Response to Reviewer 1

1. Major Comments

My major comment is that the authors should have included detailed analysis for both bias and RMS while validating. While the authors tend to emphasize the RMS statistics, I hope they know that the RMS, as a loss function, gives a relatively high weight on large errors since it is more sensitive to extreme values on long tails/outliers due to the fact that the errors are squared before they are averaged. So, sometimes very few extreme values can completely change the statistics, which is not desirable for this analysis.

In contrast, the bias is sometimes more intuitive because it tells us how much of absolute differences between the radiosonde and the reanalysis. However, the bias analysis is not perfect because the positive and negative biases will cancel out. I would separate the bias analysis into positive and negative, with positive means the reanalysis primary tropopause showing at a higher altitude, and negative means the reanalysis primary tropopause showing at a lower altitude. Meanwhile, adding the frequency (with respect to total samples considered) of positive and negative bias. In this way, we know that on average how frequent and how much the reanalysis would overestimate/underestimate the tropopause height. I think this detailed analysis is more meaningful to the community.

In this sense, it is fair to always include both bias and RMS error analysis. I think Fig. 2 should include bias on the first panel and RMS error on the second panel. For each panel, different shapes represent different months, but please do include one more statistics for all-season averages. Then, add another figure that repeats a similar analysis for the double tropopause. Having the easy visualization of the statistics, still, keep Table 1 of detailed numbers for easy reference.

Thank you for these comments. We now separate the bias analysis into positive and negative, which is listed in the new Table 1. Bias analysis is also included for primary tropopause altitudes and double tropopause frequency within different latitude bands, with statistics for all-season averages added (new Figures 2 and 3). Necessary changes have been made in Section 3.1 in the revised manuscript.

Another major comment is on the accuracy of IGRA data, and its ability to precisely document the lapse-rate tropopause is crucial for this study. It helps if the authors could iterate in more details on how the $<=$ 50m vertical resolution of raw observation are eventually reported in only 1.5-2.5 km vertical resolution at the UTLS (although in the revised manuscript the authors changed to $>$ 1km). The "$>$ 1 km" is still less desirable for studying the vertical variability of temperature records - it will miss effects of both gravity waves and the Rossby waves acting on the temperatures. Given the reported resolution, why not using GPS/COSMIC temperature records that has a better coverage? The focus of the paper is from 5-20 km, in which COSMIC is totally capable of seeing waves on temperatures.

If I understand it correctly, Figs. 1-2 and Table 1 are the only places that the authors performed apple-to-apple comparison by collocating the reanalysis to the radiosonde locations. For all other analysis, the authors just reported trends inferred from gridded results at each latitude-longitude box, so the results could be biased by sampling sizes. So, the first part of validation is more meaningful to my sense. That said, personally I am not interested in the trends reported. For example, what does a trend of +/- 50 m/decade in primary tropopause mean? What does a positive trend of double tropopause frequency mean in specific reanalysis? Unless you can elucidate the possible cause of trends with proof, I dont think the trend numbers themselves have significant meaning. On the positive side, the fact that different reanalyses showing different trends is meaningful in that they imply how unreliable the reanalyses are as to the tropopause analysis. This makes me wondering if it is necessary to include the trend analysis, especially in such a large portion of the paper. If I were the authors I would report the bias and errors in more details to help the community to understand the different performances of the reanalyses.

There is perhaps some confusion on the typical resolution of the IGRA profiles here. In the discussion paper, it is stated that reduction of the radiosonde profiles to mandatory and significant levels only *can* result in vertical resolution larger than 1 km, but this is a worst case scenario. As Figure 1 demonstrates, the vertical resolution of the IGRA data is often finer than this. We do not believe additional detail on the process of reporting mandatory and significant levels beyond what is provided in Section 2.2 is necessary, but we have clarified a few points there in the revision.

As for the suggestion to use GPS/COSMIC temperature profiles to investigate tropopause characteristics, we would like to do that in the future. However, an obvious shortcoming of the GPS/COSMIC temperature records is the limited temporal coverage, which is not suitable for studying the long-term changes in tropopause characteristics (the primary of the focus of this paper). We have added acknowledging this on P5, L21 of the revision.

Finally, the purpose of focusing on tropopause trends is motivated by the review given in the Introduction. Namely, tropopause altitude trends are believed to be an indicator of climate change as increases in tropopause altitudes often occur with increases in tropospheric temperatures. In addition, double tropopause occurrences provide a physical perspective of UTLS dynamics (most notably STE). Evaluating long-term trends provides a unique insight into these processes. Comparing model trends with observed trends are also another method of model validation, so we believe analysis of trends is well-justified and relevant to the scientific community (as also evidenced by the remaining reviews of the paper).

As for elucidating the sources of the trends, we have drawn on complementary results from other recent efforts in Section 4. We have also expanded some of the discussion on these trends there by considering potential physical and dynamical sources (most on P13 of the revision).

A last comment is that I do hope the authors could put more emphasis on the physical meaning/causes of the (large) differences among different reanalysis. So, beyond the vertical resolution, could there be any other reasons that caused the discrepancies? The current version seems to be less scientific and more like a technical report.

We have added text and additional analysis evaluating the effects of vertical resolution (see Tables 1 and Figures 2 & 3 of the revision). Apart from vertical resolution, it is quite difficult and beyond the scope of this study to identify the reasons for oftentimes subtle differences between the tropopause trends in the reanalyses. The tropopause reflects the combined impacts of a long list of choices in model design and assimilated data, so elucidating the role of each in controlling trends in tropopause characteristics is a daunting task. Vertical grid spacing is an ideal target for initial evaluation, given that the tropopause definition depends greatly on it. Thus, we have limited our detailed evaluation to this single source in the revision. To examine the role of other aspects of the model design, comprehensive sensitivity studies using a single modeling system are likely required.

2. Minor wording comments

1. P1L8, attributed > attributable

Corrected.

2. P1L9: observations

Done.

3. P1L9: and reanalyses > and the reanalyses

Done.

4. P1L9: analysis period

analysis period has been clarified.

5. P2L10: > the UTLS composition

Not changed.

6. P1L13-15: this sentence doesn't make sense

Revised to improve clarity (P1, L12-14 of the revision).

7. P3L6-7: this makes sense because of the existence of the ozone layer, but can you be more specific about it?

This point has been clarified (P3, L4-8 of the revision).

8. P4L1: I dont understand the logic here. If the radiosonde data is so limited, why bothering using them instead of COSMIC data? Plus, this part sounds like belonging to the discussion part.

See previous response to similar comments.

9. P5L10-12: all reanalyses are reported in sigma or eta coordinates. From conversion you might get temperatures on pressure levels easily, but how did you

get them in altitude coordinates? Did you use simultaneous geopotential heights to interpolate the data? Be more specific about how you preprocessed the data.

This point has been clarified at P5, L12 of the revision.

10. P5L16: > quality-controlled

Corrected.

11. P5L23-24: one comment is that this linear interpolation doesn't change the shape of the profile, at all. So, you typically end up with the same value without doing interpolation.

An interpolation is needed to verify the second criterion of the WMO definition and the criterion for identifying multiple tropopauses. We have clarified the value provided by interpolation in Section 2.2.

12. P7L18: tropopause altitudes in MERRA-2 > primary tropopause altitudes in MERRA- 2.

Done.

13. P8L16: is maximum > its maximum

Done.

14. P13L11-12: how did you reach this conclusion?

Similar to Rossby wave breaking leading to transport of tropical UT air into the extratropical LS, double tropopauses can be formed by equatorward transport of extratropical LS air into the tropical UT during wave breaking events. We have added a relevant citation to this conclusion [Liu and Barnes (2018)].

References:

Liu, C., and Barnes, E. A. (2018). Synoptic formation of double tropopauses. Journal of Geophysical Research: Atmospheres, 123, 693–707.

Response to Reviewer 2

1. General comment

1) The fairly large bias in MERRA-2 is interesting and I was surprised that it didn't receive more attention by the authors (at least not in the writeup). After all, this is a (re)analysis, i.e., it includes a modern data assimilation scheme, presumably assimilating the radiosonde observations that here used as a reference. So my expectation was that all modern reanalyses essentially reproduce the tropopause. Fig. 1 furthermore stimulates suspicion: how can a reanalysis have such large temperature biases (> 5 K!!) in the upper troposphere? Without labelling I would have guessed that this is a free-running model. Don't you expect all modern reanalyses to very closely agree about temperature in the upper troposphere? This is the case between the other three products: ERA-Interim, JRA-55, CFSR. Is this simply an outlier example or do you often find such large biases in MERRA-2? Is this something that's documented in the literature? To be honest, if this is a robust bias in MERRA-2, then this product shouldn't be used for UTLS studies . . . in any case, this requires more discussion by the authors.

After careful re-evaluation of the MERRA-2 fields we were using for the profiles in Figure 1 (added quickly after the request during initial review before passed on to open discussion), we discovered that the wrong reference levels were used. Instead of using the pressures and altitudes in the middle of the model layers that correspond to the temperatures, we were using the pressures and altitudes of the model levels (the edges of the layers). This resulted in an artificial displacement of the profile of approximately 500 m in the UTLS. We have corrected this error in Figure 1 and the remaining analyses in the paper, for which it had little impact on the results (except for the bias analysis). The revised analyses clearly show that MERRA-2 is consistent with the remaining reanalyses in its representation of UTLS temperatures. Many thanks to the reviewer for emphasizing this point.

2) Vertical resolution is mentioned at many places to potentially explain differences between radiosondes and reanalyses. Isn't this easily testable? You could degrade the radiosondes to the model resolutions and see if that really explains the differences. You could even study some of the characteristics (e.g., double tropopause frequency) as a function of vertical resolution by gradually degrading the radiosonde data. Perhaps the authors have already tried this, in any case, I would strongly suggest to include corresponding results / discussion in the paper.

Thank you for the suggestion. We have degraded the radiosonde observations to the vertical grid of each model and recomputed the bias and RMS differences. Bias and RMS differences in instantaneous primary tropopause altitudes show little sensitivity, but large reductions in both are found for double tropopause frequencies. This point has been clarified in Sections 3.1 and 4 in the revised manuscript and reflected in the revised analysis presented in new Figures 2 & 3 and Table 1.

2. Minor comments

page 2, line 16: "uncertainty that is comparable to the vertical resolution of the model" this makes intuitive sense, but is this a priori clear given that you interpolate between levels for the tropopause calculation?

The value given by interpolation of the temperature profiles has been clarified in Section 2.2. The interpolation only assists in routinely satisfying the second criterion of the WMO definition and the criterion for identifying multiple tropopauses. Therfore, yes, we do expect it to be clear a priori that uncertainty should be comparable to the vertical resolution of each model.

page 2, line 19: the lapse rate is equal to minus the vertical temperature gradient

Corrected.

page 2, line 28-29: Anel ref's

Reference has been added at P2, L25-27 of the revision.

page 3, line 7-8: sentence doesn't work like this; how about: "PV, which is conserved . . ., is commonly used for transport studies in the extratropics and often used to define a dynamical tropopause . . ."

Done.

page 3, line 10: "threshold used varies considerably" seems like an exaggeration (I'd suggest to remove "considerably"), note a lot of the STE studies (e.g., Wernli group and others) use 2 PVU and this value seems to be used mostly

Done.

page 5, lines 11-12: these are somewhat subjective choices have you checked the corresponding sensitivity? E.g., are the results sensitive to obtaining tropopause levels from the native horizontal and vertical grid, and interpolating to the 1-by-1 lon-lat grid afterwards? Im also not sure I understand the purpose of oversampling to the 200-m grid in the vertical for tropopause identification please provide rationale (relevant for line 24 as well).

We have evaluated the sensitivity to these choices and it is negligible. Interpolation in the horizontal dimension has no effect on the tropopause other than reducing the level of horizontal detail (which is advantageous for apples-to-apples comparisons of the reanalyses and is how we have locally archived the data for long-term use). Some text has been added to reflect the lack of sensitivity to the choice of synoptic time here (Section 2.1). In addition, see previous response for detail on the need for vertical interpolation prior to tropopause identification.

page 6, line 24: how do you assess whether data points are roughly evenly distributed?

We checked the length of time gaps in the tropopause altitude time series for each station, and selected the stations with maximum gap duration less than 5 years (there were only 59 stations included with gaps longer than 3 years and these were manually evaluated to confirm there were no deleterious effects on the trend analysis - e.g., missing long time chunks at the beginning and end of the 35-year analysis period). This point has been clarified at P2, L33 of the revision.

page 7, line 8-9: do you do this separately for the two hemispheres? How do you then handle the equator, which in the relative coordinates "moves" around?

Yes. For plotting, any data extending beyond the equator is trimmed. We have added some clarifying points in Section 2.4.

page 7, line 13-14: so here you suggest that you do use the native model grids for tropopause calculations, in contrast to the description on page 5 please clarify

This point has been clarified at P7, L30-31 of the revision.

page 7, line 18: is this bias a function of latitude?

This bias is derived from global observation, the variation with latitude has been included in new Figures 2 and 3.

page 8, discussion of Fig. 2: have you considered normalizing these RMS differences by a measure of internal/natural variability (e.g., interannual standard deviation)? Larger RMS differences would be expected in regions with larger internal variability, so part of the latitudinal differences could be related to different internal variability.

Indeed, the large internal variability can result in large RMS error in some regions, such as the extratropics. The variability of tropopause altitude in these regions is mainly attributed to the subtropical jet shifting latitude, which is associated with north-south migration of the tropopause break. We have not attempted to normalize these RMS differences in the revision, but have expanded the bias analysis to reflect points raised by other reviewers.

page 9, line 5: over the Atlantic trends are larger at the edges of the tropics compared to the equator, which stands in contrast to the statement of "uniformly upward trends throughout"

This has been changed to "larger upward trends".

page 10, bottom (Figs. 7, 8): not sure these Figures need to be included in the paper, perhaps as supplement is enough? They don't look that much different from the Eulerian versions (as the authors remark) and aren't discussed much either.

We believe the differences between these tropopause break-relative analyses and the Eulerian analyses, though small in some respects, are important to show in the paper and to the discussion included (despite the fact that it is relatively brief).

page 11, bottom paragraph: this discussion based on differences in how O3 is handled is useful and should be extended a bit: notably, ERA-Interim and MERRA-2 are very different in this regard with ERA-Interim using a climatological O3 product in their radiative scheme and MERRA-2 using its own O3 field so the effect of O3 on the tropopause and its trends will likely be very different between these two reanalyses.

The description of differences in ozone assimilation between reanalyses has been expanded beginning at P13, L34 of the revision.

page 11, line 33: please clarify that you are referring to anomalous upwelling and downwelling (the full residual circulation is still downward over the polar latitudes)

Corrected.

page 12, line 28: awkward sentence structure (Significant trends . . . were found to be increasing . . .) - please modify

This has been changed to "Significant increasing trends in double tropopause frequency were found nearly everywhere in the radiosonde observations ...".

Response to Juan Añel

- in page 1 line 23 I miss a citation to Anel et al. (2006). This work also deals with the trends from radiosonde data and indeed it will be useful to discuss some issues later in the paper;

Thank you for the suggestion. The citation has been added.

- in page 2, after line 17: usually there is some confusion on the issue of definition of the tropopause. Words have meanings and being fair it only exists one definition for the tropopause, the one established in 1957 by the WMO. Others are criteria to approach the behavior of the tropopause or UTLS transition according to the best fit for different studies, campaigns, etc. This does not change the reality of the complex atmospheric behavior, but using the right words is useful for those not so familiar with the topic that could waste time looking for formal definitions that do not exist anywhere. Therefore in line 18 it is not The conventional tropopause definition but The tropopause definition;

This has been changed to "The *original* tropopause definition" to retain useful context for discussing alternative definitions in the remainder of this paragraph (beginning P2, L13 of the revision).

- in page 2, line 22: in some way linked with the previous issue, I do not think that it is correct to say that there are exceptions to performance. Simply there are regions of the Earth where the UTLS structure is so complex that there is not a tropopause or transition troposphere/stratosphere as such. You mention one case where this behavior is mostly driven by the very specific tropospheric radiative balance during the austral winter. But it is not the only case. The same happens in the third-pole (the Tibetan Plateau) but because of dynamical reasons. There unstable mix of air can make impossible to get a troposphere-stratosphere distinction because of the high altitude of the plateau and its radiative balance (see Chen et al. 2013 and Chen et al. 2016);

Mentions of "performance" were removed and complex, layered stability structures were also acknowledged (P2, L18 of the revision).

- page 3 line 13: indeed fifteen years before Hoinka et al. (1998) had clearly established that the usual values of 1.6 PVU introduced in a campaing in the 1980s or the popular 2 PVU value underestimate the reality of the tropopause height (obviously in extratropics and polar regions);

The references have been cited, and text has been changed a bit at P3, L13 of the revision.

