Ocean State Estimation and Prediction in Support of Oceanographic Research

Detlef Stammer

The ECCO Consortium:

*D. Stammer (PI), R. Davis, L.-L. Fu, I. Fukumori, R. Giering, T. Lee, J. Marotzke, J. Marshall,
D. Menemenlis, P. Niiler, C. Wunsch, V. Zlotnicki*

Eric Chassignet

The HYCOM Consortium:

*E.P. Chassignet (PI), C. Barron, R. Bleck, J.F. Cayula, T.M. Chin, G. Halliwell,
P. Hogan, H. Hurlburt, G. Jacobs, A. Mariano, O.M.
Smedstad, R. Rhodes, W.C. Thacker, A. Wallcraft*

T he ocean is changing vigorously on a wide range of time and space scales. This variability leads to substantial problems in observing and modeling (simulating) the rapidly changing flow field, the ocean's temperature distribution, and more generally the consequences of those changes on a wide variety of scientific, military, and societal problems, including climate research, fisheries management, coastal pollution (spill) predictions, or ship routing. As an example, the ocean carries roughly 50% of the heat from the low latitudes to mid and high latitudes, where it is released to the atmosphere and moderates our climate. Any significant variation in this heat transport will necessarily lead to a perturbation of the climate system. Among the goals of the present ocean research are therefore to measure, understand, and eventually predict these variations by combining ocean data and ocean models.

Prototype ocean observing systems are now in place as a legacy of programs such as the World Ocean Circulation Experiment (WOCE) and the Tropical Ocean Global Atmosphere Program (TOGA), which aimed at measuring and detecting large-scale changes of various quantities in the ocean, such as temperature, salinity, velocity, nutrients, tracers, etc. At the present time, continuous global coverage is only provided for surface data via satellite measurements (sea surface height, wind, temperature, color, etc.). Measurements in the interior ocean, on the other hand, are currently quite sparse in space and time. However, the coverage of interior temperature and salinity measurements will significantly increase with the launch of the global

ARGO float program (see Roemmich and Owens, this issue), which in combination with the complimentary high-quality altimetric satellite data will be the backbone of a climate observing system.

However, in spite of these unprecedented data, the interior of the ocean will remain fairly undersampled, and much of our understanding and inference about the ocean's role in shaping our climate will come from additional sources of information. Part of this understanding is provided by numerical simulations of the ocean which are performed with realistic coastlines, bottom topography, and surface forcing. We now have the computer resources at our disposal to run such numerical models on basin-scale with horizontal resolution of 1/12° or better and globally with 1/6° resolution or better. Results show an intriguing degree of realism, but reveal also serious shortcomings in the numerical models (see Fu and Smith, 1996; Stammer et al., 1997; Paiva et al., 1999; Hurlburt et al., 2000; Smith et al., 2000, for detailed studies of General Circulation Model consistency with data).

By combining ocean observations and models, one can obtain an analysis of the time-varying ocean that, when taking into account errors in data and models, must necessarily be more complete and better than the information from either of them alone. This is the heart of ocean state estimation (often referred to as "data assimilation") the goal of which is to obtain the best possible description of the changing ocean by forcing the numerical model solutions to be consistent with the observed ocean conditions. This by itself is a very cost-

effective way to obtain a fairly complete description of the changing ocean from a limited set of observations. But at the same time it also identifies those model components that need improvement, and guides us as to where we need to extend the observing system to improve the estimated ocean state.

There is a close analogy between ocean state estimation and the data assimilation efforts ongoing in the meteorological community for weather forecasting. But there are many specific differences; e.g., oceanography still has a major emphasis on understanding, rather than the practical goal of daily forecasting. While there is a clearly identifiable user community for weather forecasting that provides regular feedback and evaluation of the products, an equivalent community is not so clearly identifiable in oceanography. Furthermore, the fact that the nature and coverage of oceanic data is usually radically different from that of the weather observing network in the atmosphere places much higher demands on the quality of ocean models and on the computing (hardware) capability.

Future oceanographic and climate research will fundamentally rely on our ability to do ocean state estimation in a regular manner. It is therefore the goal of the Global Ocean Data Assimilation Experiment (GODAE) to demonstrate the practicality of continuous ocean state estimation given all the different elements already in our hands today (the observing system, telemetry, data processing and archiving, numerical modeling and ocean state estimation). At the same time, GODAE aims to highlight the importance of these estimates by providing results and products important to build up users' groups in many areas of the community.

