**Review of 'Signs of climate variability in double tropopause global distribution from radio occultation data' by Alejandro de la Torre et al.**

There is undoubted value in the use of GNSS-RO observations to monitor and understand changes in complex tropopause characteristics including conditions of multiple tropopauses. The premise of the analysis is therefore strong. The authors are to be commended for taking this on and I would encourage them to work further on it. However, there is probably considerable work required for this to be publishable. Concerns relate to the appropriateness of the statistical approach, the lack of physical interpretation of the results in terms of fundamental processes and the overall structuring of the paper including the complete absence of a classical discussion and conclusions section.

I limit below to only major comments given the need for substantial work before this could be publishable.

**Major comments**
1. The consideration of solely double tropopauses is somewhat limiting. There are many interesting complex tropopause cases illuminated by RO and this should be at the very least acknowledged. Also, the sensitivity to the single definition of a double tropopause deployed is an obvious weakness. If you had chosen different objective criteria to define a double tropopause event how would your analysis have differed?
2. Overall paper structure is really far from the classical structure for a paper, that being introduction – methods – results – discussion-conclusion. Interleaving methods and results throughout makes for a very challenging read for a reader with new aspects of methods suddenly being dropped at random points in the text. Rewriting the paper in the more classical way would probably make for an easier read. In particular the lack of a discussion and conclusions means the 'so what' part is almost entirely missing. You need to close by placing your analysis in the broader context, highlight any caveats, and outline some potential future directions and open questions.
3. Figures in general need considerable work for clarity. In particular figure 1 is indecipherable to the reader as presented. This could instead, for example, have been presented as a stacked plot of timeseries by latitude bands N to S with the same vertical axes ranges extending vertically across a whole page enabling a reader to easily ascertain latitudinal variations. This could have avoided trying to find 18 colours which are challenging for most and indecipherable to colour-blind readers. Other figures have similar challenges but Figure 1 is by far the most challenging to comprehend as currently presented.
4. Why were the 29 indicies chosen and why do you expect these to be important in double tropopause behaviour? This married to the lack of physical interpretation is problematic. When you do compare them it currently leaves a reader with a perhaps unfortunate impression that you are proverbially throwing spaghetti at the wall in the hope that some of it sticks. I doubt this was the case but as currently written it is hard to tell on what basis you chose this set and why you think all these might, plausibly, matter. This comes to the point made in the

opening remarks that this is very statistically heavy and you really need more physical understanding in the piece as a whole.

5. In the cluster analysis work from the analysis as shown it is hard for me to really tell that there truly are six distinct clusters. In Figure 3 they just look like cuts driven by the arbitrary selection of six clusters in what is very much a continuum of behaviour with no obvious centring into distinct clusters driven by likely distinct physical behaviour. This is compounded in Figure 6 where in particular cluster 4's distribution suggests this cluster is not driven in any way by the physics with cluster placement ranging across almost all latitude bands.

6. Given significant seasonality in the latitudinal distribution of key aspects of circulation relevant to double tropopauses, the use of a seasonally varying criteria or criteria that track key features from e.g. reanalyses may have been considerably more elucidating. We know that double tropopauses are more common in key physical conditions as you have alluded to. Using a fixed lat-lon distribution when features may be repeatedly transient across such fixed grids on an annual and semi-annual basis probably explains much of the annual and higher harmonics structure in figures 4 and 5. Again, this is highlighting the need to really think about the physics here. The use of a fixed lat-lon grid vs a feature tracking approach e.g. following the sub-tropical and polar jets and the ITCZ throughout the year should be considered in revisions. A feature tracking approach which could be utilised by e.g. using ERA5 diagnostics for features of interest might give a clearer picture than your current fixed lat-lon approach.

7. The multivariate regression really needs much more physical interpretation to be of any value. At present the statistical results are presented and any physical interpretation pretty much left as an exercise for the interested reader. Statistical significance is a necessary but insufficient condition to draw robust conclusions here. It is necessary to understand physically what these results are showing us and what they mean. Why is something leading or lagging and if something is lagging does that mean that somehow double tropopauses are causing that phenomena? There is an absolute need for understanding physically what your results mean here for them to have any scientific value. I can understand how double tropopause features may lag a given phenomena, but I am unsure how to interpret a result saying they are a leading indicator. Table 1 is thus very confusing to me as a reader presently.

8. I am not really sure how I should interpret figure 9 as presented. In particular in clusters 1 and 3 the test RMSE is consistently lower than the training RMSE which makes no logical sense. This may highlight that the cluster definition is not appropriate (see earlier point) and that the behaviour within clusters is non-stationary in interesting ways as a result.

9. Table 2 again you are making the reader do the lifting of the physical understanding as to why these particular modes might matter to these particular clusters. Taken together with Table 1 I have a real challenge thinking how to interpret your results here. You need to help a reader understand how to interpret these combined results.

10. I am unclear why so much of what would nominally be considered key results is left in the supplement and not discussed at all in the main text. I may have

missed it but I failed to note a reference to it and certainly a substantive analysis and discussion of these results.

11. I am always loathed as a native English speaker to make this point as I could never even attempt to write a paper in any language other than English let alone to such as a standard, but the paper overall is a heavy read and either getting a native English proof reader or engaging a native English speaking co-author to help in the rewrite and restructure would be helpful.