- subsection 2.1 "Reanalysis output": for the purpose of this work, more relevant than this information (vertical levels and top) is to know the distribution of levels (or dz) be- tween 200 hPa and 50 hPa. I would recommend to the authors to focus the description here on this layer. This will enable them to simplify the understanding and discussion of results later, for example in section 3;

The information of model vertical resolution in the UTLS has been added in Section 2.1, as well as a more direct reference to the Fujiwara et al reanalysis comparison paper.

- page 6 lines 5-10: this is a good exercise to guarantee representation with a case study. But this had already been proved by Antuna et al. (2006) using other station at a quite similar geographical location. I recommend to cite the work to add extra support and to include in the text the coordinates for Corpus Christi (unless I have missed them);

The coordinates for Corpus Christi have been added to the text (P6, L13 of the revision). Citation has not been added because this illustration is dataset specific (i.e., showing the level of detail between full-resolution data and reduced resolution data in IGRA) and the Antuna paper focuses only on mandatory-level radiosondes and the impacts of missing mandatory-level data for climatological analyses.

- in subsection 2.4 you state the 35-year analysis period. I have not got clearly what is the period of study: 1979-2015? This is 37 years. 1981-2015?. Please, clarify it;

We have added a parenthetical reference to the time period analyzed here (1981-2015) to remind the reader (P6, L28 of the revision).

- page 8, lines 15-16: there is another basis for this (one of them briefly mentioned in the paper), the competing phenomena of tropical widening where the tropical tropopause overlaps the extratropical one and the horizontal meridional entrainment of extratropical air to tropical regions (Wang and Polvani, 2011; Ael et al. 2012; Castanheira and Gimeno, 2011).

Text modified by also acknowledging double tropopause seasonality (P9, L1-3 of the revision).

- subsection 3.2, first paragraph: this is in agreement with the results for the Scenario 1 studied by Ael et al. (2006). That is, raw series without data homogenization. Thought IGRA solved several of the problems that existed in CARDS, here you do not perform any change-point detection technique and this restricts the validity of your results. I think that the issue of not undergoing change-point detection deserves to be mentioned here and that a comparison in the text with the values obtained by Ael et al. (2006) and Santer et al. (2003a,b) would be good as it would enable readers to get a more complete picture of the state-of-the-art.

The point on Siberia deserves special attention in my view: this is also in agreement for with part of the Scenario 1, and with Scenarios 2 and 3 of Anel et al. (2006). Here I would point out two different issues:

1. some of the radiosonde series in this region show up to a 1% significant correlation with the Northern Annular Mode, this could explain partial regional trends. But as soon as in the 1960s Makhover reported that this region has a special behavior in comparison with similar latitudes in this hemisphere (check Antuna et al. 2009 or the original Russian books cited therein);

2. no doubt it deserves a deeper analysis with data homogenization techniques, but there is a potential reason that could explain bias (be aware that I talk about bias not changes in trends) over the region corresponding to the former Soviet Union. This rea- son is the use of different radiosondes with very different equipment than the extended Vaisala RS80/RS90 radiosondes for other parts of the world. A quick check of the metadata in IGRA shows how some stations over the period 1980-1990 there was up to 4 or 5 changes of radiosonde model (changes, not simple updates) and in some of them radiation corrections in 90s. This kind of problems with soundings over Russian territory with frequent radiation corrections was also pointed out by Makhover (again see Antua et al. 2009). This could have an impact on any trend computed. Therefore any statement on trends without change point detection and data homogenization should be accompanied of one on the limitations of the data analysis.

Thank you for these comments. A comparison in the text with the values obtained by Anel et al. (2006) and Santer et al. (2003a,b) has been added at P9, L17-19 of the revision. In addition, the limitations of not using change point detection or data homogenization have been acknowledged at P7, L21-25 of the revision.

- subsection 3.3, last sentence: I think that it could exist a partial explanation for this behavior in Fig.4 for CFSR. This is my hypothesis: as it has been proved by Anel et al. (2008) in presence of multiple tropopauses the first lapse rate tropopause (LRT1) is lower than when a single tropopause exist and multiple tropopauses are not present. As Xian and Homeyer show CFSR has lower bias and increased resolution at UTLS levels. This enables this dataset to better represent a bigger number of multiple tropopause events. Having more multiple tropopause events means that an increasing proportion of lower LRT1 cases should be found. This should be more clear in critical regions for the detection, such as subtropics. Therefore the positive trend in the frequency of multiple tropopauses and lower bias of CFSR would be driven an increased frequency of lower LRT1.

Nothing changed. Comparing tropopause altitude trends to the double tropopause trends in CFSR, there is no significant increasing trend in double tropopause frequency in the extratropics where decreasing tropopause altitude was found. Thus, the connection between the decreasing primary tropopause altitudes and increasing double tropopause events mentioned by the referee is not robust in the extratropics for CFSR. Moreover, the remaining reanalyses show increasing double tropopause frequencies and increasing tropopause altitude in most regions.

- page 10, lines 9-15: this is exactly what is stated in Castanheira et al. (2009) (Fig. 8) using IGRA data and a probable consequence of the energetic modes at UTLS levels. I think that the numbers here obtained should be compared to their ones and the work cited.

Thank you for bringing this work to our attention as we were not aware of this double tropopause trend analysis. We cited the work and compared our results to their trends of double tropopause frequency in two latitude-bands (30-60N and 30-60S) at P11, L11-13 of the revision.

- page 13, line 14: I do not think that ”found” is the right word here. To be fair beyond the useful contribution on comparison between state-of-the-art reanalysis, the other results here presented only confirm previous findings existing in the literature and it should be acknowledge in this way.

Replaced “found” with “shown” here (P15, L8 of the revision).

- Table 1: I understand that values in this table are computed using all the stations, independently of the hemisphere. This could provide a sense of average changes, but if you present the results for months representative of seasons, what is the point on mixing NH and SH stations?. Doing such thing does not let to appreciate the true seasonal change. In my view exposing only the values for extratropical regions of one of the hemispheres would be the right way of doing it, as there is no point on including the tropics because of the lack of seasonal variability. Moreover Double tropopauses are a phenomenon with strong seasonal dependence associated to extratropical wintertime UTLS baroclinicity (Castanheira et al. 2009) and therefore the same reasoning applies.

Good points. We have removed the seasonality from Table 1 and shown the total evaluation numbers only. The new Figures 2 and 3 summarize the results for season and location (extratropics, subtropics, and tropics).

Response to Ferreira

1. Main aspects:

1) Concerning the reproducibility of results, the paper lacks information about the radiosonde data used in the study: how were IGRA stations selected in the first place? The number of selected stations (317) and the corresponding amount of observations for 1985-2015 are given later in the results section, with their approximate locations shown in the Figures. But IGRA (version 2 released in 2016) contains temperature data from 800-900 radiosonde stations within the studied period. Nothing, however, is said about the choice of stations, concerning the homogeneity of time-series in terms of temporal and vertical features (i.e., leaving aside the much more difficult problem of instrument biases): temporal regularity and continuity; vertical resolution around the tropopause.

We selected the radiosonde observations based on both complete vertical profiles and the homogeneity of time-series, as we had previously outlined in Sections 2.2 and 2.4. We have added a few clarifications to these sections to emphasize some key points related to analysis of the radiosonde data.

2) A linear interpolation to a 200-m regular vertical grid was applied prior both to radiosonde and reanalysis temperature data before tropopause identification. The authors claim this was done in order to enable reliable tropopause identification. This phrase is potentially confusing to the reader. Evidently, an interpolation is needed to verify the second condition of WMOs definition of first tropopause, as well as to look for a second tropopause. But a linear interpolation simply does not change the lapse rate between the known data points. So, the estimation of the first and second tropopause levels is essentially limited by the resolution of data as the authors in fact recognize in other parts of the paper. The gain resulting from the interpolation scheme should be explained to make this point clear.

Thank you for identifying an opportunity to improve clarity. As correctly inferred, the value gained from linearly interpolating the radiosonde data to a higher-resolution regular grid spacing is to enable thorough evaluation of the second WMO criterion and the criterion for identifying multiple tropopauses. We have clarified these points in Section 2.2.

3) Radiosonde data were analyzed at the principal synoptic hours, 0000UT and 1200UT, whereas reanalysis data were analyzed only at 0000UT. This means that half of the time-zones on the global reanalysis fields of temperature (at latitudes outside of the polar regions, after averaging over one or more years) is represented by daylight times, while the other half is represented by nocturnal times. In this respect, in Figs. 3-6 it is not clear why some radiosonde stations show 0000UT average values while others show 0012UT values, since reanalysis-derived values refer always to 0000UT. Also, considering the diurnal variations of the tropopause

height, it should be explained how the radiosonde-reanalysis tropopause differences listed in Table 1 were exactly calculated.

Although there is a diurnal cycle of tropopause height, the long-term tropopause trends from the reanalyses at different synoptic times are consistent (not shown). The comparisons listed in Table 1 are based on 00 UTC profiles only. This point has been clarified in Section 3.1.

4) Although not obligatory, to be more informative Table 1 should depict hemispheric seasons. Or perhaps individual months, but then restricting to North Hemisphere, where the amount of radiosonde data (used as reference to errors) is much larger there than in the South Hemisphere.

Rather than restricting values in the table to North Hemisphere only, the new Figures 2 & 3 satisfy this suggestion.

5) The calculation of tropopause altitude needs a bit of clarification: is moisture included in the hypsometric equation? Tropopause altitude refers to geometric altitude or geopotential altitude?

Before tropopause identification, geopotential height was computed for each reanalysis model-level output using the moisture-included hypsometric equation. Therefore, tropopause altitude refers to the geopotential altitude. This has been clarified in Section 2.3.

6) Maybe the large discrepancies between the results obtained from CFSR and the other reanalysis models (seen in all plots) deserve a slight explanation.

We have expanded discussion of trends and their potential ties to physics/dynamics in the Conclusions and discussion section. Some additional analysis was included, but the source of the discrepancies in CFSR relative to the remaining reanalyses remains unclear.

2. Secondary aspects:

P2, L19. Where it reads (. . .) (also known as vertical temperature gradient) (. . .) it correctly should read (. . .) (negative of the vertical temperature gradient) (. . .)

Corrected.

P4, L1. (. . .) since they are only launched from land masses. Considering the radiosondes launched on whether ships and ships of opportunity (even if not used in the study) it should be better to write (. . .) since they are mostly launched from land-masses.

Good point. Corrected.

P4, L6. Reanalyses assimilate global high-quality observations (. . .). Do not forget to mention other observation platforms besides radiosondes. Moreover, I doubt that all observations assimilated in reanalysis models are of high-quality. A meteorological reanalysis is supposed to deal with inaccurate and incomplete observations to some degree. Quality-controlled is closer to reality.

Replaced with “quality-controlled”.

P4, L 30. I don't understand the words a physical perspective of the UTLS. I suppose that the authors point is that their paper provides an evaluation of reanalysis-model performance regarding the UTLS temperature structure.

Since there is a close correlation between double tropopause occurrence and STE events and the tropopause is a physical attribute of the atmosphere, tropopauses can be used to diagnose UTLS dynamics. The use of the term "behavior" seems to have been the source of confusion here, so we've replaced it with "dynamics" (P4, L31 of the revision).

P6, L12. Thus, we are confident that IGRA data are suitable for tropopause analyses following the methods employed here. How can you tell, from a demonstration with two random soundings from one site? The study uses nearly 10^5 soundings from over 300 radiosonde stations! The above assertion is not acceptable. Although Fig. 1 serves the purpose of illustration of the idea, paradoxically, expressing here some uncertainty would give more confidence to the reader.

Excellent point. We have added a few clarifying bits of information here to address this issue. We did not limit this type of evaluation to a single station and did randomly select from alternative locations and time periods where we had access to the full resolution data. The point of this comparison is to demonstrate that mandatory and significant levels are sufficient for tropopause identification. We have acknowledged that results for alternative locations and times are consistent with that shown here and that differences in tropopause identifications between full resolution and reduced resolution profiles are $\leq$100 m (P6, L16 of the revision).

P7, L14-16. It's not totally clear whether Fig. 3 (and so on) uses only four months per year or not.

It is stated throughout the paper that trend analyses are based on monthly mean fields. Since this was not a common source of confusion for the reviewers, we have decided that additional clarification is unnecessary.

P13, L16. (. . .) increases in primary tropopause altitude are associated with a warming climate (. . .). The suggested connection is supported by a very few modeling experiments until now. I'd replace "are" by something less assertive like "is probably" or "is believed to be".

Replaced by "is believed to be".

Fig. 6 and Fig. 8. If possible, the color scale legend "Double tropopause frequency" should be changed to "Double tropopause trend".

These legends have been changed to "Double Tropopause Frequency Trend".

Global Tropopause Altitudes in Radiosondes and Reanalyses

Tao Xian[1,2] and Cameron R. Homeyer[1]

[1]School of Meteorology, University of Oklahoma, Norman, OK 73072, USA

[2]School of Earth and Space Sciences, University of Science and Technology of China, Anhui, 230026, China

Correspondence: Cameron R. Homeyer (chomeyer@ou.edu)

Abstract. Accurate depictions of the tropopause and its changes are important for studies such as stratosphere-troposphere exchange and climate change. Here, the fidelity of primary lapse-rate tropopause altitudes and double tropopause frequencies in four modern reanalyses (ERA-Interim, JRA-55, MERRA-2, and CFSR) is examined using global radiosonde observations. In addition, long-term trends (1981-2015) in these tropopause properties are diagnosed in both the reanalyses and radiosondes. It is found that reanalyses reproduce observed tropopause altitudes with little bias (typically less than $\pm$ 150 m) and error comparable to the model vertical resolution. All reanalyses underestimate the double tropopause frequency (up to 30 % lower than observed), with the largest biases found in JRA-55 and the smallest in CFSR. The underestimates in double tropopause frequency are primarily attributable to the coarse vertical resolution of the reanalyses. Significant increasing trends in both tropopause altitude (40–120 m per decade) and double tropopause frequency ($\geq$3 % per decade) were found in both the radiosonde observation and the reanalyses over the 35-year analysis period (1981-2015). ERA-Interim, JRA-55, and MERRA-2 broadly reproduce the patterns and signs of observed significant trends, while CFSR is inconsistent with the remaining datasets. Trends were diagnosed in both the native Eulerian coordinate system of the reanalyses (fixed longitude and latitude) and in a coordinate system where latitude is defined relative to the mean latitude of the tropopause break (the discontinuity in tropopause altitude between the tropics and extratropics) in each hemisphere. The tropopause break-relative coordinate facilitates the evaluation of tropopause behavior within the tropical and extratropical reservoirs and revealed significant differences in trend estimates compared to the traditional Eulerian analysis. Notably, increasing tropopause altitude trends were found to be of greater magnitude in tropopause break-relative coordinates and increasing double tropopause frequency trends were found to occur primarily poleward of the tropopause break in each hemisphere.

1 Introduction

The tropopause - the boundary between the often unstable, convectively dominated troposphere and stably stratified stratosphere - is an important boundary for many studies in the atmospheric sciences. For example, long-term changes in tropopause altitude are considered to be an indicator of climate change (e.g., Santer et al., 2003a, b; Añel et al., 2006; Xian and Fu, 2017). Based on radiosonde and satellite observations, previous studies show a significant global rising trend in the tropopause altitude during the last several decades (e.g., Seidel et al., 2001; Santer et al., 2003a; Seidel and Randel, 2006; Feng et al., 2012). Santer et al. (2003a, b) employed output from climate model simulations to assess tropopause trends and attributed the long-term increase to increased greenhouse gases and stratospheric ozone depletion, which leads to a warming of the troposphere and a cooling of the stratosphere. These changes in atmospheric composition enhance the meridional temperature gradient in the upper troposphere and lower stratosphere (UTLS), which in turn strengthens the subtropical jets (to maintain thermal wind balance) and accelerates the Brewer-Dobson circulation (BDC) (e.g. Held, 1993; Kushner et al., 2001; Birner, 2010b; Butchart, 2014; Sioris et al., 2014; Abalos et al., 2015; Fu et al., 2015; Ploeger et al., 2015). The structure and variability of the tropopause also plays a key role in stratosphere-troposphere exchange (STE) studies.The fidelity of the tropopause altitude directly impacts the quantification of STE because the troposphere-stratosphere boundary itself is fundamental to identifying an exchange event and its impact on UTLS composition (e.g., Xian and Fu, 2015; Liu and Liu, 2016; Boothe and Homeyer, 2017). Moreover, the temperature of the tropopause, especially in the tropics, regulates the transport of water vapor (a powerful greenhouse gas) from the troposphere to the stratosphere (Holton et al., 1995; Fueglistaler et al., 2009).

