Real-time Ocean Data Assimilation and Prediction with Global NCOM

C. Rowley, C. Barron, L. Smedstad, and R. Rhodes

Naval Research Laboratory
Stennis Space Center, MS 39529
clark.rowley@nrl.navy.mil

Abstract – **The Naval Research Laboratory (NRL) at Stennis Space Center has developed a global implementation of the Navy Coastal Ocean Model (NCOM). Global NCOM encompasses the open ocean to 5 m depth in a curvilinear global model grid with 1/8 degree grid spacing at 45N, extending from 80 S to a complete arctic cap with grid singularities mapped into Canada and Russia. The model employs 40 vertical sigma-z levels, with sigma in the upper ocean and coastal regions, and z in the deeper ocean. The real-time system uses Navy Operational Global Atmospheric Prediction System (NOGAPS) 3-hourly wind stresses and heat fluxes. Operationally available sea surface temperature (SST) and altimetry (SSH) data are incorporated into NAVOCEANO Modular Ocean Data Assimilation System (MODAS) and Navy Layered Ocean Model (NLOM) analyses and forecasts of SSH and SST. These in turn are combined with the MODAS synthetic database to yield three-dimensional fields of temperature and salinity for assimilation into global NCOM. Here, we describe the analysis and forecast system, present selected evaluations of the model performance, and discuss planned upgrades to the model and data assimilation methods.**

I. INTRODUCTION

As part of ongoing Naval Research Laboratory research and development efforts in basin- to global-scale ocean analysis and forecast systems, a real-time data-assimilating global ocean forecast system has been developed [1]. The real-time system is currently producing daily 72 h forecasts of three-dimensional ocean circulation and thermohaline structure while undergoing evaluation prior to a transition to operational use.

The evaluation process will be based on a multiyear hindcast for the years 1997-2001. Additional model features and system upgrades are now being tested in the suite of hindcast integrations. This paper will present a description of the present real-time system, provide a sample of the model validation efforts, and discuss some of the planned improvements.

II. GLOBAL OCEAN FORECAST SYSTEM

The global ocean forecast system comprises the ocean forecast model, the data analysis component, and the data assimilation methods.

A. Navy Coastal Ocean Model

The ocean forecast model component of the real-time system is the Navy Coastal Ocean Model, NCOM [2]. NCOM is a free-surface primitive-equation model using the hydrostatic, Boussinesq, and incompressible approximations. Calculations are performed on an Arakawa C grid, using second-order, centered spatial finite differences integrated using a leapfrog technique with an Asselin time filter. The evolution of the free surface and vertical mixing are calculated implicitly. The horizontal grid is curvilinear. The vertical grid combines sigma coordinates for the upper layers and z-level coordinates for the lower layers, with the transition depth between them specified by the user. Source terms are included for riverine inputs.

The global model configuration of NCOM uses 40 vertical layers, with 19 sigma-coordinate layers in the upper 137 m, and 21 z-level coordinate levels from 137 m to 5500 m. Where the water is shallower than 137 m, the 19 sigma-coordinate layers follow the bottom. In deep water (i.e., deeper than 137 m), the vertical resolution of the sigma-coordinate layers increases from 1 m at the surface to about 20 m at the sigma-z interface. The shallowest water depth represented is 5 m. The model depth and coastline are based on a global gridded 2-minute bathymetry produced at the Naval Research Laboratory.

The horizontal grid represents the global ocean from 80 °S to a complete Arctic cap using a curvilinear grid with the grid singularities mapped onto land in Canada and Russia (Fig. 1). The grid dimensions are 2048 x 1280. For the globe south of the Arctic cap, a traditional grid is used. The longitudinal grid spacing is approximately 1/6°. The latitudinal spacing from the equator to the midlatitudes is roughly $1/6° * cos(\varphi)$, where φ is the latitude; south of 45°S, the latitudinal spacing limits to 0.1°. The midlatitude resolution is roughly 1/8° in latitude, and that is the nominal resolution of the model configuration. North of 32°N, the grid transitions to the bipolar Arctic cap (Fig. 2). The design of the polar cap follows [3]. The grid spacing is approximately 8 km in the Arctic. Boundary conditions are specified to maintain continuity across the Arctic seam.

