Response to Reviewers - Acp-2020-95

We thank anonymous Reviewers 1 and 2 for their insightful comments and discussion of the manuscript. We provide an updated document, including updated figures, that incorporates the suggested changes and address specific comments below. For select comments, we list line numbers of modified sections of the manuscript and include the most relevant excerpts. Changes address also suggestions made by the editor.

Comments from Reviewer 1

(1) As described in section 2.1, inlet fractionation is problematic for Ar/N2 data obtained by airborne observations. The authors eliminated some data considering apparent biases or bad quality. I recommend that a little more details of data quality check or a criterion will be explained. The reason I care about this is that we can see clear difference of AoA-GS relationship between SH and NH (Fig. 4 and Lines 236-240). Is there a possibility that the data is still affected by the inlet fractionation and/or sample deterioration?

While data within a single campaign are most comparable because the position, aircraft type and parts of the sampling equipment evolved between campaigns, some offsets exist between campaigns. We exclude Atom 1 due to obvious artifacts relating to both the inlet employed and the position of the inlet over the wing. The same inlet was used during ORCAS and ATom1, but in the case of ORCAS the positioning of the inlet on the fuselage of the aircraft eliminated some of the larger artifacts seen in ATom1. We think the ORCAS data are of acceptable quality for publication, but as most of the stratospheric samples from the Southern Hemisphere in our dataset are from ORCAS, we are therefore careful about interpreting the apparent difference between the Southern and Northern Hemisphere in the GS-AoA relationship, until new observations are available to confirm or refute this. This is clarified in the manuscript (line 279-282 and SI).

From manuscript: "However, almost all SH samples with mean AoA older than 1.5 years come only from the ORCAS campaign. This campaigned used a different inlet design than HIPPO and ATom which could cause a small artifactual offset and but the scatter in our observations generally makes it difficult to separate signal from noise for small interhemispheric differences."

From SI: "**$\delta(Ar/N_2)$ residual analysis**

To test the hypothesis that large artifacts or biases in $\delta(Ar/N_2)$ were introduced during flask sample collection, we look for the presence of a pressure dependency of $\delta(Ar/N_2)$ after removing what we have shown is the natural gravitational signal. Pressure serves as a proxy for sampling conditions and aircraft velocity simultaneously here because aircraft velocity also increases systematically with altitude or pressure altitude. A vertical profile of $\delta(Ar/N_2)$ (Fig. S1A) reveals no strong dependency on pressure in the troposphere below about 300 mb. Above ~300 mb, some samples represent stratospheric conditions and therefore have an age of air (AoA) greater than zero and more negative $\delta(Ar/N_2)$. For further analysis, we correct $\delta(Ar/N_2)$ for the natural gravitational signal ($\delta(Ar/N_2)_{grav}$) using observed AoA from each sample and a quadratic fit of the AoA-$\delta(Ar/N_2)$ relationship simulated in TOMCAT ($\delta(Ar/N_2)_{grav} = -8.124 \, AoA^2 - 22.73 \, AoA + 0.9529$). The result is shown in Figure S1B. Although there is scatter in the corrected data, in particular for HIPPO1–5, no systematic relationship between pressure and gravitationally-corrected $\delta(Ar/N_2)$ is evident, supporting the interpretation that aircraft sampling did not create large biases in $\delta(Ar/N_2)$. While there is some evidence of a pressure dependence in the ORCAS data from the Southern

(2) Needless to say, it is useful to show the differences of vertical distributions between the model results and the observation for the model studies. Actually, AoA calculated by TOMCAT has been compared with observations in the previous studies (e.g. Chabrillat et al., 2018). However, the comparisons with the observation results are only shown as the AoA-GS relationships in this paper and we cannot know how TOMCAT reproduces vertical distributions of GS. It would be better to show a direct comparison with the balloon GS data. Particularly, the authors emphasize that TOMCAT includes GS enhancement in the upper stratosphere and mesosphere, which is superior to the previous model in this respect. This advantage will bring a significant improvement especially in the vertical structure of GS. That is one of the reasons why it is recommended to compare with the observations for not only the AoA-GS relationships in the lower stratosphere but also the vertical distributions observed by balloons.

We thank the reviewer for the suggestion. A direct comparison of aircraft-observed and TOMCAT-simulated $\delta(Ar/N_2)$ and AoA in the lowermost stratosphere is shown in Figure S4 which we now reference in the text (lines 265-269). However, a direct comparison of our aircraft observations and modeling results in this case is not a particularly good metric of model performance or data quality because the samples represent air with unusual chemical composition for its environment. These filaments are often small-scale features that cannot be reproduced by TOMCAT with a model resolution of 2.8°. Therefore, simulated AoA in many samples is considerably lower than observed and $\delta(Ar/N_2)$ is correspondingly less fractionated. These effects at least partially cancel when studying the GS-AoA ("tracer-tracer") relationship and make it a more robust metric for model-data comparison.

TOMCAT has previously been validated against a range of AoA observations, as pointed out by the reviewer, and the configuration used for this study has a tendency to slightly underestimate AoA as discussed in the manuscript (line 262). This will also introduce a bias in $\delta(Ar/N_2)$ based on the robust relationship between $\delta(Ar/N_2)$ and AoA. Hence, we believe that a comparison to the available balloon data against pressure would not be particularly instructive and distract from the main message of the paper.

From manuscript: "A comparison of $\delta(Ar/N_2)$ and AoA between flask samples and TOMCAT is shown in Figure S4. However, the direct comparison should not be considered a good test of model performance because our stratospheric composition criteria intentionally target unusually old air in the lowermost stratosphere. For the same reason, profiles of $\delta(Ar/N_2)$ and AoA are not evaluated here. Instead, we suggest the AoA-GS relationship provides a more robust metric for comparison. Overall, our observations validate the implementation of GS for young (< 3 y) ages in TOMCAT."

(3) The authors hypothesize that an adequate representation of the mesosphere in models is critical for the AoA-GS relationship (Line 257). In fact, TOMCAT has a great advantage, because the top of model atmosphere is much higher than that in previous model study. However, I cannot fully understand that neglecting some components of molecular diffusion processes is really adequate even in mesosphere. In this model study, the molecular diffusion processes arisen from the concentration gradient and the temperature gradient are neglected. As summarized in Table A2, those components will be quite smaller

We appreciate the reviewer's comment. Following the suggestion, we have added another column to Table B2 representing mesospheric condition and generally updated the table to reflect extratropical conditions in TOMCAT better (see updated manuscript). The scaling of terms confirms that our approximation of neglecting the thermal diffusion term and mixing by molecular diffusion remains valid because changes in temperature and mixing ratio gradients between the regions are only moderate.

The Peclet number in the mesosphere is still much greater than 1 because not only molecular diffusion but also winds speeds increase with altitude (lines 467-468). This can be shown again by a back-of-the-envelope calculation. AoA is near-uniform in the mesosphere and we can conservatively assume a mesospheric turnover time of 0.5 years and height of 20 km. Comparing that to the molecular diffusivities of N_2 and Ar added to Table B2 implies a Peclet number of Pe ~500.

From manuscript:"In the mesosphere, AoA is near-uniform and we conservatively assume a turnover time of 0.5 years and height of 20 km. This implies a Peclet number of ~500 for the mesosphere."

Spatial and time derivatives of $\delta(Ar/N_2)$ are indeed not zero but even in the mesosphere remain considerably smaller than 1 when not given in per meg units. A mesospheric delta value of -5000 per meg implies deviation from a uniform mixing ratio of 0.5% which is still a good approximation.

Technical comments

We thank the reviewer for raising this question. START-08 was considered a pre-HIPPO test campaign. It was the first campaign in which the Medusa whole air sampler was used since being repackaged and partially automated following the COBRA and IDEAS campaigns, and the data quality from the campaign, as evident in the much larger scatter of vertical profiles, seems to be worse. During START-08, Medusa sampled from a relatively long polyethylene lined inlet tube, which may have contributed to adverse surface interactions. Subsequent advances in sampling have considerably improved data quality by

avoiding surface effects, artifactual fractionation, and analysis problems. Uncertainty and recent advances in $\delta(Ar/N_2)$ data quality are now discussed in more detail in a separate section of the methods (lines 87-99) and are also described in Bent (2014).

From manuscript: "Uncertainty in $\delta(Ar/N_2)$ observations arises from a combination of analytical limitations and artifactual fractionation during sampling. Replicate agreement of surface flasks shows a 1σ repeatability of ±6.1 per meg for $\delta(Ar/N_2)$ but additional scatter in the data may be introduced by small leaks in the Medusa system and thermal or pressure gradient fractionation at the sample intake (Morgan et al., 2020). For further details, see Bent (2014). These effects are challenging to separate from true atmospheric variability and differ between campaigns because sampling strategies have generally improved over time. To constrain uncertainty due to aircraft sampling, we quantify total variability in observations from the presumably fairly homogeneous tropical troposphere between 3-8 km (Fig. S1). Data from the earlier campaigns HIPPO 1–5 show a pooled standard deviation of 24 per meg whereas data from ATom 2–4 yield a pooled standard deviation of 9 per meg, illustrating advances in sampling and sample handling. While no ORCAS samples are available in the tropical troposphere, ORCAS data show similar scatter as ATom 2–4 data between 20°S and 50°S. For all campaigns, the uncertainty is small compared to the stratospheric signal of tens to hundreds per meg shown below. We also show that the stratospheric signal in $\delta(Ar/N_2)$ is not due to pressure-dependent inlet fractionation by evaluating the residual $\delta(Ar/N_2)$, which has been corrected for the natural gravitational signal (Fig. S2)."

(Fig. 1): Some symbols (rotated triangles?) do not seem to match those in the legend.

All symbols seem to match in our version of the figure. There are only 4 different 90-degree-rotated triangles pointing in each compass direction.

(Eq. 3): "λ" in the 2nd term of RHS should be $\lambda2$. $G(t'|(\Gamma, \lambda))$ should be $G(t'|(\Gamma1,\Gamma2,\lambda1,\lambda2))$, to be exact.
(L-152): "$\Gamma1, \Gamma1, \lambda1$ and $\lambda2$ are… " should be "$\Gamma1, \Gamma2, \lambda1$ and $\lambda2$ are …."
(L-179): "(ii)" should be "(iii)".
Thank you! We have corrected these oversights in the manuscript.

(Section 3.1): Parameters $\Gamma1$, $\Gamma2$, and A are shown in Fig. 3 and discussed in detail. But there is no description how the shape parameters $\lambda1$ and $\lambda2$ of the inverse-Gaussian distribution were as a result. It would be better to give us information about widths (Δ) or values of $\gamma(=\Delta2/\Gamma)$ of age spectra obtained by this method.

We thank the reviewer for pointing this out. The shape of the age spectrum is not well constrained by our method. This lack of constraint on the width of the age spectrum is now discussed in more detail in the text (see modifications to lines 236-254) and Figure S3 (see updated SI) was added to shows the pattern of all parameters fit by the algorithm.

From manuscript: "Unimodal age spectra are generally preferentially selected by the algorithm, in particular for the SH for young AoA (Fig. 3, panels (a)-(c)) but unimodal age spectra dominate the solution ensemble by a small margin except for AoA <0.5 years (Fig. S3). whereas bimodal spectra are slightly more common for older samples representing 50-80% of the solution ensemble in these bins. However, cConfidence intervals on age spectra parameters from bimodal spectra are considerably wider than for unimodal spectra. This implies that the parameters in a bimodal distribution are redundant and the shape of the age spectrum is not sufficiently constrained by the observations used. The width of the spectrum in each N_2O bin varies widely within the prescribed bounds for most unimodal and bimodal age spectra (Fig. S3). It appears that not enough data is available from the airborne campaigns to determine the amplitude of the seasonal cycle with enough confidence to constrain the width of the age spectrum and distinguish the relative contribution of the old peak (influencing only primarily the mean concentration difference to the troposphere) and the young peak (controlling the amplitude of seasonality) in bimodal age spectra. Because bimodal distributions are generated with a random value of the weighting factor A and the mean AoA of the second peak is assumed to be

~~old (5-7 years), randomly generated bimodal solutions often produce overall AoA that is quite old. Therefore, they are less likely to be selected by the MCMC algorithm for young AoA bins and a prevalent occurrence of unimodal age spectra in these bins is expected by chance. If more stratospheric data were available, the seasonality of CO₂ in each N₂O bin would be better resolved and the algorithm could derive tighter constraints on all parameters in bimodal age spectra, also allowing it to distinguish more clearly between unimodal or bimodal age spectra.~~ <u>Additional observations with less scatter or information from different age tracers are needed to properly resolve the shape of the age spectra.</u> Despite the limited resolution of the seasonal cycle, the observations are sufficient to place tight limits on the <u>mean</u> AoA in each N₂O bin and yield a well-characterised relationship between N₂O and <u>mean</u> AoA for each hemisphere."

(L-183 to Eq. 5): "....atmosphere δ_{GST} with a molecular mass 1 amu greater than that of air ..." GST seems to be a kind of virtual tracer. It may be better to mention how the molecular diffusion volume was defined for D_{GST}.

The molecular diffusion volume is important in estimating the molecular diffusivity of a gas according to eq. B5. We chose D_{GST} to be the diffusivity of Ar in air but the exact value of D_{GST} has no influence on our results because results are subsequently normalized by D_{GST} in equation 5. We clarified the phrasing of the manuscript (lines 212-218).

From manuscript: "To simplify the numerical treatment, we ~~only~~ simulate a<u>n idealized reference</u> ~~single~~ tracer of gravitational fractionation ~~in the atmosphere~~ δ_{GST} <u>, which has</u>~~with~~ a molecular mass 1 amu greater than ~~that of~~ air <u>and diffusivity equal that of Ar</u>~~which can be scaled offline~~ (see Appendix B & C)<u>. This tracer can be scaled offline to obtain the gravitational separation signal in any other species, including</u> $\delta(Ar/N_2)$<u>, e.g.</u> ~~using~~

$$\delta(Ar/N_2) \approx \frac{(M_{Ar} - M_{air}) \times D_{Ar} - (M_{N2} - M_{air}) \times D_{N2}}{(M_{GST} - M_{air}) \times D_{GST}} \times \delta_{GST}.$$

1

The appropriate diffusivity values D_{Ar} and D_{N2} for Ar and N₂ in air are derived in Appendix B for a ternary mixture of Ar, O₂, and N₂, extending previous work (Ishidoya et al., 2013; Belikov et al., 2019). <u>δ_{GST} is divided by D_{GST} in eq. 5 and therefore does not depend on the exact value chosen for D_{GST}.</u>"

(L-236 to 240): Again, with this AoA-GS plots alone, it is somewhat difficult to understand the difference between NH and SH. Vertical distributions of AoA and GS may be helpful.

Please see our response regarding model-data comparisons and data quality differences above.

(L-276 and Fig. S2): "...the small seasonal cycle amplitude of δ(Ar/N2) <6 per meg..." Certainly the average seasonal amplitude seems to be small, but we can see large scatters. Fluctuations of Ar/N2 ratio in short time will be partly smoothed by the mixing process during the upward transport from the tropical upper troposphere to the lower stratosphere via TTL. Thus, it may not be a big obstacle to AoA estimation from Ar/N2. Problem is, rather, that this scatter is real atmospheric signal or not.

We thank the reviewer for drawing our attention to this. The large scatter is present only in the HIPPO campaigns shown in Figure S2, and not in ATom 2-4, illustrating recent advances in sampling. We have updated the figure to better illustrate the campaign-dependence of precision. We have also revised the manuscript to discuss limitations in data quality of earlier campaigns (see response to comments above) and updated Fig 4 to represent these campaign differences.

(Eq. 7 in Appendix A): This number should be something like "(A1)". Also please check eq. numbers in Appendix B and C.

Thank you! We updated the equation and table numbering and referencing in all appendices.

We move the definition of ΔM_i to match its occurrence earlier in the text (lines 449-452).

From the manuscript: "Neglecting these smaller terms yields the governing Eq. (4) used in our model simulation

$$f_i \approx -ND_i\left[\left(\frac{1}{H_i} - \frac{1}{H_{air}}\right)C_i\right] = -ND_i\Delta M_i \frac{g}{RT}C_i \qquad \text{(B1)}$$

where we introduced $\Delta M_i \equiv M_i - M_{air}$ the molecular mass difference to air (kg mol^{-1}) as a short-hand. Equation (BA21) is equally valid for trace gases such as Ar and major gases N_2 and O_2 when the appropriate diffusivities given by Eqs. (BA12) and (BA14) are used."

(Eqs. B1, B2, and B3): Please check the signs (plus/minus) of the molecular diffusion terms in these equations. In the conservation equation, term of the flux divergence (3rd term RHS in eq. B1) should be a form like -∇·[f] if we put it on RHS, just the same as the eddy diffusion term. The 3rd and 4th terms of RHS in eq. B2 might have the same signs after using the chain rule. The 3rd term of RHS in eq. B3 might have the same sign with 4th term of RHS in eq. B2 after eliminating a small term.

We thank the reviewer for catching this typo. Indeed, the gravitational flux divergence term should be positive and of the same sign as the eddy diffusion term.

(L-418): "D" in the expression of Peclet number should be "De".

Yes, the manuscript was corrected.

(L-432): "Eq. (6)" should be Eq. (5).

Yes, the manuscript was corrected.

Comments from Reviewer 2

This paper uses measured and simulated ratios of Ar/N2 to deduce gravitational separation in the stratosphere and their relationship to the stratospheric age of air. The Ar/N2 relationships to AoA in this study are relatively compact compared to the previous Belikov et al. study on gravitational separation and the agreement between the observations and model relationships are improved. This suggests that gravitational separation of Ar/N2, and also potentially the ratios of heavier noble gases, could be useful as an additional diagnostic of transport. The measurements, model runs and methodology are well described and the improvement in the treatment of gravitational separation in both the measurements and model compared to previous studies is an important step forward.

My main concern with this study is in the calculation of age of air from the measurements. The technique used, although rigorous, is based on assumptions that are not consistent with our current understanding

of the shape of age spectra and the transport of air into the lowermost stratosphere. The technique and assumptions are based on a study that is nearly two decades old and there have been numerous studies since that time refining our understanding of age spectra and how they can be inferred from measurements. The results may not change substantially by using more realistic age spectra and transport assumptions but that isn't entirely clear. Regardless, the more recent studies on this topic and the newer techniques should be acknowledged and shown to be in agreement with, or ideally replace, the results from the technique used in this study. Specific references and more detail on this topic are included below.

The results of this study are significant and relevant to ACP so I recommend publication after modification of the age of air calculation and consideration of the specific comments below.

We thank reviewer 2 for a favorable review and insightful comments on our manuscript. We made changes in how we present and treat the calculation of AoA and will discuss these in more detail in response to the specific comments below.

Specific comments

Line 35: should add 'however' before 'observational'

Thank you, we changed the manuscript accordingly.

Lines 40-41: should add references here to studies that have used tracers to infer age spectra (Andrews et al., JGR, 1999; Schoeberl et al., JGR, 2005; Hauck et al., ACP, 2019; Podglajen and Ploeger, ACP, 2019).

Thank you for the suggested references. We added them to the manuscript (lines 168-173).

From manuscript:"Though more complex representations of the age spectrum have recently been proposed, accounting for multiple modes and seasonal variations (e.g., Bönisch et al., 2009; Li et al., 2012; Hauck et al., 2019, 2020; Podglajen and Ploeger, 2019), these require additional information from multiple tracers or models, and good seasonal data coverage that is not available here. In any case, we only rely on the first moment of the age distribution (i.e., the mean age) for calibrating the N_2O-AoA relationship, and the mean age is not particularly sensitive to the assumed shape of the age spectrum (Andrews et al., 2001)."

Line 94: did you mean 'CO2' instead of 'CO' or should 'CO2' be added here?

We used CO2 and CO from the Picarro and QCLS. The list was corrected (line 108).

Line 97: 'sample flask i to sample flask i+1'

Thank you, we changed the manuscript accordingly.

Line 108: change 'as a combined result of' to 'due to'

Thank you, we changed the manuscript accordingly.

Line 113: 'used' instead of 'use'

Thank you, we changed the manuscript accordingly.

