Author response to referee comments on "Tropical upwelling as seen in observations of the tape recorder signal" by Brehon et al.

We thank both reviewers for their helpful comments which have helped to improve this manuscript. We have made edits according to the comments. The comments are reproduced below followed by our responses in italics.

Anonymous Referee #1:

- I suggest using 'Tropical *stratospheric* upwelling…' in the title.
    - *Thanks for the suggestion. This change has been made.*
- It would be good to know more about the details and uncertainties of the MLS H2O upwelling calculations. The authors choose to use correlations between vertical levels ~ 4.5 km apart, so that the results represent mean upwelling over these broad layers. This detail is not well emphasized, and vertical profile results are shown with a 1-km grid (e.g. Fig. 1). What do the upwelling calculations show if narrower vertical layer differences are used (even between adjacent levels)? How sharply peaked are the lag correlations in time, and is there any corresponding information on uncertainties in derived upwelling? The 6-month window calculations look reasonable in the reanalysis upwelling comparisons, and why aren't these used throughout the paper?
    - *The choice to use broad layers for the upwelling calculation was made to minimize the uncertainty introduced through the method. Narrower separation between the layers was found in some cases to produce undesirable behaviour in the correlation curves, like peaks for very small lags, and multiple peaks. For instance, unrealistically short lags are sometimes found when applying the analysis to adjacent levels, possibly related to the data at these levels not being completely independent since the vertical resolution of MLS varies between 3.5 km to 3.8 km in the vertical range we are looking at, whereas the distance between adjacent levels is closer to about 1 km. In taking the calculation instead over broad layers, these undesirable features are avoided resulting in smaller uncertainties and fewer unphysical features in the time series. For sensitivity testing the upwelling calculation was done over more narrow altitude ranges and we find that, despite some unphysical features, the mean values and variability are largely the same and therefore it was decided to use the method that minimizes the uncertainty. More discussion on the justification for using broad layers has been added to the methods section, and the fact that the calculation is done over broad layers has been emphasized when introducing the profiles.*
    - *While the 6-month method does produce results which are reasonable, the method using shorter portions of the water vapour time series is not as robust as the 12-month method (the undesirable features in the correlation curves are encountered more frequently). Because of this, the 12-month method was used for analysis of longer term variability, and the 6-month method was applied when interested in*

    variability on a shorter timescale in order to use the more robust method when possible.
    - *The broadness of the peaks in the correlation curves varies from sharp to more broad. To obtain some measure of uncertainty in the calculated upwelling resulting from the method, we have constructed a 95% confidence interval for each of the calculated correlation coefficients. From these intervals, we find the lag corresponding to the first correlation on either side of the maximum which is statistically different from the maximum. This gives a range in the possible value for transport speed, the size of which varies based on the particular correlation curve. Sensitivity tests showed that there were negligible differences in the uncertainty if the confidence interval was instead taken at 90% or 99%. This estimate for uncertainty is shown in the manuscript through shaded uncertainty added to Fig. 2.*
- I think including the SWOOSH results prior to 2004 is interesting, but the results look problematic to me given the data gaps and much noisier character of the time series seen in Fig. 2. Figure 9 highlights poor coherence between adjacent pressure levels and poor agreement with reanalyses for the early period. What are the causes of the data gaps? Why is additional time smoothing (3rd order polynomial fits) needed for these data? A 'spike' near 2001 in the derived upwelling is discussed but this looks more like noise in the calculations to me. In my opinion the authors should be more critical of these issues in the results derived from SWOOSH.
    - *The data gaps in the upwelling calculated from SWOOSH result from the poorer correlations found between subsequent levels in SWOOSH when compared to MLS for the pre-MLS portion of the time series. For this early period, there are cases where the maximum correlation found for the varying lags is less than 0.6. In these cases, in order to avoid finding the transport speed from portions of the time series that are only weakly correlated, we assign NaN instead of a value for transport speed.*
    - *The additional smoothing is performed in order to recover a continuous signal of the tape recorder for the early portions of the time series where there are many gaps as a result of the data coming from solar occultation instruments with sparse sampling. The effect of this step is to interpolate the gaps and recover a signal that can be used with this method.*
    - *Thank you for pointing out these issues in the SWOOSH upwelling dataset, we agree that the results for 1995-2006 are not as reliable as for the later portion of the time series. We have updated the text to better emphasize the reason for the data gaps and the need for additional data processing, as well as to be more critical of the results.*
- Figure 1 left axis should be Pressure, not Altitude (also Figs. 7 and 10). Do the error bars in Fig. 1 represent the standard deviation of the calculated means, or the standard deviation of the monthly time series?