The horizontal mixing coefficient for momentum in the model is calculated using a Smagorinsky formulation with coefficient 0.1. The mixing coefficient for the tracers is presently set equal to the momentum coefficient, for a Prandtl number 1.0. Vertical mixing coefficients are based on a Mellor-Yamada level 2 turbulence closure scheme, with an additional background Richardson number-dependent mixing. A no-slip condition is applied at the lateral land-sea boundaries.

B. Modular Ocean Data Assimilation System

The Modular Ocean Data Assimilation System, MODAS [4] is the basis of the data analysis component of

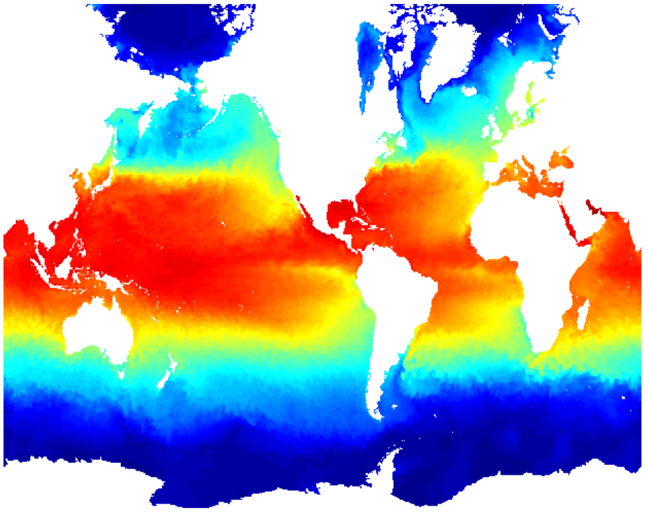

Fig. 1: The global forecast system 29 Jul 2002 72 h forecast sea surface temperature (valid time 00Z 01 Aug 2002). The shading shows the surface temperature on the model grid.

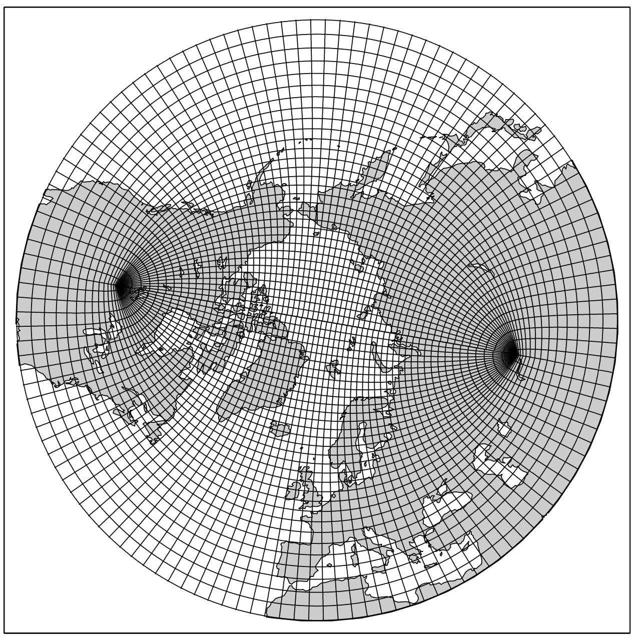

Fig. 2: The horizontal grid of the forecast model. The Arctic cap to 30 °N is shown, with only every sixteenth line of grid points plotted for clarity.

the global system. MODAS is a modular set of routines to select, sort, and quality control *in situ* and remotely sensed observations, and to produce gridded analyses of surface height, surface temperature, three-dimensional temperature and salinity, and derived fields. The observations are gridded by traditional optimal interpolation methods onto a background field (for example, a model forecast field valid at the time of the analysis, or the previous analysis field). MODAS also includes a dynamic climatology that can be used to produce synthetic temperature and salinity profiles based on observed surface variables (sea surface height anomaly and sea surface temperature). The synthetic salinity profiles are determined from the synthetic temperatures using climatological T-S relationships.

