

Total kinetic energy in four global eddy ocean circulation models and over 5000 current meter records

Robert B. Scott^{a,c,*}

^a*Institute for Geophysics, Jackson School of Geosciences, The University of Texas
at Austin, J.J. Pickle Research Campus, Bldg. 196 (ROC), 10100 Burnet Rd.
(R2200), Austin, TX 78759, United States*

Brian K. Arbic^b Eric P. Chassignet^b Andrew C. Coward^c
Mathew Maltrud^d Ashwanth Srinivasan^e Anson Varghese^a

^b*Department of Oceanography and Center for Ocean-Atmospheric Prediction
Studies, The Florida State University, Tallahassee, FL 32306, United States*

^c*James Rennell Division for Ocean Circulation and Climate, National
Oceanography Centre, Southampton, Southampton, SO14 3ZH, United Kingdom*

^d*Climate, Ocean and Sea Ice Modeling Project, Fluid Dynamics Group, Los
Alamos National Laboratory, Los Alamos, NM 87545, United States*

^e*RSMAS/MPO, University of Miami, Miami, FL, United States*

Abstract

We compare the total kinetic energy (TKE) in four global eddy ocean circulation simulations with a global dataset of over 5000, quality controlled, moored current meter records. At individual mooring sites, there was considerable scatter between

models and observations, that was greater than estimated statistical uncertainty. Averaging over all current meter records in various depth ranges, all four models had mean TKE within a factor of two of observations above 3500 m. With the exception of observations between 20 m and 100 m, the models tended to straddle the observations. However, individual models had clear biases. The free running model (no data assimilation) biases were largest below 2000 m. Data assimilation clearly improved the model-observation comparison, especially below 2000 m, despite data assimilation being only applied above this depth and only south of 47°N. Different diagnostics revealed different aspects of the comparison, though in general the models appeared to be in an eddying regime with TKE that compared reasonably well with observations.

Key words: eddying OGCM, kinetic energy, moored current meters, model validation, model intercomparison

1 Introduction

In the current decade several global ocean general circulation models, OGCMs, have been run in the eddying-regime; for a review of the state-of-the-art, see the collection of papers edited by Hecht and Hasumi (2008). Since we expect the processes generating mesocale variability are now, at least potentially, well represented by the discrete formulations of the governing equations, the question of how well the simulated mesocale currents compare with observations becomes increasingly important. Because the meoscale velocity field is observed to be highly variable and believed to be strongly turbulent, we can

* Corresponding author. Tel.: +1 512 471-0375; fax: +1 512 471-8844
Email address: rscott@ig.utexas.edu (Robert B. Scott).

only hope to compare statistical properties of the mesoscale velocity field, such as the total kinetic energy levels in the zonal and meridional velocities.

Many studies have compared OGCM surface kinetic energy, or sea level height anomaly variability, with observations from satellite-based radar altimeters, and most OGCMs with resolution about $1/6$ degrees or better compare quite favorably over most of the World Ocean (e.g. Paiva et al., 1999; Smith et al., 2000; Hurlburt and Hogan, 2000; Maltrud and McClean, 2005; Barnier et al., 2006; Chassignet et al., 2009). Less is known about the deeper ocean kinetic energy, though high-resolution North Atlantic simulations with a hydrodynamic six-layer model, NLOM, suggested that model resolution as high as about $1/32^\circ$ may be necessary to obtain convergence of the abyssal kinetic energy (Hurlburt and Hogan, 2000). Because of the poor vertical resolution of the NLOM model, and the unusual compression of topography employed, it's difficult to extrapolate these results to other models. But at the very least this study alerts us to the possibility that resolution much higher than the first baroclinic mode Rossby radius of deformation might be required to obtain realistic subsurface flows, see discussion by Treguier (2006).

In contrast with the wealth of observational information conveniently available at the ocean surface, subsurface data are much more limited, and more cumbersome to work with. Moored current meters provide the best available Eulerian data. The most comprehensive comparison of OGCM and moored current meter kinetic energy throughout the water column has been by Penduff et al. (2006). They compared 891 Atlantic current meter records from the World Ocean Circulation Experiment (WOCE) with the $1/6^\circ$ CLIPPER Atlantic OGCM. They found the model currents tend to be too baroclinic (too vertically sheared, so they are increasingly too weak at greater depths).

This suggests that even if OGCMs compare well with surface observations, important model biases may remain unnoticed at depth. This concern was substantiated by Arbic et al. (2009), who compared the abyssal flows in two versions of NLOM and global $1/10^\circ$ POP with a superset of the WOCE current meter database, maintained by Oregon State University. They found that the point-by-point comparison of time-averaged cube of bottom flow speeds in models and current meters was poor (there was a great deal of scatter in scatterplots). When the time-averages were then averaged over many mooring sites, the models performed better (within a factor of 2.7 or less for the cube of bottom flow), but still displayed a weak bias, as found by Penduff et al. (2006). The present study is more comprehensive than these two earlier studies, in that we will undertake comparisons both throughout the full-water column (as done by Penduff et al. 2006), and throughout the globe (as done by Arbic et al. 2009).

There is some suggestion that the present generation global eddying models should be able to produce realistic energy levels, and vertical structure of currents. Evidence in support of a realistic vertical structure of the currents in a $1/10^\circ$ North Atlantic POP simulation is provided by Smith et al. (2000, Fig. 8), showing good agreement between baroclinic, barotropic and total Gulf Stream transports with those from current meter moorings by Hogg (1992) and Johns et al. (1995). Further evidence is provided by comparison with eddy kinetic energy (EKE) from moored current meters along 48°N (Colin de Verdiere et al., 1989), see Smith et al. (2000, Fig. 16). Note in particular the strong EKE at depths greater than 1000 m along 35°W , with values between about 25 and 44 cm^2/s^2 in the mooring, which corresponds well with about 50 cm^2/s^2 in the model.

There are now several global OGCMs run in a realistic configuration with horizontal resolution of $1/10^\circ$ to $1/12^\circ$ (see Table 1), which resolve the first Rossby radius of deformation throughout most of the domain (*i.e.* up to $\sim 55^\circ\text{N}$) and are able to have a good representation of upper ocean baroclinic processes. How well do these models simulate total kinetic energy (TKE) throughout the full extent of the water column? Is resolution the only factor that determines a model's success? The global models in Table 1 have comparable resolution to the $1/10^\circ$ North Atlantic POP model. But their numerical formulations are quite different. The present study aims to assess the ability of these models to simulate realistic zonal and meridional TKE. This will be assessed using a large moored current meter archive, CMA. Section 2 introduces the CMA compiled for this study, the quality control we applied, and the data processing steps. Section 3 describes the global eddy OGCMs analyzed in this study. Section 4 presents the comparison of OGCMs and the CMA. We conclude with a discussion and summary of the key findings in Section 5.

2 Moored current meter records

2.1 General limitations of moored current meter records

While moored current meters provide our best knowledge of Eulerian current statistics, the reader should bear in mind their inherent limitations. These are reviewed briefly.

Subsurface moorings in deep water rarely have measurements within the upper 200 m. Their most serious limitation is that they tend to blow over in strong currents, effectively providing a measurement at varying depths that are, when

currents are stronger, lower in the water column than their nominal depth. The problem can be tracked for moorings with pressure sensors, and can be quite substantial, with hydrostatic pressure variations reaching 1000 db or more. Schemes have been devised to correct for blow-over of moorings without pressure sensors (Hogg, 1991; Meinen and Luther, 2002). To the best of our knowledge, none of the records in our archive were corrected for mooring blow-over (Wunsch, pers. comm. 2009 and Joseph Bottero, pers. comm. 2009), and we did not attempt any corrections.

Figure 1 shows the distribution of blow-over from all the records in our database that had reliable pressure recordings at the same depth as the current meters. From the distribution of interquartile range of pressure we see that most current meters spent at least half their deployment within a few dozen db of their median depth, but the most extreme case had half its measurements 170 db above or below its median depth. About 10% of the records had 25% of their measurements from depths more than 70 db below their minimum depth. For the model comparison in Section 4 we used the nominal depth given for each record, since not all records had pressure sensors to estimate the mean depth. Maximum blow-over greater than a few hundred db was rare, but the worst case reached 1860 db! This occurred for a WOCE mooring deployed in the East Australia Current in 4435 m water depth. The large blow-over event occurred around 3000 hours, see upper panel of Figure 2. The current speed recorded at this time was large, reaching over 0.5 m/s, but this is likely to be an underestimate of the true speed at the nominal deployment depth of 600 m, because currents tend to be weaker at greater depth. This blow-over effect introduces a bias in our estimates of TKE because these strong events tend to be underestimated. One should also note that observationalists tend to

avoid deploying moorings in the very strongest currents so as to avoid strong blow-over, which introduces some bias in the sampling of the ocean we have for comparison. On the other hand, moorings are often deployed in oceanographically interesting places, which we found includes some bias toward sampling stronger current regions (e.g. Sen et al., 2008).

Figure 1

Figure 2

Surface moorings can provide measurements near the surface without the problems of blow-over, but have other problems. During the MODE experiment (MODE Group, 1978) it was discovered that surface moorings can give faulty measurements (exaggerating TKE by a factor of five or more for measurements below 500 m!) due to a mode of oscillation excited by surface waves (Gould et al., 1974). The dual propeller Vector Measuring Current Meters (VMCM) (Weller and Davis, 1980) are generally used on surface moorings since field tests showed them to be less affected by wave motion (Halpern et al., 1981). Compliant elements of the mooring are also now used to tune the resonance outside the strong signals in the surface wave band. Of course current meters can only be expected to measure water velocities relative to the moving surface buoy. Plueddemann and Farrar (2006) estimate the errors this introduces to absolute near-inertial current measurements. For the TKE of interest here, the error estimates would be difficult, and not clearly of one sign. (The surface buoys must have some slack to minimize static tension in strong winds. To quote an extreme example, Plueddemann et al. (1995) found a mooring length 1.25 times the water depth to be necessary to survive the challenging environment of the subarctic North Atlantic; this allowed the surface mooring

to drift in a circle of radius $3/4$ of the water depth.)

The CMA described in the next section includes measurements from a variety of sensors from different manufacturers. Mechanical devices include the vector averaging current meter (VACM) developed in the 1960s at Woods Hole Oceanographic Institution and vector measuring current meter (VMCM) (Weller and Davis, 1980). An important limitation of these mechanical current meters is their minimum detectable current speed, below which the rotor stalls. This varies with sensor, with typical stall speeds of $1-2$ cm/s. Recorded speeds less than the stall speed are handled differently by different investigators. For instance, speeds less than the stall speed might be set to $1/2$ the stall speed, in a crude attempt to minimize the error in mean statistics. Acoustic current meters on the other hand rely on either the Doppler shift of echos from suspended particles (Acoustic Doppler Current Profilers or ADCP) or the difference in acoustic travel time between sensor pairs (VCM). Hogg and Frye (2007) compared two VACMs with two acoustic current meters, an Aanderaa RCM11 ADCP and a Nobska MAVS VCM all near 2000 m in a deep-water mooring south east of Bermuda. The ADCP measured speeds about $10-25\%$ lower (*i.e.* TKE up to 45% lower) than the reference VACMs for current speeds up to about 15 cm/s. The VCM had a constant offset from the VACMs of about $2-3$ cm/s. The investigators emphasized the challenging conditions of the comparison: low current speeds and clear water (low scattering levels). This is a reasonable interpretation since Gilboy et al. (2000) found an agreement to within statistical uncertainty between a VACM, a VCM and an ADCP at a nearby site and at 72 m instrument depth (where the ocean is more energetic and contains more suspended particles for scattering).

2.2 *Multiarchive current meter dataset*

The Current Meter Archive (CMA) was created by combining the Deep Water Archive of the Oregon State University (OSU) Buoy Group, 1901 current meter records collected by Carl Wunsch, and other sources. The current meter records were between Sept. 1973 and Feb. 2005. The OSU dataset contains over 5000 current meter records (including acoustic and mechanical devices, on surface and subsurface moorings) from many investigators and includes the WOCE archive. Most records are in deep water and typically have at least 6-month duration. Each record was visually inspected and quality controlled by the OSU Buoy Group as described on their website <http://kepler.oce.orst.edu/>. The records were checked for typical problems such as stalled rotors or sticking compass, fouling of speed sensors, tape glitches, etc. Problems were removed and gaps less than about a week long were filled in with predictive interpolation, a maximum entropy method designed to produce less distortion to the power spectrum than simple linear interpolation. Longer stretches of suspect data were flagged as missing. Further details can be found on the OSU Buoy Group webpage.

