The authors thank the reviewers for their thorough, detailed, and insightful reviews. We will use all these recommendations to improve the clarity of this paper. In the following we insert our replies directly into the review.

Review #1:

Review: General Comments
von Clarmann et al. 2019 present results from an inverse method which uses observed (MIPAS) zonal-mean tracer fields to calculate residual circulation fields which are resolved in altitude, latitude, and time. This work expands on the work of von Clarmann and Grabowski 2016 (hereafter CG16) by providing time-resolved circulation fields, and continues the line of investigation of a number of other studies which have sought to constrain the strength of the residual circulation. However, the present work seeks to provide a substantial expansion in this direction by quantifying the circulation strength in terms of two-dimensional, time-resolved velocities. Only one previous work, to my knowledge, has quantified velocities at all - that being Fu, Hu, and Yang 2007 GRL - but this was only for a single profile of upwelling, while other work has provided some sense of two-dimensional motion but without velocities, such as the work of Stiller et al. 2012.
The results show several inconsistencies with current theory. For example: The mesopheric overturning circulation is considerably higher (at least 10 km, which seems very unexpected) when southward-bound as opposed to northward-bound; the tropical pipe shows quite a bit of meridional movement rather than isolated upwelling. If reliable, such results would be of substantial and immediate interest for a large section of the middle atmospheric research community.

Reply: The authors thank the reviewer for the appreciation of our results.

Planned Action: None to this point.

Review: However, there are considerable issues with the

validity of the results, and I do not think the work should be published until they are addressed. I outline them in the following three points:
1. The inverse model robustness (specifically in terms of sensitivity to input fields) has not been explored. My impression from reading CG16 is that the inverse method requires multiple tracers but that the limit on the minimum number of tracers needed is rather soft (i.e. it is not strictly necessary to have X or more tracers). I suppose that it is possible to use a subset of the nine tracers applied, and thereby estimate the robustness of the method with respect to input data. In my opinion, having even a simple estimate is necessary. Having read this paper, I am left without an idea of how the results depend on the input tracer fields. Even if the method is mathematically sound, there may be biases in the results which extend from errors in tracer measurement or the calculation of chemical sources and sinks, and until this possibility is addressed the results remain in a somewhat skeptical position. In particular, I would propose something like a jackknife test, calculating the fields with only eight tracers, excluding one tracer for each calculation, each tracer in turn, and seeing how strongly the velocity fields vary. It may not need to be a recalculation of the entire approx. decade-long climatology, but I think a test like this (or something more advanced) in multiple seasons is necessary to establish the validity of the method.

Reply: We have meanwhile a validation paper of the ANCISTRUS (Analysis of the circulation of the stratosphere using spectroscopic measurements) method available, which is ready for submission. It includes an assessment of the relevance of sinks versus transport of patterns, the jackknife tests, model recovery tests, and an assessment of the adequacy of the regularization strength chosen. In a nutshell, the results are: below 30 km ANCISTRUS is quite accurate in a quantitative sense. Above, due to the regularization of the inversion along with less measurement information, peak velocities are underestimated. All structures and patterns, however, are nicely reproduced. Since none of the conclusions of the paper under review are fully quantitative but are related to structures and patterns, we are confident that the validity of our method for the purpose of our paper has now been sufficiently established, and it can safely be excluded that the patterns detected are mere artefacts.

Planned Action: The draft of the validation paper will be made available to the reviewers. In the climatology paper we will make reference to the results of the validation paper.

Review: 2. The inverse model accuracy has not been quantified well. In CG16, an accuracy test was performed where dynamical quantities were prescribed in simple distributions, which showed that the method had some inaccuracy in the center of the domain and much more inaccuracy on the borders.

Reply: The inaccuracies at the borders were caused by the fact that the test fields, which were chosen in an ad hoc manner, did not satisfy the continuity equation. A method that allows only solutions which do satisfy the continuity equation will never be able to reproduce such fields. That is to say, the problem was the test fields, not the method. More adequate tests are included in the validation paper mentioned above.

Planned Action: See above.

Review: I think a test with dynamical and chemical fields that are more similar to observations (i.e. less homogenous and, simply-said, messy) is necessary to ascertain the uncertainty of the results. I would suggest using CCM model data to produced inversted circulation fields and compare them with the actual, directly-calculated model fields. To be clear, the results of such a test could not be used to pin a quantity directly onto the method's results (since, for example, the inverse method includes some effects of mixing, although knowing how the "effective" velocities could differ from the actual velocities would be valuable). However, this would provide some level of confidence in the results where, at the moment, the only indication extends

from an example which simply does not resemble the observed atmosphere. As a final note on this, it would be best if the test could also examine the difference between CCM and inverted velocities, and not the residual between tracer distributions (as was done in CG16). The velocities are what matter, after all. As an example of possible data, the ESCIMO (Joeckel et al. 2016 GMD) simulations have a variety of chemical species included (all the chemicals used in this study are in the model output) and calculations of the residual circulation strength have also been performed. I am not involved in that work, but I would guess that the members of that project would be interested in sharing the data.

Reply: A validation with realistic test fields has been performed and included in the validation paper mentioned above. Comparisons to models are interesting in their own right, and this is on our agenda for the future. We think, however, that validation does not necessarily rely on model data.

Planned Action: See above.

Review: 3. The inverse model result uncertainty has not been quantified. This is a rather small point, and not as important as the previous two. In CG16, quantifications of uncertainties of wind and diffusion coefficient fields were shown (in Figure 5) for fields computed using MIPAS data. Those uncertainties seem rather small, but it seems imprudent to exclude an estimation of uncertainties when a method for estimating them is possible. I suggest including a description or depiction of those uncertainties, if the method still applies for the present work.

Reply: The method is still applicable and we have the data available, but we consider this as largely redundant in this context. We present the standard deviations of the monthly averages, and this quantity includes both the uncertainty of the inversion and the year-to-year atmospheric variability. Thus, the standard deviations characterize how well the climatology of a month can represent any particular month of the sample. The CG16 estimates of the random uncertainty represent the mapping of the measurement errors on the inferred velocity fields only, but not the representativeness problems due to natural variability. But since both types of uncertainties are independent from each other and add up quadratically, small standard deviations indicate that also the measurement-error-based uncertainties must be small. The standard deviations are even a more reliable estimate of the upper bound of the precision because they do not depend on any assumption on the measurement uncertainty.

Planned Action: We add to the manuscript: "(To diagnose this effect, also the standard deviations of the circulation vectors, which are a measure of their variability, are shown in Figs xy–xy.) This variability is caused by the natural variability of the circulation over the years and its random uncertainty. The latter is the random uncertainty of the MIPAS measurements propagated onto the circulation vectors.