Daily three-dimensional fields of temperature and salinity are produced using the dynamic climatology, and based on gridded surface height anomaly and surface temperature fields. The surface height anomaly fields are produced operationally at the Naval Oceanographic Office using the Navy Layered Ocean Model, a high horizontal resolution (about 1/16° degree) 6-layer model that assimilates altimeter-derived sea surface height anomaly observations. The gridded sea surface temperature field is produced by using MODAS to perform a two dimensional optimal interpolation of satellite-derived AVHRR temperature observations. The surface height anomaly and surface temperature fields are then passed to the dynamic climatology to produce synthetic temperature and salinity profiles at each analysis grid point. Finally, the analysis fields are interpolated to the curvilinear model grid.

C. *Forcing and data assimilation*

The model is forced primarily with 3 h wind stress and total heat flux fields from the Navy Operational Global Atmospheric Prediction System, NOGAPS. The real-time system has analysis-quality atmospheric forcing for times prior to the nowcast time, and forecast-quality forcing for the forecast period.

The daily three-dimensional fields are assimilated into the ocean model by relaxation. The coefficient that sets the strength of the relaxation is depth-dependent, increasing from 0 at the surface to 1 at depth with an e-folding scale of 200 m. The resulting time scale for the relaxation is a minimum 48 h at depth, and increases to the surface. The design is intended to allow a realistic mixed layer to develop in response to the surface forcing, but introduce the surface height anomaly data down through the thermocline via the synthetic profiles.

An additional relaxation of surface temperature and salinity is imposed. The reference temperature and salinity fields for the surface relaxation are the MODAS surface temperature analysis and the corresponding estimated surface salinity. The surface flux is imposed using a specified vertical flux velocity of 5 m/d for temperature and 0.5 m/d for salinity, so the effective time scales are approximately 5 to 10 d for temperature and 50 to 100 d for salinity, assuming mixed layer depths of 25 to 50 m. The surface salinity flux is a proxy for the surface flux of fresh water (net evaporation less precipitation), which is not included in the forcing.

Each daily forecast is made as a single 6 d integration; a 72 h hindcast brings the model up to the nowcast time, which is the initial state for a 72 h forecast period. The most recent hindcast fields for each 3 h time form an archival time-series of the model fields. Forecast fields are also archived every 3 h for verification of the model forecast skill.

The real-time system began assimilating data in October 2000 from a multi-year climatological spin-up of the 1/8° model configuration, and is up-to-date at the time of this writing (01 August 2002). Hindcast fields are archived from the start of the integration, and forecast fields are archived from 20 February 2002 on.

D. Operational considerations

Running a global domain model operationally poses many challenges even with the high performance computing resources available. With increasing resolution come increasingly burdensome storage requirements, in turn requiring a network capable of moving large files quite efficiently. The three-dimensional files created by the present system take up 1.6 GB. Storing these several times for each forecast day, along with regular 4 GB restart files and 10 GB input files, requires some logistical planning to utilize the multiple resources in a center such as the Naval Oceanographic Office (NAVO) Major Shared Resource Center (MSRC).

The real-time system uses three platforms at the NAVO MSRC. A Sun E10000 is used for pre-processing and post-processing, and an IBM SP is used for the execution of the model. Therefore, over 10 GB each day must be transferred to the IBM SP from the Sun E10000. The system presently allows 4 hours for creation of input files and their transfer. After the daily model run is finished, (usually in less than 2 hours), another 4 hours is allowed for transfer to, and post-processing on, the Sun E10000, and for a final transfer of output and restart files to the third platform, a mass storage server.