New NOPP-Supported Activities

Because of the fundamental importance of understanding the present and future states of the ocean, two consortia on ocean modeling and state estimation were proposed recently which are now supported through the U.S. National Ocean Partnership Program (NOPP) with funding provided by the National Aeronautics and Space Administration (NASA), the National Science Foundation (NSF), and the Office of Naval Research (ONR). The general goals of these two modeling and assimilation groups are to transform existing ocean state estimation efforts from their present experimental status into quasi-operational applications that can demonstrate the practical utility of their results. These activities are critical to the success of programs such as the Climate Variability Experiment (CLIVAR) and GODAE.

The projects' nature requires synergy across modeling, observation, and estimation, and the resulting scope and duration necessitate a unique form of sponsorship which is not easily available, i.e., through NSF core funding procedures. This obvious gap in the funding structure for longer-term projects was recently closed by the NOPP program, which now provides a suitable framework to support such a long-range and difficult task over five years. Through these newly formed NOPP activities, it has become possible to bring together into two consortia the various complementary elements that, until recently, were operating primarily as individual entities. The foci of those two activities are related, but distinct. The differences in emphasis can be summarized as large-scale climate syntheses versus mesoscale forecasting. They also differ by their technical approach as discussed in some detail below.

The first of the two NOPP consortia is called "Estimation of the Circulation and Climate of the Ocean" (ECCO), which builds on existing efforts at the Massachusetts Institute of Technology (MIT), the Jet Propulsion Laboratory (JPL) and the Scripps Institution of Oceanography (SIO), with additional partners at the Southampton Oceanographic Centre (SOC) and the Max-Planck Institut für Meteorologie (MPI) in Hamburg. The primary goal of this group is to provide the best possible and dynamically consistent estimate of the ocean circulation, which can serve as a basis for studies of elements important to climate (e.g. heat fluxes and variability).

The second consortium focuses on adding data assimilation capabilities to the HYbrid Coordinate Ocean Mode (HYCOM). The primary goal here is to develop accurate forecasting capabilities for the energetic mesoscale eddy fields, encompassing the establishment of a global real-time ocean forecast system with the data assimilation techniques that can be efficiently executed on massively parallel computers using a hybrid coordinate model (Bleck and Boudra, 1981; Bleck and Benjamin, 1993). Partner organizations are the university of Miami/RSMAS, the Naval Research Laboratory (NRL), the National Oceanic and Atmospheric Administration (NOAA/AOML), the University of Minnesota, Los Alamos National Laboratory (LANL), Planning Systems Inc., Orbital Image Corp., and the U.S. Coast Guard. Collaborations have also been established with the French Service Hydrographique and Océanographique de la Marine (SHOM).

A major issue for these two NOPP consortia, and generally for the wider oceanographic community, is the way in which the need for computer resources has now out-stripped their availability. No short-term solution to the computer resource bottleneck is as yet visible. However, the ongoing NOPP activity in organizing a substantial increase and improvement in computational infrastructure for oceanographic research, if successful, will have a profound impact on many future NOPP modeling and assimilation activities.

Approaches

All data assimilation methods deal, in principle, with a vector $\mathbf{x}(t)$ representing the state of the ocean at time, t, a model, $\mathbf{x}(t + 1) = L(x(t), \mathbf{B}q(t))$, and a set of data $\mathbf{y}(t) = \mathbf{E}(t)\mathbf{x}(t) + \mathbf{n}(t)$, where $\mathbf{n}(t)$ is observational noise relat-

ed to the state and $\mathbf{E}(t)$ is a known matrix. A solution $\mathbf{x}$ over time t is sought that minimizes a sum of model-data misfits and deviations from model equations, which can be written as

$$J=\sum_t(\mathbf{y}(t)-\mathbf{Ex}(t))^T\mathbf{R}(t)^{-1}(\mathbf{y}(t)-\mathbf{Ex}(t))+\sum_t(\mathbf{x}(t+1)-$$
$$\mathbf{L}[\mathbf{x}(t)])^T\mathbf{Q}(t)^{-1}(\mathbf{x}(t+1)-\mathbf{L}[\mathbf{x}(t)]). \quad (1)$$

Matrices $\mathbf{R}$ and $\mathbf{Q}$ are weights that correspond to uncertainties (error covariances) of the data and the model, respectively. In practice, a priori error estimates are often simply guesses, and the validity of the assumptions must be carefully assessed to assure the quality and integrity of the estimates.