- *The left axis labels have been corrected to read pressure instead of altitude. The standard deviations are of the monthly time series; the text has been updated to state this.*

- The regression results regarding QBO and ENSO impacts on upwelling are very similar to previous results of Abalos et al 2015 (doi:10.1002/2015JD023182)
  - *We thank the reviewer for bringing this to our attention. We added some text citing the Abalos et al. (2015) paper and stating that our results are very similar to what they found for upwelling variability explained by ENSO and QBO.*
- I'm curious about the detailed results in Fig. 4 – how can regression results with R2 less than 0.1 be statistically significant? How are the degrees of freedom evaluated for these low-frequency variations?
  - *The approach used in the paper used the same number of degrees of freedom for each regression. We have since updated this approach to adjust the degrees of freedom based on autocorrelation of the residuals following Santer et al., 2000 (their equation 6). The effective sample size is then used for the two-tailed p-test at 95% confidence. The figure has been updated with the new significance levels indicated, and the text has been updated to reflect this change in the method.*
- I like the comparisons of upwelling and ozone time series in Fig. 5, but shouldn't you use the ozone level closer to the midpoint of the upwelling calculation for the most meaningful comparison?
  - *Our approach is based on ozone at a specific level averaged over a period of time being mostly influenced by the upwelling below that level (transporting ozone anomalies from lower levels to this specific level). In other words, while the ozone at the top level of our upwelling layer 'sees' the upwelling from the total layer, ozone at a lower level within this layer would only 'see' the upwelling in the lower part of the layer. Therefore, we decided to correlate the upwelling with the ozone at the top of the layer.*
  - *We have added to the appendix a plot where the ozone is averaged over the upwelling layer, rather than simply taken at the top of the layer. In taking the average, we are incorporating the contribution from ozone everywhere that we are considering the upwelling. In doing so we find reduced correlations between upwelling and ozone.*
- The residual vertical velocity from reanalyses is given in terms of pressure vertical velocity – does this include the eddy heat flux term, or is it just the zonal mean pressure velocity? (there are small differences in the deep tropics, but this should be clarified).
  - *The residual vertical velocity is described by (Martineau et al. (2018)):*

$$\bar{\varpi}^{*} = \bar{\varpi} + \frac{1}{a\cos\phi}\frac{\partial}{\partial\phi}\left[\frac{\overline{v'\theta'}\cos\phi}{\partial\bar{\theta}/\partial p}\right] \quad (1)$$

  - *Here the eddy heat flux term is included in the form of the meridional eddy flux of potential temperature. This equation has been added to the text to clarify this point.*
- For more direct comparisons to the H2O results, I recommend calculating and plotting the reanalyses upwelling in terms of mm/s.

