V European Conference on Computational Fluid Dynamics
ECCOMAS CFD 2010
J. C. F. Pereira and A. Sequeira (Eds)
Lisbon, Portugal, 14–17 June 2010

UNIVARIATE HIGH RESOLUTION ASSIMILATION OF NON-STATE PARAMETERS INTO OCEAN MODELS

Emanuel F. Coelho[*], Germana Peggion[†], Olivier Carrière[††] and Jean-Pierre Hermand[††]

[*]Naval Research Laboratory (University of Southern Mississippi visiting scientist)
e-mail: Coelho.ctr.po@nrlssc.navy.mil

[†] Naval Research Laboratory (University of New Orleans visiting scientist)
peggion@nrlssc.navy.mil

[††]Université libre de Bruxelles (U.L.B.), OPERA, Environmental Hydroacoustics lab
jhermand@ulb.ac.be

Key words: Ocean Modeling, Data Assimilation, Ocean Acoustic Tomography

Abstract. *Environmental parameters used for planning and execution of operations at sea are usually based on in-situ and remote sensing observations combined into single analysis and numerical models providing space-time extrapolated snapshots through a forecast range. Local observations can include non-state variables (e.g. object drift, transmission loss, acoustic signal detection range or time, etc) inverted to represent synthetic state-variables as performed in Acoustic Tomography. The combination of models and observations is typically done through the assimilation of the observed state variables into the models. It starts by extrapolating the innovations through some description of the error covariances to areas and variables dynamically coupled with the observations, producing a new analysis that is used to re-initialize the model forecast. One can expect these forecasts to have improved domain wide full-state consistency when a sufficient number of observations becomes available, however they can still have large local uncertainties due to errors in the extrapolation-smoothing of the observed innovations to the resolution and extent of the domain of the model, uncertainty in the boundary and forcing fields and unrepresented physics. Under this framework this paper discusses early results of an operational cycling methodology to combine both in-situ temperature profiles and other high resolution (space and time) synthetic variables as derived from direct acoustic or other measurements to improve local consistency of the forecasts. The methodology sequentially post-process the temperature fields using localized error covariances from a Monte-Carlo simulation, and corrects near range fields. Since this local univariate fitting does not account for non-linear corrections, correlations with other variables and weights more local correlations, it might degrade the model skills in remote areas and reduce the accuracy of longer range forecasts, motivating the need to cycle this technique with standard domain wide assimilation in order to guarantee an overall convergence of the system.*

1 INTRODUCTION

During the BP07 sea trial off the west coast of Italy at the Ligurian Sea in the spring of 2007, the High Resolution Navy Coastal Ocean Modeling (HiNCOM) system was running multiple ocean forecast nests in real-time while a fairly complete set of Temperature and Salinity observations were collected inside the inner higher resolution domains [1]. Figure 1 shows the trial area and details the simulation domains of the several ocean model runs that were conducted in real-time during the trial. Satellite observations and open data bases profile observations were assimilated in nest 0 and 1. The real-time smaller nests (BP 07 and LASIE) were not assimilating any data, although they were receiving the boundary conditions from the outer nests doing assimilation.

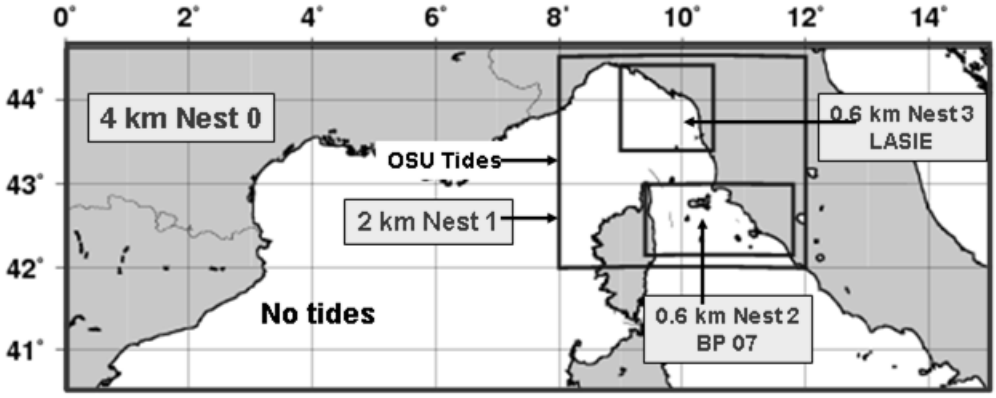

Figure 1: The triple nest configuration for BP 07. Model runs were cycling every 24 hours, producing 72 hours range forecasts, delivered in standard NeCdf file formats. Data assimilation in these real-time runs was done only in the outer (4km) and intermediate (2km) nests.

Acoustic measurements were made during some legs of the trial over the BP 07 region. Among other applications these were used through novel acoustic sequential inversion techniques to provide range resolving sound speed profiles with resolutions compatible with the high resolution nest grid spacing (order of 1km) as detailed in [2]. The description of the sequential inversion method based on a Kalman tracking algorithm and preliminary validation analysis can be found in [3]. Within this framework, this paper describes and shows preliminary testing of the approach the team will use for Multi-scale Data Assimilation combining these high resolution and frequency synthetic profiles with the other routine sparse in-situ observations in order to improve the local consistency of the high resolution ocean model fields over the regions where the acoustic measurements took place.

Early work showing improvements in correcting next 24 hour temperature bias by post-processing the real-time free runs using the local temperature profile observations of the previous day was discussed in [4]. In this paper we take an additional step and look into the impact of local high resolution temperature profile observations in post-processing the local nest also assimilating the sparse public data, local high density profile measurements and satellite observations and using improved error covariances based on a larger and well calibrated ensemble population. This approach aims to improve the short range local forecast and assumes no significant non-linearities are present during the short forecast range period (on the order of 24 hours).