Section 2.2: This section is my main concern with the paper as I mentioned above. I have two issues: (1) the assumed shape of the age spectra and (2) the tropical single entry point assumption for the lowermost stratosphere. The first issue, the shape of stratospheric age spectra has been shown in a number of studies derived from various model and reanalysis output based on trajectories or boundary impulse functions (Reithmeier et al., Clim. Dyn., 2008; Li et al., JGR, 2012; Diallo et al., ACP, 2012; Ray et al., JGR, 2014; Ploeger and Birner, ACP, 2016; Podglajen and Ploeger, ACP, 2019; Hauck et al., ACP, 2019). In these studies, it is clear that the dominant mode of variability in the age spectra in the lower stratosphere is the seasonal cycle, and for those models with a QBO, that is the second largest mode of variability. None of these results show a peak in the 5-7 year range relative to the 2-5 year range of the spectra. The use of two peak age spectra with one of the peaks in the 5-7 year range is not realistic for any part of the stratosphere and especially not in the lower stratosphere where the seasonal cycle is so large and the 5-7 year portion of the age spectra has much smaller values (often by an order of magnitude) compared to those at ages less than 5 years. The lack of reference in the manuscript to any of the papers listed above and the use of an age spectra shape that is not consistent with any of them is a clear deficiency. The use of a single trace gas, such as CO2, can only reveal a limited amount of the age spectra. It has been shown by Hauck et al. (2019) that at least 5-10 tracers with a range of lifetimes are necessary to resolve the important features and the true shape of the age spectra. The Andrews et al. (1999, 2001) studies were important steps at the time in our understanding of how trace gas observations could be used to infer features of the age spectra. But now that we have a detailed understanding of age spectra from models and reanalysis it isn't appropriate to use age spectra defined only by the fit to CO2. The second issue is related to the transport of air into the lowermost stratosphere (LMS) where most of the measurements in this study were taken. Many studies have shown that the LMS consists of a seasonally varying mixture of tropospheric and stratospheric air with some portion of the tropospheric air originating from the extratropics or tropical air that has bypassed the tropical tropopause (e.g. Ray et al., JGR, 1999; Hoor et al., ACP, 2004; Olsen et al., JGR, 2004; Boenisch et al., ACP, 2009; Skerlak et al., ACP, 2014). This precludes the use of a single-entry tropical tropopause time series to derive age of air in the LMS. I highly recommend a technique considering the above points be used here. Even if the resulting mean age values do not differ significantly, the technique would be consistent with our current understanding of transport in the lower stratosphere.

We thank Reviewer 2 for raising this important concern.

Our selection criteria for "stratospheric" samples ensures that we focus on air that is of stratospheric origin and has experienced limited mixing with the troposphere. Therefore, we believe the assumption of a single entry-point to the stratosphere when calculating AoA is imperfect but justified. Other methods (e.g., Hauck et al, 2020), rely on model results to constrain the contribution of different tropospheric source regions to air found in the lowermost stratosphere. However, these model estimates will be representative of air *typically* found in a given region, not necessarily of air in small-scale features that we specifically target here for its stratospheric signature. This is evident also in a comparison of observed AoA and AoA simulated by TOMCAT. For the oldest samples, observed AoA is typically higher than simulated AoA (Fig. S4). This key feature of our sample selection is discussed in more detail in the manuscript now (lines 141-148) and we have lowered the H_2O threshold value to 15 ppm in order to more carefully select for stratospheric composition. We now also emphasize in the text that our sampling strategy does not yield air that is fully representative of the lowermost stratosphere (lines 146-147 & 266-269).

Furthermore, the fact that our samples are typically older means that diabatic dispersion (Sparling et al., doi.org/10.1029/97JD01968, 1997) has acted on the age distribution and likely smoothed out the multiple peaks found in model studies.

Only a limited number of tracers are available across all campaigns and incorporating multiple new tracers into the method is beyond the scope of this manuscript. Therefore, constraining the full age spectrum is impossible as discussed by Reviewer 2. We adjusted the language throughout the manuscript to reflect that our method only really constrains the mean AoA of a sample and included additional references in the manuscript to studies that illustrate a more thorough treatment of the shape of the age spectrum. We also adjusted the range of possible mean AoA values of the second peak to range from 1-6 years. This and changing the H_2O threshold does not substantially change our results.

Most importantly, our objective - establishing a relationship between $\delta(Ar/N_2)$ and mean AoA – is currently not limited by our ability to determine the AoA for these samples. AoA from N_2O is still relatively more precise than our ability to sample and analyze air for $\delta(Ar/N_2)$. Moreover, our recalibration of the N_2O-AoA relationship which underlies our AoA calculations for Medusa samples is not overly sensitive to small errors introduced at young AoA where the risk of a strong tropospheric influence is greatest because the N_2O-AoA relationship is anchored at an age of zero to the mean concentration N_2O.

From manuscript (lines 141-148): "We consider the tropical upper troposphere as the single entry-point of air into the stratosphere for our AoA-calculation. Mixing between tropospheric and lower stratospheric air can occur outside the tropics leading to air masses that are transitional between stratospheric and tropospheric in the lowermost extratropical stratosphere (e.g., Škerlak et al., 2014). This additional pathway for tropospheric entry has been accounted for in previous studies via additional model information and/or multiple tracers (Ray et al., 1999; Hoor et al., 2004; Bönisch et al., 2009; Hauck et al., 2020). However, we are limited by the number of tracers available to us, and therefore instead aim to largely exclude samples influenced through this additional pathway by using threshold in N_2O, O_3, and H_2O when identifying stratospheric samples. Because of this sample selection strategy, we note that the ages calculated here should not be considered representative for the region that air was sampled in."

From manuscript (lines 168-173): "Though more complex representations of the age spectrum have recently been proposed, accounting for multiple modes and seasonal variations (e.g., Bönisch et al., 2009; Li et al., 2012; Hauck et al., 2019, 2020; Podglajen and Ploeger, 2019), these require additional information from multiple tracers or models, and good seasonal data coverage that is not available here. In any case, we only rely on the first moment of the age distribution (i.e., the mean age) for calibrating the N_2O-AoA relationship, and the mean age is not particularly sensitive to the assumed shape of the age spectrum (Andrews et al., 2001)."

From manuscript (lines 236-254): "Unimodal age spectra are generally preferentially selected by the algorithm, in particular for the SH for young AoA (Fig. 3, panels (a)-(c)) but unimodal age spectra dominate the solution ensemble by a small margin except for AoA <0.5 years (Fig. S3). whereas bimodal spectra are slightly more common for older samples representing 50-80% of the solution ensemble in these bins. However, cConfidence intervals on age spectra parameters from bimodal spectra are considerably wider than for unimodal spectra. This implies that the parameters in a bimodal distribution are redundant and the shape of the age spectrum is not sufficiently constrained by the observations used. The width of the spectrum in each N_2O bin varies widely within the prescribed bounds for most unimodal and bimodal age spectra (Fig. S3). It appears that not enough data is available from the airborne campaigns to determine the amplitude of the seasonal cycle with enough confidence to constrain the width of the age spectrum and distinguish the relative contribution of the old peak (influencing only primarily the mean concentration difference

to the troposphere) and the young peak (controlling the amplitude of seasonality) in bimodal age spectra. ~~Because bimodal distributions are generated with a random value of the weighting factor *A* and the mean AoA of the second peak is assumed to be old (5-7 years), randomly generated bimodal solutions often produce overall AoA that is quite old. Therefore, they are less likely to be selected by the MCMC algorithm for young AoA bins and a prevalent occurrence of unimodal age spectra in these bins is expected by chance. If more stratospheric data were available, the seasonality of CO_2 in each N_2O bin would be better resolved and the algorithm could derive tighter constraints on all parameters in bimodal age spectra, also allowing it to distinguish more clearly between unimodal or bimodal age spectra.~~ Additional observations with less scatter or information from different age tracers are needed to properly resolve the shape of the age spectra. Despite the limited resolution of the seasonal cycle, the observations are sufficient to place tight limits on the mean AoA in each N_2O bin and yield a well-characterised relationship between N_2O and mean AoA for each hemisphere."

From manuscript (lines 266-170): "A comparison of $\delta(Ar/N_2)$ and AoA between flask samples and TOMCAT is shown in Figure S4. However, the direct comparison should not be considered a good test of model performance because our stratospheric composition criteria intentionally target unusually old air in the lowermost stratosphere. For the same reason, profiles of $\delta(Ar/N_2)$ and AoA are not evaluated here. Instead, we suggest the AoA-GS relationship provides a more robust metric for comparison. Overall, our observations validate the implementation of GS for young (< 3 y) ages in TOMCAT."

Line 152: Gamma1 is repeated, the second one should be Gamma2.

Thank you, we corrected the typo.

Line 179: the second '(ii)' should be '(iii)'

Thank you, we corrected the typo.

Line 203: 'relationships'

Thank you, we corrected the typo.

Line 262: 'descends' instead of 'is sinking'

We adopted the suggested language.

Line 328: Is the Ar/N2 ratio 'mostly unaffected by seasonality' throughout the stratosphere?

This statement refers to the tropical upper troposphere. We clarified the phrasing to reflect this (line 371).

Lines 330-331: It's not so much that changes in AoA in the upper stratosphere and mesosphere are difficult to resolve from transient tracers but more that AoA becomes nearly constant in this region as you show in Figure 6. So AoA loses sensitivity in this region where gravitational separation gains sensitivity. This is a difficult region to make measurements of any kind so it may also be challenging to obtain enough measurements of noble gases to monitor gravitational separation.

Thank you. We adjusted the phrasing to reflect this distinction more clearly (lines 373-375). Indeed, sampling of the mesosphere is very challenging but we are hopeful that more careful noble gas

observations will become available in the next years from balloon samplers that can reach up to 30-35 km providing additional constraints on circulation in the lower and middle stratosphere.

From manuscript (lines 369-375): "Combining $\delta(Ar/N_2)$ with other tracers of circulation could lead to new insights into atmospheric mixing and transport. $\delta(Ar/N_2)$ has potential advantages over existing approaches based on transient tracers such as CO_2 or N_2O since $\delta(Ar/N_2)$ is only influenced by the physics of transport and mostly unaffected by seasonality in the tropical upper troposphere. Furthermore, because of its large gradients at high altitudes, $\delta(Ar/N_2)$ observations from the upper stratosphere and mesosphere could improve our understanding of circulation on seasonal and interannual timescales in a region where ~~changes in~~ AoA ~~from transient tracers are otherwise difficult to resolve~~ becomes near-uniform and some photochemical age tracers such as N_2O are almost fully depleted and thus lose sensitivity to circulation changes."

Gravitational separation of Ar/N₂ and age of air in the lowermost stratosphere in airborne observations and a chemical transport model

Benjamin Birner[1], Martyn P. Chipperfield[2,3], Eric J. Morgan[1], Britton B. Stephens[4], Marianna Linz[5], Wuhu Feng[2,6], Chris Wilson[2,3], Jonathan D. Bent[1,7], Steven C. Wofsy[5], Jeffrey Severinghaus[1], Ralph F. Keeling[1]

[1]Scripps Institution of Oceanography, UC San Diego, La Jolla, CA 92093, USA
[2]School of Earth and Environment, University of Leeds, Leeds, LS2 9JT, UK
[3]National Centre for Earth Observation, University of Leeds, Leeds, LS2 9JT, UK
[4]National Center for Atmospheric Research, Boulder, CO 80301, USA
[5]Department of Earth and Planetary Sciences, and School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA
[6]National Centre for Atmospheric Science, University of Leeds, Leeds, LS2 9JT, UK
[7]now at: Picarro, Inc., Santa Clara, CA 95054, USA

Correspondence to: Benjamin Birner (bbirner@ucsd.edu)

Abstract. Accurate simulation of atmospheric circulation, particularly in the lower stratosphere, is challenging due to unresolved wave-mean flow interactions and limited high-resolution observations for validation. Gravity-induced pressure gradients lead to a small but measurable separation of heavy and light gases by molecular diffusion in the stratosphere. Because the relative abundance of Ar to N_2 is exclusively controlled by physical transport, the argon-to-nitrogen ratio (Ar/N₂) provides an additional constraint on circulation and the age of air (AoA), i.e. the time elapsed since entry of an air parcel into the stratosphere. Here we use airborne measurements of N_2O and Ar/N₂ from nine campaigns with global coverage spanning 2008–2018 to calculate AoA and to quantify gravitational separation in the lowermost stratosphere. To this end, we develop a new N_2O-AoA relationship using a Markov Chain Monte Carlo algorithm. We observe that gravitational separation increases systematically with increasing AoA for samples with AoA between 0 to 3 years. These observations are compared to a simulation of the TOMCAT/SLIMCAT 3-D chemical transport model, which has been updated to include gravitational fractionation of gases. We demonstrate that although AoA at old ages is slightly underestimated in the model, the relationship between Ar/N₂ and AoA is robust and agrees with the observations. This highlights the potential of Ar/N₂ to become a new AoA tracer that is subject only to physical transport phenomena and can supplement the suite of available AoA indicators.

1. Introduction

Transport in the middle atmosphere is driven by a combination of advection by the Brewer-Dobson circulation (BDC) (Brewer, 1949; Dobson, 1956) and quasi-horizontal, two-way mixing by breaking waves (Holton et al., 1995). Models consistently

predict an acceleration of the BDC due to climate change (Butchart, 2014) but subgrid-scale mixing processes and momentum transfer by unresolved buoyancy waves limit our ability to accurately simulate circulation in the stratosphere (Haynes, 2005; Plumb, 2007). An acceleration of the BDC has important repercussions for stratosphere-troposphere exchange (STE), and thus

35 recovery of the ozone layer and the greenhouse effect of stratospheric water vapor, however observational evidence for an acceleration of the BDC is weak (Engel et al., 2009, 2017; Waugh, 2009; Ray et al., 2010, 2014).

The mean age of air (AoA) is a widely used indicator of stratospheric circulation (Hall and Plumb, 1994; Waugh and Hall, 2002; Linz et al., 2016). Air can be transported to any location r in the stratosphere via a myriad of different paths, and each path will have an associated transit time. The probability density function that describes the likelihood for air to reach location

40 r with a specific transit time is called the age spectrum. Although the age spectrum is not directly observable, some aspects of its shape can be inferred from observations of short- and long-lived tracers (Andrews et al., 1999, 2001; Schoeberl et al., 2005; Hauck et al., 2019, 2020; Podglajen and Ploeger, 2019). For tracers that are conserved in the stratosphere and whose concentrations increase approximately linearly with time in the troposphere, such as SF_6 and CO_2, the mean AoA, i.e., the first moment of the distribution, can simply be calculated as the time difference, or "lag time", to when tracer concentrations in the

45 upper troposphere last had comparable values as measured in a stratospheric sample (Hall and Plumb, 1994; Boering et al., 1996; Waugh and Hall, 2002). The stratospheric concentration of N_2O has also been calibrated as an independent AoA tracer by relating the gradual photolytic loss of N_2O in the stratosphere to AoA as determined from CO_2 (Boering et al., 1996; Andrews et al., 2001; Linz et al., 2017).

In contrast to early measurements made on rocket samples (Bieri and Koide, 1970), Ishidoya et al. (2008, 2013) have shown

50 using a balloon-borne sampling system that stratospheric air is detectably fractionated by gravitational settling (GS), with the degree of fractionation strongly correlated to AoA (Ishidoya et al., 2008, 2013, 2018; Sugawara et al., 2018; Belikov et al., 2019). GS leads to depletion of heavier gases in the stratosphere yielding lower ratios of heavy to light gases, e.g. Ar/N_2, $^{18}O/^{16}O$ of O_2 and $^{15}N/^{14}N$ of N_2, with increasing elevation. The vertical gradients induced by gravimetric separation are steeper at higher latitudes (Ishidoya et al., 2008, 2013, 2018; Sugawara et al., 2018), and consistent with patterns observed in

55 stratospheric AoA (Sugawara et al., 2018; Belikov et al., 2019). Gravimetric settling in the stratosphere has also been simulated in 1-D (Keeling, 1988; Ishidoya et al., 2008), 2-D (Ishidoya et al., 2013, 2018; Sugawara et al., 2018), and 3-D stratospheric models (Belikov et al., 2019). 2-D and 3-D models both show a pattern of GS which increases with altitude and latitude, similar to the patterns observed in tracers with a significant stratospheric sink such as N_2O, and consistent with a positive correlation with AoA (Ishidoya et al., 2013; Sugawara et al., 2018; Belikov et al., 2019). Here we attempt to calibrate GS of Ar/N_2 as an

60 AoA tracer similar to previous work on the N_2O-AoA relationship.

Observing GS in the stratosphere is challenging, however, as the signals are small and because of the need to avoid artifacts caused by temperature- and pressure-induced fractionation near the sampling inlet (Blaine et al., 2006; Ishidoya et al., 2008,

2013). The resulting scatter in existing balloon-based measurements precludes a clear evaluation of the relationship between GS and AoA (Belikov et al., 2019).

65　Here we present a dataset of gravitational fractionation of Ar/N_2 and AoA observations made on flask samples from three airborne research projects, from 9 campaigns in the lowermost stratosphere with mean AoA <3 years (Wofsy et al., 2012, 2018; Stephens, 2017). The HIAPER Pole-to-Pole Observations (HIPPO) project was a global survey of the Pacific troposphere to lower stratosphere on the NSF/NCAR Gulfstream-V aircraft (Wofsy et al., 2011), composed of 5 individual campaigns from 2008–2011. The O_2/N_2 Ratio and CO_2 Airborne Southern Ocean (ORCAS) study was conducted using the

70　same aircraft but focused on the Drake Passage and Antarctic Peninsula during Jan–Feb 2016 (Stephens et al., 2018). The Atmospheric Tomography (ATom) project was a survey of the troposphere and lower stratosphere of the Pacific and Atlantic Ocean basins on the NASA DC-8 aircraft, composed of 4 individual campaigns from 2016–2018 (Prather et al., 2017). The observations are compared to new simulations of GS with the TOMCAT/SLIMCAT (Chipperfield, 2006) tracer transport model. Our goals are to demonstrate the consistency of our data with gravitational fractionation, to evaluate model

75　performance, and to highlight the potential of Ar/N_2 as a new age tracer.

2. Methods

2.1. Measurements

Discrete 1.5L flask samples were taken with the NCAR/Scripps Medusa Whole Air Sampler (Bent, 2014) (https://www.eol.ucar.edu/instruments/ncar-scripps-airborne-flask-sampler). Medusa holds 32 borosilicate glass flasks sealed

80　with Viton o-rings and uses active pressure control to fill the flasks with cryogenically dried air to ~760 torr. Flasks are shipped to Scripps Institution of Oceanography (SIO) for analysis of Ar/N_2 ratios on an IsoPrime Mass Spectrometer. We report changes in Ar/N_2 ratios in delta notation

$$\delta = \left(\frac{R_{SA}}{R_{REF}} - 1\right) \times 10^6 (per\ meg) \qquad\qquad 1$$

where R_{SA} is the mixing or isotope ratio in the sample and R_{REF} the ratio in a reference mixture. ~~Replicate agreement shows a~~ ~~1σ repeatability of ±6.1 per meg for δ(Ar/N₂). For further details, see Bent (2014).~~ Measurements are made on the Scripps

85　O2 Program Argon Scale, as defined on 21 Jan 2020. $\delta(Ar/N_2)$ values are reported after applying an offset to the data to yield a mean of zero in the ~~equatorial~~ tropical airborne observations of the free troposphere between ~~2~~3-~~10~~8 km.

Uncertainty in $\delta(Ar/N_2)$ observations arises from a combination of analytical limitations and artifactual fractionation during sampling. Replicate agreement of surface flasks shows a 1σ repeatability of ±6.1 per meg for $\delta(Ar/N_2)$ but additional scatter in the data may be introduced by small leaks in the Medusa system and thermal or pressure gradient fractionation at the sample

90　intake (Morgan et al., 2020). For further details, see Bent (2014). These effects are challenging to separate from true

3

atmospheric variability and differ between campaigns because sampling strategies have generally improved over time. To constrain uncertainty due to aircraft sampling, we quantify total variability in observations from the presumably fairly homogeneous tropical troposphere between 3-8 km (Fig. S1). Data from the earlier campaigns HIPPO 1–5 show a pooled standard deviation of 24 per meg whereas data from ATom 2–4 yield a pooled standard deviation of 9 per meg, illustrating advances in sampling and sample handling. While no ORCAS samples are available in the tropical troposphere, ORCAS data show similar scatter as ATom 2–4 data between 20°S and 50°S. For all campaigns, the uncertainty is small compared to the stratospheric signal of tens to hundreds per meg shown below. We also show that the stratospheric signal in $\delta(Ar/N_2)$ is not due to pressure-dependent inlet fractionation by evaluating the residual $\delta(Ar/N_2)$, which has been corrected for the natural gravitational signal (Fig. S2).

