Review #1:

Comment: General comments:
This study presents a new methodology (ANCISTRUS) that provides quantitative information on stratospheric circulation in the form of effective 2D velocity fields obtained from measurements of a set of long-lived trace gases.
The paper presents clear examples of the computed fields and the transport structures they represent. It also provides a valuable illustration of the relative weights that chemical sinks and advection have on the distribution of tracers in the stratosphere. The manuscript also looks into the reliability and sensitivity of the method and shows what regions are better covered by the different chemical species.
The manuscript demonstrates the high potential the new method has to derive effective transport information for the stratospheric region that can complement other existing methods. This will help to overcome information gaps and biases that appear when applying more widely-used existing approaches.
Therefore, this study is a valuable contribution to both the modelling and the observations communities.
In its current form there are, however, several points that need further clarification or development, and editing is also required as detailed here below. Once these aspects have been satisfactorily addressed by the authors I recommend publication of the edited manuscript.

Reply: The authors thank the reviewer for their encouraging and thorough review.

Comment: Specific comments:
Some relevant scientific questions should be addressed by the manuscript:
To what extent is the ANCISTRUS-derived dataset limited by the biases in the original MIPAS measurements? How are those biases affecting the fields you derive?

Reply: This is an interesting question. Our main reply is that at first order the ANCISTRUS scheme is insensitive against such biases. The reason is this. ANCISTRUS exploits mainly the gradient information but only to a limited extent the trace gas absolute concentrations. The calculation of gradients involves differences, and biases cancel out when these differences are calculated. Surviving higher order effects are small. We learn this from the fact that the results do not change in any substantial way when a certain species is discarded from the analysis. If a possible bias had a large effect, the neglect of the related species should substantially change the results, which we do not observe.

Comment: I miss a discussion on the generalisation of the methodology: You have applied the ANCISTRUS method to MIPAS data,

but how feasible would it be to apply to other satellite products of atmospheric tracers measurements? Is there any current work on this? How would it be done? This is an important discussion to show the value of the methodology.

Reply: Indeed we plan to apply this method to other satellite data sets. However, the number of suitable data sets is quite limited. Global altitude resolved distributions of a sufficient number of species are needed. Candidate data sets under consideration are those of MLS and of ACE-FTS. The challenge with MLS is that the number of suitable species is quite limited (H_2O, CO, methyl chloride and N_2O are the most promising ones). A research proposal on this application is actually under evaluation. ACE-FTS offers data of a large number of suitable species. The challenge is the less than optimal global coverage that can be obtains with an instrument measuring in solar occultation.

Comment: Is the study of the BDC (line 31) the main application of your method or the main application in this study? Are there other applications?

Reply: Indeed the study of the BDC is the main application of the method. Other applications are thinkable but have not yet been studied.

Comment: Sentence 39-41, if I understand correctly, could indicate a limitation of the methodology rather than stability: can it be that for every year the obtained circulation patterns are the same? What about interannual variability, how does your method account for this?

Reply: We do see interannual variability, and we see it exactly where we expect it. But the general patterns are reproduced for each year. This issue is discussed in more depth in von Clarmann et al. (2019).

Comment: More information should be included on how the method copes with, and provides information on, the diffusive and dispersive characteristics of transport.

Reply: The 2D-velocities we present are effective velocities which include the effects of atmospheric diffusion and eddy mixing. How these effective 2D-velocities are related to 3D-velocities and mixing is discussed in the appendices of von Clarmann and Grabowski (2016) and von Clarmann et al. (2019). Dispersion we understand is not a characteristic of atmospheric transport but an effect of a numerical transport scheme. This has been analyzed in von Clarmann and Grabowski (2016).

Comment: Why have you chosen years 2005 and 2010 for this study? The manuscript should justify this choice over other months/years in the period covered by MIPAS observations. How representative are

these 2005 and 2010 months for the rest of the period?

Reply: As can be seen in von Clarmann et al. (2019), these years are representative for the respective months in other years. The results for a certain month are structurally very similar over the years but vary in a quantitative sense. ANCISTRUS does see the typical expected seasonal patterns. Sometimes a certain pattern may occur a little sooner or later, and the strength of a pattern may vary from year to year (we actually see, e.g., a QBO effect, which is the topic of a separate study). The choice of the test cases was influenced by technical considerations. For example, to do the Jackknife test, it is necessary to start with a month for which data of all gases were available, which was not always the case. Furthermore the Sep-Oct 2010 has a simpler structure and higher velocities while Mar-Apr 2005 has lots of detail structures. Thus we consider these examples as adequate test cases.

Comment: How much does the lack of longitudinal information affect the degree of realistic variability in your results?

Reply: The BDC is a 2D phenomenon. ANCISTRUS in its current implementation does not provide longitudinal information, but we think that this is adequate for studies of the BDC. The problem is rather that, due to correlations of 3D velocities with concentrations, our effective 2D-velocities must not be conceived as zonally averaged 3D meridional velocities. This issue is discussed thoroughly in von Clarmann and Grabowski (2016) and von Clarmann et al. (2019).

Comment: If your method does not consider SF_6 sinks (line 64) how can you overcome the biases caused by the mesospheric sink of SF6 when using mean age-of-air methods? This needs to be clearly explained and justified as the manuscript claims this is one of the main advantages of the ANCISTRUS method in the study of the BDC.

Reply: Here we have two arguments:

1. The sinks of SF_6 are most relevant above the altitude range we work with. Age-of-air based methods are sensitive to the accumulated loss of SF_6 during the air parcel's journey from the stratospheric entry point via the mesosphere and back to the stratosphere. In contrast, ANCISTRUS, for species without a sizeable stratospheric sink like SF_6, uses only the gradient information at two points within the stratosphere. Stratospheric sinks of SF_6 are negligibly small. This holds *a fortiori* for time scales as short as one month. The gradient information we use is calculated from measured concentrations. Thus, if such an air parcel contains air depleted in SF_6, this is implicitly considered in the gradient calculation. And having the gradients right, the tendency formulation of the continuity equation for SF_6 does not depend on sinks outside the domain to which

it is applied.

2. Beyond this, the influence of SF_6 on the results is surprisingly small. This is because the measurement errors of SF_6 are much larger than those of the other species we use. In the inversion more weight is given to the species where the standard error of the zonal mean is smaller. Thus, an effect of SF_6 sinks would not distort our results, even if there was one. Contrary to that, age-of-air based methods depend on species without stratospheric sinks and with a close to linear tropospheric trend, and there is not much choice. SF_6 is the only species which fulfills this requirement and for which global altitude resolved data are available.

Comment: Line 105: It is not clear what you mean by "true velocity fields are not known". Even if true velocity fields are not measured, can you use operational analyses from NWP models as a close to reality alternative? What are the assumed velocity fields you used in the tests discussed in this paragraph? How would the use of different assumed velocity fields affect your results and the spurious data you obtained at the boundaries?

Reply: We use velocity fields obtained with ANCISTRUS. For these it is guaranteed that they satisfy the continuity equation and thus belong to the set of possible solutions of ANCISTRUS. The spurious data we obtained at the boundaries in von Clarmann and Grabowski (2016) were a result of the *ad hoc* choice of a reference velocity field which did not satisfy the continuity equation. Since the continuity equation is hard-wired in ANCISTRUS, it cannot be expected that these fields are reproduced. Analysis fields cannot be directly used because our 2D effective velocity fields represent not only velocities but also atmospheric mixing as well as eddy mixing effects caused by the correlations of velocities and mixing ratios. Furthermore, realistic velocity fields alone do not help because we need also mixing ratio distributions of the relevant species which are consistent with the velocity fields.