As described, this system can be interpreted as a multi-scale solution to the assimilation of observations in ocean models, starting by correcting the large scale and longer period features in the outer nests, feeding the corrected fields into the inner nests

through the boundaries. The scales reproduced by each inner nest are then corrected internally through higher resolution analysis. The final local zoom-in into the area and variables details is then performed through post-processing of the highest resolution nests or super-ensemble consensus analysis, using the full resolution of the local observations. This nested assimilation approach allows dynamical features or instabilities as detected by the data to be sequentially projected into the scales reproduced by each domain resolution and to evolve, span and project downward to the smaller grids any new significant dynamical modes consistent with the represented scales, while keeping consistency at the boundaries for each high resolution domain. Following work will use this methodology for the analysis of the impact of high resolution and high frequency acoustic derived synthetic profiles and will test the use of the improved ocean fields into the profile inversion within Kalman tracking cycles.

2 THE NAVY COASTAL OCEAN MODEL AND DATA ASSIMILATION

The Hi-NCOM uses a standardize development and an efficient configuration management to facilitate transitions of new tools and real-time configurations of regional high resolution (up to order 1 km) ocean predictions [5]. The physics and numerical procedures of NCOM [6] are based on the Princeton Ocean Model (POM) and a Sigma/Z-level Model (SZM). It solves a three-dimensional, primitive equation, baroclinic, hydrostatic and free surface system using a cartesian horizontal grid, a combination of σ/z level (i.e. bottom-following/constant depth) vertical grid and implicit treatment of the free surface. Horizontal eddy coefficients are calculated based on maximum grid-cell Reynolds number criteria, and vertical eddy coefficients are calculated using the Mellor- Yamada Level 2 turbulence closure scheme. For meso-scale real-time applications, outer nests boundary conditions are taken from an operational run of the global NCOM (GNCOM). The global model assimilates satellite altimetry and Sea-Surface Temperature data using a combination of model analysis and data. Since the global NCOM does not include tides, these are explicitly inserted in the HiNCOM nests through the boundary conditions and local forcing terms.

The data assimilation is carried using the Navy Coupled Ocean Data Assimilation (NCODA) system, detailed in [7]. It is based on a three-dimensional multivariate optimum-interpolation (MVOI) data assimilation system, now evolving into Variational Assimilation Schemes (3DVAR and 4DVAR) that can cycle in real-time to provide new analysis and model updates of the ocean state variables (temperature, salinity, velocity and sea surface height). Additional capabilities have been built into the system, including flow-dependent background-error correlations and background-error variances that vary in space and evolve from one analysis cycle to the next. It also includes a data quality-control system with multivariate analysis using feedback of forecast fields and prediction errors in the quality control of new observations.

The NCODA interpolation problem is formulated as:

$$x_a - x_b = P_b H^T (H P_b H^T + R)^{-1} [y - H(x_b)] \quad (1)$$

where x_a is the analysis vector, x_b is the time dependent background vector, P_b is a time dependent background error covariance matrix, H is the observational functional operator, R is the observation error covariance matrix, and y is the observation vector at a specific update cycle. The forward functional H converts forecast model variables to an observed variable and, as used here, is a spatial interpolation of the forecast model grid to the observation location performed in three dimensions. Therefore, HP_bH^T is the background-error covariance between observation locations, and P_bH^T the error

covariance between observation and grid locations. The quantity $\{y - H(x_b)\}$ is referred to as the innovation vector and $x_a - x_b$ is the increment (or correction) vector. Observations are assimilated close to their measurement times within the update-cycle time window by comparison against time-dependent background fields using the first-guess at appropriate time (FGAT) method. The ocean variables are analyzed simultaneously in three dimensions such that the observation vector contains all of the synoptic temperature, salinity and velocity observations that are within the geographic and time domains of the forecast model grid and update cycle. The velocity increments are forced to be in geostrophic balance with the geopotential increments which, in turn, are in hydrostatic balance with the temperature and salinity increments. Prior to an analysis the innovation vector is normalized by the background error at the observation locations, and after an analysis the increment vector is scaled by the background error at the grid locations.

Typical implementations of NCODA (e.g. [8]) use more than 30 vertical levels, with the background error variances being computed from the increments using a recursive filter model with a time constant of 10 days and imposing geostrophic cross-correlations on the velocity errors computed from the mass variables. The first baroclinic Rossby radius of deformation is usually selected as the local spatial correlation length scale for horizontal interpolations between observations and observations and grid locations. The vertical correlation length scales is computed from local background density vertical gradients. The system uses a First Guess at Appropriate Time (FGAT) window of 24 hours, usually set as 12 hours around analysis time. When no data is available for long periods of time, error variances are relaxed towards climate variability. The NCODA system also includes a complete quality control system that assesses the error probability of each individual observation and the final profile gridded data. These profile data is then processed with the model runs to produce data match-up files that are used to run diagnostics of the model forecast errors and ensemble performances.

The runs of these systems deliver 72 to 96 hours forecasts over horizontal grids of 1 to 3km resolution and 50 vertical levels, from which the upper 30 are sigma layers. The topography is usually taken from the Naval Research Laboratory global 2 minute ocean bathymetry data base (NRL DBDB2). Atmospheric forcing usually consists of 3-hourly fields of sea level air pressure, wind stress, solar and long wave radiation, and 2m humidity 15-km resolution Coupled Ocean-Atmosphere Modular Prediction System (COAMPS) analysis/forecast runs and interpolated to the ocean model grids. Operational runs start daily 24 hours prior to analysis time from a snapshot of the previous run. The first 24 hours are to be used for the model correction by sequentially adding the increments computed by NCODA (slow insertion of model corrections), such that at the analysis time (hour 0) the model fields will reproduce the best analysis estimates as delivered by NCODA and as deailed in the nest paragraphs. Model outputs are then post-processed to standard levels and made available through NetCdf files. These files can include a full or partial range of variables from the model original runs.

3 RELO ENSEMBLES AND THE ENSEMBLE TRANSFORM DURING THE BP 07 TRIAL

The errors in the HI-NCOM variable forecasts are determined by multiple sources of uncertainty. They are associated with the model initialization and boundary conditions, numerical approximations, modeling strategies, impact of under-sampling in the assimilation process and unresolved scales. To address the initialization error this work uses the Ensemble Transform (ET) method to produce perturbations at each analysis time following the approach detailed in [9]. The ET uses the best available estimate of

analysis error covariance to transform forecast perturbations into analysis perturbations by finding K distinct linear combinations of K forecast perturbations that (a) are equally likely (b) lie within the vector subspace of forecast perturbations (c) are quasi-orthogonal although they sum to zero, and (d) have expected squared amplitudes equal to the trace of the best available estimate of the analysis error covariance matrix.