  - *We thank the reviewer for this recommendation. Figures 7, 8, and 9 have been updated to be in units of mm/s to be consistent with the rest of the paper.*
- The ANCISTRUS upwelling results (Figs. 10-11) seem overly smooth in terms of vertical structure and time evolution. While these estimates are derived from numerous chemical tracers, there are very small vertical gradients or seasonal variations in most of the tracers in the tropical lower stratosphere, and I frankly wonder about the information content in this region. Why is there no seasonal cycle in upwelling in the 100-46 hPa ANCISTRUS results in Fig. 11c? Also, the 6-month window MLS H2O results in Fig. 11 don't seem to match the corresponding results in Fig. 8 in terms of seasonal cycle variations. As with the other parts of the paper, I think it would be useful to be somewhat more critical of the ANCISTRUS comparison results.
  - *We thank the reviewer for this insightful observation. The apparent smoothness of the ANCISTRUS-derived upwelling in the tropical lower stratosphere is expected, because the 3-km ANCISTRUS grid contains only two vertical levels (21 km and 24 km) in the altitude range shown in Fig. 10. With such coarse vertical spacing, only very limited vertical structure can be represented by construction. The text has been updated to highlight the lack of vertical resolution for ANCISTRUS in this figure.*
  - *We have changed the format of Figure 11 to display the seasonal signal more clearly (see point below). The fact that the seasonal cycle in Fig. 11c is relatively weak arises since the results are averaged over six months, which inherently suppresses some of the seasonal variability. The weak seasonal signal therefore reflects the temporal averaging and not a limitation of the ANCISTRUS method. The same applies to the MLS derived upwelling.*
  - *The MLS time series plotted in Fig. 11 are the same data as has been presented earlier in the paper in Figs. 2 and 8, however here the aspect ratio of the plot has changed such that the time series are stretched farther horizontally, resulting in a visual smoothing of the signal. To try and make the figure appear more similar to prior figures, the aspect ratio of the plot has been changed to be more compressed horizontally.*

Anonymous Referee #2

MAJOR COMMENT:
As explained in the paper, the upwelling velocity estimate from the water vapour tape recorder presented here is an effective transport velocity which includes both advection by the residual mean mass circulation and mixing. I don't question that this is a valuable quantity to calculate, but I'm wondering about its comparability to the residual upwelling velocity w*. What is urgently needed for model and reanalysis validation is an observational constraint for w*. Whether the effective velocity presented here serves as a meaningful constraint is not clear to me and should be worked out in more detail.

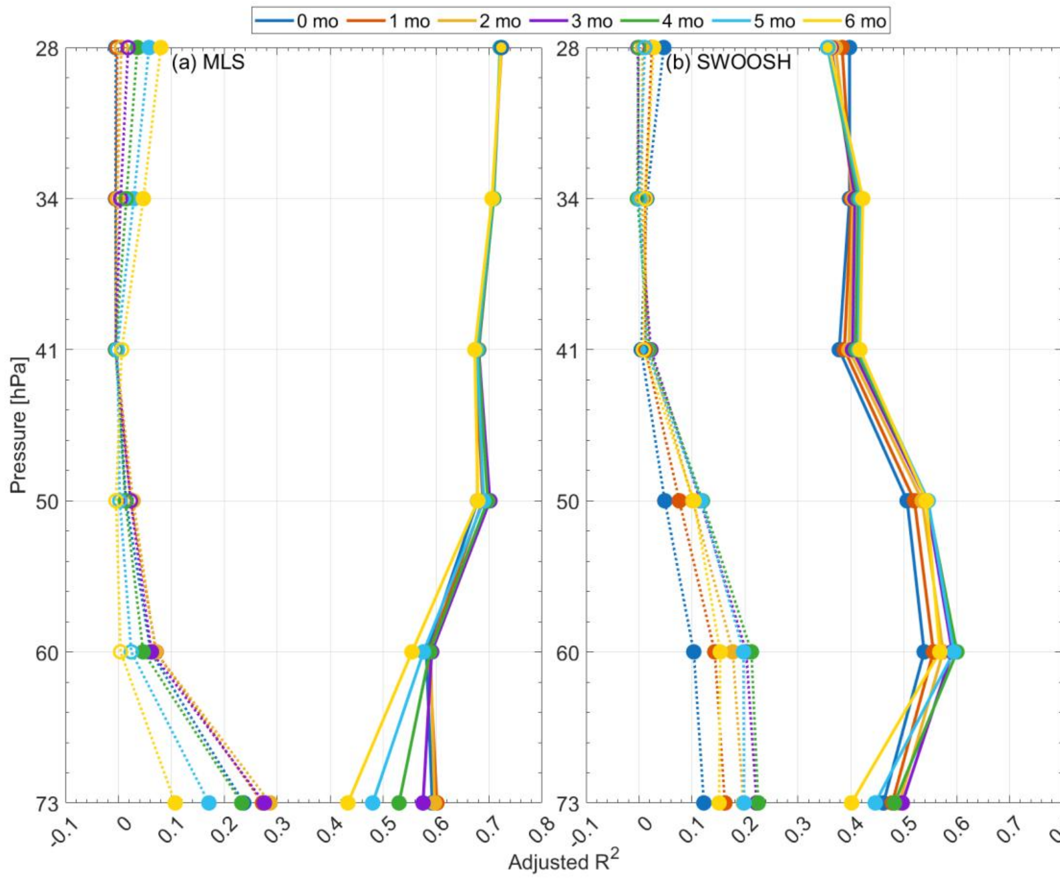
Qualitatively, the effective upwelling appears to be very consistent with residual circulation upwelling. However, there are quantitative differences, e.g. regarding the mean profiles (Fig. 7) or the seasonal cycle (Figs. 8, 9). Are these quantitative differences significant given the differences in the methods (effective upwelling vs. residual circulation), methodological uncertainties (e.g. measurement errors in H2O, lag time calculation over broad layers), and the variability around the means (e.g. there is large variability around mean profiles in Fig. 7, also see specific comment below)? In other words, can a quantitative recommendation be given on the reanalyses considered here (e.g. MERRA2 upwelling being too slow)? Or are all uncertainties together so large, that only qualitative consistency remains as a result?

