# MEASURES OF THE FIDELITY OF Eddying Ocean Models

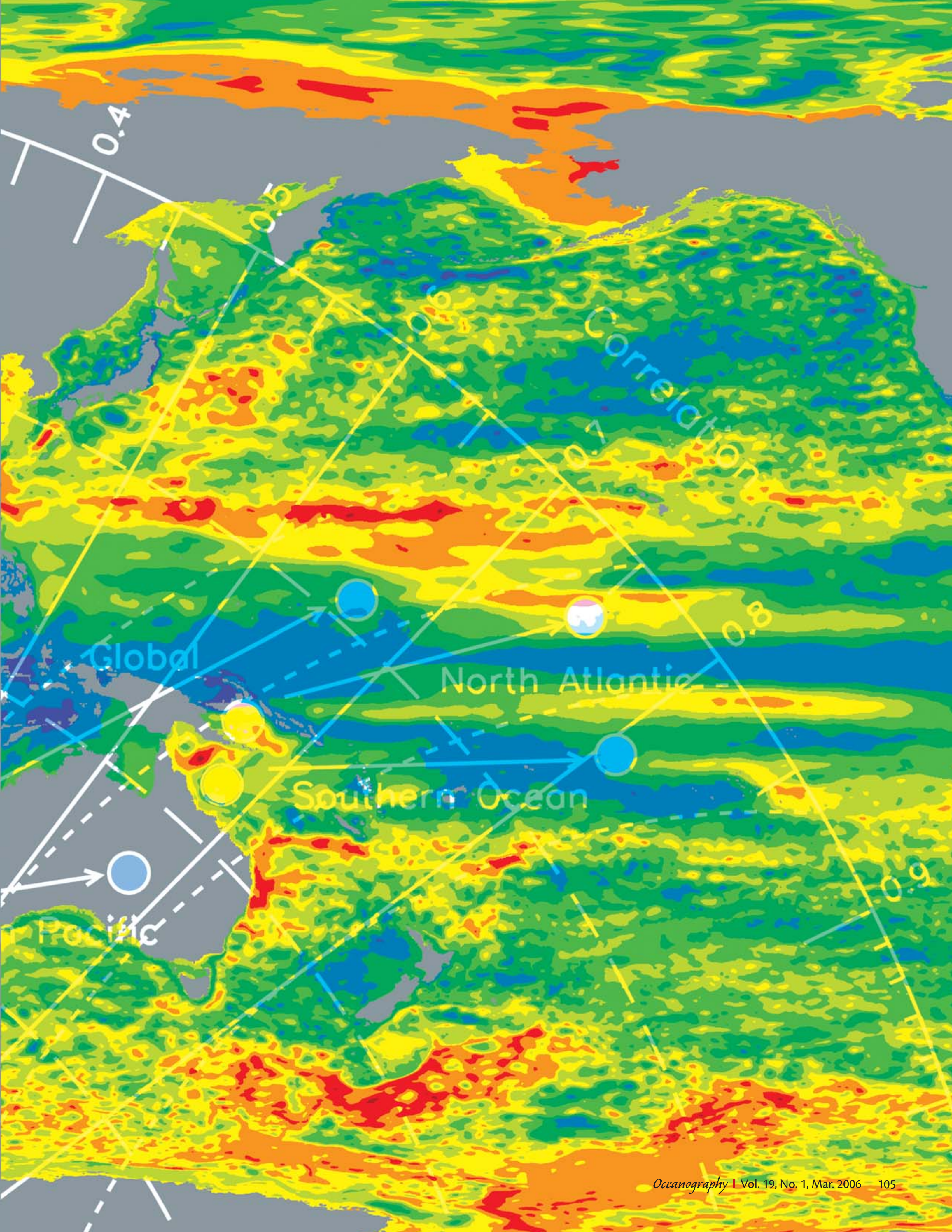
BY JULIE L. MCCLEAN, MATHEW E. MALTRUD, AND FRANK O. BRYAN

Computational simulation is now an essential methodology of science, along with theory and observation. The ability of scientists to understand and predict planetary climate variability largely depends on the veracity of the climate simulations produced by numerical models of the interacting components of the Earth system. Oceanic and atmospheric models are numerical approximations to continuous forms of the equations governing fluid flow and are "closed" by sub-grid-scale parameterizations that represent physical processes on temporal and spatial scales that are not resolved by the chosen model grid. In the past two decades, the rate at which the world's fastest computers perform floating point operations (FLOPS) has increased by a factor of 10,000. This increase in computing capability has been exploited in several ways. Longer integrations for applications such as paleoclimate (Dijkstra and Ghil, 2005) and the inclusion into models of additional processes such as biogeochemical cycles and ecosystem dynamics (Moore et al., 2004), are two such examples. Another example, and the focus of the present study, is to increase the spatial resolution of models such that a greater fraction of the physical processes are explicitly resolved, and fewer are parameterized.

Currently, such state-of-the-art decadal-time-scale global ocean simulations are being conducted using models configured on grids with horizontal resolution of 5 to 10 km and 40 to 60 vertical levels or layers. Their duration, limited to several decades by the capability of present-generation computational platforms, is sufficient to allow the flow to mostly adjust dynamically to its initial state ("spin-up" the circulation), but not to bring the model into thermodynamic equilibrium. The objective of the present study is to provide an indication of how realistically ocean models of this class are able to represent the mean and variability of the real upper-ocean general circulation in anticipation of the time when it will be possible to perform centennial climate integrations with them.



This article has been published in *Oceanography*, Volume 19, Number 1, a quarterly journal of The Oceanography Society. Copyright 2006 by The Oceanography Society. All rights reserved. Permission is granted to copy this article for use in teaching and research. Republication, systemmatic reproduction, or collective redistribution of any portion of this article by photocopy machine, reposting, or other means is permitted only with the approval of The Oceanography Society. Send all correspondence to: info@tos.org or Th e Oceanography Society, PO Box 1931, Rockville, MD 20849-1931, USA.


0.4
0.5
0.6
Correlation
0.7
0.8
0.9
Global
North Atlantic
Southern Ocean
r Pacific

The importance of simulating the ocean as accurately as possible in climate studies results from the ocean's role in storing and transporting energy, freshwater, nutrients, and dissolved gases (e.g., carbon dioxide). The ocean acts as a heat capacitor in the coupled atmosphere-ocean system and hence acts as an integrator of climate variability, introducing long time scales and slowing the rate of response to climate-change forcing. Variability intrinsic to the ocean may interact with the atmosphere over a range of time scales to produce coupled modes of climate variability such as the El Niño-Southern Oscillation and the North Atlantic Oscillation. What is likely to be gained by the use of fine-resolution grids in such applications? Such grids allow for a more realistic representation of ocean-bottom bathymetry and coastal geometry, which affects ocean dynamics and communication between ocean basins and/or marginal seas. Narrow western boundary currents (such as the Gulf Stream)—whose dynamics influence the gyre-scale circulation, jets like those in the Antarctic Circumpolar Current, and open-ocean zonal jets (Treguier et al., 2003; Maximenko et al., 2005)—are more realistically reproduced, which is essential to the proper simulation of key climatic quantities such as mass, heat, and salt transports. In a coupled climate system, more accurately positioned fronts are important in air-sea interaction processes because large sea surface temperature errors can result from biases in flow paths.

Selected metrics were chosen to assess the realism of the upper-ocean circulation in large-scale fine-resolution ocean model simulations.