Comment: Lines 120-130: Why a climate model? Why not a CTM driven by operational NWP analyses, then one has the tracers distributions obtained from the CTM and the operational velocity fields used to force the simulations?

Reply: Our arguments apply equally to a CTM. We will change the text accordingly. The problem is not how realistic the fields are, as long as they are compliant with the continuity equation, but the problem that our effective 2D-velocities must not be conceived as zonal mean 3D velocities, because they include atmospheric mixing and eddy mixing effects caused by the correlations of velocities and mixing ratios. While such model comparisons are interesting in their own right, the difficulty mentioned makes them less suited in a validation context. Not all available model runs provide distributions of all gases required, and those which do provide them often disagree with the MIPAS measurements.

In this case it is unclear which uncertainty is to be assigned to the mixing ratios. This easily tips the balance between the observation error covariance matrix and the regularization. Thus, any occurring differences cannot easily be interpreted in terms of validation. Nevertheless, we have the application of ANCISTRUS to model fields on our agenda for future work.

Comment: Line 152: Is there any alternative regularization that can better resolve adjacent opposite circulation branches? Have you tested it?

Reply: We use a Tikhonov-type regularization which reduces the differences between adjacent vectors of effective velocity. The advantage of this regularization is that the solution can be conceived as a smoothed but otherwise unbiased representation of the truth. The obvious alternative would be a stochastic regularization in the sense of 'optimal estimation'. This is, however, not neutral because the result is biased towards the assumed a priori information. Thus we stick with the Tikhonov-type regularization. After this decision we still have the choice of the strength of the regularization. Strong regularization will degrade the spatial resolution of the result (and thus lead to smaller absolute velocity values at the peak areas) while weaker regularization can give rise to numerical instabilities showing up as unphysical oscillations and leads more often to non-convergence. The optimal choice is application-dependent. For a single case study weaker regularization may be more appropriate because it reveals more details, while for a multi-annual climatological study (currently the focus of our interest) frequent cases of non-convergence shall be avoided.

Comment: Some parts of the manuscript need substantial editing: The Introduction Section needs to be rewritten: The scope and context for this research is not clearly introduced. Context should be added to the Introduction. Have other similar applications of inverse modelling been previously attempted? What are the reasons to develop this new approach? What are the advantages of the current one compared to previous ones?

Reply: Agreed; we will expand on this.

Comment: Also, some important aspects covered by the paper are not included in the Introduction, e.g. results shown in Section 2, where the effects of sources/sinks and advection are compared. These are very relevant but the Introduction does not say anything about this being an objective of the paper, this information should be added to Section 1.

Reply: Agreed, this information will be added.

Comment: Model recovery tests: what this means needs to be clearly

explained early in the text.

Reply: Agreed; we will explain this term when first mentioned.

Comment: Results shown in Figures are very interesting but on several occasions they are not sufficiently explained/ developed in the main text. An example is Fig 14: lines 217-220 should give more quantitative information on the amount of uncertainty the inclusion of CCl4 contributes to reduce, as well as explain why it does so more for the vertical field than for the horizontal one.

Reply: Agreed; the discussion of the figures, and particularly Fig. 14, will be extended. The reason why the inclusion of CCl_4 provides more information for the vertical field than for the horizontal one is that this gas has steep vertical mixing ratio gradients but not so pronounced horizontal gradients.

Comment: The Conclusions Section needs careful revision. This section should be understandable on its own as a Section that summarises the paper. Adding initial sentences summarising ANCISTRUS and why it was developed would improve its readability and completeness. Overall, statements in this section are not clearly backed up, a clearer reference to the results you have shown should be included.

Reply: Agreed, we will add some general information to make this section more stand-alone, and we will better link our conclusions to the results of the main part of the paper.

Comment: Technical corrections:
Line 4: Model recovery tests a brief explanation/definition would be helpful here.

Reply: Agreed; we will define this term when first used.

Comment: Lines 8-9: These two sentences should be merged to make the meaning clearer.

Reply: Agreed, we will merge these sentences.

Comment: line 17: citation here shouldn't use parenthesis

Reply: Thanks for spotting; this will be corrected.

Comment: Line 22: "Similar as in other applications of inverse modelling..." some citations to reference previous work and add context to this paragraph should be mentioned.

Reply: Agreed.

Comment: Line 25: Please consider including some brief information on sinks here, for completeness of text.

Reply: Agreed.

Comment: Line 37: Change "An application of" to "Applying"

Reply: Agreed. As non-native English speakers we appreciate any suggestions to improve the language. Many thanks! (This applies also to other comments in this review).

Comment: Line 39: Please develop "and so forth" to better understand what you mean here.

Reply: Agreed; we will be more specific here.

Comment: Paragraph 41-43 cannot be clearly understood in its present form. Please rewrite.

Reply: Agreed; we will rewrite this statement to make it better understandable.

Comment: Line 45: "...confidence in..."

Reply: Thanks for spotting!

Comment: line 51: Sentence needs rewriting. The word intuitively is confusing, the mechanisms described are those providing information to ANCISTRUS, it is not an intuition.

Reply: What we want to say: there are two idealized, simplified ways to look at this inverse problem, i.e. two ways to understand where ANCISTRUS takes the information from. Our tests show that none of these is fully adequate because ANCISTRUS exploits both. We will rewrite this to make clearer what we mean.

Comment: Line 64: "due to its long stratospheric lifetime, SF_6 is considered as inert in the given analysis range." If your method does not consider SF_6 sinks how can you overcome the biases caused by the mesospheric sink of SF_6 when using mean age-of-air methods? This needs to be clearly explained and justified as the manuscript claims this is one of the main advantages of the ANCISTRUS method in the study of the BDC. (See my Specific comment on this).

Reply: As explained above, this is because ANCISTRUS relies chiefly on measured gradients. For species without a sizeable stratospheric sinks like SF_6 it

relies fully on the gradients. These gradients might be affected by sink reactions in the analysis domain but not by sink reactions above the analysis domain. The age-of-air method is based on the mixing ratio difference between the stratospheric entry point and the target point. If the pathway of the air-parcel involves the mesosphere, this affects the age estimate. In contrast, ANCISTRUS uses the mixing ratio difference between two adjacent gridboxes, and the changes in these gridboxes from one month to the next, which are not affected by the mesospheric sink. If the air was in the mesosphere before, this does not matter, because the history of the air parcel is irrelevant. This is because for all gridboxes involved actual measurements are used. Will will expand on this in the manuscript.

Comment: 65-68: This paragraph cannot be clearly understood in its present form. Please rewrite and clarify.

Reply: Here the same explanation as in the reply to the comment on line 64 applies. We will rewrite this to make it clearer.

Comment: Line 70: "at a certain point at one day. Change point to location; delete "at"

Reply: Thanks for the correction.

Comment: Line 73: Do you mean "in the real atmosphere"?

Reply: Not quite. This statement is not about what happens in the atmosphere but what of this is essential for ANCISTRUS to correctly reconstruct the fields of efficient velocity. We will try to find some clearer wording. Perhaps something like "As opposed to both these simplified views where information pathways are assessed in isolation, both mechanisms contribute to the full picture."

