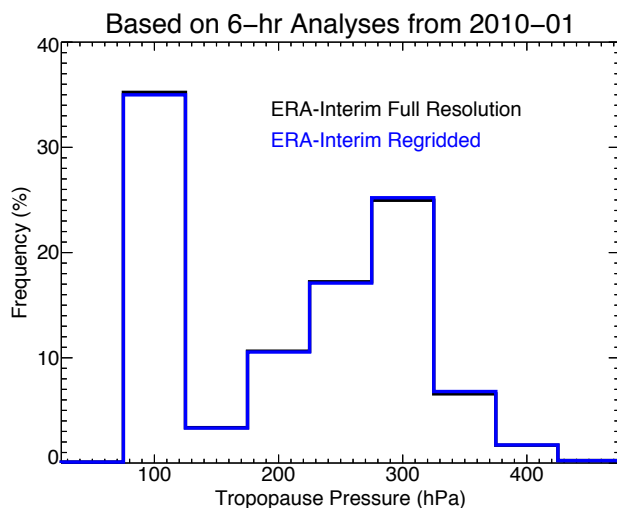


Response to Anonymous Referee #1

- 1) The thermal tropopause itself needs to be assessed for the individual data sets, before analyzing the exchange and probably before regridding (see also suggestions below.) This point is crucial, particularly for the method as applied here. Which role plays the interpolation of the fields for the results, particular for the vertical coordinate and the location of the tropopause altitude?

We now include a comparison of the reanalysis lapse-rate tropopause altitudes in the diagnostics section (see additional response below). With regard to interpolation, we only interpolate in the horizontal dimension prior to applying the WMO algorithm, so the effect of interpolation on the altitude of the tropopause is negligible (see figure here using 6-hr ERA-Interim output for an entire month). This point has been made clear in the revised manuscript at P4 L20-26.



- 2) The authors just perform a spatial classification of STE ‘lateral’ and ‘vertical’, which does not mirror the dynamical processes. For exchange between the subtropics and mid latitudes, where the tropopause a large vertical extent, this might work well. For the mid latitudes they might miss parts of the exchange (see comments below with references) since there is no ‘lateral’ STE per definition of the method. This needs to be discussed as well and potentially lead to a bias e.g. in the fluxes.

I highly suggest including the method of Škerlak et al., 2014, despite differences, since it allows for a further independent comparison also with previous results from literature.

Thank you for these comments. We have recognized following the reviews that our labeling of STE as “vertical” and “lateral” was somewhat misleading. All exchanges across the lapse-rate tropopause are considered in our trajectory analysis if they uphold the residence time and potential vorticity difference criteria; it is not dependent on the geometry of transport (save for the tropopause-relative pressure check, which is a local condition that does not require the exchange to have taken place in the

vertical dimension). We have re-named our transport classifications in the revised manuscript to reflect this issue and acknowledge the more appropriate regional distinction. The new classifications are “tropical”, “subtropical” (previously lateral), and “extratropical”. Since the previous use of “vertical” was meant to provide distinction from our so-called “lateral” exchanges across the tropopause break, its use has been removed from the revised manuscript (except for the general descriptions of individual transport processes in the Introduction).

Also, it is not clear what aspects of the Škerlak et al approach are desired/requested here. We do include a substantial comparison between their STE estimates and ours (i.e., Figs. 2 & 3 and attendant discussion), but note that beyond the use of a 3-D PV labeling methods, our Lagrangian STE approach and theirs is nearly equivalent. The only major difference is the troposphere-stratosphere boundary employed. While we appreciate the labeling method used in Škerlak et al., 2014 (and previous work), we believe it is somewhat impractical given our use of the lapse-rate tropopause in our study and provides unnecessary complexity for the overarching goal of this work: comparison of STE in multiple reanalyses.

- 3) As stated by the authors, one should expect the STE begin mass conservative. This seems however not be the case. Since this is a central point also for the long-term STE time series the authors should also discuss carefully the caveats of their method.

Yes, we do set an expectation that net STE estimates should be near zero as a result of mass conservation. As our method only counts air parcels that are “irreversibly” exchanged, we do miss many transient exchanges along the LRT (which may occur preferentially in one direction). We have made an effort to ensure that the caveats associated with our analysis are outlined well in the revised manuscript. Also, it is important to note that PV-based methods (such as Škerlak et al., 2014) show an imbalance in STE mass flux similar to that in our analysis.