We have compiled available simultaneous, high frequency measurements of a range of other trace gases including N_2O, CO_2, O_3, CH_4, and CO to identify Medusa samples with stratospheric influence and calculate mean AoA. N_2O was measured continuously with a precision of 0.09 ppb at 1 Hz frequency using the Harvard Quantum Cascade Laser Spectrometer (QCLS) (Santoni et al., 2014) during HIPPO and ORCAS, and measured every 1–3 min using the NOAA gas chromatograph PAN and Trace Hydrohalocarbon ExpeRiment (PANTHER) during ATom. The Unmanned Aircraft Systems Chromatograph for Atmospheric Trace Species (UCATS) was used to measure O_3 during HIPPO and ATom. O_3 was not measured on ORCAS. We use continuous H_2O data from the NCAR open path near-infrared multi-pass spectrometer Vertical Cavity Surface Emitting Laser (VCSEL) (Zondlo et al., 2010) for HIPPO and ORCAS whereas H_2O was measured using the NASA Diode Laser Hygrometer (DLH) (Diskin et al., 2002) during ATom flights. CH_4, CO_2, and CO were measured by QCLS during HIPPO and ORCAS, and by the NOAA Picarro (Karion et al., 2013) during ATom. An averaging kernel is applied to the continuous and semi-continuous aircraft data, such as N_2O, O_3, and H_2O, to match it to Medusa samples. The kernel multiplies a weighting function $w_i(t)$ with all continuous data before time t_s when Medusa switched from sample flask i to the next sample flask $i+1$. $w_i(t)$ for each sample i is given by

$$w_i(t) = \exp\left[\frac{-t_s - t}{\tau}\right], \qquad \qquad 2$$

where t is each 1s-increment of the continuous data and $\tau = \frac{V}{Q}$ is the flushing time of air in a Medusa flask determined by the flask volume V and airflow Q.

Stratospheric samples are identified based on their N_2O, O_3, and water vapor levels. Classification based on chemical composition rather than potential temperature or altitude effectively selects samples with a clear stratospheric signature in the lowermost stratosphere and excludes air which has experienced substantial mixing with tropospheric air masses, where mixing with the upper troposphere could be substantial, and minimizes the impact of synoptic-scale variability. We label samples as "stratospheric" if (i) water vapor levels are below 50 15 ppm and either (ii.a) O_3 values exceed 140 ppb or (ii.b) N_2O (detrended to a reference year of 2009) is below 315 ppb. These criteria yield 235 234 lower stratospheric samples with high quality N_2O

and δ(Ar/N$_2$) data, spanning a wide range of latitudes poleward of 40° in both the Northern and Southern Hemisphere (Fig. 1). We use Medusa samples from all 5 HIPPO campaigns, ORCAS, and ATom 2–4. We do not use samples from ATom 1 because inlet fractionation ~~as a combined result of~~due to the unique inlet design and location on the DC-8 on this campaign introduced apparent biases on the order of 30 per meg. An additional ~~14~~25 stratospheric samples are available from the START-08

125   campaign on the NSF/NCAR GV, but we have not used these here because of the limited number of flasks collected and greater scatter in the data~~the δ(Ar/N$_2$) data quality is considerably worse~~.

2.2. AoA calculation

Stratospheric (mean) AoA for Medusa samples is calculated from N$_2$O using an updated hemisphere-specific N$_2$O-AoA relationship. Our method broadly follows Andrews et al. (2001) who assumed a bimodal age spectrum and use~~d~~ multiple

130   observations of CO$_2$ binned by N$_2$O values to resolve the seasonal cycle of CO$_2$ in each bin. Properties of the age spectrum for each N$_2$O bin, including mean AoA were constrained by optimizing the agreement between observed CO$_2$ concentrations and concentration implied by randomly generated age spectra in each N$_2$O bin. Andrews et al. (2001) used a highly efficient "genetic algorithm" to yield the most likely relationship between mean AoA and N$_2$O in each bin. In contrast, we use a more computationally costly Markov chain Monte Carlo (MCMC) method that allows us to obtain more robust uncertainties for all

135   estimated parameters of the age spectrum. Following Malinverno (2002) and Green (1995), our algorithm builds on a Metropolis-Hasting sampler (Metropolis et al., 1953; Hastings, 1970) to evaluate probability distributions for each age spectrum parameter and automatically choses whether a unimodal or bimodal representation of the age spectrum is more appropriate in each N$_2$O bin. Finally, we constructed a new tropical upper troposphere reference time series for CO$_2$ in this study to ensure maximum consistency between all observations used. Analytical and methodological uncertainties are

140   propagated thoroughly and reported as the 95% confidence interval around a mean.

We consider the tropical upper troposphere as the single entry-point of air into the stratosphere for our AoA-calculation. Mixing between tropospheric and lower stratospheric air can occur outside the tropics leading to air masses that are transitional between stratospheric and tropospheric in the lowermost extratropical stratosphere (e.g., Škerlak et al., 2014). This additional pathway for tropospheric entry has been accounted for in previous studies via additional model information and/or multiple

145   tracers (Ray et al., 1999; Hoor et al., 2004; Bönisch et al., 2009; Hauck et al., 2020). However, we are limited by the number of tracers available to us, and therefore instead aim to largely exclude samples influenced through this additional pathway by using threshold in N$_2$O, O$_3$, and H$_2$O when identifying stratospheric samples. Because of this sample selection strategy, we note that the ages calculated here should not be considered representative for the region that air was sampled in.

Our upper troposphere reference time series (TRTS) consists of a long-term trend and a representation of the mean seasonal

150   cycle. Because direct CO$_2$ observations in the region are limited, the long-term trend and a first guess of the mean seasonal cycle are estimated from monthly mean surface observations at the Mauna Loa station in Hawaii (MLO, 19°N, 155°W, Fig 2, panel (a)) (Keeling et al., 2001) which are later adjusted to match airborne observations in the tropics (20°S-20°N) above 8

km (Fig. 2, panel (c)). The adjustment includes (i) a constant offset, (ii) a reduced amplitude of the seasonal cycle and (iii) a phase-lag of one month. As shown in Fig. 2, after these adjustments the TRTS matches the mean seasonal cycle and absolute value of the airborne data, thus accounting for known vertical gradients of CO_2 and reduced seasonality in the upper troposphere (Fig. 2, panel (d)). The amplitude of the seasonal cycle in our time series is also in good agreement with the boundary condition used by Andrews et al. (1999 and 2001).

The new TRTS is used to estimate the age spectrum in 13 N_2O bins of 5 or 10 ppb width (320-325 ppb,…, 290-295 ppb, 280-290,…, 230-240 ppb) using a Markov Chain Monte Carlo (MCMC) algorithm which compares observed CO_2 concentrations in the bin to concentration expected from the TRTS. To maximize data availability, we use high frequency data (see "2.1. Measurements") identified as stratospheric according to the criteria above from the 10 s merged products available for HIPPO, ORCAS, and ATom rather than data averaged to the lower Medusa sampling interval for this calibration exercise. Small corrections are applied to the observed CO_2 concentrations (<0.2 ppm) to account for oxidation of CH_4 and CO in the stratosphere.

The MCMC algorithm considers random noise in the TRTS and uses a unimodal or bimodal inverse Gaussian shape of the age spectrum characterized by mean ages Γ_1 and Γ_2 and shape parameters λ_1 and λ_2

$$G\big(t'|(\Gamma_1, \Gamma_2, \lambda_1, \lambda_2)\big)$$

$$= A\sqrt{\frac{\lambda_1}{2\pi t'^3}}\,\exp\left(-\frac{\lambda_1\,(t'-\Gamma_1)^2}{2\Gamma_1^2 t'}\right) + (1-A)\sqrt{\frac{\lambda_2}{2\pi t'^3}}\,\exp\left(-\frac{\lambda_2\,(t'-\Gamma_2)^2}{2\Gamma_2^2 t'}\right), \qquad 3$$

where factor A determines the relative weight of each peak and a value of $A = 1$ yields a unimodal spectrum (Hall and Plumb, 1994; Andrews et al., 2001; Bönisch et al., 2009). Though more complex representations of the age spectrum have recently been proposed, accounting for multiple modes and seasonal variations (e.g., Bönisch et al., 2009; Li et al., 2012; Hauck et al., 2019, 2020; Podglajen and Ploeger, 2019), these require additional information from multiple tracers or models, and good seasonal data coverage that is not available here. In any case, we only rely on the first moment of the age distribution (i.e., the mean age) for calibrating the N_2O-AoA relationship, and the mean age is not particularly sensitive to the assumed shape of the age spectrum (Andrews et al., 2001). By setting $A = 1$ for 50% of all tested age spectra, the MCMC algorithm automatically selects whether a unimodal or bimodal representation of the age spectrum is optimal to match observations. Although bimodal solutions with five instead of two free parameters will always be able to fit the data better, a larger number of parameters also decreases the likelihood of randomly selecting a combination of parameters that match the observations well because the fraction of the total parameter space region which yields good agreement with the observation decreases as more parameters are added (Malinverno, 2002). The algorithm is thus able to make an appropriate choice between unimodal and bimodal distributions from the data itself, without any further a priori assumptions. The MCMC algorithm is set up separately with 2000 independent chains for each N_2O bin to account for uncertainty in the TRTS and obtain best estimates of the mean AoA,

i.e., the first moment of Eq. 3. To simplify the algorithm, possible values of $\Gamma_1, \Gamma_1, \lambda_1$ and λ_2 are repeatedly sampled from the same N₂O bin-specific prior distributions. Details of the algorithm are presented in Appendix A.

Finally, the resulting relationship between mean AoA and the N₂O concentration of each bin is fit separately for each hemisphere by a quadratic polynomial and the polynomial is evaluated at the N₂O value of each Medusa sample to pair every

185   observation of δ(Ar/N₂) with an AoA. Uncertainty in N₂O and the polynomial fits are propagated by a Monte Carlo scheme. Overall, our method estimates the most likely mean AoA for each Medusa sample and improves upon previous methods by providing a ~~thorough~~ framework for the treatment of uncertainty resulting from (i) analytical error, (ii) uncertainty in the shape of the age spectrum, and (iii) uncertainty in the composition of source gas introduced into the stratosphere

2.3. TOMCAT/SLIMCAT model

190   TOMCAT/SLIMCAT (hereafter TOMCAT) is an offline 3-D chemical tracer transport model that has been used extensively for studies of stratospheric ozone depletion and circulation (e.g., Chipperfield, 2006; Chipperfield et al., 2017; Krol et al., 2018). For this study, TOMCAT was run over 31 years, from 1988 to 2018, with a timestep of 30 minutes at 2.8°×2.8° horizontal resolution forced by the ERA-Interim reanalysis (Dee et al., 2011) at 60 vertical hybrid sigma- pressure (σ–p) levels up to ~60–65 km. The first 20 years (i.e. before the first flask observation) were treated as spin-up. The TOMCAT AoA tracer

195   is initialized at the surface and corrected to a value of AoA = 0 just below the tropical tropopause. Vertical motion was calculated from the divergence of the horizontal mass fluxes. Although this approach gives slightly younger stratospheric AoA than using isentropic levels and radiative heating rates, it allows a more detailed treatment of tropospheric transport (Chipperfield, 2006; Monge-Sanz et al., 2007). The model simulation was sampled at the times and locations of the Medusa flask observations to provide a direct comparison between the measurements and model.

200   Following the methodology of Belikov et al. (2019), we include an additional vertical flux term in the model representing the GS of gases in the atmosphere. The vertical flux f_i (mol m⁻² s⁻¹) of tracer i due to molecular diffusion in Earth's gravitational field is given as (Banks and Kockarts, 1973; Belikov et al., 2019)

$$ f_i = -D_i N C_i \left[\frac{1}{C_i} \frac{\partial C_i}{\partial z} + \left(\frac{1}{H_i} - \frac{1}{H_{air}} \right) + \alpha_i \frac{1}{T} \frac{\partial T}{\partial z} \right], \qquad\qquad 4 $$

where D_i is the tracer-specific binary molecular diffusivity in air (m² s⁻¹), N is the number density of air (mol m⁻³), C_i is the mixing ratio, $H_i = \frac{RT}{gM_i}$ is the tracer-specific atmospheric equilibrium scale height (m), R is the fundamental gas constant (J K⁻

205   ¹ mol⁻¹), g is the gravitational constant (m s⁻²), M_i is the tracer-specific atomic or molecular mass (kg mol⁻¹), α_i is the tracer-specific thermal diffusivity (m² s⁻¹), and T is temperature (K). The three terms in Eq. (4) represent molecular diffusion driven by (i) vertical gradients in the mole fraction of i, (ii) pressure gradients caused by gravity and described by the barometric law, and (iii) temperature gradients (left to right). We neglect the first and third terms in the brackets in Eq. (4), leaving only the

gravitational settling term $\left(\frac{1}{H_i} - \frac{1}{H_{air}}\right)$ on the basis that both terms are orders of magnitude less important than the gravitational

210   separation term under stratospheric conditions (see Appendix B) (Ishidoya et al., 2013; Belikov et al., 2019). No fluxes are allowed through the top boundary, and C_i is held constant at the Earth's surface.

To simplify the numerical treatment, we ~~only~~ simulate a~~n~~ idealized reference ~~single~~ tracer of gravitational fractionation ~~in the atmosphere δ_{GST}~~, which has~~with~~ a molecular mass 1 amu greater than ~~that of~~ air and diffusivity equal that of Ar~~which can be scaled offline~~ (see Appendix B & C). This tracer can be scaled offline to obtain the gravitational separation signal in any other

215   species, including $\delta(Ar/N_2)$, e.g. ~~using~~

$$\delta(Ar/N_2) \approx \frac{(M_{Ar} - M_{air}) \times D_{Ar} - (M_{N2} - M_{air}) \times D_{N2}}{(M_{GST} - M_{air}) \times D_{GST}} \times \delta_{GST}. \qquad 5$$

The appropriate diffusivity values D_{Ar} and D_{N2} for Ar and N_2 in air are derived in Appendix B for a ternary mixture of Ar, O_2, and N_2, extending previous work (Ishidoya et al., 2013; Belikov et al., 2019). δ_{GST} is divided by D_{GST} in eq. 5 and therefore does not depend on the exact value chosen for D_{GST}.

2.4. NIES TM model

220   We compare our results to previous simulations of GS using the National Institute for Environmental Studies chemical transport model (NIES TM) published recently by Belikov et al. (2019). The NIES TM is a three-dimensional transport model of similar complexity as TOMCAT driven by the Japanese 25-year Reanalysis (JRA-25) with a hybrid sigma–isentropic (σ–θ) vertical coordinate up to 5 hPa or ~35 km. The model and GS results are described in detail in Belikov et al. (2013) and Belikov et al. (2019), respectively.

225 **3. Results**

3.1. N₂O-AoA calibration

Our new ~~N₂O AoA~~ relationships between N₂O and mean AoA for the Northern and Southern Hemispheres (NH & SH) are well constrained by the observations and generally follow the mid-latitude NH calibration curve of Andrews et al. (2001) (Fig. 3, panel (d)). At mean AoA>2.5, the N₂O-AoA relationship~~s~~ yields slightly younger ages in the NH than in the SH, and

230   compared to the previously published curve, suggesting there might be a latitudinal dependence of the relationship. Such a latitude-dependence is, in fact, expected based on theory (Plumb, 2007). We expect curvature and a latitudinal difference in the N₂O-AoA relationship because photolysis of N₂O depends on latitude and altitude due to local sunlight availability. Furthermore, mixing of young and old air results in a mixture with anomalously low N₂O concentration for a given age. Since

the SH, NH and tropics feature different photolysis rates and show different degrees of mixing/isolation, different N_2O-AoA

235 relationship are expected.

Unimodal age spectra are generally preferentially selected by the algorithm, in particular for the SH ~~for young AoA~~ (Fig. 3, panels (a)-(c)) but unimodal age spectra dominate the solution ensemble by a small margin except for AoA <0.5 years (Fig. S3). ~~whereas bimodal spectra are slightly more common for older samples representing 50-80% of the solution ensemble in these bins. However, c~~Confidence intervals on age spectra parameters from bimodal spectra are considerably wider than for

240 unimodal spectra. This implies that the parameters in a bimodal distribution are redundant and the shape of the age spectrum is not sufficiently constrained by the observations used. The width of the spectrum in each N_2O bin varies widely within the prescribed bounds for most unimodal and bimodal age spectra (Fig. S3). It appears that not enough data is available from the airborne campaigns to determine the amplitude of the seasonal cycle with enough confidence to constrain the width of the age spectrum and distinguish the relative contribution of the old peak (influencing ~~only~~ primarily the mean concentration difference

245 to the troposphere) and the young peak (controlling the amplitude of seasonality) in bimodal age spectra. ~~Because bimodal distributions are generated with a random value of the weighting factor *A* and the mean AoA of the second peak is assumed to be old (5-7 years), randomly generated bimodal solutions often produce overall AoA that is quite old. Therefore, they are less likely to be selected by the MCMC algorithm for young AoA bins and a prevalent occurrence of unimodal age spectra in these bins is expected by chance. If more stratospheric data were available, the seasonality of CO_2 in each N_2O bin would be better~~

250 ~~resolved and the algorithm could derive tighter constraints on all parameters in bimodal age spectra, also allowing it to distinguish more clearly between unimodal or bimodal age spectra.~~ Additional observations with less scatter or information from different age tracers are needed to properly resolve the shape of the age spectra. Despite the limited resolution of the seasonal cycle, the observations are sufficient to place tight limits on the mean AoA in each N_2O bin and yield a well-characterised relationship between N_2O and mean AoA for each hemisphere.

255 ### 3.2. The relationship between mean AoA and GS in models and observations

A comparison of the AoA-GS relationship with observations yields good agreement for the TOMCAT model results within the observational uncertainties (Fig. 4 panel (a)), but the observations fall outside the range of GS predicted by the NIES TM (Belikov et al., 2019) for mean AoA >1 year (Fig. 4, panel (b)). For young samples with mean AoA < 3 years, GS of $\delta(Ar/N_2)$ increases by roughly 35-45 per meg per year of AoA in both TOMCAT and observations, and converges to zero for the

260 youngest samples. In the upper stratosphere, TOMCAT does not obtain any ages as old as observed by Ishidoya et al., (2008, 2013, 2018) and therefore cannot reproduce these observations directly. Changing the vertical coordinate system of TOMCAT or forcing the model with a different reanalysis product could improve agreement with the observations for old ages because TOMCAT in the configuration used here is known to slightly underestimate AoA in the upper stratosphere (Monge-Sanz et al., 2007; Chabrillat et al., 2018). The very steep AoA-GS relationship for the oldest simulated air is however seen in TOMCAT

265 and the balloon observations.

A comparison of δ(Ar/N$_2$) and AoA between flask samples and TOMCAT is shown in Figure S4. However, the direct comparison should not be considered a good test of model performance because our stratospheric composition criteria intentionally target unusually old air in the lowermost stratosphere. For the same reason, profiles of δ(Ar/N$_2$) and AoA are not evaluated here. Instead, we suggest the AoA-GS relationship provides a more robust metric for comparison. Overall, our observations validate the implementation of GS for young (< 3 y) ages in TOMCAT.

The relationship between mean AoA and GS differs between TOMCAT and the NIES TM (Fig. 5). The NIES TM shows weaker curvature in the relationship overall and produces larger declines in Ar/N$_2$ at ages less than 4.5 years compared to TOMCAT. In TOMCAT's mesosphere (which is at the limit of the domain covered by the ERA-Interim reanalyses forcing the model), AoA is near uniform but GS continues to increase with increasing altitude, changing the relationship between mean AoA and GS in this region. The AoA-GS relationship in TOMCAT is very similar in all non-tropical regions (outside 15° > lat > -15°) whereas the curvature of the relationship is slightly stronger in the tropics. In contrast, the NIES TM has a clear dependence of the AoA-GS relationship on latitude. There is also some evidence in the observations for a dependence of the AoA-GS relationship on hemisphere. In the observations, δ(Ar/N$_2$) values appear to be slightly more negative in the SH than in the NH for the same age, particularly for AoA > 1.5 years. However, almost all SH samples with mean AoA older than 1.5 years come only from the ORCAS campaign. This campaigned used a different inlet design than HIPPO and AToM which could cause a small artifactual offset and but the scatter in our observations generally makes it difficult to separate signal from noise for small interhemispheric differences.

3.3. GS and AoA in TOMCAT

Annual mean δ(Ar/N$_2$) in TOMCAT follows the typical pattern of a tracer with a stratospheric sink (Chipperfield, 2006), as previously found in simulations using the NIES TM (Belikov et al., 2019) and the SOCRATES model (Ishidoya et al., 2013; Sugawara et al., 2018) (Fig. 6). δ(Ar/N$_2$) is zero in the troposphere and decreases with elevation. The most depleted δ(Ar/N$_2$) is observed at high latitudes where sinking air of the Brewer-Dobson circulation advects strongly fractionated air downward. In the tropics, δ(Ar/N$_2$) values are considerably less fractionated at the same altitude due to upwelling of unfractionated tropospheric air. Vertical gradients in δ(Ar/N$_2$) generally increase with altitude and are largest in the mesosphere (> 50km) because molecular diffusion increases with decreasing pressure and eventually dominates above the turbopause (not shown). There are strong seasonal changes in δ(Ar/N$_2$) depletion on the order of several thousand per meg, in particular in the high latitude mesosphere, with the strongest fractionation occurring during the winter season (see movie available with online version of the manuscript).