Assimilation schemes in use today vary in the degree to which they attempt to accurately find the stationary values of J; the sophistication and accuracy of their treatment of $\mathbf{R}$, $\mathbf{Q}$; in their ability to produce estimate of the reliability of the final solution; in the extent to which they encompass all of the available observations; and in regard to whether minimizing J is treated as a constrained or unconstrained optimization problem. The capabilities today range from simple, crude methods such as "nudging" to statistically rigorous methods such as the Kalman filter-plus-smoother sequential method, and the adjoint iterative type. The simple methods are computationally fast and cheap, but produce results that one can attempt to turn into fully rigorous ocean estimates. See Malanotte-Rizzoli (1996) for an overview on recent data "assimilation" in the ocean.

Ongoing Activities

The scientific goal of ECCO is to understand, describe, and forecast the global general circulation of the oceans and its role in climate. The model-based analysis of the large-scale ocean data set will enable a complete (i.e., including aspects not directly measured) dynamical description of ocean circulation. It will provide insights into the nature of climate-related ocean variability, major ocean transport pathways, heat and freshwater flux divergences (similar for tracers, oxygen, silica, and nitrate), the location and rate of ventilation, and the ocean's response to atmospheric variability. At the same time the analysis allows assessment of errors in data, surface forcing fields, and internal model parameters, including initial conditions and lateral boundary conditions. Many more applications are anticipated and have begun already, such as real-time now-casting and prediction for seasonal-to-interannual an ENSO-related changes, or studies of the marine geoid and the ocean's impact on the earth angular momentum budget (e.g. Ponte and Stammer, 1999).

The ongoing ocean state estimation is based on the MIT GCM (Marshall et al., 1997a,b) and two parallel optimization efforts; the adjoint method (Lagrange multipliers or contstrained optimization method), exploiting the Tangent-linear and Adjoint Compiler (TAMC) of Giering and Kaminsky (1997), and a reduced

state Kalman filter, e.g. Fukumori et al. (1999). Ongoing computations include a 6-year estimate of the time-evolving ocean circulation (1992 through 1997) with up to 1° spatial resolution, a complete mixed layer model (Large et al., 1994), and an eddy parameterization scheme (Gent and McWilliams, 1990). At present, data employed in J include the absolute and time-varying temperature/pressure data from October 1992 through December 1997, sea surface height anomalies from the ERS-1 and ERS-2 satellites, monthly mean sea-surface temperature data (Reynolds and Smith, 1994), time varying NCEP reanalysis fluxes of momentum, heat and freshwater, and NSCAT estimate of wind stress errors. Monthly means of the model state are required to remain within assigned bounds of the monthly mean Levitus et al. (1994) climatology. To bring the model into consistency with the observations, the initial potential temperature (Θ) and salinity (S) fields are modified, as well as the surface forcing fields. Changes in those fields (often referred to as "control" terms) are determined as a best-fit (in a least-squares sense) of the model state to the observation and their uncertainties over the full data period. In the current configuration, there are 108 elements in the control vector.
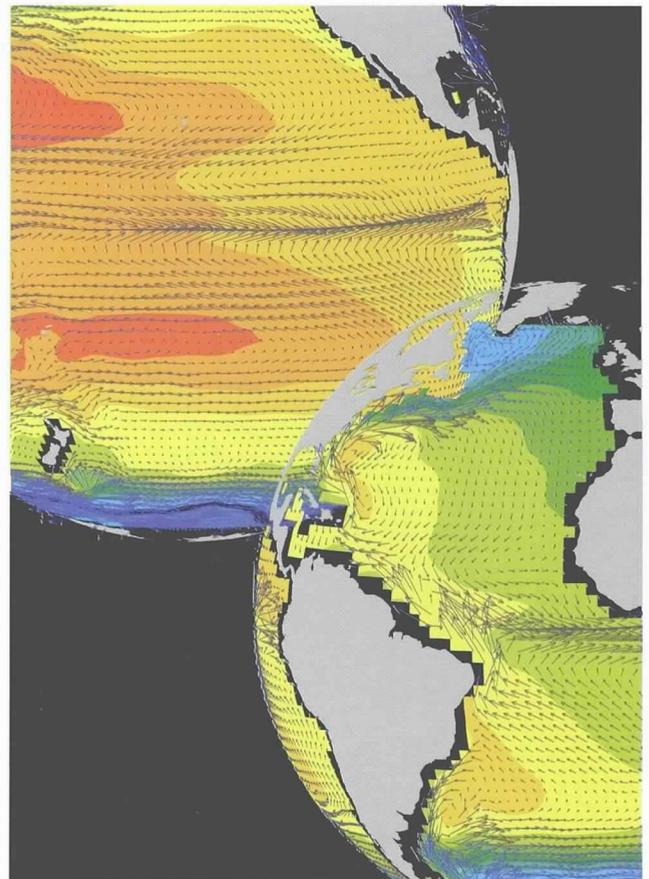

