We thank the reviewer for her/his comments. Below are our responses in blue.
Here is a summary of the major changes in the revised manuscript:

- We enlarged Figure 4, 5, 7, 8, A1 and A2 to improve readability.
- Figure A1 and A2 instead of showing Lauder lidar now show Mauna Loa so it can be compared against Hilo ozonesonde.
- We added two more figures in the appendix to highlight other datasets (SPURT, TACTS, Summit and South Pole ozonesondes and the Lauder lidar measurements).
- We change the title to: Multi-Parameter Dynamical Diagnostics for Upper Tropospheric and Lower Stratospheric Studies.
- We added a figure showing the range covered by the ozonesondes and lidars in EqL and latitude wrt STJ.
- We added discussions on the differences on sampling.
- We expanded the discussion of reanalysis discontinuities
- We emphasized the future OCTAV-UTLS studies with a figure showing some illustrative examples.

Reviewer 2 suggest that the manuscript is not suitable for this journal. The authors feel the material is appropriate for AMT, but if, after the revisions, the consensus among the reviewers and editor is that it is not, we will consider resubmitting (or transferring if that is possible) to ESSD. In any case, we thank the editor and reviewers in advance because we feel the paper is more complete after their suggestions.

Review by Reviewer 2

The paper is dedicated to several diagnostics for the UTLS studies, which are derived with the JETPAC software and MERRA2 reanalysis data. The authors introduce the SPARC OCTAV-UTLS project, which aims to reduce uncertainties in the UTLS trend estimates by accounting for dynamically induced variability. The authors state that “suitable coordinates should increase the homogeneity of the air masses analyzed together, thus reducing the uncertainty caused by spatio-temporal sampling biases in the quantification of UTLS composition trends”.
The paper has a comprehensive introduction and describes well the objectives of the OCTAV-UTLS project. However, the scientific information content of the paper is low. The paper describes in detail the computing of dynamical diagnostics using the well-established JETPAC software, which is already described in several papers.
We agree on the fact that some of the JETPAC software is documented elsewhere, however, this manuscript offers a complete description of the current algorithm, that is, this manuscript offers a review and a compendium of all the changes / tweaks made since the original developments (i.e., Manney et al 2007 and Manney et al 2011). Originally, this algorithm was developed for satellite measurements but now it can process several type of datasets, including aircraft campaings, ozonesondes, and lidars. We believe there is value in having one manuscript describing this algorithm as opposed to having to read five or more papers to get the up-to-date picture. Our manuscript provides details that were not given in previous papers about the products that are produced at measurement locations, as opposed to the already well-established general JETPAC algorithm used for studies of the jets and tropopauses themselves in reanalysis data.

To emphasis these points we added the following sentences: Originally, these algorithms were developed to process satellite measurements, but they have since been adapted to accommodate a diverse range of atmospheric measurements, including ozonesondes, lidars, and aircraft campaigns. Notably, this is the first time that all these records will be characterized consistently.

We also change ‘For completeness, a full discussion is given here’ to ‘For completeness, a review of the previously published use of JETPAC products for characterizing composition measurements is given here along with an update describing the current capabilities (for example, processing high resolution datasets) and their applications. This review also describes the JETPAC algorithm as applied to characterizing measurement environments, as opposed to referring the reader to Manney et al. (2011) that describes a previous iteration of the algorithms for two satellite datasets, and Manney et al. (2014, 2017, 2021a), and Manney and Hegglin (2018) that use JETPAC products only on the reanalysis grids for dynamical studies. In other words, the previous publications do not give a full view of the current capabilities of JETPAC for characterizing composition measurements.’

Although examples of climatological ozone distributions in different coordinate systems are provided, there is no demonstration of advantages of using the alternative coordinates (references to published papers are not sufficient). In the current form, the paper does not meet the scope of AMT: “the development, intercomparison, and validation of measurement instruments and techniques of data processing and information retrieval for gases, aerosols, and clouds”. The information related to atmospheric measurements, which is presented in the paper, is insufficient for AMT.

The authors (and reviewer 1) believe that the manuscript is in the scope of AMT: it shows intercomparisons and it shows techniques of data processing. We have expanded some discussions and highlighted additional datasets as detailed below.

From my point of view, there are two ways of improving the paper. The first one (preferable) is to add more illustrations and discussion on datasets in various coordinate systems and their agreement.

As requested by reviewer 1 we added two more figures in the appendix to highlight other datasets (SPURT, TACTS, Summit and South Pole ozonesondes and the Lauder lidar measurements).

We also added a figure showing the equivalent latitude and latitude wrt STJ coverage for each of the ozonesondes and lidars.