The AoA tracer in TOMCAT shows a similar pattern to δ(Ar/N$_2$) with younger ages at low altitude and in the tropics and oldest ages in the mesosphere. Vertical gradients in mean AoA are largest at high latitudes close to the tropopause. At high latitudes above 20 km and at low latitudes above 50 km, vertical gradients of AoA mostly disappear and AoA becomes nearly uniform.

4. Discussion

4.1. Difference between TOMCAT and the NIES TM

300 We hypothesize that an adequate representation of the mesosphere in models is critical in determining the curvature of the AoA-GS relationship. The residence time of air above ~40 km is rather short in TOMCAT and AoA is nearly constant with altitude in this region. GS in contrast continues to increase with altitude because of the diffusivity dependence on pressure allowing gases to separate more effectively. The much lower top of the NIES TM (~35 km vs 60–65 km) reduces its ability to capture this effect, which impacts the AoA-GS relationship because the mesospheric signal is exported into the stratosphere,

305 in particular in polar regions where mesospheric air ~~is sinking~~<u>descends</u>. TOMCAT furthermore produces a less negative slope of the AoA-GS relationship for young air (<u>mean</u> AoA <3 years) and greater similarity in the AoA-GS relationship between latitude bands than the NIES TM as shown in Fig. 5. This could in part be a consequence of using a different meteorological reanalysis product for forcing the two models (Chabrillat et al., 2018), or could indicate differences between the models in vertical and horizontal mixing intensity.

4.2. Estimating <u>mean</u> AoA from observed $\delta(Ar/N_2)$

Different tracers of AoA all have unique strengths and weaknesses. Estimating AoA in the lowermost stratosphere from CO_2 for example is limited by complexities involving seasonality and the possibility of multiple entry points of air into the stratosphere due to isentropic mixing with the midlatitude troposphere (Hall and Plumb, 1994; Andrews et al., 2001; Waugh and Hall, 2002). SF_6-derived AoA in contrast is biased high at mid and high latitudes due to the influence of mesospheric air

315 ~~in~~ which SF_6 has been <u>photochemically</u> depleted <u>in SF_6</u> ~~by electron attachment, photolysis, and chemical reactions~~ (Kovács et al., 2017; Linz et al., 2017). N_2O as an AoA tracer relies on the photolytic destruction of N_2O in the stratosphere which may depend strongly on location and had only been empirically calibrated for young ages at mid latitudes so far (Andrews et al., 2001; Linz et al., 2017).

Thanks to the robust relationship between <u>mean</u> AoA and $\delta(Ar/N_2)$, and the small seasonal cycle amplitude of $\delta(Ar/N_2)$ <~~6~~ <u>5</u>

320 per meg in the upper troposphere (Fig. ~~S2~~<u>S1</u>), <u>mean</u> AoA could also be estimated from $\delta(Ar/N_2)$. Using the current analytical $\delta(Ar/N_2)$ uncertainty of 12.2 per meg (2σ) and the AoA-$\delta(Ar/N_2)$ relationship seen in TOMCAT (including variability from seasonal and latitudinal differences), we estimate that <u>mean</u> AoA could be calculated to within about ±0.4 years (2σ, Fig. S~~5~~<u>4</u>). This is still considerably worse than the ±0.1 years confidence interval in <u>mean</u> AoA estimated from N_2O. However, the uncertainty ~~is almost exclusively analytical and~~ can be improved with future improvements in sampling<u>,</u> and the accuracy and

325 precision of the measurements.

The heavy noble gases krypton and xenon will be roughly 3.6× and 5.8× more strongly fractionated in the stratosphere than Ar, but are also more challenging to measure due to their ~8000× and 100000× lower abundance in the atmosphere. Nevertheless, future analysis of these gases in stratospheric air samples might further improve our ability to estimate AoA

11

from the gravitational fractionation of gases and help diagnose artefactual fractionation, because heavier noble gases are more

330 strongly fractionated under the influence of gravity and less sensitive to thermal fractionation (Seltzer et al., 2019).

4.3. Future directions

An open question in climate applications of noble gases is whether there could be a stratospheric influence on tropospheric $\delta(Ar/N_2)$. Tropospheric observations of $\delta(Ar/N_2)$ and other noble gas-elemental ratios have been used to infer ocean heat content changes by capitalizing on the temperature-dependency of gas solubility in the oceans (Keeling et al., 2004; Headly

335 and Severinghaus, 2007; Ritz et al., 2011; Bereiter et al., 2018). However, long-term trends (Butchart, 2014) and natural variability in stratospheric circulation and stratosphere-troposphere exchange (STE) such as the Quasi-Biennial Oscillation (QBO) (Baldwin et al., 2001) could advect a stratospheric GS signal into the troposphere and alias onto surface observations of these gases. To this end, we calculate the volumetric Ar deficit in the atmosphere in moles using TOMCAT (Fig. 7) relative to a hypothetical, unfractionated atmosphere with a homogenous mixing ratio ($C_{o_{Ar}}$) at all elevations. Ar deficit is

$$Ar_{deficit} = \left(C_{Ar} - C_{o_{Ar}}\right) \times N_{air} \qquad\qquad 6$$

340 where C_{Ar} is the simulated Ar mixing ratio and N_{air} is the number density of air.

The Ar deficit is concentrated in the lower stratosphere at around 20 km and at mid to high latitudes. Although the GS signal is considerably stronger in the mesosphere, the potential for perturbing the troposphere is low given the low molar density of air in the mesosphere. The region with greatest potential to influence the troposphere therefore lies in the lower stratosphere. The total Ar deficit of the atmosphere above 200 hPa is approximately -32.9 ×10^{13} mol and the atmosphere below 200 hPa

345 contains roughly 1.3 ×10^{18} mol of Ar. Perturbing STE and/or the stratospheric circulation by 10%, consistent with interannual to decadal variability of STE in models (Salby and Callaghan, 2006; Flury et al., 2013; Ray et al., 2014; Montzka et al., 2018), thus should lead to a detectable signal of roughly 3 2.2 per meg (-3.9 2.9×10^{13}/1.3×10^{18}×10%×10^6) in tropospheric $\delta(Ar/N_2)$. Corresponding advection of stratospheric GS signals in N_2 amplifies the pure Ar signal by roughly 10%. A careful investigation of such a signal in the $\delta(Ar/N_2)$ surface network data (Keeling et al., 2004) is needed because secular trends of $\delta(Ar/N_2)$ caused

350 by degassing of Ar and N_2 from a warming ocean are also expected to be on the order of 2-3 per meg per decade. Previous studies have also used ratios involving heavier noble gases (Xe/N_2, Kr/N_2) to reconstruct mean ocean temperature changes over glacial-interglacial timescales (Headly and Severinghaus, 2007; Bereiter et al., 2018; Baggenstos et al., 2019). Simultaneous changes in stratospheric circulation and STE affect heavier noble gases more strongly than Ar/N_2 and will need to be accounted for in such applications of noble gas thermometry.

5. Conclusions

With improvements in data treatment, measurement quality, and modelling constraints, we have shown that gravitational fractionation of Ar relative to N_2 in the stratosphere and mesosphere is a potentially powerful constraint on circulation. High-precision observations of $\delta(Ar/N_2)$ in air samples of the lowermost stratosphere from 9 airborne campaigns are well captured by the 3-D chemical transport model TOMCAT/SLIMCAT, which has been updated to account for the influence of gravity on air composition through molecular diffusion. In the observations and the model, $\delta(Ar/N_2)$ is directly related to stratospheric mean age of air (AoA) derived here using a new calibration of N_2O. Our observations for mean AoA <3 years produce a slope of roughly 35-45 per meg of $\delta(Ar/N_2)$ per year of AoA. TOMCAT/SLIMCAT shows better agreement with the new observations than the NIES transport model (Belikov et al., 2019) and we speculate that the model disagreement could be explained by (i) the factor of 2 lower top of the NIES transport model, (ii) the use of different reanalysis products, and/or (iii) differences in vertical and horizontal mixing. In this context, further work is also needed to study the influence of unresolved turbulence on AoA and $\delta(Ar/N_2)$ in chemical transport models.

As the importance of stratospheric circulation for ozone recovery, climate projections, and evaluation of tropospheric trends in halocarbons is increasingly recognized, a need for new observations from the still undersampled stratosphere is becoming evident. Combining $\delta(Ar/N_2)$ with other tracers of circulation could lead to new insights into atmospheric mixing and transport. $\delta(Ar/N_2)$ has potential advantages over existing approaches based on transient tracers such as CO_2 or N_2O since $\delta(Ar/N_2)$ is only influenced by the physics of transport and mostly unaffected by seasonality in the tropical upper troposphere. Furthermore, because of its large gradients at high altitudes, $\delta(Ar/N_2)$ observations from the upper stratosphere and mesosphere could improve our understanding of circulation on seasonal and interannual timescales in a region where changes in AoA from transient tracers are otherwise difficult to resolve becomes near-uniform and some photochemical age tracers such as N_2O are almost fully depleted and thus lose sensitivity to circulation changes.

6. Data availability

All Data from the ORCAS, HIPPO and ATom airborne campaigns are freely available at: doi:10.3334/CDIAC/HIPPO_014, doi:10.5065/D6SB445X, doi:10.3334/CDIAC/HIPPO_014, and doi:10.3334/ornldaac/1581. All of the primary data used here is consistent with the associated merged data products in Wofsy et al. (2012), Stephens (2017), and Wofsy et al. (2018). For Medusa samples identified as stratospheric, CO_2, $\delta(Ar/N_2)$, AoA, and kernel-averaged N_2O, O_3, H_2O, CO, CH_4, as well as additional metadata are available as a supplement to this manuscript at https://doi.org/10.3334/ORNLDAAC/1788xxxx (doi: in prep.).

NIES TM modelling results (Belikov et al., 2019) and stratospheric observations of GS from the cryogenic balloon sampling-system (Ishidoya et al., 2008, 2013, 2018; Sugawara et al., 2018) were provided directly by the authors.

7. Appendix A: Description of the Markov chain Monte Carlo algorithm

The following list outlines key steps in our MCMC algorithm to calculate AoA from CO_2 for each N_2O bin:

1) Start a new Markov Chain and allow for uncertainty in TRTS by adding white noise with an amplitude given by the scatter of upper tropospheric observations around the mean TRTS in Fig. 2.

2) With a 50% chance, peak weighting factor A is set to zero and a unimodal spectrum is tested (k=1). Alternatively, A is allowed to vary between 0 and 1 for a bimodal distribution (k=2).

3) The other age spectrum parameters are selected from N_2O bin-dependent prior distributions: Γ_1 is sampled from a uniform prior distribution with a mean AoA predicted by the N_2O-AoA relationship of Andrews et al. (2001) and with generous width (>1.5 years). Γ_2 is sampled from a uniform prior distribution with values between ~~5~~1-~~7~~6 years ~~based on the nearly invariant value of ~6 years previously found by Andrews et al. (2001)~~ representing older stratospheric air. The shape parameters are defined as $\lambda_i \equiv \frac{\Gamma_i^2}{2\gamma_i}$, and γ_i is chosen randomly for each peak with values between 0.1 and 1, as observed in previous studies (Hall and Plumb, 1994; Andrews et al., 2001; Waugh and Hall, 2002).

4) Convolve the age spectrum calculated from the parameters ($\boldsymbol{m} \equiv [\Gamma_1, \Gamma_2, \lambda_1, \lambda_2, A]$) with the perturbed TRTS to obtain a possible time series of CO_2 in the N_2O bin and calculate the misfit ($\boldsymbol{e}$) between the CO_2 time series and observations.

5) Calculate the likelihood function $P(\boldsymbol{d}|k,\boldsymbol{m})$ for the set of parameters $\boldsymbol{m}$ and k, given a total of n observations ($\boldsymbol{d}$)

$$P(\boldsymbol{d}|k,\boldsymbol{m}) = \frac{1}{[(2\pi)^n \det(\widehat{\boldsymbol{C}}_e)]^{0.5}} \exp\left(-\frac{1}{2}\boldsymbol{e}^T \widehat{\boldsymbol{C}}_e^{-1}\boldsymbol{e}\right) \qquad (A1)~~7~~$$

where $\widehat{\boldsymbol{C}}_e$ is the covariance matrix. ~~Because~~ We assume that all n observations are independent. Thus, $\widehat{\boldsymbol{C}}_e$ has only diagonal entries of $\sigma_{CO_2}^2$ and $\det(\widehat{\boldsymbol{C}}_e)$ simplifies to $(\sigma_{CO_2}^2)^n$. The value of $\sigma_{CO_2}^2$ is different for each bin and determined iteratively as the approximate root mean square error of the observations around the final time series for each bin obtained at the end of the MCMC algorithm. Typical values of $\sigma_{CO_2}^2$ are between 0.18 and 1.28 ppm and generally decrease with increasing AoA of a N_2O bin.

6) If this is the first pass of the chain, define k_{saved} and $\boldsymbol{m}_{saved}$ to equal k and $\boldsymbol{m}$. Otherwise, calculate the selection criterion $\alpha \equiv \min\left(1, \frac{P(\boldsymbol{d}|k,\boldsymbol{m})}{P(\boldsymbol{d}|k_{saved}, \boldsymbol{m}_{saved})}\right)$ and accept k and $\boldsymbol{m}$ as new saved values ($k_{saved}, \boldsymbol{m}_{saved}$) with probability α. Sampling from the same prior distributions on each pass of the chain simplifies our expression of α compared to that presented by Malinverno (2002), making it only dependent on the likelihood ratio.

7) Repeat steps 2-6 1000 times sampling parameter values from the same prior distributions and store the final values of k_{saved} and $\boldsymbol{m}_{saved}$ obtained after the 1000th iteration (i.e., a plausible solution produced past the burn-in period) for later use.

8) To sample the full posterior pdf (i.e., the full uncertainty about the age spectrum parameters), initialize 2000 different Markov chains by repeating steps 1-7. Each stored value of $\mathbf{m}$ characterizes one age spectrum that is likely not far from the best solution, given the data $\mathbf{d}$, yielding an ensemble of 2000 age spectra from which statistics can be computed. Note that each Markov chain is fully independent, so the algorithm can be easily parallelized to minimize computational costs.

8. Appendix B: Derivation of Eq. (4) from the Maxwell-Stefan Equations

We start by approximating air as a ternary mixture of N_2, O_2 and Ar, and later generalize to consider additional trace species. According the Maxwell-Stefan equations (Taylor and Krishna, 1993) diffusion in this ternary mixture is governed by:

$$d_{N2} = \frac{C_{N2}\, f_{Ar} - C_{Ar}\, f_{N2}}{N \times D_{N2:Ar}} + \frac{C_{N2}\, f_{O2} - C_{O2}\, f_{N2}}{N \times D_{N2:O2}} \tag{B1}$$

$$d_{Ar} = \frac{C_{Ar}\, f_{N2} - C_{N2}\, f_{Ar}}{N \times D_{Ar:N2}} + \frac{C_{Ar}\, f_{O2} - C_{O2}\, f_{Ar}}{N \times D_{Ar:O2}} \tag{B2}$$

$$f_{O2} = -f_{N2} - f_{Ar} \tag{B3}$$

$$C_{N2} + C_{O2} + C_{Ar} = 1 \tag{B4}$$

where $C_i \equiv n_i/N$ is the mole fraction, n_i is molar or number density (mol m^{-3}), N is the total number density (mol m^{-3}), f_i is the molecular diffusion flux (mol m^2 s^{-1}) relative to the molar average velocity of the mixture, d_i is the thermodynamic driving force for molecular diffusion (m^{-1}), and $D_{i:j}$ is the binary diffusion coefficient of the $(i:j)$ pair. An equation for O_2 that is analogous to Eq. (B~~A~~1) is not needed because changes in O_2 are governed by the conservation Eqs. (B~~A~~3) and (~~A~~B4). Binary diffusivity coefficients (cm^2 s^{-1}) can be calculated using the method of Fuller et al. as reported in (Reid et al., 1987)

$$D_{ij} = 0.001 \frac{T^{1.75} \left(\frac{1}{M_i} + \frac{1}{M_j} \right)^{0.5}}{P \left[v_i^{1/3} + v_j^{1/3} \right]^2} \tag{B5}$$

where P is pressure (atm), v_i is the molecular diffusion volume of a trace gas or air (Table ~~A~~B1), and M_i is the molecular mass (g mol^{-1}) of a gas.

For an ideal gas, d_i is given by

$$d_i \equiv \nabla C_i + (C_i - \omega_i)\frac{\nabla P}{P} - \frac{\rho_i}{P}\left(\frac{F_i}{M_i} - \sum_{j=1}^{n}\omega_i\frac{F_j}{M_j}\right) + \frac{k_i^T}{T}\nabla T \tag{B6}$$

where $\omega_i \equiv \frac{\rho_i}{\rho} = \frac{M_i C_i}{M_{air}}$ is the mass fraction of gas i, ρ_i is density of i (kg m^{-3}), P is pressure (Pa), T is temperature (K), k_i^T is the thermal-diffusion ratio of i, M_i is the molecular mass of i (kg mol^{-1}), and F_i the external body force per mole (N mol^{-1}) for i (Chapman et al., 1990; Taylor and Krishna, 1993).

In the atmosphere, (vertical) pressure gradients are caused by gravity and well approximated by hydrostatic balance

$$\nabla P \approx \frac{\partial P}{\partial z} \approx -\rho g = -\frac{P M_{air}}{RT}g \tag{B7}$$

The gravitational force per mole F_i is

$$F_i = -\frac{\rho_i}{n_i}g \tag{B8}$$

Substituting Eqs. (~~A~~B7) and (~~AB~~8) into Eq. (~~AB~~6) yields

$$d_i \approx \nabla C_i - \left(1 - \frac{M_i}{M_{air}}\right)C_i\frac{1}{P}\frac{P M_{air}g}{RT} + \frac{\rho_i}{P}\left(\frac{\frac{\rho_i}{n_i}g}{M_i} - \sum_{j=1}^{n}\omega_j\frac{\frac{\rho_j}{n_j}g}{M_j}\right) + \frac{k_i^T}{T}\nabla T$$

$$= \nabla C_i + (M_i - M_{air})\frac{g}{RT}C_i + \frac{\rho_i}{P}\left(g - \sum_{j=1}^{n}\omega_i g\right) + \frac{k_i^T}{T}\nabla T \tag{B9}$$

$$= \nabla C_i + \left(\frac{1}{H_i} - \frac{1}{H_{air}}\right)C_i + \frac{k_i^T}{T}\nabla T$$

where we use the definition of the scale height $H_i \equiv \frac{RT}{gM_i}$. The two terms involving the body force cancel because all species experience the same gravitational force per unit mass. The tendency for gravimetric separation instead arises from the pressure gradient term proportional to $\left(\frac{1}{H_i} - \frac{1}{H_{air}}\right)$.

Equations (~~A~~B1) and (~~A~~B2) can be inverted to solve for f_{N2} and f_{Ar} (Taylor and Krishna, 1993)

$$f_{N2} = -ND_{N2}^{air}d_{N2} - ND_{N2\times(Ar,O2)}^{air}d_{Ar} \tag{B10}$$

$$f_{Ar} = -ND_{Ar\times(N2,O2)}^{air}d_{N2} - ND_{Ar}^{air}d_{Ar} \tag{B11}$$

16

where

$$D_{N2}^{air} = \frac{D_{N2:O2}(C_{N2}D_{Ar:O2} + (1 - C_{N2})D_{N2:Ar})}{S} \tag{B12}$$

$$D_{N2\times(Ar,O2)}^{air} = \frac{C_{N2}D_{Ar:O2}(D_{N2:O2} - D_{N2:Ar})}{S} \tag{B13}$$

$$D_{Ar}^{air} = \frac{D_{Ar:O2}(C_{Ar}D_{N2:O2} + (1 - C_{Ar})D_{N2:Ar})}{S} \tag{B14}$$

$$D_{Ar\times(N2,O2)}^{air} = \frac{C_{Ar}D_{N2:O2}(D_{Ar:O2} - D_{N2:Ar})}{S} \tag{B15}$$

$$S = C_{N2}D_{Ar:O2} + C_{Ar}D_{N2:O2} + C_{O2}D_{N2:Ar} \tag{B16}$$

440   Here D_{N2}^{air} and D_{Ar}^{air} are the effective diffusivities of N_2 and Ar in air, while $D_{N2\times(Ar,O2)}^{air}$ and $D_{Ar\times(N2,O2)}^{air}$ reflect ternary cross interactions, such as the tendency of N_2 to be impacted by any process that drives a diffusive flux of Ar.