In my opinion, the "golden way" would be to carry out a "proof-of-concept" study within a controllable model environment first, where all quantities are well-known. This would imply calculating effective velocities from model water vapour tape recorder and comparing with the model w* velocity. If such an analysis is beyond the scope of this paper, at least a thorough discussion should be included, as to what degree the effective upwelling can be considered a quantitative constraint for the residual circulation, and hence how to interpret the differences to reanalysis w* presented here.

- *We thank the reviewer for this comment, whether the tape recorder derived vertical velocity is a good observational constraint on $\overline{w}^*$ is an important question.*
- *We began to investigate this question but found that unrealistically strong dispersion of water vapour in reanalysis (Glanville and Birner, 2017; Linz et al., 2019) creates an incompatibility between the residual circulation and the tape recorder derived vertical transport. The reanalysis used did not assimilate water vapour in the stratosphere, but the strong impact of the dispersion made the suggested 'proof-of-concept' impossible for this particular data set. Similar issues could be encountered in other reanalysis and model simulations, where the question of dispersion of the water vapour signal would complicate the comparison with observational data. In consequence, a 'proof-of-concept' study would only be possible with a dataset in which the mixing and dispersion processes are highly realistic and all quantities are known, we therefore decided not to include a study of this kind in the revised version of the manuscript.*
- *We have added a discussion to the manuscript to address this point, using the results of previous studies to explain how successfully the tape recorder derived transport speeds simulate the residual circulation.*

- I was wondering how much the assumption of a fixed lag in the ENSO regression would influence the results (e.g. L195ff). There are other studies, as e.g. cited the Diallo et al. (2018, 2022), which use variable lag time for the MEI index. I'd find it helpful if potential influences of the fixed lag in the regression were discussed. A full MLR with lag optimization is not needed here in my opinion, but perhaps some sensitivity test of the shown results regarding different lag times.
    - *The value of a fixed 2 month lag was determined based on the timing that would produce the best results for minimizing the residuals (improving the R^2_{adj}) across the six different levels. Testing a lag of between 0-6 months, the variation between the maximum and minimum R^2_{adj} was 0.175, 0.036, 0.023, 0.007, 0.005, and 0.003 for the 100-46 hPa, 83-38 hPa, 68-32 hPa, 56-26 hPa, 46-22 hPa, and 38-18 hPa*

levels for MLS, and 0.099, 0.064, 0.041, 0.045, 0.031, and 0.037 for the same levels for SWOOSH. This information is also shown in the plot below, where the dotted lines are the R^2_{adj} profiles for regression with only ENSO, and the solid lines are the R^2_{adj} profiles for the full regression with all four regressors. The length of the lag chosen does not significantly alter the results here, and therefore we chose to use the fixed lag of 2 months across all levels. Some explanation for the justification of this choice was added to the manuscript.