Several tropopause criteria have been used in prior studies and are based on thermal, dynamical, and chemical characteristics of the atmosphere (WMO, 1957; Holton et al., 1995; Kunz et al., 2011, 2015; Pan et al., 2004). The original tropopause definition is based on the temperature lapse rate (the negative of the vertical temperature gradient) according to criteria put forth by the World Meteorological Organization (WMO, 1957). This lapse-rate tropopause is globally reliable and found to commonly coincide with the sharpest stability and chemical transitions between the troposphere and stratosphere (Pan et al., 2004; Gettelman et al., 2011). The only known exceptions to this reliability of the lapse-rate tropopause occur within complex UTLS stability environments (e.g., layered UT and LS air near the subtropical jet during Rossby wave breaking events; Homeyer et al., 2011) and over the Antarctic during austral winter, where there exists an erroneously high lapse-rate tropopause altitude due to weakened stability in the lower stratosphere (Zängl and Hoinka, 2001). The issue over the Antarctic is confined to latitudes poleward of 60°S for about 3 months out of the year. The WMO definition also allows for identification of more than one tropopause if low stability layers are observed for a substantial depth above the primary tropopause. Double lapse-rate tropopauses have been the focus of many studies during the past two decades and have been found to be largely related to Rossby wave breaking events and associated STE above the subtropical jets (e.g., Shapiro, 1980; Seidel and Randel, 2006; Pan et al., 2009; Homeyer et al., 2011; Añel et al., 2012; Peevey et al., 2014; Schwartz et al., 2015; Manney et al., 2017; Liu and Barnes, 2018).

Tropopause definitions based on either the maximum static stability gradient in the vertical dimension or curve fitting to the static stability profile, which is typically characterized as a step function from uniformly low stability in the troposphere and high stability in the stratosphere, have been increasingly used in global tropopause studies (Birner, 2010a; Homeyer et al., 2010; Gettelman and Wang, 2015). These approaches often provide unique information on tropopause structure, such as its sharpness (the depth of the troposphere-stratosphere stability transition). However, static stability definitions frequently fail in the subtropics where the stability profile is often layered (e.g., double lapse-rate tropopauses), with maxima at multiple altitudes or at altitudes well removed from the most prominent transition, or the transition from troposphere to stratosphere occurs over a deep (≥3 km) layer.

Alternative definitions to the lapse-rate tropopause or one based on static stability are often applied in specific locations, separated mostly by latitude. In the tropics, the UTLS temperature minimum, known as the cold point, is often used to define the tropopause. This cold point tropopause is typically employed in studies that examine troposphere-to-stratosphere transport of water vapor (Holton et al., 1995; Mote et al., 1996; Fueglistaler et al., 2009). While easily defined, cold point tropopause altitudes are only reliable within the deep tropics (between 20°S and 20°N), because outside of this region the coldest temperature in a vertical profile is not always associated with the transition between tropospheric and stratospheric air (indicated by stability or composition). Profiles containing multiple lapse-rate tropopauses are a good example of when the cold point tropopause fails. Potential vorticity (PV), which is conserved in an adiabatic and friction-less flow, is commonly used for transport studies in the extratropics and often used to define a dynamical tropopause (Holton et al., 1995; Kunz et al., 2011, 2015; Homeyer and Bowman, 2013; Boothe and Homeyer, 2017). The PV threshold used ranges from ±1–4 PVU (where 1 PVU = 10^{-6} km^2 kg^{-1} s^{-1}) in previous studies, with most using ±2 PVU (e.g. Wernli and Bourqui, 2002; Sprenger et al., 2003; Sprenger and Wernli, 2003). The PV value that best coincides with the lapse-rate tropopause varies with latitude and season and, if this variability is not accounted for, it can result in large differences in quantitative transport studies (Hoinka, 1998; Homeyer and Bowman, 2013). For example, Homeyer and Bowman (2013) found that changing the PV iso-surface from ±2 to ±4 PVU resulted in a reversal of the net STE between the tropics and extratropics from Rossby wave breaking. The dynamical tropopause is only reliable in close proximity to and poleward of the subtropical jets, since PV approaches a value of 0 near the equator and iso-surfaces diverge from the lapse-rate and cold point tropopauses in the deep tropics.

One final type of tropopause definition that has been used in many studies is that based on chemical composition. Multiple studies have used ozone (O_3) profiles to define the tropopause, where the O_3 tropopause is defined using absolute thresholds for O_3 concentration and vertical gradients of O_3 (Bethan et al., 1996; Wild, 2007). Unique limitations exist for O_3 tropopause altitudes due to the seasonality and location-dependent variability of the ideal O_3 concentration threshold value (Logan, 1999). Another chemical tropopause definition exists that uses multiple coincident trace gas concentrations to define the UTLS chemical transition layer, often leveraging O_3 and carbon monoxide (Zahn et al., 2004; Pan et al., 2004). However, such coincident chemical observations are uncommon, making the method impractical for climatological analysis. Artificial tracers have become increasingly used in numerical models to allow for a chemical tropopause definition, with the 90-day lifetime tracer (known as e90) being a widely used choice in the chemistry climate model community (Prather et al., 2011). Primary limitations of this approach are that it cannot be applied to observations and it is not incorporated into reanalysis models, which are commonly used for climatological analyses.

In summary, multiple tropopause definitions exist in the community and are often chosen based on the goals of the study. The primary goal of this study is to evaluate long-term changes in tropopause characteristics globally. Since the lapse-rate tropopause is a global definition that can be easily applied to both conventional observations and model output, agrees with the sharp stability and chemical transitions between troposphere and stratosphere, and enables the unique opportunity to study multiple tropopause structures, we employ this definition for analysis in this study.

Radiosondes have been the traditional source of thermodynamic profiles of the atmosphere from the near surface up to 30 km since 1950s, and have been widely used for studying long-term changes in the tropopause (e.g. Seidel and Randel, 2006;

Añel et al., 2007, 2008; Xian and Fu, 2015). A primary shortcoming of radiosonde observations is the limited spatial coverage, since they are mostly launched from land masses. Another limitation is that not all radiosonde flights from a given location are successful, leading to discontinuities in the data record. Moreover, despite the fact that the number of sites providing operational radiosonde observations has increased over time, the number of locations with long-term records suitable for trend studies is relatively small. More recently, modern high-resolution reanalysis models have been used to evaluate tropopause characteristics globally since they provide data that is spatially and temporally continuous (Manney et al., 2014, 2017; Boothe and Homeyer, 2017). Reanalyses assimilate global quality-controlled observations to provide best estimates of past three-dimensional atmospheric states and are used to develop an understanding of a wide range of atmospheric processes, which is often not possible using observations alone (Fujiwara et al., 2017). Several reanalyses are publicly available and cover historical periods of 30 years or longer, with modern reanalyses such as MERRA and MERRA-2 (the National Aeronautics and Space Administration Modern-Era Retrospective analysis for Research and Applications, Versions 1 and 2), ERA-Interim (the European Centre for Medium-Range Weather Forecasts interim reanalysis), JRA-55 (the Japanese Meteorological Association 55-year reanalysis), and CFSR (the National Centers for Environmental Prediction Climate Forecast System Reanalysis) being widely used for research today.

While tropopause altitudes can be similarly evaluated in observations and reanalyses, the reanalyses can provide us with a broader understanding of tropopause behavior given the spatial and temporal limitations of the observations outlined above. For example, Manney et al. (2017) compared the frequency of double tropopauses in five modern reanalyses during 1980–2014 and highlighted the sensitivity of the double tropopause identifications to model vertical resolution. Boothe and Homeyer (2017) further argued that differences in STE estimates based on reanalysis output are partly due to the differences in vertical grid spacing. Moreover, reanalyses are highly dependent on the underlying global forecast models, data input sources, and assimilation systems (Fujiwara et al., 2017). For instance, although both ERA-Interim and MERRA-2 assimilate O_3 profiles from the Aura Microwave Limb Sounder, they use climatological and prognostic O_3 fields for radiation calculations, respectively (Dethof and Hólm, 2004; Rienecker et al., 2008). This, in turn, may have a significant impact on assimilated ozone and stratospheric temperatures during winter and spring, which could impact tropopause calculations.

Here, we investigate the accuracy of primary tropopause altitudes and double tropopause frequency in four modern reanalyses (ERA-Interim, JRA-55, MERRA-2, and CFSR) and diagnose long-term (1981-2015) trends in tropopause altitude and double tropopause frequency using both radiosonde observations and reanalyses. In particular, we address three research questions: (1) How well do modern reanalyses represent the lapse-rate tropopause? (2) What are the recent trends in tropopause altitude and double tropopause frequency and how do they vary spatially?, and (3) How sensitive are tropopause altitude trends to the geographic coordinate system used? This research provides a unique evaluation of model performance, of climate variability and change, and a physical perspective of UTLS dynamics, including STE that is commonly associated with double tropopause events.

2 Data and Methods

2.1 Reanalysis output

ERA-Interim output is available from 1979 to the present on an approximately 80 km horizontal grid and with 750-1250 m vertical resolution in the UTLS (Dee et al., 2011), where the UTLS is defined here as the 8-18 km altitude layer. JRA-55 is available from 1958 to the present on a ∼60 km horizontal grid (Kobayashi et al., 2015). Similar to ERA-Interim, JRA-55 has 750-1250 m vertical grid spacing in the UTLS. MERRA-2 is available from 1979 to the present on a 0.5°× 0.625° longitude-latitude grid and at ∼1100 m vertical resolution throughout the UTLS (Bosilovich, 2015). CFSR is available from 1979 to 2010 on a 0.5°× 0.5° longitude-latitude grid and at 700-900 m vertical resolution in the UTLS. CFSR output is extended to the year 2015 in this study using analyses from the Climate Forecast System version 2 (CFSv2) model. For a more detailed discussion of these reanalyses and their differences (including profiles of vertical resolution), see Fujiwara et al. (2017). All tropopause analyses are done using daily 0000 UTC fields from each reanalysis (though not shown, analyses using alternative synoptic times for shorter time periods are consistent). Geopotential height was computed for each reanalysis model-level output using the moisture-included hypsometric equation. Meteorological parameters are interpolated linearly to a regular 1°× 1° longitude-latitude grid in the horizontal for analysis to enable 1-to-1 comparison (this choice has a negligible impact on the reported results). Temperature profiles are linearly interpolated to a regular 200-m vertical resolution prior to tropopause identification.

2.2 Global radiosonde data

Radiosonde data used in this study were obtained from the Integrated Global Radiosonde Archive (IGRA) Version 2 (Durre et al., 2016). The IGRA database provides historical radiosondes from locations around the world that have been comprehensively quality controlled and corrected for gross errors. These data are the best long-term historical record available for studies of the vertical structure of the UTLS (global satellite-based observations, such as radio occultation, are available for only the last ∼18 years). IGRA radiosonde observations are mainly available twice daily at 0000 and 1200 UTC. Given that a small number of stations launch radiosondes at non-standard times, we included launches that occurred between 2100 and 0300 UTC in the 0000 UTC analysis, and those between 0900 and 1500 UTC in the 1200 UTC analysis. The IGRA radiosondes are provided at mandatory (conventional pressure levels) and significant levels, which preserves any substantial lapse-rate changes in the original data (observations are typically taken at 6-s intervals, resulting in ≤50-m vertical resolution), but can result in coarse vertical resolution in the UTLS (>1 km). Such coarse resolution of the profile can prevent successful application of the lapse-rate tropopause by limiting the number of observations used to satisfy the second criterion of the WMO definition and the criterion for identifying multiple tropopauses (see Section 2.3 below), as is true for the reanalysis output. Thus, in order to enable thorough evaluation of the WMO criteria and reliable tropopause identification, radiosonde data are also linearly interpolated to a 200-m regular vertical grid prior to tropopause identification. Only soundings that have valid observations between 5 km and 22 km altitude are used to identify the tropopause. The original high vertical resolution (∼5 m) profiles from select National Weather Service sites in the United States were also retrieved for illustration purposes only (see Figure 1).

2.3 Lapse-rate tropopause identification

As outlined in the Introduction, we employ the WMO lapse-rate tropopause definition for analysis in this study. The WMO definition defines the first tropopause in a profile as "the lowest level at which the lapse rate falls to 2 °C km^{-1} or less, provided also the average lapse rate between this level and all higher levels within 2 km does not exceed 2 °C km^{-1}" (WMO, 1957). The WMO definition allows for a secondary tropopause "if above the first tropopause the average lapse rate between any level and all higher levels within 1 km exceeds 3 °C km^{-1}, then a second tropopause is defined by the same criterion." To avoid boundary layer inversions and false tropopause identification, as well as secondary tropopauses above the altitude of that typically observed for the primary tropopause in the tropics, the algorithm is applied only to altitudes ranging from 5 km to 22 km. Lapse rates are calculated for each profile using a forward (upward) difference scheme in the form $\Gamma(z_i) = -\partial T/\partial z \approx -(T_{i+1} - T_i)/(z_{i+1} - z_i)$, where Γ is the lapse rate in °C km^{-1}, T is temperature in °C, and z is altitude in km. Tropopause altitudes from reanalysis are geopotential altitudes.

Example WMO lapse-rate tropopause altitudes calculated for two randomly selected radiosonde profiles launched at the Corpus Christi, Texas National Weather Service office (97.5°W longitude, 27.78°N latitude) in the United States are shown in Figure 1. Both of these profiles have two WMO tropopauses. Here, we show both the full resolution temperature profile obtained by each radiosonde and the reduced resolution profile from the IGRA archive to demonstrate common differences in the level of detail between native data and that reported at mandatory and significant levels. As demonstrated in both profiles, the IGRA data preserves all significant lapse rate transitions and results in nearly equivalent tropopause altitude definitions to those computed using the full resolution profile (differences are ≤100 m). Consistent results are found when selecting profiles randomly from other stations and time periods (not shown). Thus, we are confident that mandatory and significant levels included in the IGRA data are suitable for tropopause analyses following the methods employed here. Coincident temperature profiles and tropopause altitudes from each reanalysis model are superimposed in Figure 1 and show much less detail than the higher resolution observations and a general inability to capture multiple tropopause altitudes. These example profiles have shallow inversion layers above the primary lapse-rate tropopause altitude, which are often not well captured in the coarse resolution reanalysis model profiles. As a result, the secondary tropopause is often erroneously classified as the primary tropopause altitude in model output (e.g., see also Figs. 4 & 6 and discussion from Homeyer et al., 2010). Extensive comparisons of tropopause altitudes and multiple tropopause frequencies between the radiosondes and reanalyses are provided in Section 3.1.

2.4 Trend analyses

Trends over the 35-year analysis period (1981-2015) of this study are calculated using monthly mean primary tropopause altitudes and the monthly fraction of profiles with double tropopauses. For radiosondes, we require a station to have at least 20 days of suitable profiles (all from either 00 UTC or 12 UTC and including mandatory and significant levels up to an altitude of 22 km or higher) to compute a monthly mean tropopause altitude or double tropopause fraction. Trends are only calculated for radiosonde stations that have sufficient observations for at least half of the total number of months in the 35-year period and a roughly even distribution of data points throughout the period (data gaps $<$ 5 years and no missing data near the beginning and end of the time series) for adequate trend analysis. Monthly tropopause time series from both radiosondes and reanalyses are then deseasonalized using a high-pass filter that removes variability at time scales less than or equal to 1 year. Linear regression is used on the filtered time series to measure trends over the 35-year period. Trends are deemed significant if they exceed the 3-σ uncertainty of the measured slope, which is analogous to statistical significance at the 99% confidence level for a Gaussian distribution.

For reanalyses, trends are also calculated in an alternative coordinate system. A sharp discontinuity in the primary lapse-rate tropopause altitude is found near the subtropical jet and known as the "tropopause break" (e.g., Randel et al., 2007; Pan and Munchak, 2011; Homeyer and Bowman, 2013). Tropopause altitudes are uniformly high ($\geq$15 km) in the tropics and uniformly low (mostly 8–12 km altitude) in the extratropics. In fact, as shown by Birner (2010b), Boothe and Homeyer (2017) and others, global and hemispheric frequency distributions of tropopause altitude are bimodal. As a result, the tropopause break is easily defined in model output as globe circling contours of a threshold tropopause altitude or pressure coinciding with the frequency minimum between the tropical and extratropical modes (typically $\sim$14 km or $\sim$150 hPa; e.g., see Homeyer and Bowman, 2013). Because the location (latitude) of the subtropical jets and tropopause breaks varies considerably in space and time, trend analyses in the vicinity of the tropopause breaks can be adversely impacted by their variability and potential long-term changes in latitude. Thus, to remove this variability and evaluate trends within the tropical and extratropical reservoirs separately, we also analyze trends in the reanalyses using a tropopause break-relative latitude coordinate. Tropopause break latitudes are identified at each 00 UTC analysis using contours of the tropopause altitude that coincides with the frequency minimum between tropical and extratropical modes in each hemisphere. Monthly means are then calculated by averaging the instantaneous tropopause fields on the relative latitude grid in each hemisphere. For plotting, tropopause break-relative analyses are mapped using the long-term mean break latitudes and any data extending beyond the equator and pole is trimmed from each hemisphere.