Review: In my view, the first two points are necessary before these results should be published. Without a clear indication of the model validity, the novel results shown in this work seem as likely to be artifacts of the inverse method (or the calculation of chemical balances) as they are to correspond to reality.

Reply: The validation is reported in a separate paper, see above.

Planned Action: As stated above, the draft of the validation paper will be made available to the reviewers.

Review: Furthermore, I find that the figures in the manuscript are difficult to interpret, and should be changed before publication. I have made more specific comments on this topic below. In general, the figures provide a qualitative idea of the circulation, in that I can examine the figures and know where the circulation is the strongest and which way it is headed at any one time and know how that strongest location changes with time. That does provide a general characterization of the circulation, but knowing how the

circulation strength changes with time in each location is of equal value to knowing where the circulation is strongest. The changing color scales in the first 12 panels makes that assessment difficult.

Reply: While we think that a qualitative empirical representation at this detail level is still unprecedented, we do agree that the figures should be improved.

Planned Action: Circulation patterns will be represented on a common color scale. The original data will be made publically available, allowing each user to plot them in their preferred representation.

Review: The figures showing standard devations need to be explained clearly, because they are so unusual, but do not seem to be properly explained anywhere. I think the figures should be remade showing contours or heatmaps, or some combination of both methods, so that readers can assess the strength of the vertical and meridional velocities and the variability of them. Maybe including stream lines would maintain the ease in understanding the flow direction. I understand that this likely means a doubling of the number of figures (assuming there is not a clever way to combine the meridional and vertical information), but I think that it is warranted for the clarity of interpretation.

Reply: We have decided to follow the suggestion to represent the variabilities in a different way.

Planned Action: Separate plots for variabilities in v and w will be provided.

Review: Furthermore, I have the following two points which I think warrant explanation, but are not addressed in the text: Why do boundaries seem to be included in results, when they were problematic following CG16? Are they somehow excluded?

Reply: No, boundaries are not excluded. These problems in CG16 were not a problem of our transport and inversion schemes but they were due to the simplified velocity fields used there which were not consistent with the continuity equation. E.g., if I have a poleward flow but no backward flow it does not come as a surprise that funny things happen near the pole. In CG16 such simple velocity fields were used because they allowed much simpler tests whether the transport scheme does what it is supposed to do, e.g. if a feature is transported with the right velocity etc. Applied to real data this type of problem does not exist because the true concentration fields are well described by the continuity equation, and the Earth's air is not forced to accumulate at the pole because of unrealistic velocity fields, or to do other funny things.

Planned Action: None specific to this point. We will, however, summarize the results of the validation paper and discuss which kind of conclusions are supported by the data.

Review: How were the sources and sinks of each species calculated?

Reply: The general description of sources and sinks was already described in Section 2.1.2 of the original paper. Some missing information will be included, particularly the assumptions on the abundancies of involved species.

Planned Action: See comment on line 78.

Review: To conclude, I would like to stress that results from this method should be published in the future if the reliability of the method can be addressed. They would be of very substantial interest, and therefore I wholeheartedly recommend resubmission.

Reply: A validation paper is ready for submission. It confirms that the conclusions in the paper under review are robust.

Planned Action: A draft validation paper is now available, see above.

Review: Specific Comments
Line 14: Neither Brewer nor Dobson suggest upper and lower branches to the BDC, nor the mesospheric overturning circulation. I think it would be best to cite somebody here who talks about that. Maybe Butchart's 2014 review paper would be good, to provide a general reference.

Reply: Agreed.

Planned Action: The related paragraph will be rewritten and the Butchart (Rev. Geophys. 52(2), doi10.1002/2013RG000448, 157-184, 2014) reference will be included.

Review: Line 20: The new sentence in this line seems to dismiss the value of having a single estimate of the upwelling mass flux / intensity. Perhaps it's just the use of the words "merely" and "far" (more/too). Im not sure if anyone has ever suggested that a single quantity could sufficiently characterize the circulation, but certainly the upwelling mass flux has provided a lot of value as a broad estimate of the circulation. I suggest simplifying the sentence - "These studies suggest that the ... is too complicated and detailed to be fully characterized by a scalar intensity." - or something similar.

Reply: agreed to remove the words which make the sentence sound dismissive.

Planned Action: These words will be deleted.

Review: Line 56: I don't understand what is meant by point b. Stiller et al. also uses MIPAS data. What does the present method do differently? Since the chemistry of SF_6 is not considered, any chemistry that is actually happening would create biases in those regions where it occurs (but I am no expert on this, and cannot say if those regions are included here). As a secondary point, there is no over-aging involved in this method because ages are not

calculated. What would you expect in the case of your inversion method, if you had this bias? I would assume the result would be a faster lower-strat to upper-middle atmosphere circulation.

Reply: The main SF_6 depletion happens in the mesosphere. Thus in the following we refer only to air parcels travelling through the mesosphere and back to the stratosphere. The method by Stiller et al., as well as earlier studies using this approach, are sensitive to the destruction of SF_6 along the entire trajectory of an air parcel from the stratospheric entry point, through the stratosphere and mesosphere and back into the stratosphere, because the age is calculated by comparison of actual SF_6 stratospheric mixing ratios and past SF_6 mixing ratios at the entry-point. Our method is different in that we use SF_6 mixing ratios measured at the upper boundary as a reference for gradient calculations in the uppermost model layer. Thus, any calculation of differences used for the calculation of the gradients (needed to solve the transport equation) is based on SF_6 mixing ratios which are, if coming from the mesosphere, already depleted in SF_6. Thus, mesospheric SF_6 destruction cannot lead to artefacts in the gradients. The only SF_6 loss that we possibly miss is SF_6 destruction within the **stratosphere** from one month to the other, which is a minor inaccuracy compared to the problem of mesospheric SF_6 depletion in age-of-air applications. In short: The reference SF_6 used by Stiller is the (past) tropospheric SF_6 while our reference SF_6 is the depleted SF_6 in airmasses intruding from the mesosphere.

Planned Action: A clarifying sentence will be added.

Review: Line 78: Where did the data for OH, $O^1(D)$, and Cl come from? As I understand it, those JPL publications only contain reaction rate and cross-section information. Did you obtain that data from somewhere else (does MIPAS have those species?) or did you model those in some other way?

Reply: We estimate OH using the parametrization by Minschwaner et al. (Atmos. Chem. Phys. 11(3), 955-962, doi10.5194/acp-11-

955-2011, 2011). $O^1(D)$ is estimated using the equilibrium equation 5.38 in Brasseur & Solomon (2005, Springer), applied to MIPAS ozone; Cl is estimated by interpolating a climatological noon profile (Brasseur & Solomon, 2005, Fig 5.50) to the actual atmospheric state. Inaccuracies in the latter are considered to be less important because the atomic Cl sink is of much less relevance than the other sinks.