Comment: Figure 1 caption: for clarity, spell out month in the units (deg month-1)

Reply: Agreed.

Comment: Line 81: "not so much interested in the explanation of the atmospheric features", but this is the main scope of the methodology, right? If you do this in the companion Part 1 paper at least you should mention that here.

Reply: This study is meant as a technical validation study. A first step towards the scientific analysis of ANCISTRUS results is reported in von Clarmann et al. (2019). We will mention this.

Comment: Line 85: delete "broadly speaking" or substitute by "on first approximation"

Reply: Agreed; we will delete this without loss of information.

Comment: Line 88: (right panels) does not correspond to figures layout

Reply: Yes, indeed. This will be corrected. Thanks for spotting.

Comment: Line 90: "regardless if sinks are estimated.." to "regardless of sinks being estimated.."

Reply: Agreed, thanks.

Comment: Line 99: More information should be included on how the method copes with, and provides information on, the diffusive and dispersive characteristics of transport. (also Specific Comment)

Reply: The method provides effective velocities which include diffusive (physical diffusion, not numerical diffusion!) and eddy mixing effects. We understand dispersion to be a numerical effect of transport modelling but not a characteristic of what really happens in the atmosphere.

Comment: Line 101 and 120: The word severe is not the best one here, perhaps exhaustive, strict, tough..?

Reply: 'Severe tests' is a technical term (c.f., e.g., Mayo (1996). A test is called severe if the likelihood is large that it will refute a hypothesis if false.

Comment: Most of page 6, if I understand correctly, is mainly a summary of results in vCG16. If vCG16 shown the validity of the method, this should not be repeated here in a lengthy way, but perhaps written in a way that is more clearly related to the results you show in the current study, e.g. linked to the arguments you use to choose further tests.

Reply: We are afraid that, without this text, it might not be clear which conclusions of vCG16 are still valid (and why) and which are misleading due to a sub-optimal test setup. We will carefully check what can be shortened but in the case of doubt we prefer to give the full information in the paper under review.

Comment: Line 140: "SeptemberOctober 2005" does not correspond to what Fig 3 labels indicate. Please resolve.

Reply: Thanks for spotting!

Comment: Figure 3 and related discussion: what you mean by reference fields needs to be more clearly explained in the main text.

Reply: The reference case is the field of effective velocities used to calculate the trace gas distributions which then are used as "surrogate truth". In this test setup, ANCISTRUS uses these trace gas distributions to calculate "surrogate measurements" of trace gases. These trace gas distributions are then used to retrieve the velocity field. The comparison of the resulting field and the reference field contains information about the robustness of the ANCISTRUS method. We agree that the logic of model recovery tests and the involved specific terminology should be explained in the text.

Comment: Lines 143-144: how do these underestimation values compare to biases/uncertainties obtained with other methodologies?

Reply: To the best of our knowledge, there exist no other observational methods which provide a picture of meridional middle atmospheric circulation at a spatial and temporal resolution comparable to ANCISTRUS. The SF_6-based age-of-air method has uncertainties in the order of a couple of years due to the neglect of the mesospheric sink. Ray et al. (2016) report a difference between SF_6-based and CO_2-based age measurements of 14-6=8 years, which indicates a bias of more than 100%. Unfortunately no dense global vertically resolved middle atmospheric CO_2 measurements are available. To these the sink problem would not apply. Another problem with the use of CO_2 as an age tracer is the annual cycle, which causes ambiguities in the transformation of mixing ratios into ages.

Comment: Line 149: Please check labels of Figures and corresponding references in the main text match each other.

Reply: Thanks for spotting.

Comment: Line 151: "...are underestimated by about 25% but broadly speaking, the inversion is successful also in quantitative terms." Not clear what you mean, a 25% underestimation does not sound like a quantitative success. Please rephrase or explain further.

Reply: Given the possible large bias of the age-of-air based method and the lack of any other method which provides middle atmospheric meridional circulation at a comparable temporal and spatial resolution we find our results not so bad. Furthermore, the numbers quoted do not describe the typical underestimation but features that stand out as particularly problematic. And beyond this, all the patterns are recovered in a very robust way.

Comment: Line 155: Move "(Figs 4 and 6)" somewhere else within the sentence, it is not clear whether these two figures refer to the August-September-October 2010 cases or the previous tests.

Reply: Agreed.

Comment: Line 156: Include some quantitative information on the slight underestimation to put it into context with the results presented earlier. Overall, in the discussion of Figs 3 to 6, more information/explanation should also be included on the reasons for the under/overestimation of fields.

Reply: Agreed to present numbers. The damping of the amplitudes is quite a natural thing when a regularization is used which constrains the differences of values at adjacent gridpoints. We will expand on this in the text.

Comment: Line 157: But has it removed existing fields in any occasion? Please add some sentence on this.

Reply: You mean existing patterns? We have not observed that any sizeable circulation pattern has been removed.

Comment: Figures 3 to 6 use different color scales for the differences (lower panels in the figs.), wouldn't it be better to use the same color scale to facilitate comparison?

Reply: The values in the difference plots are much too small to be shown in a common color scale. For the related discussion it is essential to clearly resolve the differences.

Comment: Line 172: Why this particular year?

Reply: We could have chosen almost any year and any month. The only months less suitable for such tests are those where some gases had to be discarded. There is no further particular reason why we have used 2010; September-October is interesting because of the pronounced structures and large velocity contrasts. With respect to that, Sep-Oct is a particularly severe test which is supposed to react quite sensitive to the choice of the regularization parameter.

Comment: From results in Figure 7 it seems as if weaker regularization produced better results (middle right panel), why haven't you chosen that regularization instead of the nominal one? If it is due to convergence problems, wouldn't it be useful to show also results for other month/year where the stronger regularization does not work so well?

Reply: Indeed we find that our chosen regularization is a fair compromise between accuracy and stability. Since currently long-term studies where data gaps are to be avoided are in the focus of our research interest, we have chosen the stronger regularization. For case studies it may be worthwhile to optimize the regularization strength to the particular case. The regularization is not hardwired but a user-defined input and can easily be adjusted to the actual needs. To the second part of the question: Do you mean "where the weaker regularization does not work so well"? In these cases we simply have no results. We do not consider the intermediate results of a non-converged iteration as meaningful result. They are simply a data gap.

Comment: Line 196-197: If you mean that low sensitivity to the omission of a single species shows the robustness of the methodology, I agree and suggest rephrasing this sentence to make it clearer.

Reply: This is exactly what we mean, and we agree to state this conclusion more clearly.

Comment: 198: "respective" to "corresponding"

Reply: Agreed, thanks.

Comment: Line 199: "similar to a jackknife method", not sure what this means in this context and not sure this part of the sentence is necessary, the set-up is clear.

Reply: The reviewer of von Clarmann et al. (2019) explicitly demanded 'Jackknife' tests to be performed.

Comment: Line 202: gradients between regions

Reply: Agreed, thanks.

Comment: Some of the Figures 8-13 could/should be combined as multi-panel figures (6 or 9 panels/fig) to reduce the number of Figures and facilitate looking at results in a more straightforward way.

Reply: Agreed.

Comment: When describing the figures in the main text, some quantitative data should be added, e.g. percentage contribution for each species.

Reply: Agreed to provide some quantitative information. However, this has to be limited to selected examples. The percentage contribution of a gas is different for each gridpoint and each time, thus the full quantitative information

cannot be communicated in the text.