The archive provided by Carl Wunsch contained 1901 records on 525 moorings, of which over 100 moorings were visually inspected (Wunsch, pers. com. 2009) and analyzed by Wunsch (1997). Literature citations to the first published work on the various moorings were tabulated by Wunsch (1997). The records were further quality controlled for the present study by visual inspections and comparing with records that also appeared in the OSU dataset. Some outliers were identified through comparison with the models in the current study, and problems traced to various sources such as units errors.

We also obtained 59 current meter records from several experiments in the online archive maintained by the Upper Ocean Processes Group at Woods Hole Oceanographic Institution, <http://uop.whoi.edu/index.html>.

When working with large volumes of data, Murphy's Law comes into effect. One problem was removing repeated records both within the large OSU archive, and between archives. We searched for redundancies with the following chosen tolerances for deciding if records were co-located: 0.005° latitude and longitude, and 5 cm in the vertical. This revealed over 500 redundant records, though more redundancies could be found with less strict tolerances. Records overlapping in time were dealt with when computing the mean statistics.

Our CMA and the software interface are described at <http://www.ig.utexas.edu/research/projects/cma/data.html>, and is available upon request.

2.3 Data selection and preprocessing

We found 5814 unique current meter records that were moored in more than 10 m of water, and had more than 90 days of good data. These 5814 records were further checked with automated routines for problems such as rotor or vane being stuck, or suspicious gaps in the data. This was accomplished by dividing each time series into sections of at least 10 days of consecutive points (or at least 30 points for records with sampling period greater than 8 hrs) and checking if the standard deviation of either u , v , or speed was less than $\epsilon = 5 \times 10^{-4}$ m/s, or the standard deviation of direction was less than $\epsilon = 0.05^\circ$. This identified a further 46 records as suspect. Reducing the threshold by $\times 5$ lower (less strict) yielded 40 bad records. However, doubling the time window

to 20 days cut the rejected records to only 23, so there remains some inevitable arbitrariness in choosing tolerances. After these quality control steps, there remained 5741 records on 1363 mooring locations. Sampling frequencies varied from 8 minutes to 1 day, so many records contained strong tidal and near inertial gravity wave components. To produce homogeneous records for comparison with mesoscale motions simulated in the OGCMs, the current meter (u, v) time series were reduced to 5-day averages. Records with missing data in a 5-day period were averaged over available good data if at least 5/2 days of good data were available; otherwise that time period was flagged as missing. No temporal interpolation was performed. Applying a 3-day Butterworth filter to the records, to more cleanly remove the tidal and other high-frequency signals, prior to taking 5-day averages led to negligible differences.

The POP model was compared with an earlier version of the CMA (CMA 1.0) containing only 1198 moorings of the 1363 contained in the current version (CMA 2.0). All other models were compared with the 1363 mooring database. Evaluating HYCOM and OCCAM with CMA 1.0 instead of CMA 2.0 led to changes in the correlation coefficients (Table 2) of at most 0.01, and scatter plots difficult to distinguish.

At some locations current meter moorings were re-deployed at the same location and depth multiple times. From the 5741 quality controlled current meter records we found 5339 unique depth bins with at least 90-days of accumulated 5-day averaged velocities, on 1361 mooring sites. Hereafter, we'll refer to these as "records" even those they were actually combinations of current meter records deployed at very similar locations and depths at different dates, often with considerable temporal gap. There were 4466 records with accumulated duration greater than 180 days, and these were deployed on the

1213 moorings plotted in Fig. 3 with black X's. The remaining 148 red o's indicate the moorings with records between 90 and 180 days. However, only 5055 of these unique depth bins or "records" had TKE greater than $2 \text{ cm}^2/\text{s}^2$, a threshold we imposed in the analysis below, but not the scatter plots.

Figure 3

2.4 Statistical limitations

The CMA spanned over 30 years while the model output was analyzed over a single year or 4 years. The assumption is that the models and the ocean are in a quasi-statistical equilibrium. This is certainly not strictly the case. A more serious limitation is how well the current meters records of only 90-days duration or longer represent the local long-term statistics. Assuming approximately Gaussian velocity statistics (Gille and Llewellyn-Smith, 2000) and a decorrelation time of about 10 to 20 days (LeTraon et al., 1998) we can infer the 95% confidence intervals on the velocity variance for a record of a given length as follows. Recall the $(1 - a)$ confidence intervals for the variance of a Gaussian random variable are

$$P \left(\frac{(n-1)S^2}{\chi_{(n-1), (1-a)/2}^2} < \sigma^2 < \frac{(n-1)S^2}{\chi_{(n-1), a/2}^2} \right)$$

where $\chi_{n-1, a}^2$ is the a^{th} quantile of the χ^2 distribution with $(n - 1)$ degrees of freedom. The ratio of the confidence bounds is then,

$$\frac{\text{Upper confidence bound}}{\text{Lower confidence bound}} = \frac{\chi_{n-1, 1-a/2}^2}{\chi_{n-1, a/2}^2}.$$

Values of the ratio of upper confidence bound to lower confidence bound are plotted in Fig. 4. For instance, for the upper 95% confidence bound to be less than, say, 20% more than the lower 95% confidence bound, we find $n = 926$, or at least 25 years of data! Relaxing the confidence intervals to 80% still requires $n = 397$, or about 11 years of data assuming (optimistically) 10-day decorrelation. For some of the models we had 4 years of data. Assuming a 10-day decorrelation time, we can have 80% confidence that the upper bound will be no more than 35% larger than the the lower bound. Unfortunately very few current meter records were 4-years long. For records of only 90 days we have between about 3.5 and 8 degrees of freedom, and the expected 67% confidence intervals have a ratio between about 2.75 and almost 6. Because of this large range of confidence bounds for 90-day records, we also considered the comparison of models with current meter records 180 days and longer. For the 180-day records, the upper 67% confidence bound was somewhere between 2.25 and 2.75 times the lower bound. While statistical uncertainty is larger than the measurement errors referred to above, it does not lead to a bias.

Figure 4

3 Models

Table 1

3.1 Description of POP model runs

POP (<http://climate.lanl.gov/Models/POP/>) is a publicly available, z-level, hydrostatic, Boussinesq, primitive equation ocean model that allows for

generalized orthogonal horizontal grids. It is an implicit free surface derivative of the original Bryan-Cox model with vastly improved numerics that have been widely tested. It has a wide user base and is the ocean component of NCAR CCSM (one of two U.S. contributions to the Fourth Assessment Report of the IPCC). It is being used in ultra-high resolution coupled climate simulations as part of the LLNL grand challenge project.

We analyzed an updated version of the $1/10^\circ$ grid POP model described by Maltrud and McClean (2005) that has several significant changes. The horizontal grid was modified in the Northern Hemisphere, moving from a dipole to a tripole version, resulting in more uniform resolution in the Arctic. There are 42 vertical levels, ranging from 10 m at the surface to 250 m at depth. Full-cell bottom topography (generated from ETOPO2; ETOPO2 (2006)) was replaced with partial bottom cells (Adcroft et al., 1997). Vertical diffusion coefficients were calculated using KPP (Large et al., 1994), and biharmonic representations of subgrid momentum and tracer diffusion were used, but with lower values than in Maltrud and McClean (2005). Surface forcing was calculated from the “normal year” of the CORE dataset (Large and Yeager, 2004), averaged in time from 6-hourly to monthly. Mesoscale variability agrees well with satellite altimetry in regions of high eddy variability such as the Antarctic Circumpolar Current and Western Boundary Current systems (Maltrud et al., 2008). Under-representation of mesoscale variability in the more quiescent regions such as in the central and eastern basins, however, continues to be a problem.

We converted the daily averages from years 48 to 52 to 5-day averages for comparison with the CMA. See further discussion in Section 3.4 below.

3.2 Description of HYCOM model runs

HYCOM is widely used by the ocean community (<http://www.hycom.org>) and is the backbone of the global eddy-resolving ($1/12^\circ$ horizontal resolution) real-time nowcast/forecast system at the Naval Oceanographic Office. We analyzed runs of the global $1/12^\circ$ HYCOM model run at the Naval Research Laboratory, Stennis Space Center with and without data assimilation (outputs available at <http://www.hycom.org/dataserver>). The model has nominal $1/12^\circ$ Mercator grid horizontal resolution (6.5-km grid at mid-latitudes with a bipolar patch north of 47°N , *i.e.*, 3.5-km grid spacing at the North Pole) and 32 hybrid layers in the vertical (pressure-coordinates are used in the mixed layer, isopycnal coordinates are used in the ocean interior, and terrain-following coordinates are used in shallow areas) (Bleck, 2002; Chassignet et al., 2003, 2006, 2009).

The global model was initialized with January climatology derived from GDEM3 (Generalized Digital Environment Model version 3.0; Carnes (2003)) and spun-up for 13 years with daily climatological wind and monthly mean thermal forcing derived from the 1.125 European Center for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA15) over the 1979-1993 time frame. The daily climatological wind forcing contains higher frequency variability needed for realistic simulation of the surface mixed layer and is constructed by adding 6-hourly variability from the ECMWF operational model over the period September 1994 to September 1995 to the climatological monthly winds. After the spin-up phase, two overlapping runs, a non assimilative interannual run for years 2003-2007 and a three-year data assimilative hindcast for years Nov 2003 - Dec 2006, were performed starting from January 2003. These runs

were forced by 3-hourly Navy Operational Global Atmospheric Prediction System (NOGAPS, http://www.nrlmry.navy.mil/nogaps_his.htm) wind stress, wind speed, heat flux (using bulk formula) and precipitation. Runoff from 986 rivers is included as virtual salinity flux with no mass exchange.

A multivariate Optimal Interpolation scheme (MVOI) is used for data assimilation (Daley, 1991; Cummings, 2005). The hindcast run assimilates available satellite altimeter observations, satellite and in-situ Sea Surface Temperature (SST) as well as in-situ vertical temperature and salinity profiles from XBTs, ARGO floats and moored buoys. The ARGO data was the greatest bulk of the data, and was available mostly above 2000 m. No data assimilation was performed in the bipolar grid north of 47°N.

The performance of HYCOM over the North Atlantic, North Pacific, Indian Ocean, and globally are documented by Chassignet et al. (2003); Kelly et al. (2007); Chassignet et al. (2009); Srinivasan et al. (2009).

Currently, daily snapshots of the model's 3-D state are archived and can be downloaded from the already mentioned public access site. The 3-D fields have been interpolated to the same horizontal axes, and geopotential depth levels (the depth levels used by the Levitus dataset (Locarnini et al., 2006)). We obtained all global 3D daily snapshots of zonal and meridional velocities (u, v) and potential temperature referenced to 20 MPa and salinity (θ, S) for the year 2006 for both the data assimilative and free running versions of HYCOM. The daily snapshots contain high-frequency motions, as well as the mesoscale motions of interest in this study. To reduce the influence of the high-frequency motions we averaged 5 consecutive snapshots to form 5-snapshot averages. It's important to distinguish this from a true 5-day mean

that would be formed from the Riemann sum of all time steps within a 5-day period.

To verify that the resulting 5-snapshot averages were not too contaminated with high-frequency motions, we compared them with the corresponding geostrophic velocities computed from 5-snapshot averaged (θ, S) . Figure 5 shows the comparison for the zonal velocity component for an arbitrarily chosen mooring in the midlatitude North Pacific. (Similar results, not shown, were found for the meridional velocity.) The geostrophic velocities were computed from the horizontal pressure gradients at all levels. Pressure was not provided with the model output, but was calculated from (θ, S) by simply converting θ to in situ temperature, and in situ density using MatLab Seawater routines (http://www.cmar.csiro.au/datacentre/ext_docs/seawater.htm), which contain the fully nonlinear equation of state. We vertically integrated the density using the simple trapezoidal rule. The horizontal gradients were computed using a weighted least squares fit of a 2D plane to all the points within a disk of radius 9 km. (The weights are described below, see Eq. 2.) The sea level gradient was used to obtain the surface geostrophic flow. At the surface, see upper panels, the full 5-snapshot average velocity (black lines) clearly had more variability than the geostrophic velocity (blue lines), presumably due to the direct wind forcing in the Ekman layer. At 200 m and 1000 m depth (two middle panels) the comparison was remarkably good. At 3000 m the velocity was quite weak, and the comparison was not as good. We are unsure why the deep velocities appear less geostrophic, though smoothing of the density field resulting from mapping from the hybrid coordinates to the regular grid at Levitus depth levels may be responsible.

