Data Assimilation in High Resolution Numerical Simulations of the Ocean Circulation: A Progress Report

E. Chassignet, Z. Garraffo, A. Mariano, M. Chin, and R. Baraille

RSMAS/MPO, U. of Miami, 4600 Rickenbacker Cway, Miami, FL, 33149

Corresponding author: echassignet@rsmas.miami.edu

Abstract

The main scientific goal of this Challenge project is to generate optimal estimates of the time-varying ocean state in support of the NAVY's needs on synoptic time scales on the order of weeks to months and on spatial scales typically on the order of 10-1000 km (mesoscale). Doing this in real time requires interplay between large varied data sets, numerical ocean circulation models, and data assimilation algorithms. Due to the large demand placed by near-optimal assimilation techniques on raw computing power, this work fits most naturally under the Grand Challenge label.

Introduction

Numerical modelers at the University of Miami have long pursued the goal of studying the ocean circulation using models formulated in density (isopycnic) coordinates (Bleck and Chassignet, 1994). Because many physical processes in the ocean are rather intimately related to isopycnal surfaces and to the way in which they deviate from the horizontal, isopycnic models have much to contribute in elucidating oceanic circulation features with scales ranging from frontal to global. The association of vertical shear with isopycnal packing and tilting in the ocean makes these models appropriate for studies of strong baroclinic currents, such as the Gulf Stream. However, the fundamental reason for modeling ocean flow in density coordinates is that this system suppresses the "diapycnal" component of numerically caused dispersion of material and thermodynamic properties (temperature, salinity, ...). It is this characteristic that allows isopycnic models to keep deep water masses near the freezing level for centuries – in agreement with observations – while surface waters can be as warm as 30 degree Celsius. Models framed in Cartesian coordinates suffer from vertical "heat leakage" which causes the ocean to act as a giant heat sink in climate simulations (Chassignet *et al.*, 1996).

The main objective of this Challenge project is to perform a realistic, truly eddy-resolving, wind- and buoyancy-forced numerical simulation of the North Atlantic basin with data assimilation capabilities and assess the nowcast/forecast capabilities of such a high-resolution ocean model to support the NAVY's needs on synoptic time scales on the order of weeks to months and on spatial scales typically on the order of 10-1000 km (mesoscale). Specifically, the primary research objectives are real-time forecasting of both Lagrangian trajectories and 3-D Eulerian fields associated with such physical parameters as velocity, temperature, salinity, and density. The five major components of the effort are (i) MICOM, the Miami Isopycnal Coordinate Ocean Model, (ii) data from Lagrangian drifters and satellite-derived sea surface temperature and height fields, (iii) an Extended Kalman Filter (EKF) with a Gauss-Markov Random Field (GMRF) model for spatial covariances, (iv) a random flight turbulence model for Lagrangian trajectory prediction, and (v) contour-based parameter estimation and assimilation techniques. The computational requirements for basin-scale ocean modeling at the resolutions of interest (less than 10 km) are extreme (Bleck *et al.*, 1995). Each time that the horizontal resolution is increased by a factor of n, the computational load goes up by a factor n^3 since the n-fold reduction in linear mesh size requires n times more time steps to integrate the model over a given time interval. Assimilation techniques increase this computational load by a factor 2 to 20 depending on the optimality of the technique.

Approach

The Ocean Model

The computational domain for the Miami Isopycnal Coordinate Ocean Model (MICOM) consists of the North and Equatorial Atlantic Ocean basin from 28^oS to 65^oN, including the Caribbean Sea and the Gulf of Mexico, but excluding the Mediterranean Sea (see Figure 1), with a horizontal resolution of 1/12 degree (mesh size on the order of 6 km) and 16 layers in the vertical. The vertical grid was chosen to provide maximum resolution in the upper part of the ocean. The bottom topography is derived from a digital terrain data set with 2.5' latitude-longitude resolution. A Kraus-Turner mixed layer parameterization is included. All surface boundary conditions are based on seasonal climatological data sets from COADS (Comprehensive Ocean-Atmosphere Data Set). The surface thermal boundary conditions are specified by a linear bulk formula as in Han (1984) while those for fresh water flux are a direct prescription of the E-P (Evaporation - Precipitation) climatology with a small relaxation to climatological surface salinities. Open ocean boundaries are treated as closed, but are outfitted with buffer zones in which the temperature T and salinity S are linearly relaxed toward their seasonally varying climatological values. These buffer zones restore the T and S fields to the initial Levitus climatology in order to approximately recover the vertical shear of the currents through geostrophic adjustment. The three buffer zones are located at the northern boundary and at the southern boundary (3 degree latitude band) and in the Gulf of Cadiz (representing the Mediterranean Sea outflow). The restoring time scale for the boundaries varies linearly from 30 days at the inner edge to 5 days at the walls.

