

EVALUATION CAPABILITY FOR THE NAVY OCEAN MODELING AND PREDICTION PROGRAM

By John A. Leese, Robert C. Willems,
and Lanny A. Yeske

THE NAVY Ocean Modeling and Prediction (NOMP) Program has emphasized the need to evaluate ocean models and prediction systems for operational and tactical applications. The Institute for Naval Oceanography (INO) was assigned a leading role in the evaluation and validation of ocean models and in providing recommendations to the Navy for adoption of models. The INO placed a high priority on this role in its efforts to design and implement an Experimental Center for Mesoscale Ocean Prediction (ECMOP) as described by Leese *et al.* (1992, this issue).

The evaluation of ocean modeling and prediction received considerable emphasis at the ECMOP Workshop (INO, 1988). The workshop recommended that guidelines be established for conducting an evaluation or determining success. It further recommended that the setting of evaluation criteria and standards be considered among the high-priority items in the early efforts of ECMOP.

Model Evaluation

The INO Summer Colloquium in 1989 initiated the development of evaluation criteria (Willems, 1989). A working group was asked to define the role of data within the context of ECMOP as a specialized facility for model evaluation. In this context, ECMOP's purpose is to provide standard data sets for model evaluation. These should fit within the framework of limited space and time resources and of limited data availability for the evaluation of mesoscale eddy-resolving models.

There are a finite number of phenomena that have observational bases from which quantitative

criteria can be developed. In addition, there are few field observation programs that can be used for model evaluation as part of an oceanographic-integrity test. Guidelines need to be established to identify the existence of phenomena within a data set. In constructing criteria for phenomenological model evaluation, one can designate two main areas requiring examination: 1) physical integrity and 2) oceanographic integrity. Physical integrity considers specific properties of the model itself. How well does the model conserve mass? What are its numerical and stability characteristics? Oceanographic integrity considers the evaluation of a model relative to ocean observation. This area applies to a model's performance run solely in a simulation mode, all the way to a model's performance as part of a nowcast/forecast system.

Data sparseness, Gaussian versus non-Gaussian statistics, and statistical assessment of events were viewed as generic problems. These must be considered in the evaluation of ocean models for mesoscale ocean nowcasting/forecasting. Incorporation of statistical procedures was recommended in the following: predictability studies; statistical comparisons (model/observations); distortion due to biased sampling; error effects on statistics; decomposition of root-mean-square-error; regional (sub-domain) evaluation; inter- and intramodel comparison (simulation properties, i.e., climate drift); and second-order statistics.

Two classes of model evaluation were considered for the standards of model performance. First was scientific evaluation with the requirement that an accuracy measure was performed using benchmark phenomena and case studies. A model then can be quantitatively assessed using agreed upon accuracy measures. The model also can be qualitatively assessed based on the inherent knowledge of the circulation in the ocean region from which the benchmark phenomena and case studies are taken. Second was operational evaluation, which

. . . a leading role in
the evaluation and
validation of ocean
models . . .

J.A. Leese, R.C. Willems, and L.A. Yeske, Institute for Naval Oceanography, Stennis Space Center, MS 39529-5005, USA. The Institute for Naval Oceanography is operated by the University Corporation for Atmospheric Research (UCAR) for the Naval Research Laboratory (NRL).

. . . a series of experiments that began with a very simple case study . . .

is a much more demanding evaluation. Specification of climatology, input data, computer resources, manpower/skill level, and interface with other modules must be made. As a model moves through an operational evaluation, there must be routine quantitative measures of success. These are in addition to qualitative assessment of how well it performs with observational data available within the time constraints of a routine operational schedule.

Ocean Model and Prediction Evaluation Experiments

The INO recognized the need to start evaluation experiments for ocean modeling and prediction as soon as possible, to provide information and recommendations to the NOMP program about existing ocean models and to provide a learning experience for evaluating ocean prediction systems. INO is carrying out a series of experiments that began with a very simple case study in June 1990 and is evolving to the Data Assimilation and Model Evaluation Experiments during 1992.