Fig. 5

3.3 OCCAM model run

The global $1/12^\circ$ OCCAM is, like the POP model, based upon the z-level coordinate representation of the Boussinesq primitive equations, first used in GFDL's MOM model. The model horizontal resolution for OCCAM is non-isotropic, employing a regular $1/12^\circ$ latitude-longitude grid outside the Arctic and North Atlantic. In the North Atlantic and Arctic, the grid has been rotated to avoid the polar singularity, as explained at www.noc.soton.ac.uk/JRD/OCCAM/EMODES/info/coord.php3. MatLab routines to do the rotation are provided at www.ig.utexas.edu/people/staff/rscott/personal.htm. The vertical resolution was 66 levels with spacing ranging from 5 m near the surface to 200 m in the deep ocean. Topography was based upon Smith and Sandwell (1997), and the representation included partial bottom cells. Vertical viscosity was $1 \times 10^{-4} \text{ m}^2/\text{s}$ and the horizontal viscosity was $50 \text{ m}^2/\text{s}$. The forcing fields and other model details are described by Lee et al. (2007).

The model is integrated from rest with initial tracer fields from the annual mean World Ocean Circulation Experiment Special Analysis Center climatological values (Gouretski and Jancke, 1996). The model has been run starting with model year 1985, and we analyzed 5-day means from 1999 to 2003 and, separately, from 2004.

3.4 Processing model data

Our main diagnostic of interest was the mean squared velocity components formed from the 5-day (or 5-snapshot for HYCOM) averaged time series of the CMA and models. We'll denote these as

$$\begin{aligned}\langle u^2 \rangle &= \frac{1}{P} \sum_{p=1}^P u_p^2 \\ \langle v^2 \rangle &= \frac{1}{P} \sum_{p=1}^P v_p^2\end{aligned}\tag{1}$$

where $P = 72$ for the (default) yearly averages of the model data, and P is the number of pentads available for the individual current meter records (recall minimum $P = 18$). We'll refer to $\langle u^2 \rangle$ and $\langle v^2 \rangle$ as the mean squared zonal and meridional velocities, and of course $1/2$ their sum is the total kinetic energy (TKE) from the 5-day means.

It's worth mentioning that we did *not* remove the time mean from the velocity time series in Eq. 1. While it is common practice in physical oceanography to consider the time mean and anomalies separately, we avoided that herein because we were especially interested in identifying model biases using large amounts of observational data. Only about $1/3$ of the current meter records in our CMA had a duration of one year or more, implying that if we wanted to calculate the yearly means and corresponding anomalies we would be restricted to $1/3$ of the data. Using shorter time series to calculate the mean and anomalies, is possible, but introduces a stronger bias; it overestimates (underestimates) the energy in the mean (anomalies). To avoid these biases, we restricted attention to components of the TKE, for which Eq. 1 is an unbiased estimate, albeit an increasingly noisy estimate for shorter records.

Before we could form $\langle u^2 \rangle$ and $\langle v^2 \rangle$ we had to obtain the model output corresponding to the current meter locations. We compared results for two methods: using the closest horizontal grid point, and using a weighted average of all the points within a disk surrounding the mooring site. The weighted average was

of the form:

$$u_p(x, y, z_i, t) = \frac{\sum_i u_p(x_i, y_i, z_i, t) w_i}{\sum_i w_i} \quad (2)$$

where $w_i = \exp(-((x - x_i)^2 + (y - y_i)^2)/a^2)$, (x, y) are coordinates of the mooring site, $a = 9$ km, and (x_i, y_i, z_i) are the model grid points within a disk of radius $a\sqrt{3}$. Grid points below the sea floor were omitted (*i.e.* w_i set to zero). The choice between the two methods was a compromise. The closest grid point method involved less averaging of velocity, and therefore higher energy levels for most (85%) points, and was closer to the point-like measurements of the current meter records. However, the averaging method had the advantage that the resulting velocity time series were less noisy. These two methods are illustrated in Figure 6 and Figure 7. Figure 6 shows the two time series from the surface grid point of the HYCOM DA model run at the JADE mooring in the Banda Sea. This mooring represents the worst case in that the absolute value of the difference in mean squared velocities between the two methods was the largest for this location and depth ($0.44 \text{ m}^2/\text{s}^2$). Figure 7 shows the cumulative distribution of difference in $\langle u^2 \rangle$ computed from the closest and averaged velocity time series, divided by $\langle u^2 \rangle$ from the closest. Only about 15% of the points had more energy in the averaged time series, and about 88% of points had energy from the two methods that agreed to within 25%. We present the results below only for the weighted average case. Using the nearest grid point gave visually similar plots, and reduced the correlations between the two HYCOM model runs and CMA $\langle u^2 \rangle$ in Table 2 below but only by 0.01 in each case.

Figure 6

Figure 7

For the POP model, daily averaged currents were linearly interpolated to the latitude and longitude of the mooring sites, while the model was run. Model points below the sea floor had zero velocity (in keeping with no-slip boundary conditions), and this may have reduced the interpolated flow speed near topography. As mentioned earlier, the POP model was compared with CMA 1.0 which didn't contain all the moorings of CMA 2.0.

For all models the $\langle u^2 \rangle$ and $\langle v^2 \rangle$ values were interpolated linearly between vertical grid points to the current meter depths. Current meter records below the deepest mode grid point were omitted. For the HYCOM model data were only available to the deepest Levitus grid point of 5500 m. For OCCAM and POP the vertical axes extended to 6366 m and 5875 m respectively.

4 Results

We start with simple log-log scatter plots of mean squared velocities from Eq. 1 with the CMA values on the x -axis, and model on the y -axis. Figures 8 and 9 show the results for $\langle u^2 \rangle$ and $\langle v^2 \rangle$. For perfect agreement between models and observations all the points would fall on the thin 45° line. For all models there was considerable scatter about this line, much larger than the measurement error. None-the-less there was clearly some skill for each of the models. The Pearson correlation coefficient for $\langle u^2 \rangle$ and $\langle v^2 \rangle$ between models and the CMA are presented in Table 2. The scatter plots of $\langle u^2 \rangle$ and $\langle v^2 \rangle$ were visually similar but the corresponding correlation coefficients revealed that the zonal component was generally better than the meridional component of

TKE. This was especially apparent for the OCCAM model, which had the most anisotropic grid. Restricting the comparison to CMA records of at least 180 days of accumulated data led to similar, though generally slightly higher, correlations, see Table 3. Similarly, using four years of model data improved the correlations slightly, see lower part of Table 3. With at least hundreds (if not thousands) of degrees of freedom, even tiny correlations would be highly statistically significant.

Figure 8

Figure 9

Does the large scatter imply we cannot trust any of the models to be accurate at arbitrarily chosen locations, or does it arise from statistical uncertainty? The large scatter results from a combination of measurement errors for the current meter records (see Section 2.1), statistical uncertainty in both the current meter records and model time series, and model errors. We need to estimate how much of the scatter resulted from statistical uncertainty. Fig. 4 alerts us to the possibility of very long records being necessary to produce narrow confidence bounds on EKE, and TKE is dominated by EKE in most places. For Gaussian velocity statistics one could in principle infer the confidence bounds on the ratio of sample variances using the F-distribution, but that approach is complicated here by the large number of records with different degrees of freedom. Instead we approached this numerically as follows. Consider the case of a perfect model and perfect current meter records, hampered only by statistical uncertainty with degrees of freedom chosen to correspond to the actual records we had to analyze. To simulate this situation we created two sets of Gaussian random pseudo time series for each of the 5339 current meter

records, one set $G_1 = \{g_{1,1}, g_{1,2}, \dots, g_{1,N}\}$, $N = 5339$ to represent perfectly accurate CMA records and the other set $G_2 = \{g_{2,1}, g_{2,2}, \dots, g_{2,N}\}$, $N = 5339$ to represent a corresponding perfect OGCM. The pseudo CMA records and the pseudo OGCM are considered to measure and simulate a statistically stationary ocean. The difficulty was in choosing appropriate degrees of freedom for each pair of the 5339 pseudo records. Assuming that the real ocean 5-day mean time series had about a 15-day decorrelation time, we chose appropriate corresponding degrees of freedom pseudo times. That is, G_2 all had 26 independent Gaussian random variables, and each g_1 pseudo record had 1/3 the number of 5-day averages in the corresponding real CMA record. Not all of the CMA records were independent, since current meters are deployed on moorings with typically several current meters on a single mooring, which reduces the total number of degrees of freedom, and therefore increases the scatter. To account for this we tried four assumptions for the Pearson correlation coefficient for pseudo time series $u(t, z)$ on the same mooring: $r = 0, 1/2, 1, c$, where c was chosen randomly from the interval $[0, 1]$ for each mooring. For each pair of pseudo records the two normally distributed time series were scaled by the same amplitude, σ_n , implying identical population variance σ_n^2 but not necessarily identical sample variances $s_{1,n}$ and $s_{2,n}$. The amplitudes σ_n were chosen from the interval $[0, \exp(z/H)]$ m/s where z was the depth of the instrument, $H = 900$ m was the decay scale chosen so that roughly 4100 points were above the threshold of 2×10^{-4} m²/s². The scatter plots of the two sets of sample variances S_1 vs. S_2 are shown in Fig. 10. There was little sensitivity to the choice of r .

Fig. 10

The scatter in the simulated records Fig. 10 is clearly smaller than for the real

CMA and OGCM records Figs. 8 and 9. While the CMA measurement error was not negligible, it was likely in the range $0\% < E < 50\%$, see discussion in Section 2. Because there are many points in the scatter plots of Figs. 8 and 9 that are much more than a factor of two beyond the statistical scatter of Fig. 10, we must conclude that modeling errors contributed significantly to the discrepancy between OGCM and CMA mean squared velocities.

How were the modeling errors distributed? Often we can live with errors if they are random, so what we're especially concerned about is identifying biases. For free-running HYCOM and especially OCCAM, more of the points were below the line than above indicating that the model was often too weak. And there were perhaps more points below the line at weaker levels. These biases were difficult to see in these scatter plots, so we turn to another measure.

The model biases were more clear when we examined the following statistic, akin to that used by Scott et al. (2008) to compare zonal and meridional velocity variances:

$$D \equiv \frac{\text{TKE}_{\text{CMA}} - \text{TKE}_{\text{MOD}}}{\text{TKE}_{\text{CMA}} + \text{TKE}_{\text{MOD}}}. \quad (3)$$

where the subscripts CMA and MOD refer to the Current Meter Archive and OGCM values respectively. This normalization maps the discrepancy to the interval $[-1, 1]$. We prefer D to considering just the numerator of D (the unnormalized discrepancy) because it gives a measure of the magnitude of the discrepancy relative to the local energy level. We prefer not to normalize the discrepancy by just the CMA TKE because the latter also has some error, and when the CMA TKE happens to be very small, the result becomes very

large. That is, the quantity

$$\frac{\text{TKE}_{\text{CMA}} - \text{TKE}_{\text{MOD}}}{\text{TKE}_{\text{CMA}}}$$

maps the discrepancy to the interval $(-\infty, \infty)$, and can exaggerate the importance of smaller errors. D is defined for each record in our CMA, so we can consider distributions for each model. Perfect agreement between models and observations would result in D distributed like the Dirac- δ function. So the more narrowly the histogram of D is distributed about the Dirac- δ function, the better the model agrees with observations. We are especially concerned with the skewness of the distribution. For if D were distributed symmetrically about zero, there would be no bias, while a distribution skewed to the right indicates the model was too weak relative to observations. We found the histograms of D to be a sensitive diagnostic to reveal the model biases.

Figure 11 presents the histograms of D for all models, and the mean D is quoted in Tables 2 and 3. In these plots, we have eliminated points with $\text{TKE}_{\text{CMA}} < 2 \times 10^{-4} \text{ m}^2/\text{s}^2$ because the current meter records are unreliable at these weak current speeds, see Section 2. The distributions have been stratified by depth, with the darker colors indicating deeper levels. The upper ocean was clearly much better simulated than the abyssal ocean. The HYCOM run with data assimilation was actually slightly too strong in the upper ocean. In contrast to the upper ocean, at depth the models were generally biased toward being too weak relative to the CMA. Clearly none of the models were free of any bias, but the POP model and the HYCOM model run with data assimilation stand out as having a much less obvious bias. The bias in POP likely represents the few outlier points that are much weaker than observations. The mean D in Table 2 suggested a similar conclusion as Figures 11; the smaller

mean D of the HYCOM data assimilative model, and POP, are consistent with less model bias. Comparing Tables 2 and 3, we found very similar mean D , suggesting mean D was not strongly influenced by statistical uncertainty.