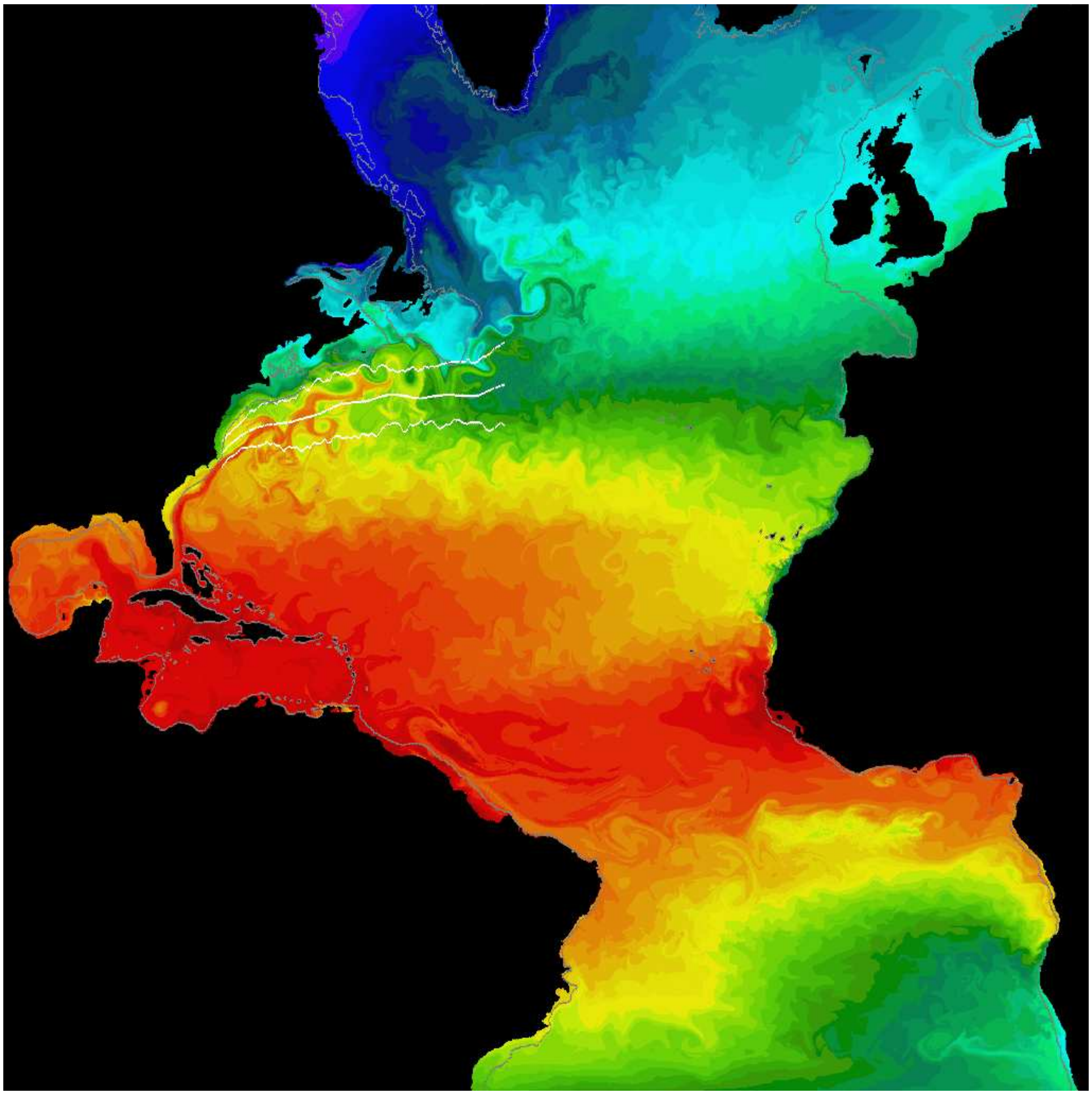

Figure 1: Model SST field year 14 day 273 (red denotes warm temperatures, blue cold temperatures). The observed mean position of the Gulf Stream is represented by the dotted line. The envelope is defined by the solid lines.