Comment: Lines 214-216: If the information coming from the mentioned species contributes to reduce uncertainty, then it is neither useless nor redundant; please consider rewriting these sentences to avoid confusion.

Reply: We did not mean 'redundant' in a dismissive sense. We agree to rephrase this for clarity.

Comment: This is also a general suggestion for the whole of Section 5, results in this section show the importance of the different species and the different role they play in forming the final resulting fields, therefore I would suggest not using the word "redundant". Otherwise, why would you use, and show here, redundant information? As far as I understand you have included all species to obtain the final ANCISTRUS results, right? If not, this should be more clearly stated early in the manuscript.

Reply: Indeed we have used information of all species. We agree that 'redundant' is misleading in this context.

Comment: Figure 14: How does the standard deviation responds to the omission of some of the other "minor" species? It would be worth adding one sentence to the main text and perhaps some additional panels to this figure.

Reply: Omission of other species has a larger effect. We have chosen CCl_4 because for this species the information it provides is least evident from the Jackknife test. For CCl_4 we felt the largest pressure to justify why we consider it at all. We agree to add some information on the other species.

Comment: Line 221: Please introduce ANCISTRUS at the start of this Section. See also my Specific Comment about Conclusions. Some sentences read as contradictory. For example you say "fairly accurate", then "perfectly reproduced", and then again that there is still room for fine-tuning for a better retrieval of velocities. Overall statements in this section are not clearly backed up, a clearer reference to the results you have shown should be included. The meaning of the last sentence is not fully clear.

Reply: We agree to include some general information in the conclusion. We do not see a contradiction in our statements. Accuracy refers to the quantitative results, while perfect reproduction refers to the recovery of structure, which is another category. Further, there is nothing principally wrong with results obtained with a strong regularization. They just represent a smoothed version of

the true state. The fine tuning does not make the inversion better in a general sense but more adequate for a particular purpose. We will rewrite the conclusion to make this clearer, and we will better link our conclusions to the main part of the paper.

Review #2:

Comment: This manuscript is meant to demonstrate the robustness of the "Analysis of the Circulation of the Stratosphere Using Spectroscopic Measurements (ANCISTRUS) algorithm described in Part 1 several years ago (von Clarmann and Grabowski, ACP,2016; vCB16). ANCISTRUS is a continuity equation inversion methodology that relies on monthly differences in trace gas distributions to derive "effective velocities" that describe trace gas transport. I very much appreciate the concept and it would be a great boon to the community if it were demonstrated to be successful in providing information about the stratospheric transport circulation.

Reply: We do not understand why the form of the counterfactual conditional has been chosen here.

Comment: The paper is mostly well-written and easy to follow and the model recovery tests and sensitivity tests do indeed demonstrate that the model is relatively robust in terms of being able to reproduce its own results.

Reply: We are afraid that the reviewer has misunderstood the model recovery tests. The test did not merely show that the model reproduces its own results. The tests have shown that the algorithm, applied to tracer distributions related to a **known** field of effective velocities does reproduce these. This is a standard procedure in testing inverse methods.

Comment: However, the lead author [...]

Reply: Does the reviewer suggest that this paper is not co-written by both authors and that its content is not agreed by both authors? Is there any indication of this? Why this *ad personam* comment?

Comment: [...] has referred to this manuscript as a "validation" of the method in the interactive discussion of a second paper under review at ACP (von Clarmann etal., ACP, 2019; vC19), and I find that it falls far short of that description. The model recovery tests, in particular, demonstrate only that the model will retrieve more or less the same effective velocities from more or less the same tracer distribution [...]

Reply: We disagree. In one case, we use MIPAS tracer fields; in the other case

we use tracer fields generated by the model. That these are similar is simply another positive instantiation of the validity of the method. It proves that the velocity fields chosen are actually a solution of the problem. Otherwise the tracer fields would be quite different from the measured ones.

Comment: [...] but do not provide any assessment of whether those effective velocities have any physical meaning or are a unique solution to the continuity equation (these comments are explained in more detail below).

Reply: The physical meaning of the effective velocities is quite clear: The resulting effective velocities are those 2D velocities which best reproduce the changes in zonal mean mixing ratio distributions, under consideration of the continuity equation. The physical meaning is identical to age-of-air differences over distance, except that we derive this quantity at better temporal and spatial resolution. Although we apply a lot of diagnostic tools, we have not found any indication of problems with non-unique solutions. For detail, see below, under 'specific comments'.

Comment: If the ANCISTRUS method is to be used to study stratospheric transport in a meaningful way (and the authors indeed attempt use the method to provide a climatology of the meridional circulation in vC19), then those properties must be demonstrated. I therefore cannot recommend publication of this manuscript in ACP without major revisions that address these concerns.

Reply: We have to respectfully disagree. If we understand the reviewer correctly, they say that this discussion paper should be rejected just to block publication of vC19. We do not think that this is the regular reviewing procedure of ACP. This manuscript should be judged by its own content, independently of vC19.
In the manuscript under discussion we have applied the neccessary tests to show that ANCISTRUS does exactly what it is supposed to do. The physical meaning of the velocity fields, provided of equations containing the transport variables of a 3D atmosphere has been provided both in the appendices of vCG16 and vC19. If the reviewer would take the effort to look into these appendices, they would better understand what the physical meaning of the velocities provided is. The fact that these are not the same the reviewer is used to is no reason to dismiss this scientific work.

Comment: Major technical comments: Lines 32-33: It has been demonstrated several times (Neu and Plumb, 1999; Linz etal., 2016; Linz et al., 2017) that the age of air is not a good measure of the meridional circulation, but that the age difference between upwelling and downwelling regions is, in fact, equivalent to the diabatic circulation.

Reply: We do agree that age differences are a useful measure, but they cannot be measured globally without relevant uncertainties. The only global age measurements are based on SF_6; these measurements, however, are strongly biased due to the the mesospheric sink of SF_6. This bias is different in different regions. Thus the age differences as a measure of the meridional circulation will be biased.

Comment: This methodology does not require assumptions about the age spetra.

Reply: Ploeger and Birner (2016) have shown that age spectra have a strong interannual and seasonal dependence. We think that substracting mean ages associated with different age spectra will also be affected by the differences of the respective age spectra and thus means comparing apples and oranges.

Comment: Certainly if ANCISTRUS were able to successfully retrieve the BDC then it would have some advantages over the age difference, but to compare it to the use of age itself as a circulation diagnostic is somewhat disingenuous.

Reply: Above we have put forward arguments why age differences are affected by the mesospheric SF_6 sink. In the case of the mesospheric SF_6 sink, the age differences between different latitudes are even more affected than, e.g., trends at one latitude, as estimated, e.g. by Stiller et al. (2012) or Haenel et al. (2015). We agree that age differences, based on an ideal age tracer, might be an appropriate diagnostic of the circulation in the model world; in the real world, however, where one depends on observational data, this method is deficient, and we thus see no reason why our criticism shall be "somewhat disingenuous".

Comment: Lines 39-41: The fact that the interannual variability in the ANCISTRUS-derived circulation is small, particularly in the tropics (from having looked at the figure in vC19), is a red flag for me. We know that the QBO's secondary meridional circulation has a large influence on trace gases in the tropics and subtropics, and any tracer-derived circulation should pick up this variability. It is a very clear signal in trace gas anomalies.