As a complementary way to present the bias in vertical structure of horizontal TKE, we also plotted the average of TKE from all locations within depth bins, see Fig. 12. The depth bins were chosen to give roughly similar numbers of observations within each bin, see thin line with circles. The circles indicate the centre of the depth bin. The resulting vertical structure of the currents should not be over interpreted since it represents the mean from different latitudes and longitudes for different depth bins. In forming the average, we first averaged over all records on each mooring that fell within the same depth bin, and then averaged over moorings. The number of observations refers to the number of moorings (not the total number of records). We could then resample these different mooring averages, which were statistically independent, and form the 95% confidence limits using bootstrapping. All models agreed reasonably well with the CMA observations in the upper 300 m, but at greater depths the OCCAM and free-running HYCOM models were systematically too weak. Below about 4500 m the number of available observations (see thin line with circles) dropped off quickly and the comparison is much less significant. In Fig. 13 we plotted the same information, but with a log scale so that the upper ocean was more clear. Furthermore, we divided model TKE in each bin by the corresponding value in the CMA. All four models capture the averaged kinetic energy to within factors of 2 to 3 or better. The OCCAM and free-running HYCOM models were around 1.5 times too weak at 300 m depth, about 2 times too weak at 3000 m, and about 3 times too weak at 4000 m. However, the POP model had mean TKE profile that agreed with observations to within

statistical uncertainty for all depths below 1000 m. HYCOM with data assimilation agreed almost as well as POP with observations below 1000 m, and perhaps slightly better than POP above 1000 m.

5 Summary and Discussion

A large collection of moored current meter records from a recently assembled current meter archive (CMA) was used to assess the ability of four eddying OGCMs to simulate the time-averaged total kinetic energy (TKE) throughout the water column. While moored current meters provide the best available Eulerian data for this purpose, they do have inherent limitations, as discussed in Section 2, and Figs. 1 and 2. Measurement errors can be as large as 45% for TKE for acoustic current meters under challenging conditions of low current speeds and low suspended particle density at depth. However, generally measurement errors were much smaller. Mooring blow-over was also a concern, since about 10% of the records had 25% of their measurements from depths more than 70 m below their nominal deployment depth. This results in strong currents being underestimated by subsurface mooring current meters.

There was found to be considerable scatter in the comparison between TKE in the eddying OGCMs and the CMA, even larger than the statistical uncertainty, implying that models cannot be trusted to give reliable TKE at arbitrary locations. This is similar to the findings of Arbic et al. (2009), who compared models and an earlier version of our CMA at the bottom (rather than throughout the water column, as done here), and of Penduff et al. (2006), who looked at the Atlantic CLIPPER model and WOCE current meters. However, there was highly significant correlation between the simulated and ob-

served TKE, suggesting some model skill. We were mostly concerned with identifying any model bias in the vertical distribution of TKE. Scatter plots were only slightly useful for revealing model biases. A much more sensitive diagnostic was the distribution of discrepancy between simulated and observed TKE, scaled by their sum, see Eq. 3. All models had some bias, and the bias tended to be toward too weak simulated currents at greater depths. However, the HYCOM model with data assimilation, and the (free running) POP model had much less obvious bias. An average taken over all current meters, plotted versus binned depths, also showed that these two simulations yielded the least bias over the full water column.

Penduff et al. (2006) suggest that horizontal resolution may be the most important factor limiting the CLIPPER OGCM in generating realistic EKE, since the first Rossby radius of deformation was best resolved at low-latitudes, and the model-data comparison was most favorable there. Space did not allow us to investigate this systematically here. A preliminary analysis suggests the situation is complicated. Figure 14 shows the distribution of D for low, mid and high latitudes. While Penduff et al.'s suggestion appears to be consistent with the results for the HYCOM model with data assimilation, one must keep in mind that no data assimilation were performed north of 47°N . The POP model was actually bimodal, revealing both strong and weak biases at high latitudes. This issue clearly deserves more attention.

The reasons why POP appears to do better than the other non-assimilating models is unclear. The sensitivity of model results to the model formulation and associated numerical treatment is not fully documented (Griffies et al., 2000, 2009) and several studies show that the modeled circulation remains quite sensitive to the choices made for subgrid scale parameterizations (Chas-

signet and Marshall, 2008; Hecht and Smith, 2008; Hecht et al., 2008). While bathymetry is an important consideration in general, the differences in its numerical treatment between the three models appear minor; like POP, OCCAM also used partial bottom cells with a z-grid as well (though OCCAM was also hindered by decreasing horizontal resolution at higher latitudes by an anisotropic grid), and the generalized vertical coordinate of HYCOM, while different in approach, leads to comparable flexibility. Important differences in the forcing were the monthly forcing of POP, which did not include the higher frequency forcing that tends to excite high frequency vertical motions that increase numerical diffusion. Dramatic examples of the dependence of vertical velocity on forcing frequency are shown by Klein (2008, Plates 4 and 5). One might speculate that this diffusion damps the mesoscale kinetic energy, though the net result would be somewhat counter intuitive (less high-frequency forcing leads ultimately to more energy?). The biggest difference, besides numerics and physical parameterization choices, appears to be in the length of the spinup (48 years for POP) therefore suggesting that several decades may be required to spin up the TKE in the deep ocean.

Acknowledgements

We benefitted from conversations with Carl Wunsch. John Toole pointed us to an interesting study of Bill Schmitz that motivates a nice follow-up study, but unfortunately we didn't have space to pursue that herein. We are grateful to Carl Wunsch for contributing generously to our CMA, and to the Buoy Group at OSU for maintaining the OSU DWA.

RBS thanks the NOC for hosting an extended visit. The authors acknowledge

the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing High-Performance Computing (HPC) resources that have contributed to the research results reported within this paper, <http://www.tacc.utexas.edu>. RBS was supported by National Science Foundation (NSF) grant OCE-0526412, a grant from KAUST, and the National Oceanography Centre, Southampton (NOCS). BKA and AV were supported by NSF grant OCE-0623159 and a Jackson School of Geosciences Development Grant. MM was supported by the DOE Office of Science Climate Change Prediction Program. This is UTIG contribution #.

This paper is dedicated to the memory of the great physical oceanographer and journal editor Peter Killworth.

References

- Adcroft, A., Hill, C., Marshall, J., 1997. Representation of topography by shaved cells in a height coordinate ocean model. *Monthly Weather Review* 125, 2293–2315.
- Arbic, B. K., Shriver, J. F., Hogan, P. J., Hurlburt, H. E., McClean, J. L., Metzger, E. J., Scott, R. B., Sen, A., Smedstad, O. M., Wallcraft, A. J., 2009. Estimates of bottom flows and bottom boundary layer dissipation of the oceanic general circulation from global high-resolution models. *J. Geophys. Res.* 114, C02024.
- Barnier, B., Madec, G., Penduff, T., Molines, J.-M., Treguier, A.-M., Le Sommer, J., Beckmann, A., Biastoch, A., Bning, C., Dengg, J., Derval, C., Durand, E., Gulev, S., Remy, E., Talandier, C., Theetten, S., Maltrud, M., McClean, J., De Cuevas, B., 2006. Impact of partial steps and momentum advection schemes in a global ocean circulation model at eddy-permitting resolution. *Ocean Dynamics* 56, 543–567.
- Bleck, R., 2002. An oceanic general circulation model framed in hybrid isopycnic cartesian coordinates. *Ocean Modelling* 4, 55–88.
- Carnes, M. R., 2003. Description and evaluation of GDEM-v 3.0. Naval Oceanographic Office Technical Note.
- Chassignet, E., Hurlburt, H., Smedstad, O., Halliwell, G., Hogan, P., Wallcraft, A., Bleck, R., 2006. Ocean prediction with the hybrid coordinate ocean model (HYCOM). In: Chassignet, E., Verron, J. (Eds.), *Ocean Weather Forecasting*. Springer, pp. 413–426.
- Chassignet, E. P., Hurlburt, H. E., Metzger, E. J., Smedstad, O. M., Cummings, J. A., Halliwell, G. R., Bleck, R., Baraille, R., Wallcraft, A. J., Lozano, C., Tolman, H. L., Srinivasan, A., Hankin, S., Cornillon, P., Weis-

- berg, R., Barth, A., He, R., Werner, F., Wilkin, J., JUN 2009. US GODAE Global Ocean Prediction with the HYbrid Coordinate Ocean Model (HYCOM). *Oceanography* 22 (2, Sp. Iss. SI), 64–75.
- Chassignet, E. P., Marshall, D. P., 2008. Gulf stream separation in numerical ocean models. In: Hecht, Hasumi (Eds.), *Ocean Modeling in an eddying regime*. AGU Monograph Series. AGU, pp. 39–62.
- Chassignet, E. P., Smith, L. T., Halliwell, G. R., Bleck, R., 2003. North Atlantic simulations with the HYbrid Coordinate Ocean Model (HYCOM): Impact of the vertical coordinate choice, reference pressure, and thermobaricity. *J. Phys. Oceanogr.* 33, 2504–2526.
- Colin de Verdiere, A., Mercier, H., Arhan, M., 1989. Mesoscale variability from the western to the eastern Atlantic along 48°N. *J. Phys. Oceanogr.* 19, 1149–1170.
- Cummings, J. A., 2005. Operational multivariate ocean data assimilation. *Quart. J. R. Met. Soc.* 131, 3583–3604.
- Daley, R., 1991. *Atmospheric Data Analysis*. Cambridge Atmospheric and Space Science Series. Cambridge University Press.
- ETOPO2, 2006. 2-minute gridded global relief data (ETOPO2v2). U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Geophysical Data Center.
- Gilboy, T., Dickey, T., Sigurdson, D., Yu, X., Manov, D., 2000. An intercomparison of current measurements using a vector measuring current meter, an acoustic Doppler current profiler, and a recently developed acoustic current meter. *J. Atmos. Ocean Tech.* 17 (4), 561–574.
- Gille, S. T., Llewellyn-Smith, S. G., 2000. Velocity probability density functions from altimetry. *J. Phys. Oceanogr.* 30, 125–136.
- Gould, W. J., Schmitz, W. M., Wunsch, C., 1974. Preliminary field results for

- a mid ocean dynamics experiment (MODE-O). *Deep-Sea Res.* 21, 911–932.
- Gouretski, V. V., Jancke, K., 1996. A new hydrographic data set for the South Pacific: Synthesis of WOCE and historical data. WHP SAC Tech. Rep. 2, WOCE Rep. 143/96, 110 pp.
- Griffies, S. M., Biastoch, A., Boening, C., Bryan, F., Danabasoglu, G., Chassignet, E., England, M., Gerdes, R., Haak, H., Hallberg, R., Hazeleger, W., Jungclaus, J., Large, W., Madec, G., Pirani, A., Samuels, B., Scheinert, M., Gupta, A., Severijns, C., Simmons, H., Treguier, A.-M., Winton, M., Yeager, S., Yin, J., 2009. Coordinated ocean-ice reference experiments (COREs). *Ocean Modelling* 26, 1–46.
- Griffies, S. M., Bning, C., Bryan, F., Chassignet, E., Gerdes, R., Hasumi, H., Hirst, A., Treguier, A.-M., Webb, D., 2000. Developments in ocean climate modelling. *Ocean Modelling* 2, 123–192.
- Halpern, D., Weller, R. A., Briscoe, M. G., Davis, R. E., McCullough, J. R., 1981. Intercomparison tests of moored current measurements in the upper ocean. *J. Geophys. Res.* 86 (NC1), 419–428.
- Hecht, M. W., Hasumi, H. (Eds.), 2008. Ocean modeling in an eddying regime. AGU Monograph Series. AGU.
- Hecht, M. W., Hunke, E., Maltrud, M. E., Petersen, M. R., Wingate, B. A., 2008. Lateral mixing in the eddying regime and a new broad-ranging formulation. In: Hecht, Hasumi (Eds.), *Ocean Modeling in an eddying regime*. AGU Monograph Series. AGU, pp. –.
- Hecht, M. W., Smith, R. D., 2008. Towards a physical understanding of the North Atlantic: A review of model studies in an eddying regime. In: Hecht, Hasumi (Eds.), *Ocean Modeling in an eddying regime*. AGU Monograph Series. AGU, http://public.lanl.gov/mhecht/preprints/NA_pp.pdf.
- Hogg, N. G., 1991. Mooring motion corrections revisited. *J. Atmos. Ocean*