Planned Action: This missing information will be provided in (old) Section 2.1.2.

Review: Line 86: How necessary is the stabilization of the inversion, that mixing coefficients were assumed to be zero? Can you compute mixing coefficients for even a single pair of months, or perhaps for a pair of boreal summer and a pair of boreal winter months? Otherwise, it's difficult to say how the effective velocities compare to advective velocities.

Reply: We think that the effective velocities represent the essence of the Brewer-Dobson circulation in the sense that they conflate all the effects (advection, correlation effects, mixing) that bring, in a 2D world, a trace gas from here to there. From comparison with zonal mean advective velocities we can learn about the relative contribution of the non-advective terms.
Technically speaking, all monthly results are inferred independently. That is to say, the instability does not come from accumulation of errors over the months but is inherent in the analysis of each single month. The cause of the instability is the following: The system of equations solved tends towards linear dependencies as soon as velocities and mixing coefficients are to be retrieved simultaneously. The matrix to be inverted has an extremely high condition number. This can, in principle, be remedied by regularization. We found out, however, that in this case (contrary to the velocity-only analysis) the result depends strongly on the chosen regularization and is, thus, not robust. As a consequence, we have decided to constrain the mixing coefficients to zero and to re-interpret the resulting velocities as those 2D-velocities which best describe the combined effect of advection, eddy transport and eddy mixing. It cannot be expected that these effective velocities equal the zonal mean advective

velocities.

Planned Action: A clarifying sentence will be added.

Review: Line 88: What does the word “efficient” in ”efficient 2D circulation” mean? Do you mean effective? If not, the meaning should be clarified.

Reply: This is a wording error and should read ‘effective’ instead. Thanks for spotting!

Planned Action: This will be corrected.

Review: Line 89: I have never seen the term “Fickian mixing” before. Having done some searching, I think what you are referring to is more commonly called diffusion. If that’s the case, I would use that term. Otherwise, the meaning should be explained.

Reply: We intentionally avoid the term ‘diffusion’ in this context for the following reason: ’Diffusion’ we understand is a physical process happening on a molecular scale. The processes we describe still abide to Fick’s law of diffusion but are macroscopic processes. Thus we consider the term ‘diffusion’ in this context as misleading.

Planned Action: We will add a footnote explaining the meaning of ‘Fickian mixing’.

Review: Section 2.3: Why do you average every pair of months? I would guess that is due to interannual variability of the phase (and I use that word very loosely here, perhaps it would be better to say timing) of the circulation. Whatever the case is, it should be stated.

Reply: There seems to be a fundamental misunderstanding. We do not average over two months. The velocity field labelled, say, March-April, is the velocity field that best reproduces the change of the monthly mean March mixing ratio field to the monthly mean April mixing ratio field.

Planned Action: We will add some clarification to avoid this misunderstanding. We think, however, that this clarification fits much better in the Section 'The General Approach' than here.

Review: Figures 1,2,3, and 4: I find these figures very difficult to interpret. First, the changing color scale does not seem necessary, as the maximum values do not vary strongly between the figures - most of them are around 11/12 - although it would create an issue for the December-January figure. At the moment, however, it is difficult to assess how the magnitude of the circulation changes at each point, month-to-month, except in the most starkly contrasting cases.

Reply: We do agree that the changing colour scales are disadvantageous.

Planned Action: New plots with fixed colour scales will be provided.

Review: Second, it is difficult (if not for most cases pratically impossible) to obtain even a rough magnitude of the vector components because the color scale shows a norm of the vectors. I suggest using contours or heatmaps of the separate vector components instead.

Reply: The length and direction of the arrows represent the velocity components. The colour scale was meant just as an additional guide of the eye. The length units of the arrows are ad hoc, that is to say, they are not consistent with the ϕ and z axes intervals of the plots.

Planned Action: The original data will be made available.

Review: Third - and this has nothing to do with interpretation - you show the boundary velocities of these results although it is clear from CG16 that the vectors at the boundaries are difficult for the inverse model.

Reply: As already stated above, within the reply to the general comments, the problem at the boundaries is not a problem of the inversion scheme but a problem of inconsistent test data which represent non-realistic circulations where air accumulates at the boundary of the domain. By the way, the arrows in the boundary tiles refer to transport in the 80-90 degrees latitude band (and not to the 90 deg. latitude; i,e., we have no northward transport at the North pole.)

Planned Action: None.

Review: Figures 5 and 6: These figures are unusual and require some considerable effort to comprehend. I assume the width versus height of the bubbles shows the stardard deviation in the meridional and vertical directions, and that the colors show the standard devation of the norm, but no information is given about that. I would suggest simply replacing these with contours and/or heat maps.

Reply: Agreed.

Planned Action: We will show variabilities in v and w in separate figures.

Review: All figures: I suggest using a perceptually uniform colorscale, which makes viewing much easier for those who do not have a standard perception of color.

Reply: We have tried many different colour scales but the alternatives did not seem convincing to us.

Planned Action: We will present the same plots in other color scales in the supplement. Beyond this, we will make the original data available, thus the readers can plot then in their preferred representation.

Review: Line 133: I am not sure that I can agree that the vertical motion over this range creates a transport barrier.

I would rather say it suggests one, at the most. But it could also be interpreted as a latitude (or a section of a latitude) where the circulation splits. In that case, there's not really a barrier to horizontal transport, but only relatively little forcing towards horizontal transport / a stronger forcing towards vertical transport. I'm not sure what the case truly is, but I dont think it's certain that this represents a barrier.

Reply: We will rephrase this sentence to avoid misunderstanding.

Planned Action: We will rephrase lines 133-134 to: "The direct vertical motion over 30°S suggests the existence of a region where horizontal transport is minimal compared to vertical transport; the location of this region is in good agreement with the location of the subtropical transport barrier (e.g. Stiller et al., 2017)."

Review: Line 138: "would likely look" - I think it's highly likely, even, but no definitive statement can be made until the analysis is done. On that topic, I think that analysis would be very interesting for future work.

Reply: Agreed.

Planned Action: We will reword this: "Representation in equivalent latitudes would be more adequate to analyze this phenomenon but since that representation would not be optimal for global analyses, it is deferred to a future study."

Review: Line 142: You might consider showing the tropopause and stratopause levels in your figures. I think that would be very helpful, and could alleviate a lot of confusion.

Reply: Monthly averaged tropopause altitudes can be very misleading.