The majority of the energy in the ocean circulation is generally accounted for by flows with spatial scales of 50 to 500 km, the so-called oceanic mesoscale (Stammer, 1997). The resolution of these energetic mesoscale flows, such as eddies and fronts, on short spatial scales impacts the mixing processes represented by the model. Finer resolution allows explicit dissipation to be reduced to a level whereby realistic eddy growth rates and decay time scales are possible. At what horizontal and vertical resolution these models can be considered truly "eddy-resolving" is still to be understood. Smith et al. (2000) argue that because the first baroclinic Rossby radius is resolved up to about 50° N in a 0.1° North Atlantic ocean simulation using the Parallel Ocean Program (POP) and that typical length scales for mesoscale eddies are somewhat larger than the Rossby radius, then eddies would be reasonably well resolved in most of the model domain. This assertion is based on the two-grid point criteria of sampling theory; hence, the eddy spectrum is not fully resolved.

So what is the fidelity of large-scale, fine-resolution simulations? In this study we provide a brief answer to this question by evaluating the realism of the mean and variability of the upper-ocean circulation in 0.1°, 40-level global and basin-scale POP simulations. We will compare selected quantitative and semi-quantitative statistical analyses of available observations and consistent model representations of the upper-ocean circulation. These techniques are both Eulerian (repeated sampling in time at a particular location such as along a satellite track or by a moored instrument) and Lagrangian (measurements collected in time while following a water particle). This choice is motivated by the length of the simulations available (a few decades) and the fundamental nature of the mean flow and eddy variability in ocean transport processes. Comparisons of dynamical balances and budgets calculated using consistent model output and data are other examples of metrics that can be used to assess the realism of an ocean model. The selection of these metrics

**Julie L. McClean** *(jmcclean@ucsd.edu) is Associate Researcher of Oceanography, Scripps Institution of Oceanography, La Jolla, CA, USA, and Participating Guest Scientist, Lawrence Livermore National Laboratory, Livermore, CA, USA.* **Mathew E. Maltrud** *is Technical Staff Member, Fluid Dynamics Group, Los Alamos National Laboratory, Los Alamos, NM, USA.* **Frank O. Bryan** *is Scientist III, Climate and Global Dynamics Division, National Center for Atmospheric Research, Boulder, CO, USA.*

and the data upon which they are based must largely be designed to address the problem to which the model is being applied. As longer-length experiments become available, the ability of the models to simulate water-mass properties and the distribution of both natural and anthropogenic tracers will become important metrics for establishing the skill of the simulation of ocean heat uptake, sea-level rise, and perturbations to the carbon cycle under climate-change forcing.

A significant challenge lies in finding data with sufficiently high temporal and spatial resolution with which to conduct these comparative analyses. Two data sets are chosen for use in this study: sea surface height anomalies (SSHA) from the Archiving, Validation, and Interpretation of Satellite Oceanographic data (AVISO)-merged *TOPEX/POSEIDON* (*T/P*) and *ERS 1* and *2* altimeters (for more information see http://www.aviso.oceanobs.com) and velocities from surface drifting buoys at 15 m. Both data sets have spatial coverage on a near-global basis with sufficiently high temporal resolution for statistical comparisons. Subsurface float data, although sparser than that from the surface drifters, also exist in sufficient quantities in some regions for statistical comparisons; we present an example here. Finally, two time series of vertical profiles of upper-ocean temperature data collected at particular geographical locations, the so-called "Ocean Time Series" that are being monitored on a long-term regular basis, are also used in this study.

The ocean model used in this study, POP, is a three-dimensional, *z*-level, primitive equation ocean general circulation model. The code is a descendant of those developed by Bryan (1969), Cox (1970, 1984), and Semtner (1974). Smith et al. (1992) rewrote the code to run on massively parallel computers (platforms on which the model domain is decomposed into sub-regions, in this case geographical regions, which are spread over a group of processing units that perform the same calculations at about the same time). The global configuration is forced with synoptic National Center for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) atmospheric surface fluxes of momentum, heat, and freshwater for the period 1979–2003. The global model spin-up period is 1979–1993 (Maltrud and McClean, 2005); hence, analyses of the global model are conducted for the period 1994–2001. Further details of this simulation are found in Maltrud and McClean (2005). Two 0.1° North Atlantic simulations are also analyzed—that of Smith et al. (2000) and a later run which is forced with the same NCEP/NCAR atmospheric fluxes that are used to force the global model for the period 1979–1999 subsequent to a five-year spin-up.

The choice of metrics was driven by our desire to understand the fidelity of these simulations in the context of their potential use in future fine-resolution, coupled climate-system studies.

## PSEUDO-EULERIAN MEAN UPPER-OCEAN CURRENTS AND LAGRANGIAN STATISTICS

Lagrangian statistics have long been recognized as a most sensitive statistical means of testing the validity of eddying models. If the Lagrangian statistics from the model output are comparable to those from drifting buoys or floats, then one can have confidence in the choice of model parameters (Krauss and Böning, 1987). Both Garraffo et al. (2001) and McClean et al. (2002) compare pseudo-Eulerian and Lagrangian statistics from surface drifting buoys and fine-resolution (0.1° and greater) basin simulations of the North Atlantic. Treguier et al. (2005) compare velocities from surface drifters and floats at 700 m in the sub-polar North Atlantic with velocities from two eddy-resolving North Atlantic simulations.

The sub-polar North Atlantic is a key region for heat and freshwater exchange between the mid-latitudes and the polar seas. The North Atlantic Current transports heat poleward, while boundary currents around Greenland bring freshwater of Arctic origin into the Labrador Sea. The representation of these sub-polar upper-ocean currents is therefore important in climate models. Here we use a James test (Seber, 1984) to statistically