III. RESULTS

The nominal 1/8° resolution of the global system may be considered eddy resolving, but is not sufficient to properly treat the behavior of separating western boundary currents [5]. In coarse resolution models of the Gulf Stream, the current typically overshoots the climatological separation point, and the separated jet is often too weak to penetrate the basin. The assimilation of data is expected to improve the model results, at least to a degree. In a comparison of the mean surface velocity kinetic energy from a multi-year climatological run, with the year 2001 mean calculated from the real-time (data-assimilating) system, the introduction of observations significantly improves the simulation of both the separation latitude of the Gulf Stream, and the downstream path (Fig. 3). The introduction of data assimilation also improves the jet penetration in the model; this is apparent in the distribution of the surface eddy kinetic energy for the same two integrations (Fig. 4). Similar improvements are noted in the behavior of the Kuroshio and Agulhas Current.

In a first evaluation of forecast skill, the root mean square error (RMSE) of the model at predicting its future nowcast surface temperature analysis at lead forecast times of 48 h and 72 h is compared with the RMSE from persisting the previous nowcast for the same lead time. Using 41 validation times in the period 10 March to 31 May 2002 when all forecast times available, global fields of forecast and persistence error were calculated. For the

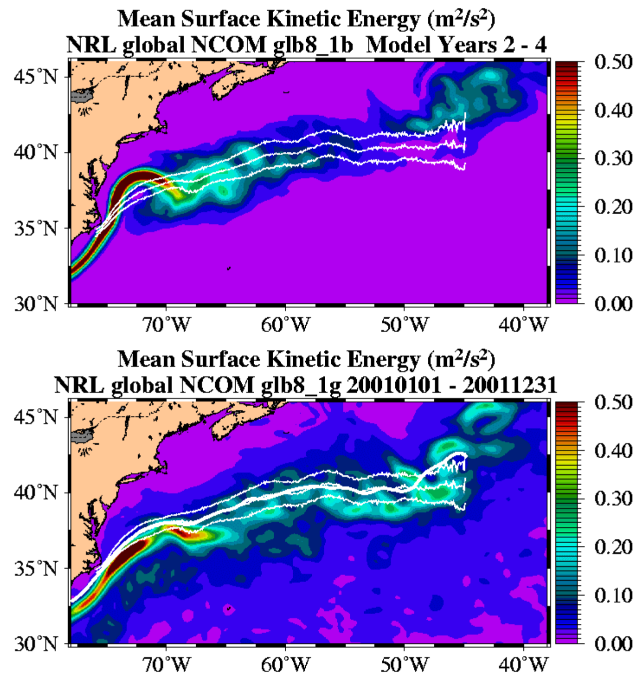

Fig. 3: Mean surface kinetic energy fields for three years of a climatological integration of the model (upper panel) and for year 2001 of the data-assimilating real-time run. For reference, the DAMEE mean Gulf Stream path and its envelope (both panels) and the 2001 mean NAVOCEANO frontal analysis (bold line, lower panel only) are overlaid.