During operational implementations the analysis error variance is obtained from the NCODA analysis error (e.g. [10]), combining observed and background errors to create a new ET ensemble at each new analysis cycle. The ET analysis perturbations are then added to the best available analysis (in this case produced by the HiNCOM-NCODA) to generate K initial states. These K initial states are then integrated forward in time using the non-linear model to produce the next ensemble forecast. This forecast will then be the starting point for the new ET computation that will generate the initial conditions for the subsequent ensemble forecast once the subsequent analysis is available. As such, the ET ensemble generation algorithm is a cycling procedure with strong similarities to a breeding scheme. As in the breeding scheme, creating initial condition perturbations from linear combinations of forecast perturbations ensures that, after a few cycles, the structures of the ensemble perturbations reflect those of the leading Lyaponuv or most rapidly growing normal modes of the dynamical system.

The HiNCOM ensembles also include the errors due to uncertain atmospheric forcing by producing an ensemble of atmospheric perturbations and allow each ocean run during its forecast period to have an independent atmospheric forcing, therefore augmenting the domain span by the perturbed initial states ensemble. This is done using atmospheric forcing perturbations as developed by the methodology detailed in [11]. This set of perturbed forcing fields uses spatially-varying time shifts of the atmospheric forecast, with a choice of parameters to provide a well developed spread of atmospheric variables perturbations. If there are no atmospheric data time-series to calibrate these perturbations, their accuracy may not be guarantee over the whole simulation period. This method is based on the observation that predicted atmospheric fields often contain the forecast feature of interest, but they can be misplaced in space and time, therefore it does not take into account any possible significant bias or strong non-linearity in the atmospheric forecasts. The ensembles resulting from combining the ET and the perturbed atmospheric forcing are then used to predict how uncertainties of the ocean fields will evolve in space and time (local variances and covariances). Typical implementations of this technique for targeting observations (adaptive sampling, e.g. [10]) or other reduced order or highly localized problems use 40 independent members perturbed from a control run performing the full data assimilation cycle. Each one of the simulations uses independent atmospheric forcing and perturbed initial conditions as detailed above. This technique still does not account for additional error sources that could develop during the forecast period through the boundaries and/or through unrepresented numerical errors and approximations.

This paper will focus on the period April 16 to May 4 2007 in the BP07 domain as detailed in Fig. 1. During this period the 32 Ensemble runs were done starting each day from perturbed initial conditions of a control run assimilating satellite and other profile measurements made available through the public global data bases, following the procedures detailed above. Figure 2-a show a snapshot of the temperature field at 00:00 May, 1st at 0m depth and Fig. 2-b shows the corresponding ensemble spread defined as the standard deviation of the population of possible temperature values delivered by the Monte-Carlo runs. The small values in the south boundary are due to the fact the system is still not using perturbed boundary conditions, such that the ensemble estimates near this region are not to be considered accurate. Future versions of the system will account

for this limitation, along with the uncertainties in the bathymetry and in the physics approximations being used by the model.

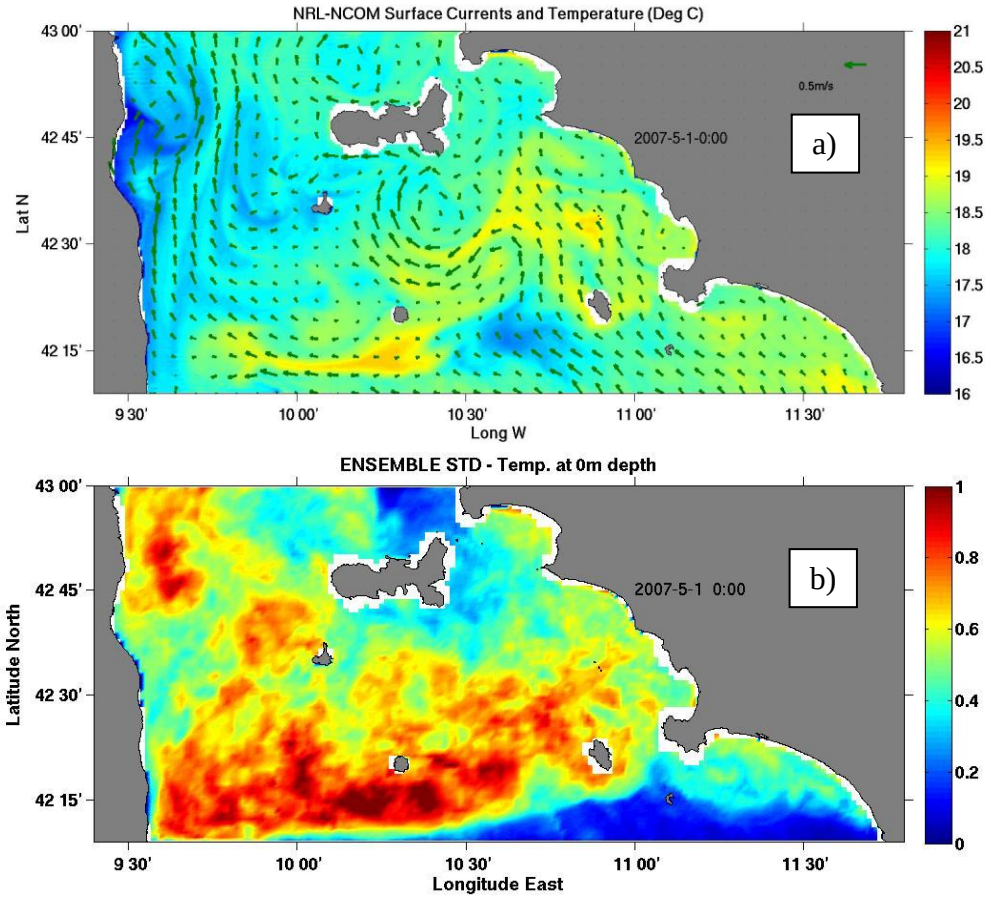

Figure 2: a) HiNCOM water temperature and currents snapshot at the surface at 00:00 May, 1st 2007. Land on the East corresponds to the coast of Italy and o the West to the island of Corsica. The large island between Italy and Corsica is the island of Elba. b) Ensemble spread (or predicted error) of the temperature fields displayed in 1.a) computed as the standard deviation of 32 independent perturbed runs of HiNCOM.