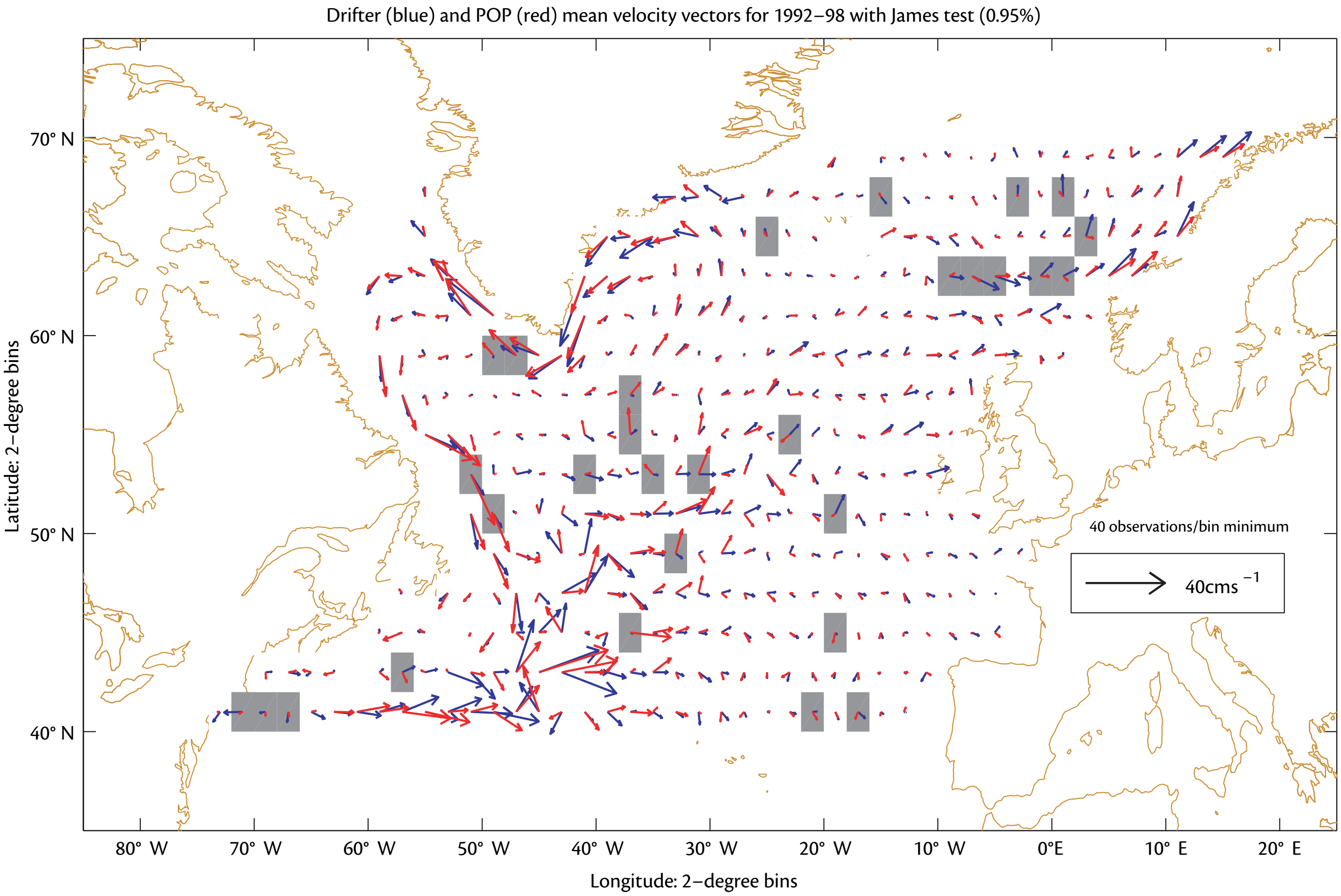



Figure 1. James test (0.95 percent confidence level) using 2° x 2° binned mean surface (15 m) drifting buoy (blue) velocity vectors (cm $s^{-1}$) and 0.1° North Atlantic POP (red) velocities (cm $s^{-1}$). The scale arrow applies to both sets of velocities. Gray shaded bins are those where the null hypothesis of equal means is rejected with the possibility of being wrong 5 percent of the time. Bins containing fewer than 40 observations are not included. Only model velocities that were collocated with observations are included in the calculation. Good agreement is seen in the boundary currents around Greenland, which transport freshwater of Arctic origin into the Labrador Sea and where the Gulf Stream bringing warm water from lower latitudes forms the Northwest Corner (40°–50°W, 40°–50°N).

compare mean vector fields calculated from surface drifting buoy data at 15 m and the 0.1° North Atlantic POP model forced with NCEP/NCAR fluxes for 1992–1998 in the sub-polar North Atlantic. Only model velocities collocated in space and time with the observations are included in the calculation. Both simulated and observed velocities are spatially averaged into 2° x 2° bins. The James test compares populations with different variances and provides information as to where vectors at a particular location are significantly different. Gray shaded bins in Figure 1 are those where the null hypothesis of equal means is rejected with a possibility of being wrong 5 percent of the time. In general, there is good agreement between the observed (blue arrows) and simulated velocities (red arrows), particularly in the boundary currents around Greenland and where the Gulf Stream forms the Northwest Corner (40°–50°W, 40°–50°N). Significant differences occur in the mid sub-polar gyre and in the Norwegian Basin. See Garraffo et al. (2001) or McClean et al. (2002) for further explanation of the technique as applied to fine-resolution models and drifter data.

Accumulating drifter data in Eulerian bins is very useful because that is the model's natural reference frame.

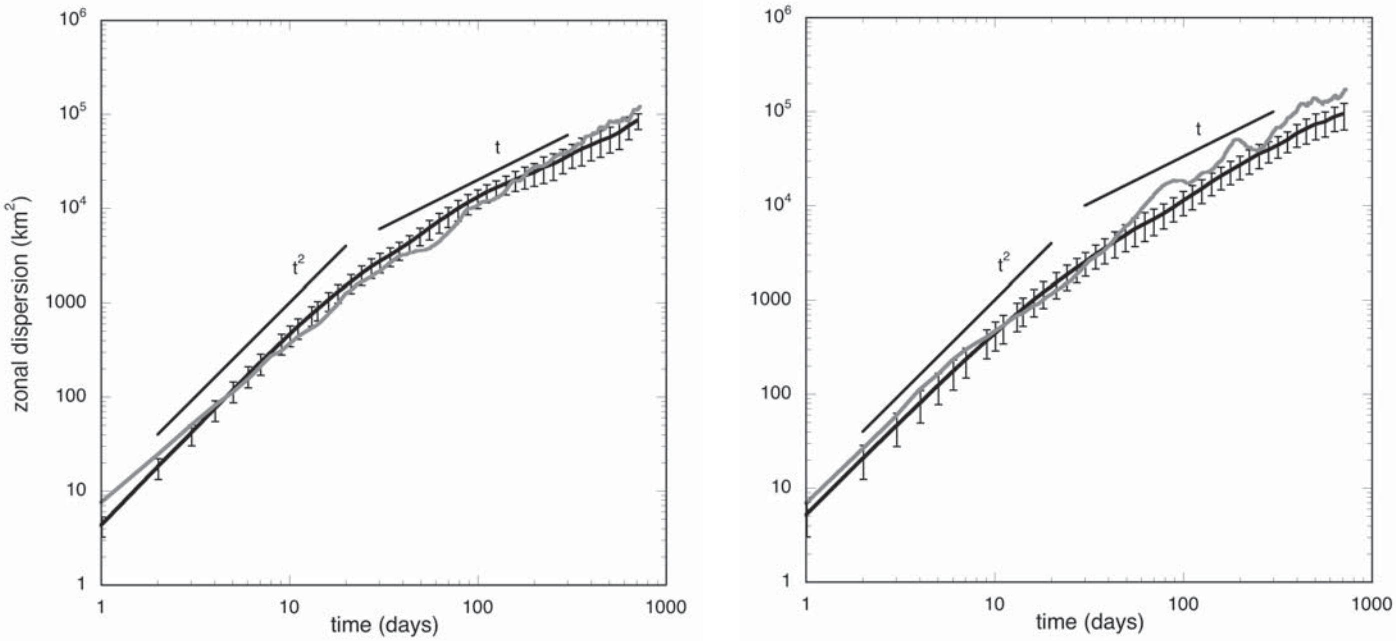



Figure 2. Zonal (east-west) and meridional (north-south) dispersion of model floats (black curves) and EUROFLOATS (grey curves) as a function of time, where the dispersion is the average squared distance that a deployment of floats moves from the group's center of mass. The error bars denote the standard deviation of the ensemble of model floats. For reference, the $t^2$ line denotes the theoretical behavior at early time, and the $t$ line denotes the later time random walk prediction.

However, in some instances it is advantageous to do the reverse, and represent the model output in the Lagrangian frame for comparison with observational data such as sub-surface floats. As mentioned above, good agreement between simulated and observed Lagrangian statistics is an indicator of the appropriateness of the choice of model parameters. Here, output from the Smith et al. (2000) 1/10° North Atlantic simulation is compared with data from the EUROFLOAT experiment (Speer et al., 1999), which consisted of 21 floats deployed in the northeast Atlantic at 1750-m depth. The model floats were released at the same locations as the real oceanic floats, but instead of just one deployment, we were able to release 20 independent groups of model particles at different times of year in order to estimate the variability of the statistics.