In air, Ar is a minor gas ($C_{Ar} \ll C_{N2} \sim C_{O2}$) and therefore interactions of N_2 with Ar can be neglected in the N_2 flux and the diffusive fluxes of N_2 and O_2 must balance approximately as in the case of a binary mixture of the two gases

$$f_{N2} \approx -ND_{N2}^{air}d_{N2} \tag{B17}$$

$$f_{O2} \approx -f_{N2} \tag{B18}$$

Combining Eqs. (B10) and (B17) with Eq. (B9), yields

$$f_{N2} \approx -ND_{N2}^{air}\left[\nabla C_{N2} + \left(\frac{1}{H_{N2}} - \frac{1}{H_{air}}\right)C_{N2} + \alpha_{N2:O2}^{T}\frac{C_{N2}(1 - C_{N2})}{T}\nabla T\right] \tag{B19}$$

$$
\begin{aligned}
f_{Ar} \approx &-ND_{Ar\times(N2,O2)}^{air}\left[\nabla C_{N2} + \left(\frac{1}{H_{N2}} - \frac{1}{H_{air}}\right)C_{N2} + \alpha_{N2:O2}^{T}\frac{C_{N2}(1 - C_{N2})}{T}\nabla T\right] \\
&-ND_{Ar}^{air}\left[\nabla C_{Ar} + \left(\frac{1}{H_{Ar}} - \frac{1}{H_{air}}\right)C_{Ar} + \alpha_{Ar:air}^{T}\frac{C_{Ar}}{T}\nabla T\right]
\end{aligned}
\tag{B20}
$$

445   where we have replaced the thermal diffusion ratio k_i^T with the better empirically constrained thermal diffusion factor $\alpha_i^T \equiv \frac{k_i^T}{c_i c_j}$ defined as such only in binary mixtures. Therefore, $k_{N2}^T \approx \alpha_{N2:O2}^T C_{N2} C_{O2}$ and $k_{Ar}^T \approx \alpha_{Ar:air}^T C_{Ar}$. Table B2 presents rough estimates of the magnitudes of the terms in Eqs. (B19) and (B20) in the stratosphere, showing that the thermal diffusion

17

and cross-diffusion terms involving $D_{Ar\times(N2,O2)}^{air}$ are at least two orders of magnitude smaller than the remaining terms. Neglecting these smaller terms yields the governing Eq. (4) used in our model simulation

$$f_i \approx -ND_i \left[\left(\frac{1}{H_i} - \frac{1}{H_{air}} \right) C_i \right] = -ND_i \Delta M_i \frac{g}{RT} C_i \tag{B21}$$

where we introduced $\Delta M_i \equiv M_i - M_{air}$ the molecular mass difference to air (kg mol^{-1}) as a short-hand. Equation (BA21) is equally valid for trace gases such as Ar and major gases N_2 and O_2 when the appropriate diffusivities given by Eqs. (BA12) and (BA14) are used. In our case $D_{N2} \equiv D_{N2}^{air} = D_{O2}^{air} \approx D_{N2:O2}$ for N_2 and $D_{Ar} \equiv D_{Ar}^{air} \approx \frac{D_{Ar:O2}D_{N2:Ar}}{C_{N2}D_{Ar:O2}+C_{O2}D_{N2:Ar}}$ for Ar.

9. Appendix C: Calculating δ(Ar/N₂) from δ(GST)

The conservation equation of gas i with mixing ratio C_i accounting for advection (1st term RHS), eddy mixing (2nd term RHS), and molecular diffusion (3rd term RHS, using the simplified Eq. (BA21)) is given by

$$\frac{\partial}{\partial t}(C_i) = -\vec{u} \cdot \vec{\nabla}[C_i] + \vec{\nabla} \cdot \left[\boldsymbol{D}_e \vec{\nabla} C_i \right] + \frac{\partial}{\partial z} \left[D_i \Delta M_i \frac{g}{RT} C_i \right] \tag{C1}$$

where $\vec{u}$ is the velocity vector (m s^{-1}), $\boldsymbol{D}_e$ is the eddy diffusivity tensor (m^2 s^{-1}), D_i is the molecular diffusivity of species i in air (i.e., D_i^{air} m^2 s^{-1}), ΔM_i the molecular mass difference to air (kg mol − 1), g the gravitational acceleration, R is the fundamental gas constant (J K^{-1} mol^{-1}), and T is temperature (K). "$\vec{\nabla} \cdot$" represents the divergence operator. $\vec{u}, \boldsymbol{D}_e, D_i, N,$ and T depend on x,y,z, and t and the largest gradients for these variables generally occur in the vertical direction.

Dividing Eq. (CB1) by a reference value $C_{i,0}$, separating C_i into a constant and a perturbation component (i.e., $\frac{C_i}{C_{i,0}} = 1 + \frac{C_i'}{C_{i,0}} = 1 + \delta_i$), and using the chain rule yields

$$\frac{\partial}{\partial t}(\delta_i) = -\vec{u} \cdot \vec{\nabla}[\delta_i] + \vec{\nabla} \cdot \left[\boldsymbol{D}_e \vec{\nabla} \delta_i \right] + D_i \Delta M_i \frac{g}{RT} \frac{\partial}{\partial z}[\delta_i] + (1 + \delta_i) \frac{\partial}{\partial z} \left[D_i \Delta M_i \frac{g}{RT} \right] \tag{C2}$$

To simplify Eq. (B2C2), we assume that the perturbations in the mixing ratio are small (i.e., $\delta_i \ll 1$) which is true at the 1% level for δ(Ar/N₂) even in the mesosphere. Furthermore, we assume that $U + \frac{D_e}{L} \gg \frac{D_i}{\Delta H_i}$ (U is a characteristic velocity scale, D_e is a characteristic eddy diffusivity, and L is a characteristic length scale; $\Delta H_i = \frac{RT}{\Delta M_i g}$) or equivalently stated in terms of the Peclet number $Pe \sim \frac{U \Delta H_i + D_e \frac{\Delta H_i}{L}}{D_i} \gg 1$. We estimate typical values of the vertical Peclet number in the stratosphere to be between 1000 to 10000 based on the height (~20 km) and turnover time (~5 years) of the stratosphere and the range of molecular diffusivities given in Table A1B1. In the mesosphere, AoA is near-uniform and we conservatively assume a turnover time of 0.5 years and height of 20 km. This implies a Peclet number of ~500 for the mesosphere. Under these conditions, the third

18

term on the RHS can be eliminated because it is $O(Pe)$ smaller than the first and second terms and because it is $O(\delta_i)$ smaller than the fourth term. Thus, we obtain

$$\frac{\partial}{\partial t}(\delta_i) \approx -\vec{u} \cdot \vec{\nabla}[\delta_i] + \vec{\nabla} \cdot \left[\boldsymbol{D}_e \vec{\nabla}\delta_i\right] + -\frac{\partial}{\partial z}\left[D_i \Delta M_i \frac{g}{RT}\right] \tag{C3}$$

Where the last term approximates the vertical divergence of the gravitational flux. At the top of the atmosphere and Earth's surface, the gravitational flux abruptly vanishes, and its divergence becomes large. If its divergence is small in the interior, Eq. (C3B4) can be conceptually interpreted as an advection-diffusion problem with large apparent sources or sinks at the bottom and top boundary due to diffusive flux divergence. The steady-state solution to Eq. (CB3) for a tracer δ_i with no additional sources and sinks (other than the apparent sources from diffusive flux divergence) yields the steady-state profile for species i. This solution can be scaled to yield the steady-state solution for any other tracer δ_j as

$$\delta_j \approx \frac{D_j \Delta M_j}{D_i \Delta M_i} \delta_i \tag{C4}$$

as can easily be validated by solving Eq. (CB4) for δ_i and substitution into Eq. (CB3). Note that $\frac{D_j \Delta M_j}{D_i \Delta M_i}$ does not depend on x,y, and z.

Furthermore, ratios of passive tracers k and l can be calculated directly from δ_i by recognizing that $\delta_{k/l} \approx \delta_k - \delta_l$ for $\delta_k + \delta_l \ll 1$. Hence, we obtain the general version of Eq. (65) in the main text

$$\delta_{k/l} \approx \delta_k - \delta_l \approx \frac{D_k^{air} \Delta M_k - D_l^{air} \Delta M_l}{D_i \Delta M_i} \delta_i. \tag{C5}$$

10. Author contributions

MPC, WF and CW set up and ran the TOMCAT model simulations. EJM, JDB, BBS, and SCW collected samples and curated data. BB carried out the data analysis with important input from MPC, RFK, SCW, ML, and EJM. BB prepared the manuscript with contributions from all co-authors.

11. Competing Interests

The authors declare that they have no conflict of interest.

19

12. Acknowledgements

13. References

Andrews, A. E., Boering, K. A., Daube, B. C., Wofsy, S. C., Hintsa, E. J., Weinstock, E. M. and Bui, T. P.: Empirical age spectra for the lower tropical stratosphere from in situ observations of CO2: Implications for stratospheric transport, Journal of Geophysical Research Atmospheres, 104(D21), 26581–26595, doi:10.1029/1999JD900150, 1999.

Andrews, A. E., Boering, K. A., Wofsy, S. C., Daube, B. C., Jones, D. B., Alex, S., Loewenstein, M., Podolske, J. R. and Strahan, S. E.: Empirical age spectra for the midlatitude lower stratosphere from in situ observations of CO2: Quantitative

evidence for a subtropical "barrier" to horizontal transport, Journal of Geophysical Research Atmospheres, 106(D10), 10257–10274, doi:10.1029/2000JD900703, 2001.

Baggenstos, D., Häberli, M., Schmitt, J., Shackleton, S. A., Birner, B., Severinghaus, J. P., Kellerhals, T. and Fischer, H.: Earth's radiative imbalance from the Last Glacial Maximum to the present, Proceedings of the National Academy of Sciences of the United States of America, 116, 14881–14886, doi:10.1073/pnas.1905447116, 2019.

Baldwin, M. P., Gray, L. J., Dunkerton, T. J., Hamilton, K., Haynes, P. H., Randel, W. J., Holton, J. R., Alexander, M. J., Hirota, I., Horinouchi, T., Jones, D. B. A., Kinnersley, J. S., Marquardt, C., Sato, K. and Takahashi, M.: The quasi-biennial oscillation, Reviews of Geophysics, 39, 179–229, doi:10.1029/1999RG000073, 2001.

Banks, P. M. and Kockarts, G.: Aeronomy, Part B, 1st ed., Academic Press, New York and London., 1973.

Belikov, D., Sugawara, S., Ishidoya, S., Hasebe, F., Maksyutov, S., Aoki, S., Morimoto, S. and Nakazawa, T.: Three-dimensional simulation of stratospheric gravitational separation using the NIES global atmospheric tracer transport model, Atmospheric Chemistry and Physics, 19, 5349–5361, doi:10.5194/acp-19-5349-2019, 2019.

Belikov, D. A., Maksyutov, S., Sherlock, V., Aoki, S., Deutscher, N. M., Dohe, S., Griffith, D., Kyro, E., Morino, I., Nakazawa, T., Notholt, J., Rettinger, M., Schneider, M., Sussmann, R., Toon, G. C., Wennberg, P. O. and Wunch, D.: Simulations of column-averaged CO2 and CH4 using the NIES TM with a hybrid sigma-isentropic (σ-θ) vertical coordinate, Atmospheric Chemistry and Physics, 13, 1713–1732, doi:10.5194/acp-13-1713-2013, 2013.

Bent, J. D.: Airborne Oxygen Measurements over the Southern Ocean as an Integrated Constraint of Seasonal Biogeochemical Processes, University of California San Diego., 2014.

Bereiter, B., Shackleton, S., Baggenstos, D., Kawamura, K. and Severinghaus, J.: Mean global ocean temperatures during the last glacial transition, Nature, 553, 39–44, doi:10.1038/nature25152, 2018.

Bieri, R. H. and Koide, M.: Noble gases in the atmosphere between 43 and 63 kilometers, Journal of Geophysical Research, 75(33), 6731–6735, doi:10.1029/jc075i033p06731, 1970.

Blaine, T. W., Keeling, R. F. and Paplawsky, W. J.: An improved inlet for precisely measuring the atmospheric Ar/N2 ratio, Atmospheric Chemistry and Physics, 6, 1181–1184, doi:10.5194/acp-6-1181-2006, 2006.

Boering, K. A., Wofsy, S. C., Daube, B. C., Schneider, H. R., Loewenstein, M., Podolske, J. R. and Conway, T. J.: Stratospheric mean ages and transport rates from observations of carbon dioxide and nitrous oxide, Science, 274(5291), 1340–1343, doi:10.1126/science.274.5291.1340, 1996.

Bönisch, H., Engel, A., Curtius, J., Birner, T. and Hoor, P.: Quantifying transport into the lowermost stratosphere using simultaneous in-situ measurements of SF6 and CO2, Atmospheric Chemistry and Physics, 9, 5905–5919, doi:10.5194/acp-9-5905-2009, 2009.

21

Brewer, A. W.: Evidence for a world circulation provided by the measurements of helium and water vapour distribution in the stratosphere, Quarterly Journal of the Royal Meteorological Society, 75(326), 351–363, doi:10.1002/qj.49707532603, 1949.

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

550   Chabrillat, S., Vigouroux, C., Christophe, Y., Engel, A., Errera, Q., Minganti, D., Monge-Sanz, B. M., Segers, A. and Mahieu, E.: Comparison of mean age of air in five reanalyses using the BASCOE transport model, Atmospheric Chemistry and Physics, 18, 14715–14735, doi:10.5194/acp-18-14715-2018, 2018.

Chapman, S., Cowling, T. G. and Burnett, D.: The mathematical theory of non-uniform gases: an account of the kinetic theory of viscosity, thermal conduction and diffusion in gases, 3rd ed., Cambridge University Press., 1990.

555   Chipperfield, M. P.: New version of the TOMCAT/SLIMCAT off-line chemical transport model: Intercomparison of stratospheric tracer experiments, Quarterly Journal of the Royal Meteorological Society, 132(617), 1179–1203, doi:10.1256/qj.05.51, 2006.

Chipperfield, M. P., Bekki, S., Dhomse, S., Harris, N. R. P., Hassler, B., Hossaini, R., Steinbrecht, W., Thiéblemont, R. and Weber, M.: Detecting recovery of the stratospheric ozone layer, Nature, 549, 211–218, doi:10.1038/nature23681, 2017.

560   Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., Mcnally, A. P., Monge-Sanz, B. M., Morcrette, J. J., Park, B. K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J. N. and Vitart, F.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Quarterly Journal of 565   the Royal Meteorological Society, 137, 553–597, doi:10.1002/qj.828, 2011.

Diskin, G. S., Podolske, J. R., Sachse, G. W. and Slate, T. A.: Open-path airborne tunable diode laser hygrometer, in Diode Lasers and Applications in Atmospheric Sensing, edited by A. Fried, pp. 196–204, Proceedings of the Society of Photo-Optical Instrumentation Engineers (SPIE)., 2002.

Dobson, G. M. B.: Origin and distribution of the polyatomic molecules in the atmosphere, Proceedings of the Royal Society 570   of London. Series A. Mathematical and Physical Sciences, 236, 187–193, doi:10.1098/rspa.1956.0127, 1956.

Engel, A., Möbius, T., Bönisch, H., Schmidt, U., Heinz, R., Levin, I., Atlas, E., Aoki, S., Nakazawa, T., Sugawara, S., Moore, F., Hurst, D., Elkins, J., Schauffler, S., Andrews, A. and Boering, K.: Age of stratospheric air unchanged within uncertainties over the past 30 years, Nature Geoscience, 2, 28–31, doi:10.1038/ngeo388, 2009.

Engel, A., Bönisch, H., Ullrich, M., Sitals, R., Membrive, O., Danis, F. and Crevoisier, C.: Mean age of stratospheric air 575   derived from AirCore observations, Atmospheric Chemistry and Physics, 17, 6825–6838, doi:10.5194/acp-17-6825-2017, 2017.

22

Flury, T., Wu, D. L. and Read, W. G.: Variability in the speed of the Brewer-Dobson circulation as observed by Aura/MLS, Atmospheric Chemistry and Physics, 13, 4563–4575, doi:10.5194/acp-13-4563-2013, 2013.

Green, P. J.: Reversible jump Markov chain monte carlo computation and Bayesian model determination, Biometrika, 82(4), 580 711–732, doi:10.1093/biomet/82.4.711, 1995.

Hall, T. M. and Plumb, R. A.: Age as a diagnostic of stratospheric transport, Journal of Geophysical Research, 99(D1), 1059–1070, doi:10.1029/93JD03192, 1994.

Hastings, W. K.: Monte carlo sampling methods using Markov chains and their applications, Biometrika, 57(1), 97–109, doi:10.1093/biomet/57.1.97, 1970.

585 Hauck, M., Fritsch, F., Garny, H. and Engel, A.: Deriving stratospheric age of air spectra using an idealized set of chemically active trace gases, Atmospheric Chemistry and Physics, 19, 5269–5291, doi:10.5194/acp-19-5269-2019, 2019.

Hauck, M., Bönisch, H., Hoor, P., Keber, T., Ploeger, F., Schuck, T. J. and Engel, A.: A convolution of observational and model data to estimate age of air spectra in the northern hemispheric lower stratosphere, Atmospheric Chemsitry and Physics Discussions, in review, doi:10.5194/acp-2020-167, 2020.

590 Haynes, P.: Stratospheric Dynamics, Annual Review of Fluid Mechanics, 37, 263–293, doi:10.1146/annurev.fluid.37.061903.175710, 2005.

Headly, M. A. and Severinghaus, J. P.: A method to measure Kr/N2 ratios in air bubbles trapped in ice cores and its application in reconstructing past mean ocean temperature, Journal of Geophysical Research, 112(D19105), 1–12, doi:10.1029/2006JD008317, 2007.

595 Holton, J. R., Haynes, P. H., Mcintyre, M. E., Douglass, A. R. and Rood, B.: Stratosphere-Troposphere, Review of Geophysics, (95), 403–439, 1995.

Hoor, P., Gurk, C., Brunner, D., Hegglin, M. I., Wernli, H. and Fischer, H.: Seasonality and extent of extratropical TST derived from in-situ CO measurements during SPURT, Atmospheric Chemistry and Physics Discussions, 4, 1427–1442, doi:10.5194/acpd-4-1691-2004, 2004.

600 Ishidoya, S., Sugawara, S., Morimoto, S., Aoki, S. and Nakazawa, T.: Gravitational separation of major atmospheric components of nitrogen and oxygen in the stratosphere, Geophysical Research Letters, 35(L03811), doi:10.1029/2007GL030456, 2008.

Ishidoya, S., Sugawara, S., Morimoto, S., Aoki, S., Nakazawa, T., Honda, H. and Murayama, S.: Gravitational separation in the stratosphere - A new indicator of atmospheric circulation, Atmospheric Chemistry and Physics, 13, 8787–8796, 605 doi:10.5194/acp-13-8787-2013, 2013.

Ishidoya, S., Sugawara, S., Inai, Y., Morimoto, S., Honda, H., Ikeda, C., Hashida, G., Machida, T., Tomikawa, Y., Toyoda, S., Goto, D., Aoki, S. and Nakazawa, T.: Gravitational separation of the stratospheric air over Syowa, Antarctica and its connection with meteorological fields, Atmospheric Science Letters, 19, 1–7, doi:10.1002/asl.857, 2018.

Karion, A., Sweeney, C., Wolter, S., Newberger, T., Chen, H., Andrews, A., Kofler, J., Neff, D. and Tans, P.: Long-term greenhouse gas measurements from aircraft, Atmospheric Measurement Techniques, 6, 511–526, doi:10.5194/amt-6-511-2013, 2013.

Keeling, C. D., Stephen, C., Piper, S. C., Bacastow, R. B., Wahlen, M., Whorf, T. P., Heimann, M. and Meijer, H. A.: Exchanges of atmospheric CO2 and 13CO2 with the terrestrial biosphere and oceans from 1978 to 2000, Global Aspects, SIO Reference Series, 01–06, 83–113, 2001.

Keeling, R. F.: Development of an interferometric oxygen analyzer for precise measurement of the atmospheric O2 mole fraction, Harvard University., 1988.

Keeling, R. F., Blaine, T., Paplawsky, B., Katz, L., Atwood, C. and Brockwell, T.: Measurement of changes in atmospheric Ar/N_2 ratio using a rapid-switching, single-capillary mass spectrometer system, Tellus, 56B, 322–338, doi:10.1111/j.1600-0889.2004.00117.x, 2004.

Kovács, T., Feng, W., Totterdill, A., Plane, J. M. C., Dhomse, S., Gómez-Martín, J. C., Stiller, G. P., Haenel, F. J., Smith, C., Forster, P. M., García, R. R., Marsh, D. R. and Chipperfield, M. P.: Determination of the atmospheric lifetime and global warming potential of sulfur hexafluoride using a three-dimensional model, Atmospheric Chemistry and Physics, 17, 883–898, doi:10.5194/acp-17-883-2017, 2017.

Krol, M., De Bruine, M., Killaars, L., Ouwersloot, H., Pozzer, A., Yin, Y., Chevallier, F., Bousquet, P., Patra, P., Belikov, D., Maksyutov, S., Dhomse, S., Feng, W. and Chipperfield, M. P.: Age of air as a diagnostic for transport timescales in global models, Geoscientific Model Development, 11, 3109–3130, doi:10.5194/gmd-11-3109-2018, 2018.