Note that time series used for trend analysis were not adjusted for potential discontinuities and/or biases owing to changes in instrumentation or other factors (e.g., see Añel et al., 2006; Antuña et al., 2009). Thus, some elements of the trend analyses (especially for the radiosonde observations) may be impacted by these artifacts. However, a substantial number of time series with large, significant trends (see Section 3.2) were manually evaluated and no discontinuities were found (not shown). Thus, we expect such factors to have a minimal impact on the results outlined below.

3 Results

3.1 Tropopause validation

To evaluate the fidelity of tropopause altitudes in the reanalyses, we first compare tropopause altitudes from the gridded reanalysis output with the radiosonde data at 00 UTC only. Instantaneous tropopause altitudes computed on the $1^\circ \times 1^\circ$ longitude-latitude grids are interpolated linearly in space to the locations of the radiosondes for comparison. Results of these comparisons are shown for individual months during each season: January, April, July and October between 1981 and 2015. In total, 317 radiosonde stations and approximately 9.9×10^4 profiles are used for this validation, with their geographic locations shown as circles in Figure 4. Biases and errors (r.m.s. differences) in reanalysis primary tropopause altitudes from this comparison are listed in Table 1. More than half of the primary tropopause altitudes in MERRA-2 are found to have a positive bias, whereas the majority of biases in the remaining reanalyses (ERA-Interim, JRA-55, and CFSR) are negative. Errors in the tropopause altitudes range from 950 m to 1200 m, which is comparable to the vertical grid spacing of the reanalyses.

All reanalyses produce too few double tropopauses compared to the radiosonde observations. Double tropopause frequency in the reanalyses is significantly underestimated, with 67-85 % of the sample having a negative bias. Biases are largest in JRA-55, with over 22 percentage points fewer double tropopauses than observed, and smallest in CFSR, which has 13 percentage points fewer double tropopauses than observed. Errors in double tropopause frequency show similar differences among the reanalyses. The lower double tropopause frequency in reanalyses is likely due to the coarse vertical resolution of the models. As outlined in Section 2.3 and illustrated in Figure 1, accurate tropopause identification requires vertical resolution of $\leq$1 km to detect shallow low and high stability layers that are often responsible for the occurrence of multiple tropopauses. CFSR, which has the lowest bias and error also has the finest vertical grid resolution in the UTLS, while MERRA-2 typically has finer resolution than ERA-Interim and JRA-55 at the altitude of the secondary tropopause. The differences in bias and r.m.s. error between models with approximately equivalent vertical grids, such as ERA-Interim and JRA-55, suggest that the representation of atmospheric processes and/or the data assimilation system used could also be responsible for some of the under-representation of double tropopause events.

To better evaluate the role of model vertical resolution as a source for the observed tropopause differences, we repeated the validation with degraded radiosonde observations. In particular, each IGRA radiosonde profile was linearly interpolated to the fixed model-level grid of each reanalysis in order to limit the detail of the observed temperature profile to that available from each model. These degraded radiosonde profiles were then used to calculate unique observation-based tropopause altitudes to compare with each reanalysis and determine the resulting bias and error, which is expected to be reduced (especially for double tropopauses). Table 1 shows these evaluations using the degraded radiosonde profiles and confirms that the biases and errors in both primary tropopause altitude and double tropopause frequency largely decrease. However, some bias and error remains (and for MERRA-2 tropopause altitudes, increases), which suggests that alternative sources of error (e.g., data assimilation, model physics/dynamics) are significant.

In addition to overall performance of the reanalyses, there are seasonal regional variations in the tropopause biases and errors. In order to better understand the source of these variations, we group the comparisons by latitude to reduce potential sources of uncertainty from the non-uniform global distribution of the radiosonde locations. Figures 2 and 3 show the bias and r.m.s. error for primary tropopause altitudes and double tropopause frequency, respectively, within five latitude bands: two in the extratropics of each hemisphere (45–90°N and 45–90°S), two in the subtropics of each hemisphere (20–45°N and 20–45°S), and one in the deep tropics (20°S–20°N). The southern hemisphere subtropics and extratropics tend to have larger errors than their counterparts in the northern hemisphere, likely due to (in part) fewer data sources for assimilation and fewer profiles used for analysis. Regionally, the largest primary tropopause errors are found within the subtropics of each hemisphere - commonly associated with the location of the subtropical jet and tropopause break. The complicated stability structure within the subtropics is known to lead to large errors in tropopause altitudes within models and is primarily the result of inadequate representation of multiple tropopauses and the precise location of the tropopause break (Homeyer et al., 2010). The largest differences within the subtropics occur during the winter season of each hemisphere, which is consistent with the time period during which the subtropical jet reaches its maximum intensity (Manney et al., 2017) and double tropoapuses are more frequent (e.g. Randel et al., 2007; Añel et al., 2008; Manney et al., 2017). The tropopause break is also sharpest in the vicinity of the subtropical jets at this time due to thermal wind balance (i.e., the latitudinal temperature gradients in the vicinity of the jet are largest). Biases and errors in double tropopause frequency are largest in the subtropics and extratropics of each hemisphere and small within the tropics, where multiple tropopauses are generally infrequent. Biases and errors are also largest in the winter of each hemisphere and smallest during the summer.

3.2 Eulerian mean tropopause altitude trends

The long-term trend in the tropopause altitude is considered to be an indicator of climate change. Here, we evaluate trends in Eulerian monthly mean primary tropopause altitudes using both radiosondes and reanalyses (Figure 4). Statistically significant trends in tropopause altitude are found in the radiosonde observations for many locations across the globe. Most trends point to an increase in tropopause altitude over time, with the largest increases (100-200 m per decade) found over western China, the contiguous United States, eastern Europe, and Indonesia. A small region of significant decreasing primary tropopause altitudes is found over Siberia and the subtropical Pacific.

Significant tropopause altitude trends in the reanalyses are in broad agreement with those identified from the radiosonde observations. In particular, regions with dense radiosonde coverage and large, positive trends are well represented in each reanalysis. The small, negative trend identified over Siberia is also reproduced in each reanalysis and agrees with previous studies (Santer et al., 2003a, b; Añel et al., 2006), but this behavior may be an artifact of changes in instrumentation and quality control of the radiosonde data over time (Añel et al., 2006). Trends in the reanalyses are generally larger and more variable outside of the regions with dense radiosonde coverage. Some notable features are the upward/downward tropopause altitude trend dipoles found over the eastern subtropical Pacific in ERA-Interim, JRA-55, and MERRA-2. This dipole is not found in CFSR and the magnitudes of the trends and spatial extent of the dipole vary considerably in the remaining reanalyses. Moreover, the dipole is consistent with significant narrowing of the tropics over the eastern Pacific that has been identified in the reanalyses via subtropical jet and tropopause break analysis (Manney and Hegglin, 2018; Martin et al., 2018).

Another notable difference among the reanalyses is the depiction of tropopause trends across the tropics, where MERRA-2 and CFSR show considerably larger upward trends than ERA-Interim and JRA-55. ERA-Interim also depicts a downward trend over the central Pacific, which is not observed in the remaining reanalyses. Trends in JRA-55 and MERRA-2 appear to be more consistent with the limited number of radiosondes available in the tropics, especially over Indonesia, the central Pacific, and Northern Australia. Finally, the tropopause altitude trends over Antarctica are found to be inconsistent among the reanalyses. Namely, a rising trend is found over Antarctica in ERA-Interim and MERRA-2, while a less extensive decreasing trend is found in JRA-55 and CFSR.

3.3 Break-relative tropopause altitude trends

As discussed briefly in the previous section, recent studies have identified significant regional changes in the width of the tropics that introduce some uncertainty to the precise nature of tropopause changes when diagnosing trends in an Eulerian framework. Therefore, in order to mitigate the effects of a meandering tropopause break and focus on tropopause changes in the tropical and extratropical reservoirs alone, we employ a tropopause break-relative coordinate here. Figure 5 shows the geographic distribution of primary tropopause altitude trends in a tropopause break-relative altitude coordinate for each reanalysis from 1981 to 2015. Considerable differences are found in the diagnosed trends here compared to those using the Eulerian monthly mean fields. In particular, significant trends are larger in magnitude in the tropopause break-relative coordinate and are more consistent amongst the reanalyses, especially over the eastern subtropical Pacific. ERA-Interim, JRA-55, and MERRA-2 are broadly consistent with one another, with large upward trends in tropopause altitude in the extratropics over the Pacific and similar patterns and signs of significant trends elsewhere.

MERRA-2 shows substantially different trends throughout the tropics relative to ERA-Interim and JRA-55, with both larger upward magnitudes and unique patterns over the western Pacific and Indian Ocean. CFSR shows some consistency with MERRA-2 in the tropics in terms of the trend magnitude, but is inconsistent in pattern (especially over the Americas). In ERA-Interim, JRA-55, and MERRA-2, trends in the tropics are largest immediately equatorward of the tropopause break latitude in each hemisphere, while CFSR shows the greatest trends in the deep tropics. However, the Eulerian mean trends in Figure 4 show more consistent behaviour amongst the reanalyses in the depiction of these poleward maxima in altitude trends within the tropics, which is in agreement with previous analyses (e.g. Seidel and Randel, 2006). Outside of the tropics, CFSR is in broad disagreement with the remaining reanalyses and suggests largely decreasing trends in tropopause altitude.

3.4 Double tropopause climatology and trends

As outlined in Section 3.1, double tropopause frequencies often have large bias and error in the reanalyses, which is largely attributed to their coarse vertical grid resolution. To better understand the nature of these errors, maps of annual-mean double tropopause frequency are presented for the radiosondes and reanalyses in Figure 6. These maps are consistent with similar analyses of radiosondes, satellite observations, and reanalyses in prior studies (e.g., Randel et al., 2007; Añel et al., 2008; Peevey et al., 2012; Manney et al., 2017), and show that there are belts of high double tropopause frequency in the northern subtropics and midlatitudes of each hemisphere, largely near and poleward of the subtropical jets and tropopause breaks. The patterns and spatial extent of these belts are consistent between the radiosonde observations and reanalyses, but there are considerable differences in the frequency values. Radiosondes show that these high-frequency belts are characterized by values $\geq$40 %, while the reanalyses are at least 10–20 % lower. The magnitudes of the differences between the radiosonde and reanalysis frequencies within the high-frequency belts are consistent with that found in the overall bias and error evaluation (Fig. 3). Outside of the high-frequency belts, there is a unique feature found within the tropics in CFSR. Namely, a narrow band of double tropopause frequency between 10 and 30 % is seen along the equator stretching from central Africa to eastern Indonesia. This feature does not exist in the remaining analyses and is poorly sampled by the radiosonde network. However, the small number of stations available in western Indonesia do show consistent double tropopause frequencies. These double tropopauses may be driven (in part) by shallow, lateral transport of extratropical lower stratospheric air into the tropical upper troposphere on the eastern edge of the Asian monsoon anticyclone (e.g., Konopka et al., 2010), but the lack of continuity of this feature between the midlatitude high-frequency belt and the enhanced frequency belt along the equator suggests that the dynamics of the monsoon anticyclone may also be important to their formation.

Eulerian mean trend maps for double tropopause frequency during 1981-2015 are shown for the radiosondes and reanalyses in Figure 7. Trends in the double tropopause frequency are found to be statistically significant at almost all of the radiosonde locations and show substantial increases in frequency (≥2 % per decade) nearly everywhere. The largest increasing trends (≥3 % per decade) for double tropopause frequency are found in the midlatitudes in each hemisphere and poleward of the high-frequency belts in the long-term climatology, with some small (mostly <1 % per decade) decreasing trends over Siberia, southern China and the Caribbean (locations with climatologically low double tropopause frequency). The midlatitude increasing trends are comparable to those diagnosed in Castanheira et al. (2009) between 1970 and 2006 for the 30–60°N and 30–60°S latitude belts, which were 3.3 % and 6.6 % per decade, respectively. Taken together with the climatological double tropopause distribution, these trends imply that the area of frequent double tropopause environments is increasing in each hemisphere, mostly indicating a northward expansion of the high-frequency belts.

Areas of significant increasing trends in double tropopause frequency are largely consistent between the radiosondes and reanalyses, with CFSR being the only exception. In particular, ERA-Interim, JRA-55, and MERRA-2 all show large increases in double tropopause frequency along and mostly north of the tropopause break. CFSR shows mostly decreasing trends in double tropopause frequency across the globe and is broadly inconsistent with the radiosonde observations. Where areas of significant trends agree in pattern and sign between the reanalyses and radiosondes, the magnitudes are largely consistent near the tropopause break and inconsistent poleward of the break. ERA-Interim, JRA-55, and MERRA-2 do not reproduce well the poleward extent of the significant increasing trends in observations, especially over North America. A unique increasing trend is found along the equator in MERRA-2, which is consistent with the location of the narrow band of moderate double tropopause frequency found in the CFSR climatology. While none of the remaining reanalyses show this feature, it is consistent with trends observed in the small number of radiosonde stations over western Indonesia and the western Pacific. Finally, the double tropopause frequency trends over Antarctica are decreasing in all reanalyses, but the area and magnitude of the trend varies considerably. The most consistent element of this feature is found over western Antarctica, but there are no radiosonde observations in this region to validate such a trend and the climatological frequencies in this region are small.

Maps of annual-mean double tropopause frequency and frequency trends in tropopause break-relative coordinates from the reanalyses are shown in Figures 8 and 9, respectively. In general, there is less variation in the patterns of both high-frequency double tropopause regions and increasing double tropopause frequency trends between the Eulerian and tropopause break-relative analyses. Annual-mean frequencies are higher in tropopause break-relative coordinates and maximize poleward of the mean tropopause break latitude throughout each hemisphere. Significant increasing double tropopause frequency trends are also found primarily poleward of the tropopause breaks in each hemisphere, but are otherwise mostly consistent with the Eulerian analysis. Two notable exceptions are the areas of significant increasing trends over the Northern and Southern east

Pacific and east Atlantic in ERA-Interim, JRA-55 and MERRA-2, where the largest increasing trends were found to be mostly equatorward of the mean tropopause break latitude in the Eulerian analysis. In tropopause break-relative coordinates, the areas of greatest increasing trends are generally centered on or poleward of the tropopause break in the Northern Hemisphere and poleward of the tropopause break in Southern Hemisphere.

4 Conclusions and discussion

In this study, we examined the fidelity of primary tropopause altitudes and double tropopause frequency in four modern reanalyses (ERA-Interim, JRA-55, MERRA-2, and CFSR) using the WMO lapse-rate tropopause definition. Long-term trends in the primary tropopause altitude and double tropopause frequency over a 35-year period (1981-2015) were also examined using both radiosonde observations and reanalyses. All reanalyses were found to reproduce observed primary tropopause altitudes with little bias and error comparable to the vertical grid resolution of the models, which is consistent with previous model tropopause evaluations (e.g., Homeyer et al., 2010; Solomon et al., 2016a). Bias and errors in the primary tropopause altitude were found to vary regionally, with the largest magnitudes of both routinely found in the subtropics of each hemisphere (Fig. 2). Double tropopause frequencies are broadly underrepresented in the reanalyses, with biases of up to 30 percentage points lower than observed. JRA-55 consistently showed the largest double tropopause bias, while CFSR consistently showed the lowest. The majority of error in double tropopause frequency was found in the subtropics and high latitudes of each hemisphere, where double tropopause environments are most common. Based on the differences in vertical grid resolution of the models and the necessary conditions for multiple tropopause identification using the WMO definition, the underestimates in double tropopause frequency in the reanalyses are argued to primarily be the result of too coarse vertical grid spacing. When the radiosonde data are degraded to the vertical grid resolution of each reanalysis, the biases and errors in double tropopause frequency are greatly reduced, while the biases and errors in primary tropopause altitude show little sensitivity to this change (Table 1; Figs. 2 & 3).