From Fig. 2 we can see that regions of larger ensemble spread concur with the areas of larger spatial gradients (and time variability). In particular one can see the higher ensemble spread along the gyre south of the Island of Elba, along the small re-circulation cells west of Elba and along the high temperature shape close to the southern boundary, noting these are areas of high uncertainty in the model. Since these estimates use an ensemble population larger than 30, through the central limit theorem one can assume there is a normal probability distribution density envelope capturing the perturbations of the temperature fields around the control run (or ensemble mean) that can be estimated by this ensemble population. Based on the observed spread for the example displayed in Fig. 2-b and taking into account a number of samples equal to 32, we can then expect the ensemble mean temperatures errors to be smaller than 0.7^{o}C up to 95% when the ensemble spread is 2^{o}C and 0.3^{o}C when the ensemble spread is 1^{o}C. This analysis can provide an immediate application of the ensemble as a proxy of the error variance. This is checked in Fig. 3 by showing how the predicted ensemble spread compares with the magnitudes of the 0-48 hour forecast-observations mismatches (|ei|). The upper scatter plots show the temperature and salinity ensemble spread vs the

observed |ei| for days May, 1st to May 3rd. The maps below each scatter diagram show the ensemble spread of the surface fields at the analysis time and the white crosses show the locations were data was available and used to compute the model mismatches |ei|.

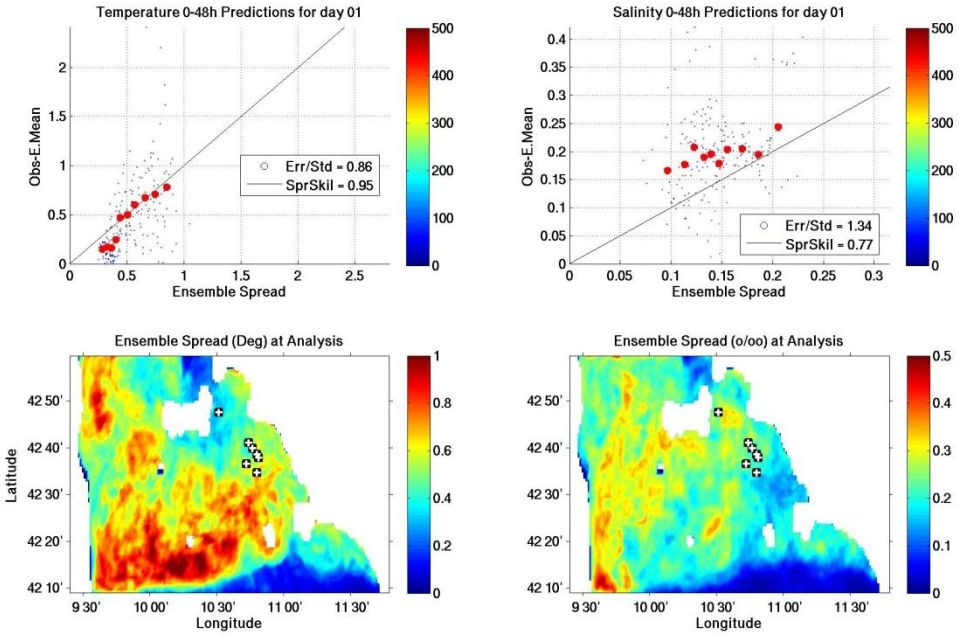

Figure 3: The upper scatter diagrams show TEMPERATURE and SALINITY ensemble spread vs the observed forecast errors (e) for the ensemble run of May 1st 2007. The maps below each scatter diagram show the mean ensemble spread at the analysis time and the white crosses show the locations were data was available and used to compute the model-data mismatches. The color of each individual estimate displayed as the small dots corresponds to the depth of the observation, accordingly to the color code displayed in the bars on the right (in meters).

Since the ensemble variance is designed to be a predictor of the true error variance its square root it is not comparable to single measurements of the magnitude of model-data mismatches. To account for this fact the observed model mismatches were used to estimate observed error variances by ordering the data pairs from smallest ensemble spread to largest ensemble spread and then grouping the data pairs into 10 approximately equally populated bins. Within each bin, the bin averaged magnitudes of model-data mismatches were computed together with the bin-averaged ensemble spread. Since both of these quantities are sample means, they are comparable. These quantities are displayed in Fig. 3 as the large red dots in the scatter diagrams. For the ensemble to be accurate, these large red dots should be aligned along a positive slope and ideally along the main diagonal, highlighted as a black line on the plots. The skill is represented by the metric "SprSkil" corresponding to the bin correlation changing between 0 and 1. A value of SprSkil=1 represents a perfect spread-skill relation and values above 0.5 can be considered a good ensemble spread-skill (i.e. the ensembles have the ability to correctly separate the small ensemble spread values correlated with the smaller observed errors, from the larger ensemble spread values correlated with the larger observed forecast errors). Another relevant metric is the mean ratio between measured magnitudes of model-data mismatches and ensemble spread, (Err/Std in the plots), which represents the ensemble skill in terms of correctly predicting the magnitudes of the errors. Values higher than 1 indicate that the ensemble is under-dispersive (under-predicts the error magnitude and error bins are above the diagonal

line) and values below 1 indicate that the ensemble is over-dispersive (over predicts forecast error magnitudes such that error bins are below the main diagonal). We can see for this case that for temperature the ensemble was slightly over-dispersive and under-dispersive for salinity.