The reviewer is correct in pointing out that we could have shown some demonstrations of reduced error bars (i.e., due to correctly accounting for ozone gradients) or timeseries. However, those are the topics for OCTAV-UTLS studies that are in progress as described in the summary and future directions section.

To emphasize this, that section was rewritten as shown below. We also included illustrative examples (Figure 10 in the current manuscript), but we note that the actual analysis of those topics is left for future papers.

Future OCTAV-UTLS studies on ozone will:
– Investigate which dynamical coordinates better homogenize the mapped measurements, that is, which ones best segregate air in regions separated by geophysical transport barriers (for example, the tropopause) and thus provide the most comparable results within a given bin for non-coincident measurements. A paper on this topic is already on preparation. An illustrative example is shown on Figure 10-a for the WISE campaign using the relative standard deviation (RSTD, i.e., the standard deviation

divided by the mean) as metric to assess the variability in different coordinate systems. As expected, the dynamical coordinates display a smaller RSTD.
– Explore how sampling biases can confound the intercomparability of these records. These biases will be explored by comparing the ozone reanalysis fields interpolated to the measurement times and locations of the dynamical diagnostics discussed here to the raw reanalysis fields. Additionally, this study will explore which coordinates can, if any, reduce sampling biases. As an example, Figure 10-b shows the MLS sampling biases for January 2005.
– Analyze the impact of dynamical coordinates on quantification of long-term trends. The aim is to identify regions where dynamical coordinates can reduce the uncertainty associated with such trends. As an example, Figure 10-c shows MLS 20S-20N time series at 100 hPa as well as at the WMO tropopause. As shown, the time series at 100 hPa displays much larger amplitude variations that are related to the annual cycle of the 100 hPa pressure surface with respect to the tropopause, which results in sampling different fractions of stratospheric and tropospheric air in different seasons. Interannual differences in these variations are also likely related to changes in the relative altitudes of the 100 hPa isobaric surface and the tropopause.

Another way is shortening the paper and submitting it and the dataset of dynamical diagnostics to ESSD.
We believe the manuscript is within the scope of AMT (since it provides intercomparisons and describes new techniques for data processing). By outlining the methods for data analysis and the complexities of intercomparing the diverse datasets (topics that are within the scope of AMT), this paper also provides necessary foundations for other OCTAV-UTLS studies that are in preparation and will need to cite it.

Reviewer 1 believed the manuscript to be publishable after some corrections. We have implemented those as detailed at the top of this document. We hope that these additions which provide more complete descriptions of the datasets and their analysis and relationships to physical phenomena, are sufficient to change reviewer 2's opinion on the manuscript.

OTHER COMMENTS

- The link to the dynamic diagnostics data does not work.
  The reviewer is correct, due to some JPL changes, we cannot longer host the dynamical diagnostics for the satellites under a JPL webpage. We are currently working to deliver the MLS dynamical diagnostics to a NASA disc. The dynamical diagnostics will be delivered to zenodo.

- It is mentioned that the collection of the datasets used in the paper is limited. With the selected subset of available data, it is probably possible to study some processes in the UTLS. However, for the ambitious objectives of OCTAV-UTLS, the dynamical diagnostics should be provided also for other available satellite and in-situ measurements.
  The reviewer is correct in suggesting that ideally the dynamical diagnostics should be process and made public for other satellites, ozonesondes, lidars, and aircraft campaigns to better attempt to fulfill the OCTAV-UTLS objectives. However, processing the dynamical diagnostic is time consuming and computationally expensive and currently there is no funding to process any more records. A MEASURES proposal was explicitly written to compute the dynamical diagnostics for many limb sounders, but the results of the panel review are still pending. Other avenues of research will be sought to process more datasets but, in the meantime, no more records can be process.

Alternatively, JETPAC can be made as a free software.
Unfortunately, at this time, the JETPAC software cannot be made publicly available due to JPL regulations, but we will explore pathways to make it public in the future.

- P.7 " When computing dynamical diagnostics as discussed in this section it is important to use the same reanalysis fields for all the datasets to be used in a given study." However, possible discontinuities in reanalyses data, which are caused by changes in assimilated datasets (your discussion on Page 8), will also affect the dataset of dynamical diagnostics. When the same reanalysis is used for all datasets, these discontinuities will appear as an artificial drift in the dynamical parameters and thus in evaluated trends (if these dynamical parameters are used for data transformation). When using different reanalyses, one may hope that the timing of discontinuities will be different, and thus the overall evaluation of the trends using multiple datasets will have a reduced reanalyses-related drift. These issues should be discussed/mentioned in the paper.
  The reviewer is right, that discontinuities could affect the interpretation of data. Here, first, we wanted to make sure that the different observational characteristics of the different data sets are related to the same reanalysis fields to ensure a consistency at this point. Those issues were discussed in the paper at the end of the same paragraph discussed in the reviewer comment. However, to emphasis the discussion we broke up the paragraph and added: Thus, given these discontinuities, it is crucial to use the same reanalysis for all datasets throughout a given study. While using only one reanalysis can't completely eliminate the impact of these discontinuities, it can ensure that their effects are consistent across all datasets studied. That is, the timing of these discontinuities is well documented and it can be identified. If more than one reanalysis were used, the timing of discontinuities will be different, complicating the interpretation of such studies. That said, there is also value in repeating a study with meteorological fields from different reanalyses to study the sensitivity of the results to the uncertainties in the dynamical variables as recommended by S-RIP (Fujiwara et al., 2022).