- Tech. 8, 289–295.
- Hogg, N. G., 1992. On the transport of the Gulf Stream between Cape Hatteras and the Grand Banks. *Deep-Sea Res.* 39A, 1231–1246.
- Hogg, N. G., Frye, D. E., 2007. Performance of a new generation of acoustic current meters. *J. Phys. Oceanogr.* 37 (2), 148–161.
- Hurlburt, H. E., Hogan, P. J., 2000. Impact of $1/8^\circ$ to $1/64^\circ$ resolution on Gulf Stream model-data comparisons in basin-scale subtropical Atlantic Ocean models. *Dyn. Atmos. Oceans* 32, 283–329.
- Johns, W. E., Shay, T. J., Bane, J. M., Watts, D. R., 1995. Gulf stream structure, transport, and recirculation near 68°W . *J. Geophys. Res.* 100, 817–838.
- Kelly, K. A., Thompson, L., Cheng, W., Metzger, E. J., 2007. Evaluation of HYCOM in the Kuroshio Extension region using new metrics. *J. Geophys. Res.* 112, C01004.
- Klein, P., 2008. High-frequency winds and eddy-resolving models. In: Hecht, Hasumi (Eds.), *Ocean Modeling in an eddying regime*. AGU Monograph Series. AGU, pp. 83–100.
- Large, W. G., McWilliams, J. C., Doney, S. C., 1994. Oceanic vertical mixing—a review and a model with a nonlocal boundary layer parameterization. *Rev. Geophys.* 32, 363–403.
- Large, W. G., Yeager, S. G., 2004. Diurnal to decadal global forcing for ocean and sea-ice models: the datasets and flux climatologies. NCAR Technical Note TN-460+STR, National Center for Atmospheric Research.
- Lee, M.-M., Nurser, A. J. G., Coward, A. C., De Cuevas, B. A., 2007. Eddy advective and diffusive transports of heat and salt in the Southern Ocean. *J. Phys. Oceanogr.* 37, 1376–1393.
- LeTraon, P.-Y., Nadal, F., Ducet, N., 1998. An improved mapping method of

- multisatellite altimeter data. *J. Atmos. Oceanic Technol.* 15, 522–534.
- Locarnini, R. A., Mishonov, A. V., Antonov, J. I., Boyer, T. P., Garcia, H. E., 2006. World ocean atlas 2005, volume 1: Temperature. S. Levitus, Ed. NOAA Atlas NESDIS 61, U.S. Government Printing Office, Washington, D.C., 182 pp.
- Maltrud, M., Bryan, F., Hecht, M., Hunke, E., Ivanova, D., McClean, J., Peacock, S., 2008. Global ocean modeling in the eddying regime using POP. *CLIVAR Exchanges* 13, 5–8.
- Maltrud, M. E., McClean, J. L., 2005. An eddy resolving global 1/10 degree ocean simulation. *Ocean Modelling* 8, 31–54.
- Meinen, C. S., Luther, D. S., 2002. Mooring motion when pressure sensors fail: a method employing inverted echo sounders. *J. Atmos. Ocean Tech.* 19, 1451–1460.
- MODE Group, 1978. The mid-ocean dynamics experiment. *Deep-Sea Res.* 25, 859–910.
- Paiva, A. M., Hargrove, J. T., Chassignet, E. P., Bleck, R., 1999. Turbulent behavior of a fine mesh ($1/12^\circ$) numerical simulation of the North Atlantic. *J. Mar. Sys.* 21, 307–320.
- Penduff, T., Barnier, B., Molines, J.-M., Madec, G., 2006. On the use of current meter data to assess the realism of ocean model simulations. *Ocean Modelling* 11, 399–416.
- Plueddemann, A., Farrar, J. T., 2006. Observations and models of the energy flux from the wind to mixed-layer inertial currents. *Deep-Sea Res. II* 53, 5–30.
- Plueddemann, A. J., Weller, R. A., Stramska, M., Dickey, T. D., Marta, J., 1995. Vertical structure of the upper ocean during the marine light-mixed layers experiment. *J. Geophys. Res.* 100 (C4), 6605–6619.

- Scott, R. B., Arbic, B. K., Holland, C. L., Sen, A., Qiu, B., 2008. Zonal versus meridional velocity variance in satellite observations and realistic and idealized ocean circulation models. *Ocean Modelling* 23, 102–112.
- Sen, A., Scott, R. B., Arbic, B. K., 2008. Global energy dissipation rate of deep-ocean low-frequency flows by quadratic bottom boundary layer drag: Computations from current-meter data. *Geophys. Res. Lett.* 35, L09606 doi:10.1029/2008GL033407.
- Smith, R. D., Maltrud, M. E., Bryan, F. O., Hecht, M. W., 2000. Numerical simulation of the North Atlantic ocean at 1/10 degrees. *J. Phys. Oceanogr.* 30, 1532–1561.
- Smith, W. H. F., Sandwell, D. T., 1997. Global sea floor topography from satellite altimetry and ship depth soundings. *Science* 277, 1956–1962.
- Srinivasan, A., Garraffo, Z., Iskandarani, M., 2009. Abyssal circulation in the Indian Ocean from a 1/12° resolution global hindcast. *Deep-Sea Res. I*, in press.
- Treguier, A. M., 2006. Models of the ocean: Which ocean? In: Chassignet, E., Verron, J. (Eds.), *Ocean Weather Forecasting*. Springer, Ch. 3, pp. 75–108.
- Weller, R. A., Davis, R. E., 1980. A vector measuring current-meter. *Deep-Sea Res. A* 27 (7), 565–582.
- Wunsch, C., 1997. The vertical partition of oceanic horizontal kinetic energy and the spectrum of global variability. *J. Phys. Oceanogr.* 27, 1770–1794.

6 Tables

Table 1

Models analyzed in this study. FR = free running (no data assimilation); DA = data assimilation; B = Boussinesq; NB = non-Boussinesq ; PE = primitive eqns.

Quantity	OCCAM	POP	HYCOM FR	HYCOM DA
Equations	B, PE	B, PE	NB, PE	NB, PE, DA
Vert. coord.	Z	Z	ρ, Z, σ	ρ, Z, σ
Vert. res.	66 levels	42 levels	32 layers	32 layers
Horiz. res.	1/12° lon- lat	1/10°lon	1/12° lon	1/12° lon
Output	5-day mean	daily mean	daily snapshot	daily snap- shot

Table 2

Summary of comparisons between models and current meter records. Slightly higher correlations were found using 5-day low-pass Butterworth filtered CMA records than for the 5-day averages presented here. Criteria imposed on current meters: ≥ 90 days of good data (*i.e.* 18 5-day means), $\text{TKE} > 2 \text{ cm}^2/\text{s}^2$. Model results were averaged over one year. Correlations for the sets of simulated gaussian random variables G_1 and G_2 were between 0.84 and 0.88 with similar number of points, typically 4150 to 4250 (see discussion surrounding Fig. 10) . D is defined in Eq. 3

Quantity	OCCAM	POP	HYCOM FR	HYCOM DA
Correlation $\langle u^2 \rangle$	0.68	0.72	0.59	0.77
Correlation $\langle v^2 \rangle$	0.50	0.61	0.66	0.72
Average D	0.26	0.08	0.17	0.01
Number of points	4211	4211	4211	4211

Table 3

As in Table 2, but now for current meter records at least 180 days long, and different model durations.

Quantity	OCCAM	POP	HYCOM FR	HYCOM DA
Model duration	1 year	1 year	1 year	1 year
Correlation $\langle u^2 \rangle$	0.68	0.76	0.58	0.80
Correlation $\langle v^2 \rangle$	0.48	0.67	0.70	0.76
Average D	0.27	0.09	0.18	0.01
Number of points	3510	3510	3510	3510
Model duration	4 years	4 years		
Correlation $\langle u^2 \rangle$	0.68	0.77		
Correlation $\langle v^2 \rangle$	0.52	0.70		
Average D	0.26	0.10		
Number of points	3692	3692		

7 List of Figures

- (1) Figure 1. Distribution of statistics for 1559 current meters with co-located pressure sensors. Left: Histogram of the interquartile range of pressure from each pressure record. Right: Histogram of the minimum to maximum range of pressure (revealing maximum vertical displacement occurring in each pressure record).
- (2) Figure 2. Top: Time series of pressure for an instrument in the East Australia Current, where the seafloor depth was 4435 m. Nominal instrument depth was 600 m. Note the strong blow-over event around 3000 hrs from the start of the record. Bottom: Recorded current speed. Current speeds and TKE at the nominal depth are likely underestimated during strong blow-over events.
- (3) Figure 3. Locations of the 1361 current meter mooring sites used in this study. The 1213 black X's indicate the moorings with a record at least 180 days long, while the red o's are the moorings with records between 90 and 180 days. There were a total of 5339 unique depth bins from current meter deployments between Sept. 1973 and Feb. 2005, with at least (accumulated) 90 days of good data. Multiple deployments at the same mooring site and depth are counted as one record or "unique depth bin", despite the temporal gap. The sea-floor depth for 76% of moorings was more than 2000 m.
- (4) Figure 4. Ratio of the upper confidence bound to lower confidence bound vs. degrees of freedom for various levels of confidence. Current meter records had at least 90 days duration, so between 3.5 and 8 degrees of freedom.
- (5) Figure 5. Time series of 5-snapshot averages of full zonal velocity (black)

and geostrophic velocity (blue) for the free-running HYCOM model. The upper panel is for the surface, and the lower panels are for 200 m, 1000 m, and 3000 m respectively. The surface velocities were clearly influenced by wind forcing, while the intermediate depths showed excellent agreement with geostrophic balance. It remains unexplained why the deeper velocities are not as well balanced.

- (6) Figure 6. Time series of 5-snapshot averaged zonal velocity u_p from the HYCOM DA model run. The solid line used a weighted average of points within a disk of radius about 16 km, see Eq. 2. The dashed line is the closest grid point time series. This location represents the most extreme difference in mean squared values $\langle u^2 \rangle$ computed from the two methods.
- (7) Figure 7. Cumulative distribution of the difference in $\langle u^2 \rangle$ from the HYCOM DA run calculated using the weighted averaged (subscript 'oi') and nearest grid point (subscript 'nearest') methods, divided by the latter.
- (8) Figure 8. Scatter plot of $\langle u^2 \rangle$ from the CMA, and the four OGCMs. The thin vertical dashed line is the threshold above which the correlations quoted in Table 2 were computed.
- (9) Figure 9. As in Fig. 8 but for meridional velocity $\langle v^2 \rangle$.
- (10) Figure 10. Scatterplot of simulated $\langle S_1^2 \rangle$ vs. $\langle S_2^2 \rangle$, the sample variances of pseudo random pseudo CMA records and corresponding pseudo OGCM time series, see Section 4. Subplots are for different values of r , the correlation between pseudo time series at different levels on the same mooring. The scatter here is meant to depict the effect of statistical uncertainty and is the bench mark of perfect CMA records and perfect OGCMs, to which Figs. 8 and 9 should be compared. The thin vertical dashed line is the threshold above which the correlations quoted in Table 2 were computed.
- (11) Figure 11. Distribution of D for the four OGCMs for different depth

- ranges: 0 to 750 m (cyan); 750 to 2000 m (blue); below 2000 m (black).
- (12) Figure 12. TKE averaged over all current meter locations vs. depth for all models. Thin line with circles is the number of observations in each depth bin $/10^5$ (so there were just over 300 observations in the depth bin centred just above 4000 m depth). There were only 56 observations in the lowest depth bin.
- (13) Figure 13. As in Fig. 12 but now normalized by the CMA TKE, a log vertical axis, and the number of observations in each depth bin is in thousands. Using the closest grid point time series rather than the averaging method shifted the lines less than about 0.05 to the right. The thin dashed lines represent the 95% confidence bounds obtained using bootstrapping.
- (14) Figure 14. As in Fig. 11 but the results are stratified by latitude (as opposed to depth): 0 to 20° (cyan); 20 to 45° (blue); $> 45^\circ$ (black).

8 Figures

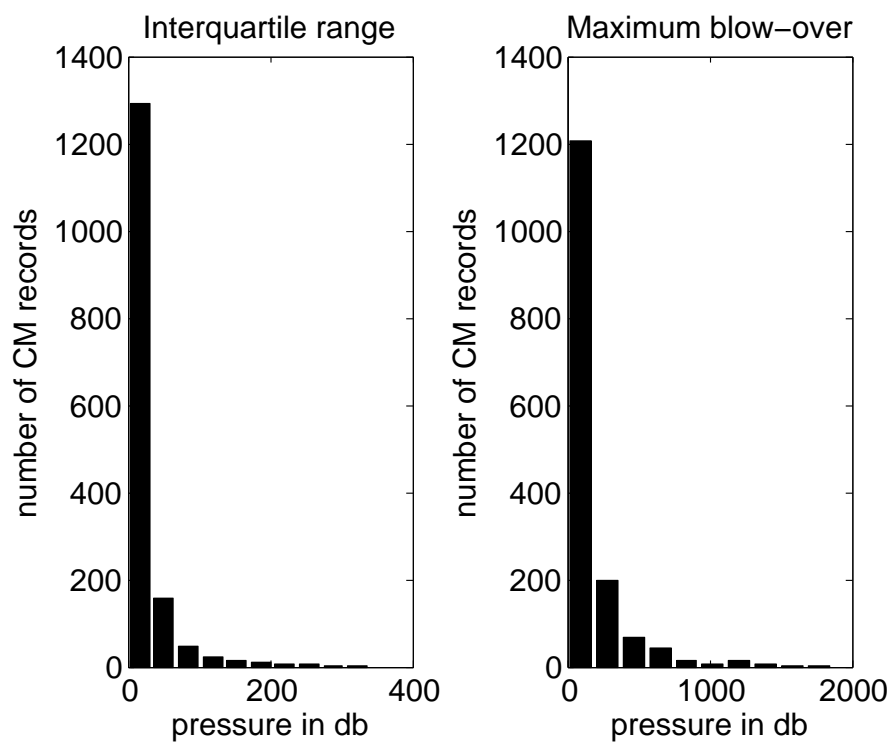


Fig. 1. Distribution of statistics for 1559 current meters with co-located pressure sensors. Left: Histogram of the interquartile range of pressure from each pressure record. Right: Histogram of the minimum to maximum range of pressure (revealing maximum vertical displacement occurring in each pressure record).

