Measuring the Skill of an Ocean Model under Eddy-Topographic Effects, Based on a Global Inventory of Long-Term Current Meters

Greg Holloway and Tessa Sou

Institute of Ocean Sciences, Sidney B.C. V8L4B2 Canada

Abstract. From a global inventory of current meters, each of more than 100 days duration, we calculate skills of a global ocean model with and without a parameterization of eddy-topography interaction ("Neptune"). Skills are measured by the kinetic energy of the difference between modeled and observed flow and by error in direction. We assess confidence in the results by repeated tests in which half of the observations are rejected. Without Neptune, the model achieves small but significantly nonzero skills. Inclusion of Neptune improves skills by an increment roughly twice as large as the basic skills without Neptune.

Measuring Model Skill

Advances in computing power have increasingly allowed large scale ocean models to execute with more realistic detail of topography. However, mechanics of eddy-topography interaction may yet require much finer resolution before they can be treated explicitly. An alternative to parameterize eddy effects is an active research topic (Roubicek, Chassignet, and Griffa, this volume). When developing parameterizations for practical application, one seeks "skill measures" based upon observations against which one may refine uncertain model aspects. One tries to identify those ocean attributes which are well observed and which exhibit sensitivity to model components one seeks to refine.

Recently Alvarez et al. (1994), Eby and Holloway (1994, hereafter EH), Fyfe and Marinone (1995) and Holloway et al. (1995) have implemented a representation "Neptune effect") of eddy-topography forcing. These studies considered the western Mediterranean, global ocean, Georgia Strait, and Japan Sea domains, respectively, reporting apparent improvement in fidelity of modeled circulations. Other than modest parameter exploration in the Japan Sea study, little has been done to systematically adjust uncertain Neptune parameters against directly observed flows. Rather, the various authors have only cited examples of currents that may be "better" with Neptune. We seek more quantitative measure of such putative "improvement", perhaps providing a basis for subsequent optimization.

Oceans have been measured in many ways. Distributions of temperature (T), salinity (S), oxygen, and nutrients species are archived. Surveys of transient tracers are reported. Drift bottles, ship drift, current meters, and surface and subsurface floats provide information on currents. Satellite altimetry, electromagnetics, and acoustic methods provide further information.

We considered datasets that provide global coverage, with an aim to measure performance of a global model such as EH rather than a more region-specific study. We first appraised the EH model against T and S from the Levitus (1982) atlas. This proved frustratingly inconclusive. While the EH output exhibited large departures from Levitus, those discrepancies were insensitive to inclusion of Neptune. Errors in temperature and salinity are more dependent upon other factors such as uncertain surface forcing and model misrepresentation of mixing, stirring, and convection. We turned to datasets that directly address circulation features. (Although circulation and tracer distributions are coupled, inference of one from the other is unclear. A model which might get T and S "right" may still have circulation quite wrong, as seen in uncertainty of inverse models, while conversely "right" circulation can produce quite wrong T and S by misrepresentation of mixing, forcing, etc.)

Altimetric products offer an approach to global circulation, at least on larger scales. However, as models tend to produce similar large scale gyres under Sverdrup dynamics (to within eddy-driven recirculation and artifacts of grid-scale smoothing), altimetry may not be decisive with respect to topographic effects upon mean flow We expect some of the swiftest narrow flows to overly steep topographic slopes where geoid uncertainty will contaminate altimetric estimation of mean flow. Other approaches based upon transient tracers or drifters offer promise which we've not yet pursued. We turn to long-term current meter records.

From published results as well as privately contributed archives, we've assembled more than 2000 records based on minimum duration of 100 days. Unhappily, the global distribution is quite nonuniform. Moreover, one may anticipate that any current meter, even if its duration is sufficient to sample some long-term "mean" (itself an ambiguous idea), may be quite unrepresentative of flow resolved on a model with grid spacing of 200 km (as EH). Equally long-term records obtained only some few kilometers apart may differ markedly depending upon specifics of local topography. Our hope is that, if there is not a systematic bias in the unrepresentativeness of each current meter, then large numbers of records should provide useful skill measure.