- For relating upwelling and water vapour variability it would be particularly useful to consider upwelling at a level lower down in the tropical stratosphere, say 100-80hPa, as this should be most closely linked to cold point temperature variability. The Randel et al. (2006) study shows a persistent strengthened upwelling at 100hPa during the years after 2000 (e.g. their Fig. 9) correlated with anomalously low cold point temperatures. Perhaps, the fact that this is not seen here is just related to the broad vertical region over which the upwelling estimate is averaged, and because of that a missing link to tropopause temperatures? It would be good to discuss these points in more depth here, and perhaps even try the upwelling velocity calculation over a narrower vertical layer 100-80hPa.
    - *Unfortunately, the lag between the signal at 100 and 83 hPa is too small to be useful for this calculation method. These levels are only spaced about 1 km apart, whereas the vertical resolution of the MLS instrument in this region is around 3.5 km*

producing signals which are very close with minimal lag. For that reason, it is necessary to calculate the upwelling over these larger vertical regions.

Specific comments:

- L20: Anti-correlation holds only for long-lived tracers with stratospheric sources. For long-lived tropospheric tracers, e.g. N2O, correlation holds. Please clarify.
    - *Sorry for the confusion. The word "stratospheric" was removed from the sentence at line L20 in an effort to meet the character limit for the abstract. We thank the reviewer for pointing out the confusion this creates, and it has been added back to emphasize that this anti-correlation is only valid for long-lived tracers with stratospheric sources.*
- L94: As far as I know, SWOOSH v02.7 in addition includes ACE-FTS and SAGE III. Please clarify (also at later places in the paper).
    - *Sorry for the confusion. Yes, SAGE III/M3M, SAGE-III/ISS, and ACE-FTS are included in the SWOOSH dataset. The text has been updated to make this more clear when discussing the SWOOSH dataset following the start of the MLS data in 2004.*
- L127: According to Von Clarmann et al. (2016, 10.5194/acp-16-14563-2016), the ANCISTRUS model is based on an effective transport formulation (not a TEM residual circulation framework) and therefore provides effective transport velocities also if mixing is explicitly taken into account. If so, this should be clearly explained here (and text later needs to be adjusted accordingly).
    - *We would like to point out that ANCISTRUS does not diagnose the Transformed Eulerian Mean (TEM) residual circulation. Instead, ANCISTRUS derives effective transport velocities that represent the combined influence of advection and mixing needed to match the observed tracer tendencies. This is true regardless of whether explicit mixing coefficients are included in the inversion. In the current implementation, regularization suppresses the diagnosed mixing term, so mixing effects are folded into the transport term implicitly, but the diagnosed velocities remain effective velocities in any case.*
    - *To avoid suggesting that ANCISTRUS would produce TEM-comparable, mixing-free velocities, we revised and clarified the description in the manuscript. We also replaced the misleading "Therefore" in the original text with "In any case" to emphasize that the ANCISTRUS velocities are always effective velocities and not TEM residual circulation components.*
- L140: Please give an approximate level here, e.g. "up to about 10hPa".
    - *Done.*
- L142: "water vapour altitude" --> "pressure altitude"
    - *Done.*
- L154: When first reading I got confused here as to whether 12 or 6-month intervals are used in the study (became clear later that both are used). Perhaps make this directly clear by writing something like: "In addition to the calculation based on 12 months intervals, we also ..."
    - *Sorry for the confusion. The text has been updated to make this distinction more clear.*