- The influence of sampling patterns should be discussed in more detail. In particular, it should be mentioned /discussed also in the text related to Figure 5 and Figure 7. The distributions for CARIBIC-2 data are different from those for other datasets.
  We added the following after the latitude altitude zonal mean discussion for Figure 7:
  The sampling of each record is evident in the zonal means presented here. MLS, with its dense sampling and near-global coverage (82◦S-82◦N), provides a comprehensive perspective of global ozone fields, capturing the climatological view of the subtropical jet and subvortex jet (black contours, see Figure 6 for reference). However, MLS suffers from a lack of coverage below ∼10 km and coarser vertical resolution compared to the other datasets used here. On the other hand, SAGEIII/ISS offers sampling in the upper troposphere with measurements down to 5 km, with better vertical resolution, but is limited to measurements between 60◦S and 60◦N. The distorted wind contours in comparison to MLS illustrate the greater impact of sampling biases than for MLS.

  CARIBIC-2 measurements are restricted to flight levels, mostly situated at altitudes below 12 km, near the extratropical tropopause. These measurements, due to their high frequency observation rate, capture much finer details that are not resolved by satellite products. As such, it provides an exceptional perspective for studying tropopause-related processes in this region, but its coverage is limited elsewhere. Finally, ozone measurements obtained by ozonesondes and lidars provide superior vertical resolution, but are confined to their specific measurement locations. To enhance their visibility, we have replicated the ozonesonde and lidar measurements in the adjacent latitude bins in Figure 7-top.

The different sampling of each of these datasets may contribute to biases / artifacts and confound the interpretation of the results. To better understand this issue, OCTAV-UTLS aims to characterize the sampling biases of these datasets by comparing ozone reanalysis fields that are interpolated to the measurement times and locations (included in the dynamical diagnostics discussed herein) with the raw reanalysis fields (similar to the approach used in previous studies e.g., Toohey et al. (2013); Millán et al. (2016)).

After the tropopause mapping discussion, we added:

The impact on coverage of the tropopause mapping can be seen in particular in MLS and in CARIBIC-2. Although MLS shows very few measurements below the tropopause in the extratropics in the latitude/altitude view, mapping with respect to the tropopause allows separation of individual measurements that were taken above and below the tropopause. This shows that MLS does on occasion sample below the extratropical tropopause. Similarly, in the extratropics, CARIBIC-2 covers only a few kilometers above the tropopause in the latitude/altitude mapping, but the use of tropopause coordinates shows that it covers air masses that are high above the local tropopause (e.g., above deep folds) and thus samples deep stratospheric air.

In the Equivalent latitude mapping discussion, we added:

Figure 8-left depicts a normalized count of the ozonesonde and lidar coverage in equivalent latitude. The vertical dashed lines display the geographical latitude for the ozonesonde launches or the lidar locations (noting that geographical and equivalent latitude have completely different meanings and implications). Any deviations observed between these lines and the maximum normalized count (i.e., 1) indicate that the respective datasets are situated in a region of significant dynamical variability, which leads to diabatic PV modification and thus a shift in equivalent latitude, such as the South Pole and Summit, Greenland ozonesondes, or the Table Mountain lidar measurements. As shown, the expanded space can cover from 11 to 30^o (determined as the half width of the normalized counts) from the measurement location. Note that, in Figure 7, all other records show coverage improvements, with SAGEIII/ISS covering most of the globe, CARIBIC-2 extending its coverage throughout the northern hemisphere, and MLS covering the entire globe.

In the subtropical jet mapping discussion, we added:
Note that when referring to latitude with respect to the STJ, all instruments show an expanded measurement space, which, is again most noticeable for the ozonesonde and lidar datasets. MLS, with its denser sampling, has full sampling of the region near the subtropical jets, as indicated by the concentric wind contours (black line) around the zero line. Figure 8-right also shows the extent of this coverage for the ozonesondes and lidars.