Trends in primary tropopause altitudes were found to be significant and increasing (i.e., upward) across most of the globe in the radiosonde observations, largely ranging from 40 to 120 m per decade (Fig. 4). Some similar, but significant decreasing altitude trends were found for a few radiosonde stations in Siberia. The reanalyses broadly reproduce the patterns and signs of significant trends in the radiosonde observations, with some disagreement in the areas of significant upward trends over China, Australia, and northern Antarctica. Outside of the regions with dense radiosonde coverage, there are significant upward and downward primary tropopause altitude trend dipoles over the central and eastern subtropical Pacific in ERA-Interim, JRA-55, and MERRA-2. Trend patterns in CFSR in this region are not consistent with the remaining reanalyses. To limit the impact of frequent meandering of the tropopause break and long-term trends in its location to the diagnosed primary tropopause altitude trends within the tropical and extratropical reservoirs, we also computed trends in a tropopause break-relative latitude coordinate (Fig. 5). The break-relative analysis revealed larger trends ($\geq$120 m per decade) in both the tropics and extratropics, which were increasing nearly everywhere and greatest within the tropics immediately equatorward of the tropical break latitudes in each hemisphere and in the extratropical reservoir over the eastern Pacific. As found in the Eulerian tropopause trend analysis, ERA-Interim, JRA-55, and MERRA-2 showed consistent patterns and magnitudes of significant trends, except for some locations within the tropics. CFSR was once again broadly inconsistent with the remaining reanalyses and showed decreasing tropopause altitude trends throughout most of the extratropical reservoir.

Depending on the reference frame (Eulerian or tropopause break-relative), the maxima in tropopause altitude trends in the tropics immediately equatorward of the mean tropopause break latitudes may indicate widening of the tropics and/or changes in the strength of the subtropical jets. Changes in jet speed can impact tropopause altitudes through changes in the magnitude of the associated vertical ageostrophic circulations around the jets. These circulations advect tropical upper troposphere air poleward above the jet altitude and extratropical lower stratosphere air downward and equatorward below in regions where the jet speed is increasing from west to east. Thus, changes in tropopause altitude near the tropopause break latitudes can be dynamically forced, with lower extratropical tropopause altitudes poleward of the jet and higher tropopause altitudes equatorward of the jet in regions where the west-to-east gradient in jet wind speed is increasing and vice versa in regions where the west-to-east gradient is decreasing. Manney and Hegglin (2018) find decreasing trends in subtropical jet wind speeds in the eastern Pacific within the Northern Hemisphere and increasing trends elsewhere, and increasing trends in subtropical jet wind speeds in the eastern Pacific and decreasing trends over the Indian Ocean within the Southern Hemisphere (e.g., see their Figure 8). Thus, dynamically driven changes in tropopause altitude near the subtropical jet are expected to be upward within the tropics from the eastern Pacific across North America and decreasing from Asia across the central Pacific in the Northern Hemisphere, while dynamically driven trends are expected to be smaller in the Southern Hemisphere. There are some patterns tropopause altitude trends that are consistent with this expectation, but it does not appear to be a major source of the diagnosed trends.

Patterns in tropopause altitude are also expected to be driven (in part) by the geographic distribution of observed surface (and tropospheric) warming during the 1981-2015 period. Figure 10 shows changes in global surface temperatures during the 35-year analysis period from the NASA Goddard Institute for Space Studies (GISS) surface temperature analysis (GISTEMP Team, 2018; Hansen et al., 2010). Patterns of long-term surface warming are consistent with the diagnosed trends in tropopause altitude here, particularly within the midlatitudes. For example, the two prominent regions of tropopause altitude increases found in the eastern midlatitude Pacific within each hemisphere coincide with locally enhanced surface warming during this time period. In addition, the patterns of increasing tropopause altitude trends over North America and Greenland also closely resemble patterns in surface warming there. Decreasing trends in the mid-to-high latitudes of the Southern Hemisphere found in the reanalyses are also consistent with decreasing trends in surface temperatures. These similarities imply that surface (and tropospheric) warming/cooling may be a significant source of diagnosed tropopause altitude trends in the extratropics. Sources of the upward trends within the tropics and their patterns are less clear.

Primary tropopause altitude trends over Antarctica were found to vary considerably among the reanalyses, with increasing trends in ERA-Interim and MERRA-2 and decreasing trends in JRA-55 and CFSR. These conflicting trends over Antarctica may be a result of different ozone input sources and assimilation as well as the dynamical responses to ozone concentration changes (e.g., Martineau et al., 2016; Polavarapu and Pulido, 2017; Fujiwara et al., 2017). For example, ERA-Interim and MERRA-2 assimilate ozone retrievals (both profiles and total column ozone (TCO)) from SBUV and SBUV/2, as well as TCO from OMI and profiles from Aura MLS, while JRA-55 only assimilates TCO and CFSR assimilates TCO and relatively coarse vertical resolution profiles from SBUV and SBUV/2. In particular, the rising Antarctic tropopause for ERA-Interim and MERRA-2 may be associated with a strengthening of the stratospheric polar vortex, which results in an elevated tropopause altitude due to the anomalous residual meridional upwelling in the polar latitudes and downwelling in the midlatitudes (Kidston et al., 2015). The decreasing tropopause altitudes in JRA-55 and CFSR may be associated with stratospheric ozone recovery (Son et al., 2009; Solomon et al., 2016b) and an acceleration of the BDC. Future work is needed to better elucidate the contributions from these known processes to the diagnosed long-term tropopause altitude trends. If possible, additional research on tropopause characteristics and variability using observations in this region would also be helpful.

Significant increasing trends in double tropopause frequency were found nearly everywhere in the radiosonde observations, with the largest trends near and poleward of the tropopause break ($\geq$3 % per decade; Fig. 7). Considering the long-term climatology of double tropopauses (Fig. 6), the observed trends imply that the high-frequency double tropopause belt in the subtropics and midlatitudes of each hemisphere is expanding poleward over time. The ERA-Interim, JRA-55, and MERRA-2 reanalyses showed consistent regions of long-term increasing double tropopause frequency trends, but underestimated the poleward extent of these trends compared to the observations. As found in the primary tropopause altitude trend analyses, CFSR showed trends in double tropopause frequency that were largely inconsistent with the remaining reanalyses and observations.

Given the relationship between double tropopauses, Rossby wavebreaking, and STE above the subtropical jets, the increasing frequency of double tropopauses over time implies that similar increases in Rossby wave breaking and STE have occurred during this time period and are mostly poleward. The tropopause break-relative trend analysis (Fig. 9) further confirmed that these increasing trends are almost entirely on the poleward side of the tropopause break in each hemisphere. Consistent increases in Rossby wave breaking and transport have been found in recent studies. In particular, increases in Rossby wave breaking frequency using MERRA-2 output at an altitude between the primary and secondary tropopauses, where one would expect the closest relationship between Rossby wave breaking and the occurrence of double tropopauses, have been documented in Jing and Banerjee (2018). Modeling studies suggest that transport of air from the tropics into the extratropical lower stratosphere has also increased in both hemispheres, which has been related to recently observed decreases in lower stratospheric ozone in the extratropics (Ball et al., 2018; Wargan et al., 2018). In comparison, the spatially limited increasing trends for double tropopause frequency found equatorward of the tropopause break in this study may indicate an increase in equatorward transport of stratospheric air from the extratropics into the tropical upper troposphere (Liu and Barnes, 2018), but more work is needed to better understand the impact of these tropopause changes.

Recognizing the lack of long-term radiosonde observations over the oceans and throughout much of the southern hemisphere, it is not surprising that reanalysis tropopause altitude errors and altitude trends differ the most in these regions. In addition to impacts directly related to data assimilation, differences in long-term trends between the reanalyses are likely the result of (1) differences in vertical grids, (2) differences in the representation of physical and dynamical process that impact both short- and long-term tropopause change, and/or (3) differences in the accuracy of multiple tropopause identification. It is not clear which of these factors is the most significant contributor to the observed differences, but differences in multiple tropopauses are likely responsible for much of the disagreement within the subtropics and high-frequency double tropopause belts in each hemisphere. In particular, since failing to identify a multiple tropopause is most often the result of misidentifying the secondary tropopause as the primary tropopause, long-term trends may be enhanced or reduced as a result of these errors (especially if they have a time dependence). As suggested by Gettelman and Wang (2015), the primary tropopause inversion layer depth has been changing over time, increasing in some regions and decreasing in others. Such changes will limit accurate identification of primary lapse-rate tropopause altitudes in the regions where it is getting shallower and improve identification in the regions where it is getting deeper. These changes may induce false trends in primary tropopause altitude and double tropopause frequency in the reanalyses, which we have not attempted to diagnose here. Future studies should investigate the factors responsible for differences in reanalysis trends in further detail.

In summary, this work has shown that global tropopause altitudes and the frequency of double tropopauses have largely increased between 1981 and 2015. These changes are relevant to climate and UTLS composition since increases in primary tropopause altitude are believed to be associated with a warming climate and double tropopause events often provide a physical indication of STE between the tropical upper troposphere and extratropical lower stratosphere. Broad agreement between three out of four of the modern reanalyses included in this study provides some confidence in their depictions of UTLS change. In addition, the consistency between reanalysis tropopause identifications and those from available radiosonde observations suggest that the tropopause and its behavior are well represented in modern reanalyses. Future work is needed to examine long-term variability and trends in tropopause characteristics using additional observations and models, including existing model output from future climate projections. Longer time periods and a greater number of potential solutions from available models may provide increased confidence in the sign, magnitudes and locations of the trends diagnosed in this study and those projected to occur in the future.

Acknowledgements. This work is funded by the National Natural Science Foundation of China Grant 41505033 and China Scholarship Council. The authors thank the agencies that provided reanalysis data used in this study: ERA-Interim from the European Centre for Medium Range Weather Forecasts (ECMWF) and JRA-55 from the Japan Meteorological Agency (JMA), both obtained from Reanalysis Data Archive (RDA) managed by Computational and Information Systems Laboratory (CISL) at the National Center for Atmospheric Research (NCAR), CFSR/CFSv2 from the National Centers for Environmental Prediction (NCEP), and MERRA-2 from the National Aeronautics and Space Administration (NASA) Global Modeling and Assimilation Office (GMAO). We also thank the National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Information (NCEI) for providing IGRA radiosonde data and the full-resolution radiosonde data used for comparison in Figure 1 and GISTEMP Team for providing GISS Surface Temperature Analysis.

References

Añel, J. A., Antuña, J. C., de la Torre, L., Nieto, R., and Gimeno, L.: Global statistics of multiple tropopauses from the IGRA database, Geophysical Research Letters, 34, L06709, https://doi.org/10.1029/2006GL029224, 2007.

Añel, J. A., Antuña, J. C., de la Torre, L., Castanheira, J. M., and Gimeno, L.: Climatological features of global multiple tropopause events, Journal of Geophysical Research (Atmospheres), 113, D00B08, https://doi.org/10.1029/2007JD009697, 2008.

Abalos, M., Legras, B., Ploeger, F., and Randel, W. J.: Evaluating the advective Brewer-Dobson circulation in three reanalyses for the period 1979-2012, Journal of Geophysical Research (Atmospheres), 120, 7534–7554, https://doi.org/10.1002/2015JD023182, 2015.

Añel, J. A., Gimeno, L., de La Torre, L., and Nieto, R.: Changes in tropopause height for the Eurasian region determined from CARDS radiosonde data, Naturwissenschaften, 93, 603–609, 2006.

Añel, J. A., de la Torre, L., and Gimeno, L.: On the origin of the air between multiple tropopauses at midlatitudes, The Scientific World Journal, 2012, 2012.

Antuña, J. C., Añel, J. A., Sterin, A., and Gimeno, L.: Professor Zalman Makhover: a relevant contributor to early tropopause studies, Meteorologische Zeitschrift, 18, 573–584, https://doi.org/10.1127/0941-2948/2009/0414, 2009.

Ball, W. T., Alsing, J., Mortlock, D. J., Staehelin, J., Haigh, J. D., Peter, T., Tummon, F., Stübe, R., Stenke, A., Anderson, J., Bourassa, A., Davis, S. M., Degenstein, D., Frith, S., Froidevaux, L., Roth, C., Sofieva, V., Wang, R., Wild, J., Yu, P., Ziemke, J. R., and Rozanov, E. V.: Evidence for a continuous decline in lower stratospheric ozone offsetting ozone layer recovery, Atmos. Chem. Phys., 18, 1379–1394, https://doi.org/10.5194/acp-18-1379-2018, 2018.

Bethan, S., Vaughan, G., and Reid, S. J.: A comparison of ozone and thermal tropopause heights and the impact of tropopause definition on quantifying the ozone content of the troposphere, Quarterly Journal of the Royal Meteorological Society, 122, 929–944, https://doi.org/10.1002/qj.49712253207, 1996.

Birner, T.: Residual Circulation and Tropopause Structure, Journal of Atmospheric Sciences, 67, 2582–2600, https://doi.org/10.1175/2010JAS3287.1, 2010a.

Birner, T.: Recent widening of the tropical belt from global tropopause statistics: Sensitivities, Journal of Geophysical Research (Atmospheres), 115, D23109, https://doi.org/10.1029/2010JD014664, 2010b.

Boothe, A. C. and Homeyer, C. R.: Global large-scale stratosphere-troposphere exchange in modern reanalyses, Atmospheric Chemistry & Physics, 17, 5537–5559, https://doi.org/10.5194/acp-17-5537-2017, 2017.

Bosilovich, M. G.: MERRA-2: Initial evaluation of the climate, National Aeronautics and Space Administration, Goddard Space Flight Center, 2015.

Butchart, N.: The Brewer-Dobson circulation, Reviews of Geophysics, 52, 157–184, https://doi.org/10.1002/2013RG000448, 2014.

Castanheira, J., Añel, J., Marques, C., Antuña, J., Liberato, M., Torre, L., and Gimeno, L.: Increase of upper troposphere/lower stratosphere wave baroclinicity during the second half of the 20th century, Atmospheric Chemistry and Physics, 9, 9143–9153, 2009.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, https://doi.org/10.1002/qj.828, 2011.

Dethof, A. and Hólm, E. V.: Ozone assimilation in the ERA-40 reanalysis project, Quarterly Journal of the Royal Meteorological Society, 130, 2851–2872, https://doi.org/10.1256/qj.03.196, 2004.

Durre, I., Vose, R. S., Yin, X., Applequist, S., and Arnfield, J.: Integrated Global Radiosonde Archive (IGRA) Version 2, https://doi.org/10.7289/v5x63k0q, 2016.

Feng, S., Fu, Y., and Xiao, Q.: Trends in the global tropopause thickness revealed by radiosondes, Geophys. Res. Lett., 39, L20706, https://doi.org/10.1029/2012GL053460, 2012.

Fu, Q., Lin, P., Solomon, S., and Hartmann, D.: Observational evidence of strengthening of the Brewer-Dobson circulation since 1980, Journal of Geophysical Research: Atmospheres, 120, 2015.

Fueglistaler, S., Dessler, A. E., Dunkerton, T. J., Folkins, I., Fu, Q., and Mote, P. W.: Tropical tropopause layer, Reviews of Geophysics, 47, RG1004, https://doi.org/10.1029/2008RG000267, 2009.

Fujiwara, M., Wright, J. S., Manney, G. L., Gray, L. J., Anstey, J., Birner, T., Davis, S., Gerber, E. P., Harvey, V. L., Hegglin, M. I., Homeyer, C. R., Knox, J. A., Krüger, K., Lambert, A., Long, C. S., Martineau, P., Molod, A., Monge-Sanz, B. M., Santee, M. L., Tegtmeier, S., Chabrillat, S., Tan, D. G. H., Jackson, D. R., Polavarapu, S., Compo, G. P., Dragani, R., Ebisuzaki, W., Harada, Y., Kobayashi, C., McCarty, W., Onogi, K., Pawson, S., Simmons, A., Wargan, K., Whitaker, J. S., and Zou, C.-Z.: Introduction to the SPARC Reanalysis Intercomparison Project (S-RIP) and overview of the reanalysis systems, Atmospheric Chemistry & Physics, 17, 1417–1452, https://doi.org/10.5194/acp-17-1417-2017, 2017.

Gettelman, A. and Wang, T.: Structural diagnostics of the tropopause inversion layer and its evolution, Journal of Geophysical Research (Atmospheres), 120, 46–62, https://doi.org/10.1002/2014JD021846, 2015.

Gettelman, A., Hoor, P., Pan, L. L., Randel, W. J., Hegglin, M. I., and Birner, T.: The Extratropical Upper Troposphere and Lower Stratosphere, Reviews of Geophysics, 49, RG3003, https://doi.org/10.1029/2011RG000355, 2011.

GISTEMP Team: GISS Surface Temperature Analysis (GISTEMP), 2018.

Hansen, J., Ruedy, R., Sato, M., and Lo, K.: Global surface temperature change, Rev. Geophys., 48, 121–132, https://doi.org/10.1029/2010RG000345, 2010.

Held, I. M.: Large-Scale Dynamics and Global Warming., Bulletin of the American Meteorological Society, 74, 228–242, https://doi.org/10.1175/1520-0477(1993)074<0228:LSDAGW>2.0.CO;2, 1993.

Hoinka, K. P.: Statistics of the Global Tropopause Pressure, Monthly Weather Review, 126, 3303, https://doi.org/10.1175/1520-0493(1998)126<3303:SOTGTP>2.0.CO;2, 1998.