Figure 2 shows the zonal and meridional dispersion of the model floats and the EUROFLOAT data, where the dispersion is defined as the average squared distance that a deployment of particles moves from the group's center of mass as a function of time. The error bars denote the standard deviation of the 20-member model ensemble at each time, so it is clear that the data are almost always within the standard deviation of the model. In addition, both the model and data show two distinct power-law ranges with a similar transition time between them (around 10 to 20 days). At early times, the dispersion is proportional to $t^2$, which is the classical result for a turbulent fluid when movements of water parcels are still highly correlated with their initial displacements (Taylor, 1921). At later times, we find a regime close to a classical random walk, which predicts that the dispersion is proportional to $t$. Both the model and data deviate somewhat from the idealized prediction, indicating that the particle trajectories are being significantly influenced by the local bathymetry, such as the Mid-Atlantic Ridge (which is to the west of the deployment region), and Earth's rotation. Other standard Lagrangian statistics (such as integral length and time scales) also agree well with the data. Increased model resolution plays a major role in improving the fidelity of the simulation in these types of comparisons (not shown).

## EULERIAN MEASURES OF UPPER-OCEAN VARIABILITY

The ocean exhibits variability over a broad range of scales, and the degree to which ocean models reproduce these fluctuations is a gauge of the usefulness of such models in a variety of applications, including climate and synoptic forecasting. Altimeter-derived SSHA are the primary Eulerian data set used to gauge variability levels in models at these resolutions (Paiva et al., 1999; Smith et al., 2000; Brachet et al., 2004; Masumoto et al., 2004; Maltrud and McClean, 2005). Other comparison studies use tide gauges (Tokmakian and McClean, 2003; Barron et al., 2004), while Donohue et al. (2002) compare Acoustic Doppler Current Profiler (ADCP) velocity sections and an eddying model with high vertical resolution.

To address the realism of the variability, we compare representative meridional sections of root-mean-square (RMS) SSHA from the AVISO-blended altimetry product to the 0.1° global POP simulation averaged over 10° longitude bands in the Indian (60°–70°E), Pacific (150°–160°E; 140°–130°W), and Atlantic (40°–30°E) Oceans (Figure 3). These sections include the variability associated with the Antarctic Circumpolar Current (ACC) and western boundary currents and their extensions. As well, the eastern boundary current region of the Pacific Ocean is examined. The POP output and data used in this calculation were both for the period 1998-2000 and the model fields were spatially binned to match the resolution of the 1/3° x 1/3° Mercator AVISO grid.

The agreement between the observed and simulated variability is generally good in all sections for both peak and background values. In the western Indian Ocean, all notable peaks have the same location, width, and magnitude except for the variability associated with the ACC at about 45°S; here POP overestimates the observed values. The variability associated with the ACC is dominated by the mesoscale eddy band (see Figure 4), which is defined as consisting of time scales of 20–150 days and spatial scales of 50–500 km. Le Traon et al. (2001) found that in regions dominated by the mesoscale, a North Atlantic 0.1° POP simulation forced with daily varying atmospheric surface forcing produced high-wavenumber, high-frequency (<20 days) processes that even multi-satellite configurations were unable to resolve. These signals explained more than 5 percent of the total SSHA variance, so it is very possible that the discrepancy in the ACC can be similarly explained. In the western Pacific, the match between the observed and simulated variability is good except at about 22°S where the model fails to reproduce the observed peak. As well, it slightly underestimates the variability in the tropics. The model underestimates the amplitude of the annual cycle in these locations (figure not shown) so it is likely that the mismatches are partially due the representation of the annual cycle by the model. The model biases in the eastern tropical Pacific and in the Brazil Current/Malvinas Confluence are similarly explained. The Gulf Stream mismatch is due to the Gulf Stream failing to form the Northwest Corner (Maltrud and McClean, 2005).

The mesoscale eddy band generally dominates ocean variability. Because eddies transport heat and momentum, and interact with the mean flow, it is important to understand their contribution to total ocean variability, and also their transport properties and dynamics. The measurement of the eddy contribution, however, is an extremely challenging observational problem on basin and global scales. SSHA fields from altimetry provide a near-global perspective. Figure 4 shows the ratio of mesoscale variability relative to the total RMS SSHA from (a) AVISO altimetry and (b) global 0.1° POP; both fields were calculated for the

> The availability of data on a near-global basis of sufficiently long duration for statistical analyses posed another constraint leading to the use of both Eulerian and Lagrangian methods for the comparisons of consistent analyses of model output and observations.