- 4) The thermal tropopause in general, but especially in high latitudes is problematic, how does this affect the results (see e.g. Zangl and Hoinka, 2000)?

We disagree that the lapse-rate tropopause is generally problematic, but do acknowledge that there are issues with its use in the Antarctic polar region (south of 60S) during SH winter, as outlined in Zangl and Hoinka, 2001. We have mentioned this potential issue in the paper, but note that it is limited to ~3 months per year and about 12% of the area of the Southern Hemisphere. Regardless of its limitations in that sense, hemispheric STE fluxes show similar seasonal variability when comparing our lapse-rate tropopause based methods and PV-based methods.

5) How do the results relate to other approaches?

Only Lagrangian studies of STE (very few exist) can be used to compare spatial distributions, which are consistent with our results, aside from those outlined in the detailed comparison with Škerlak et al., 2014 included in the paper (Section 3). Our discussion (updated section 5.2) now includes more quantitative comparison of globally integrated net mass fluxes between our approach and previous ones (see also response to Reviewer 2).

MAJOR: Which role plays interpolation of the fields for the results? Did the authors interpolate also in the vertical? If yes I think a sensitivity for at least one model should be done to assess the effect of interpolation of the STE results.

As outlined above in response to a similar comment, interpolation to a coarser grid was only performed in the horizontal (which was made clear in the revised manuscript). As for its impact on the result of the trajectory calculations to determine STE, there have been many studies in the past that examine this problem (please see the Stohl, 1998 review on this topic). At wind field temporal resolution similar to that available in this study (6 hr), the effects of spatial resolution on the trajectory result are minimal. Temporal resolution (which we cannot change) is considerably more important for reducing errors in large-scale trajectories. Horizontal resolution is more important when the temporal resolution is sufficiently fine (< 4-6 hr). We have included these details on P4 L20-26 in the revised manuscript.

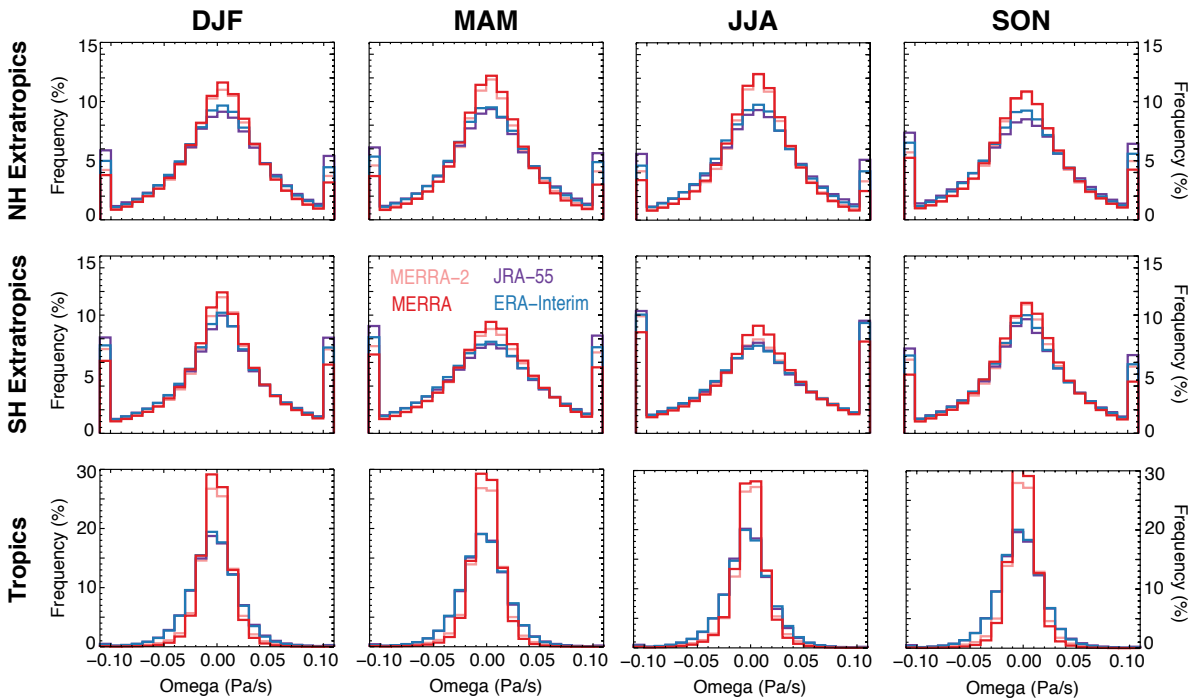
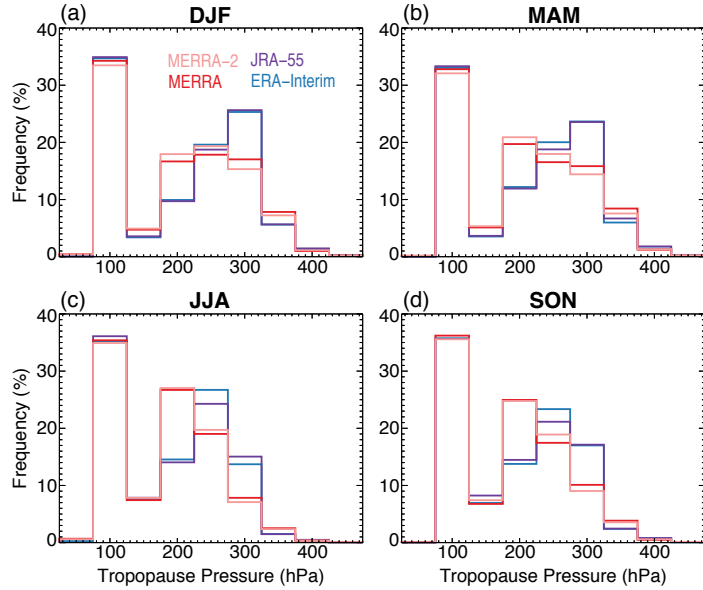
References:

Stohl, A., 1998: COMPUTATION, ACCURACY AND APPLICATIONS OF TRAJECTORIES -- A REVIEW AND BIBLIOGRAPHY, *Atmos. Environ.*, **31** (6), 947-966.

Further the authors find the largest differences between the data sets for the 'vertical' exchange. This is not surprising, since it might be related to differences in the vertical resolution or the variability of the vertical wind in the models. Also the differences in the representation of the thermal tropopause might contribute to these differences, which in turn depends on the vertical resolution of the specific data set. I missed an assessment of this particularly for the extratropics (e.g. a monthly pdf of vertical wind for each month the extratropics)

We have rewritten the diagnostics section in the paper and provided PDFs of vertical motion (separated by region and season) and tropopause pressure (separated by season). These figures are also included here. The differences in the reanalyses are clearer from these comparisons and show consistency among the models with similar vertical grids (as seen in STE). Vertical wind PDFs of MERRA and MERRA-2 should reveal higher frequencies of negative omega (ascent) at the tropopause as a result of TST-dominant exchange in the extratropics, however the TST-Dominant reanalyses do not show a positive skew, rather the shape of the PDFs are

similar throughout the annual cycle among each reanalysis. More information can be drawn from the tropics. PDFs show a slight positive skew indicating a larger frequency of ascent at the tropopause in the tropics. Moreover, the vertical winds are weaker in both MERRA and MERRA-2 compared to ERA-Interim and JRA-55 (STT-dominant).



Since the spatial coordinates play such a crucial role the authors need to systematically assess this: They should add plots (PDFs) of the tropopause height separated for the extratropics (seasonally resolved) and tropics for each data set. This should be done for the original data as well as for the interpolated data to get both, the differences between the data sets and the effect of regridding.

[See previous comments.](#)

For STE: Evaluating the differences $TP_{\text{tropical_press}}$ minus $TP_{\text{extratropical_press}}$ between the different data sets is important since differences of the diagnosed separation between extratropical and tropical tropopause will directly affect STE results (see also comments further below).

[PDFs of tropopause pressure from the reanalyses were shown above. MERRA and MERRA-2 reveal a shallower transition from tropical to extratropical tropopause pressures in each season. A frequency minimum between tropical and extratropical tropopause modes of 150 hPa is shown to occur in each season for each reanalysis \(as outlined previously in the manuscript\). Tropopause altitudes are frequently higher \(i.e., lower pressures\) in the extratropics for the MERRA and MERRA-2 reanalyses. Tropopause pressures are largely consistent in the tropics, with differences between the reanalyses mostly due to slight offsets in the location of vertical model levels \(not apparent here, but note that additional resolution in the PDFs is not appropriate given the native resolution of each reanalysis\).](#)

Criteria: ‘Lateral’ STT: Is exchange across the extratropical tropopause possible, which is not ‘vertical’? How are particles counted, which start in the troposphere, but follow downward sloping isentropes into the stratosphere? These parcels (initially lying below the extratropical TP) descend (e.g. from above the polar jet and are mixed into the adjacent stratosphere above a trough). Such a parcel will descend, but not gain PV? This is not an exotic process and does occur quite frequent (e.g. Pan et al., 2007, Pan and Konopka, 2012, see also Juckes, 2000). How is quasi-isentropic mixing in the extratropics treated? According to the classification no ‘lateral’ exchange is possible if the tropopause is below 200 hPa (i.e. at higher pressures). Also: Why does exchange above and below the jets need to be ‘vertical’ (p.6, L.33)? This is a limitation of the method and needs to be clearly discussed, also in comparison to Škerlak et al., 2014.