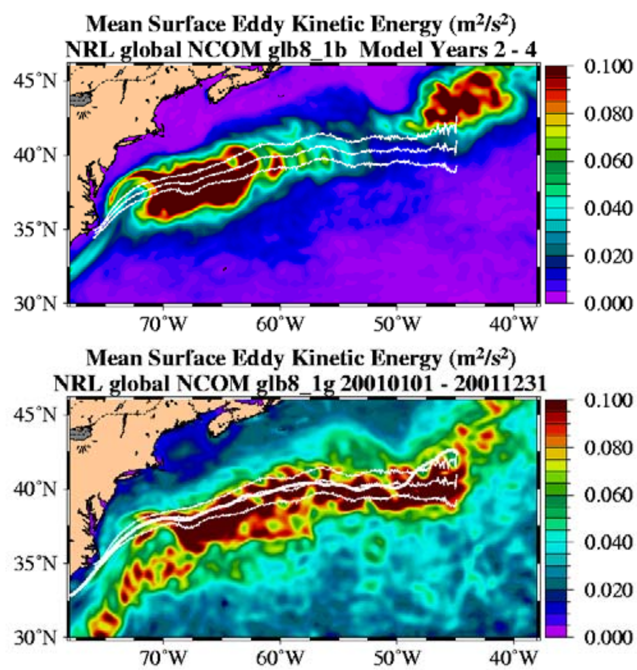

Fig. 4: Mean surface eddy kinetic energy fields for three years of a climatological integration of the model (upper panel) and for year 2001 of the data-assimilating real-time run. For reference, the DAMEE mean Gulf Stream path and its envelope (both panels) and the 2001 mean NAVOCEANO frontal analysis (bold line, lower panel only) are overlaid.

Kuroshio, Agulhas, and Gulf Stream regions, the model forecast is much more skillful than persistence at predicting its own future nowcasts (Figs. 5 to 7). The model might be expected to provide the greatest improvement in surface temperature forecast skill in areas of strong currents along sharp gradients; this is indeed what we see in the significant decreases in the largest errors along the fronts. The model skill is not limited to areas close to the strong fronts, though, as there is also a considerable reduction of error away from the jets. There is a falloff in the improvement over persistence; in the Kuroshio region, the area-average forecast RMSE increases from 0.22 °C at 48 h to 0.30 °C at 72 h, but the persistence error increases from 0.34 °C to 0.40 °C, a smaller percentage increase (Figs. 5 and 6).

A direct comparison with observations is a stricter measure of forecast skill. In matching nowcasts and 24 h, 48 h, and 72 h forecasts against satellite-derived AVHRR temperature observations (the same measurements used to produce the assimilated analyses), the Agulhas and Gulf Stream regions presented above, and the Arabian Sea (45ºE to 78ºE, 0ºN to 25ºN), have significant numbers of observations available within +/- 6 h of the valid times of the 00 UTC forecasts. Forecast skill is measured against persistence of the MODAS two-dimensional optimum interpolation analysis of sea surface temperature valid at the start time of the forecast, and referenced to the observations. (This analysis is subject to biases. For example, the 00 UTC forecasts for the Gulf Stream region are generally being compared with 18 UTC data, and the horizontal distribution of data is not uniform over the specified analysis region. A routine global matching of forecasts with MCSST observations using the 3 h forecast fields is being developed).

A skill score is defined as

$$SS=(1 - MSE^F/MSE^M),$$

where MSE^F and MSE^M are the mean square error of the forecast and MODAS analysis, respectively [6]. A positive (negative) skill score implies the forecasts are more (less) skillful than persistence of the MODAS analysis at predicting the AVHRR observations.

A sample of 86 (85, 85, 84) 00 UTC validation times in the period 01 Apr to 01 Aug 2002 is used, when forecasts were available for lead time 00 h (24 h, 48 h, 72 h). A subset of 29 validation times was used for comparison, when all four lead times were available for each valid time. This has the effect of eliminating forecast integrations where the continuous model time series was interrupted, and the model nowcast degraded by the interruption ("good" validation times). Skill scores are presented in Table 1 for both samples, for the three regions, at the three forecast lead times and the 00 h lead nowcast (we note that the nowcast MODAS and forecast fields have both used some of the observational data used here for reference).

The initial negative skill scores at the nowcast time (00 h lead) suggest that the initial states for the model forecasts