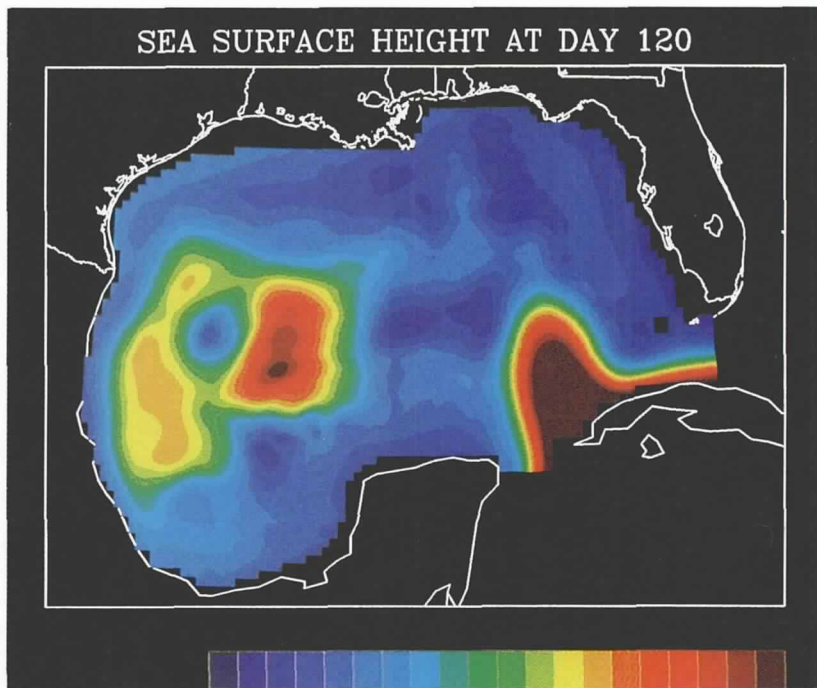


Fig. 1: TRial Ocean Prediction Experiment forecast of sea-surface height in the Gulf of Mexico using synthetic temperature profiles derived from the Geodetic Earth Orbiting Satellite Exact Repeat Mission, conventional observations for initialization of the Princeton model and assimilation by the Derber technique. The scale ranges from -25 cm (left side) to 45 cm (right side). (Courtesy of Dr. Dong-Shan Ko.)

Results from Case Study

A nowcast/forecast system using the Princeton ocean general circulation model (Blumberg and Mellor, 1987) was adapted to the Gulf of Mexico and the Derber data-assimilation technique (Derber and Rosati, 1989) by L. Kantha and R. Passi. The infrastructure of data management, visualization, and evaluation were provided by the ECMOP facility (Leese *et al.*, 1992, this issue) to conduct a TRial Ocean Prediction Experiment (TROPE) in the Gulf of Mexico. Real data from satellite and conventional observations were used as input data (Ko, unpublished data).

The Gulf of Mexico forecast output fields (Fig. 1) were evaluated (Waters *et al.*, 1990). The evaluations used mainly the statistical techniques that were formulated from the Colloquium on Model Evaluation Criteria and that were implemented as a verification and evaluation module in ECMOP (Anand *et al.*, 1991).

The primary emphasis of TROPE has been a software engineering-type evaluation. This successfully proved the validity of the ECMOP concept to provide a specialized off-line computation center to facilitate the development, demonstration, and evaluation of a mesoscale ocean prediction capability. A number of scientists made use of the TROPE nowcast/forecast system and the output fields to conduct further analysis and evaluations. A specific example is the work of Lewis *et al.*, (1991) to evaluate the movement of eddies by the Primitive Equation Data Assimilating Model (PEDAM) in the Gulf of Mexico.

An Ocean Nowcast/Forecast Experiment

The results from the ECMOP Workshop (INO, 1988), along with preparatory work for the Colloquium on Model Evaluation Criteria (Willems, 1989), brought forth an interest among several modeling groups to conduct an assessment experiment in ocean modeling and prediction. INO hosted a separate meeting after the Colloquium to consider a proposal for a nowcast/forecast assessment experiment during 1992. There was a consensus at the meeting that it was particularly timely to conduct such an assessment experiment with the following general characteristics:

Objective: To estimate the state-of-the-art for data assimilation and nowcast/forecast production, using the primitive-equation ocean general circulation model.

Data period: 1987–1988.

Geographical area: Within the region bounded by 25° – 45° N latitude and 40° – 80° W longitude.

Scenarios: Case studies (up to 10) where data quality and quantity were sufficient.

Participants: Modeling groups working in the North Atlantic Region.

The term assessment means, in general, the evaluation of objectively analyzed and short-term forecast fields to determine how well they describe the real ocean at the valid times and, to the extent possible, to determine the effects of input data, the data-assimilation techniques, and the performance of the ocean circulation model.