Definitions of Skill

At each current meter we obtain mean flow $\mathbf{d}_i = (u,v)_i$, where "mean" means time-average over the duration of the i-th record. We also note the total variance σ^2_i of departures from $\mathbf{d}_i$. (In many cases variances of the u- and v-components, and the uv-correlation, or principal axes of a variance ellipse, are available. However, to provide uniformity over as large a global dataset as possible, we use only σ^2_i.)

In the case of EH, the GFDL model "MOM 1.1") was run in global domain on a grid 1.875° longitude by 1.856° latitude by 31 levels. Two runs were made, integrated for 800 years each under climatological mean wind (Hellerman and Rosenstein, 1983) with surface relaxation of T and S to Levitus (1982). One run ("AsUsual") was done in a conventional manner; the other run ("Neptune") centered the lateral viscosity operator about a non-zero flow field $\mathbf{U}^*$, thus $A_m\nabla^2(\mathbf{u}\text{-}\mathbf{U}^*)$, where $\mathbf{U}^*$ is obtained from a transport streamfunction $\psi^* = -fL^2H$, where f is Coriolis parameter, H is water depth and L is length parameter given by EH as a weak function of latitude. The velocity fields from EH runs were then interpolated to current meter locations to obtain model velocity mi. Difficulties arose due to coarseness of grid and the "staircase" topography of the GFDL model. Current meters which might be near steep topography were sometimes seen by the model as interior to "earth". Ad hoc rules "rescued" as many current meters as possible, for example by applying the bottom-most model velocity to be located at the current meter when such current meters occurred "not too far" below the model ocean bottom.

We wish to measure the skill of $\{\mathbf{m}_i\}$ against $\{\mathbf{d}_i\}$. A natural choice is to measure energy of the difference **m-d**, thus an error kinetic energy eKE = $0.5(\mathbf{m}\text{-}\mathbf{d})\bullet\mathbf{V}^{-1}\bullet(\mathbf{m}\text{-}\mathbf{d})$ where **V** is a diagonal matrix of elements σ^2_i normalized to trace **V** = 1. Ideally **V** would be a matrix of standard error of estimates $\mathbf{d}_i$ of some "true mean" $< \mathbf{d}_i >$. To make this calculation we would need to estimate numbers of degrees of freedom in the current meter records. In some cases, investigators produced such information; but in many cases it is not available. For uniformity while retaining the largest dataset, we've kept only σ^2_i.

As reference for eKE, we compare the weighted KE of the data (dKE = $0.5\mathbf{d}\bullet\mathbf{V}^{-1}\bullet\mathbf{d}$). Then it is convenient to form a ratio which we call "skillE" ("E" is for "energy"):

$$\text{skillE} = (\text{dKE} - \text{eKE}) / (\text{dKE} + \text{eKE})$$

such that an error-free model (eKE = 0) yields skillE = 1 whereas a model with huge eKE >> dKE yields skillE approaching -1. Within the range -1 <skillE <1 it is also important to note the skill of a completely skill-less model, one whose flows **m** are randomly unrelated to **d**. With the weighted kinetic energy of the model (evaluated at the current meter locations) given by mKE = $0.5\mathbf{m}\bullet\mathbf{V}^{-1}\bullet\mathbf{m}$, the value of skillE for the skill-less model is called "skillF" ("F" is for "floor"):

$$\text{skillF} = -\,\text{mKE} / (2\text{dKE} + \text{mKE})$$

so that the achieved energetic skill of the model is the difference skillE - skillF.