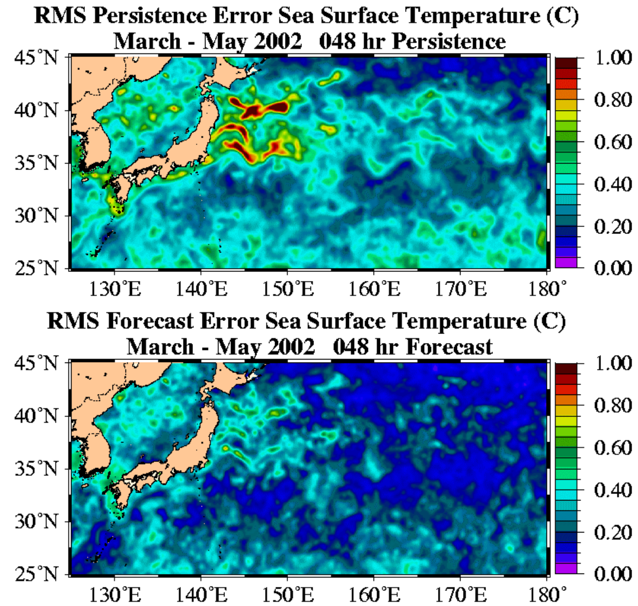

Fig. 5: The persistence root-mean-square error at 48 h for 41 forecasts from March to May 2002 measured against the nowcast at the valid time of the forecast for the Kuroshio region (upper panel). The forecast error at 48 h for the same period, measuring a 48 h forecast against the nowcast at the valid time (lower panel).

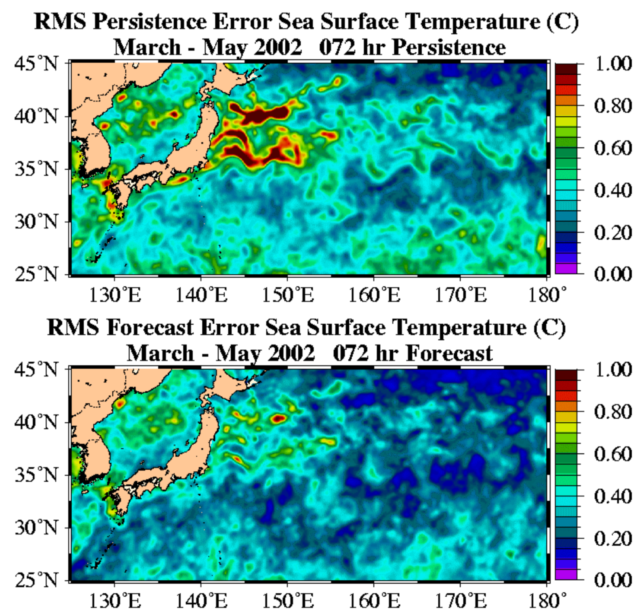

Fig. 6: The persistence root-mean-square error at 72 h for 41 forecasts from March to May 2002 measured against the nowcast at the valid time of the forecast for the Kuroshio region (upper panel). The forecast error at 72 h for the same period, measuring a 72 h forecast against the nowcast at the valid time (lower panel).

are not as accurate as the MODAS gridded analysis of the observations. At later times, the model gains skill relative to the persisted analysis, and for the Gulf Stream region is more skillful than the persisted analysis at 48 h and 72 h. In the Agulhas region, the nowcast appears to be quite