[As outlined above, the use of “vertical” and “lateral” was misleading. Since these are three-dimensional trajectories, any geometric evolution can result in STE if the tropopause-relative pressure changes during a parcel’s transit through the atmosphere. What matters for exchange is the passing of a local condition \(i.e., a parcel must lie above/below the local tropopause at its initial time and below/above the local tropopause at its final time\). This could result from transport that is mostly vertical in nature, horizontal in nature and, possibly, that which deviates from the conceptual framework of upward = TST and downward = STT. We have](#)

revised the text where necessary to avoid this confusion, including the renaming of our STE classifications (as outlined above).

SPECIFIC: p.5, L.10-12: It has been shown in many studies (Gettelman et al., and references therein) that in the extratropics away from the subtropical jet very well represents tracer isopleths. This is due to the fact that PV is materially conserved under adiabatic conditions, which is not the case for the LRT. Notably the vertical gradient is included in the PV definition, which therefore inherently includes the thermal definition. Note further, that the -2K/km are an arbitrary definition for the thermal tropopause gradient in a similar way as a fixed PV threshold. Therefore, the above argument is not valid.

This text has been modified a bit here. The intended message is that the advantage of using the lapse-rate tropopause is its ability to characterize the tropopause globally and its common coincidence with the largest chemical gradients between troposphere and stratosphere, while the use of a fixed PV threshold/gradient requires a second definition for the tropics (a limitation this paper seeks to avoid).

p.6, L.19: Isn't convection 'vertical'? How is it considered?

Convection is not a resolved process in the reanalyses due to the coarse nature of their native grid resolutions, and therefore not considered in this study, which was previously mentioned at the end of the Introduction.

p.7, L.15: What is 'systematic' upwelling?

Replaced with "ubiquitous".

p.7, L.17: 'according to our knowledge...LRT method agree more closely with known transport mechanisms'. Please give a few references here, such a statement without references is very superficial. How does this compare with e.g. Juckes et al., 2000? Maybe the LRT method is very well suited for identifying exchange at location of the subtropical jet. At mid latitudes the LRT altitude criteria probably fail in regions?

Done.

p.13, L.20: What is 'equivalent dynamics'?

This is no longer relevant given the significant revisions to the diagnostics section in the paper.

p.13, L.13-22: The arguments are confusing as well as the use of 'dynamical and physical differences': Why is the jet location a dynamical difference, the fold and tropopause physical differences? Both are related to the same physical processes, which control temperature gradients and pressure etc. and finally the location of these structures – based on the representation of physics in the respective reanalysis model.

See previous.

Further: Why can TP altitude lead to changes in STE between the different models? If the tropopause location (and jet, folds) in each reanalysis is the result of differences in the respective model physics, it still might be self consistent within each reanalysis data set. One could get differences of tropopause height, location, jets etc. between different data sets without differences in STE.

[See previous.](#)

Since ‘vertical’ exchange is so important – which role plays the variability of the wind in the data sets?

[See previous responses to similar comments.](#)

p.14, L.13: Vertical and lateral are inappropriate terms to characterize physical processes, it’s just a spatial direction in (Cartesian) coordinates, but STE is more complex as you show. Therefore please change the ‘processes’ to ‘direction’.

[See previous.](#)

p.14, L.25-27: This statement as it stands here is incorrect or at least misleading. In geometrical coordinates the direction could be downward, although the PV change can be positive. This were a TST in the physical view accounting for thermodynamics, but an STT from geometrical aspects.

[This has been revised in line with changes outlined in response to previous comments.](#)

p.14, L.30: How is transport ‘stratified’?

[Replaced with “separated”.](#)

p.15, L.12: Why is poleward transport the same as TST?

[This point has been clarified. This refers to transport across the tropopause break, which is inherently poleward if TST and equatorward if STT.](#)