Planned Action: none

Review: Line 150: It would be more useful to replace the

values shown in Figures 5 and 6 with the values discussed in this sentence. That would be a more direct indication for the reader of where the circulation is consistent. Otherwise, they need to compare these values themselves, which is rather tedious if a thorough comparison is desired.

Reply: If we understand correctly, the reviewer recommends to present the ratio between variability and absolute velocity instead of the variabilities themselves. (in the sense "inferred velocities exceed their variabilities by a factor of ..."). We tried this but due to the often small velocities, this representation is not easy to interpret either. The plot would be dominated by large but meaningless ratios related to tiny velocities while regions of interest with large absolute variabilities would no longer be obvious. That is why we didn't choose this representation.

Planned Action: none

Review: Line 189: Leaving aside the term "latitudinal barrier" (again, I am not sure how to distinguish between a barrier to latitudinal transport or a region of weak latitudinal transport), I do not see that agree that with the term "contribute to the formation".

Reply: We will rephrase this sentence.

Planned Action: Lines 189-190 will be rewritten as "This feature will evolve in the following months as a region where uplift motion clearly overtakes horizontal transport around 60°N."

Review: Line 234: I've mentioned this already, but I think this case shows clearly why the variability should be depicted in another way. It's too difficult to compare the standard deviations here to the circulation strength, for the most part. But your argument does seem plausible.

Reply: We agree that this is difficult.

Planned Action: We will quote the numbers in the text and and rephrase to make it clearer that the figure illustrates our argument rather than quantify it.

Review: Line 237: It seems like you wanted to specify a figure in Ploeger et al. 2017. I suppose you mean Figure 5? You should specify the figure.

Reply: Agreed. Indeed we mean Figure 5.

Planned Action: The Figure will be specified.

Review: Line 323: What is independent about the ANCISTRUS results?

Reply: ANCISTRUS results are independent from each other in the sense that the result of an ANCISTRUS run for one month is never used a first guess, a priori, or similar of an ACISTRUS run of another month. All ANCISTRUS runs can thus be performed independently, and any artificial autocorrelation of the results is thus excluded.

Planned Action: We will slightly reword this: " from the results of independent ANCISTRUS runs".

Review: Line 323: "resulting fields are stable" The statement regarding field stability should be more nuanced. Some parts of the fields do seem to be stable, sure, but this statement suggests that the fields are generally/always consistent.

Reply: Agreed.

Planned Action: This statement will be made more specific.

Review: Line 324: "increases confidence in the robustness of the analysis method" - I do not agree. If the method is robust, then a rather stable circulation field over the approx. decade of measurements in one region would suggest that the circulation field is a typical phenomenon for that

region, at least for that decade. However, the robustness of the inverese method cannot be assessed by seeing consistency in its results without a second point of reference (preferably, other observation-based circulation estimates, which - to my knowledge - do not exist).

Reply: Well, this depends what how the term ‘robustness’ is understood. We understand ‘robustness’ as the characteristic that the solution is not overly sensitive to varying input. We do not claim here to have shown that the method is accurate. That is also why our initial wording was “increases confidence in the robustness” instead of “shows the results are robust”.

Planned Action: We will clarify what we mean: “ The stability of results from independent ANCISTRUS runs increases confidence in the robustness of the analysis method in the sense that it produces similar results for similar input fields.”

Review: Line 330: It’s not clear to me if a clear separation between these two pathways would be expected or not. Could you provide any context on that, in terms of earlier literature?

Reply: Agreed

Planned Action: we insert: “..., as suggested by the schematic shown in Fig. 1 of Boenisch et al. (2011)...”; Reference: Atmos. Chem. Phys. 11(8), 3937-3948, doi10.5194/acp-11-3937-2011, 2011),

Review: Line 333: “consistent with the assumption” - Has anyone previously suggested this idea, or are you saying that your results would only be consistent with a northward pole-to-pole circulation if it was above the domain of MIPAS? If nobody has suggested this, then this statement should be written differently to clarify the novelty of the result.

Reply: Well, we think that the existence of the pole-to-pole overturning circulation is well established. We see velocities going up in the north and downward velocities in the south, but the meridional velocities which would close this circulation are above our data domain. We neither claim to have found a novel circulation path nor do we refer to a specific assumption written in the literature. Thus our careful wording.

Planned Action: We will rewrite this without the term ‘assumption’: “Our data are consistent with - but do not directly support - a southward pole to pole circulation from March to May at altitudes not covered by MIPAS data”

Review: Line 337: To my understanding, this intrepretation of the tropical pipe would be novel. I only wish to note that here.

Reply: Yes, indeed. We have intentionally chosen a very careful wording here (‘suggests’...‘may not be’...). We are willing to make the wording even more careful.

Planned Action: We will change the wording to “This seems to suggest that...not always...

Review: Line 341: This could be consistent with some earlier results. See Butchart 2014 (The Brewer-Dobson Circulation, Rev. Geophys.) Figure 6 and discussion of that figure. If mean downwelling during winters where the polar vortex is not disturbed is the same between both hemispheres, then one would expect stronger climatological descent in the southern hemisphere because the vortex is disturbed less often there.

Reply: We understand our (old) lines 340-341 as an introductory sentence for the five following more specific points. Thus the quite specific comment seems to refer to line 342 (#1 in the list) rather than to line 341. We will add a sentence there.

Planned Action: We shall add to #1: “This is consistent with stronger southern than northern polar winter subsidence which is associated with less perturbed polar vortices there (Butchart 2014, Section 5.1)." Reference: Rev. Geophys. 52(2), doi10.1002/2013RG000448, 157-184 (2014)

Review: Line 343: To my knowledge, this result is not expected.

Reply: Ok, we will mention this.

Planned Action: We shall add: "To the best of our knowledge this has not been reported before either."

Review: Line 347: I've mentioned it already, but I do not think the term "barriers" is justified in this context.

Reply: Agreed

Planned Action: We will replace the term 'barriers' by 'areas with near zero ... velocities.'

Review: Line 348: Same to point 2.

Reply: ok, we will mention this.

Planned Action: We shall add: "To our best knowledge this also this has not been reported before."

Review: Line 364: In the broad stroke, I agree with this statement. However, the absence of a southward mesospheric overturning circulation means that this statement cannot be written in the absolute. Furthermore, the results do not characterize these patterns in an expected fashion (tropical pipe, for example). This statement should be rewritten to reflect those points.

Reply: The overturning circulation is not absent but just not covered by the MIPAS measurements. But we agree to reword our statement.

Planned Action: We will write: "The ANCISTRUS method applied to MIPAS data broadly reproduces well the known atmospheric meridional circulation patterns, although with some unexpected features. Additional information ..."

Review: Technical Comments

Review: Abstract, line 1: HCFC-22.