Reply: First, we would have expected here a review of this manuscript and not one of vC19. And secondly, we have **not** said that the "interannual variability in the ANCISTRUS-derived circulation is small", but that "for each year similar circulation fields were found for any particular time of the year." We do see, for example, a clear QBO signal in the inferred fields of effective velocity. This QBO signal is currently under investigation but it belongs neither in a technical validation paper nor in a paper which focuses on the climatology in the sense of multi-annual mean circulation.

Comment: Lines 55-59: I feel that the entire concept of the meridional circulation in this manuscript is highly oversimplified, and this is one example of such oversimplification.

Reply: We are afraid that the reviewer grossly misunderstands the purpose of Section 2. We clearly state that here we do **not** describe the "concept of the meridional circulation" but that we investigate two "candidate mechanisms [that] can explain where ANCISTRUS takes the information from..." (line 51). This is not our view of the circulation but an assessment of the sensitivity of ANCISTRUS to the various information sources.

Comment: The stationarity condition can, in fact, define any number of circulation fields with different mixes of horizontal and vertical advection. In principal, these components might be separable with the right set of trace gases, but there is no evidence presented here that the suite of trace gases used is sufficiently orthogonal to separate horizontal and vertical advection unequivocally.

Reply: The evidence is in that the system of equations has a solution at all. The fact that the condition number stays within reasonable bounds proves that the system of equations we solve has no problem with collinearity. If ambiguity due to insufficient orthogonality were a problem, the inversion would face a singular matrix and we would not get any solution at all. If we faced major ill-posedness of the inverse problem, this would show up as huge errors in the error covariance matrix of the result (which is calculated routinely). We do agree that the solution would be ambiguous if we had data at two places only, but we have many data points and the continuity equation has to be satisfied everywhere. Since we do not have only mixing ratios at two points but full vertically and latitudinally resolved distributions of air density and a series of trace gases, the inverse problem is better constrained than one might think. We have by far more equations than unknowns, and we reduce the effective degree of freedom of the system further with the regularization term. Ill-posed inverse problems going along with ambiguous solutions are terribly sensitive to noise and are instable in the sense that infinitesimal changes in the input entail huge differences in the output. We observe the opposite. If the solutions were indeed ambiguous due to the lack of orthogonality, it would not be possible that ANCISTRUS finds similar structures independently for many years. If the inverse problem really was ill-posed, it would be over-sensitive to variations in the mixing ratios. It would produce very different results when we apply ANCISTRUS to, say, the same month of a different year. We observe the opposite. Further, it would not be explainable that patterns evolve smoothly from one month to the next. Also it would not be possible that discarding a gas has only minor effect on the result. Beyond this, non-orthogonality would lead to a solution-space instead of a point-solution. The mathematical and diagnostic tools we use are well established standard and widely used in many fields of science.
Of course the steady state assumption provides less information than the regular

case where structures are transported. But this is exactly the point we want to make in this section. This test case is an investigation of this information pathway in isolation. With this test case we show that the idealized steady state assumption does not provide sufficient information to recover the circulation field in full. Thus we do not understand how the criticized lack of information in this test case can be put forward as an argument against ANCISTRUS which, in its normal application, uses both pathways.

Comment: Lines 69-72: This is another example of oversimplification. The change in amplitude of the structures is also affected by mixing in the real atmosphere.

Reply: We do agree, and it is for this reason why we call the velocities effective velocities. By the way, the age of air differences as a measure of the circulation share the same characteristic. Our effective velocities can be conceived as age of air differences far better resolved in space and time.

Comment: More importantly, while the simplest (not necessarily best) explanation might be a southward velocity, another explanation would be a shift in the upwelling region (which brings high mixing ratios upward from the tropopause) by 5 degrees south. This would indeed appear as a change in effective southward velocity based on the tracer inversion, but that southward velocity is not a meaningful description of the meridional circulation.

Reply: We did not expect that the reviewer (or any reader) would take this simple example in the paper literally. We tacitly assumed that it is clear that the continuity equation is satisfied at any point in the system. We analyze the mixing ratio changes at all points simultaneously, and an unphysical velocity vector which would be the simplest solution for one point in the system would cause increased residuals at the other gridpoints. Since ANCISTRUS minimizes the residuals at all gridpoints simultaneously, it would not accept such a solution but search for the global minimum of residuals.

Comment: In fact, if anything [...],

Reply: Is there any evidence that the effective velocities might not represent anything? Or is this just a rhetoric trick to dispraise our paper and our method?

Comment: [...] the effective velocities seem to represent anomalies in the meridional circulation rather than the circulation itself. The effective velocities are derived from the change in trace gas distributions from one month to the next, but that distribution for any given month already reflects the mean meridional circulation when using real stratospheric trace gases.

Reply: We strongly disagree. At places where the change of mixing ratios is zero, the equations provide the information from the balance of advection and sinks; where we have patterns which are transported and thus go along with local changes in mixing ratios, these provide additional information. The patterns themselves may be considered as anomalies, but how these are transported is controlled by the total (i.e. mean plus anomalies) circulation. The signal is imprinted by any – random or seasonal or whatever – effect. The most prominent such effect is the atmospheric water vapour tape recorder. The imprinted signal is an anomaly in the sense that the water vapour amount at the stratospheric entry point has a pronounced seasonality. But how this signal is transported upwards just reflects the total circulation, not only its anomaly. We have many more species than water vapour only, and due to the natural variability of the atmosphere, there is a huge number of anomalies in the mixing ratio distributions. And these patterns are transported, horizontally and vertically; and similarly as the tape recorder, where the ascent of H_2O anomalies provides information on tropical uplift, the displacement of other anomalies we see provides additional information on the actual circulation. All these "additional quasi-tape recorders" contain signal about the total circulation, not only on circulation anomalies.

Comment: The familiar shape of tracer isopleths, with an upward bulge in the tropics, strong gradients in the subtropics, relatively flat isopleths in midlatitudes, strong gradients at the vortex edge, and a downward bulge in the vortex are all a reflection of the balance between sinks and the mean meridional circulation and effects of mixing.

Reply: Yes, and within ANCISTRUS the interplay between sinks and advection is an important information source. This information, however, is not exploited for annual mean fields but for actual ones and is complemented by the information contained in the pattern transport. ANCISTRUS provides the total actual circulation field and not the steady state field in isolation.

Comment: When you look at the change in this trace gas distribution from one month to the next, it reflects at best [...]

Reply: What does the reviewer intend to say with the words "at best", and on which evidence is this based?

Comment: [...] the month-to-month change in the circulation, but not the overall circulation itself.

Reply: On top of the steady state the trace gas distributions in the real world change from instance to instance. This is because of the time-dependence of sinks, the time-dependent lower boundary condition, and a natural variability of circulation. As said above, what we get is the total circulation at a certain time. Who denies the information content of pattern transport on the total cir-

culation commits oneself to also deny that the atmospheric tape recorder bears any information on the circulation. What we see is the total actual circulation, composed of the background circulation and its anomalies.

Comment: All of this highlights the absolute need to understand how well the ANCISTRUS method retrieves an actual circulation field rather than an idealized one (as in Part 1 of the manuscript) or one that it has already generated itself (as in the model recovery tests in this manuscript).