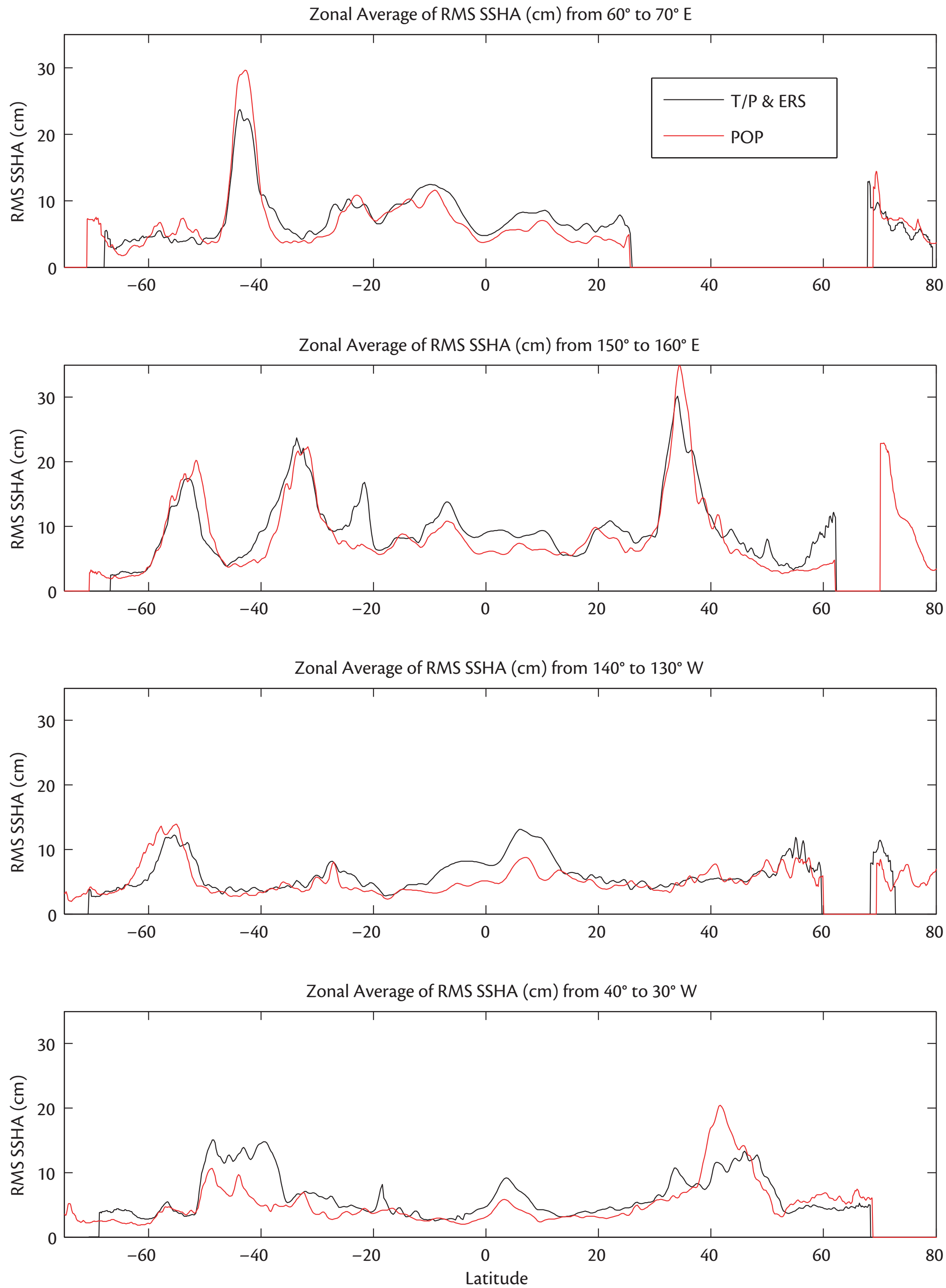



Figure 3. Meridional sections of root-mean-square (RMS) sea surface height anomaly (SSHA) from the AVISO-blended altimetry (*TOPEX/POSEIDON* and *ERS 1* and 2) product (black lines) and the 0.1° global POP simulation (red lines) averaged over 10° longitude bands in the Indian (60°–70°E), Pacific (150°–160°E; 140°–130°W), and Atlantic (40°–30°E) Oceans for 1998-2000. These sections include the variability associated with the Antarctic Circumpolar Current (ACC) and western boundary currents and their extensions. As well, the eastern boundary current region of the Pacific Ocean is examined. The agreement between the simulated and observed variability is generally good; underestimated variability by the model is likely due to the representation of the annual cycle in the model while overestimation in the ACC just eastward of the Agulhas Retroflection may be due to the model's ability to resolve high-frequency (<20 days), high-wavenumber processes that cannot be resolved by multi-altimeter data as originally found by Le Traon et al. (2001) in the Gulf Stream.

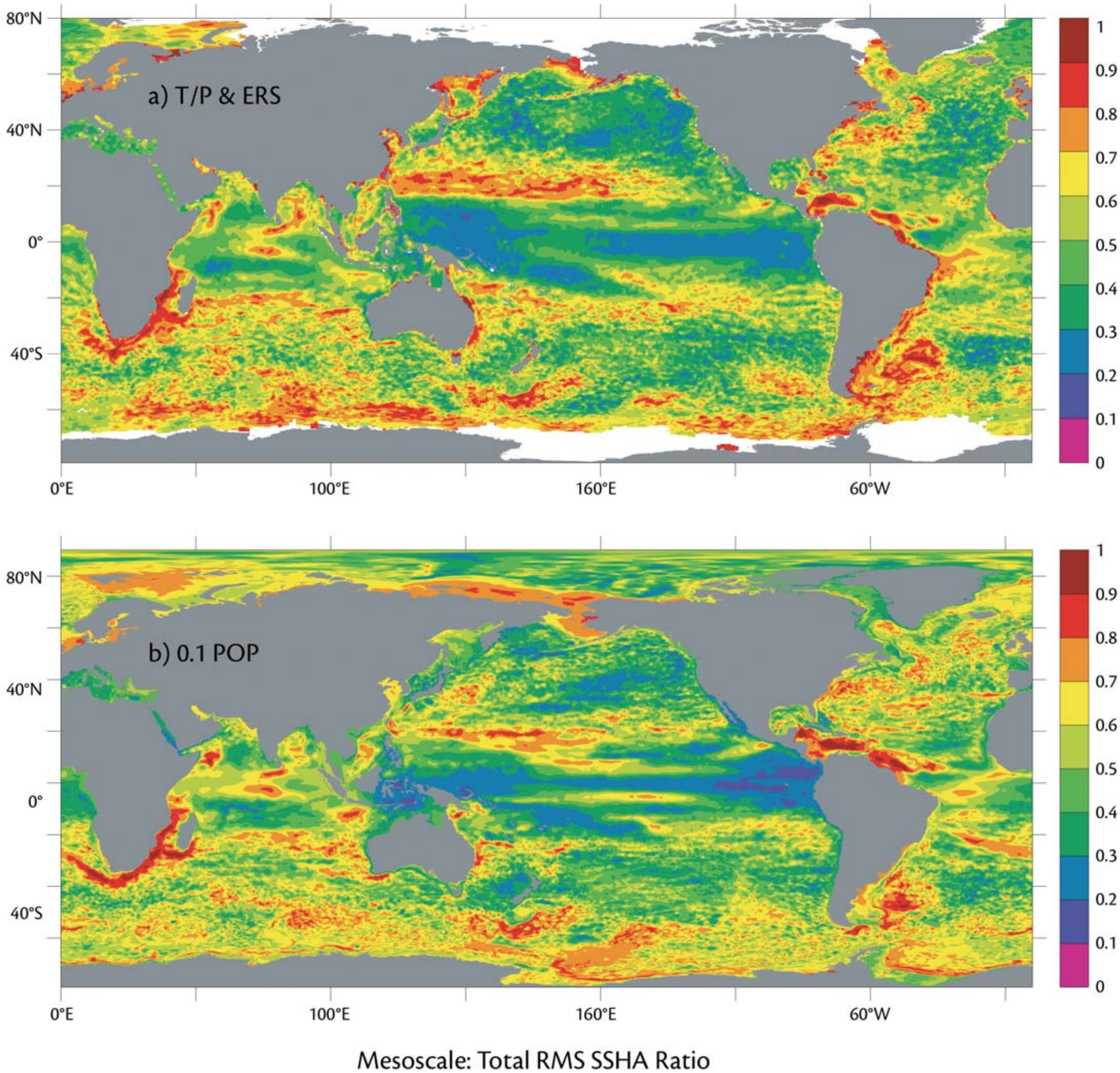



Figure 4. The ratio of mesoscale to total root-mean-square (RMS) sea surface height anomaly (SSHA) from (a) the AVISO-blended altimetry (*TOPEX/POSEIDON* and *ERS 1* and 2) product and the (b) global 0.1° POP model for 1997–2001. The mesoscale variability is obtained by band-pass filtering the SSHA between 20 and 150 days. The agreement is generally good except along the pathway of the Agulhas eddies in the South Atlantic where the variability is overestimated relative to the observed results. This bias may be due the representation of bottom topography in the model.

period 1997-2001 and the model was sub-sampled every 7 days and binned to match the temporal and spatial nature of the data. The mesoscale energy is obtained by band-pass filtering the SSHA between 20 and 150 days; we are aware that it also contains energy at scales longer than a few hundred kilometers. The overall comparison shows generally good agreement in terms of the magnitude and spatial distribution of the ratio. Biases can be identified south of 60°S in the Southern Ocean where the simulation appears to be underestimating the mesoscale variability while it is overestimated along the pathway of the Agulhas eddies in the South Atlantic. The former bias may be due to the quality of the atmospheric surface fluxes used to force POP at these latitudes, while the representation of the bottom topography in the model maybe responsible for the latter bias.

Often is it useful to represent how well a model field compares with a given observational data set as concisely as possible, especially if one has several different models or model variables to compare. Single numbers such as the correlation coefficient and the RMS difference provide an overall measure of how well two fields agree, with the former giving information about the structure (or phase) of the two fields, and the latter about relative amplitude. In fact, these two measures are related to each other through the following relation:

$$E^2 = 1 + \sigma^2 - 2\sigma R$$

where $\sigma$ is the ratio of the standard deviation of the model field to the standard deviation of the data, $E$ is the RMS difference between the model and data (normalized by the data standard deviation), and $R$ is the correlation coefficient. By noting that this relationship is simply the triangle law of cosines (with one side of length unity and $cos[angle] = R$), Taylor (2001) introduces a concise way of representing these quantities as a single point on a two-dimensional diagram.