Figure 1: Mean surface velocity and sea surface height field for the Atlantic and Pacific Oceans estimates from 6 years of absolute TOPEX/POSEIDON data, sea surface temperature fields, surface fluxes and the monthly mean Levitus hydrography fields. See Stammer et al. (1997) for details.

A representative preliminary ECCO result is summarized in Figure 1, showing the estimated mean surface velocity and sea surface height field for the Atlantic and Pacific Oceans. Eventually, the estimated time-varying model state and consistent surface flux fields will be available for the entire estimation period and will be the basis for a wide variety of climate and societal applications. See Stammer et al. (1997) and the web page http://www.ecco.ucsd.edu for details on this work.

To properly simulate dynamical processes in the ocean, a high degree of realism is required for the simulated ocean flow field. This is generally accompanied by the need for substantial spatial resolution of ocean circulation models, which must be much higher than we have been able to use for ocean state estimation approaches. As an example, results shown in Figure 1 are fairly smooth in space due to the limited spatial resolution of 2° but actually carry all temporal variability present on the simulated scale. However, with increasing computational resources, the horizontal resolution of the model will increase as well, which eventually will lead to much more realistic simulations. And it is anticipated that, in two to three years, the project will be able to address the U.S. CLIVAR and GODAE related objective of depicting the time-evolving ocean state with spatial resolution up to 1/4° globally and with substantially higher resolution in nested regional approaches which are required for quantitative studies of the ocean circulation.
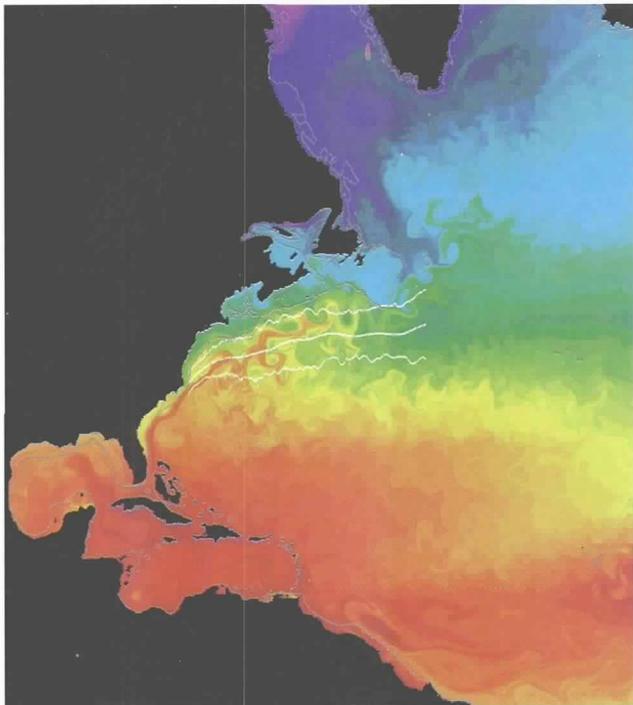

Figure 2: Sea Surface Temperature (SST) field from the fine-mesh simulation at day 273, year 14, performed with the Miami Isopycnic Coordinate Ocean Model (MICOM). Red denotes warm temperatures, blue cold temperatures. The observed mean position of the Gulf Stream is represented by the center line. The envelope is defined by the outer lines.