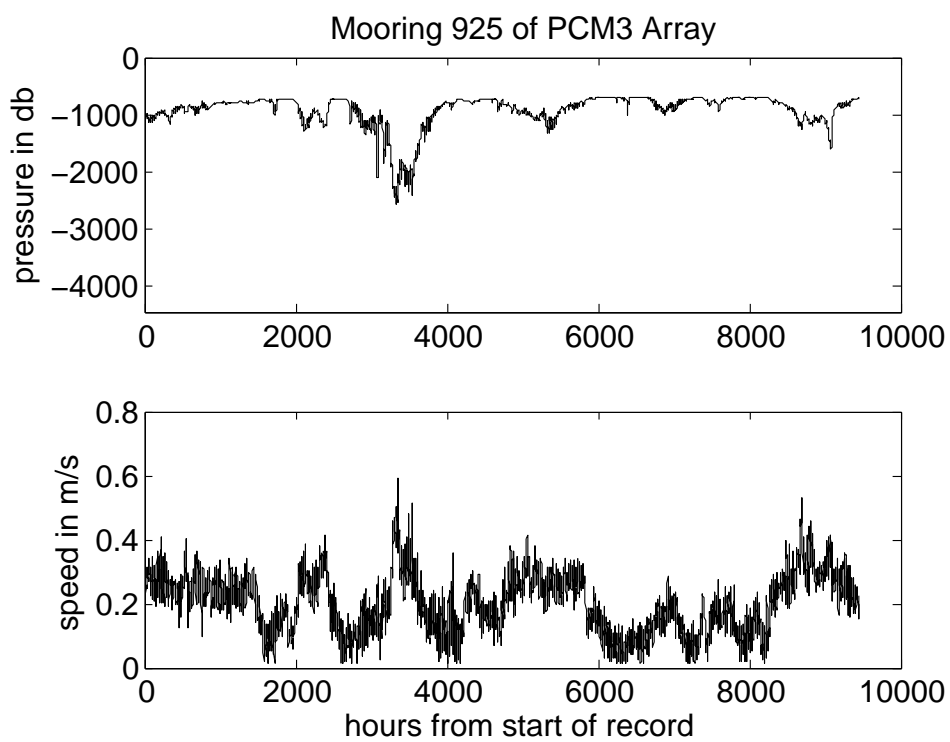


Fig. 2. Top: Time series of pressure for an instrument in the East Australia Current, where the seafloor depth was 4435 m. Nominal instrument depth was 600 m. Note the strong blow-over event around 3000 hrs from the start of the record. Bottom: Recorded current speed. Current speeds and TKE at the nominal depth are likely underestimated during strong blow-over events.

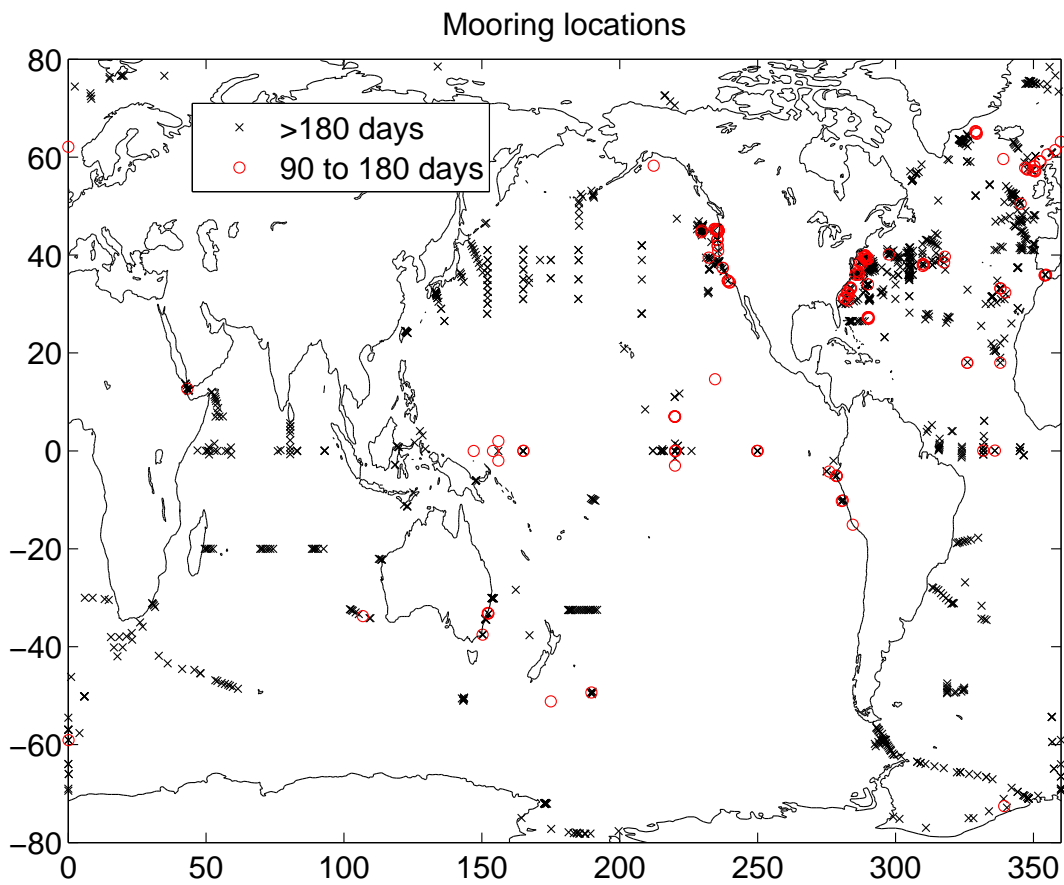


Fig. 3. Locations of the 1361 current meter mooring sites used in this study. The 1213 black X's indicate the moorings with a record at least 180 days long, while the red o's are the moorings with records between 90 and 180 days. There were a total of 5339 unique depth bins from current meter deployments between Sept. 1973 and Feb. 2005, with at least (accumulated) 90 days of good data. Multiple deployments at the same mooring site and depth are counted as one record or "unique depth bin", despite the temporal gap. The sea-floor depth for 76% of moorings was more than 2000 m.

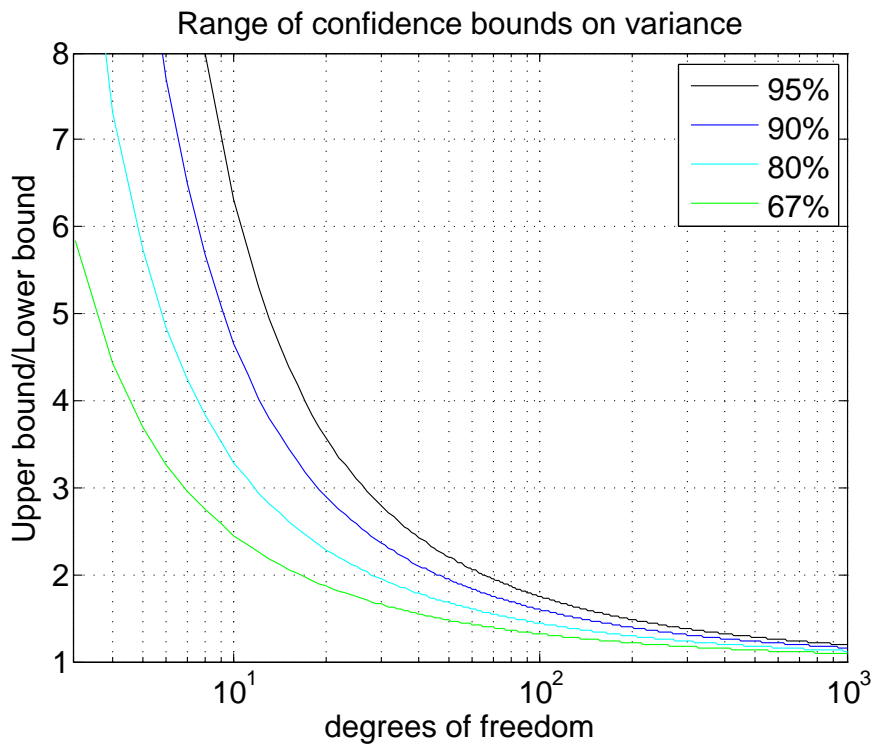


Fig. 4. Ratio of the upper confidence bound to lower confidence bound vs. degrees of freedom for various levels of confidence. Current meter records had at least 90 days duration, so between 3.5 and 8 degrees of freedom.

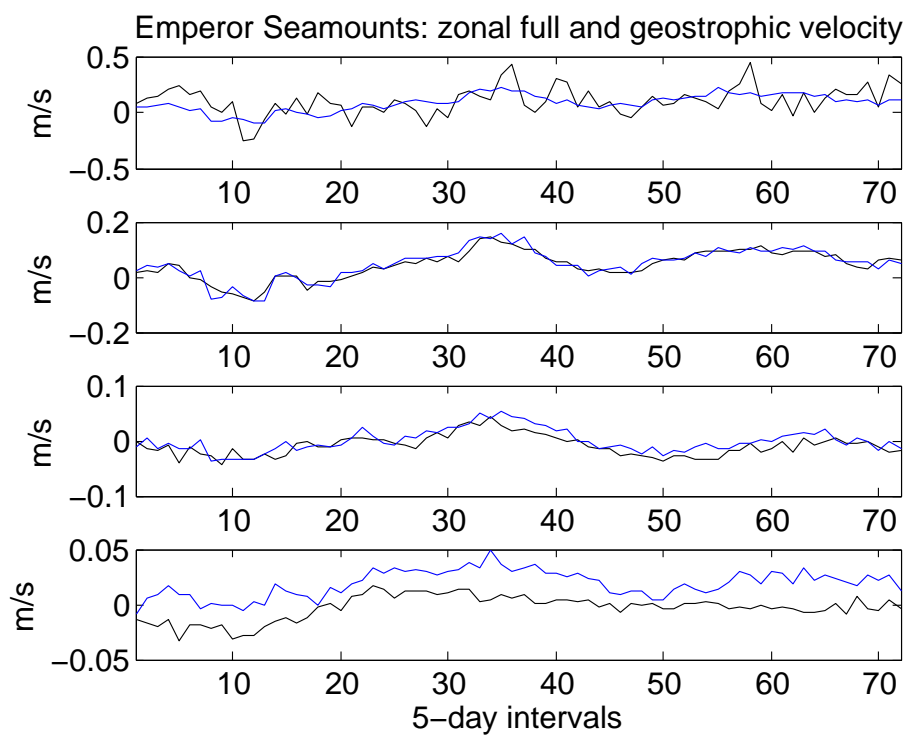


Fig. 5. Time series of 5-snapshot averages of full zonal velocity (black) and geostrophic velocity (blue) for the free-running HYCOM model. The upper panel is for the surface, and the lower panels are for 200 m, 1000 m, and 3000 m respectively. The surface velocities were clearly influenced by wind forcing, while the intermediate depths showed excellent agreement with geostrophic balance. It remains unexplained why the deeper velocities are not as well balanced.