One particularly useful application of the Taylor diagram is in interpreting changes in the fidelity of a set of simulations after some kind of significant change has been made to the model. Here we consider how well the SSHA variability of the 0.1° global simulation compares with the AVISO data, and how it relates to a another POP simulation that differs only in the horizontal resolution (0.4°) and the magnitude of horizontal friction coefficients (Figure 5). The red/blue circles denote the agreement of the 0.4°/0.1° simulation with the data over the entire globe (from 70°S to 70°N) and three specific oceanic regions representing differing levels of eddy activity: high eddy intensity (Southern Ocean), relatively low intensity (open Pacific Ocean), and mixed intensity (North Atlantic Ocean). In this diagram, the distance from the origin to a given point is equal to $\sigma$, the angle between the point and the x-axis is related to $R$, and the distance of the point from unity on the x-axis is equal to $E$. The black semicircle represents the location of perfect agreement (based on these measures) between the simulated field and the data, because such a comparison would yield $\sigma = 1$ (both have the same standard deviation), $R = 1$ (perfect correlation), and $E = 0$ (no RMS difference).

We can see from Figure 5 that in all four geographical regions there is a marked improvement in the agreement with data when going from 0.4° to 0.1° resolution. In particular, the standard deviation ratio ($\sigma$) for the global domain increases from 0.7 to just over 1, but that is only part of the story—the correlation is also improved from $R = 0.5$ to $R = 0.65$. Only in the North Atlantic is there not necessarily an improvement in $\sigma$ (because the 0.1° run overshoots unity by about the same amount as the 0.4° undershoots), while all regions show an increase in correlation ($R$) and reduction in normalized RMS difference ($E$). It also appears that much of the apparent

Overall, these quantitative and semi-quantitative metrics provided a gauge of the veracity of the simulated upper-ocean circulation, indicating that these fine-resolution models will be useful in future climate simulations.

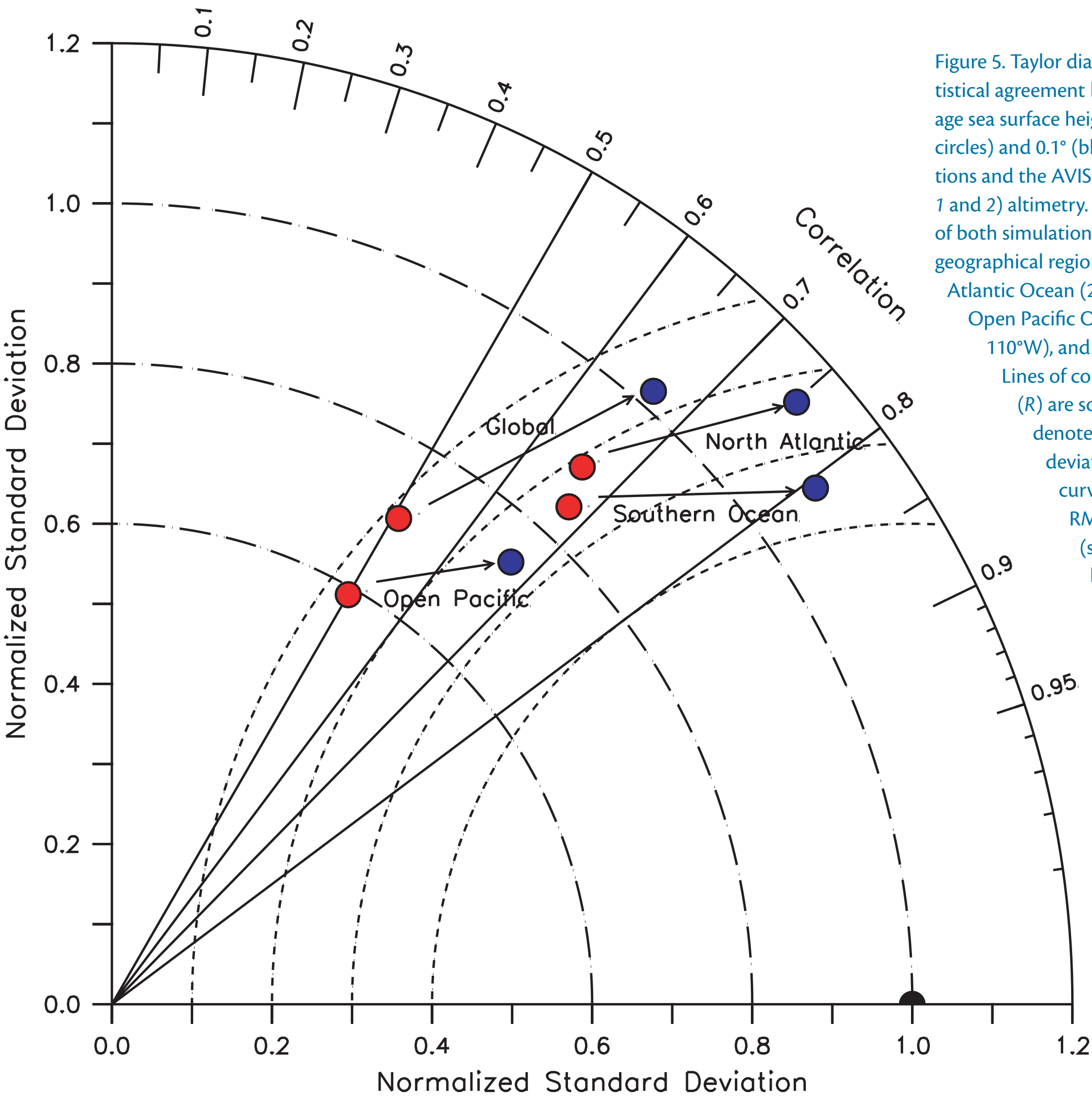



Figure 5. Taylor diagram showing the level of statistical agreement between the 1994–2001 average sea surface height anomaly from the 0.4° (red circles) and 0.1° (blue circles) global POP simulations and the AVISO (TOPEX/POSEIDON and *ERS 1* and 2) altimetry. The arrows connect the results of both simulations evaluated over the following geographical regions: Global (70°S–70°N), North Atlantic Ocean (20°N–55°N, 100°W–20°W), Open Pacific Ocean (30°S–30°N, 150°E–110°W), and Southern Ocean (65°S–40°S). Lines of constant correlation coefficient (*R*) are solid; the long dashed curves denote lines of constant standard deviation ratio (σ); the short dashed curves denote lines of constant RMS difference, varying from 0.6 (small radius) to 0.9 by 0.1. The black semicircle represents the location of perfect agreement between the simulation and the comparison data set.

improvement in σ for the global domain is likely due to an overestimate of variability in eddy-active regions (also seen in Figure 3), while the more quiescent regions still have too low variability.

## MIXED-LAYER DEPTHS

Up to this point, we have focused on the representation of the horizontal circulation of the ocean, but it is also important to investigate the time-varying vertical structure of the model. Hydrographic and moored observations, and data from profiling (e.g., PALACE) floats and acoustic Doppler current profilers (ADCPs) all can be used for this purpose. In some cases, locations are monitored on a regular basis, providing a more complete analysis of trends and variability. Two such places are the Bermuda Ocean Time-series Study (BATS) and the Hawaii Ocean Time series (HOT), nominally located at 32°N, 64°W and 22°N, 158°W, respectively.

Because the ocean communicates with the atmosphere only at its surface, the fidelity of the model's mixed layer is very important for correctly simulating ventilation processes. The horizontal resolution may not be as important for this, but the vertical resolution is crucial, as is the choice of a high-quality vertical mixing scheme, such as the K-Profile Parameterization (KPP) developed by Large et al. (1994) that is used in the global and North Atlantic 0.1° POP simulations forced with NCEP/NCAR atmospheric fluxes. Figure 6 shows time series of daily averaged mixed-layer depth from the global 0.1° POP for 1998 at the location of BATS and for 1999 at that of HOT. Mixed-layer depth calculated from data using the same algorithm as for the model (the maximum vertical gradient in near-surface temperature) is also shown for the dates when actual surveys were performed. The comparison at BATS is extremely good except for winter, where it appears that the model is able to reproduce only the *maximum* depth of the mixed layer.