Data Assimilation Techniques

The data assimilation techniques consist of an (1) Extended Kalman Filter, hereafter EKF, with a Gauss-Markov Random Field, hereafter GMRF, parameterization of the covariance functions needed for assimilating SSH data and (2) the Parameter

Matrix Objective Analysis, hereafter PMOA, algorithm (Mariano and Brown, 1992) for assimilating SST and mixed layer data. The PMOA is a data-driven analogue of a KF that uses an assumed horizontal covariance model with nine adjustable parameters. The novelty of this approach is that the nine covariance parameters are allowed to vary in time and space as a function of the data.

A GMRF model is the spatial analogue of an Auto-Regressive model in time-series modeling. This model allows us to reduce the $O(n^2)$ covariance information needed for data assimilation to O(10n) for an ocean circulation model with a state-space (number of model prognostic variables times the number of horizontal grid points times the number of isopycnal layers) of size n. In addition, a fast updating algorithm for the inverse of the covariance matrix, known as the information matrix, and an EKF based on the information matrix has been formulated, tested, and verified (Chin *et al.*, 1999). In this study, a number of revelant dynamical models: advection, diffusion, reduced-gravity shallow water, and multi-layer wind-forced double gyre were tested under a variety of worst-case initial conditions and over a range of signal-to-noise ratios. In all of these studies, our proposed method produced the desired results of an exponential decrease in the forecast error over time for all of the state variables. The method is very robust to analysis assumptions.

The 3-D correlation structure, needed for optimal data assimilation, is decomposed into horizontal correlations and vertical correlations by our algorithm. The EKF with a GMRF model determines the horizontal correlation structure and the correlations between the state variables in an isopycnal layer. The vertical correlations are based on ensemble averages of the vertical covariances and variances determined from the model output. We are presently determining optimal updating strategies for the spatially varying vertical correlations. When the assimilation of real data is started, vertical correlations between Topex-Poseidon SSH data and the layer thicknesses will be computed and compared to the model correlations. Consequently, our proposed methodology is a blend of an Ensemble Kalman Filter in the vertical and an EKF based on the tangent linear approximation in the horizontal.

Performance

The message-passing version of MICOM called MP-MICOM was developped by computer scientists (http://www-mount.ee.umn.edu/˜okeefe/micom/) at the University of Minnesota. MP-MICOM uses SHMEM on the Cray T3E and MPI on other machines. A shared-memory, multi-threaded version (SC-MICOM) is also available for SMP clusters that uses MPI to communicate between machines and direct shared memory within each machine.

The model integration is performed on the NAVO T3E at the Stennis Space Center. MICOM's scalability on the T3E is excellent. In Bleck *et al.* (1995), linear scalability to 512 processors was shown for this type of machine on a 512 x 512 x 11 grid. In an unpublished study by Cray Analyst Mike O'Neill, MICOM was also shown to scale linearly to 1024 processors. When switching from the T3D to the T3E, a speed ratio of 3 was obtained with the streams option turned on the T3E. On the NAVO 128 T3E nodes (number of processors for the 1/12 degree mesh North Atlantic run), about 5 GFlops are achieved.

One simulated day requires about 90 SUs of T3E. This number becomes substantial once data assimilation capabilities are added. One simulated day with data assimilation increases the number of SUs by a factor of 10 to 20 depending on the order of the GMRF model used in the EKF. The simpler PMOA approach only adds 25% to the calculation. One simulated year with data assimilation would therefore require between 300,000 and 600,000 SUs for the near-optimal EKF data assimilation.

The breakdown of T3E usage per year is estimated as follows:

- FY 99 - Completion of the spin-up (210,000 SUs) and test of the data assimilation techniques (90,000 SUs)
- FY 00 - Production run (400,000 SUs)
- FY 01 - Production run (400,000 SUs)

Progress Report

The time-table of our project is the following

- FY 99: Run forward ocean model to statistical equilibrium to generate autocorrelation statistics needed for the data assimilation; write parallel assimilation code. Assemble historical data sets for data base.
- FY 00: Demonstrate assimilation model on historical data. Initial model simulations and hindcasts.
- FY 01: Implement near-real time system. Use data mining techniques with model output and verification data to check the Kalman filter estimates. Distribute products to the NAVY labs (FMNOC, NRL, ...).