During the period discussed in this work the spread-skill was always above 0.9 and the ensemble was slightly under-dispersive in temperature such that we can consider the Monte-Carlo runs as a consistent model for error variances. However, one still needs to check if the ensemble was also providing reliable error covariances, i.e. if the correlation between errors from different locations and times, as delivered by the model, were consistent with the observations. This is a critical skill in allowing extrapolating the measured innovations from the locations and times were we had observations to the regions and times of interest. To accomplish this tests one would like to have long records of time-series data through many data points and well define dynamical features that one could easily track from the control run, which was not feasible to achieve during this trial. In any case, using the results of a different application where the ensemble was used to design optimal sampling strategies, the ensemble based covariances proved to be providing accurate estimates of the areas of enhanced sensitivity to selected target regions as described in [10]. From these results one could expect that the ensemble once providing consistent error variance estimates will be also providing reliable covariances within the error modes detected by the system

The assimilation of local observations and the ensemble transform applied to the initial conditions perturbation reduced the ensemble spread in temperature (smaller predicted errors) and the ratio Err/Std became closer to 1 throughout this leg of the trial. As time evolved the larger spread became more well defined along the regions with larger space-time variability and were observations were more sparse. The larger differences between the ensemble estimates and the observed model-data mismatches were seen between the 20m and 50m isobaths, near the mixed layer depth.

4 HIGH RESOLUTION LOCAL ANALYSIS USING THE ENSEMBLE TRANSFORM KALMAN FILTER (ETKF) DURING THE BP 07

Numerical models estimates contain multi-scale physics with different statistical properties and local data that can represent the dynamics up to the scales solved by the numerical schemes and constrained by the observations when using data assimilation. However, the local observations networks are usually insufficient to provide synoptic fields of the high frequency dynamics, such that certain scales and processes will not be reproduced accurately subject to the scales and time spans not well resolved by the observation network.

Figure 4 shows an example of the mean profile RMS errors for the several grids running in real-time. The dashed and solid blue lines correspond to the 2 km resolution nest 1 grids as shown in Figure 1, running without NCODA analysis and with NCODA analysis respectively. The dashed and solid green lines correspond to the 0.6km resolution nest 2 grids taking boundary conditions from the nest 1 free run and assimilative run respectively. The nest 2 run shown as solid green was also running a high resolution NCODA analysis and assimilating the local profile and remote sensing data. The nest 1 free run RMS errors were more noticeable in the deeper layers and we can see that these errors were significantly reduced through the assimilation of the local profile data (solid blue and green lines). However, the high resolution nest 2 by allowing the existence of higher frequency modes not constrained by the observations was displaying larger errors near the mixed layer depth. The assimilation of the local profiles in nest 2 does not seem to have been able to account for this limitation.

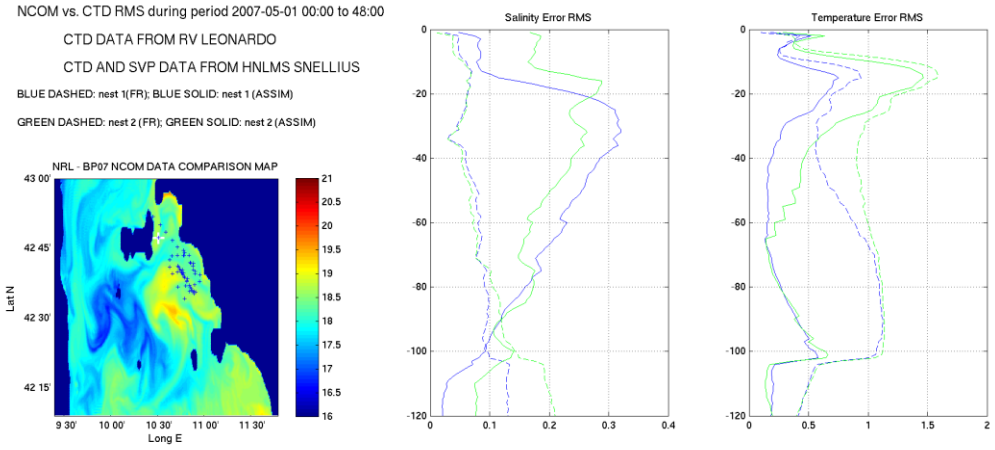

Figure 4: HiNCOM model-data comparisons for nest 1 free and assimilative runs and for nest 2 free and assimilative runs. The map on the left show the locations where observations were available during the simulation period (May, 1st and May, 2nd 2007). The background colour corresponds to the surface temperature at the end of the simulation (00:00 May, 3rd). The plots on the right show the RMS error profile for salinity and temperature.

Variational approaches (e.g. [12]), and Ensemble based Error Sub-Space assimilation techniques (see [13] for a review) can account for the different scales detected in the model simulations, but they cannot separate these scales within the observations networks (i.e. filter the data up to the scales for which the observed data is enough to constraint the dynamical structures) such that unrepresented observations will always be mapped into the other scales detected by the model runs. These effects become more relevant when data is sparse in resolution but has high sampling rates and there are energetic high frequency dynamics in the observations. These limitations suggest treating the assimilation and error prediction altogether as a multi-scale problem that might require several analysis steps. This paper shows preliminary results of an alternative combining a standard analysis (presently based in a MVOI) and a post-processing step using a Kalman filter formulation and Error covariances estimated through a Monte-Carlo simulation, imposing a localization of the analysis into local data as explained below.

When using highly localized or post-processing analysis one can expected to improve the short range fitness skills for the observed variables, constrained by the scales solved by the observations and the accuracy of the covariances estimated by the ensemble. Therefore, if the corrections are not adequately localized they may degrade the accuracy of the near and far range (in space and time) or the analysis of non observed state variables that could be poorly correlated with the measurements. One way to mitigate this problem is to apply the post-processing analysis to single output variables of interest (e.g. sound speed, temperature or any other combination of state parameters) and relax or impose external constraints for any balance that may be physically required (e.g. static stability, etc). This single variable (or non-state variables) analysis will not allow a direct insertion of the analyzed fields into new forecasts, limiting the feedback into the next forecast cycle. Therefore, to guarantee the model runs will remain consistent and improve the near range fields, this technique should always be complemented with other standard multi-variable scheme and allow the same data (or derived synthetic profile data) to be included into the routine data assimilation process and used into the overall smoother regional analysis.