The INO agreed to coordinate experiment planning and design, provide infrastructure facilities through the ECMOP facility, assist modeling groups in conducting experimental runs, and complete arrangements for meetings and workshops. The INO has worked, since the August 1989 meeting, with the NOMP Program Manager and the ocean modeling groups working in the Northwest Atlantic region to coordinate the planning and design of the nowcast/forecast assessment experiment. A number of modeling groups provided assistance in the planning and design for the nowcast/forecast experiment and are now participating in the experiment:

- a. Group led by A. Robinson, using the Harvard open-ocean model as a basis for an ocean nowcast/forecast system known as FLEXCAST;
- b. Group led by P. Rizzoli (Massachusetts Institute of Technology) and D. Haidvogel (Rutgers University), using a nowcast/forecast system based on the Semi-Spectral Primitive Equation Model (SPEM);
- c. Group led by G. Mellor, using the Princeton model as a basis for a nowcast/forecast system known as PEDAM for the Gulf Stream;
- d. Group led by D. Fox, using a nowcast/forecast system based on the NRL primitive equation (PE) model and referred to as Data Assimilation Research and Transition (DART).

The geographical boundaries used by each of the modeling groups is shown in Figure 2. Model attributes are shown in Table 1. The design of the numerical models was determined, implemented, and verified independently and prior to this nowcast/forecast experiment (see Blumberg and Mellor, 1987; Robinson *et al.*, 1988; Haidvogel *et al.*, 1991). Clearly, similarities and differences exist in

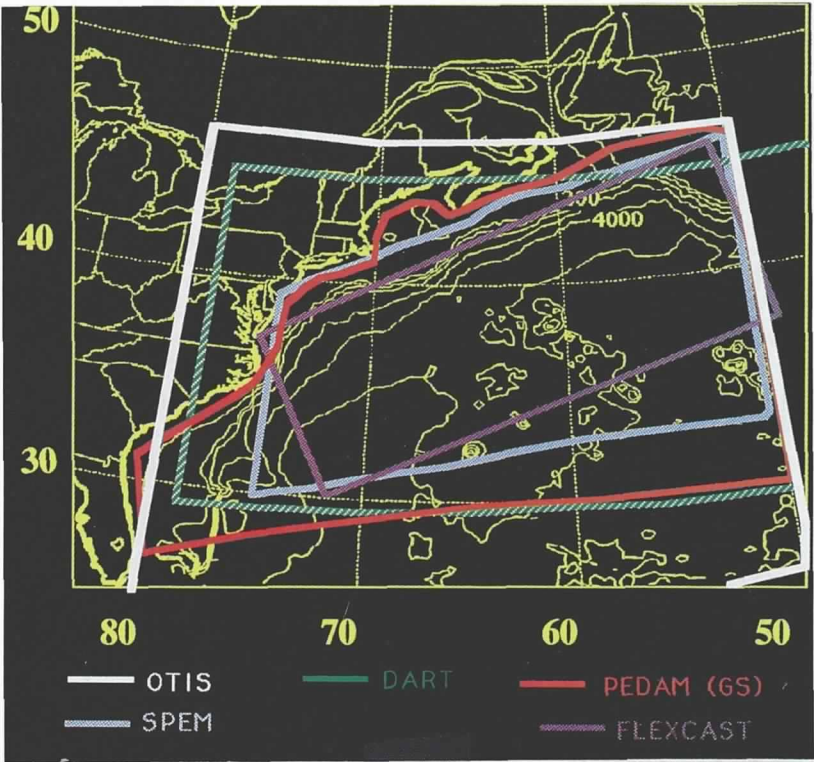


Fig. 2: Optimal Thermal Interpolation System (OTIS 3.0) and ocean model domains for the Data Assimilation and Model Evaluation Experiments in 1992.

methodologies and design. Numerical solution methods; surface, lateral, and bottom boundary specifications; and mixed-layer representation (whether resolved or not) are dealt with differently in the various models. Table 1 articulates the more substantial differences to be considered in developing model evaluations. The infrastructure facilities of data management, visualization, and evaluation are being prepared by the ECMOP project in INO led by J. Corbin. A data group, led by S. Glenn (Rutgers University), prepared the research-quality data sets needed to conduct the experiment.