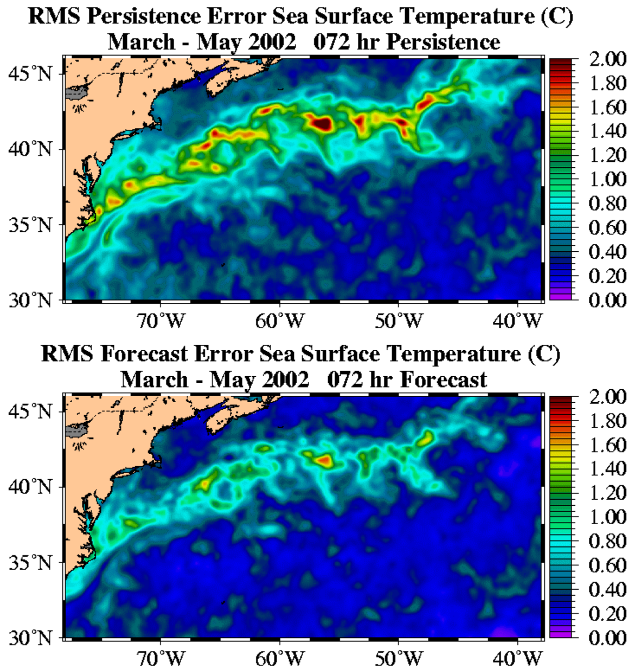

Fig. 7: The persistence root-mean-square error at 72 h for 41 forecasts from March to May 2002 measured against the nowcast at the valid time of the forecast for the Gulf Stream region (upper panel). The forecast error at 72 h for the same period, measuring a 72 h forecast against the nowcast at the valid time (lower panel).

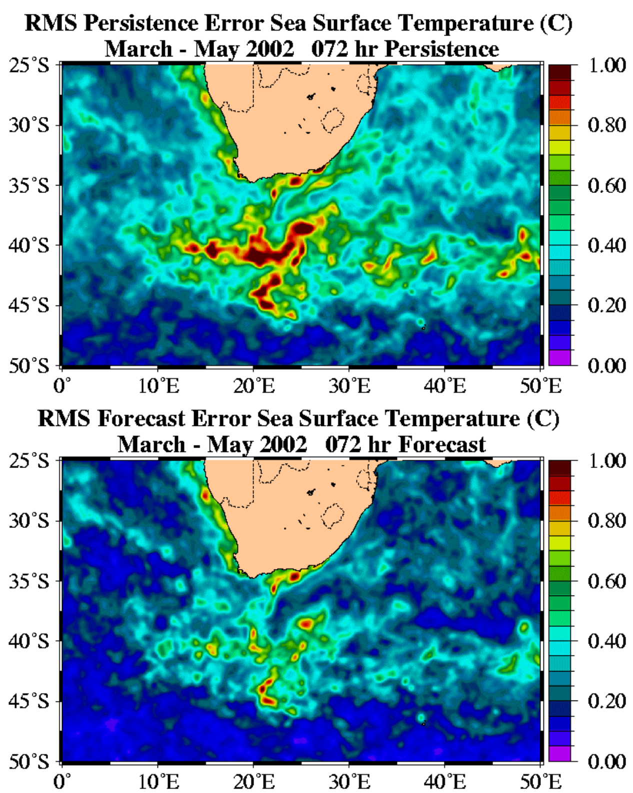

Fig. 8: The persistence root-mean-square error at 72 h for 41 forecasts from March to May 2002 measured against the nowcast at the valid time of the forecast for the Agulhas region (upper panel). The forecast error at 72 h for the same period, measuring a 72 h forecast against the nowcast at the valid time (lower panel).

Table 1: Skill scores of the global forecast system compared to persistence of the MODAS analysis for sea surface temperature, for the Gulf Stream region, the Agulhas region, and the Arabian Sea, at lead times of 00, 24, 48, and 72 h.

Valid times	Region	Lead time			
		00 h	24 h	48 h	72 h
All	Gulf Stream	-1.92	-0.02	+0.07	+0.12
	Agulhas	-4.09	-0.42	-0.15	-0.11
	Arabian Sea	-2.51	-0.24	-0.01	+0.05
Good	Gulf Stream	-1.97	+0.01	+0.17	+0.21
	Agulhas	-5.64	-0.35	-0.06	+0.03
	Arabian Sea	-2.45	-0.21	+0.02	+0.16

poor. The subset of "good" forecasts, where the model nowcast was not degraded by an interruption in the daily integration cycle, shows positive skill over persistence at 72 h for all three regions (and, hence, the improvement that will result from an operational system that is monitored to avoid interruptions to the continuity of the hindcast time series).

The increase in skill with lead time compared to the persisted MODAS analyses, and the falloff in skill compared to the persisted nowcast, both suggest that the model is drawing toward a better forecast state from a less accurate nowcast. Effort is being directed at improving the model initialization.