Our fine-mesh simulation of the North Atlantic circulation is presently in its 17th year of integration. This simulation has generated considerable interest in both the computing and the oceanographic community. The term "fine-mesh" describes a horizontal grid resolution – typically of order 10 km – which allows barotropic/baroclinic instability, shear instabilities typical for geophysical flows, to be modeled. Since these instabilities cause ocean currents to meander and break up into individual eddies, fine-mesh ocean models are also referred to as "eddy-resolving".

In this configuration, a realistic result for the Gulf Stream separation is achieved (Figure 1). This result supports the view that an inertial boundary layer (which results from the fine resolution) is an important factor in the separation process. It

Figure 2: Two-week composite of surface float trajectories.

was also the first simulation with a fully thermodynamic basin-scale model to simulate the separation in a realistic fashion. This simulation allows direct and detailed comparisons with observations such as satellite data (sea surface height, sea surface temperature), moorings measurements, inverted echo sounders, and free-floating drogues. For example, a total of 15,000 model floats were released at 3 levels (surface, 400 meters, and 1500 meters) (Figure 2) and their statistics are presently being compared to their observed counterparts. For a detailed look at the results, please check the web site: http://www.rsmas.miami.edu/groups/micom.html

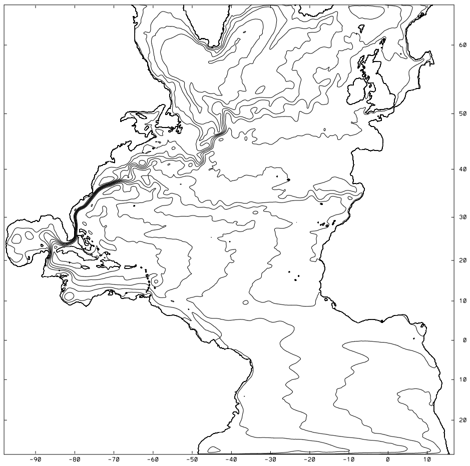

Figure 3: Mean model SSH field with Laplacian viscosity operator

During the first half of FY99, the ocean model was integrated to year 17 and is expected to reach a statistical equilibrium at year 20. During this spin-up phase, the impact of the the subgrid-scale parameterization on the Gulf Stream separation and penetration into the domain interior was investigated. With a viscosity Laplacian operator, the Gulf Stream penetration was found to be relatively small (Figure 3) and the eddies to decay quite rapidly. With a higher order operator (biharmonic), the eddies were found to keep their structure for long period of times, but the Gulf Stream on the other hand separated early from the coast (Figure 4). A combination of these two operators led to a satisfactory Gulf Stream behavior (Figure 5). These results illustrates the high sensitivity of the solution to the choices one makes to parameterize the processes not represented by the model grid.

Once statistical equilibrium is reached at year 20, the COADS wind forcing will be modified to include daily wind variations based on a 1990-1995 European Center for Medium Weather Forecasting (ECMWF) dataset. The turbulent behavior of the simulation was recently assessed by Paiva *et al.* (1999). Sea surface height variability spectra in the North Atlantic subtropical gyre were computed from the model results and compared to observations and previous models, within the framework of the geostrophic turbulence theory. Despite higher eddy activity and a correct Gulf Stream separation, there is no geographic variation in spectral slope in the inertial ranges as in previous simulations with coarser resolution. Observations, on the other hand, show a flattening in the spectra derived from altimeter data when moving from the western to the eastern side of the Atlantic. The surface forcings used in this simulation are based on monthly climatologies which resolve only the seasonal cycle and spatial scales on the order of several hundred kilometers. Consequently, mechanisms such as fluctuations in wind forcing on the spatial and temporal scales of atmospheric fronts