By applying expression (1) to the local observations of water temperature (x(t)) during an analysis period (e.g. first 24 hours of the simulation) and using local time dependent estimates of the temperature ensemble covariance one can estimate an analysis field $x_a(t) = x_b(t) + G(t)\Delta x$, where Δx corresponds to the local observed innovations and G is the time dependent mapping of these innovations into the model grid. Because the small number of ensemble members and to account for processing times, in this experiment we localized the vertical profiles over six levels at (10-20-30-40-50-75 m) and corrections were computed independently the each level bins using spatial correlations directly estimates from the ensemble temperature fields, therefore significantly reducing the number of degrees of freedom of the problem. The red bars on the right displayed in Figure 5 shows the mean residual errors and RMS between analysis (trained) and observations for the period from 00:00 May, 1st to 00:00 May 2nd. The blue left bars correspond to the nest 2 (assimilative run) forecast errors. As expected the residual mean errors are significantly smaller but some background standard deviation error remains due to the representation error (R in expression 1) and fine scale structure not solved by the error covariance based interpolation.

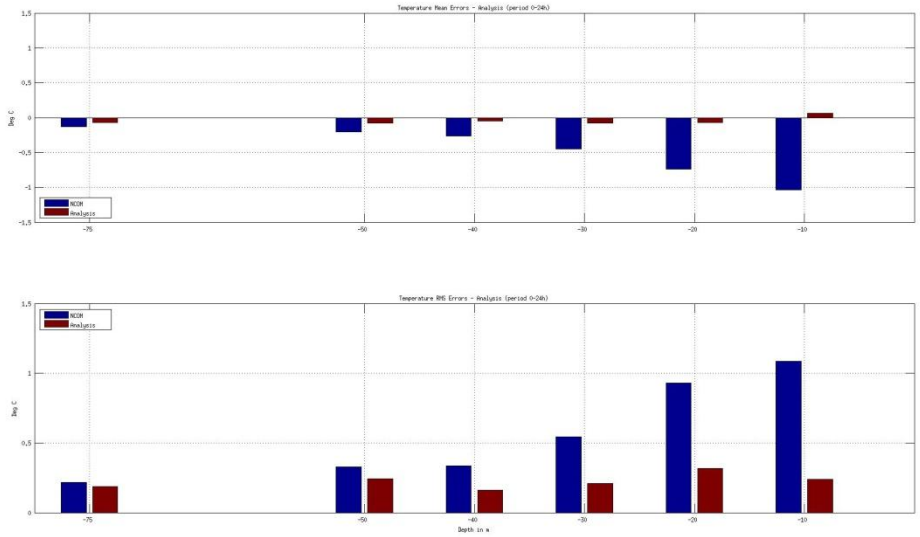

Figure 5: Bar diagrams with the residual mean level bin errors (upper plot) and standard deviation error (lower plot) before and after the analysis (training) for the data set from 00:00 May, 1st to 00:00 May 2nd and the run of May, 1st.

The extrapolation of these innovations to the data at different locations and times was then done using correction persistency and error covariance corrections estimates from a technique proposed by [14] based on the Ensemble Transform Kalman Filter (ETKF) and applied by [15] to adaptive sampling in atmospheric modelling applications and by [10] in ocean applications. This technique uses the ensemble forecast, as detailed in the previous section, and rapid low rank solutions of the Kalman filter equations to estimate the impact of observation networks in the errors covariance.

Following [14], this technique assumes the analysis error covariance at the observation time can be estimated by $\mathbf{P}_e^r = \frac{\mathbf{X}^r \mathbf{X}^{rT}}{K-1}$ where Xr, O(NxK) is the matrix with the ensemble N state-variables at the observation time and K is the number of ensemble

members. However, these perturbations should be consistent with the best available guess of the error variance of an analysis made using all of the observations apart from the observations that will be targeted. To account for this constraint and as described in [15] one could apply a Transformation matrix $\mathbf{T}^r$, such that $\mathbf{X}^r = \mathbf{X}^f \mathbf{T}^r$, and where $\mathbf{X}^f$ are the raw forecast perturbations. The transformation matrix $\mathbf{T}^r$ can be computed using the ET technique and a guess of the analysis error covariance associated with the local observations.

The posterior analysis error covariance $\mathbf{P}_i^a$ after assimilating the ith deployment of observations $\mathbf{y}_i^a$ will then be given by

$$\mathbf{P}_i^a = \mathbf{P}_r^e - \mathbf{P}_r^e \tilde{\mathbf{H}}_i^{aT} \left(\tilde{\mathbf{H}}_i^a \mathbf{P}_r^e \tilde{\mathbf{H}}_i^{aT} + \mathbf{I} \right)^{-1} \tilde{\mathbf{H}}_i^a \mathbf{P}_r^e \tag{2}$$

where $\tilde{\mathbf{H}}_i^a$ describes the mapping from the model state vector to the observation vector normalized by the inverse square root of the observation error covariance $\mathbf{R}_i^{-1/2}$ associated with the ith feasible deployment. Using this result, we can also estimate $\mathbf{P}_i^a$ as shown in [14], through $\mathbf{P}_i^a = \frac{\mathbf{X}_i^a \mathbf{X}_i^{aT}}{K-1}$ where $\mathbf{X}_i^a$ is the ensemble state matrix after the assimilation of the observed variables.

The columns of $\mathbf{X}_i^a$ may be interpreted as transformed ensemble perturbations such that their covariance gives the analysis error covariance at the observation time assuming that the ith deployment of targeted observations had been assimilated i.e. $\mathbf{X_i^a} = \mathbf{X^r}.\mathbf{T_{ETKF}}$ where T_{ETKF} is now determined by the eigenvectors and eigenvalues of the projections of the magnitude of the analysis perturbations, corresponding to the possible observations strategies, into $\tilde{\mathbf{H}}_i^a$. As shown in [14], this transform will not change along the forecast cycle; therefore the same transform matrix can be applied to the forecast perturbations to estimate the error covariance at the verification time $\mathbf{P}_i^v$. These results can then be used to infer the impact of observations taken at a given observations period in changing the error variance (and covariances), for the selected variables at a later time and at specific target locations. This technique allow us then to project the mapping of the innovations over time, taking into account a prognostic of the changes in variances and covariances that will be imposed by the new observations, prior to executing a new full run of the models. To check this procedure, the target locations were set as the positions of the observations on the following 24 hours such that we could then compared the forward projected analysis based on the first 24 hours measurements with new independent measurements.

