A review of egusphere-2023-310 :

The Mixed Layer Depth in the Ocean Model Intercomparison Project (OMIP): Impact of Resolving Mesoscale Eddies

by Anne-Marie Treguier et al.

Summary

The aim of this MS is to estimate the effect of resolved mesoscale processes on the ocean mixed-layer depth (MLD). To do this, the authors have diagnosed the ocean mixed-layer depths in a set of similarly-forced runs of ocean models submitted to phase 2 of the Ocean Model Intercomparison Project (OMIP2) available at both eddy resolving lateral resolution (~0.1°) and coarse CMIP-style 1° resolution. They are then compared with the observation-based dataset of de Boyer Montegut et al. (2022); henceforth BM22, an update of de Boyer Montegut et al. (2004); henceforth BM04.

The authors first discuss how best to define mixed-layer depth. This is always an ambiguous concept, but they follow BM04 in trying to capture the depth over which "surface fluxes have been recently mixed … meaning a timescale of at minimum a daily cycle, and no more than a few daily cycles."

This depth is essentially the nighttime MLD, and they again follow BMO4 in taking the MLD to be the shallowest depth where the density ρ is 0.03 kg m^{-3} greater than $\rho_{10\mathrm{m}}$, that at 10m, i.e. $\rho - \rho_{10\mathrm{m}} \geq 0.03$kg m^{-3}, so as to exclude shallow (1–3 m) daytime stratification. They follow Griffies at al. (2016) (G16) in using this simple density criterion rather than the density difference implied by a temperature difference of 0.2° at the surface (which depends on the thermal expansion coefficient and hence the surface temperature and salinity) that is used by e.g. Holte et al. (2017). Monthly-mean outputs for May and September from the NEMO coarse-resolution model that had fine vertical resolution (~1 m near the surface) showed that indeed taking the density difference from that at the surface (uppermost grid level); i.e. $\rho - \rho_1 \geq 0.03$ kg m^{-3} rather than $\rho - \rho_{10\mathrm{m}} \geq 0.03$ kg m^{-3} gave MLD that were widely up to 40m shallower. For the NCAR-POP model with 5 m vertical resolution that did not resolve the diurnal cycle properly, differences were less marked.

They then discuss the well-known importance of using spatial averages of MLD calculated from individual profiles rather than calculating MLD from the spatially-averaged climatology of temperature and salinity. By comparing the BM22 climatology calculated from individual profiles with the a MLD field calculated from the recent ISAS temperature and salinity climatology they show that the latter underestimates the zonally-averaged MLD by typically 5 m in summer and 25 m in winter.

Next they investigate the impact of non-linearity in the time averaging. In both high and low resolution versions of a couple of the submitted model-pair runs, the monthly average of the online-averaged MLD calculated every time step was generally shallower than the MLD calculated from the monthly-average temperature and salinity: by about 2–5 m in the summer and by 20–40 m in winter. These differences, though significant, were generally not as great as those arising from using the surface-layer instead of the 10-m density.

Motivated by the above they settle on a definition of the model MLD as the shallowest depth where as $\rho - \rho_{10m} \geq 0.03$ kg m^{-3}, calculating the density from the monthly-average temperature and salinity: this definition has the advantage that it can be calculated for all models in the same way, rather than using the online-averaged MLD that are defined using different criteria in different models.

They than address the question of the impact of resolution on MLD by examining the summer and winter MLD defined as above for the various models both at eddy-resolving and coarse resolution, and comparing against the observational dataset of dBM22.

The IAP-LICOM model pair is a clear outlier: at both coarse and high resolution it has far too shallow (by >30 m) summer MLDs in the ACC, and at high resolution way too deep (by >200–300 m) winter MLDs in the Nordic Seas and parts of the ACC. The too-shallow MLDs are ascribed to the use of the Canuto ML model that is normally used with short time steps (a few minutes) in models with longer time steps (up to ~1 hr). The FSU-HYCOM model has too deep (by >200–300 m) winter MLDs in the ACC at both coarse and fine resolution; the reasons for this are unclear.

The other models (ACCESS-MOM, CNCC-NEMO, NCAR-POP and AWI-FESOM) all tend to have excessively deep winter MLDs over the Labrador and Nordic Seas (and over parts of the ACC) at coarse resolution, but this deep bias is considerably reduced at high resolution. Summer MLD biases seem to be generally less sensitive to resolution, though the excessive near-equatorial MLDs produced at coarse resolution by all four models (and FSU-HYCOM) are shallower at high resolution. It is suggested that the better resolved near-equatorial flow at high resolution transports salty water masses more correctly, giving more realistic vertical stratification. CMCC-NEMO does however give deeper summer ML over the ACC at high resolution than at coarse resolution; this may be associated with the coarser vertical grid used at low resolution.