Reply: Here the reviewer seems not to distinguish between the tests of the forward model and the test of the inversion scheme. We refer to Part 1 of the manuscript only for the tests of the forward model. For this purpose, idealized tests are the most severe ones, because diffusive and dispersive characteristics of the transport model show up clearly, and the results can be verified by analytical calculations. We do **not** refer to Part one of the manuscript for the tests of the inversion scheme. The requirement to use an "actual circulation field" is unfulfillable because the actual circulation field is unknown and unknowable. Using ANCISTRUS-derived fields as reference fields guarantees that the reference field satisfies the continuity equation and thus can be recovered by the scheme. Related mixing ratio fields at the end of the time step are **not** the same as used in the first analysis. Thus, the model recovery test is **not** a repetition of the first inversion.

Comment: Lines 76-77: I am not sure I agree with the statement that the circulation fields roughly match our expectations of the meridional circulation. For one thing, it is extremely difficult to tell whether this is the case or not from the vector plots. The streamfunction should be plotted instead, with the vectors superimposed over the streamfunction contours if desired. From the plots in the manuscript, the only thing that fairly clearly matches expectations is the circulation in the mesosphere,[...]

Reply: To us the vector representation is more instructive. We appreciate that different people have different preferences, thus we will make all the data of this paper available via the KITopen portal. Then everybody can plot the data in their own preferred representation.

Comment: [...] though the seasonal differences in the height of the circulation are odd (but might more accurately be called interannual differences since two different years are used).

Reply: Does there exist any observational evidence against this altitude difference?

Comment: I certainly do not clearly see the "branches of the BDC"

(line 80, and this phrase should be referenced and defined) – in fact it is hard to see any coherent tropical upwelling region at all. Again, plotting the streamfunction would make the circulation characteristics much clearer.

Reply: We see a lot of the expected features in, e.g., the top panel of Figure 1:

1. subsidence in the Antarctic in early Austral winter;
2. a small but coherent upward component above the equator (the tropical upwelling is a very slow process; one cannot expect to see it as clearly as, say, polar winter subsidence);
3. poleward velocities at about 20–30 km and above 35 km altitude in the Southern hemisphere;
4. poleward velocities at about 15–20 km altitude at Northern midlatitudinal and polar latitudes;
5. a signal of a sudden stratospheric warming.

But all this discussion is only about a little side remark and has little to do with the test we present. The purpose of this figure is to demonstrate how both the advection-sink balance and the pattern transport contribute to the full picture. We see, for example, in Fig. 1, middle panel, that without pattern transport (viz.: tape recorder!) the tropical updraft cannot be retrieved.

Comment: Lines 77-79: This is another example of an important difference between the effective circulation based on tracers and the BDC. This upward velocity is not meaningful as part of the meridional circulation, which is still downward but weaker than prior to the vortex displacement.

Reply: We agree that the physical velocity vectors of an air parcel point downward. The problem is that in the 2D world in a polar coordinate system the displacement of an initially perfectly symmetric vortex off the pole cannot be represented, and there exist no latitudinal velocities that could generate the observed effect of increasing VMRs of most trace species above the pole. To retrieve a velocity which does not exist in the 2D world is too much to ask from a scheme that is based on the 2D continuity equation. The counter-intuitive result does not hint at a problem with ANCISTRUS but it does hint at a problem with any 2D representation of the 3D world. Given the characteristics of the 2D world, ANCISTRUS retrieves exactly the perfect solution, i.e., the only 2D velocity field which is able to reproduce the observed trace gas observations. As we understand that the BDC is a 2D description of stratospheric circulation, we do not quite agree that this is a "difference between the effective circulation based on tracers and the BDC".

Comment: Again, it may be more appropriate to view the effective velocity not as a proxy for the BDC, but as anomalies on the background BDC circulation. But this must be demonstrated using an actual circulation field.

Reply: We disagree; we do not see anomalies but we see the total 2D-circulation (background plus anomalies), which must, however, not be conceived as the average of the 3D velocities, due to the eddy terms and effects discussed above.

Comment: Lines 87-94: The plots using annual mean tracer values are, in fact, the only ones that look like the prototypical middle atmospheric circulation to me. The authors seem to indicate that the lack of a pole-to-pole circulation is a deficiency, when, in fact there is no coherent pole-to-pole circulation in the annual mean (nor is there one during the equinoxes, from which the sink terms were used). I also see evidence of the "tropical pipe" ending at25 km, where there is strong poleward advection, rather than "reaching up to the mesosphere". The "pipe" is not defined by upwelling, but rather by a lack of communication with the midlatitudes.

Reply: Figures 1 and 2 are not meant as a discussion of atmospheric processes. They are meant to show that both information paths, advection-sink-balance and pattern transport, are important to reconstruct the full picture. Since by now no time-resolved global measurements of the meridional circulation were available, it is no surprise that the annual mean example looks more familiar. This, however, does not imply that it is closer to truth for the actual month. Similarly, prior to the invention of the telescope, when the human eye could not resolve the satellites of Jupiter, the prototypical sky was one without Jupiter's satellites.

Comment: Lines 105-112: The authors assert that many tests of this nature were performed for vCG16, [...]

Reply: This rephrasing of our text by the reviewer does not at all capture what we say in the lines quoted. The reviewer's paraphrasing sounds as if we wanted to suggest that we have made enough tests in vCG16. But what we actually say is quite the opposite. It is hard to believe that this is unintentional. We consider this as a rhetoric of which the only purpose is to create some animus against the authors. The wording "assert" seems to suggest that the authors are lying.

Comment: [...] but the only ones described or shown used very simplified velocity and tracerfields.

Reply: We have to distinguish two cases: The tests of the forward model and the tests of the inverse model.

A transport forward model is best tested with very simple and extreme cases (large gradients and gradient changes in the fields). This is the only reasonable way to test the diffusive and dispersive characteristics of a transport model. With realistic cases multiple effects are superimposed, and we have no reference to compare with. We thus consider these tests as severe and valid. And we clearly state that the tests of the inversion tool made in vCG16 were insufficient because the reference fields to be retrieved did not comply with the continuity equation. Since the continuity equation is a hard constraint, large differences between the results and the reference fields were unavoidable. To remedy this deficiency is the main purpose of this manuscript. Thus we do not understand why exactly this deficiency is criticised here. Here (and elsewhere) our arguments are torn out of context to twist our words.

Comment: What is required is a model recovery test using a realistic meridional circulation (with vertical and horizontal components and satisfying the continuity equation) and realistic trace gas distributions with both vertical and horizontal gradients. I am not convinced that ANCISTRUS can successfully retrieve a unique solution to the continuity equation that does not alias horizontal and vertical components of the circulation into one another.

Reply: The tests we present are based on realistic trace gas distributions and use reference fields of effective velocity that satisfy the continuity equation. As described above, ambiguities between horizontal and vertical components of the circulation would show up in the respective off-diagonal elements of the error covariance matrix of the results, in very different solutions for slightly different situations, and a failed model recovery test. All these diagnostics are established standard.
We have meanwhile model recovery tests available based on the annual mean states (considered as more realistic by the reviewer; not by us, however). ANCISTRUS recovers the velocity fields even better than in the model recovery tests presented in the manuscript. This is because the test cases chosen for the paper were particularly difficult cases with a lot of structure.

