Gallmeier et al. apply a relatively novel statistical method Ulmo to assess a global submesoscale permitting simulation of the ocean (LLC4320). With the increase in computational power, such high-resolution ocean simulations have started to emerge. However, due to issues related to storing and disseminating petabytes of data these simulations produce, a comprehensive assessment on the reality of such simulations has remained a challenge for the ocean modeling community. In this manuscript, the authors compare the fine scale structure in sea-surface temperature (SST) simulated by LLC4320 to satellite observations. They show that while LLC4320 captures the overall basin-scale features in SST, there are discrepancies in the equatorial region and eddy active regions. To my knowledge, this is the first study to attempt to provide a quantitative evaluation of such high-resolution simulations and I believe that their Ulmo method provides a method in evaluating current and future ocean simulations. I recommend their manuscript for publications with minor revisions.

- Lines 46-47: This is more of a comment rather than a criticism but the statement "*the simulated mesoscale and submesoscale features of free- running models are not expected to match, one-to-one, the observations*" is an obvious one. Even if we had the observational tools, free-running models will never match up perfectly to observations due to the chaotic variability of the ocean (e.g. Serazin et al., 2015; Penduff et al., 2018; Leroux et al., 2022; Uchida et al., 2022).

- Lines 223-224 and 375-376: Can the authors comment on how tidal forcing could affect the simulated SST structures that are too energetic? I am asking this because LLC4320 was inadvertently overly forced by tides and consequently is too energetic about the semi-diurnal frequencies (Yu et al., 2019; Arbic et al., 2023).

- Lines 265-272: Is the comparison between LL_{VIIRS} and LL'_{LLC} a fair one given that the latter is corrected by the former and hence the two are no longer independent with each other?

- Lines 285-287: Can the authors explain why the thresholds provide a criteria to judge bad measurements of SST or deficiencies in the simulation?

- Lines 317-318: Could the fact that LLC4320 is a forced ocean simulation and not coupled with the atmosphere be affecting the lack of structure in the equatorial region? In other words, could the atmospheric forcing be damping out the SST structures or could the relatively low temporal resolution in the atmospheric forcing not be exciting sufficient levels of inertial variability (Arbic et al., 2023)? It would be interesting to see if we'd see the same lack of SST structures in the DYAMOND simulation for example, which is air-sea coupled (https://gmao.gsfc.nasa.gov/global_mesoscale/dyamond_phaseII/docs/GEOS-vPICO-EGU21-12782.pdf).

- Line 324 and FIgures 13, 15, 17: Is dT equivalent to ΔT introduced in line 199?

References

- Arbic, B. et al., (2023). Frequency dependence and vertical structure of ocean surface kinetic energy from global high-resolution models and surface drifter observations. *ArXiv*.
- Leroux, S. et al., (2022). Ensemble quantification of short-term predictability of the ocean dynamics at a kilometric-scale resolution: a Western Mediterranean test case. *Ocean Sci*.
- Penduff, T. et al., (2018). Chaotic variability of ocean heat content: climate-relevant features and observational implications. *Oceanogr*.
- Serazin, G. et al., (2015). Intrinsic variability of sea level from global ocean simulations: Spatiotemporal scales. *J. Clim*.
- Uchida, T. et al., (2022). An ensemble-based eddy and spectral analysis, with application to the Gulf Stream. *J. Adv. Mod. Earth Sys*.
- Yu, X. et al., (2019). Surface kinetic energy distributions in the global oceans from a high-resolution numerical model and surface drifter observations. *Geophys. Res. Lett*.