Holton, J. R., Haynes, P. H., McIntyre, M. E., Douglass, A. R., Rood, R. B., and Pfister, L.: Stratosphere-troposphere exchange, Reviews of Geophysics, 33, 403–439, 1995.

Homeyer, C. R. and Bowman, K. P.: Rossby Wave Breaking and Transport between the Tropics and Extratropics above the Subtropical Jet, Journal of Atmospheric Sciences, 70, 607–626, https://doi.org/10.1175/JAS-D-12-0198.1, 2013.

Homeyer, C. R., Bowman, K. P., and Pan, L. L.: Extratropical tropopause transition layer characteristics from high-resolution sounding data, Journal of Geophysical Research (Atmospheres), 115, D13108, https://doi.org/10.1029/2009JD013664, 2010.

Homeyer, C. R., Bowman, K. P., Pan, L. L., Atlas, E. L., Gao, R.-S., and Campos, T. L.: Dynamical and chemical characteristics of tropospheric intrusions observed during START08, Journal of Geophysical Research (Atmospheres), 116, D06111, https://doi.org/10.1029/2010JD015098, 2011.

Jing, P. and Banerjee, S.: Rossby wave breaking and isentropic stratosphere-troposphere exchange in 1981–2015 in the Northern Hemisphere, J. Geophys. Res. Atmos., in press, https://doi.org/10.1029/2018JD028997, 2018.

Kidston, J., Scaife, A. A., Hardiman, S. C., Mitchell, D. M., Butchart, N., Baldwin, M. P., and Gray, L. J.: Stratospheric influence on tropospheric jet streams, storm tracks and surface weather, Nature Geoscience, 8, 433–440, https://doi.org/10.1038/ngeo2424, 2015.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., et al.: The JRA-55 reanalysis: General specifications and basic characteristics, Journal of the Meteorological Society of Japan. Ser. II, 93, 5–48, 2015.

Konopka, P., Grooß, J.-U., Günther, G., Ploeger, F., Pommrich, R., Müller, R., and Livesey, N.: Annual cycle of ozone at and above the tropical tropopause: observations versus simulations with the Chemical Lagrangian Model of the Stratosphere (CLaMS), Atmos. Chem. Phys., 10, 121–132, 2010.

Kunz, A., Konopka, P., Müller, R., and Pan, L. L.: Dynamical tropopause based on isentropic potential vorticity gradients, Journal of Geophysical Research (Atmospheres), 116, D01110, https://doi.org/10.1029/2010JD014343, 2011.

Kunz, A., Sprenger, M., and Wernli, H.: Climatology of potential vorticity streamers and associated isentropic transport pathways across PV gradient barriers, Journal of Geophysical Research (Atmospheres), 120, 3802–3821, https://doi.org/10.1002/2014JD022615, 2015.

Kushner, P. J., Held, I. M., and Delworth, T. L.: Southern Hemisphere Atmospheric Circulation Response to Global Warming., Journal of Climate, 14, 2238–2249, https://doi.org/10.1175/1520-0442(2001)014<0001:SHACRT>2.0.CO;2, 2001.

Liu, C. and Barnes, E.: Synoptic formation of double tropopauses, Journal of Geophysical Research: Atmospheres, 123, 693–707, 2018.

Liu, N. and Liu, C.: Global distribution of deep convection reaching tropopause in 1 year GPM observations, Journal of Geophysical Research (Atmospheres), 121, 3824–3842, https://doi.org/10.1002/2015JD024430, 2016.

Logan, J. A.: An analysis of ozonesonde data for the troposphere: Recommendations for testing 3-D models and development of a gridded climatology for tropospheric ozone, J. Geophys. Res., 104, 16, https://doi.org/10.1029/1998JD100096, 1999.

Manney, G. L. and Hegglin, M. I.: Seasonal and Regional Variations of Long-Term Changes in Upper-Tropospheric Jets from Reanalyses, Journal of Climate, 31, 423–448, https://doi.org/10.1175/JCLI-D-17-0303.1, 2018.

Manney, G. L., Hegglin, M. I., Daffer, W. H., Schwartz, M. J., Santee, M. L., and Pawson, S.: Climatology of Upper Tropospheric-Lower Stratospheric (UTLS) Jets and Tropopauses in MERRA, Journal of Climate, 27, 3248–3271, https://doi.org/10.1175/JCLI-D-13-00243.1, 2014.

Manney, G. L., Hegglin, M. I., Lawrence, Z. D., Wargan, K., Millán, L. F., Schwartz, M. J., Santee, M. L., Lambert, A., Pawson, S., Knosp, B. W., Fuller, R. A., and Daffer, W. H.: Reanalysis comparisons of upper tropospheric-lower stratospheric jets and multiple tropopauses, Atmospheric Chemistry & Physics, 17, 11 541–11 566, https://doi.org/10.5194/acp-17-11541-2017, 2017.

Martin, E. R., Homeyer, C. R., McKinzie, R. A., McCarthy, K. M., and Xian, T.: Regionally Varying Assessments of Tropical Width in Reanalyses and CMIP5 Models Using a Tropopause Break Metric, Journal of Climate, submitted, 2018.

Martineau, P., Son, S.-W., and Taguchi, M.: Dynamical consistency of reanalysis datasets in the extratropical stratosphere, Journal of Climate, 29, 3057–3074, 2016.

Mote, P. W., Rosenlof, K. H., McIntyre, M. E., Carr, E. S., Gille, J. C., Holton, J. R., Kinnersley, J. S., Pumphrey, H. C., Russell, III, J. M., and Waters, J. W.: An atmospheric tape recorder: The imprint of tropical tropopause temperatures on stratospheric water vapor, J. Geophys. Res., 101, 3989–4006, https://doi.org/10.1029/95JD03422, 1996.

Pan, L. and Munchak, L.: Relationship of cloud top to the tropopause and jet structure from CALIPSO data, Journal of Geophysical Research, 116, D12 201, 2011.

Pan, L. L., Randel, W. J., Gary, B. L., Mahoney, M. J., and Hintsa, E. J.: Definitions and sharpness of the extratropical tropopause: A trace gas perspective, Journal of Geophysical Research (Atmospheres), 109, D23103, https://doi.org/10.1029/2004JD004982, 2004.

Pan, L. L., Randel, W. J., Gille, J. C., Hall, W. D., Nardi, B., Massie, S., Yudin, V., Khosravi, R., Konopka, P., and Tarasick, D.: Tropospheric intrusions associated with the secondary tropopause, Journal of Geophysical Research (Atmospheres), 114, D10302, https://doi.org/10.1029/2008JD011374, 2009.

Peevey, T. R., Gille, J. C., Randall, C. E., and Kunz, A.: Investigation of double tropopause spatial and temporal global variability utilizing High Resolution Dynamics Limb Sounder temperature observations, Journal of Geophysical Research (Atmospheres), 117, D01105, https://doi.org/10.1029/2011JD016443, 2012.

Peevey, T. R., Gille, J. C., Homeyer, C. R., and Manney, G. L.: The double tropopause and its dynamical relationship to the tropopause inversion layer in storm track regions, Journal of Geophysical Research (Atmospheres), 119, 10, https://doi.org/10.1002/2014JD021808, 2014.

Ploeger, F., Riese, M., Haenel, F., Konopka, P., Müller, R., and Stiller, G.: Variability of stratospheric mean age of air and of the local effects of residual circulation and eddy mixing, Journal of Geophysical Research: Atmospheres, 120, 716–733, 2015.

Polavarapu, S. and Pulido, M.: Stratospheric and mesospheric data assimilation: The role of middle atmospheric dynamics, in: Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. III), pp. 429–454, Springer, 2017.

Prather, M. J., Zhu, X., Tang, Q., Hsu, J., and Neu, J. L.: An atmospheric chemist in search of the tropopause, Journal of Geophysical Research (Atmospheres), 116, D04306, https://doi.org/10.1029/2010JD014939, 2011.

Randel, W. J., Seidel, D. J., and Pan, L. L.: Observational characteristics of double tropopauses, Journal of Geophysical Research (Atmospheres), 112, D07309, https://doi.org/10.1029/2006JD007904, 2007.

Rienecker, M., Suarez, M., Todling, R., Bacmeister, J., Takacs, L., Liu, H.-C., Gu, W., Sienkiewicz, M., Koster, R., Gelaro, R., et al.: The GEOS-5 Data Assimilation System—Documentation of Versions 5.0. 1, 5.1. 0, and 5.2. 0, 2008.

Santer, B. D., Sausen, R., Wigley, T. M. L., Boyle, J. S., Achutarao, K., Doutriaux, C., Hansen, J. E., Meehl, G. A., Roeckner, E., Ruedy, R., Schmidt, G., and Taylor, K. E.: Behavior of tropopause height and atmospheric temperature in models, reanalyses, and observations: Decadal changes, Journal of Geophysical Research (Atmospheres), 108, 4002, https://doi.org/10.1029/2002JD002258, 2003a.

Santer, B. D., Wehner, M. F., Wigley, T. M. L., Sausen, R., Meehl, G. A., Taylor, K. E., Ammann, C., Arblaster, J., Washington, W. M., Boyle, J. S., and Brüggemann, W.: Contributions of Anthropogenic and Natural Forcing to Recent Tropopause Height Changes, Science, 301, 479–483, https://doi.org/10.1126/science.1084123, 2003b.

Schwartz, M. J., Manney, G. L., Hegglin, M. I., Livesey, N. J., Santee, M. L., and Daffer, W. H.: Climatology and variability of trace gases in extratropical double-tropopause regions from MLS, HIRDLS, and ACE-FTS measurements, Journal of Geophysical Research (Atmospheres), 120, 843–867, https://doi.org/10.1002/2014JD021964, 2015.

Seidel, D. J. and Randel, W. J.: Variability and trends in the global tropopause estimated from radiosonde data, Journal of Geophysical Research (Atmospheres), 111, D21101, https://doi.org/10.1029/2006JD007363, 2006.

Seidel, D. J., Ross, R. J., Angell, J. K., and Reid, G. C.: Climatological characteristics of the tropical tropopause as revealed by radiosondes, Journal of Geophysical Research (Atmospheres), 106, 7857–7878, https://doi.org/10.1029/2000JD900837, 2001.

Shapiro, M. A.: Turbulent Mixing within Tropopause Folds as a Mechanism for the Exchange of Chemical Constituents between the Stratosphere and Troposphere., Journal of Atmospheric Sciences, 37, 994–1004, https://doi.org/10.1175/1520-0469(1980)037<0994:TMWTFA>2.0.CO;2, 1980.

Sioris, C. E., McLinden, C. A., Fioletov, V. E., Adams, C., Zawodny, J. M., Bourassa, A. E., Roth, C. Z., and Degenstein, D. A.: Trend and variability in ozone in the tropical lower stratosphere over 2.5 solar cycles observed by SAGE II and OSIRIS, Atmospheric Chemistry & Physics, 14, 3479–3496, https://doi.org/10.5194/acp-14-3479-2014, 2014.

Solomon, D. L., Bowman, K. P., and Homeyer, C. R.: Tropopause-Penetrating Convection from Three-Dimensional Gridded NEXRAD Data, Journal of Applied Meteorology and Climatology, 55, 465–478, https://doi.org/10.1175/JAMC-D-15-0190.1, 2016a.

Solomon, S., Ivy, D. J., Kinnison, D., Mills, M. J., Neely, R. R., and Schmidt, A.: Emergence of healing in the Antarctic ozone layer, Science, 353, 269–274, https://doi.org/10.1126/science.aae0061, 2016b.

Son, S.-W., Polvani, L. M., Waugh, D. W., Birner, T., Akiyoshi, H., Garcia, R. R., Gettelman, A., Plummer, D. A., and Rozanov, E.: The Impact of Stratospheric Ozone Recovery on Tropopause Height Trends, Journal of Climate, 22, 429, https://doi.org/10.1175/2008JCLI2215.1, 2009.

Sprenger, M. and Wernli, H.: A northern hemispheric climatology of cross-tropopause exchange for the ERA15 time period (1979-1993), Journal of Geophysical Research (Atmospheres), 108, 8521, https://doi.org/10.1029/2002JD002636, 2003.

Sprenger, M., Croci Maspoli, M., and Wernli, H.: Tropopause folds and cross-tropopause exchange: A global investigation based upon ECMWF analyses for the time period March 2000 to February 2001, Journal of Geophysical Research (Atmospheres), 108, 8518, https://doi.org/10.1029/2002JD002587, 2003.

Wargan, K., Orbe, C., Pawson, S., Ziemke, J. R., Oman, L. D., Olsen, M. A., Coy, L., and Knowland, K. E.: Recent decline in extratropical lower stratospheric ozone attributed to circulation changes, Geophysical Research Letters, 2018.

Wernli, H. and Bourqui, M.: A Lagrangian “1-year climatology” of (deep) cross-tropopause exchange in the extratropical Northern Hemisphere, Journal of Geophysical Research (Atmospheres), 107, 4021, https://doi.org/10.1029/2001JD000812, 2002.

Wild, O.: Modelling the global tropospheric ozone budget: exploring the variability in current models, Atmospheric Chemistry & Physics, 7, 2643–2660, 2007.

WMO, M.: A three-dimensional science, WMO Bull, 6, 134–138, 1957.

Xian, T. and Fu, Y.: Characteristics of tropopause-penetrating convection determined by TRMM and COSMIC GPS radio occultation measurements, Journal of Geophysical Research (Atmospheres), 120, 7006–7024, https://doi.org/10.1002/2014JD022633, 2015.

Xian, T. and Fu, Y.: A hiatus in the tropopause layer change, International Journal of Climatology, 37, 4972–4980, https://doi.org/10.1002/joc.5130, 2017.

Zahn, A., Brenninkmeijer, C. A. M., and van Velthoven, P. F. J.: Passenger aircraft project CARIBIC 1997-2002, Part I: the extratropical chemical tropopause, Atmospheric Chemistry & Physics Discussions, 4, 1091–1117, 2004.

Zängl, G. and Hoinka, K. P.: The Tropopause in the Polar Regions., Journal of Climate, 14, 3117–3139, https://doi.org/10.1175/1520-0442(2001)014<3117:TTITPR>2.0.CO;2, 2001.

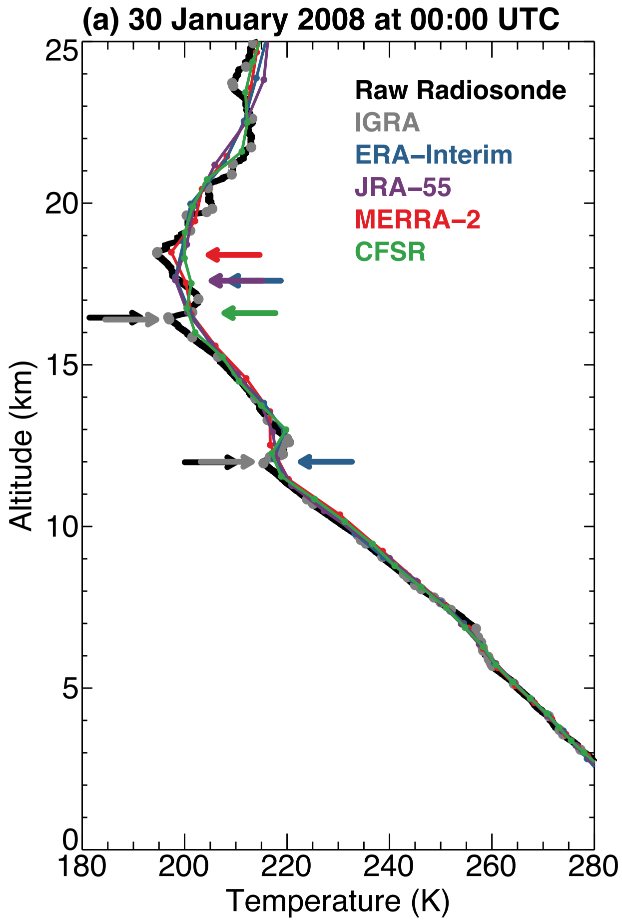

Figure 1. Example temperature profiles from the National Weather Service radiosonde site in Corpus Christi, Texas, at 0000 UTC on (a) 30 January 2008 and (b) 12 May 2008. Black lines show the original, full high-resolution radiosonde profile and gray dots show the reduced levels saved in the IGRA data. Coincident temperature profiles from ERA-Interim are shown in blue, JRA-55 in purple, MERRA-2 in red, and CFSR in green, with circles along these lines denoting each native model level. Colored arrows denote the locations of primary and secondary lapse-rate tropopause altitudes calculated using each temperature profile.