We've examined another measure of skill based only on unit vectors $\mathbf{d}_i = \mathbf{d}_i /|\, \mathbf{d}_i \,|$ and $\mathbf{m} = \mathbf{m} /|\, \mathbf{m} \,|$ by forming the weighted inner product "skillD" ("D" is for "direction"):

$$\text{skillD} = \mathbf{d}\bullet\mathbf{V}^{-1}\bullet\mathbf{m}$$

so that skillD simply asks if the model knows which way the water goes, regardless of speed. SkillD also falls within bounds -1 <skillD <1

Results

As we gather current meter records from various regions, we ask how stable will be the results for skills E, F and D, given the plausible unrepresentativeness of current meters relative to a coarse resolution model. We found that after we had several hundred records in the database, results became more stable with respect to adding further records. With more than 1000 records, similar results were obtained and were like those found in the present paper using nearly 2000 records as given below:

	E	F	E - F	D
AsUsual	0.003	-0.065	<u>0.068</u> (0.034)	<u>0.103</u> (.080)
Neptune	0.093	-0.100	<u>0.193</u> (0.037)	<u>0.288</u> (.065)
U* only	0.087	-0.080	0.167	0.289

None of the skills are very large. Although this reflects in part model infidelity, both from unfaithful internal dynamics as well as imperfect applied forcing, the small values of skill also reflect the difference between pointwise current meter records and a model representation on vastly coarser scale (even if such a model were "perfect" on its resolved scales).

Results in the first two rows are from EH, without and with Neptune parameterization. To assess stability of the results, we performed ten trials in which individual current meter records were rejected with probability 0.5. Roughly half the data were randomly discarded for each because we cannot know which current meters are influencing skill. The standard deviations of skills E-F and D from these ten independent trials are shown in parentheses.

Although skills are small, some results emerge. All of the E-F and D are positive, which may be encouraging for the numerical modeling enterprise in general! Moreover, the skills are positive by more than one standard deviation across the ten trials that randomly reject half the data. Of concern here is that the *increment* in skill from "AsUsual" to Neptune is roughly *twice* as great as the basic skill of "AsUsual". This increment is substantially larger than the standard deviation across the ten trials.

One sees clear suggestion that eddy-topographic forcing is a major part of ocean dynamics, hitherto omitted in non-eddy-resolving models and possibly quite corrupted by marginally eddy-resolving "eddy-admitting") models. The specific parameterization employed by EH appears to contribute skill. One could imagine repeating the skill calculations above to "tune" the EH parameterization; however, the computational cost to do so is large and may not be warranted given the presently uncertain basis on which the EH parameterization was proposed.

A third line in the table adds a chilling footnote. Our idea with Neptune parameterization is that internal eddy tendencies compete with externally imposed forcing (wind, thermal, freshwater, ...) We suppose that combining the two tendencies by means of parameterization should yield superior results, as the table indeed indicates. Now we "turn the table" by retaining only the parameterized internal tendencies while omitting all external forces. [One may object that without external forcing there wouldn't be eddies hence shouldn't be any such parameterization. However, first, we only mean to pose an "interesting" remark and, second, we omit only mean forcing. Or one could bandy words about stochastic forcing by unresolved monster goldfish.] Without mean external forcing, Neptune simply brings the flow to **u** = U*. Following conventional wisdom that wind and sun and such cause ocean currents, we should expect to see skill markedly reduced when we remove (in the mean) the wind and sun and such. Surprisingly and perhaps distressingly, the table does not support this. Skills E - F and D under U* only are insignificantly different from Neptune (which includes conventional forcing). While the global inventory of current meters shows statistical mechanics at work, after taking this into account the inventory of current meters cannot tell "which way the wind blows". [We haven't taken a next step to blow the mean wind backwards and see if it really doesn't matter.]

There is another footnote. While we have drawn upon statistical mechanics to improve the modeling of mean flow, there is an intimate connection between equilibrium statistical mechanics and nonlinear stability as discussed by Carnevale and Frederiksen (1987), leading us to suggest that flows nearer to **U*** should be more stable hence more steady.