Reply: Thanks for spotting

Planned Action: This will be corrected.

Review: Line 14: The citation of Brewer and Dobson 1949 is incorrect. That's a single-author publication, just from Brewer.

Reply: Thanks for spotting

Planned Action: This will be corrected.

Review: Line 14: I think it's better to write abbreviations in a separate set of parentheses, just so it's clear that the abbreviation isn't some part of a citation.

Reply: This comment has become obsolete after rewriting in reply to a specific comment (see above).

Planned Action: No additional action.

Review: Line 16: The last part of this sentence seems to suggest that only aerosols affect major chemistry-climate processes. It would be more clear with "as aerosols, all of which affect major chemistry".

Reply: Agreed.

Planned Action: This will be corrected

Review: Line 19: This sentence is a bit of a run-on. It would be better to write one sentence for Engel's balloon studies and another for the satellite studies.

Reply: We agree that the original sentence is too long. We prefer, however, to split the sentence immediately after the Butchart reference.

Planned Action: Will be changed to "...Butchart et al. (2006). This triggered..."; Reference: Clim. Dyn. 27(7-8), 727-741, doi10.1007/s00382-006-0162-4 (2006).

Review: Line 22: "Offline model simulations ... analysis data have also confirmed ...", or add a comma after "Also".

Reply: Thanks!

Planned Action: A comma will be inserted.

Review: Line 28: It looks like you meant to write "has been investigated" or something similar. At the moment, the sentence doesn't make sense.

Reply: Agreed.

Planned Action: This will be corrected.

Review: Line 30: If Funke et al. (from all those years) showed this, then I would write "has been" not "could be".

Reply: Agreed.

Planned Action: This will be corrected.

Review: Line 35: It would be more precise to simply say that the picture of the middle atmospheric circulation is better resolved in space and time, cutting out the "more detailed" part.

Reply: Agreed.

Planned Action: This will be corrected.

Review: Line 38: The ending, "over the years", isn't necessary here.

Reply: We had added 'over the years' to make clear that we do not talk about the inter-annual, not the intra-annual (month to month) variability. We will try to reword this in a clearer clumsy way.

Planned Action: 'over the years' will be deleted, and 'inter-annual' will be inserted before 'variability'.

Review: Line 46: I would say not just monthly but "monthly-mean". Furthermore, the sentence suggests that this is the only way to infer the circulation, so you should specify "is inferred in this work".

Reply: Agreed for "in this work". The original text already reads "monthly zonal mean" and we think that it is clear that averaging was made in both domains.

Planned Action: "in this work" will be added.

Review: Line 49: "The resulting circulation fields..."

Reply: Thanks!

Planned Action: This will be corrected.

Review: Line 62: I don't understand what "related software" refers to. That's the inversion method, right?

Reply: Yes, it is.

Planned Action: This will be reworded as suggested.

Review: Line 79: "source reactions were also considered"

Reply: Thanks for spotting!

Planned Action: This will be corrected.

Review: Line 83: “the neglect of sinks above that altitude”

Reply: Thanks!

Planned Action: This will be corrected.

Review: Line 91: See comment on line 14 about abbreviations.

Reply: As far as we know, our way to set the parentheses here is the one which complies with the COPERNICUS rules. I think here the copy editors have the last word.

Planned Action: We'll wait what the copy editors say.

Review: Line 103: “From MIPAS, measurements”

Reply: MIPAS here is a sort of attribute or specifications. With the comma inserted, the meaning would change towards “measurements are calculated from MIPAS”, which is not what we mean.

Planned Action: None

Review: Line 104: You explained the data gaps in the last section.

Reply: Yes, indeed.

Planned Action: we will include the 2006 data gap in line 95 and reword here: “... with data gaps as reported above .”

Review: Line 116: “up to 30 km”

Reply: Agreed.

Planned Action: This will be corrected.

Review: Line 123: move "also" from "also the standard..." to "are also shown"(...). Furthermore, it should be clear to all readers that standard deviations are a measure of variability, so the "which are a measure of their variability" is not necessary.

Reply: Thanks!

Planned Action: This will be corrected.

Review: (...). Furthermore, it should be clear to all readers that standard deviations are a measure of variability, so the "which are a measure of their variability" is not necessary.

Reply: A standard deviation is a measure of the width of a distribution but it is not clear if it represents variability, uncertainty, probability, or other. In particular in our community, uncertainties and estimated errors are often reported in terms of standard deviation. Thus, we find it necessary to be specific here.

Planned Action: None.

Review: Line 127: You can just say northern hemisphere or winter hemisphere.

Reply: Agreed.

Planned Action: "local winter" will be deleted.

Review: Line 146: "has its maximum ... and at 30°S"

Reply: Agreed.

Planned Action: This will be reworded as suggested.

Review: Line 155: You are clearly comparing this month-

pair with the previous, but this sentence should make that explicit.

Reply: Agreed.

Planned Action: "seen in January-February" will be inserted.

Review: Line 167: "will give rise"

Reply: Thanks!

Planned Action: This will be corrected.

Review: Line 170: The abbreviation SH was already used before this point. The notification of the abbrevation should be shifted to the first usage of "southern hemisphere". Ditto for NH.

Reply: Thanks for spotting.

Planned Action: This will be corrected.

Review: Line 379: "their figure" - This part of the sentence addresses a particular figure, but the earlier part speaks generally of schematics. I suggest sticking to one approach or the other.

Reply: Agreed.

Planned Action: We will split the sentence to make clear what refers to such schematics in general and what refers to this particular figure.

Review #2:

Review: The authors present an estimate of meridional circulation patterns in the middle atmosphere based on measurements from 2002 though 2012 by the MIPAS instrument of a range of trace gas species. The estimate is based on an inverse method that infers an effective flow field in the meridional plane from the continuity equation along with an estimate of chemical sources and sinks. The methodology is updated from previous work by the first two authors through inclusion of further chemical sources and sinks, and by inferring only an 'effective' meridional flow that includes the effects of mixing/eddy transport.
The main results shown are the month by month estimates of the decadal-averaged meridional flow, as well as an estimate of the interannual variability of the flow. In as much as this estimate is a relatively direct observational estimate of a difficult to measure quantity, this result is of potential value to the broader community.

Reply: We thank the reviewer for this encouraging evaluation.

Planned Action: None to this point.

Review: My main concern is that if this is to be the case, enough quantitative details should be given in order to facilitate comparisons with these results; this is largely the case but there are a few ambiguities that should be addressed (see below).

Reply: Agreed.

Planned Action: See specific actions below.

Review: Beyond this I have a few questions and comments about the presentation of these results (in particular I find the presentation of the interannual variability difficult to understand), but otherwise feel this is appropriate for publication with some minor revisions.