Other verifications of model skill being performed but not shown here include comparisons of model transports to observed values, direct comparisons of model temperature, salinity, and mixed layer depth with long mooring time series, and model errors referenced to observed profiles (i.e., BTs, profiling floats).

IV. DISCUSSION

During the model evaluation process prior to the planned transition to operational use, several improvements to the forecast system configuration are being tested in the multiyear hindcast integrations. These include an updated bathymetry, a correction to the Arctic climatology, improvements to the application of surface heat fluxes, and changes in the data assimilation.

An updated version of the NRL 2-minute bathymetry is available, and it was noted that the model depth in some key transport locations (e.g., the Florida Straits) is improved with the new bathymetry. The effect on the model transports is not expected to be large, but there should be an overall improvement in the accuracy of the model bottom depths.

The model Arctic temperatures and salinities have been improved by replacing the background values from 80°N to the pole with interpolated monthly fields from a new version of the Naval Oceanographic Office Generalized Digital Environmental Model (GDEM). Previously, MODAS climatological fields were smoothed to remove discontinuities at the pole, and artifacts of the process remained in the fields. This change has already been implemented in the real-time run.

The imposition of heat fluxes calculated from an atmospheric model with no feedback from the ocean can create large biases in the net surface fluxes and the resulting ocean heat content. An updated scheme [7] has been implemented in NCOM and should produce more realistic surface flux fields. The addition of surface freshwater flux would improve the model surface salinity, but is not planned at this time. A new global database of river discharges is being tested in the hindcast run, with the volume fluxes of over 900 rivers worldwide being added (C. Barron and L. Smedstad, this issue).

Adjustments are being made to both the surface and subsurface data assimilation methods. A weaker relaxation to the reference surface fields is being tested, and the ability to allow the timescale of the subsurface relaxation to vary geographically as well as in the vertical is planned. This will make it possible to reduce the impact of the data assimilation in locations where the model shows more skill.

V. SUMMARY

The real-time system has now completed almost two years of daily data-assimilative forecast runs, and continues to serve both as a useful development and testing tool, and as a benchmark against which to measure future system performance.

ACKNOWLEDGMENTS

This work is supported by the Oceanographer of the Navy through the Space and Naval Warfare Systems Command PMW155, as part of the NRL 6.4 Large-scale Models project and the NRL 6.4 Ocean Data Assimilation project, under program element 0603207N. Computational resources were provided by the Department of Defense High Performance Computing Modernization Program through the Naval Oceanographic Office Major Shared Resource Center.

REFERENCES

[1] R. Rhodes, et al., “Navy real-time global modeling systems,” *Oceanography*, vol. 15, pp. 29-43, 2002.

[2] P. Martin, “Description of the Navy Coastal Ocean Model Version 1.0,” NRL Report No. NRL/FR/7322/00/9962, 45 pp. Available from NRL Code 7322, Bldg. 1009, Stennis Space Center, MS 39529.

[3] R. Murray, “Explicit generation of orthogonal grids for ocean models,” *J. Comp. Phys.,* vol. 126, pp. 251-273, 1996.

[4] D. Fox, W. Teague, C. Barron, M. Carnes, and C. Lee, “The Modular Ocean Data Analysis System (MODAS)”, *J. Atmos. Ocean. Tech.,* vol. 19, pp. 240-252, 2002.

[5] H. Hurlburt and P. Hogan, “Impact of 1/2° to 1/64° resolution on Gulf Stream model-data comparisons in basin-scale subtropical Atlantic Ocean models,” *Dyn. Atmos. Oceans*, vol. 32, pp 283-329, 2000.

[6] Murphy, A., and E. Epstein, “Skill scores and correlation coefficients in model verification,” *Mon. Weather Rev*., vol. 117, pp. 572-581, 1989.

[7] B. Kara, P. Rochford, and H. Hurlburt, “Efficient and accurate bulk parameterizations of air-sea fluxes for use in general circulation models,” *J. Atmos. Ocean. Tech*., vol. 17, pp. 1421-1438, 2000.