Comment: Lines 120-130: I am unable to understand why one cannot take the 2-D Transformed Eulerian Mean circulation from a CCM and use it to advect an initial MIPAS trace gas distribution and then retrieve the circulation using ANCISTRUS to see if you recover anything like the model circulation. This would be similar to the tests in vCG16, but using realistic velocity and tracer fields. Some sort of test of this type must be performed before ANCISTRUS can be found to inform our knowledge of the middle atmosphere meridional circulation.

Reply: MIPAS mixing ratios cannot be combined with modeled velocity fields because these are typically not consistent with each other. As we have learned

from the tests in Section 2, the full information is not contained in the difference fields alone because of the sink-advection-balance. To combine CCM 2D velocity fields with MIPAS might be adequate for SF_6 which has no stratospheric sinks. Only for SF_6 we have $\hat{\vec{v}} = f(\Delta vmr)$, where $\hat{\vec{v}}$ is the estimated field of effective velocity vectors, and where Δvmr is the field of mixing ratio differences between the beginning and the end of the time step. SF_6 alone, however, is not sufficient to constrain the inverse problem. For gases with stratospheric sinks we have, due to the compensation of sinks by advection, $\hat{\vec{v}} = f(\Delta vmr, vmr_1)$, where vmr_1 is the initial velocity field. That is to say, for other concentrations, other velocities are necessary to balance the sinks. The velocity information is not provided by the mixing ratio differences alone. Model velocities are not identical to the real velocities which made the trace gas contributions as they are. Thus, one cannot expect that ANCISTRUS retrieves the modeled velocities, because there is a 'hidden' velocity term in the absolute values of the mixing ratios. From a validation which will result in differences between the result and the reference velocity field which can be explained by such inconsistencies we do not learn anything about the reliability of ANCISTRUS. We need a test setup which allows to unambiguously attribute each discrepancy to ANCISTRUS.

Comment: Lines 131-136: As far as I can see, all this demonstrates is that ANCISTRUS is capable of retrieving the same answer when you invert the same field.

Reply: We are afraid that the reviewer has grossly misunderstood the logic of the model recovery test. The key point is that we need (a) a field of effective velocities satisfying the continuity equation, and (b) tracer fields which are perfectly consistent with this velocity field. We achieve this by generating the tracer fields with our own model. This guarantees that we can attribute all differences between the result and the reference field to our inversion and that there is no other "excuse".

Comment: The effective velocity fields were generated based on the change in trace gases between two months. There is no reason that applying those velocity fields to the initial trace gas distribution should result in a different change in the trace gases than what was used in the initial retrieval, and so for the same distribution, ANCISTRUS essentially gets the same answer twice.

Reply: We disagree. If ANCISTRUS was defective, it would not get the same answer twice. If, e.g., ANCISTRUS would alias vertical into meridional velocities, this effect would also be visible when the fields resulting from the tests are compared to the reference fields.
Our test is by no means redundant with the initial inversion. In the initial inversion the mixing ratio distributions were measured ones. In the model recovery test the mixing ratio distributions are calculated ones. Since the system of equations is over-determined, these two cannot be the same. The measure-

ment space is of a far larger dimension than the retrieval space, and in the least squares inversion this excess information is lost, we will not get it back with the forward calculation. The forward model will thus not be able to exactly reproduce the initial, measured, mixing ratio distributions. The fact that the distributions are similar is just another proof that what we got first is indeed a valid solution of the inverse problem.
If something went wrong with the inversion, we would **NOT** get anything similar to the reference velocity distribution in the model recovery test. We do not claim that the model recovery tests are meant as a test of the forward model involved. This has been tested independently in vCG16.

Comment: This test does not in any way validate that the effective velocities derived are in anyway related to actual transport velocities, [...]

Reply: This is not the purpose of the model recovery test. The testing has been split up into two logical steps. The forward model test in vCG16 provided evidence that the forward model involved models the transport in a realistic manner. The model recovery test provides evidence that a solution consistent with the forward model in use is found, and only both these tests together validate that the effective velocities derived are in anyway related to actual transport velocities. The model recovery test does show that we get the reference velocity field back if we feed ANCISTRUS with the associated mixing ratio data. This is exactly the purpose of a model recovery test, and ANCISTRUS has passed this test. If ANCISTRUS aliased vertical and horizontal velocities when applied to MIPAS data, there is no reason why it should not alias these again when fed with simulated data and cause differences from the reference field of effective velocity.

Comment: [...] nor does it demonstrate that the retrieved circulation is a unique solution to the continuity equation that properly resolves both the vertical and horizontal components of the circulation.

Reply: If the solution was not unique the model recovery test would not reproduce the reference field. We do not observe suspicious error covariances between vertical and horizontal velocities.

Comment: Lines 137-139: Even with the reduced vertical scale plots, it is again very difficult to see and interpret these results from vectors. The streamfunction should be plotted, as well as difference plots between the initial and final streamfunction.

Reply: As stated above, we will make the data available. Every interested reader can then plot the data in their preferred way.

Comment: Lines 142-143: Again, I do not easily see the "strato-

spheric branches of the BDC". Please plot streamfunctions and define what you mean be "branches" (I do understand what is meant, but many readers may not).

Reply: For the representation, we will provide the original data to allow each reader to plot them in their preferred representation. To avoid quibbling about words we will replace "stratospheric branches of the BDC" with "poleward transport in the SH subtropics at 25 km altitude and in the NH subpolar region at 15 km altitude."

Comment: Lines 147-148: I'm not sure I agree that the "the slow circulation patterns in the tropopause region and the lower stratosphere are well recovered". If plotted as percent differences, I think some very large discrepancies would emerge.

Reply: And if the true value was zero, even the best recovered velocity would have an infinite error. Percentage values can be very misleading when the reference values are small.

Comment: Lines 153-154: The right panels of Figure 5 are the only figures in the manuscript that seem to resemble the canonical stratospheric meridional circulation. They show rapid poleward transport by the shallow branch (below 15 km here) and strong tropical upwelling, poleward transport, and high latitude downwelling. No other plot shows a coherent upwelling region like this one does. Of course differences are to be expected given the seasonality of the circulation, but the upwelling branch moves back and forth across the year rather than disappearing.

Reply: We are happy to hear from the reviewer that the panels on the right of Fig. 5 satisfy the picture they are used to. The top right panel of Fig. 5 is just a zoomed version of the top right panel of Fig. 3. It is a result of ANCISTRUS, restricted to the altitude range the BDC is usually looked at, with a velocity scale that is adjusted to the low velocities appearing here (in contrast to the high velocities that dominate the upper stratosphere and mesosphere). The tropical upwelling is an extremely slow process and is easily masked by the seasonality. It can be seen in the third panel of Figure 2 that we do have the tropical upwelling as a background signal. In the individual months, this signal is, however, superimposed by other processes related to, e.g., seasonality or inter-annual variation (QBO, ENSO, ...).

Comment: Lines 157-158: While it is true that the second retrieval did not create significantly different patterns than the first, it has not been established that the patterns retrieved in the first place are not artificial given that the algorithm does not appear to have ever been tested with a realistic circulation pattern and realistic tracer

distribution.

Reply: What 'realistic' velocity fields do we have available? Models? Funke et al. (2011, their Fig. 14) have fed 10 different models with the same measured distribution of NO_y, which can be considered as an inert tracer on the relevant time-scale. Already after 3 days, 10 very different distributions were predicted, and the differences were attributed to transport modelling. If model fields are realistic, which of these realities is the real reality? Or do we have parallel universes?
And does the reviewer intend to label our tracer distributions as unrealistic? These are based on MIPAS measurements, and a lot of validation studies have provided evidence of their reliability. What more realistic global tracer distributions do we have available?

