Large Uncertainty in Observed Meridional Stream Function Tropical Expansion

Baldassare et al.

General

This is my second review of this work, where the authors analyze the uncertainty in the SF metric for estimating the extent of the Hadley circulation (HC) based on ensemble members of the ERA5 reanalysis. As before, the key findings are a reduction over time of uncertainty, which the authors associate with better quality of the assimilated data, and that the SF metric has a relatively high uncertainty due to less observationally constrained upper wind and relatively weak meridional gradients near the zero-crossing latitude, which increase the uncertainty of the SF metric compared to other zero-crossing metrics. My main previous concerns were that (i) using only the ERA5 ensemble limits the generality of the results, (ii) the authors provided recommendations on the metrics which are not generally justified , (iii) the statistical implications of using only 9 members were not adequately discussed, and (iv) the motivation and results were lacking context. Of these comments, I can accept the revisions regarding the two first items, but I am not satisfied with the latter two items. As before, I think the analysis may merit publication. However, some critical issues remain. Detailed comments are provided below.

General comments

- On second reading, I suggest to rephrasing the title. Mainly, this is because "Meridional stream function tropical expansion" is obscure. First, the analysis applies to the edge of the HC as quantified by the meridional stream function. Second, tropical expansion is a tricky term since the tropical rain belt and tropical climate land area is actually narrowing (see recent paper by Adam et al. 2023, "reduced tropical climate land area under global warming"). For example, "Large uncertainty in observed estimates of Hadley cell expansion based on the meridional stream function" would be more appropriate (and consistent with the wording of the abstract).
- I still think that a qualitative and quantitative discussion of the limitations of using only 9 ensemble members is lacking.
- Figure 6a looks odd. How many data points are in this plot? I an able to identify 8 (4 seasons X 2). Of these 8, 3 lie along some line and 5 have essentially the same value of Δ, and yet $R^2 = 0.97$? Clearly this is not a statistically robust result (for example, it would fail a cross validation test.). I would omit Fig. 6a and only keep 6b.

Comments by line

17-18 As I stated in the previous review, I don't agree with this statement. Previous analyses have also considered the uncertainty of HD edge metrics. The current paper is likely the first to consider using the ERA5 ensemble members to estimate

the uncertainty. But other estimates exist, for example, like those now mentioned int he text (Davis and Rosenlof, 2012; the TropD paper; Davis and Birner, 2017; Waugh et al. 2018; Seviour et al. 2018; etc.). Specifically, using the ERA5 ensemble members to estimate 'observed' uncertainty is very much like estimating the differences across en ensemble of climate models, i.e., 'modeled uncertainty'. The authors should do a better job of delineating their work from existing works. This is done in the revised discussion, and to some degree in the introduction, but not in the revised abstract.

33   Instead of "tropical width" I would use "the extent of the tropical circulation" (or Hadley cell extent).

36   I suggest refraining from such recommendations.

59   Note the critical differences between this statement (which I agree with), and the one in the abstract (see above comment on lines 17-18)

69-72 As mentioned in the previous review, this is an odd choice for motivating the present analysis, as Chemke and Polvani find the models to generally agree on the extent trends. Surely you can find additional works to motivate the analysis. For example, the importance of having good estimates of observational uncertainty is demonstrated in State et al. (2020, BAMS). Specifically, in the early 2000's works like Johanson and Fu (2009) found large discrepancies between modeled and observed trends, which sent many researchers rushing to identify the errors in the models, only to realize a decade later than the error was in the observed trends.

188   I believe kernel density estimates require some assumptions. Please provide a reference and provide the reader with all the information required to exactly reproduce your results.

241   remove "sometimes"

282   In the revised text you added that the change from 6% to 4% is not statistically significant. According to what criteria?