Li, F., Waugh, D. W., Douglass, A. R., Newman, P. A., Pawson, S., Stolarski, R. S., Strahan, S. E. and Nielsen, J. E.: Seasonal variations of stratospheric age spectra in the Goddard Earth Observing System Chemistry Climate Model (GEOSCCM), Journal of Geophysical Research Atmospheres, 117(D05134), 1–14, doi:10.1029/2011JD016877, 2012.

Linz, M., Plumb, R. A., Gerber, E. P. and Sheshadri, A.: The relationship between age of air and the diabatic circulation of the stratosphere, Journal of the Atmospheric Sciences, 73, 4507–4518, doi:10.1175/JAS-D-16-0125.1, 2016.

Linz, M., Plumb, R. A., Gerber, E. P., Haenel, F. J., Stiller, G., Kinnison, D. E., Ming, A. and Neu, J. L.: The strength of the meridional overturning circulation of the stratosphere, Nature Geoscience, 10, 663–667, doi:10.1038/ngeo3013, 2017.

Malinverno, A.: Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem, Geophysics Journal International, 151, 675–688, 2002.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E.: Equation of state calculations by fast computing machines, The Journal of Chemical Physics, 21(6), 1087–1092, doi:10.1063/1.1699114, 1953.

Monge-Sanz, B. M., Chipperfield, M. P., Simmons, A. J. and Uppala, S. M.: Mean age of air and transport in a CTM: Comparison of different ECMWF analyses, Geophysical Research Letters, 34(L04801), 1–5, doi:10.1029/2006GL028515, 2007.

Montzka, S. A., Dutton, G. S., Yu, P., Ray, E., Portmann, R. W., Daniel, J. S., Kuijpers, L., Hall, B. D., Mondeel, D., Siso, C., Nance, J. D., Rigby, M., Manning, A. J., Hu, L., Moore, F., Miller, B. R. and Elkins, J. W.: An unexpected and persistent increase in global emissions of ozone-depleting CFC-11, Nature, 557, 413–417, doi:10.1038/s41586-018-0106-2, 2018.

Morgan, E. J., Stephens, B. B., Birner, B. and Keeling, R. F.: Observational Constraints on the Vertical and Interhemispheric Gradient of the Ar/N2 Ratio in the Troposphere, in Abstract CT14A-0845 presented at Ocean Sciences Meeting 2020, AGU, San Diego., 2020.

Plumb, R. A.: Tracer interrelationships in the stratosphere, Reviews of Geophysics, 45(RG4005), 1–33, doi:10.1029/2005RG000179, 2007.

Podglajen, A. and Ploeger, F.: Retrieving the age of air spectrum from tracers: Principle and method, Atmospheric Chemistry and Physics, 19, 1767–1783, doi:10.5194/acp-19-1767-2019, 2019.

Prather, M. J., Zhu, X., Flynn, C. M., Strode, S. A., Rodriguez, J. M., Steenrod, S. D., Liu, J., Lamarque, J. F., Fiore, A. M., Horowitz, L. W., Mao, J., Murray, L. T., Shindell, D. T. and Wofsy, S. C.: Global atmospheric chemistry - Which air matters, Atmospheric Chemistry and Physics, 17, 9081–9102, doi:10.5194/acp-17-9081-2017, 2017.

Ray, E. A., Moore, F. L., Elkins, J. W., Dutton, G. S., Fahey, D. W., Vömel, H., Oltmans, S. J. and Rosenlof, K. H.: Transport into the Northern Hemisphere lowermost stratosphere revealed by in situ tracer measurements, Journal of Geophysical Research Atmospheres, 104(D21), 26565–26580, doi:10.1029/1999JD900323, 1999.

Ray, E. A., Moore, F. L., Rosenlof, K. H., Davis, S. M., Boenisch, H., Morgenstern, O., Smale, D., Rozanov, E., Hegglin, M., Pitari, G., Mancini, E., Braesicke, P., Butchart, N., Hardiman, S., Li, F., Shibata, K. and Plummer, D. A.: Evidence for changes in stratospheric transport and mixing over the past three decades based on multiple data sets and tropical leaky pipe analysis, Journal of Geophysical Research Atmospheres, 115(D21304), 1–16, doi:10.1029/2010JD014206, 2010.

Ray, E. A., Moore, F. L., Rosenlof, K. H., Davis, S. M., Sweeney, C., Tans, P., Wang, T., Elkins, J. W., Bönisch, H., Engel, A., Sugawara, S., Nakazawa, T. and Aoki, S.: Improving stratospheric transport trend analysis based on SF6 and CO2 measurements, Journal of Geophysical Research: Atmospheres, 119, 14110–14128, doi:10.1002/2014JD021802, 2014.

Reid, R. C., Prausnitz, J. M. and Poling, B. E.: The properties of gases and liquids, 4th ed., edited by B. Sun and G. H. Fleck, McGraw-Hill, New York., 1987.

Ritz, S. P., Stocker, T. F. and Severinghaus, J. P.: Noble gases as proxies of mean ocean temperature: Sensitivity studies using a climate model of reduced complexity, Quaternary Science Reviews, 30, 3728–3741, doi:10.1016/j.quascirev.2011.09.021, 2011.

Salby, M. L. and Callaghan, P. F.: Influence of the Brewer-Dobson circulation on stratosphere-troposphere exchange, Journal of Geophysical Research Atmospheres, 111(21), 1–9, doi:10.1029/2006JD007051, 2006.

Santoni, G. W., Daube, B. C., Kort, E. A., Jiménez, R., Park, S., Pittman, J. V., Gottlieb, E., Xiang, B., Zahniser, M. S., Nelson, D. D., Mcmanus, J. B., Peischl, J., Ryerson, T. B., Holloway, S., Andrews, A. E., Sweeney, C., Hall, B., Hintsa, E. J., Moore, F. L., Elkins, J. W., Hurst, D. F., Stephens, B. B., Bent, J. and Wofsy, S. C.: Evaluation of the airborne quantum cascade laser spectrometer (QCLS) measurements of the carbon and greenhouse gas suite - CO2, CH4, N2O, and CO - during the CalNex and HIPPO Campaigns, Atmospheric Measurement Techniques, 7, 1509–1526, doi:10.5194/amt-7-1509-2014, 2014.

Schoeberl, M. R., Douglass, A. R., Polansky, B., Bonne, C., Walker, K. A. and Bernath, P.: Estimation of stratospheric age spectrum from chemical tracers, Journal of Geophysical Research Atmospheres, 110(D21303), 1–18, doi:10.1029/2005JD006125, 2005.

Seltzer, A. M., Ng, J., Danskin, W. R., Kulongoski, J. T., Gannon, R. S., Stute, M. and Severinghaus, J. P.: Deglacial water-table decline in Southern California recorded by noble gas isotopes, Nature Communications, 10, 1–6, doi:10.1038/s41467-019-13693-2, 2019.

Škerlak, B., Sprenger, M. and Wernli, H.: A global climatology of stratosphere-troposphere exchange using the ERA-Interim data set from 1979 to 2011, Atmospheric Chemistry and Physics, 14, 913–937, doi:10.5194/acp-14-913-2014, 2014.

Stephens, B. B.: ORCAS Merge Products. Version 1.0, , doi:10.5065/D6SB445X, 2017.

Stephens, B. B., Long, M. C., Keeling, R. F., Kort, E. A., Sweeney, C., Apel, E. C., Atlas, E. L., Beaton, S., Bent, J. D., Blake, N. J., Bresch, J. F., Casey, J., Daube, B. C., Diao, M., Diaz, E., Dierssen, H., Donets, V., Gao, B. C., Gierach, M., Green, R., Haag, J., Hayman, M., Hills, A. J., Hoecker-Martínez, M. S., Honomichl, S. B., Hornbrook, R. S., Jensen, J. B., Li, R. R., McCubbin, I., McKain, K., Morgan, E. J., Nolte, S., Powers, J. G., Rainwater, B., Randolph, K., Reeves, M., Schauffler, S. M., Smith, K., Smith, M., Stith, J., Stossmeister, G., Toohey, D. W. and Watt, A. S.: The O2/N2 ratio and CO2 airborne Southern Ocean study, Bulletin of the American Meteorological Society, 99, 381–402, doi:10.1175/BAMS-D-16-0206.1, 2018.

Strahan, S. E., Douglass, A. R., Stolarski, R. S., Akiyoshi, H., Bekki, S., Braesicke, P., Butchart, N., Chipperfield, M. P., Cugnet, D., Dhomse, S., Frith, S. M., Gettelman, A., Hardiman, S. C., Kinnison, D. E., Lamarque, J. F., Mancini, E., Marchand, M., Michou, M., Morgenstern, O., Nakamura, T., Olivié, D., Pawson, S., Pitari, G., Plummer, D. A., Pyle, J. A., Scinocca, J. F., Shepherd, T. G., Shibata, K., Smale, D., Teyssèdre, H., Tian, W. and Yamashita, Y.: Using transport diagnostics to understand chemistry climate model ozone simulations, Journal of Geophysical Research Atmospheres, 116(17), 1–18,

doi:10.1029/2010JD015360, 2011.

Sugawara, S., Ishidoya, S., Aoki, S., Morimoto, S., Nakazawa, T., Toyoda, S., Inai, Y., Hasebe, F., Ikeda, C., Honda, H., Goto, D. and Putri, F. A.: Age and gravitational separation of the stratospheric air over Indonesia, Atmospheric Chemistry and 700 Physics, 18, 1819–1833, doi:10.5194/acp-18-1819-2018, 2018.

Taylor, R. and Krishna, R.: Multicomponent mass transfer, 2nd ed., John Wiley & Sons., 1993.

Waldmann, L.: Die Temperaturerscheinungen bei der Diffusion in ruhenden Gasen und ihre meßtechnische Anwendung, Zeitschrift für Physik, 124(2), 2–29, doi:10.1007/BF01374919, 1947.

Waugh, D.: Atmospheric dynamics: The age of stratospheric air, Nature Geoscience, 2, 14–16, doi:10.1038/ngeo397, 2009.

705 Waugh, D. W. and Hall, T. M.: Age of stratospheric air: Theory, observations, and models, Reviews of Geophysics, 40(4), doi:10.1029/2000RG000101, 2002.

Wofsy, S. C., the HIPPO science team, cooporating modelers and satellite teams: HIAPER Pole-to-Pole Observations (HIPPO): Fine-grained, global-scale measurements of climatically important atmospheric gases and aerosols, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 369, 2073–2086, 710 doi:10.1098/rsta.2010.0313, 2011.

Wofsy, S. C., Daube, B. C., Jimenez, R., Kort, E., Pittman, J. V., Park, S., Commane, R., Xiang, B., Santoni, G., Jacob, D., Fisher, J., Pickett-Heaps, C., Wang, H., Wecht, K., Wang, Q.-Q., Stephens, B. B., Shertz, S., Watt, A. S., Romashkin, P., Campos, T., Haggerty, J., Cooper, W. A., Rogers, D., Beaton, S., Hendershot, R., Elkins, J. W., Fahey, D. W., Gao, R. S., Moore, F., Montzka, S. A., Schwarz, J. P., Perring, A. E., Hurst, D., Miller, B. R., Sweeney, C., Oltmans, S., Nance, D., Hintsa, 715 E., Dutton, G., Watts, L. A., Spackman, J. R., Rosenlof, K. H., Ray, E. A., Hall, B., Zondlo, M. A., Diao, M., Keeling, R., Bent, J., Atlas, E. L., Lueb, R. and Mahoney, M. J.: HIPPO MEDUSA Flask Sample Trace Gas and Isotope Data (R_20121129), , doi:10.3334/CDIAC/HIPPO_014, 2012.

Wofsy, S. C., Afshar, S., Allen, H. M., Apel, E., Asher, E. C., Barletta, B., Bent, J., Bian, H., Biggs, B. C., Blake, D. R., Blake, N., Bourgeois, I., Brock, C. A., Brune, W. H., Budney, J. W., Bui, T. P., Butler, A., Campuzano-Jost, P., Chang, C. S., Chin, 720 M., Commane, R., Correa, G., Crounse, J. D., Cullis, P. D., Daube, B. C., Day, D. A., Dean-Day, J. M., Dibb, J. E., Digangi, J. P., Diskin, G. S., Dollner, M., Elkins, J. W., Erdesz, F., Fiore, A. M., Flynn, C. M., Froyd, K., Gesler, D. W., Hall, S. R., Hanisco, T. F., Hannun, R. A., Hills, A. J., Hintsa, E. J., Hoffman, A., Hornbrook, R. S., Huey, L. G., Hughes, S., Jimenez, J. L., Johnson, B. J., Katich, J. M., Keeling, R., Kim, M. J., Kupc, A., Lait, L. R., Lamarque, J. F., Liu, J., McKain, K., McLaughlin, R. J., Meinardi, S., Miller, D. O., Montzka, S. A., Moore, F. L., Morgan, E. J., Murphy, D. M., Murray, L. T., 725 Nault, B. A., Neuman, J. A., Newman, P. A., Nicely, J. M., Pan, X., Paplawsky, W., Peischl, J., Prather, M. J., Price, D. J., Ray, E., Reeves, J. M., Richardson, M., Rollins, A. W., Rosenlof, K. H., Ryerson, T. B., Scheuer, E., Schill, G. P., Schroder, J. C., Schwarz, J. P., St.Clair, J. M., Steenrod, S. D., Stephens, B. B., Strode, S. A., Sweeney, C., Tanner, D., Teng, A. P.,

27

Thames, A. B., Thompson, C. R., Ullmann, K., Veres, P. R., Vizenor, N., Wagner, N. L., Watt, A., Weber, R., Weinzierl, B., et al.: ATom: Merged Atmospheric Chemistry, Trace Gases, and Aerosols, , doi:10.3334/ornldaac/1581, 2018.

730   Zondlo, M. A., Paige, M. E., Massick, S. M. and Silver, J. A.: Vertical cavity laser hygrometer for the National Science Foundation Gulfstream-V aircraft, Journal of Geophysical Research Atmospheres, 115(D20309), 1–14, doi:10.1029/2010JD014445, 2010.
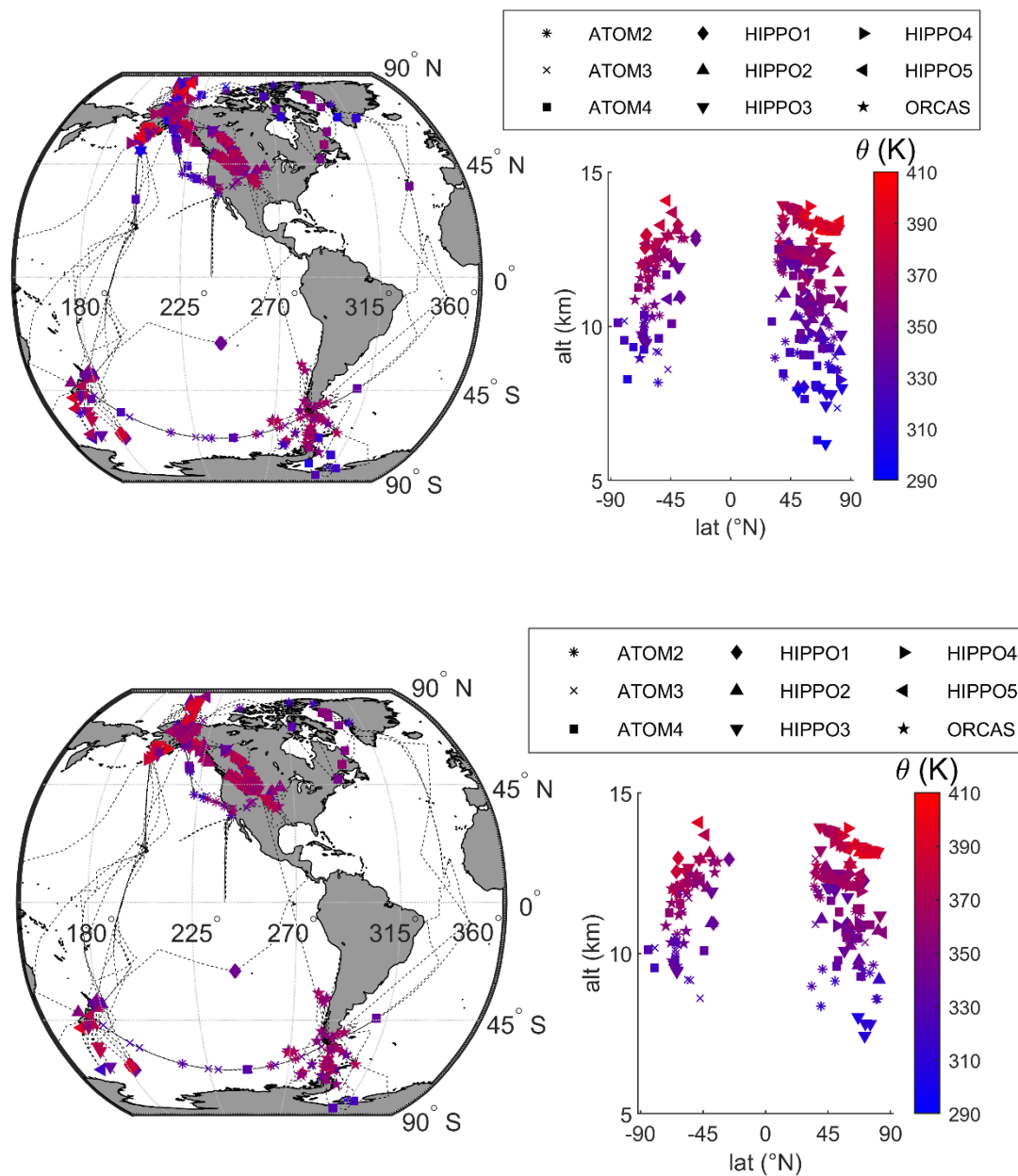
Figure 1. Horizontal (left) and vertical (right) distribution of airborne flask sample locations identified as being of stratospheric origin (see text). Thin dashed black lines on the map illustrate the flight tracks of all 9 campaigns. Symbols indicate the campaigns during which the stratospheric samples were collected, and colours show the potential temperature at which the sample was taken.

