Review of Dynamically downscaled seasonal ocean forecasts for North American East Coast ecosystems.

August 20, 2024

This article discusses the skill of historical seasonal forecasts utilizing a regional North Atlantic ocean model initialized with the GLORYS12 ocean analysis (approximately the same resolution as the ocean model) and forced by atmospheric forecast conditions from the SPEAR seasonal forecast. Results are then compared with the global, low resolution, coupled forecasts of SPEAR.

I have always been an advocate for the usuage of ocean reanalysis as a tool for both initialization of forecasts – and for their use as a diagnostic tool to assess the ocean in regions where observations are sparse, or non-existant. However, the usage here, to use the GLORYS12 product as initialization for the regional seasonal forecasts – and then to assess skill against the GLORYS12 reanalysis seems somewhat incestuous to me – especially since the SPEAR ocean analysis likely differs substantially from the GLORYS12 analysis The study tnen really becomes one of assessing the initialization of a high resolution ocean model with GLORYS12 versus initializing with the ocean model component of SPEAR, and not particularly a "downscaling" of a seasonal forecast.

My question to the authors: If this system was to become an operational forecast system for the U.S. East Coast, would the goal be to intialize such forecasts with the real time GLORYS12 analysis *(Skill assessed in this manuscript)*, or initialized by downscaling the SPEAR ocean analysis *(Skill **not** assessed in this manuscript)*. If it is the former, then perhaps utilizing the full multi-model ensemble of NMME, as opposed to only SPEAR atmospheric component forcing, would be a more prudent approach, as that likely would increase the underdispersiveness of only using SPEAR atmospheric forecast forcing.

Reommendation: Despite my trepitation with regards to the skill assessment primarily against the system initialization product, **I would recommend publication after the authors answer my question and comments.**

Itemized Comments:

1. I believe other studies (not seasonal forecasts, however) have been undertaken with $\frac{1}{12}$th degree North Atlantic systems, although admittedly I could not find a particularly relevent study in my quick search. Perhaps the authors could more explicit with regards to the definitiion of their $\frac{1}{12}$th grid: Is the grid identical to a North Atlantic subset of the GLORYS12 grid, or how does it differ from the ORCA12 grid utilized by GLORYS?

2. It is not the responsibility of the authors to discuss the ocean initialization of SPEAR, but nonetheless, how it is initialized, and in particular, how its ocean state estimation approach differs from GLORYS12 is an important component of this study. More information is required to assess this, preferably with some explicit text in the manuscripts, but minimally by explicit citations of the SPEAR ocean initialization procedure. The manuscripts does show "0 lead" (actually 0.5 lead I believe) results that can be used to assess these differences somewhat, but some more explicit comparisons, particularly for the reemergence discussion would be useful – for instance, the manuscript shows the reemergence in the GLORYS12 reanalysis – is it also present in the SPEAR ocean analysis (or concatenation of 0 leads).

3. The statement in the conclusion, " Finally, full data assimilation, rather than nudging towards a reanalysis, could improve prediction skill through better initial conditions $\cdots$," might be true – but not necessarily when basing that skill on the reanalysis being nudged towards. This may be particularly true if not many observations are going into the ocean analysis in the areas of skill assessment, which unfortunately may be true for the coastal region, strong ocean current (short ARGO float retention) regions under study in this manuscript.

4. The spread error discussion was interesting, and the skill versus ensemble size (including CRPS results) did expand on this. But I am always interested in expanding on the probabalistic nature of the ensemble – and it would seem the reemergence diagnostics utilized here might be a natural way to expound on this. I assume the reemergence diagnostics are perform on the ensemble mean? Could an member by member diagnostics be performed that might lead to a "probabality" of re-emergence that could be accessed for skill (Brier Sc.ore)?

5. I remind the authors that Atlantic Overturning Circulation variability can be driven by atmospheric variability as well (Jackson et al, 2016; https://doi.org/10.1038/ngeo2715), with particular implications to density anomalies along western Atlantic.

6. The authors should highlight their skill assessment of ocean currents is performed using an independent "observation" source, and therefore is not as sensitive to the

initial conditions as their temperature and salinity skill assessment. Although it then may be instructive to give some evidence of current skill in the GLORYS12 analysis (Aijaz et al, 2023; https://doi.org/10.1016/j.ocemod.2023.102241)

7. In light of the previous two points, I wonder why authors did not present a more detailed skill assessment of currents beyond just a trend analysis?

8. I found the divergenging color schemes used in plots to not be particularly easy to distinguish null results, particulary with the red/green scheme used in figures 1-4 (plus I believe it is not particularly colour blind friendly). The purple/green scheme of figures 5-7 seems somewhat better – but either an explicitly white color marker for 0 difference, or 3 colour scheme (i.e. yellow as zero difference) might be preferable.