Figure 6.a) displays the accumulated model corrections at 00:00 May, 2nd computed from the nest 2 assimilative run starting 00:00 May, 1st, and using the observations made during the first 24 hours. Figure 6.b) shows the time extrapolated corrections at 00:00 May, 3rd computed using the estimates of the error covariance changes as detailed in the following paragraphs. The black dots in Fig. 6.a) correspond to the observation positions used for the analysis and the black dots in Fig. 6.b) to the independent observation positions used for validation.

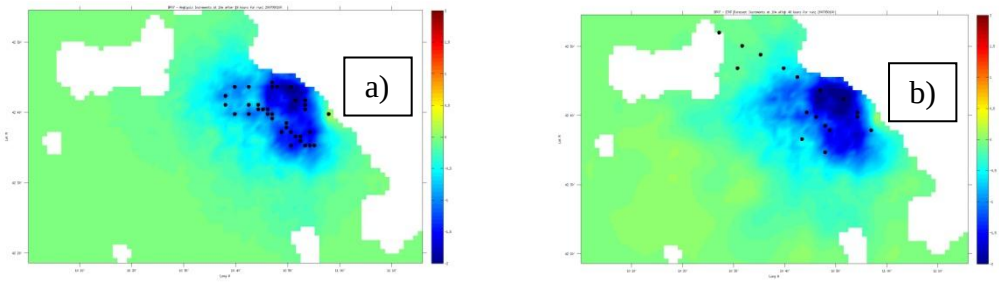

Figure 6: a) accumulated model corrections at 00:00 May, 2nd computed from the nest 2 assimilative run starting 00:00 May, 1st, and using the observations made during the first 24 hours. The black dots in Fig. 6.a) correspond to the observation positions used for the analysis b) time extrapolated corrections at 00:00 May, 3rd. The black dots correspond to the independent observation positions used for validation.

Equation (3) summarizes the procedure. At the observations time the new analyses $x_a(t)$ are computed from the NCOM run background (Nest 2 assimilative in this case) and the observed corrections interpolation through G(t) as displayed in the upper expression. For a time in the future (t +Δt) the background and corrections evolve according to the non-linear model M as shown in the expression in the middle, has we should obtain through the routine assimilation. This step is approximated and localized assuming an observed error persistency plus a linear tangent approximation of the error evolution, plus the contribution of the change in error covariance S(t/H) based on the observations H made at time t and estimated through the ETKF (bottom expression).

$$\begin{aligned} x_a(t) &= x_b(t) + G(t)\Delta x(t) \\ \ddot{x}_a(t+\Delta t) &= M^{\Delta t} x_b(t) + M^{\Delta t}\ddot{G}(t)\Delta x(t) \\ x_a(t+\Delta t) &\approx x_b(t+\Delta t) + G(t+\Delta t)\Delta x(t) + G(t)\bar{\varepsilon}\Delta t + S'(t/H)\Delta x(t) \end{aligned} \tag{3}$$

Figure 7 shows the resultant profile RMS errors of the nest 1 runs and nest 2 assimilative without and with post-processing, starting at 00:00 May, 1st, and during the post-processed forecast period from 00:00 May 2nd to 00:00 May 3rd. The panel on the left show the independent observations used for the error calculations (similar to those displayed in Fig. 6.b). One can see the post-processing brings the RMS profile errors of the nest 2 assimilative to levels comparable with the nest 1, although there are still some large errors in the upper layers. Following tests will include a larger vertical population and different localization algorithms to attempt improving these results. One should note that this is done without any direct filtering, while keeping the high frequency modes present in this high resolution grid.

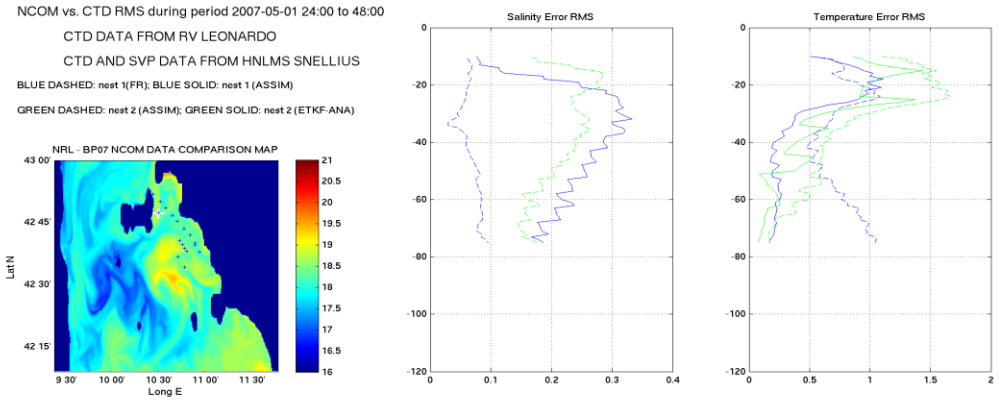

Figure 7: HiNCOM model-data comparisons for nest 1 free and assimilative runs and for nest 2 assimilative and post-processed runs. The plots on the right show the RMS error profile for salinity and temperature. The map on the left show the locations where observations were available during the validation period (May, 2nd 2007). The background colour corresponds to the surface temperature at the end of the simulation period (00:00 May, 3rd).

Figure 8 shows how the mean profile RMS errors evolved during the period from April, 24th to May, 2nd. One can see that during all testing cases the nest 2 post-processed runs had significantly smaller errors than the original forecast, bringing the error values close to the larger grid and smoother nest 1 run. The large magnitudes of the errors were all in the upper levels. Future work will include more levels in the upper layer to bring these values further down.