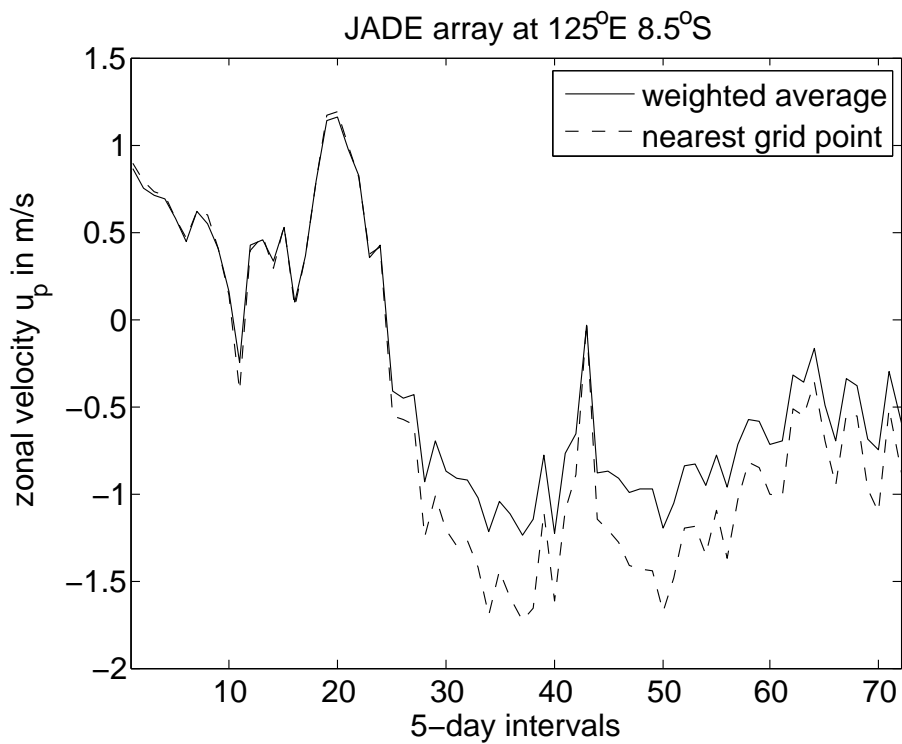


Fig. 6. Time series of 5-snapshot averaged zonal velocity u_p from the HYCOM DA model run. The solid line used a weighted average of points within a disk of radius about 16 km, see Eq. 2. The dashed line is the closest grid point time series. This location represents the most extreme difference in mean squared values $\langle u^2 \rangle$ computed from the two methods.

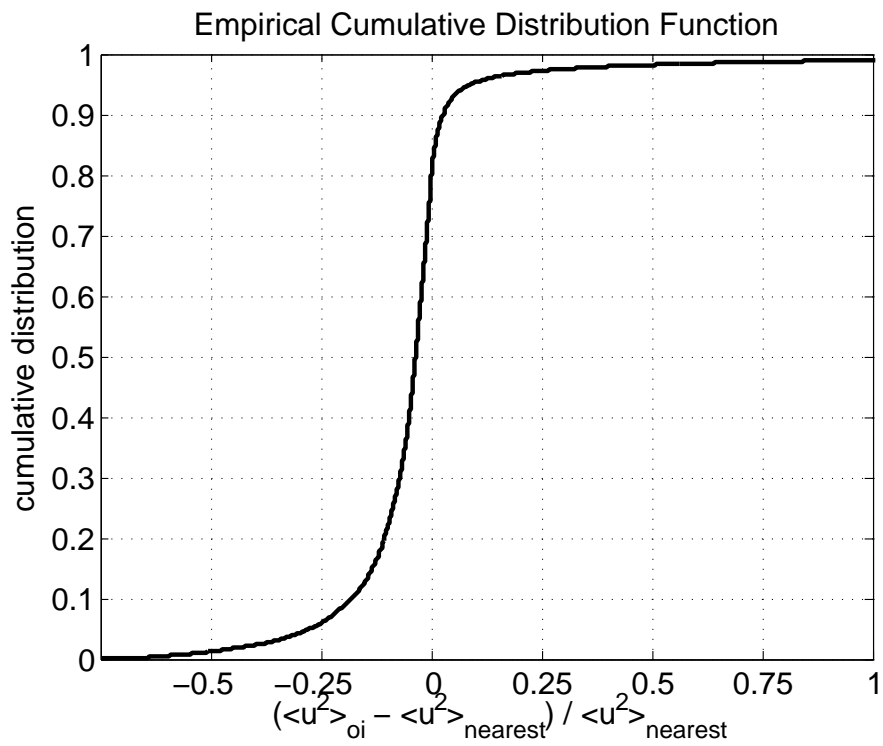


Fig. 7. Cumulative distribution of the difference in $\langle u^2 \rangle$ from the HYCOM DA run calculated using the weighted averaged (subscript 'oi') and nearest grid point (subscript 'nearest') methods, divided by the latter.

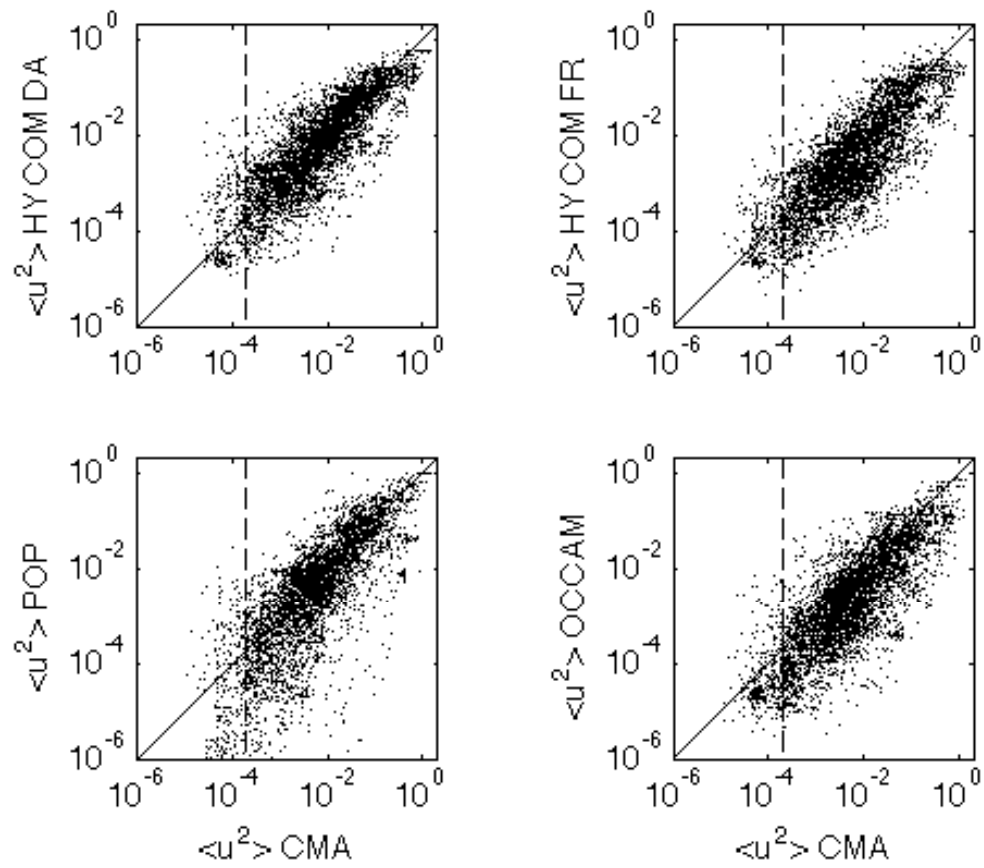


Fig. 8. Scatter plot of $\langle u^2 \rangle$ from the CMA, and the four OGCMs. The thin vertical dashed line is the threshold above which the correlations quoted in Table 2 were computed.

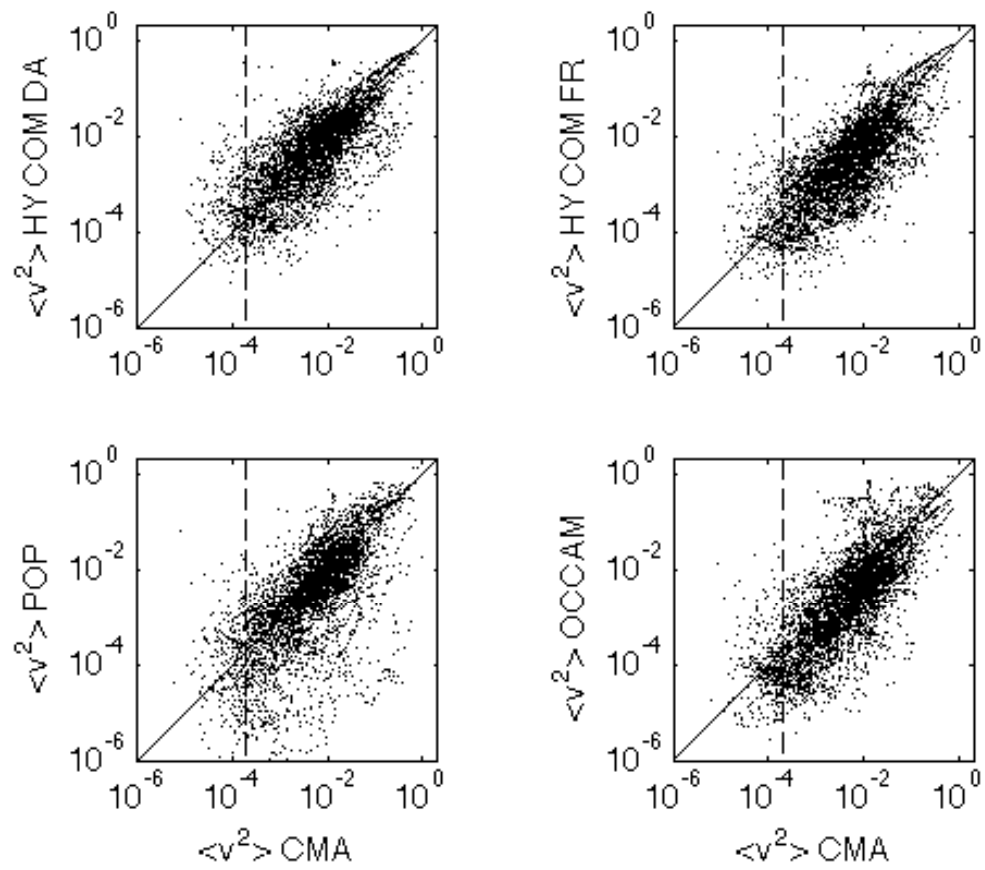


Fig. 9. As in Fig. 8 but for meridional velocity $\langle v^2 \rangle$.

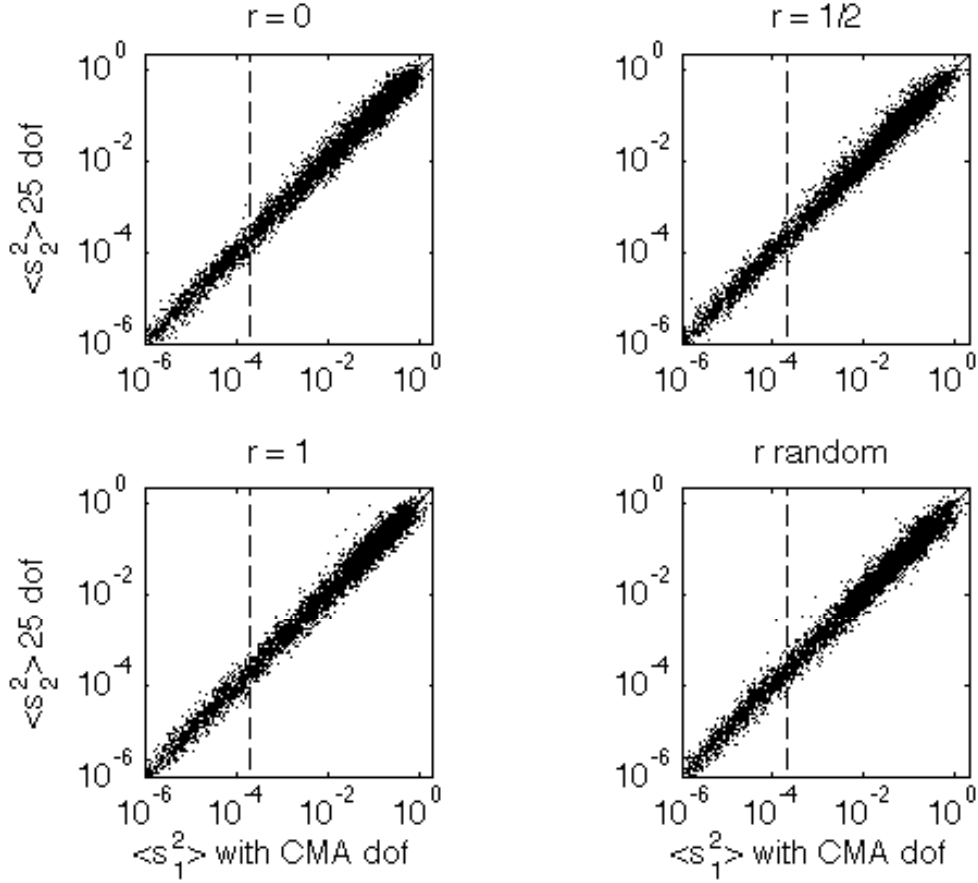


Fig. 10. Scatterplot of simulated $\langle S_1^2 \rangle$ vs. $\langle S_2^2 \rangle$, the sample variances of pseudo random pseudo CMA records and corresponding pseudo OGCM time series, see Section 4. Subplots are for different values of r , the correlation between pseudo time series at different levels on the same mooring. The scatter here is meant to depict the effect of statistical uncertainty and is the bench mark of perfect CMA records and perfect OGCMs, to which Figs. 8 and 9 should be compared. The thin vertical dashed line is the threshold above which the correlations quoted in Table 2 were computed.