Do current meter records support this? From the inventory, we formed two bins: simply "with" or "against" U* (as sign **d**•**U***). We found (for the number of current meters then available) there were 677 "against" and 1156 "with" **U***, consistently with skillD. Averaged over each bin, we formed the kinetic energy of fluctuations (eddy KE, "EKE") and the kinetic energy of mean flows "MKE"). Ratios EKE/MKE were 1.78 "with" and 3.03 "against". These are ratios of average quantities. We also considered the ratio EKE_i/MKE_i at each current meter, averaging these ratios over each bin. The results are wilder (less stable) numbers: 35.1 "with" and 99.1 "against". We see evidence that when external forcing admits flows closer to statistical mechanical equilibrium, these tend to exhibit less variability.

Aconclusion

It is too early to draw conclusion about such a "new" idea as the role of entropy gradients forcing ocean flows. The important and rather exciting observation is that it seems possible to make substantial advances in the skill of ocean models (or theory) by recognizing internal eddy tendencies as playing a role far greater than in usual eddy viscosity. We consider a probability distribution of possible oceans, gradients of distribution entropy appearing as forces acting upon realized moments of the distribution. Approximating such forces in practice has consisted of estimating the entropy gradient by (linear) departure from an approximate state of higher entropy (a maximum under idealized quasi-geostrophic (QG) dynamics). In particular we anticipate that eddy-topographic effects ("Neptune") should drive flows toward a non-zero mean state rather than "as usual" state of rest. A poorly determined eddy "fudge factor" appears as the L^2 in the definition of **U***. We surely guess that this prescription is not "right"; only it may be less wrong than "as usual".

Comparing effects of including Neptune tendency with observations from a global inventory of current meters, one may be encouraged by a striking increment in skill. Measured either by kinetic energy of the difference between model flow and observed flow, or simply by agreement in direction, the increment in skill is nearly twice as large as the basic skill "as usual". A caution is needed: These results, based upon model integrations of EH, test skill against a particular configuration of the GFDL ocean model, with coarse resolution and sundry internal parameters under particular conditions of external forcing. Such model outcomes are distressingly sensitive to "fiddles" with internal parameters and external forcing, with "details" of topography, and with respect to underlying model formulation (for example in layers rather than levels). The danger of appearances of right answers for wrong reasons is ever a concern. What we can see is the possibility for substantial progress, both at theoretical understanding and at practical model skill. This motivates fresh attention to such basics as statistical mechanics in QG dynamics and to extensions such as explored by Roubicek et al. While further efforts are made at fundamentals, we may also learn from try-and-see practical application, ranging from global integrations through marginal sea studies (Alvarez et al., 1994; Holloway et al., 1995) to estuarine scales (Fyfe and Marinone, 1995).

Acknowledgment. GH has been supported in part by the National Science Foundation (OCE 93-11550) and TS by the Office of Naval Research (N00014-92-J-1775).

References

Alvarez, A., J. Tintore, G. Holloway, M. Eby and J. M Beckers, 1994, Effect of topographic stress on the circulation in the western Mediterranean, *J. Geophys. Res.*, 99, 16053-64.

Carnevale, G. F. and J. S. Frederiksen, 1987, Nonlinear stability and statistical mechanics of flow over topography, *J. Fluid Mech.*, 175, 157-181.

Eby, M. and G. Holloway, 1994, Sensitivity of a large scale ocean model to a parameterization of topographic stress, *J. Phys. Oceanogr.*, 24, 2577-2588.

Fyfe, J. and G. Marinone, On the role of unresolved eddies in a model of the residual currents in the central Strait of Georgia, B.C., *Atmos-Ocean*, to appear

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

Holloway, G., 1992, Representing topographic stress for large scale ocean models, *J. Phys. Oceanogr.*, 22, 1033-46.

Holloway, G., T. Sou, and M. Eby, Dynamics of circulation of the Japan Sea, J. Mar. Res., to appear.

Levitus, S., 1982, *Climatological Atlas of the World Ocean*, NOAA Prof. Paper 13, Washington, DC.