Reply: We agree that the presentation of the inter-annual variability should be better explained.

Planned Action: See specific actions below.

Review: Specific comments
1) As mentioned above there are a few points that would be helpful for making quantitative comparisons with these results. Firstly, does the inferred circulation conserve mass?

Reply: Yes, it largely does, except for transport into and out of the model domain at the upper and lower boundary, which may not necessarily be balanced. But besides the continuity equation of species, also the continuity equation of air density is one of the determining equations of our system.

Planned Action: None.

Review: If so the authors may want to consider showing a mass stream function instead of the vector plots.

Reply: Any representation which involves weighting by air density suffers from the large dynamical range of air density with altitude. We have tried various representations of this type but always all structure and information in the middle atmosphere was lost. Only values at the lowermost layer were discernable.

Planned Action: none

Review: If not, the choice of units for Figs. 1 through 4 are a bit confusing; surely the velocities should be homogeneous in units (e.g. m/s)?

Reply: With homogeneous units the vertical velocities would be invisible, although important. The norm we use for the colour scale and for the direction of the arrows roughly corresponds to the aspect ratio of the plots where the vertical axis does not represent the true geometric conditions either but is heavily exaggerated.

Planned Action: The original data will be made available in digital form. With this the user can represent the data in the preferred way, most suitable for the respective application.

Review: The note regarding the colour scales in the 2) Figures 5 and 6 show standard deviations of the inferred effective velocities, but the visualization is not explained.

Reply: The explanation is indeed missing in the original submission.

Planned Action: An explanation of the visualization will be included.

Review: One assumes that the axes of the ellipses are scaled relative to the variance of the y and z components of the velocities but it seems no account is being taken of their covariance if that is the case. How are the colors chosen?

Reply: The axes of the ellipses represent the standard deviations in the the y and z components. The covariances are not represented. The colours are chosen by adjusting the colour scale to the maximum value of the individual plot.

Planned Action: We have decided to represent the variabilities in separate plots for v and w. We will use a common colour scale for the entire series of plots. Further we will mention that the same norm is applied to the standard deviations as for the effective velocities.

Review: More importantly, are these estimates of the standard deviation of the mean (implied by the figure caption) or sample standard deviations?

Reply: The title was indeed misleading. We present the sample standard deviations, not the standard error of the mean. This is because we are interested in the variability, not in the uncertainty. The standard error of the mean would become lower for a larger sample and is thus not the adequate measure.

Planned Action: Titles in the plots will be corrected.

Review: In sum the interannual variability is difficult to assess in comparison to the mean circulation from these figures and is not very satisfyingly discussed in the text.

Reply: Agreed

Planned Action: New plots will be provided and adequately described.

Review 3) In all figures, different years are included in each panel with no discussion; why?

Reply: First we have some data gaps in the MIPAS data, and second, a (small) number of the inversions did not converge.

Planned Action: This information will be provided.

Review: 4) The methodology used in the present work includes the role of chemical sources and sinks; this has been updated from von Clarmann and Grabowski (2016). The value of these updates should be demonstrated.

Reply: The impact is indeed substantial, and we are happy to show respective plots. However, we think that this fits in much better with the paper on validation and sensitivity studies discussed in reply to reviewer #1.

Planned Action: This issue will be deferred to the validation paper of which the draft will be made available to the reviewers and reference to that paper together with a brief explanation will be included to the current manuscript.

Review: It would also be useful to make some assessment of the role of mixing, again in order to facilitate quantitative comparisons with the mean meridional flow from models,

for instance.

Reply: We agree that this is interesting, and we have actually submitted a research proposal which will tackle exactly this question. We think that this is a research topic in its own right and defer this to a future paper.

Planned Action: none for this paper.

Review #3 (This reply refers to the uploaded comment, not to the comment sent to the editor. There are differences between these):

Review: In this paper, the authors use measurements of a variety of trace gases from MIPAS to infer the stratospheric and mesospheric circulation. They calculate a climatology and determine that the deep branch of the Brewer Dobson circulation is connected to the mesospheric pole-to-pole circulation. They verify a number of known characteristics of the circulation, such as sudden stratospheric warmings increasing variability.

Reply: We are happy that the reviewer confirms that the presented climatology verifies a number of characteristics of the BDC.

Planned Action: none to this point.

Review: Using chemical tracers to infer the circulation is an excellent idea. Tracers are what we can measure from space, so to validate any model, we need to quantitatively relate the tracers to the dynamical output from climate models.

Reply: We agree that tracer measurements are essential for an empirical assessment of circulation. However, validating modelled tracer distributions is not enough, particularly if there are discrepancies between the model prediction and the observed atmospheric state. We are primarily studying the atmosphere but not (climate) models that try to reproduce the atmospheric processes correctly. Adjusting the models to the observed processes is a second, nevertheless necessary step that is, however, not our primary concern in this paper.

Planned Action: none to this point.

Review: The inverse methods used here are promising. Unfortunately, the approach from the authors is lacking in a number of ways. 1) The validity of the method has not

been established.

Reply: Our method is clearly based on the established validity of the continuity equation. And as the Reviewer acknowledges here above the results we obtain successfully reproduce a number of BDC characteristics. In addition a validation paper is ready for submission.

Planned Action: none to this point. However, a validation of the algorithm will be published in a separate paper; see reply to reviewer #1.

Review: 2) Uncertainties are not calculated and [...],

Reply: Reviewer #1 has raised a similar concern before. We consider the estimated uncertainties as largely redundant. We present the variabilities of the results. These include both the precision of the results and the natural variability. The variance describing the precision of the results cannot be larger than the variance describing the variability of the results. Thus, our presented variabilities are an upper estimate of the precision of our method.

Planned Action: See our related reply to Reviewer #1.

Review: perhaps most importantly, 3) The utility of the resulting product from the method is unclear.

Reply: The utility is that we provide an empirical diagnostic of the stratospheric circulation which does not suffer from the main drawbacks of either the direct comparison of modeled versus observed trace gas fields or the age-of-air based methods. While the former is very unspecific with respect to causes of discrepancies, the latter's drawbacks are the dependence of assumed age of air spectra and artificial overaging to unaccounted mesospheric sinks of tracers. Our results contain considerably more information than the trace gas fields and their variation with time, and it provides a better time-resolved understanding of the circulation than the age-of-air method (which integrates over the time an air parcel spent in the stratosphere).

Planned Action: We will include a couple of sentences specifically stating the utility of our product and the capability of our method.

Review: The authors have not done any test that would demonstrate that this inversion does actually recover velocity fields in a model.