The main reason for the wintertime discrepancy is that the model forcing and output are averaged over a day, while the data are essentially instantaneous samples taken over a day or two for each cluster of observations. This sampling allows measurement of the rapidly evolving mixed layer in different stages of formation, while the model has aver-

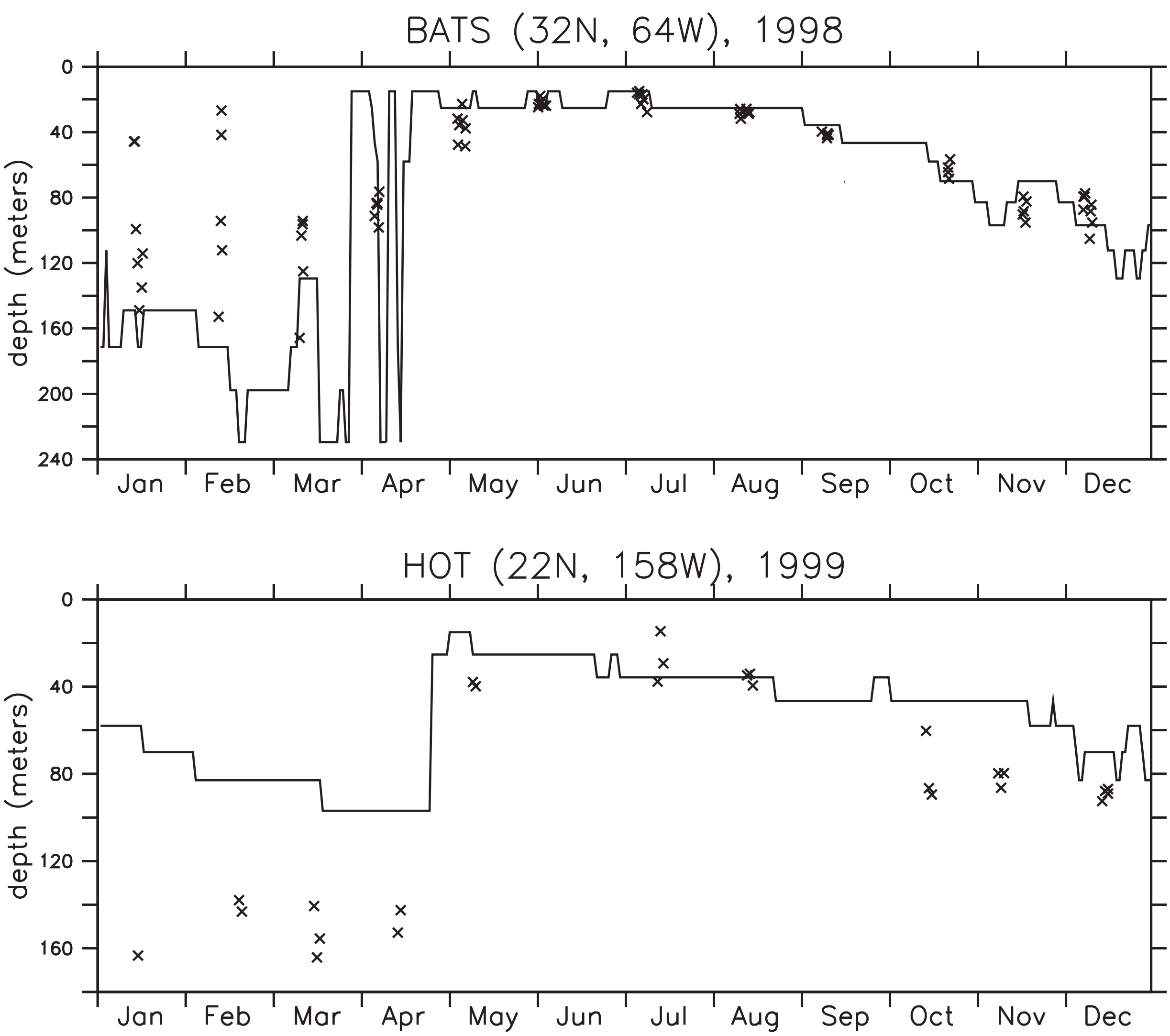



Figure 6. Time series of the mixed-layer depth (solid curves) from the global POP 0.1° simulation for the Bermuda Atlantic Time-series Study (BATS, nominally located around 32°N, 64°W) for 1998 (top) and the Hawaii Ocean Time-series (HOT, nominally located around 22°N, 158°W) for 1999 (bottom). The X's mark the mixed-layer depth derived from data taken at these locations. The step-like results are a manifestation of the vertical grid discretization.

aged out such high-frequency motions. At HOT, it appears that the model's mixed layer is consistently too shallow in the winter and autumn. Here, the density of the water column is too stably stratified to allow realistically deep penetration of the mixed layer, which is possibly due to deficiencies in the surface forcing.

The realistic representation of sea surface temperature, particularly its annual cycle, is important in ocean models to be used for climate studies.

## DISCUSSION AND CONCLUSION

Selected metrics were chosen to assess the realism of the upper-ocean circulation in large-scale fine-resolution ocean model simulations. The choice of metrics was driven by our desire to understand the fidelity of these simulations in the context of their potential use in future fine-resolution, coupled climate-system studies. The availability of data on a near-global basis of sufficiently long duration for statistical analyses posed another constraint leading to the use of both Eulerian and Lagrangian methods for the comparisons of consistent analyses of model output and observations. Overall, these quantitative and semi-quantitative metrics provided a gauge of the veracity of the simulated upper-ocean circulation, indicating that these fine-resolution models will be useful in future climate simulations.

Examining SSHA variability from altimetry and POP provided insight into the representation of mesoscale processes and the annual cycle in POP, and data limitations. The total simulated variability was lower than the observed values in regions where the amplitude of the annual cycle was underestimated by the model. It was overestimated in eddy energetic regions such as to the east of the Agulhas Retroflection in the ACC. We speculate that part of the difference may be due to the altimeter being unable to resolve high-wavenumber, high-frequency (<20 days) variability. Model biases associated with the positioning of energetic currents were also identified. A Taylor diagram of SSHA variability from altimetry and two global POP simulations with horizontal resolutions of 0.4° and 0.1° showed a measurable increase in realism using finer resolution. As well, it showed the 0.1° variability to be overestimated in eddy-active regions while the more quiescent regions still had too low variability (as in the coarser-resolution simulation) relative to the altimetry values.

Locations of differences (at the 95 percent significance level) between pseudo-Eulerian means, calculated from North Atlantic sub-polar gyre surface drifter data, and collocated North Atlantic POP output, were determined using a statistical James test. The climatically important Greenland boundary currents and the North Atlantic Current were generally not significantly different. The zonal and meridional dispersion of floats at 1750 m and North Atlantic POP floats were in good agreement; both showed two distinct power-law ranges with a similar transition time between ranges. These results indicated that the dispersive characteristics of the model in this location were realistic—a combined result of the choice of horizontal resolution and mixing parameters. Comparisons of two observational time series of mixed-layer depth showed some discrepancies that may be related to the applied surface forcing.

