Review of

Large Uncertainty in Observed Meridional Stream Function Tropical Expansion

Baldassare et al.

General

The authors analyze the uncertainty in metrics for estimating the extent of the Hadley circulation (HC) based on nine ensemble members of the ERA5 reanalysis, focusing in particular on the commonly-used streamfunction (SF) metric. 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.

I think the analysis may merit publication. However, I find some critical issues with the methodology and conclusions. In particular, the authors should do a better job of constraining the uncertainty in their results and better contextualize the results. I also don't think that the authors should provide recommendations or try to rank the different metrics, but rather focus on providing the objective uncertainty estimates and discuss the associated implications. Detailed comments are provided below.

Comments

1. I partly disagree with the recommendation by the authors to use surface winds instead of the SF metric. Surface winds are affected by many processes and therefore the extent index based on surface winds captures information which may be fundamentally different from SF-based indices. In addition, as shown in the TropD paper (Adam et al.2018), the variance across models in the PSI and UAS metrics is roughly the same. The recommendation by the authors is strictly based on the apparent reduced uncertainty in ERA5. This point should be clarified, and the relevance of their recommendation better outlined. Similarly, there is great variance across reanalyses, which suggests that the uncertainty estimates by the authors are specific to the ERA5 dataset. This should be discussed.
2. Standard deviation (STD) is calculated across 9 ensemble members. The uncertainty in STD is inversely proportional to the number of degrees of freedom (N). Specifically, the fractional uncertainty in STD is $\frac{\delta\sigma}{\sigma} = \frac{1}{\sqrt{2(N-1)}}$ which for $N = 9$ gives an uncertainty of 25% in STD estimates. There is therefore significant uncertainty in the uncertainty estimates based on STD using only 9 members. This is a critical point in the discussion of changes in uncertainty over time. For example, in line 250, the change in STD from 6% to 4% is not statistically significant at 95% confidence.

3. Please better explain the y-axis in figures 1 and 2. It is not clear what these value signify and why the distributions are so smooth.
4. In figures 1 and 2 the trend itself is subject to uncertainties due to natural variability. In this respect, the mean extent is more revealing. Thinking about the trend as proportional to the difference between two mean values, the metric-uncertainty in a trend is therefore simply $\sqrt{2}$ times the uncertainty of the mean in a given period. Basing the estimates on regression trends adds more noise, since there is additional uncertainty associated with the trend estimation.

Comments by line

11 I would refrain from using the term 'tropical expansion' as there are recent indications that the tropics are actually narrowing (while 'tropical expansion' is commonly used, it is a bad choice of words since it is the subtropics that are in effect expanding). Hadley cell expansion is the subject of this analysis, and is a more appropriate term in this case.

21 State the period of the calculated trend.

41 in some cases there are conflicting results, but not as a rule.

54-58 This is not correct. Davis and Rosenlof (2012) do an excellent job of demonstrating the variance across datasets and methods. The failure to cite Davis and Rosenlof (2012) is particularly upsetting, as this paper was pivotal in convincing the community that there is a need to better constrain estimates of HC expansion. Similarly, in the TropD paper, variance across models and sensitivity to grid spacing are examined. The authors examine variability within a particular dataset, and should better delineate their analysis from previous works.

65 I don't agree. Chemke and Polvani study HC intensity discrepancies and actually specifically state that reanalysis and model extent trends generally agree.

85-92 Over the past few decades, variance across reanalyses in HC extent estimates has increased significantly (Adam et al. 2014), despite "better data". There is therefore every reason to assume that metric uncertainties vary across reanalyses. In other words, estimates based on the ERA5 ensemble cannot be assumed to generally hold for other datasets.

230 Doesn't this contradict the preceding assumption that the reduction in uncertainty is related to improved data quality?

232 There are more stationary waves in the NH but there is significant transient variability in both hemispheres, so it is not clear that this is a valid argument.

References

Adam, O., Schneider, T., & Harnik, N., 2014: Role of Changes in Mean Temperatures versus Temperature Gradients in the Recent Widening of the Hadley Circulation, *Journal of Climate*, **27**(19), 7450-7461

Davis, S. M. and Rosenlof, K. H., 2012: A Multidiagnostic Inter-comparison of Tropical-Width Time Series Using Reanalyses and Satellite Observations. *J. Climate*, **25**, 1061–1078