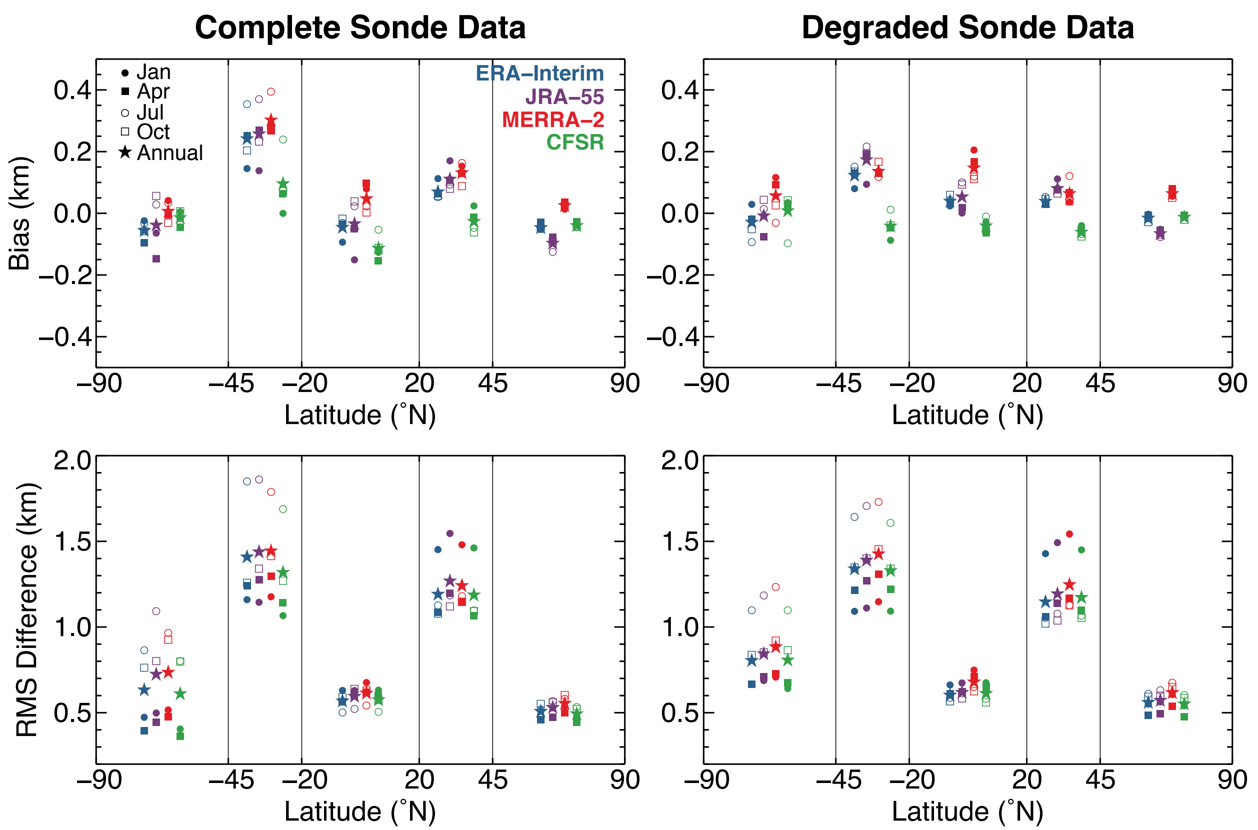

Figure 2. (Top panel) Average bias and (bottom panel) root-mean-square differences in instantaneous primary tropopause altitudes between radiosonde and reanalyses within five latitude bands. Open and closed symbols show the statistics as a function of season, with January (July) given as closed (open) circles, April (October) given as closed (open) squares and the annual values as closed stars. Results for ERA-Interim are shown in blue, JRA-55 in purple, MERRA-2 in red, and CFSR in green. Latitude values along the x-axis and vertical lines denote the boundaries of each latitude band. The number of radiosonde stations (profiles) used for each band are as follows: 113 (3.5×10^4) in the northern hemisphere extratropics (45–90°N), 152 (5.0×10^4) in the northern hemisphere subtropics (20–45°N), 25 (7,000) in the deep tropics (20°S–20°N), 20 (5,000) in the southern hemisphere subtropics (20–45°S), and 7 (2,000) in the southern hemisphere extratropics (45–90°S). Results using the complete IGRA radiosonde observations are shown on the left and results following the degradation of IGRA profiles to the native vertical grid of each reanalysis prior to tropopause calculation are shown on the right.

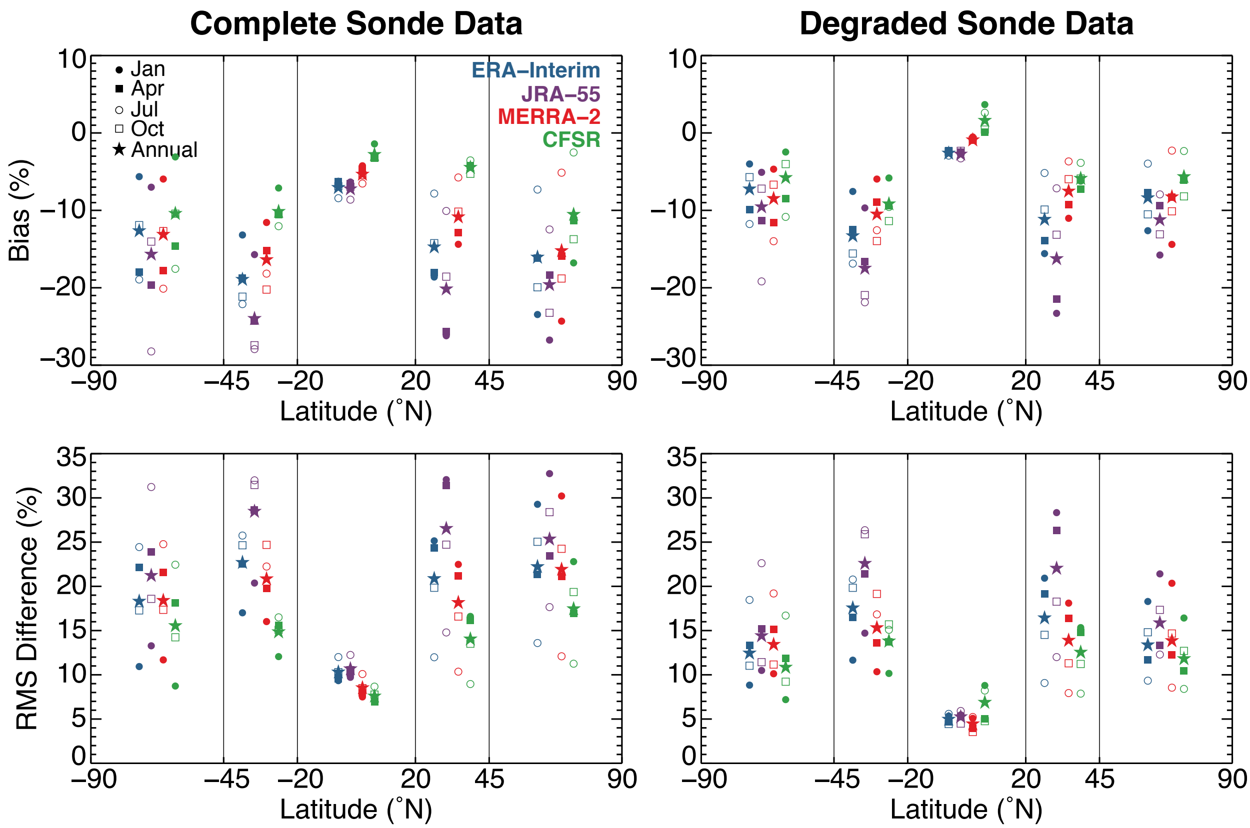

Figure 3. As in Fig. 2, but for the monthly double tropopause frequency.

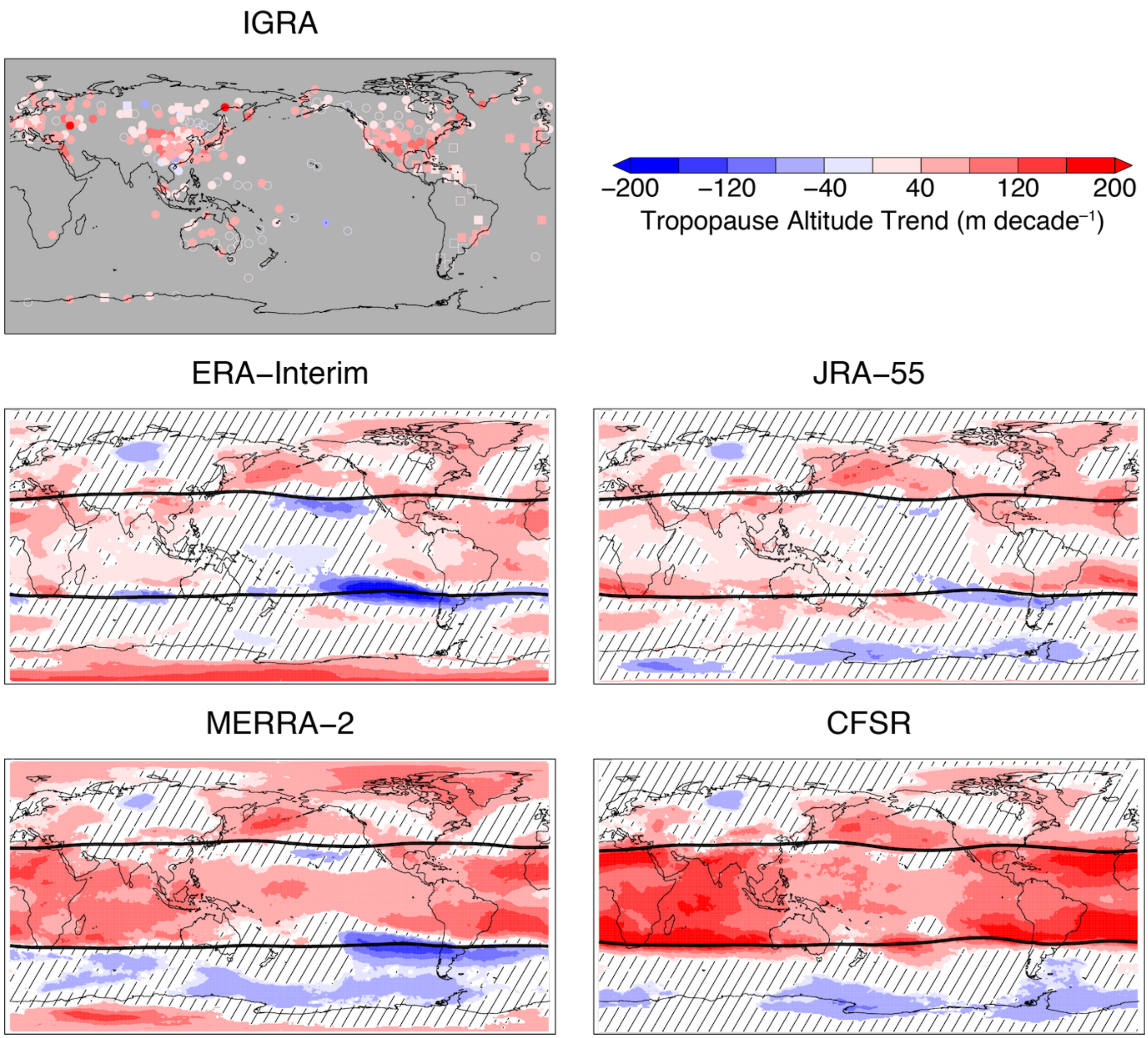

Figure 4. Eulerian mean trends of the primary tropopause altitude from 1981 to 2015 for IGRA radiosonde observations and the reanalyses. For IGRA trends, circles denote 0000 UTC trends and squares denote 1200 UTC trends, with filled symbols denoting statistical significance. Colored areas of the reanalysis maps are statistically significant, while line-filled regions are not. Thick black lines in each reanalysis map show the 35-year mean tropopause break latitudes.

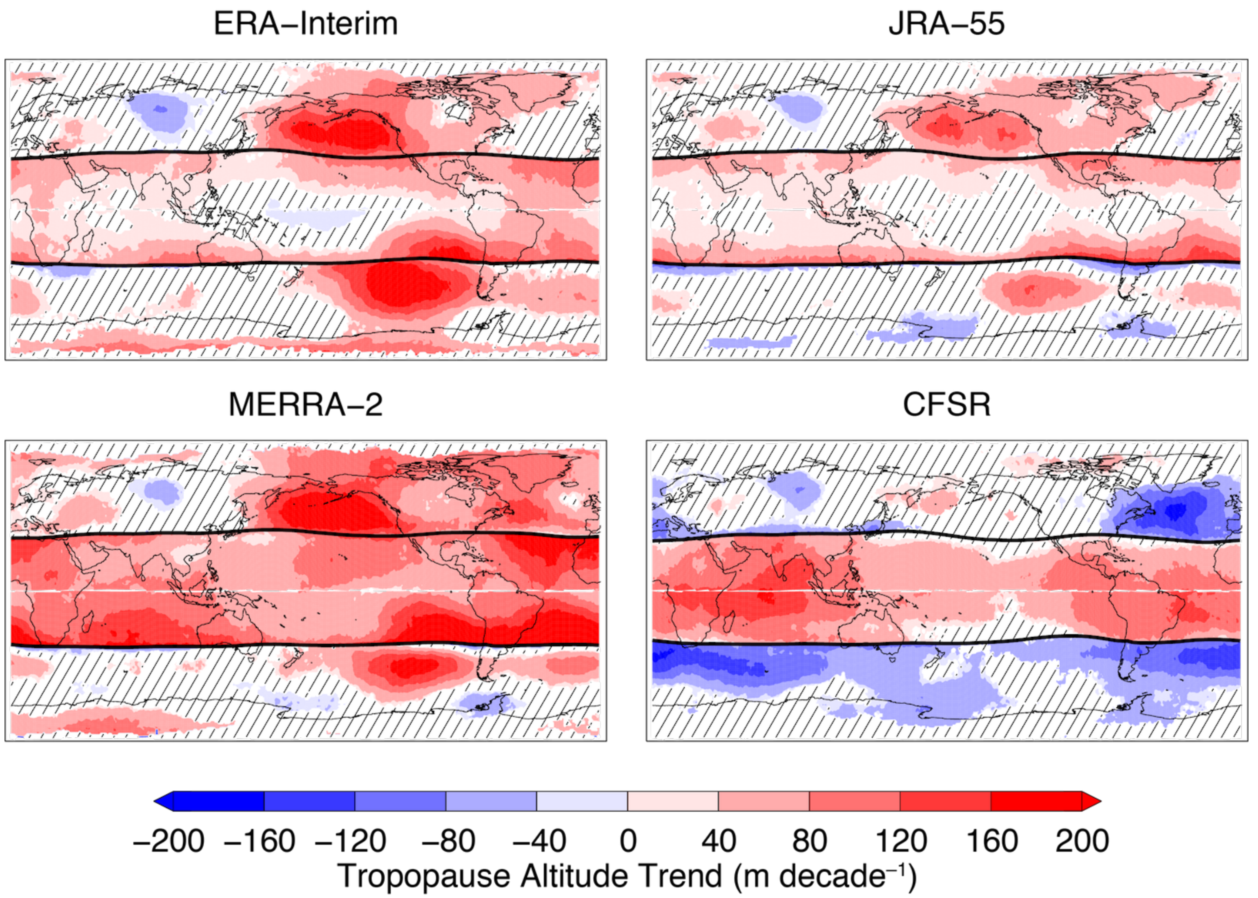

Figure 5. As in Fig. 4, but for tropopause-break relative primary tropopause altitudes from reanalyses only.