The metrics used in this study are far from exhaustive. The realistic representation of sea surface temperature, particularly its annual cycle, is important in ocean models to be used for climate studies. Profiling ALACE and ARGO floats provide measures of the variability of the vertical varying structure of temperature and salinity. Comparisons with geochemical, biological, and anthropogenic tracers are also useful in discovering deficiencies in the simulated circulation. On a community-wide basis the need for more accurate synoptic and climate-prediction systems has precipitated a workshop whose goal is the further development of ocean model metrics. It was held in February 2006 at the East-West Center in Honolulu; many new and novel approaches resulted from this meeting.

## ACKNOWLEDGEMENTS

The Office of Naval Research (N000140610112), the National Science Foundation (OCE-0221781), and the Department of Energy (CCPP Program) supported this work. Computer time was provided through the Department of Defense High Performance Computing Modernization Office at the Maui High Performance Computing Center and the Naval Oceanographic Office (NAVO) in Mississippi as part of a Grand Challenge Award. Detelina Ivanova (Scripps Institution of Oceanography) prepared the NCEP/NCAR forcing fields for the periods of the POP simulations analyzed here and assisted with the altimetry calculations. The altimeter products were produced by the CLS Space Oceanography Division as part of the Environment and Climate EU ENACT project (EVK2-CT2001-00117) and with support from CNES. The surface-drifter data were provided by Peter Niiler (Scripps Institution of Oceanography). Carmyn Priewe assisted the with model/surface drifter comparison. Kevin Speer (Florida State University) provided the EUROFLOAT data. The authors also thank all those involved in the collection of drifter and float data in the North Atlantic and in the Nordic Seas. Ivan Lima (Woods Hole Oceanographic Institution) provided the data from BATS and HOT in a readily usable form. The three reviewers are thanked for their comments. Preparation of this paper was performed using Naval Postgraduate School computers.

## REFERENCES


Barron, C.N., A.B. Kirol, H.E. Hurlburt, C. Rowley, and L.F. Smedstad. 2004. Sea surface height predictions from the Global Navy Coastal Ocean Model during 1998-2001. *Journal of Atmospheric and Oceanic Technology* 21:1,876–1,893.

Brachet, S., P.Y. Le Traon, and C. Le Provost. 2004. Mesoscale variability from a high-resolution model and from altimeter and from data in the North Atlantic Ocean. *Journal of Geophysical Research* 109, doi: 10.1029/2004JC002360.

Bryan, K. 1969. A numerical method for the study of the circulation of the world ocean. *Journal of Computational Physics* 4:347–376.

Cox, M.D. 1970. A mathematical model of the Indian Ocean. *Deep Sea Research* 17:45–75.

Cox, M.D. 1984. A primitive equation three dimensional model of the ocean. *GFDS Ocean Group Technical Report. 1*, *Geophysical*. National Oceanic and Atmospheric Administration, Fluid Dynamics Laboratory and Princeton University, New Jersey, 250 pp.

Dijkstra, H., and M. Ghil. 2005. Low-frequency variability of the large-scale ocean circulation: A dynamical systems approach. *Reviews of Geophysics* 43(RG3002): doi:10.1029/2002RG000122.

Donohue, K.A., E. Firing, G.D. Rowe, A. Ishida, and H. Mitsudera. 2002. Comparison between observed and Modeled Pacific Equatorial Subsurface Countercurrents. *Journal of Physical Oceanography* 32:1,252–1,264.

Garraffo, Z.D., A.J. Mariano, A. Griffa, C. Veneziani, and E.P. Chassignet. 2001. Lagrangian data in a high resolution numerical simulation of the North Atlantic. I: Comparison with *in-situ* drifter data. *Journal of Marine Systems* 29:157–176.

Krauss, W., and C.W. Böning. 1987. Lagrangian properties of eddy fields in the northern North Atlantic as deduced from satellite-tracked buoys. *Journal of Marine Research* 45:259–291.

Large, W.G., J.C. McWilliams, and S.C. Doney. 1994. Oceanic vertical mixing: A review and a model with a nonlocal boundary layer parameterization. *Reviews of Geophysics* 32:363–403.

Le Traon, P.Y., G. Dibarboure, and N. Ducet. 2001. Use of a high-resolution model to analyze the mapping capabilities of multiple-altimeter missions. *Journal of Atmospheric and Oceanic Technology* 18:1,277–1,287.

Maltrud, M.E., and J.L. McClean. 2005. An eddy resolving global 1/10° ocean simulation *Ocean Modelling* 8(1–2):31–54.

Masumoto, Y., H. Sasaki, T. Kagimoto, N. Komori, A. Ishida, Y. Sasai, T. Miyama, T. Motoi, H. Mitsudera, K. Takahashi, H. Sakuma, and T. Yamagata. 2004. A fifty-year eddy-resolving simulation of the world ocean—Preliminary outcomes of OFES (OGCM for the Earth Simulator). *Journal of the Earth Simulator* 1:35–56.

Maximenko, N.A., B. Bang, and H. Sasaki. 2005. Observational evidence of alternating zonal jets in the world ocean. *Geophysical Research Letters* 32(L12607):doi:10.1029/2005GL022728.

McClean, J.L., P-M. Poulain, J.W. Pelton, and M.E. Maltrud. 2002. Eulerian and Lagrangian statistics from surface drifters and a high-resolution POP simulation in the North Atlantic. *Journal of Physical Oceanography* 32:2,472–2,491.

Moore, J.K., S.C. Doney, and K. Lindsay. 2004. Upper ocean ecosystem dynamics and iron cycling in a global three-dimensional model. *Global Biogeochemical Cycles* 18(GB4028): doi:10.1029/2004GB002220.

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. *Journal of Marine Systems* 21:307–320.

Seber, G.A.F. 1984. *Multivariate Observations*. John Wiley and Sons, Inc., Hoboken, NJ, USA, 686 pp.

Semtner, A.J. 1974. An oceanic general circulation model with bottom topography. *Technical Report 9*. University of California-Los Angeles, Department of Meteorology, Los Angeles, CA, USA, 99 pp.

Smith, R.D., J.K. Dukowicz, and R.C. Malone. 1992. Parallel ocean circulation modeling. *Physica D: Nonlinear Phenomena* 60:38–61.

Smith, R.D., M.E. Maltrud, F.O. Bryan, and M.W. Hecht. 2000. Numerical simulation of the North Atlantic Ocean at 1/10°. *Journal of Physical Oceanography* 30:1,532–1,561.

Speer, K.G., J. Gould, and J. LaCasce. 1999. Year-long float trajectories in the Labrador Sea Water of the eastern North Atlantic Ocean. *Deep-Sea Research II* 46(1–2):165–179.

Stammer, D. 1997. Global characteristics of ocean variability estimated from regional *TOPEX/POSEIDON* altimeter measurements. *Journal of Physical Oceanography* 27:1,743–1,769.

Taylor, G.I. 1921. Diffusion by continuous movements. *Proceedings of the London Mathematical Society* 20:196–212.

Taylor, K.E. 2001. Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research* 106(D7):7,183–7,192.

Tokmakian, R., and J.L. McClean. 2003. How realistic is the high frequency signal of a 0.1° resolution ocean model? *Journal of Geophysical Research* 108(C4):3115, doi:10.1029/2002JC001446.

Treguier, A.M., N.G. Hogg, M. Maltrud, K. Speer, and V. Thierry. 2003. The origin of deep zonal flows in the Brazil Basin. *Journal of Physical Oceanography* 33:580–599.

Treguier, A.M., S. Theetten, E.P. Chassignet, T. Penduff, R. Smith, L. Talley, J.O. Beismann, and C. Böning. 2005. The North Atlantic subpolar gyre in four high-resolution models. *Journal of Physical Oceanography* 35:757–774.