- L173: I think MLS is not the only dataset contributing to SWOOSH v02.7 (see my comment above).
    - *Yes, similar to the above comment the text has been updated to remove confusion around this point.*
- Fig. 3 and L203ff: Why are different instruments shown for the different layers? If there is some reason, please explain. If not, I'd suggest to either show SWOOSH or MLS for both layers.
    - *The choice of instrument and layer was made to illustrate how the regression fit captured a large portion of the variability in the upwelling time series. In particular, at the lower layer for SWOOSH there is a good fit for the majority of the time series, except for the times where there are known sharp changes in water vapour. Similarly, the higher level for MLS illustrates the skill of the regression in capturing the variability.*
    - *The fact that we have chosen only a couple of examples has been made more clear in the text, and the regression fits and residuals for both instruments and every level have been added to the appendix.*
- L212: I was wondering why the decline in regression performance was not discussed here at all, unless I saw that it is then in detail discussed later. Perhaps mention that already briefly here and refer to the later sections.
    - *A sentence was added to make it clear that discussion on this point occurs slightly later in the text.*
- L256: "... such as found for ..:"
    - *Done.*
- L263ff (throughout Sect. 3.2): I miss some discussion of the effects of mixing in ozone-rich extratropical air into the tropical lower stratosphere. This process has been shown to substantially influence the tropical lower stratospheric ozone budget just above the tropopause (e.g. Abalos et al., 2013, 10.5194/acp-13-10787-2013), strongest in the NH tropics (Stolarski et al., 2014, 10.1002/2013JD021294).
    - *We thank the reviewer for pointing out this gap in the discussion. The effect of mixing on the ozone budget has been added to the discussion of how upwelling influences ozone here.*
- L295: "... reduced variability after 2015 ..."
    - *Done.*
- Fig. 7, L327: There is remarkably larger variability in the reanalysis profiles compared to the observational estimate in Fig. 7. Has the seasonal cycle variability filtered in the reanalysis data before? If the variability in reanalysis is about 5-times as large as in the observational estimate, this needs to be discussed.
    - *The large difference in the seasonal cycle variability between the reanalysis and observational estimate is likely a result of a combination of the impact of mixing on the observational estimate and the seasonal variation in this effect as a result of the seasonal cycle in mixing. Discussion addressing this point has been added to the text.*

- L347: Is the agreement in seasonal cycles really "good"? To me it seems that the seasonal cycle in reanalyses in Figs. 8, 9 is substantially larger than from observations. Perhaps a comparison figure or table with seasonal cycle amplitudes could make this point clearer.
  - *Good point, the wording has been changed from "good" to "reasonable". Here we are focusing on checking that the timing of maxima and minima align between the observational estimate and the reanalysis circulation. As we do find agreement in this regard we consider this good agreement. In addition, the differences in the amplitude of seasonal variability have been addressed with added comments to the text in response to the comments for Fig. 7, L327.*
- Fig. 8: It's hard to see the lines behind the symbols in the seasonal cycle plots and therefore almost not possible to distinguish between ERA5 and MERRA2. It would be helpful to include a legend for the symbols.
  - *Thank you for this suggestion to help with readability of the figures. The suggested change has been made.*
- L348: I find the relation to the shallow branch here somewhat confusing. As far as I know the deep branch of the stratospheric circulation shows a clear seasonal cycle with maximum upwelling during boreal winter, while the seasonality is not so clear for the shallow branch (c.f. Lin and Fu, 2013, 10.1029/2012JD018813; Baikhadzhaev et al., 2025, 10.5194/acp-25-12753-2025). In this context, it is important to note that the upwelling across a given level in the lower stratosphere, as is presented and discussed here, includes the mass flux related to both the shallow and deep branch. Please clarify.
  - *Sorry for the confusion, this is a good point, since this is the only place where specific mention is made to the shallow/deep branches of the circulation. The text has been changed to more generally refer to the stratospheric meridional overturning circulation.*
  - *The mass flux for the lower stratosphere does include contributions from both the deep and shallow branches of the BDC, with different sources in the literature defining different boundaries for the separation between the branches. The seasonal cycle of this circulation should be as described in the manuscript, with maxima in NH winter and minima in NH summer (e.g. Rosenlof, 1995).*
- L377: As said before, isn't this just a cause of the ANCISTRUS formulation and not only due to the neglection of mixing.
  - *Sorry for the confusion, the text has been updated to make this point more clear consistent with the previous comment for L127.*
- Fig. 10: Why are there only 2 error bars for ANCISTRUS?
  - *Sorry for the confusion. There are only 2 error bars because there are only 2 data points in the plotted region. The 3-km ANCISTRUS vertical grid contains points only at 21 km and 24 km within the range of the upwelling calculated from water vapour here.*
- L439: "... with long-lived stratospheric tracers ..." (see comment above).
  - *Done.*
- L458: "... about six measurement days ..."?
  - *Done.*