Reviewer #2 comments in plain font.
Author response in bold font.

We thank the reviewer for carefully reading our manuscript and pointing out a number of points that needed clarification. We specifically would like to thank this reviewer for sharing her/his perspective on the issue raised by reviewer 1.

We have incorporated more detail about the specifics of the model settings and its evaluation, along the lines of the reviewers' major comments. Our specific changes and responses to the reviewer are summarized below.

We also attached a revised manuscript with highlighted changes as supplement.

Major:

One primary conclusion of this paper, that vertical mixing in the tropical lower stratosphere must be four times larger than the value estimated by Mote et al (1998), rests on the 1-d model simulations. While the model is fairly simple conceptually and is likely to be a useful tool for this kind of analysis, the devil is in the details which are not completely explained, and the impact of the certain assumptions embedded in the model are not fully explored.

1. Seasonality is introduced in the parameters omega, K, and alpha by prescribing reductions and enhancements of 50% over the course of a seasonal cycle. There is no discussion of why a 50% variation is a valid assumption. Is there observational evidence to support fixing this amplitude? How are results impacted if one chooses, say 30% or 70% amplitudes? Are these prescriptions sinusoidal seasonal variations? For the phases of seasonal cycles, the paper has some discussion that justifies the choices based on models or observations; however, "boreal winter" and "boreal summer" are given instead of dates or months, which would be preferable. For example, saying that horizontal mixing maximizes during boreal summer likely refers to the July-August period and not the NH summer solstice. Gettelman et al (2011) discuss the importance of horizontal mixing during July-August (e.g. the Asian monsoon anticyclone).

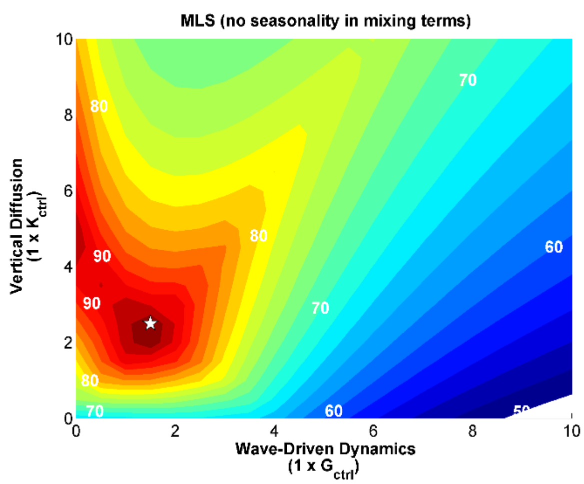
> **The following was motivation for choosing 50% seasonal variance in the 3 terms:**
>
> **For vertical advection we oriented ourselves at Rosenlof (1995) and Abalos et al. (2013). We feel that these and other references on the subject constrain the variations for vertical advection (w*) strongly enough that 50% seems a solid choice. Also note, our results are very similar in terms of the w* tendency in ERA-i (dashed red lines in Figure 7).**

For horizontal mixing (α) we oriented ourselves at Gettelman et al. (2011) and Ploeger et al. (2012, see reference in revised manuscript). It is clear from these and other references in the literature that horizontal mixing is stronger during (late) boreal summer.

For vertical mixing (K) we primarily referred to Flannaghan and Fueglistaler (2014), who indicate more vertical mixing during DJF but the seasonal cycle amplitude is uncertain.

The 50% choice is admittedly less obvious for both mixing terms. We used it for simplicity but also note the following.

We tested the model without seasonality in the two mixing terms (but still with 50% seasonality in vertical advection) and the resulting scores for the MLS tape recorder (at 80 hPa) can be seen in the figure below. Removing the seasonal cycle in K and α results in needing 50% stronger w* for the top-scoring solution (which seems unrealistic based on literature listed above). The best simulations (>90%) still require amplifying K by at least a factor of two. There is a slight "fork" with the warm colors seen in the figure below, both requiring amplified K.



Overall, this and other tests we performed reveal that the seasonal cycle in vertical advection is most crucial and this is the one that is also best constrained by past literature. Our main qualitative result that vertical mixing is as important as vertical advection does not seem to be very sensitive to the choice in seasonality-strength in K and/or alpha. Incorporating a 50% seasonal cycle to the mixing terms slightly narrows down the solutions, perhaps bringing them closer to reality.

Also, the cycles are based on a sine wave, and the peaks occur in the middle days of January and July – this has been clarified in the text.

2. The model's score is based on comparisons with the amplitude, phase, and annual mean of the observed water vapor mixing ratio at 80 hPa (or 400 K), but it is not clear how these are derived from the MLS data. Is this from a simple FFT analysis? If so, Fig 4 indicates that the seasonal variations are not exactly sinusoidal, so how does this impact the analysis if a different functional form is used, one that better simulates the seasonality of the effective transport velocity? In this regard, is there any explanation for why the MLS velocity in Fig 4 has a double minimum, or is the spring dip just noise?

Thanks for pointing out need for clarification. Correct, the phase is calculated using a simple FFT analysis. But the amplitudes are obtained from the minimum and maximum values. Clarifying sentence has been added.

We feel that the climatological seasonal evolution of water vapor is sufficiently sinusoidal (e.g. Fig. 6 bottom) that this simple FFT analysis to obtain the phase is adequate. Note that Fig. 4 is a plot of vertical velocities, not the water vapor evolution (the latter is used to obtain the score).

We believe that the spring dip in MLS effective vertical velocity in Fig. 4 is due to noise. By testing the wEff method on synthetic tape recorders with different vertical resolutions, we found that coarser resolutions resulted in more noise, especially for the transition between the wet and dry signals. Note added in section 4.1.