Comment: Figure 3: There are obviously large differences in the velocities at 60S, 60 km for Feb-Mar. Why don't these show up in the difference plots? There are also other examples where the difference plots do not seem to reflect the visual differences between the top and middle plots.

Reply: At 60S, 60 km for Feb-Mar, the reference field shows values slightly larger than 8; the retrieved field shows values around 6 to 7, and the difference field shown values around 2. We do not see what the problem is. We have randomly checked other instances and did not find any inconsistency either. We do not know what the reviewer is speaking about.

Comment: Lines 201-204: Why does withholding CFC11 give the opposite signal to CFC12 in the Arctic? If the sinks are properly accounted for, the effective transport for these two species should be similar.

Reply: These gases have their strongest vertical gradients in different altitudes and have quite different lifetimes. Furthermore, we solve an overdetermined system of equations, and measurements are not always perfectly consistent. One gas may try to push the solution into one direction, and the other gas in the opposite one, and the least squares solution is a compromise. Removing one species in some way slightly tips the balance. It should be noted that these differences are quite small compared to the effective velocities (Note the factor 10^{-3} in the titles of the panels).

Comment: Line 208-211: I do not understand what is meant by "compressed colour scale". Again, the streamfunction and percent differences might be more useful for seeing the stratospheric changes.

Reply: This means that a larger range of values is covered by the colours.

Comment: Lines 211-212: The water vapor "tape recorder" has been extensively used for deriving vertical transport in the tropics, yet water seems to do nothing to inform the tropical upwelling. Can this be explained?

Reply: Yes, it can be explained. The answer is that the other species include so much information already that adding consistent information from water vapour does not change a lot. This just means that the information from water vapour and that of the other species is pretty much consistent. By the way: as said above, the ascent rate of the tape recorder is exactly the same concept as our pattern displacement concept discussed in Section 2.

Comment: Lines 232-233: If this is meant to refer to circulation patterns and structures, then I have to say I strongly disagree that there is evidence that ANCISTRUS is fit for purpose. It does indeed generate a consistent set of patterns and structures from a given set of trace gas fields, but there is no evidence that these patterns and structures are physically meaningful in any way. Until this is demonstrated using a known, realistic circulation field with the MIPAS tracer measurements, I cannot recommend publication of this manuscript.

Reply: Model recovery tests as we have performed them are the standard procedure to test inverse schemes. It is the fundamental logic of model recovery tests that some input is generated using some 'surrogate truth' with the forward model and to test if the model is able to reproduce the 'surrogate truth'. Model recovery tests alone do not demonstrate that the structures are physically meaningful, but complemented with the forward model tests in vCG16 they do. The model recovery tests demonstrate that the inversion procedure does what it is supposed to do. All the physics (which makes the results 'physically meaningful') is in the forward model which has been tested independently.

Comment: Minor comments: Line 1: The wording "allows to infer" is not grammatically correct (it needs a subject). I suggest "provides an inference of".

Reply: agreed.

Comment: Line 2: The phrase "both given by" should be "given by both"

Reply: agreed.

Comment: Line 4: Using "have shown" in the past tense makes it sound as if these tests were performed in another paper rather than here.

Reply: agreed.

Comment: Abstract in general: The abstract does not provide sufficient context for this work or provide any indication of the meaningfulness of the results.

Reply: Context will be added and the meaningfulness of results will be stated.

Comment: Lines 66-67: The phrase "does effectively not work" should be "effectively does notwork"

Reply: agreed.

Comment: Line 84: Should "or equatorward transport" be "of equatorward transport"?

Reply: agreed.

Comment: Line 88: The reference should be made to "bottom panels" rather than "right panels".Line 140: I believe "September October 2005" should be "March April 2005"

Reply: agreed.

Summary Reply to Review #2: This is a technical paper which presents tests of the ANCISTRUS analysis tool. The review is dominated by a dispraisal of the trace gas and velocity distributions we present. These, however, are not the topic of this paper. The topics of this paper are:

1. Which are the dominating information pathways explored by ANCISTRUS?
2. Can ANCISTRUS reproduce reference fields when it is fed with trace gas distributions consistent with these fields?
3. To which degree do ANCISTRUS results depend on the regularization chosen?
4. Which is the information content provided by different trace gases?

For the few comments which are related to these key questions we think to have shown that these are based on a fundamental misunderstanding of the purpose and the rationale of the related tests. Many of the comments do not discuss these tests at all. Furthermore, we recognize a certain disbalance between the will of the reviewer to accept well-founded criticism of the established methods they use and the manner in which they deal out criticism.

References:

Funke, B., Baumgaertner, A., Calisto, M., Egorova, T., Jackman, C. H., Kieser, J., Krivolutsky, A., Lpez-Puertas, M., Marsh, D. R., Reddmann, T., Rozanov, E., Salmi, S.-M., Sinnhuber, M., Stiller, G. P., Verronen, P. T., Versick, S., von Clarmann, T., Vyushkova, T. Y., Wieters, N., and Wissing, J. M.: Composition changes after the "Halloween" solar proton event: the High Energy Particle Precipitation in the Atmosphere (HEPPA) model versus MIPAS data intercomparison study, Atmos. Chem. Phys., 11, 90899139, https://doi.org/10.5194/acp-11-9089-2011, 2011.

F. J. Haenel, G. P. Stiller, T. von Clarmann, B. Funke, E. Eckert, N. Glatthor, U. Grabowski, S. Kellmann, M. Kiefer, A. Linden, and T. Reddmann Reassessment of MIPAS age of air trends and variability, Atmos. Chem. Phys., 15, 13161-13176, 2015.

Deborah G. Mayo, Error and the Growth or Experimental Knowledge, The University of Chicago Press, Chicago and London, 1996.

Felix Ploeger and Thomas Birner, Seasonal and inter-annual variability of lower stratosphericage of air spectra, Atmos. Chem. Phys., 16, 1019510213, 2016, www.atmos-chem-phys.net/16/10195/2016/doi:10.5194/ acp-16-10195-2016

Eric A. Ray, Fred L. Moore, James W. Elkins, Karen H. Rosenlof, Johannes C. Laube, Thomas Röckmann, Daniel R. Marsh, and Arlyn E. Andrews, Quantification of the SF6 lifetime based on mesospheric lossmeasured in the stratospheric polar vortex, J. Geophys.Res. Atmos.,122, 46264638,doi:10.1002/2016JD026198, 2016.

G. P. Stiller, T. von Clarmann, F. Haenel, B. Funke, N. Glatthor, U. Grabowski, S. Kellmann, M. Kiefer, A. Linden, S. Lossow, and M. Lpez-Puertas, Observed temporal evolution of global mean age of stratospheric air for the 2002 to 2010 period, Atmos. Chem. Phys., 12, 3311-3331, 2012.

Thomas von Clarmann and Udo Grabowski, Direct inversion of circulation and mixing from tracer measurements, part 1: Method, Atmos. Chem. Phys. 16, 14563-14369, 2016.

Thomas von Clarmann, Udo Grabowski, Gabriele P. Stiller, Beatriz M. Monge-Sanz, Norbert Glatthor, and Sylvia Kellmann, An observation-based climatology of middle atmospheric meridional circulation, Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2019-704, 2019.