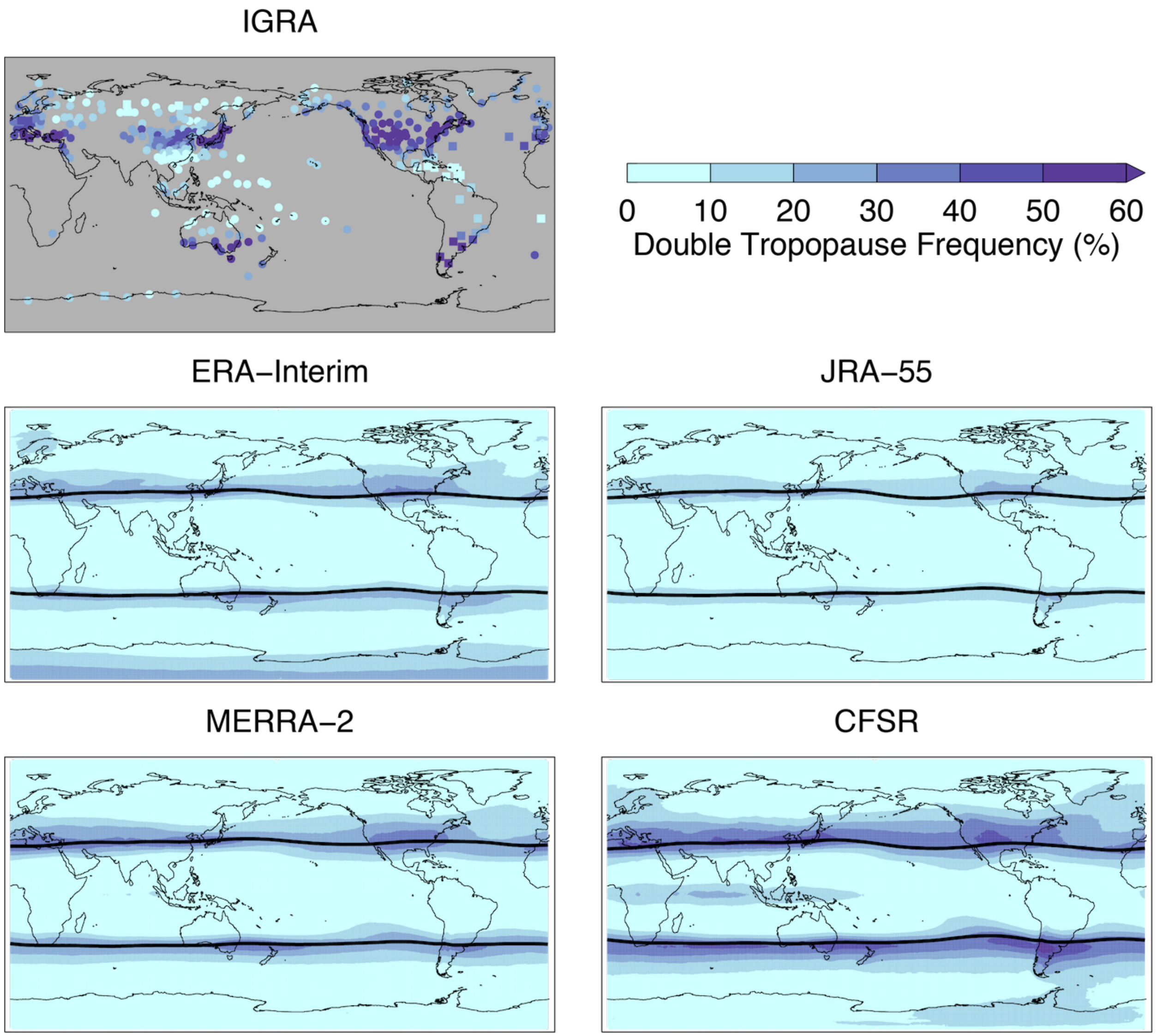

Figure 6. Average double tropopause frequency from 1981 to 2015 for IGRA radiosonde observations and the reanalyses. For IGRA, circles denote 0000 UTC frequencies and squares denote 1200 UTC frequencies. Thick black lines in each reanalysis map show the 35-year mean tropopause break latitudes.

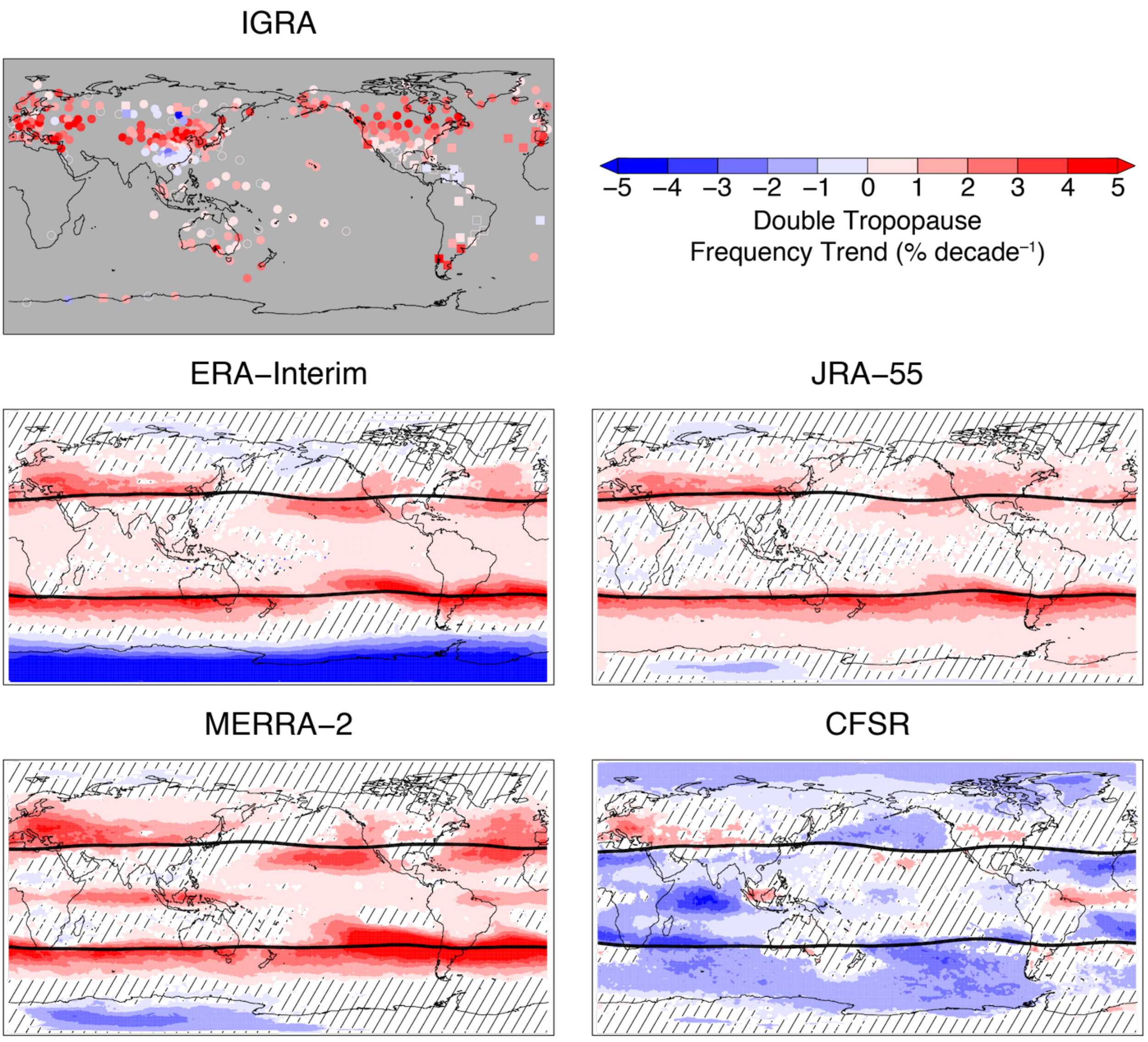

Figure 7. As in Fig. 4, but for trends in double tropopause frequency.

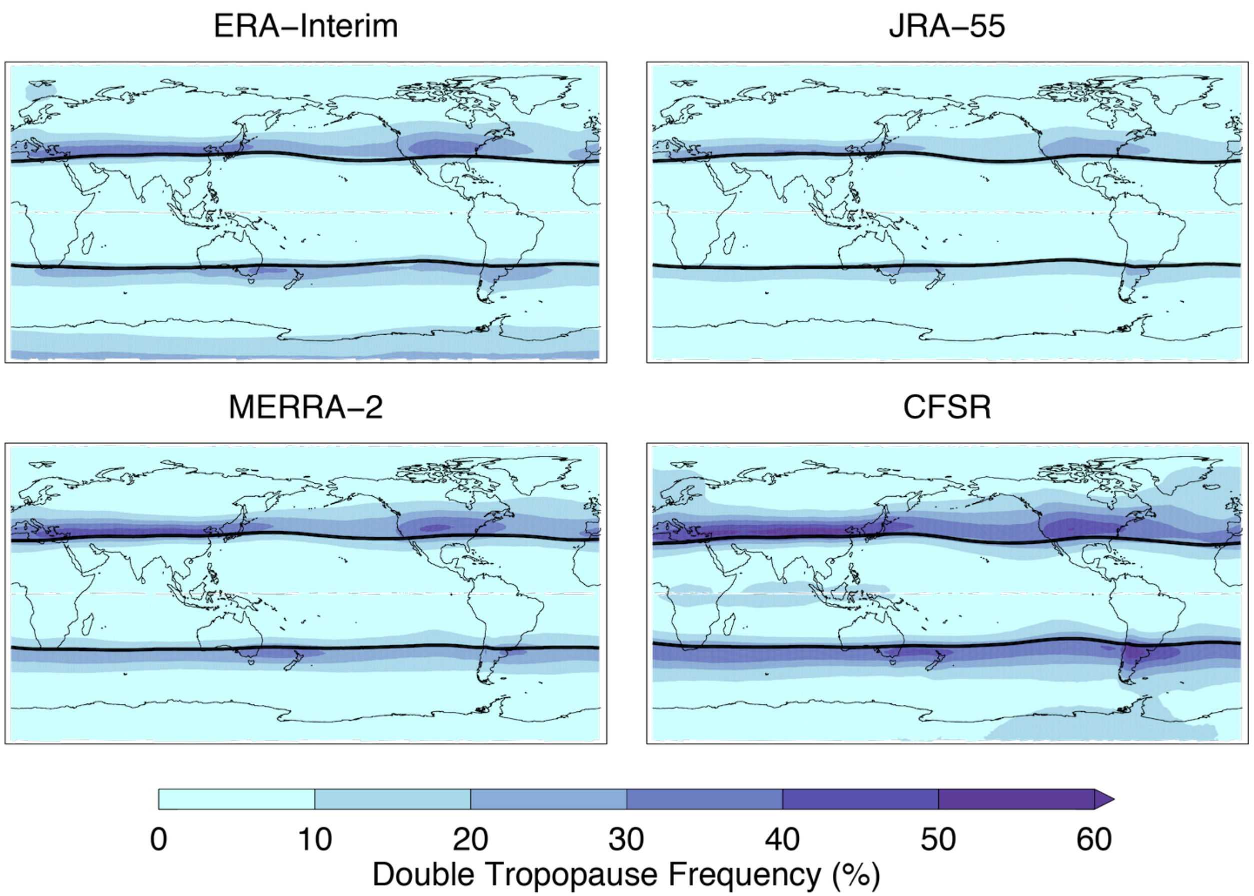

Figure 8. As in Fig. 6, but for reanalyses only in a tropopause break-relative coordinate.

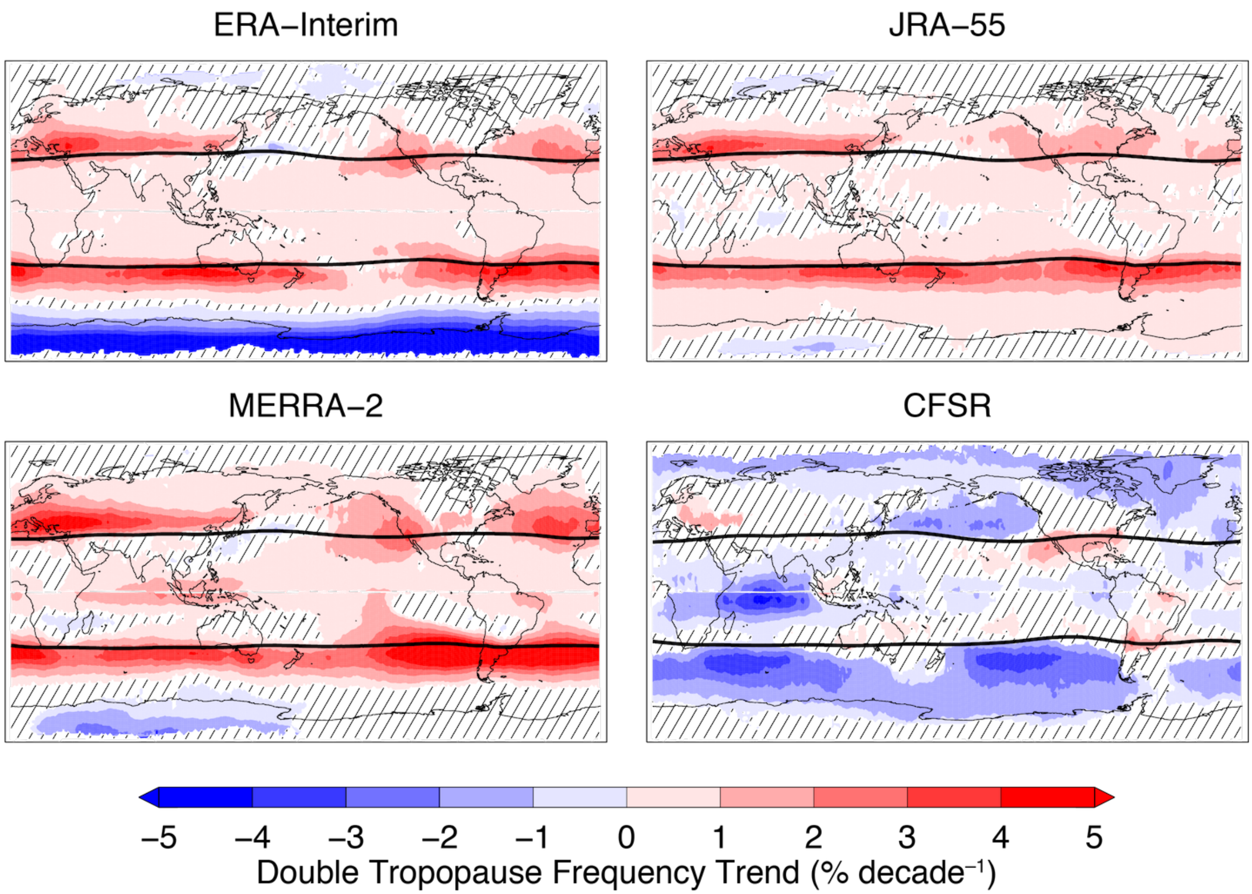

Figure 9. As in Fig. 7, but for reanalyses only in a tropopause break-relative coordinate.

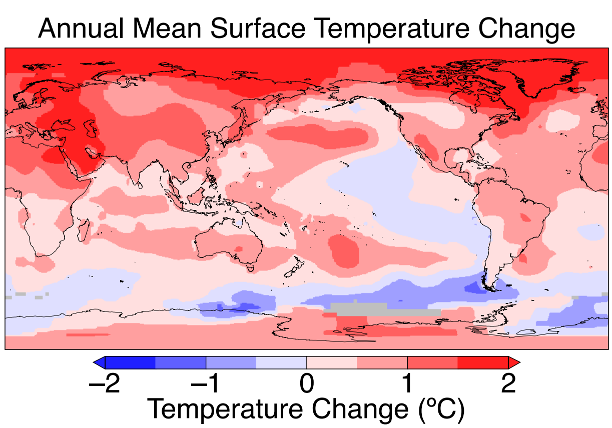

Figure 10. Observed 35-year (1981-2015) annual mean surface temperature changes from the NASA GISS surface temperature analysis.

Table 1. Bias and root-mean-square differences between reanalysis tropopause identifications and radiosonde tropopause identifications. Bias is defined as reanalysis − radiosonde, with the frequencies of positive and negative bias in parentheses. The number of radiosonde profiles used for primary tropopause altitudes and double tropopause frequencies is 99,023 and 45,181, respectively.

	Primary Tropopause Altitude			Double Tropopause Frequency		
	Positive Bias (km)	Negative Bias (km)	RMS difference (km)	Positive Bias (%)	Negative Bias (%)	RMS difference (%)
			Complete Sonde Data			
ERA-Interim	0.584 (45.589 %)	-0.436 (54.411 %)	0.973	2.156 (18.816 %)	-18.696 (81.184 %)	20.922
JRA-55	0.657 (43.565 %)	-0.447 (56.435 %)	1.030	0.684 (14.786 %)	-22.414 (85.214 %)	25.291
MERRA-2	0.599 (54.552 %)	-0.506 (45.448 %)	1.019	3.725 (23.332 %)	-17.271 (76.668 %)	19.386
CFSR	0.508 (45.944 %)	-0.485 (54.057 %)	0.961	6.525 (32.662 %)	-13.570 (67.336 %)	15.200
			Degraded Sonde Data			
ERA-Interim	0.464 (50.844 %)	-0.428 (49.156 %)	0.939	1.117 (32.772 %)	-14.454 (67.228 %)	14.720
JRA-55	0.538 (47.669 %)	-0.429 (52.331 %)	0.974	0.265 (26.641 %)	-17.877 (73.359 %)	18.912
MERRA-2	0.515 (57.610 %)	-0.511 (42.391 %)	1.021	2.109 (39.919 %)	-13.561 (60.081 %)	13.451
CFSR	0.437 (49.782 %)	-0.507 (50.218 %)	0.950	3.421 (44.586 %)	-12.193 (55.414 %)	12.001

The values given are root-mean-square differences and mean differences between reanalyses and radiosonde observations.