3. Use of a constant, 7-km scale height to convert from pressure velocity: This is not appropriate for a couple of reasons. First, temperatures near 70 hPa are about 200-210 K in the tropical lower stratosphere, so that the scale height is closer to 6 km. Second, there is a well-documented seasonal cycle in temperature that causes variations of 3-4% in the scale height, and this should be included in the calculation of effective transport velocities, particularly in examining there seasonal behavior.

We prefer to work with log-p coordinates, as this makes comparisons to models most straightforward (which usually run in p-coo.). This means that H needs to be a constant (no seasonal variations, otherwise we would not be working in a p-coo. anymore). We use H = 7 km simply because this seems to be the standard value that people use in the literature (and in text books, e.g. Andrews 1987), despite the fact (well-taken by reviewer) that 7 km is off in the tropical LS. We've included a clarifying comment in the manuscript and modified the Fig. captions.

Minor:

1. Abstract, lines 8-9: This seems to state that the seasonal cycle of residual velocity derived from MLS has a larger amplitude than that in ERA-i, which conflicts with results shown in Figure 4.

> **Thanks for pointing out - the sentence has been reworded.**

2. Abstract, lines 20-21: "as opposed to" implies an either/or scenario, whereas I think this paper finds that a combination of slow upward transport *and* rapid vertical mixing play a role in shaping the tape recorder signal.

> **Our "as opposed to" refers to the term "tape recorder", for which we do in fact mean to imply an either/or scenario: if transport is dominated by slow (vertical) advection then "tape recorder" is a justifiable term, but if mixing plays an important role (regardless of how important advection still is) then the term "tape recorder" becomes misleading. So we wish to leave the sentence as is.**

3. p 3, first paragraph: The latitude averaging for MLS data should be presented here, along with a discussion/justification of the choice of latitude bounds (appears to be 10S-10N from figure captions).

> **Thanks for pointing out lack of clarity. 10S-10N is a common choice for the inner tropics – in our case it makes sure we have sufficient sampling and cover the latitudinal variations in the location of maximum upwelling. We didn't find much sensitivity to making the latitude band slightly bigger (15S-15N). Text has been added in section 2 to clarify.**

4. p. 4, lines 30-32: As correctly noted, the effect of methane oxidation is primarily an additive constant. This can be easily accommodated by looking at anomalies for the MLS data analysis, or by a simple parameterization of "S" in equation 1 for the 1-d model. Thus, this reason alone does not seem to be a valid motivation for restricting the analysis to altitudes less than 21 km (~40 hPa).

> **We agree with the reviewer and appreciate the idea how to circumvent the complications due to methane oxidation at higher levels. However, we are particularly interested in the region just above the tropopause, which has been less studied from a tape recorder perspective and where vertical mixing may play a bigger role. We agree that the way we stated our motivation is misleading and have reworded the statement accordingly.**

5. p. 5, line 32: The midlatitude reference mixing ratio should be allowed to vary seasonally for a correct model simulation. If that is the case, it should be clearly stated here.

We have added to the text: it does vary seasonally.

6. p. 8, lines 18-20: “while its phase relies more on string enough vertical advection and on allowing for transport seasonality” is unclear. Is this saying something about simulating the phase of the tape recorder? If so, what is “strong enough” and for which transports (advection, vertical mixing, or horizontal mixing) are the seasonality important?

Thanks for bringing up need for clarification. Yes the statement refers to simulating the phase on its own. “Strong enough” refers to scores over 90% for each individual measures (phase, amplitude, and annual mean). We found that different swaths (of factors beyond the control) can satisfy those measures when assessing their scores individually. For example, the amplitude alone scored best with 3xK_ctrl while the phase alone scored best with a variety of factors (2-6xK_ctrl). However, the phase had a narrower swath of best simulations when analyzing it in terms of vertical advection (w*). It’s also the seasonality of advection that matters most. Sentence has been reworded to clarify.

7. p. 10, lines 7-12: A lot of the notation needs to be clarified in the equations, e.g., what do the hat symbols represent?

Thanks for pointing this out; notation has been clarified in the revised text. However, beyond the hat symbols (and the primes earlier in the text), we didn’t find any other notation that needed clarification. Overbars and asterisks had already been introduced after Eq. (2).

8. p. 10, lines 27-32, and Figure 9: First, it is not obvious why we should care much about vertical profiles of derived vertical eddy fluxes. The “sanity check” rationale is a stretch, as the vertical gradient only gives consistency with the 1-d model to within a factor of 10, and upon closer inspection, the negative tendency shown in Fig 7 for the vertical eddy mixing in boreal winter should correspond to a negative slope in Fig 9 for DJF at 80 hPa, which is clearly not the case. Thus, it appears that there are very large errors in the calculated eddy fluxes (perhaps as expected when taking differences between two quantities with large inherent uncertainties). A more robust discussion of the uncertainties in these results is warranted, along with a more complete analysis (e.g. comparison with previous studies, or what has been used in the past in 1-d models) of calculated eddy fluxes.

Fair enough, we agree with the reservation by the reviewer about this section. Our primary motivation to include it is that observational estimates of vertical eddy tracer fluxes on a zonal-mean scale are essentially non-existent. But they are required to be able to quantify more accurately the role of vertical mixing. Despite the large errors in our estimated fluxes, we feel it’s useful to include these results as they might inspire future research in that direction. We are not aware that our theoretical approximate formula derived in the appendix has been pointed out or used before, so the hope is

that it could be useful for future studies. At the least we feel that the idea to parse out information about vertical mixing by comparing pressure (or height) to isentropic coordinates is novel and the related theoretical discussion may be insightful to some readers.

The section has been revised, emphasizing the uncertainties more.