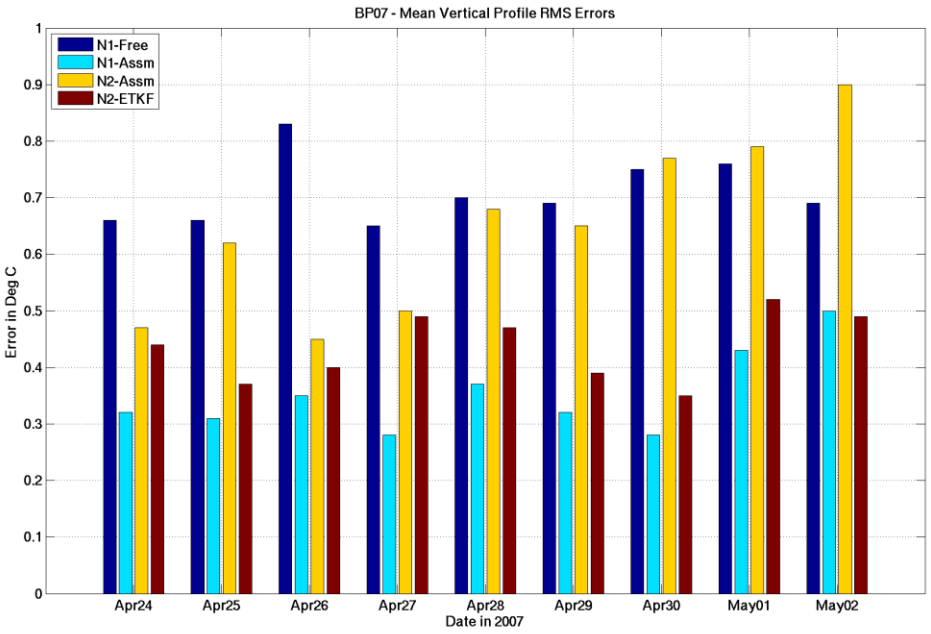

Figure 8: HiNCOM model-data comparisons during the period April, 24 to May, 2 2007 for nest 1 free (N1-free) and assimilative (N1-assm) runs and for nest 2 assimilative (N2-assm) and post-processed (N2-ETKF) runs. The bar values correspond to the mean of RMS error profiles computed over a 24 hours validation period (24-48 hours forecast ranges) using independent data.

5 CONCLUSIONS

This work proposes a method to combine models and local observations to be used in operational ocean models together with routine data assimilation of local state-parameters observations, leading to improved consistency of single forecasts variables. In the recent past some work has also been done in bringing synthetic ocean profiles derived from the acoustic anomalies into the model assimilation process, though requiring a special care due to their difference in the representation and instrumental when compared to local in-situ profile measurements some problems. The proposed methodology seems well suited to perform this task in post-processing by combining both in-situ temperature and salinity profiles and high resolution (space and time) temperature or sound-speed profiles derived from acoustic measurements and because of its fast turn-out time is well suited for sequential approaches in the acoustic inversions like Kalman filters (e.g. [3]).

AKNOWLEDGMENT

This work was partially supported by the Office of Naval Research under the grants N00014-08-2-1146 and N00014-07-1-1069. The authors thank the MREA07 technical and science teams, the ships crews that carried the in-situ data collection and Dr. Jim Richman for his comments on an early version of the manuscript.

REFERENCES

[1] E.F. Coelho, Peggion, G., Rowley,C., Jacobs, G., Allard, R., Rodriguez, E., "A Note on NCOM Temperature Error Calibration Using The Ensemble Transform", Journal Marine Systems , Rapid Environmental Assessment special issue, 78, S272-S281 (2009)

[2] J.-C. Le Gac and J.-P. Hermand, MREA/BP07 cruise report. Tech.Rep. Technical Document. NURC-CR-2007-04-1D1.NATO Undersea Research Centre, La Spezia, Italy, (2007)

[3] O. Carrière, J.-P. Hermand, and J. V. Candy, "Inversion for time-evolving sound-speed field in a shallow ocean by ensemble Kalman filtering," IEEE J. Ocean. Eng., vol. 34, no. 4, pp. 586–602, doi:10.1109/JOE.2009.2033954, (2009)

[4] E.F. Coelho, Hermand, J-P, Peggion, G., and Carriere, O., "An Operational Approach to Sound Speed Data Assimilation in High Resolution Ocean Models", Underwater Acoustic Measurements & Results Conference – Nafplion, Greece, 21-26 June (2009)

[5] C. Rowley, P. J. Martin, and James A. Cummings, "The Naval Research Laboratory Relocatable Ocean Nowcast/Forecast System", US-Navy, ONR Journal of Underwater Acoustics, accepted (2010).

[6] P.J. Martin, "Description of the Navy Coastal Ocean Model Version 1.0", NRL/FR/7322-00-9962. Naval Research Laboratory. 42 pp (2000).

[7] J. Cummings, “Operational multivariate ocean data assimilation”. Q. J. R. Meteorol. Soc., 131, pp. 3583–3604 (2005).

[8] Lunde, B., Coelho, E.F., “Implementations Of The Navy Coupled Ocean Data Assimilation System At The Naval Oceanographic Office” MTS/IEEE Oceans’09, 090602-015 (2009).

[9] J.G. McLay, C.H. Bishop, and C.A. Reynolds, “Evaluation of the Ensemble Transform Analysis Perturbation Scheme at NRL”. Mon. Wea. Rev., 136, 1093–1108 (2008)

[10] E.F. Coelho, Rowley, C., Jacobs, G., “Ocean Data Assimilation Guidance Using Uncertainty Forecasts”, MTS/IEEE Oceans’09, 090611-014 (2009)

[11] X. Hong, and C.H. Bishop,”Ensemble and Probabilistic Forecasting”. In Proceedings IUGG XXIV General Assembly 2007, Perugia, Italy, 2-13 July (2007).

[12] H. Ngodock, S. Smith and G. Jacobs, “Cycling the representer algorithm for variational data assimilation with a nonlinear reduced gravity ocean model”, Ocean Modelling, Volume 19, Issues 3-4, (2007).

[13] L. Nerger, W. Hiller and J. Schroter, “A comparison of error subspace Kalman filters”, Tellus, 57A, 715–735, (2005)

[14] C.H. Bishop, B.J., Etherton and S.J. Majumdar, “Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects”. Mon. Wea. Rev. 129, 420-436 (2001)

[15] S.J. Majumdar, C.H. Bishop, B.J.. Etherton and Z. Toth, “Adaptive Sampling with the Ensemble Transform Kalman Filter. Part II: Field Program Implementation”. Mon. Wea. Rev., 130, 1356-1369 (2002).