Forecasting the mesoscale eddy field is a central, complementary focus of the second NOPP assimilation activity. The global model configuration is the one adopted by the LANL in a comparison between HYCOM and Parallel Ocean Program (POP). The hybrid coordinate, in the context of the proposed work, is one that is isopycnal in the open, stratified ocean, but smoothly reverts to a terrain-following coordinate in shallow coastal regions, and to z-level coordinates in the mixed layer and/or unstratified seas. The capability of assigning additional coordinate surfaces to the oceanic mixed layer allows for sophisticated closure schemes, such as K-Profile Parameterization (KPP) (Large et al., 1994, 1997). A hierarchy of assimilation techniques is now being evaluated, from a simple optimal interpolation scheme, to a Kalman-filter-like algorithm. Since POP will become the oceanic component in the next generation of the National Center for Atmospheric Research (NCAR) Climate System Model (CSM), this intercomparison effort ensures that HYCOM's set-up and forcing parameterizations conform to the latest consensus on climate modeling. Coarse resolution model-based re-analysis of archived observational data will provide a comprehensive picture of the dynamics and thermodynamics of the global ocean during recent decades, and eddy-resolving basin-scale configurations will be performed using lateral boundary conditions provided by the global simulations, initially focusing on the Atlantic Ocean. By 2003, the project should be able to address the U.S.-GODAE principal objective of depicting/forecasting the three-dimensional ocean state at fine resolution in near real time with primary focus on the evolving eddy field.

The project will first assess the nowcast/forecast capabilities of truly eddy-resolving numerical simulations of the North Atlantic basin on synoptic time scales on the order of weeks to months and on spatial scales typically on the order of 10-100 km (mesoscale). Specifically, the primary research objective are real-time forecasting of both Lagrangian trajectories and 3-D Eulerian fields associated with such physical parameters as velocity, temperature, salinity, and density. An example of a fine-mesh numerical simulation is shown in Figure 2. The computational domain for the simulation consists of the North and Equatorial Atlantic Ocean basin from 28°S to 65°N, including the Caribbean Sea and the Gulf of Mexico, but excluding the Mediterranean Sea, with a horizontal resolution of 1/12° (mesh size on the order of 6 km) and 16 isopycnic layers in the vertical. The surface boundary conditions are based on monthly climatological data sets from Comprehensive Ocean-Atmosphere Data Set (COADS). Open ocean boundaries are treated as closed, but are outfitted with buffer zones. This simulation allows direct and detailed comparisons with observations such as satellite data (sea surface height, sea surface temperature) moorings measurement, inverted echo sounders, and free-floating drogues. For example, a total of 25,000

model floats were released at five levels (surface, 400 meters, 1000 meters, 1500 meters, and 3000 meters) (Figure 3). Comparison of the surface drifter statistics with their observed counterparts was recently performed by Garraffo et al. (2000) which show that the model correctly represents the eddy kinetic energy in the main current systems (with the exception of the Gulf Stream extension), but underestimates it in the ocean interior. The model's Lagrangian time scales are found to be longer in the interior than the observed *in situ* time scale by a factor of two, probably because of the lack of high-frequency winds in the model forcing.
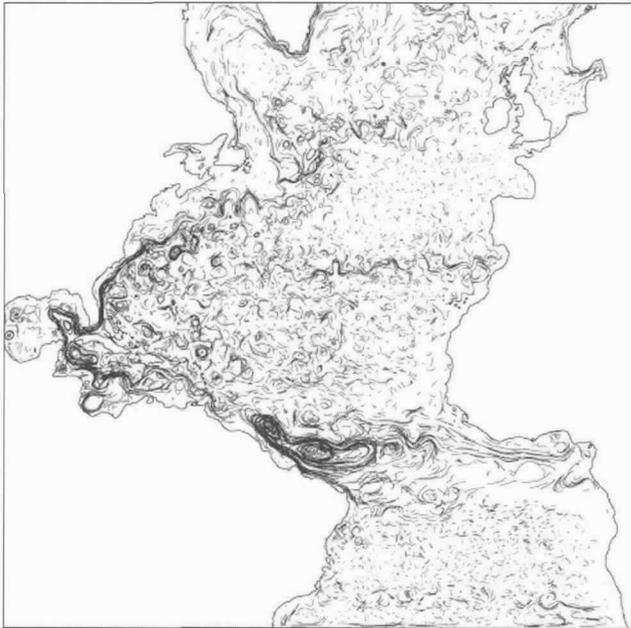

Figure 3: Two-week composite of surface float trajectories.