Table 1
Ocean model attributes

Model	Numerical Solution Method	Surface Boundary Condition	Vertical Coordinate Specification	Mixed Layer	Lateral Boundary Condition	Bottom Boundary Condition
SPEM	Implicit	Rigid lid	Sigma level Chebyshev polynomials	No	Coastline/open boundary	Smoothed topography
PEDAM	Implicit	Free surface	Sigma level	Yes Mellor-Yamada closure scheme	Coastline/open boundary	Smoothed topography
FLEXCAST	Implicit	Rigid lid	Sigma level	No	Open boundary	Smoothed topography
DART	Semi-implicit	Free surface	Layer/reduced gravity	No	Coastline/open boundary	Smoothed topography

... assess the impact of several data-assimilation techniques on the various models using the same case-study data.

The Navy has recently completed an evaluation of two ocean circulation models as components of the Naval Operational Gulf Stream Forecast System (NOGUS). The case studies used for the evaluation were during data-rich periods. Some of the cases contained ambiguities in the analyses of Gulf Stream and ring locations (due to clouds, etc.), as determined from the Advanced Very High Resolution Radiometer (AVHRR). A re-analysis was performed to prepare research-quality data sets (Glenn *et al.*, 1991). From this re-analysis, six case studies were identified for the assessment experiments.

Pre-Experiment Tests in 1991

A pre-experiment to quantify the forecast capability of an ocean model was conducted in 1991 and involved initializing an ocean model domain that was configured for the Gulf Stream region. The temperature, salinity, and sea-surface height fields were prepared using the Optimal Thermal Interpolation System (OTIS), which is currently being used by the Navy (Clancy *et al.*, 1990). Figure 3 is an example of the temperature field at

500 m. At 1- and 2-week intervals, the model forecast position of the north wall of the Gulf Stream was compared with the observed position. The best comparison was with the position determined from the National Oceanic and Atmospheric Administration (NOAA) satellite image available at the end of each case study.

Data Assimilation and Model Evaluation Experiments in 1992

The Data Assimilation and Model Evaluation Experiments (DAMEE) are being carried out in 1992. This is a follow-on to the 1991 pre-experiment tests and will assess the impact of several data-assimilation techniques on the various models using the same case-study data. Additionally, two month-long case studies are identified, wherein one has meandering characteristics and the other has a ring formation/pinchoff or ring coalescence. The data available for assimilation and verification are the NOAA satellite infrared images, Airborne Expendable Bathythermograph (AXBT), Expendable Bathythermograph (XBT), Multi-Channel Sea-Surface Temperature (MCSST), and Geodetic Earth Orbiting Satellite (GEOSAT) Exact Repeat Mission sea-surface height. OTIS 3.0 fields from the six case studies are available as a data set as well. These case studies occur when the GEOSAT satellite provided useful sea-surface height data and also during the Synoptic Ocean Prediction (SYNOP) observational period. The assessment will be conducted in four phases, each based on different data sets from the case studies and building upon the results from previous phases.

Quantitative measures of ocean-model performance with data assimilation are relative to persistence and using forecast positions of the Gulf Stream, compared with a best-available analysis position produced by AVHRR on the forecast day at the end of each week.

Concluding Remarks

The progress in developing and implementing an ocean modeling and prediction evaluation capability for NOMP has evolved over the last two years. This capability is now being combined with the ECMOP Modular System in the INO to provide a special purpose facility. The Data Assimilation and Model Evaluation Experiments in 1992 will serve as a good test of the overall concepts that provided the basis for the design and implementation of this facility.

Acknowledgements

The INO is preparing the ocean modeling and prediction evaluation capability for NOMP with the very capable assistance of experts from the ocean modeling community. We particularly acknowledge the assistance received from all the participants at the 1989 INO Summer Collo-

