Summary

I would like to thank the authors for having taken all my comments into account and for their considerable efforts in revising the manuscript. This revised version of the manuscript is much improved, especially through the additions of the new observational analysis & reanalysis products, and the more robust statistical assessment of the LR & HR ensembles.

The authors have resolved all my original concerns. Upon addressing my small number of minor comments below, I would recommend this manuscript be accepted for publication in Ocean Sciences.

General Comments

- **Manuscript Length:**

  Whilst I'm grateful to the authors for addressing the comments of Reviewer #2 and myself so thoroughly, I'm slightly concerned that the revised manuscript has increased significantly in length from 34 to 47 total pages. I would strongly recommend reviewing the manuscript for opportunities to consolidate the existing figures / text, whilst preserving the central messages. I'd like to highlight one possible approach to do this below (although this is simply a suggestion):

  - Distributing the findings of **3.6 Testing the significance of differences between ensembles** amongst the relevant sections earlier in the Results. As a reader, I felt that this statistical analysis would have been most useful when the diagnostics were first discussed. This would provide an opportunity to reduce the length of the text, and the summary table could be moved to Supplementary Information / Appendix A.
  - Similarly, the contents of **3.7 Characteristic features in the HR-HIST models** could be distributed between the previous Results sections (for quantitative comparisons – which would then be helpfully located near each large multi-panel figure) and the excellent Discussion & Conclusions section (for the more speculative discussion points).

- **3.8 Relations between dynamical and physical properties:**

  This is an interesting addition to the manuscript, which directly addresses my previous concerns regarding the originality of the study. My only concern is whether it is appropriate to use a composite of the LR-HIST and HR-HIST ensembles to explore the correlations between state variables, since earlier you highlight important differences between these ensembles. For example, might it be the case that the relationship between the SPG strength and the maximum overturning at 26.5N is different between the LR & HR ensembles? Hirschi et al. (2020) highlighted that, at HR, the more realistic SPG circulation projects more strongly onto the diapycnal rather than vertical overturning at subpolar latitudes

(horizontal circulation across sloping isopycnals), whereas LR models exhibit a more classical vertical "conveyor" like overturning cell.

Suggest commenting that the relationships between dynamical & physical properties may hence also depend on horizontal resolution (although a larger HR ensemble would be needed to perform this analysis). This may be a case of bringing some of the excellent discussion on Lines 765-774 forward in the text.

Specific Comments

Methods

Use of GLORYS12v1: I'm grateful to the authors for including the mesoscale eddy resolving GLORYS12v1 reanalysis product in their model evaluation, however, I have some concerns regarding the implications of using the regridded outputs for the purpose of calculating meridional overturning and barotropic stream functions. Given that GLORYS12v1 is originally simulated on a curvilinear ORCA12 grid, the use of linear interpolation to regrid the model velocity will inevitably introduce multiple sources of error into a volume transport calculation (including estimation of grid cell areas, interpolation errors at high-latitudes and considerations of bottom topography). This likely also applies to the ORAS5m outputs shown. Given that GLORYS12v1 is available on its original NEMO model grid, it would at least be worth briefly emphasising to readers the limitations of using a regridded velocity field, although I suspect this will not alter your qualitative results.

Results

Figure 5: Suggest using a different colour / marker to identify the LR-HIST mean profile than dark grey, this is quite difficult to distinguish from the large number of LR-HIST ensemble member profiles (Fig. 5b especially) - although I do recognise the need to relate these visually.

Lines 392-419: Suggest revising this discussion on the methodological concerns regarding comparisons between models and observations at the RAPID array. I would highlight two important points that are missing:

1. There is an existing approach to compare ocean models with RAPID observations in an equivalent manner, accounting for the four separate AMOC components: using the Meridional ovErTuring ciRculation diagnostIC (**METRIC**) package originally developed by Castrucio et al. (https://github.com/AMOCcommunity/metric).
2. The chosen treatment of the net throughflow across the RAPID 26.5N section can be a important control on the magnitude of the depth-space overturning calculated in models. It would be useful to inform readers whether a uniform

volume transport compensation term was applied to each of the models (as done in observations) prior to the stream function calculation or if these diagnostics include the net throughflow across the section. This would be especially relevant for the reanalysis data used, since the regridded velocity field does not properly represent bathymetry.