HYCOM is now being configured over a larger domain (northern boundary at 70°N, inclusion of the Mediterranean Sea) with surface forcing based on the 6-hourly 1979-1999 European Center for Medium Weather Forecasting (ECMWF) dataset. The data assimilation capabilities that are presently being evaluated are: (1) an Optimal Interpolation scheme (mostly for reference) combined with a Cooper-Haines vertical projection of the surface information; (2) the Adaptive Filter (Hoang et al., 1997), which estimates unknown parameters by minimizing the forecast error–this technique requires the model's adjoint (developed and parallelized in collaboration with SHOM)–and (3) the Markhov Random Field Information filter (Chin et al., 1999), a derivative of the Kalman filter that exploits the sparseness of the information matrix to dynamically update the forecast error covariance matrix.

Outlook

Given sufficient access to ever growing computer power, we anticipate that we will soon be able to con-

strain models on a routine basis globally and with resolution sufficient to address urgent dynamical and climate related questions in the ocean. At the same time, coupled climate models are being developed, and we can expect that in due time we will be able to run climate models over time intervals and with spatial resolution sufficient to simulate the climate system in process-oriented settings. However, considerable experience is needed in assessing the realism of ocean and climate models and in assessing the performance and design of climate observing systems. Both areas will benefit greatly from global data sets and from fully-developed global ocean state estimation efforts being developed now with NOPP support. Among the outcomes anticipated from these efforts, will be:

- an integrated analyses/reanalyses effort for research programs,
- a foundation for hypothesis testing and model improvement,
- an assessment of the observing system and of the utility of new ocean data sets,
- improved predictability of coastal, shelf, and regional models by providing suitable open-ocean boundary conditions, and
- improved open ocean nowcasts and forecasts (with application to search and rescue, iceberg paths, oil spills, shipping routes, fisheries, etc.).

The biggest challenge will be the evaluation of the results by cross-validation. For that purpose it will be important to establish performance measures, similar to those used in weather prediction. But the difficulty will be that performance measures very much depend on the application, and that the long intrinsic timescales of climate variability will make it difficult to test the model's predictive skill in an immediate manner.

Acknowledgements

REFERENCES

Bleck, R. and S. Benjamin, 1993: Regional weather prediction with a model compbining terrain-following and isentropic coordinates. Part I: Model description. *Mon. Wea. Rev.*, 121, 1770-1785.

Bleck, R. and D. Boudra, 1981: Initial testing of a numerical ocean circulation model using a hybrid (quasi-isopycnic) vertical coordinate. *J. Phys. Oceanogr.*, 11, 755-770.

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.

Fu, L.-L. and R.D. Smith, 1996: Global ocean circulation from satellite altimetry and high-resolution computer simulation. *Bulletin of the American Meteorological Society*, 77, 2625-2636.

Fukumori, I. R. Raghunath, L.-L. Fu and Y. Chao, 1999: Assimilation of TOPEX/POSEIDON altimeter data into a global ocean circulation model: How good are the results? *J. Geophys. Res.*, 104, 25647-25665.

Garraffo, Z.D., A.J. Mariano, A. Griffa, C. Veneziani and E.P. Chassignet, 2000: Lagrangian data in a high-resolution numberical simulation of the North Atlantic. I: Comparison with *in-situ* drifter data. *J. Mar. Sys.*, submitted.

Gent, P.R. and J.C. McWilliams, 1990: Isopycnal mixing in ocean models. *J. Phys. Oceanogr.*, 20, 150-155.

Giering, R. and T. Kaminsky, 1997: Recipes for adjoint code construction. *ACM Trans. Math. Software*, 24, 437-474.