The performance of the models in simulating the annual cycle of MLD over key deep-water formation sites (the Greenland, Labrador and Irminger Seas) is then considered, focusing on the high-resolution runs, since they perform better than the coarse runs in these regions. Wide model to model variation (e.g. in the Labrador Sea ACCESS-MOM MLD > 1500 m while NCAR-POP MLD ~800 m) is evident in the maximum MLDs in late winter/early spring. NCAR-POP perhaps performs the best. MLDs in intermediate water generation sites are then considered. The models perform very well in the eastern sub-polar Atlantic, and fairly well in the Sargasso Sea and Kuroshio, with relatively little model spread but with a consistent deep bias of ~50 m, though ACCESS-MOM is a deeper outlier (~30 m deeper). However over both the ACC south of Australia and in the ACC in the eastern South Pacific, there is considerable model spread, with ACCESS-MOM again a deeper outlier.

Conclusions are then:

1. Model MLD should be defined as the shallowest depth where as $\rho - \rho_{10\mathrm{m}} \geq 0.03$ kg m-3 and then compared against consistently calculated datasets like dBM22.

2. Model MLD should be calculated online and then statistical moments such as averages and standard deviations calculated over longer periods.

3. The impact of the diurnal cycle could and should be investigated in models with sufficient vertical resolution (~1 m near the surface) by considering the density difference from the top level $\rho - \rho_1 \geq 0.03$ kg m-3.

4. Finer lateral resolution mostly gives shallower, generally more realistic, winter MLDs.

5. Finer lateral resolution does not always lead to more consistent results between models (e.g. over the ACC) , though more consistent maximum winter MLDs are seen in the eastern subpolar Atlantic, the Sargasso Sea and Kuroshio.

Major comments

This is an important and timely piece of work on a key topic that is only going to become more important for climate science. The work presented here is rigorous and careful, and the authors have taken care to point out the inconsistencies in the model datasets employed here —there are not many eddy resolving runs available forced by the same atmospheric forcing datasets, and different vertical resolutions and parameterisations are used by different models. Hopefully this MS will stimulate further work on this subject.

The suggested definition of the model MLD as the shallowest depth where as $\rho - \rho_{10\mathrm{m}} \geq 0.03$ kg m-3 (at least when calculating MLD online) should be followed in future work.

It is an important finding that finding that (generally) the MLD thus defined is indeed shallower at eddy-resolving resolution, presumably because the eddy parameterisations at coarse resolutions are not as effective at fluxing buoyancy upwards as are the resolved eddies at coarse resolution.

It is also interesting to see how sensitive the MLD are to the actual ML model and to model vertical resolution, especially in summer.

I was not fully convinced, however, of the argument that the MLD calculated from the monthly-average temperature and salinity, $\mathrm{MLD}(\langle T \rangle, \langle S \rangle)$, was a totally adequate reflection of the monthly-mean of the MLD calculated online at each time step $\langle \mathrm{MLD}(T, S) \rangle$, and thus a fully comparable field to the MLD calculated from individual observational profiles. For the zonally averaged $\langle \mathrm{MLD}(T, S) \rangle - \mathrm{MLD}(\langle T \rangle, \langle S \rangle)$ in Fig. 6 seem to reach ~30–40 m in late spring. Moreover of the two models plotted in Fig. 6, IAP-LICOM is very much an outlier, and FSU-HYCOM has notably coarse vertical resolution (~ 36 layers) and so would not expect to show much of an effect, especially in summer. Fig. 6 would be more convincing if it included output from a higher resolution, "good" model, like CMCC-NEMO which has 98 levels. Why not use NCAR-POP and CMCC-NEMO as were used in Fig. 3 to show the impact of using different reference levels?

I agree that for purposes of comparability it was necessary to use MLD($\langle T \rangle$, $\langle S \rangle$), but suggest that it would have been better to have tried a larger density difference criterion when using monthly-mean averages. For it is conventional when calculating MLD using monthly-mean observational climatologies to use a larger density difference, e.g. WOA 2018, which uses $\rho - \rho_{10\mathrm{m}} \geq 0.125$ kg m^{-3} s. It would be relatively straightforward to calculate MLD using monthly-mean T and S, with a range of density differences starting up from 0.03 kg m^{-3} and ranging up to 0.125 kg m^{-3} using e.g. CMCC-NEMO, and choose the density difference that minimised deviations from the monthly-average of the MLD calculated online with the 0.03 kg m^{-3} density criterion (i.e minimise the difference field plotted out in the panels in Fig. 6).

Recommendation

This manuscript will be very useful for the field and should be published subject to minor corrections.

Detailed comments

P6, l138. Does IAP-LICOM include the wave generated mixing of Qiao, et al., GRL, 2004?
P6, l164 “NEMO version 3.64” Is this what you mean?
P14, Fig. 4. Caption. Observation datasets ⇒ observational datasets.

P17, Fig. 4. Why not use NCAR-POP and CMCC-NEMO for consistency?

George Nurser