Reply: Any comparison with model results would suffer from the fact that in the case of discrepancies it is not clear if they are caused by a failure of the new method or the inadequacy of the model. Furthermore, the interpretation of 2D fields inferred from 3D model output depends on certain approximations (see Appendix of von Clarmann and Grabowski, Atmos. Chem. Phys. 16(22), 14563-14584, doi10.5194/acp-16-14563-2016", 2016). Nevertheless, we have tested how far velocity fields can be recovered by our inversion.

Planned Action: These tests are included in a separate validation paper whose draft will be made available to the reviewers.

Review: The closest is a very idealized case in their 2016 paper and even in that case, idealized tracers are used instead of real tracers.

Reply: The idealization was made on purpose. Only in these simple cases it is possible to predict (without the help of another model) what the result should be, and to check if, e.g., the transport scheme does what it is supposed to do. In a simple test, with a constant velocity field of x degrees per month it is straight forward to check if a structure has actually moved by x degrees in a month; how the shape of the structure is conserved; what kinds of wiggles are created. With any close-to-realistic fields it is virtually impossible to judge if any wiggles are caused by diffusion or dispersion or are real phenomena. Also the over-exaggerated structures in the idealized test are, due to the large gradients in the fields, a particular challenge for the transport scheme, and can be considered almost as a worst case study. All methods used (the McCormack transport scheme, matrix inversion, etc) are well established methods.

Planned Action: The validation paper with further validation test cases will be made available to the reviewers.

Review: In order to use the method on data, essentially an entire separate modeling study needs to be performed: a) using a CCM, with full knowledge of the tracer fields used here, the inversion needs to be performed and compared to the model velocity and stream function.

Reply: We agree that such a modelling study is interesting, and it is actually under way. We object, however, that such a modelling study is the only possible approach to corroborate the validity of our scheme.

Planned Action: Model recovery tests are included in the validation paper mentioned above.

Review: If this is successful, then the next step is: b) with the same CCM, the sampling characteristics of MIPAS (coverage and averaging kernels) need to be applied to the tracer fields so that now the limitations due to sampling and retrieval characteristics are applied. This seems especially important for vertical and horizontal resolution.

Reply: MIPAS provides dense sampling. Of course the sampling varies slightly from year to year. Still the year-to-year variability of the inferred circulation is quite small. If sampling was an issue, how can it be then explained, that in different years so similar circulation patterns are obtained?
By the way: Different standards seem to be applied to different methods. MIPAS sampling is dense, and we have representative zonal means. Other observational studies use single snapshots of the atmosphere obtained by balloon observations (e.g. Engel et al., Engel et al., 2009, 10.1038/ngeo388; cited approvingly by the reviewer) to infer the strength of the Brewer-Dobson circulation. What is the purpose of applying such a different standard to different methods? The issue of vertical resolution and related implications have already been discussed and solved in von Clarmann and Grabowski, Atmos.

Chem. Phys., 16, 14563-14584, Section 3.5.
Again, why is our work judged by a different standard than other work? We do not know any age-of-air related work where vertical resolution has been considered. In the work of the reviewer, quoted in her review, vertical resolution is not even mentioned. We expect our work to be judged by the same standards as previously published work on this field. This preaching-water-and-drinking-wine stance does not fit into a neutral review!

Planned Action: none to this point.

Review: Then the inversion needs to be done and compared once again to the model velocity and stream function output. This test will illuminate what the method actually means.

Reply: What the method actually means is quite clear: it provides the most possible direct observational access to the temporally and spatially resolved effective 2D circulation. The appendix of the paper explains how the same quantities shown in the paper can be derived from 3D model fields. This allows a direct comparison when a model validation is the topic of further work.

Planned Action: see the discussion of the model recovery tests and the reply to reviewer #1.

Review: The errors caused by the method with full tracer fields and then with the limited sampling can then be characterized for the model as well, hopefully beginning to address point 2) above.

Reply: The sampling in one month over the years does vary. Still we get small standard deviations. This furnishes evidence that the method is not sensitive to MIPAS sampling issues.

Planned Action: We will include this argument in the paper.

Review: This would also work towards addressing point 3) above. This "transport circulation" that the authors

obtain is not obviously relevant. Without being able to meaningfully compare values to model output or reanalysis products, this quantity does not seem to be of interest,[...]

Reply: Our study does provide a new dataset based on observations to characterise atmospheric circulation processes, which will also additionally serve for model and reanalysis comparisons. The appendix of the paper explains how the same quantities as presented from observational data can be produced from model data or reanalyses. This allows a far more detailed comparison/model validation than age of air or the quantity "strength of the BDC". Since the BDC was posited to explain the effective transport of trace gases from low to high latitudes, our effective velocities capture the essence of the BDC, and every move towards quantities represented in 3D models would move away from this nature of the BDC. Actually the models fall short to predicting temporally and spatially resolved measures of the circulation which can be directly validated by observations. The stance that only observations which are related to model output are relevant is not tenable. Since Ian Hacking (1983, Representing and Intervening, Cambridge University Press) it is established that the task of observations goes beyond the mere verification of model predictions and that empirical science is a science in its own right.

Planned Action: none

Review: [...] and so the authors claims of being able to assess quantitatively the variability of the circulation fall flat.

Reply: In this paper we have focussed on the structures and the seasonal variations of these structures of the BDC. We would like to point out that between purely qualitative work and fully quantitative work, there is the wide field of work on structures (often ignored. Too many people misconceive qualitative vs. quantitative as a dichotomy!). Further, the conclusions of our paper do not depend on the absolute accuracies of the inferred effective velocities.

Planned Action: none

Review: In fact, this is the reason that age of air is such

a useful tracer – it has been quantitatively related to the circulation of the stratosphere in a way that allows direct comparison of data (including the MIPAS data) to models (Linz et al. 2017).

Reply: The age of air cannot be directly observed, it must be inferred from tracer measurements. This inference of the age of air from tracer measurements is based on assumptions, some of which we challenge. The method we present in this paper does not make these assumptions. We do not state that age-of-air based methods are not useful or should not be used by models, but we have used age-of-air based methods long enough to know their weaknesses and to find legitimate to search for alternative methods which avoid some known weaknesses. And as said above, the quantities derived from observations in this paper can all be calculated from models and reanalyses as well, as described in the Appendix.

Planned Action: none

Review: I would strongly encourage the authors to perform such a study and then to rethink their results for this work in the context of the information provided by their validation study. That would be an excellent paper that I would be truly excited to see.

Reply: A validation study (model recovery test) will be presented but we do not consider a model-based validation study as the optimal approach.

Planned Action: see reply to reviewer #1.

