characterization for adaptive sampling. *Naval Research Laboratory Reviews*.
- Gneiting, T., Westveld III, A., Raftery, A., Goldman, T., 2004. Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Technical Report no. 449*, Department of Statistics, University of Washington-USA.
- Hamill, T.M., Whitaker, J., 2007. Ensemble calibration of 500-hPa geopotential height and 850-hPa and 2-m temperatures using reforecasts. *Monthly Weather Review* 135, 1414–1430.
- Hodur, R.M., 1997. The Naval Research Laboratory's Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS). *Monthly Weather Review* 125, 1414–1430.
- Hoffman, R.N., Liu, Z., Louis, J.-F., Grassotti, C., 1995. Distortion representation of forecast errors. *Monthly Weather Review* 123, 2758–2770.
- Judd, K., Smith, L.A., Weisheimer, A., 2007. How good is an ensemble at capturing truth? Using bounding boxes for forecast evaluation. *Quarterly Journal of the Royal Meteorological Society* 133, 1309–1325.
- LeGac, J.C., Hermand, J.P., 2007. MREA/BP07 cruise report. *Tech. Rep. Technical document NURC-CR-2007-04-1D1*. NATO Undersea Research Centre, La Spezia, Italy.
- Lermusiaux, P., Chiu, C.-S., Gawarkiewicz, G., Abbot, P., Robinson, A.R., Miller, R., Haley, P., Leslie, W., Majumdar, S., Pang, A., Lekien, F., 2006. Quantifying uncertainties in ocean predictions. *Oceanography* 19 (1), 80–93 Mar.
- Martin, P.J., 2000. Description of the Navy Coastal Ocean Model Version 1.0, NRL/FR/7322-00-9962. *Naval Research Laboratory*, 42 pp.
- Ngodock, H., Smith, S., Jacobs, G., 2007. Cycling the representer algorithm for variational data assimilation with a nonlinear reduced gravity ocean model. *Ocean Modelling* 19 (3–4).
- Rixen, M., Coelho, E.F., 2006 (2). Operational prediction of acoustic properties in the ocean using multi-model statistics. *Ocean Modelling* 11, 428–440.
- Rixen, M., Coelho, E., Signell, R., 2008. Surface drift prediction in the Adriatic Sea using hyper-ensemble statistics on atmospheric, ocean and wave models: uncertainties and probability distribution areas. *Journal of Marine Systems* 69 (1–2), 86–98.
- Rosmond, T.E., Teixeira, J., Peng, M., Hogan, T.F., Pauley, R., 2002. Navy Operational Global Atmospheric Prediction System (NOGAPS): forcing for ocean models. *Oceanography* 69, 99–108.
- Hong, H.X., Bishop, C., 2007. Ensemble and probabilistic forecasting. *IUGG XXIV General Assembly 2007, Perugia, Italy*, 2–13 July.