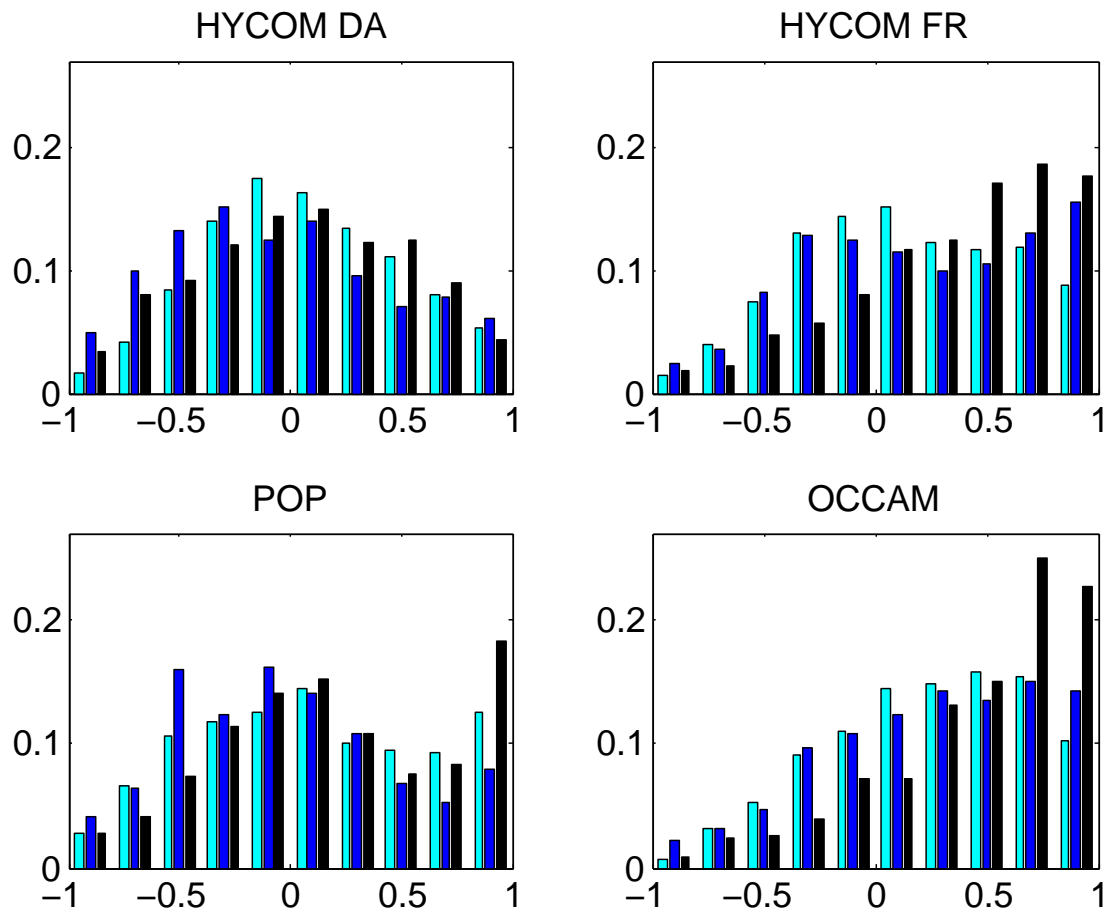


Fig. 11. Distribution of D for the four OGCMs for different depth ranges: 0 to 750 m (cyan); 750 to 2000 m (blue); below 2000 m (black)

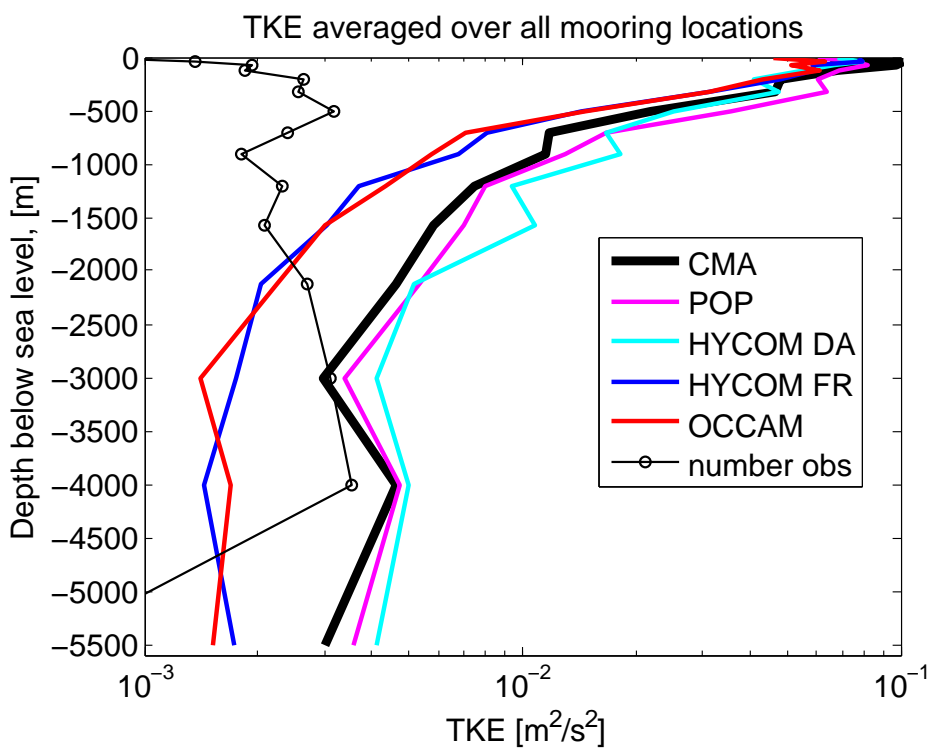


Fig. 12. TKE averaged over all current meter locations vs. depth for all models. Thin line with circles is the number of observations in each depth bin / 10^5 (so there were just over 300 observations in the depth bin centred just above 4000 m depth). There were only 56 observations in the lowest depth bin.

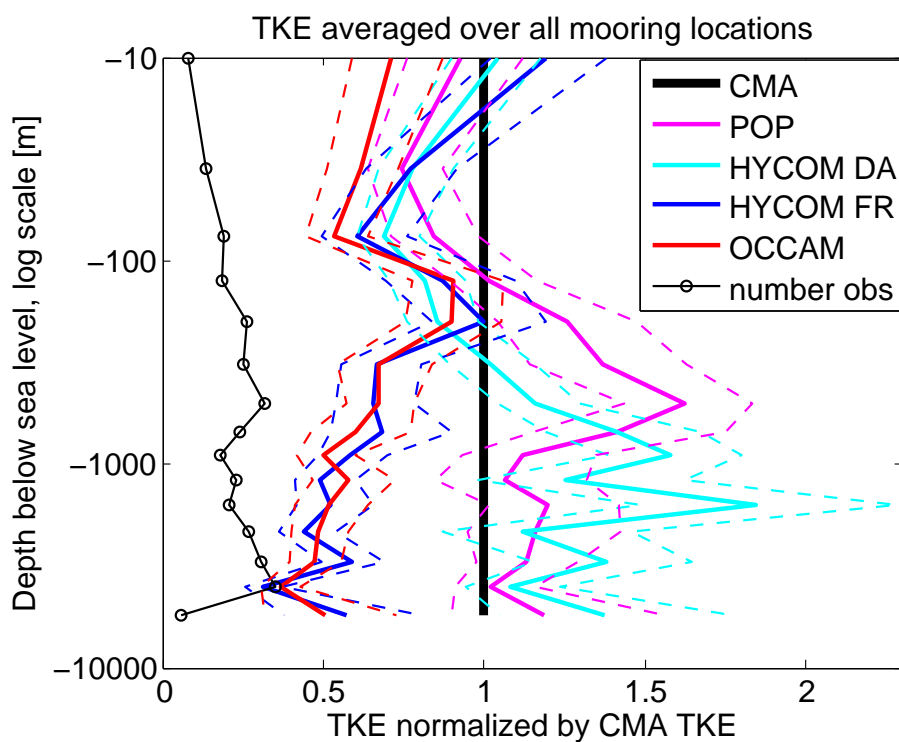


Fig. 13. As in Fig. 12 but now normalized by the CMA TKE, a log vertical axis, and the number of observations in each depth bin is in thousands. Using the closest grid point time series rather than the averaging method shifted the lines less than about 0.05 to the right. The thin dashed lines represent the 95% confidence bounds obtained using bootstrapping.

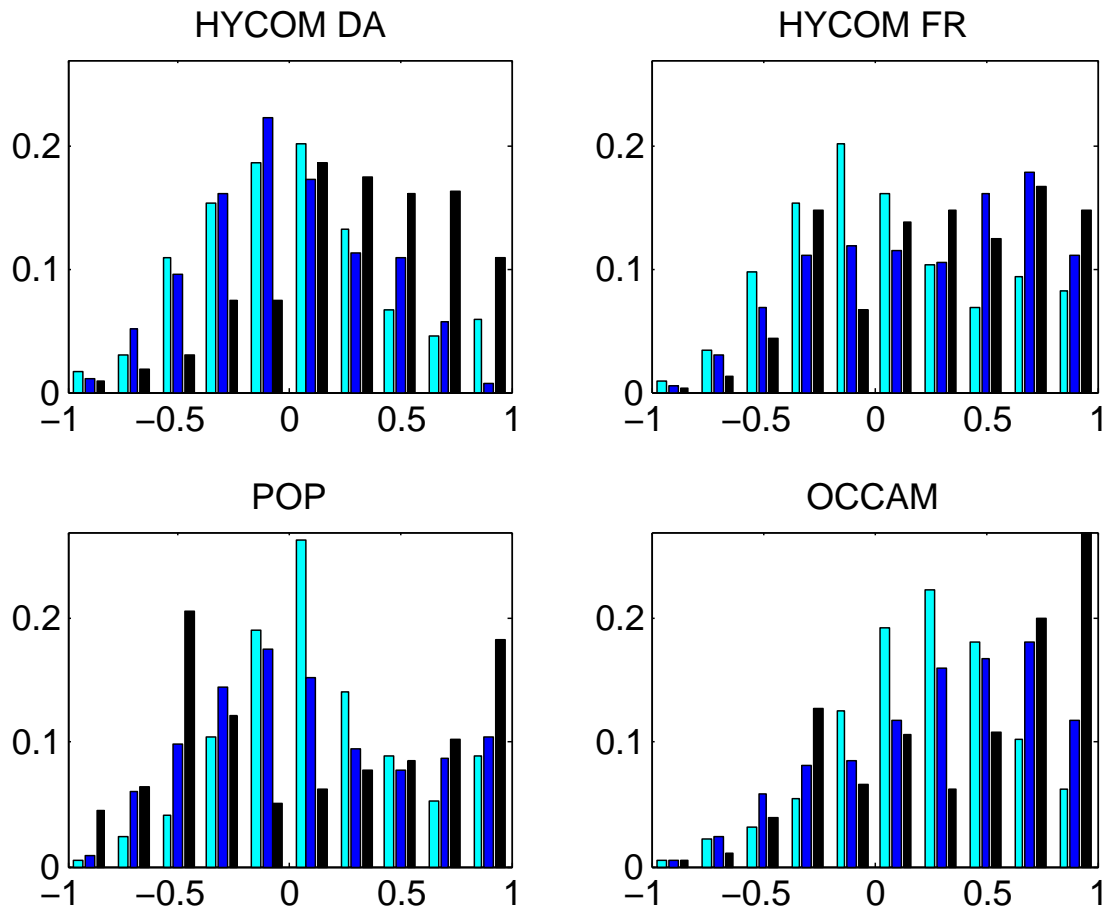


Fig. 14. As in Fig. 11 but the results are stratified by latitude (as opposed to depth):
 0 to 20° (cyan); 20 to 45° (blue); > 45° (black)