Review: Beyond this overall assessment, I have included more detailed comments below: 24: What about the lifting of the circulation? (e.g. Oberlander-Hayn 2016)

Reply: In our paper, we deal with the climatology of transport vectors and their year-to-year variability. Long-term trends as tackled in the Oberländer-Hayn et al. paper are beyond the scope of our paper.

Planned Action: Nevertheless, we will mention the paper in the introduction.

Review: Overall introduction: What is the gap that this research is filling? The introduction reviews the literature but does not identify any motivation for the present study.

Reply: It does. The motivation is clearly stated in the sentence criticised before: "In this study we aim to provide a picture of the meridional middle atmospheric circulation (better resolved in space and time than that provided by age-of-air based methods.)"

Planned Action: Include "(better resolved in space and time than that provided by age-of-air based methods.)"

Review: 2.1 This discussion of age of air is surprising. What is this so called "traditional observation-based characterization of the circulation"? The authors do not provide a citation.

Reply: The review paper by Waugh and Hall (Rev. Geophys. 40(4), doi10.1029/2000RG000101, 2002) gives an excellent introduction.

Planned Action: We will include this reference.

Review: Some recent work that uses age of air observations to characterize the circulation is Linz et al. 2017. Recent work by Ray et al 2016 combines the TLP with chemistry to examine transport, and the improvement this offers over that method should be addressed. Ray et al. 2010 is also a relevant comparison here.

Reply: Linz et al (10.1038/NGEO30132017, 2017) have characterized the BDC by a single profile. Ray et al. (10.1029/2010JD014206, 2010 and 10.1002/2015JD024447, 2016) introduce the leaky pipe model to explain age of air, ozone, CFCs, and their trends. However, since we do not deal with age of air in this paper, we do not see how these papers are related to our work. The reviewer seems to assume that we are not familiar with the current literature and the approaches used so far. We wish to state that the contrary is true, and our previous work with age of air has made us aware of the weaknesses of this approach and the need to find an improved observation-based access to the BDC.

Planned Action: none

Review: Furthermore, it is not clear how the method can reveal causes of "discrepancies" between these age and chemical tracer based methods since there are no error estimates.

Reply: The standard deviations representing the precision are by definition smaller than the standard deviations that we show. Thus the standard deviations can serve as upper estimates of the random errors of each monthly field. The uncertainty of the average over the years is accordingly smaller. The rationale behind our claim that our quantities are better suited to determine causes of discrepancies than age of air is fairly trivial. If, say, in the polar stratosphere there is a discrepancy, we still do not know when (since the air entered the stratosphere) or where (along the trajectory of the air parcel) the discrepancy was caused. Our method provides quantities that are resolved latitudinally, vertically and temporally. Thus a much clearer idea can be developed where the model world and the observational world begin to diverge. This is exactly where our method provides an advantage over the age-of-air based methods.

Planned Action: see reply to reviewer #1

Review: 2.1.2 There is no discussion of degrees of freedom. How much independent information is gained by including additional tracers?

Reply: Tests have been shown that inclusion of further species predominantly reduce the error estimates. This effect is seen even in cases where the resulting circulation does not change. For more details, see response to Reviewer #1.

Planned Action: see validation paper.

Review: Specifically, how are sinks included?

Reply: This is described in Section 2.1.2.

Planned Action: Some clarifying amendments as requested by reviewer #1 will be included in the description of the sinks.

Review: 3.1: Plots are very hard to understand. Streamfunctions would be much better.

Reply: We concede that the changing colour scales between the panels of a figure were not optimal. To better serve the needs of the data users, we will make the data available. Then every user can represent the data in their favourite way. Our vector representation offers the advantage that it can directly be compared to the often reproduced schematic by Boenisch et al.

Planned Action: Colour scales will be homogenized.

Review: 3.1.1 How are horizontal transport barriers identified? Why, physically, are they associated with this vertical motion? Is this purely a result of continuity and the fact that this is a 2-D calculation? If so, this should not be referred to as a barrier.

Reply: We have identified horizontal transport barriers as consecutive latitude/altitude bins where the meridional transport velocity is zero, while meridional transport vectors point in opposite directions in the two meridionally adjacent bins. We agree with Reviewer #1 that this might indeed just identify a splitting or bifurcation and that our wording needs editing.

Planned Action: We will rephrase lines 133-134 as specified in reply to reviewer #1.

Review: 3.1.6 How precisely do you identify that this circu-

lation is associated with the monsoon? Are there particular tracers (e.g. water vapor) that mark this as a monsoon signal? Or is it just about the timing and the fact that it's in the Northern Hemisphere, in which case the link is suggested at best. "Our results show overall agreement with the one shown by Ploeger et al. (2017)[...].";

Reply: Indeed it is the timing, the altitude range and the fact that it appears in the NH only, that we link this to the monsoon.

Planned Action: We will edit the main text to make this clearer.

Review: What "one"?

Reply: Their Fig. 5

Planned Action: see reply to reviewer #1.

Review: What is meant by "overall agreement"?

Reply: We mean agreement related to the structures.

Planned Action: We will write "structural agreement."

Review: 3.2.1 337: What is meant by this? How does this reconcile with the well-established water vapor tape recorder results?

Reply: We concede that our wording may lead to misunderstanding.

Planned Action: We will reword this statement.

Review: 368-374: This seems to be saying that this study is a good validation of the method. That may be, but it's not the stated goal, and more stringent validation is needed especially so as to be able to actually interpret the resulting "effective velocities".

Reply: We think that these indicators of robustness are an important piece of information. They are by no means obsolete, even with a model recovery test in place. The validation will be published in a separate paper.

Planned Action: Model recovery tests will be performed and published in a separate paper which will be made available to the reviewers.

Review: 384: What applications would use these effective velocities?

Reply: The Brewer-Dobson circulation explains large ozone amounts over the poles while ozone is predominantly generated at low latitudes. In other words, it uses the effective 2D transport as an explanation of the trace gas distributions in the stratosphere. Thus, the effective 2D transport velocities are the natural measure of the BDC because they directly capture the essence of the BDC.
Further, these effective velocities can be understood as inverse age increments per segment of a mean trajectory. They can thus be related to the age of air but are time-resolved. The effective velocities are an empirical diagnostic in their own right. And in the appendix we relate them to the usual model quantities. Review #2 provides evidence that a significant part of the community regards this quantity as useful.
Regardless if models are able to produce such quantities or not, we think that we can learn a lot from them: Trends, variabilities etc. In this first scientific application paper we have restricted ourselves to climatologies, because these depend less on quantitative validation. The observed transport patterns and their variation contain a wealth of information in a structural sense even if one is sceptical about the associated numbers.

Planned Action: Some sentences on possible future work will be included.