Hellerman, S. and M. Rosenstein, 1983: Normal monthly wind stress over the world ocean with error estimates. *J. Phys. Oceanogr.*, 13, 1093-1104.

Hoang, S., Baraille, R. Talagrandd, X. Carton and P. DeMey, 1997: Adaptive filtering: Application to satellite data assimilation in oceanography. *Dyn. Oceans Atmos.*, 27, 257-281.

Hurlburt, H.E and P. Hogan, 2000: Impact of 1/8 to 1/64(resolution on Gulf Stream model-data comparisons in basing-scale subtropical Atlantic Ocean models. *Dyn. Atmos. Ocean.*, in press.

Large, W.G., J.C. McWilliams and S.C. Doney, 1994: Oceanic vertical mixing: a rview and a model with nonlocal boundary layer parameterization. *Rev. Geophys.*, 32, 363-403.

Large, W.G., G. Danabasoglu, S.C. Doney and J.C. McWilliams, 1997: Sensitivity to surface forcing and boundary layer mixing in a global ocean model: annual mean climatology. *J. Phys. Oceanogr.*, 27, 2418-2447.

Lemoine, F. and 17 others, 1997: The development of the NASA, GSFC and NIMA Joint Geopotential Model. In *Gravity, Geoid and Marine Geodesy*, International Association of Geodesy Symposia, 117, ed. Segawa et al, Springer-Verlag, Berlin Heidelberg.

Levitus, S., R. Burgett and T. Boyer, 1994: *World Ocean Atlas 1994*, vol. 3, *Salinity*, and vol. 4, *Temperature*, NOAA Atlas NESDIS 3 & 4, U.S. Dep. of Comm., Washington, D.C..

Malanotte-Rizzoli, P, 1996: *Modern Approaches to Data Assimilation in Ocean Modeling*. Elsevier, Amsterdam, The Netherlands, 445pp.

Marotzke, J. R. Giering, Q.K. Zhang, D. Stammer, C.N. Hill and T. Lee, 1999: Construction of the adjoint MIT ocean general circulation model and appiication to Atlantic heat transport sensitivity. *J. Geophys. Res.*, 104, 29529-29548.

Marshall, J., A. Adcroft, C. Hill, L. Perelman and C. Heisley, 1997a: A finite-volume incompressible navier-stokes model for studies of the ocean on parallel computers. *J. Geophys. Res.*, 5753-5766.

Marshall, J., C. Hill, L. Perelman and A. Adcroft, 1997b: Hydrostatic, quasi-hydrostatic and non-hydrostatic ocean modeling. *J. Geophys. Res.*, 5733-5752.

Paiva, A.M., J.T. Hargrove, E.P. Chassignet and R. Bleck, 1999: Turbulent behavior of a fine mesh (1/12() numerical simulation of the North Atlantic. *J. Mar. Sys.*, 21, 307-320.

Ponte, R.M. and D. Stammer, 1999: The role of regional mass and current changes in the ocean on seasonal polar motions fluctuations. *J. Geophys. Res.*, 104, 23393-23409.

Reynolds, R.W. and T.M. Smith, 1994: Improved global sea surface temperature analyses using optimum interpolation. *J. of Climate*, 7, 929-948.

Roemmich, D. and B. Owens, 2000: The ARGO Progect: Global ocean observations for understanding and prediction climate variability. *Oceanography*, this issue.

Smith, R.D., M.E. Maltrud, F.O. Bryan and M.W. Hecht, 2000: Numerical simulations of the North Atlantic Ocean at 1/10°. *J. Phys. Oceanogr.*, in press.

Stammer, D., and C. Wunsch, 1994: Preliminary assessment of the accuracy and precision of TOPEX/POSEIDON altimeter data with respect to the large-scale ocean circulation. *J. Geophy. Res.*, 99, 24000-25500.

Stammer, D., C. Wunsch, R. Giering, Q. Zhang, J. Marotzke, J. Marshall, and C.N. Hill, 1997: The global ocean circulation estimated from TOPEX/POSEIDON altimetry and the MIT general circulation model. MIT Center for Global Change Science, Report 49.

Wunsch, C. and D. Stammer, 1998: Satellite Altimetry, the Marine Geoid and the Oceanic General Circulation. *Annual Reviews of Earth and Planetary Sciences*, 26, 219-253. ▣