29
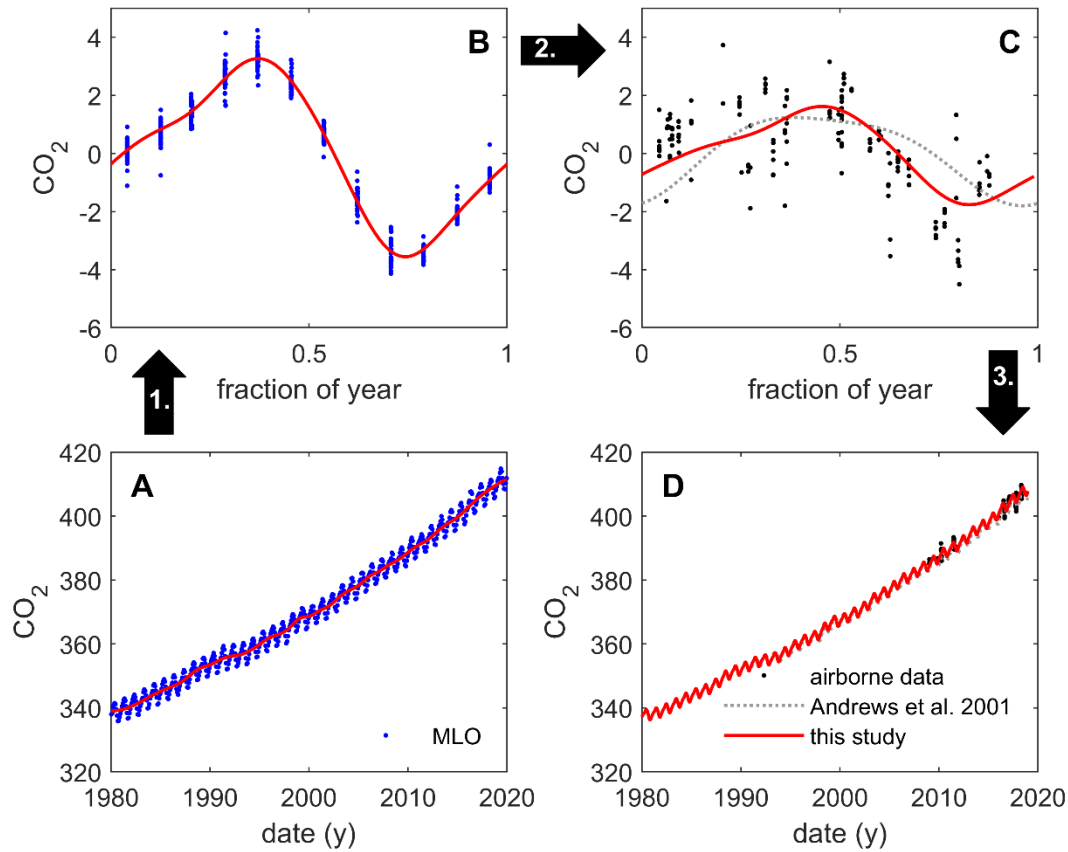
740 **Figure 2. Illustration of the major steps in deriving a tropical upper tropospheric reference time series. Panel (a) shows the monthly mean surface CO_2 concentrations at Mauna Loa (MLO) (Keeling et al., 2001) with a stiff spline trend shown in red. Panel (b) shows a fit of the mean seasonal cycle at MLO compared to detrended observations from 1980–2019. Panel (c) shows the 1-month-lagged seasonal cycle (red line) rescaled in amplitude to match detrended airborne observations in the equatorial upper troposphere (20°S< lat < 20°N, alt >8 km, black). The seasonal cycle derived by Andrews et al. (2001) is presented as a dotted grey line for reference.**
745 **Panel (d) shows the resulting upper tropospheric mean reference time series (TRTS, red) used for CO_2 in the age of air algorithm with the airborne observations from the tropical upper troposphere (black dots) and the time series (grey dotted line) of Andrews et al. (2001).**

750
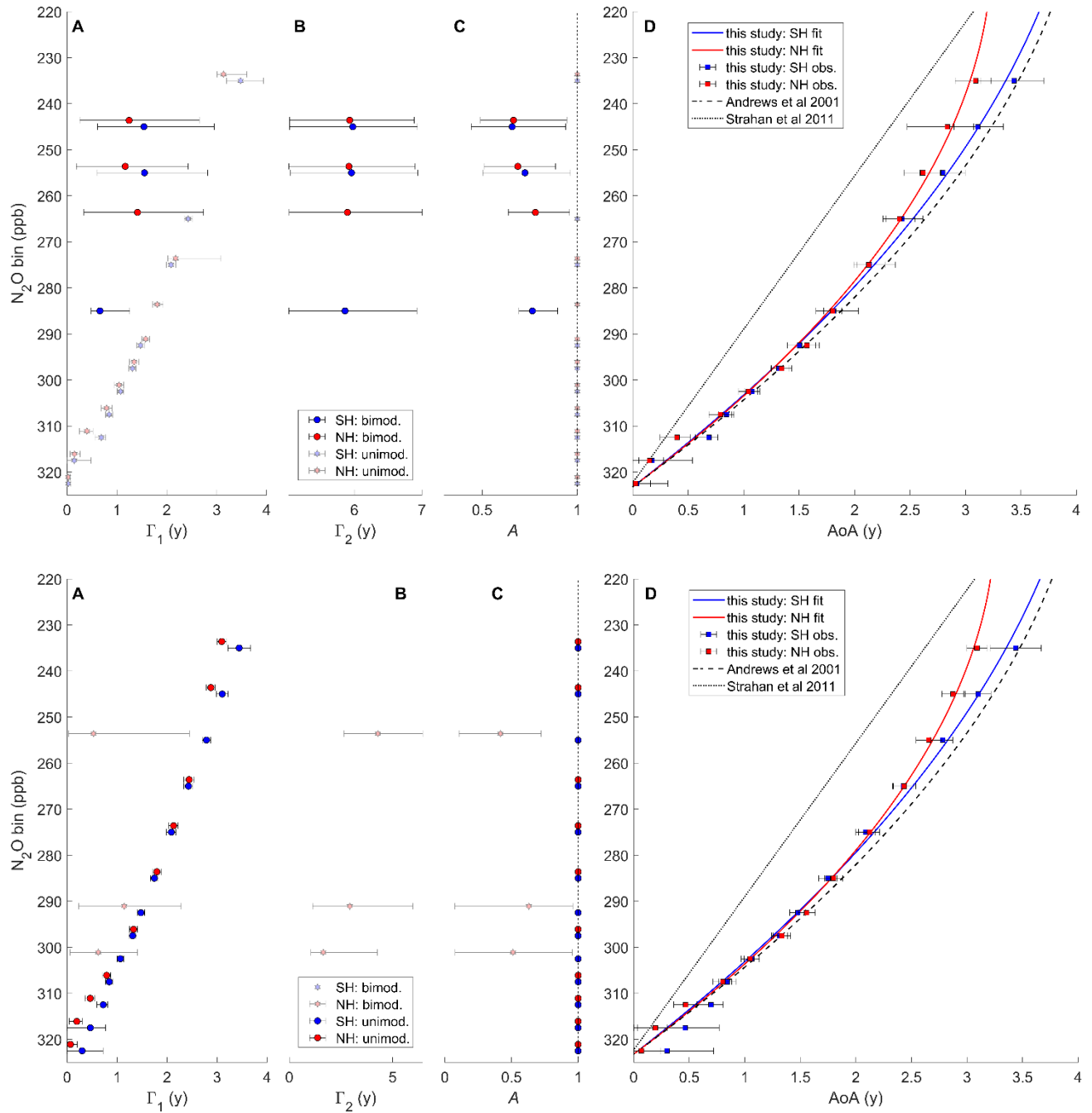
Figure 3. Hemisphere-specific properties of the age spectra in different N$_2$O bins from the 10s-averaged airborne observations estimated by Markov Chain Monte Carlo. Panels (Aa)-(Cc) show the value and 95% confidence interval of Γ_1, Γ_2, and A in each bin for either a unimodal (faint blue and faint red starscircles) or bimodal (faint blue and faint red circlesstars) age spectrum depending

755 on which type was preferentially selected by the algorithm. N$_2$O values are offset by 1.4 ppb for all Northern Hemisphere estimates (NH, red) for visual clarity in panel ~~(A~~a~~)~~-~~(C~~c~~)~~. Panel ~~(D~~d~~)~~ shows the mean AoA and confidence interval of the age spectra ensemble in each bin (i.e., unimodal and bimodal together). These data are fit by a quadratic polynomial for each hemisphere with a fixed y-intercept (NH AoA = -0.~~0002361~~ 0002381 (323.23 - N$_2$O)^2 + 0.~~05530~~ 05567 (323.23 - N$_2$O); SH AoA = -0.~~0001754~~ 0001723 (323.23 - N$_2$O)^2 + 0.~~05359~~ 05317 (323.23 - N$_2$O)). Previous relationships published by Andrews et al. (2001) and Strahan et al. (2011) are given

760 as a dashed and dotted line for reference. Y-intercept values of the previously published N$_2$O-AoA calibrations have been updated to reflect the gradual increase in tropospheric N$_2$O between 1997 and the new reference year 2009.
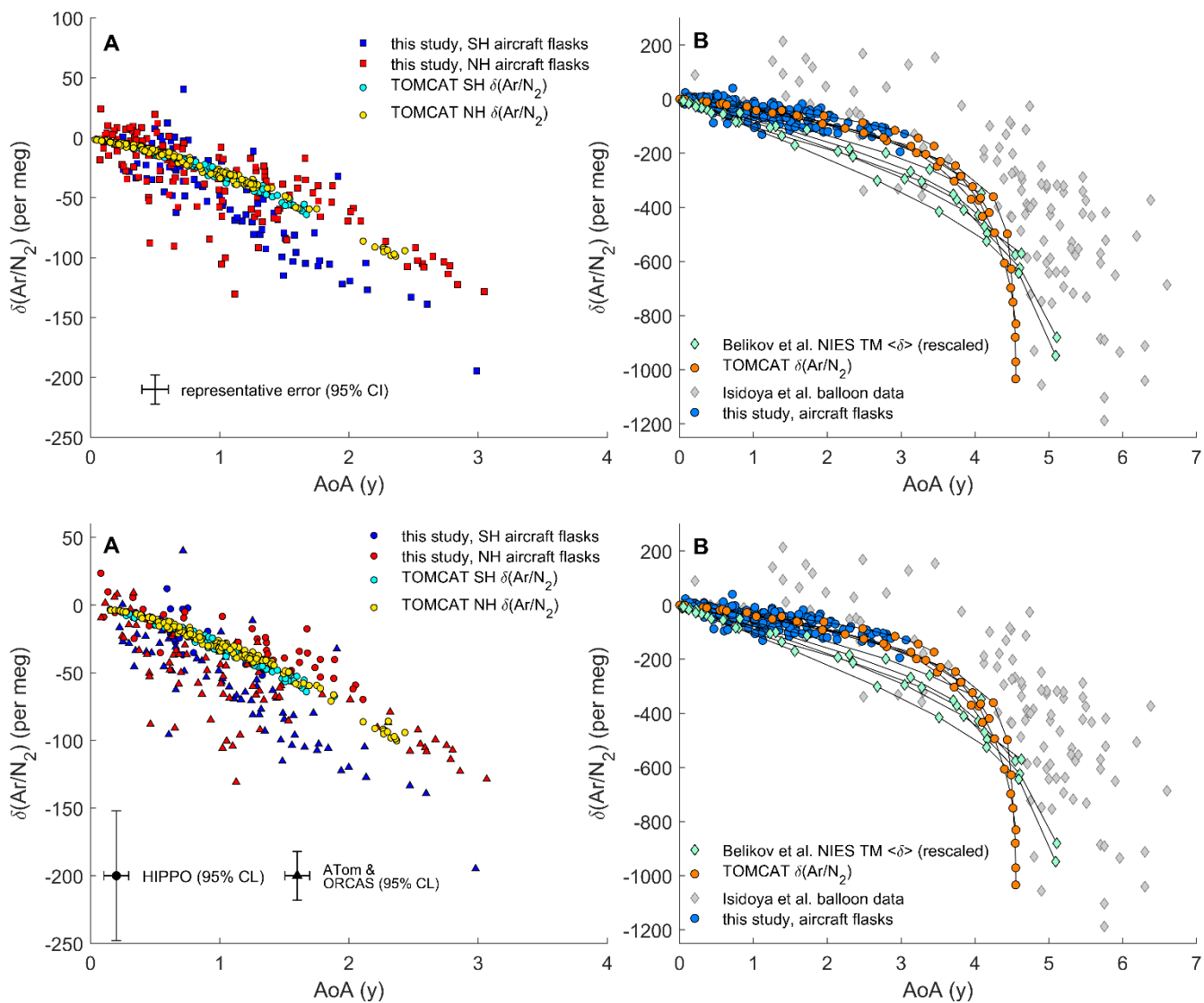
Figure 4. Comparison of age of air (AoA) and gravitational settling (GS) of δ(Ar/N₂) in ~~the~~ models and observations. Panel (a) shows observations from the Northern (red ~~squares~~) and Southern (blue) Hemisphere~~s~~ ~~(blue squares)~~ together with TOMCAT model output (SH: cyan ~~circles~~; NH: yellow ~~circles~~) selected from the closest model grid box in time and space. Because ~~uncertainties~~ uncertainty for ~~δ(Ar/N₂) and~~ AoA ~~are~~ is similar for all stratospheric samples but δ(Ar/N₂) uncertainty differs between campaigns, representative error bars (95% confidence interval) are shown for HIPPO (triangles) and ATom/ORCAS (circles) at an arbitrary AoA ~~of ~ 0.5 years~~. δ(Ar/N₂) is normalized to yield a delta value of zero in the equatorial free troposphere. In panel (b), observations from airborne campaigns (this study) and the balloon sampling system (Ishidoya et al., 2008, 2013, 2018; Sugawara et al., 2018) are plotted with the AoA-GS relationship observed in TOMCAT and the NIES TM (Belikov et al., 2019) as lines for each of the latitude bands shown in Fig. 5 and points for different altitude bins between 10 and 35 km. To yield an equivalent estimate of δ(Ar/N₂), δ($^{13}CO_2$/$^{12}CO_2$) results from the NIES TM (Belikov et al., 2019) have been rescaled according to Eq. (5) and offset by ~10 per meg to account for the different tropospheric reference region in the definition of δ.
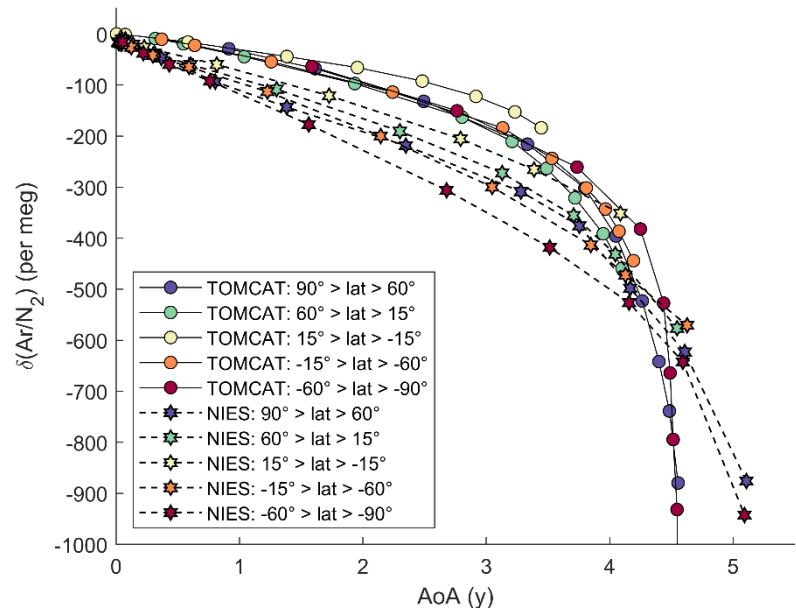
33

Figure 5. Comparison of the annual and zonal mean age of air (AoA) relationship to gravitational settling simulated in TOMCAT and the NIES TM. The relationship is plotted as lines for the latitude bands indicated by marker colour. For the NIES TM, markers show the vertical profile using all grid boxes available between ~10-35 km. For TOMCAT, markers instead correspond to binned altitude bands between 10 and 35 km with a spacing of 2.5 km because of the finer vertical resolution of the model. To yield an equivalent estimate of δ(Ar/N₂), δ($^{13}CO_2$/$^{12}CO_2$) results from the NIES TM (Belikov et al., 2019) have been rescaled according to Eq. (5) and offset by ~10 per meg to account for the different tropospheric reference region in the definition of δ.

790
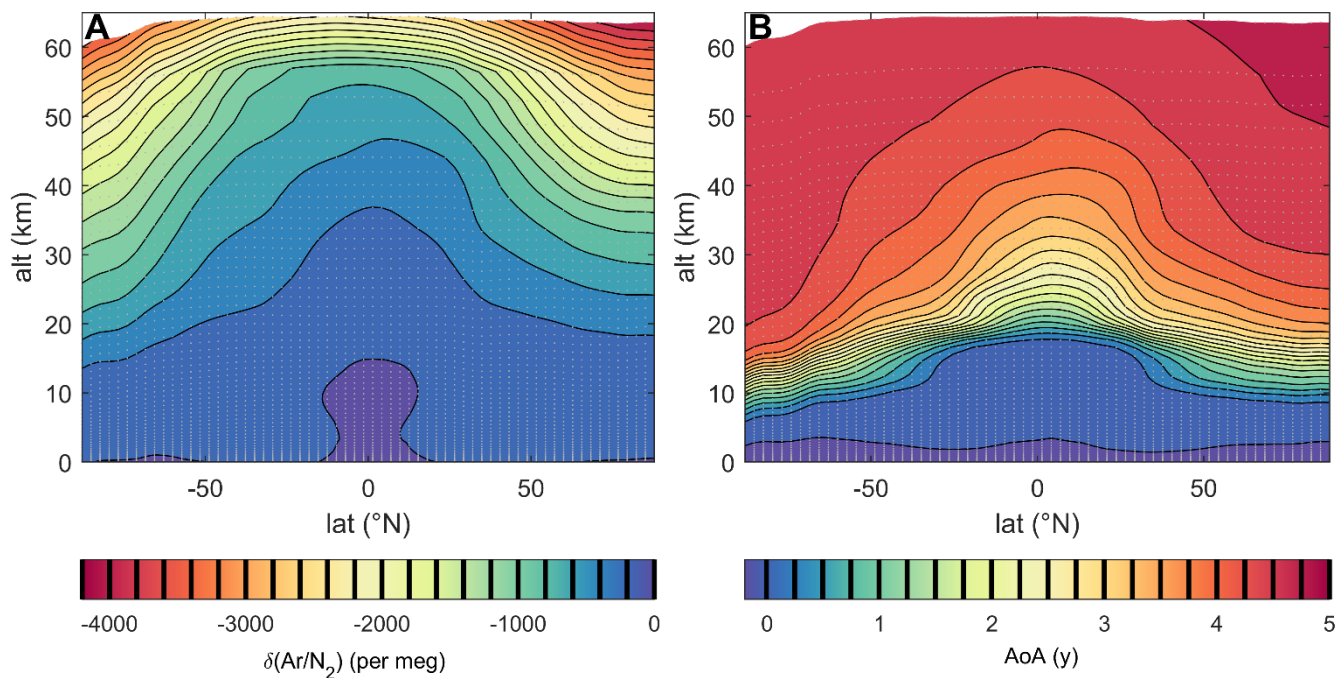
Figure 6. Annual and zonal mean of δ(Ar/N₂) (panel ~~A~~a) and age of air (AoA) (panel ~~B~~b) simulated by TOMCAT. Values of the contour lines are shown as black vertical lines on the colour bar. Grey dots indicate the centre of a grid boxes in TOMCAT. A video of this figure highlighting the natural variability in monthly-mean values is available online with this manuscript.
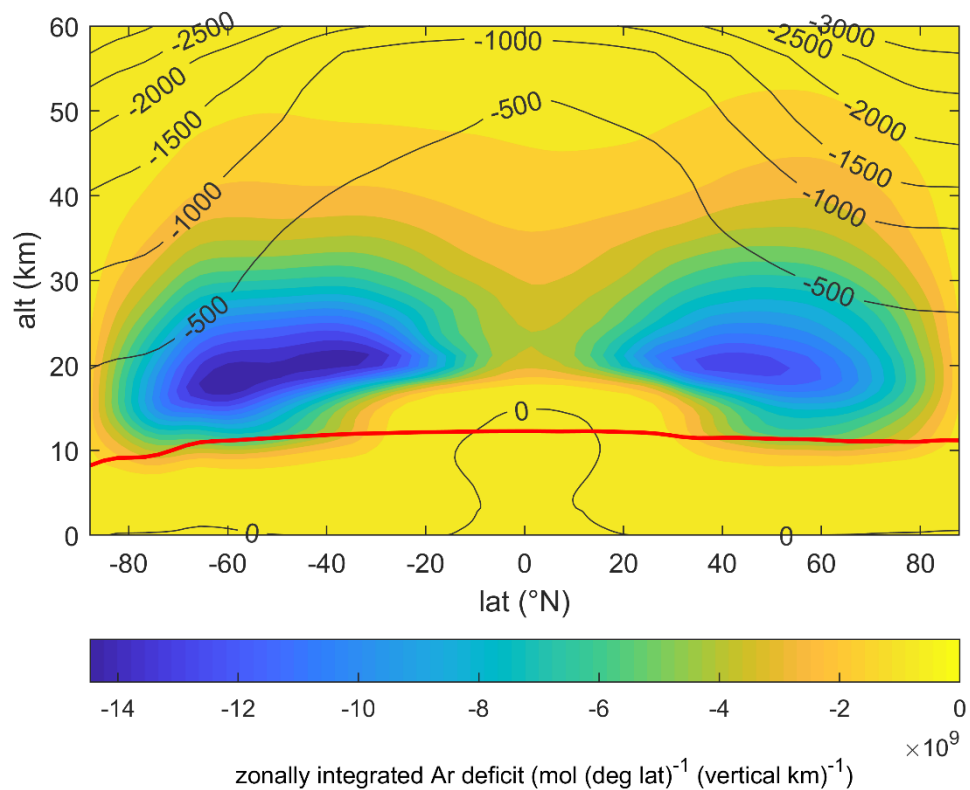
zonally integrated Ar deficit (mol (deg lat)$^{-1}$ (vertical km)$^{-1}$)
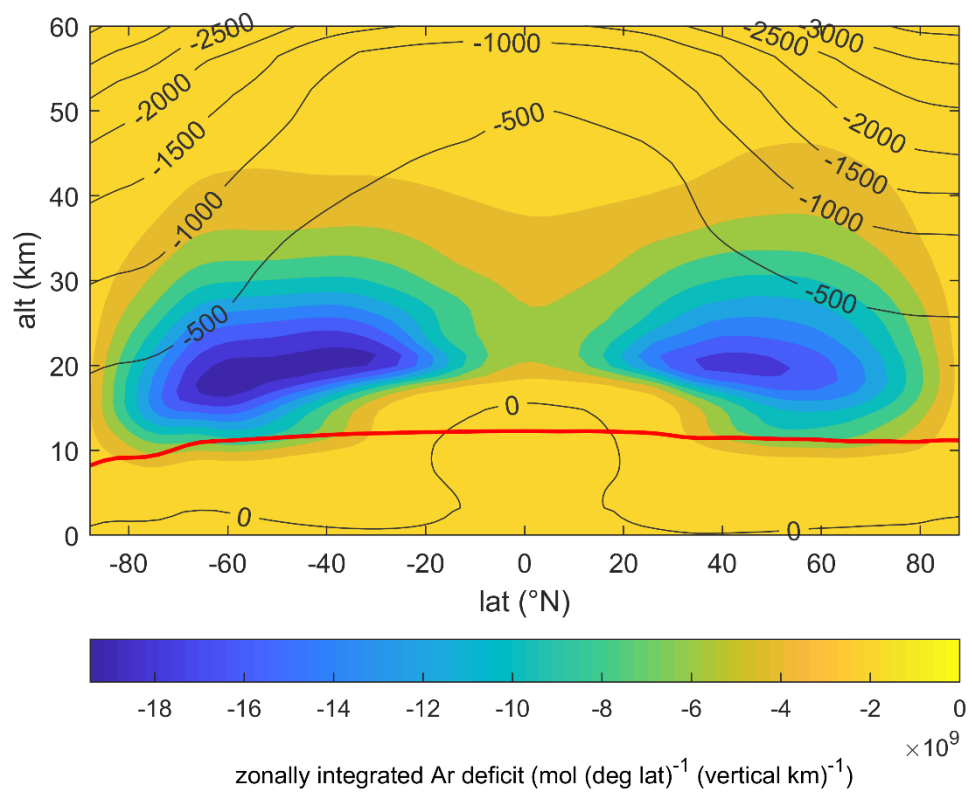
Figure 7. Colormap of annual mean, zonally integrated Ar deficit in TOMCAT. Overlain black contour lines show δ(Ar/N₂) (per meg). The red solid line highlights the mean position of the 200 hPa isobar.