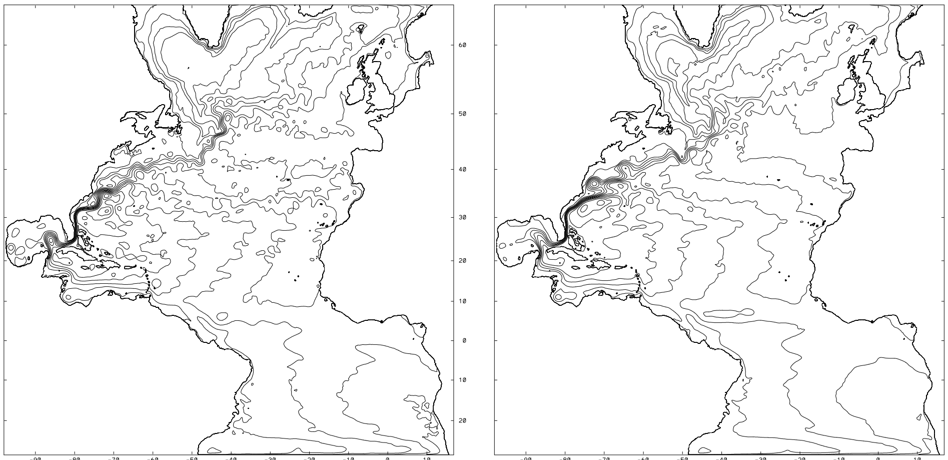

Figure 4: Mean model SSH field with biharmonic viscosity operator. (a) small coefficient; (b) large coefficient.

and eddies are not represented in these simulations. The model SSH variability only reflects the contribution from the internal oceanic instabilities. The impact of high frequency wind forcing will be investigated once the simulation has reached a steady state. The wind field consists of a superposition of monthly COADS data, ECMWF (European Center for Medium-Range Weather Forecasting) synoptic scales variability, and high frequency synthetic wind stress variability based on ERS-1 scatterometer statistics.

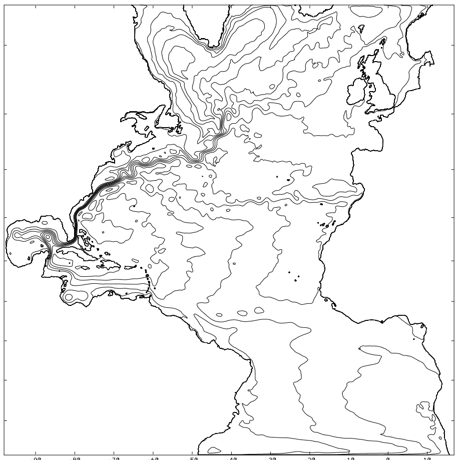

Figure 5: Mean model SSH field with Laplacian/biharmonic viscosity operator

Data assimilation capabilities are presently being added to the model and the parallel version are being tested on the T3E with a coarser version of the ocean model (1/3 degree grid spacing) using 1993/1994 Topex-Poseidon altimeter data. In addition to sea surface height (SSH), the impact of assimilating sea surface temperature (SST) fields on the surface and the interior fields will also be assessed. Several forecast experiments will be performed in FY 01 in collaboration with the Fleet Numerical Meteorology and Oceanography Center (FNMOC).

References

Bleck, R., and E.P. Chassignet, 1994: Simulating the oceanic circulation with isopycnic coordinate models. *The Oceans: Physiochemical Dynamics and Resources.* The Pennsylvania Academy of Science, 17-39.

Bleck, R., S. Dean, M. O'Keefe, and A. Sawdey, 1995: A comparison of data-parallel and message-passing versions of the Miami Isopycnic Coordinate Ocean Model (MICOM). *Parallel Computing*, **21**, 1695-1720.

Chassignet, E.P., L.T. Smith, R. Bleck, and F.O. Bryan, 1996: A model comparison: Numerical simulations of the North and Equatorial Atlantic oceanic circulation in depth and isopycnic coordinates. *J. Phys. Oceanogr.*, **26**, 1849-1867.

Chin, T.M., A.J. Mariano, and E.P. Chassignet, 1999: Spatial regression with Markov random fields for Kalman filter approximation in least-squares solution of oceanic data assimilation problems. *J. Geophys. Res.*, **104**, 7991-8014.

Mariano, A.J. and O.B. Brown, 1992: Efficient objective analysis of dynamically heterogeneous and nonstationary fields via the parameter matrix. *Deep-Sea Res.*, **39**, 1255-1271.

Paiva, A.M., J.T. Hargrove, E.P. Chassignet, and R. Bleck, 1999: Turbulent behavior of a fine mesh (1/12 degree) numerical simulation of the North Atlantic. *J. Mar. Sys.*, in press.