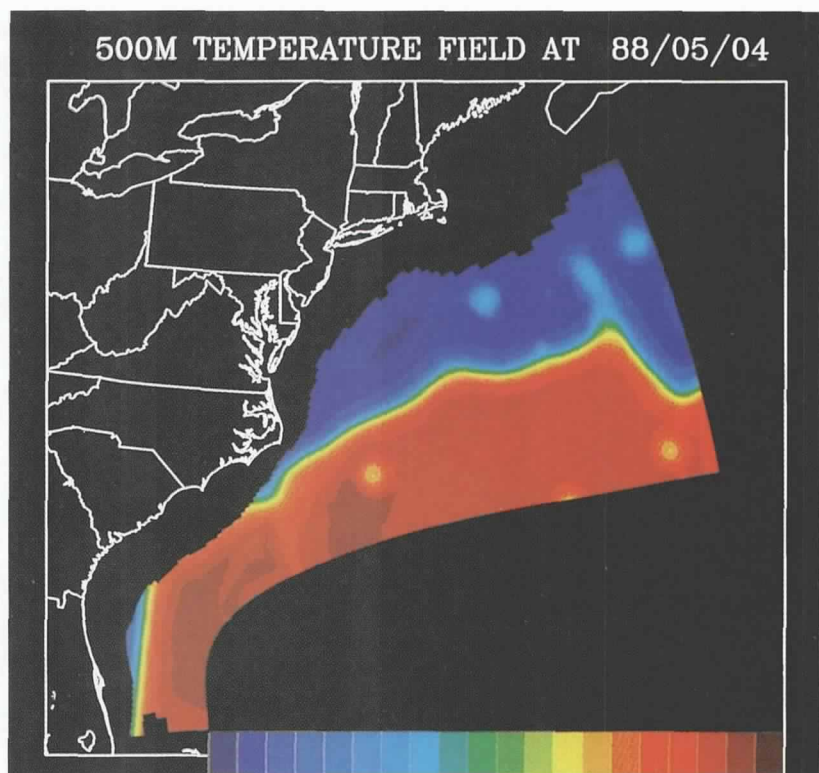


Fig. 3: Optimal Thermal Interpolation System (OTIS) 3.0 temperature field at 500 meters for 4 May 1988 as applied to the Princeton model domain in the Gulf Stream (for the 1991 preexperiment). Advanced Very High Resolution Radiometer analysis of the stream position and warm/cold eddy locations, climatology, and conventional observations are input to the calculation of OTIS fields. The scale ranges from 5°C (left side) to 18°C (right side). (Courtesy of Dr. Dong-Shan Ko.)

quium on Model Evaluation Criteria. We also acknowledge the contributions from the modeling groups participating in the Ocean Modeling and Prediction Evaluation Experiments being coordinated by the INO.

References

- Anand, H., M.S. Foster, R. Krishnamagaru and R.M. Passi, 1991: VERMOD Capabilities: VERMOD 1.0 and Future. *INO Technical Memo TM-4*.
- Blumberg, A.F. and G.L. Mellor, 1987: A description of a three-dimensional coastal ocean circulation model. In: *Three-Dimensional Coastal Ocean Models*, vol. 4. N. Heape, ed., American Geophysical Union, Washington, DC, 208 pp.
- Clancy, R.M., P.A. Phoebus and K.D. Pollok, 1990: An operational global-scale ocean thermal analysis system. *J. Atmosph. Oceanic Technol.*, 7, 233–254.
- Derber, J. and A. Rosati, 1989: A global oceanic data assimilation system. *J. Phys. Oceanogr.*, 19, 1333–1347.
- Glenn, S.M., D. Crout and L. Perkins, 1991: Comparison of Gulf Stream Forecast Model Initialization and Verification Analysis. *INO Technical Memo TM-6*, 37 pp.
- Haidvogel, D.G., J.L. Wilkin and R. Young, 1991: A semi-spectral ocean circulation model using vertical sigma and orthogonal curvilinear horizontal coordinates. *J. Comp. Phys.*, 94, 151–185.
- INO, 1988: Experimental Center for Mesoscale Ocean Prediction (ECMOP), Report on a Workshop. *INO Report 89-2*. University of Southern Mississippi, Long Beach, MS.
- Leese, J.A., M.S. Foster and J.H. Corbin, 1992: An experimental center for mesoscale ocean prediction. *Oceanography*, 5, 60–63.
- Lewis, J.K., L. Kantha, A. Gallegos and R. Passi, 1991: Verification and calibration of an eddy-resolving model of the Gulf of Mexico. Science Applications International, Technical Report 91/1134, 14 pp.
- Robinson, A.R., M. Spall and N. Pinardi, 1988: Gulf Stream simulations and the dynamics of ring and meander processes. *J. Phys. Oceanogr.* 18, 1811–1853.
- Waters, M.P., R.G. Kelley, B.S. Carlson and M.J. Roth, 1990: Trial ocean prediction experiment of the ECMOP PEDAM (GOM); Unpublished Final Report of the U.S. Naval Academy/INO Summer Research Program.
- Willems, R.C., 1989: Summer colloquium on mesoscale ocean science and prediction: model evaluation criteria. *INO Technical Note (PROC-1)*, National Center for Atmospheric Research, Boulder, CO. □