Table BA1. Molecular diffusion volumes (Reid et al., 1987) and masses used in this study

Chemical species	Molecular mass (g mol^{-1})	Molecular diffusion volume
Ar	40	16.2
N_2	28	18.5
O_2	32	16.3
air	28.95	19.7

805

Table ~~B~~A2. Characteristic values of variables in Eqs. (~~B~~A14) and (~~B~~A15)

Variable/term	Magnitude in extratropics at:			units	Notes ~~or~~ or Eq~~.~~. numbe~~r~~
	~~magnitude at~~ ~20 km	~~magnitude at~~ ~35 km	~57 km		
P	50	5	0.3	mb	
T	215	~~250~~235	255	K	3
$\dfrac{\partial T}{\partial z}$	~~3~~5e-04	~~3e~~2.5e-03	2.4e-03	K m^{-1}	3
H_{N2}	6.51e+03	~~7.57~~7.11e+03	7.57e+03	m	
H_{Ar}	4.56e+03	~~5.30~~4.98e+03	5.30e+03	m	
H_{air}	6.29e+03	~~7.32~~6.88e+03	7.32e+03	m	
$\alpha^T_{N2:O2}$		<1.8e-02		-	1
$\alpha^T_{Ar:air}$	~5.6e-02	~6.~~6e~~3e-02	~6.6e-02	-	2
C_{N2}		78.09		%	
C_{O2}		20.95		%	
C_{Ar}		0.93		%	
$\dfrac{\partial C_{N2}}{\partial z}$	3.~~3~~2.0e-09	7.~~2~~2.6e-09	8.7e-09	m^{-1}	3
$\dfrac{\partial C_{Ar}}{\partial z}$	3.~~3~~2.7e-10	7.~~4~~3.4e-10	11.6e-10	m^{-1}	3
D^{air}_{N2}	2.3e-04	~~3.0~~2.7e-03	5.3e-02	m^{2}s^{-1}	
$D^{air}_{Ar\times(N2,O2)}$	9.8e-09	1.~~1~~3e-07	2.2e-06	m^{2}s^{-1}	
D^{air}_{Ar}	2.2e-04	2.~~6~~9e-03	5.0e-02	m^{2}s^{-1}	
$D^{air}_{N2}\dfrac{\partial C_{N2}}{\partial z}$	6.~~3~~4.7e-13	8.~~2~~7.1e-12	4.6e-10	m s^{-1}	A19
$D^{air}_{N2}\left(\dfrac{1}{H_{N2}}-\dfrac{1}{H_{air}}\right)C_{N2}$	9.6e-10	1.~~1~~1.0e-08	1.9e-7	m s^{-1}	A19
$D^{air}_{N2}\alpha^T_{N2:O2}\dfrac{C_{N2}(1-C_{N2})}{T}\dfrac{\partial T}{\partial z}$	<~~1.6~~5.9e-12	<~~1.1~~8.7e-1~~0~~1	≤1.6e-09	m s^{-1}	A19
$D^{air}_{Ar\times(N2,O2)}\dfrac{\partial C_{N2}}{\partial z}$	2.~~6~~2.0e-17	3.~~4~~3.0e-16	1.9e-14	m s^{-1}	A20
$D^{air}_{Ar\times(N2,O2)}\left(\dfrac{1}{H_{N2}}-\dfrac{1}{H_{air}}\right)C_{N2}$	4.0e-14	4.~~3~~5e-13	7.8e-12	m s^{-1}	A20
$D^{air}_{Ar\times(N2,O2)}\alpha^T_{N2:O2}\dfrac{C_{N2}(1-C_{N2})}{T}\dfrac{\partial T}{\partial z}$	<~~6.8~~2.5e-1~~7~~6	<~~3~~4.6e-15	≤6.8e-14	m s^{-1}	A20
$D^{air}_{Ar}\dfrac{\partial C_{Ar}}{\partial z}$	7.~~8~~6.0e-14	9.~~7~~8.9e-13	5.8e-11	m s^{-1}	A20
$D^{air}_{Ar}\left(\dfrac{1}{H_{Ar}}-\dfrac{1}{H_{air}}\right)C_{Ar}$	1.3e-10	1.~~4~~1.3e-09	2.4e-08	m s^{-1}	A20
$D^{air}_{Ar}\alpha^T_{Ar:air}\dfrac{C_{Ar}}{T}\dfrac{\partial T}{\partial z}$	~~2.7~~9.8-13	~~2.1~~1.6e-11	~3.2e-10	m s^{-1}	A20

[1] 1.8e-02 observed at 293 K (Waldmann, 1947)

[2] assuming $C_{Ar} \approx \frac{N_{Ar}}{N_{N2}}$

[3] estimated from TOMCAT results ~~and observations~~

Supplementary information to: Gravitational separation of Ar/N₂ and age of air in the lowermost stratosphere in airborne observations and a chemical transport model

Benjamin Birner[1], Martyn P. Chipperfield[2,3], Eric J. Morgan[1], Britton B. Stephens[4], Marianna Linz[5], Wuhu Feng[2,6], Chris Wilson[2,3], Jonathan D. Bent[1,7], Steven C. Wofsy[5], Jeffrey Severinghaus[1], Ralph F. Keeling[1]

[1]Scripps Institution of Oceanography, UC San Diego, La Jolla, CA 92093, USA
[2]School of Earth and Environment, University of Leeds, Leeds, LS2 9JT, UK
[3]National Centre for Earth Observation, University of Leeds, Leeds, LS2 9JT, UK
[4]National Center for Atmospheric Research, Boulder, CO 80301, USA
[5]Department of Earth and Planetary Sciences, and School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA
[6]National Centre for Atmospheric Science, University of Leeds, Leeds, LS2 9JT, UK
[7]now at: Picarro, Inc., Santa Clara, CA 95054, USA

Correspondence to: Benjamin Birner (bbirner@ucsd.edu)

δ(Ar/N₂) residual analysis

To test the hypothesis that large artifacts or biases in δ(Ar/N₂) were introduced during flask sample collection, we look for the presence of a pressure dependency of δ(Ar/N₂) after removing what we have shown is the natural gravitational signal. Pressure serves as a proxy for sampling conditions and aircraft velocity simultaneously here because aircraft velocity also increases systematically with altitude or pressure altitude. A vertical profile of δ(Ar/N₂) (Fig. S1A) reveals no strong dependency on pressure in the troposphere below about 300 mb. Above ~300 mb, some samples represent stratospheric conditions and therefore have an age of air (AoA) greater than zero and more negative δ(Ar/N₂). For further analysis, we correct δ(Ar/N₂) for the natural gravitational signal ($\delta(Ar/N_2)_{grav}$) using observed AoA from each sample and a quadratic fit of the AoA-δ(Ar/N₂) relationship simulated in TOMCAT ($\delta(Ar/N_2)_{grav} = -8.124\,AoA^2 - 22.73\,AoA + 0.9529$). The result is shown in Figure S1B. Although there is scatter in the corrected data, in particular for HIPPO1−5, no systematic relationship between pressure and gravitationally-corrected δ(Ar/N₂) is evident, supporting the interpretation that aircraft sampling did not create large biases in δ(Ar/N₂). While there is some evidence of a pressure dependence in the ORCAS data from the Southern Hemisphere (SH), this could also be the result of an incomplete correction for gravitational fractionation if there are differences in the true AoA-δ(Ar/N₂) relationship for the SH not captured by TOMCAT. However, stratospheric samples from ORCAS were almost exclusively taken at the same pressure, so our analysis of the δ(Ar/N₂)-AoA relationship should be robust regardless.

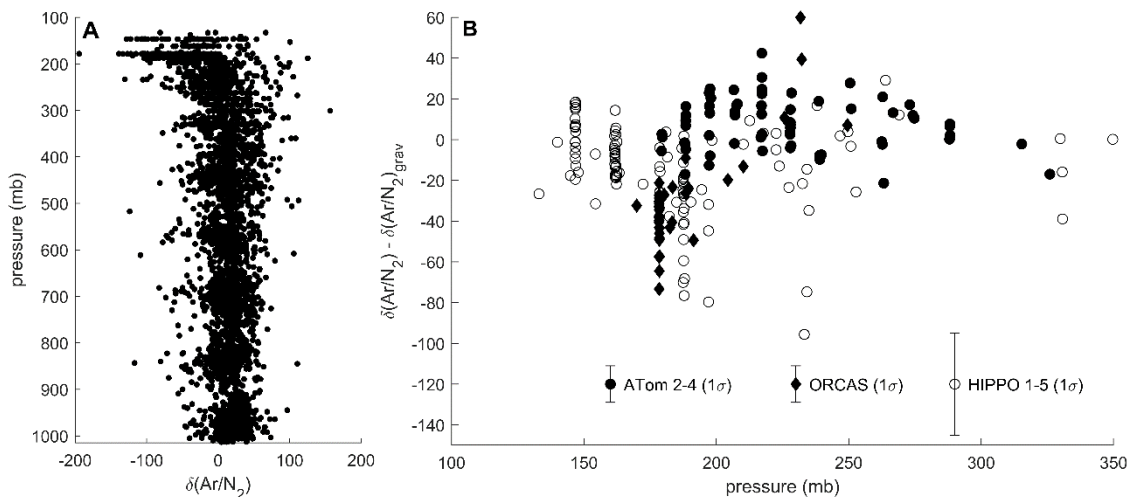**Figure S1. Pressure relationship of Medusa flask δ(Ar/N₂) (A) and residual δ(Ar/N₂) (B), i.e., δ(Ar/N₂) corrected for the influence of gravity in the stratosphere (see text). Data from HIPPO (open circles) and ATom (solid circles), and ORCAS (solid diamonds) are shown separately in (B) with their respective error bars.**
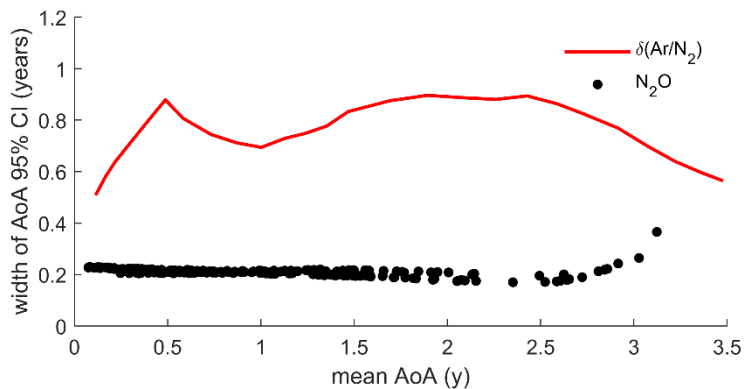

~~**Figure S1. Width of the 95% confidence interval for AoA estimates from N₂O based on the Markov chain Monte Carlo algorithm (points, see methods) and from δ(Ar/N₂) using the AoA-δ(Ar/N₂) relationship in TOMCAT (red line).**~~

35

40

2

Figure S2. Seasonal cycle of δ(Ar/N₂) in the topical ~~upper~~ <u>free</u> troposphere (20°>lat>-20° & altitude ~~>6~~<u>3-8</u> km).
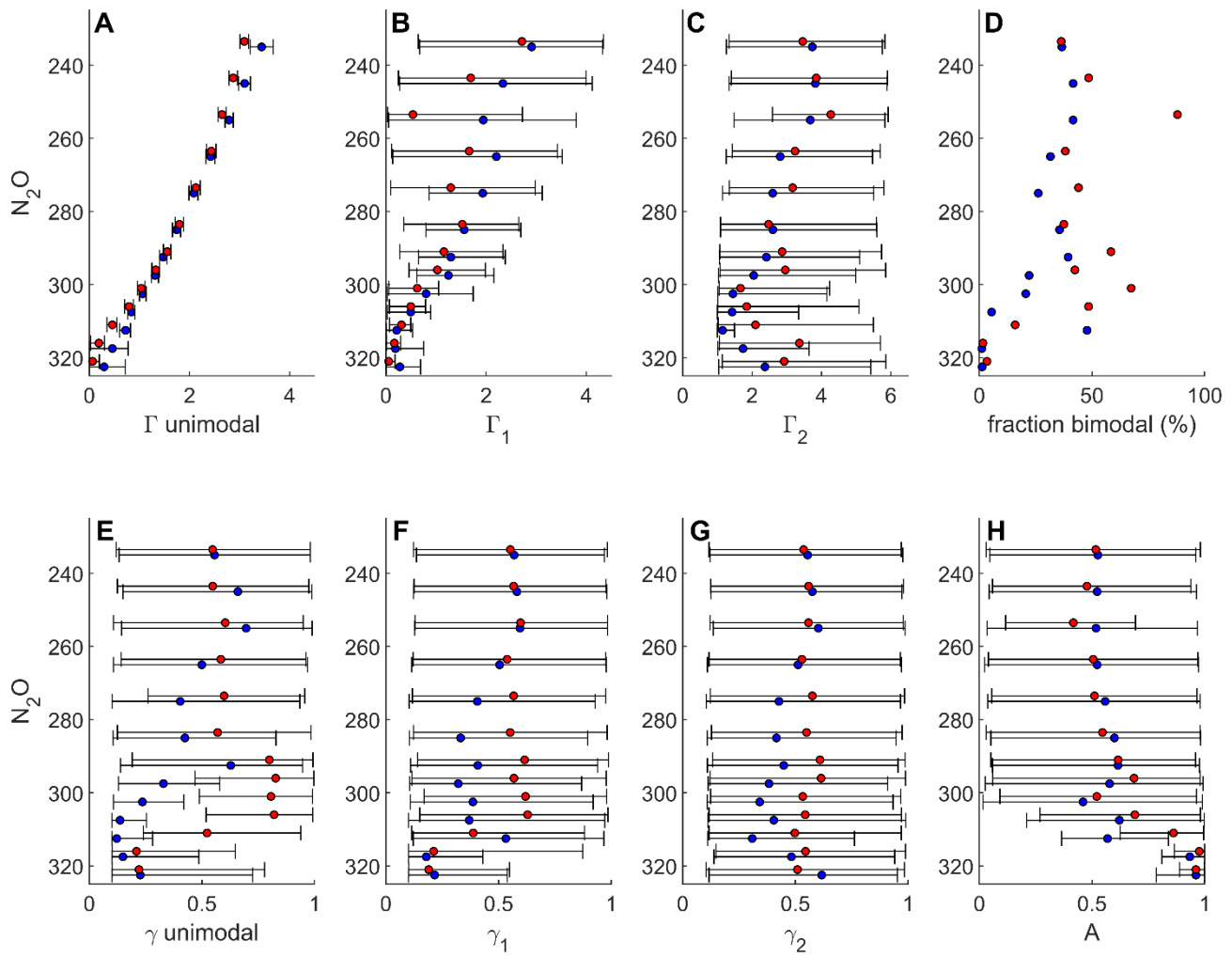
45 **Figure S3. Age spectrum parameter values selected by the Markov chain Monte Carlo algorithm in each N$_2$O bin for the Southern (blue) and Northern (red) Hemispheres with error bars showing 95% confidence interval. NH N$_2$O bin values were adjusted by -1.5 ppb here for visual clarity only. The mean AoA is well constrained but the algorithm fails to clearly distinguish between unimodal and bimodal age spectra and places poor constraints on the width of unimodal and bimodal age spectra.**
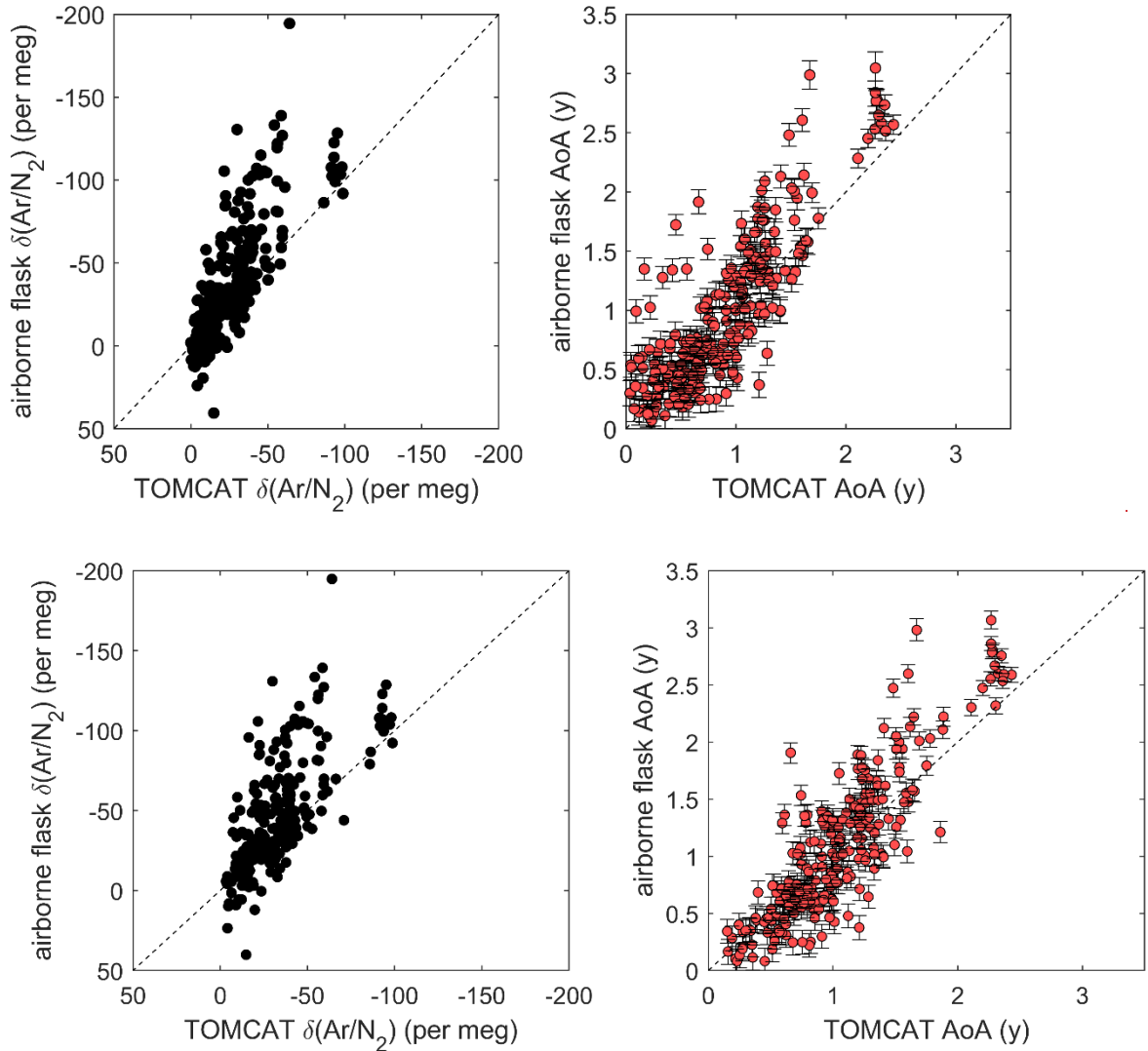
Figure S3S4. Comparison of ~~AoA and~~ δ(Ar/N₂) (left) and AoA (right) ~~in~~between stratospheric HIPPO, AToM and ORCAS observations and TOMCAT ~~to the airborne campaign data~~ in the lowermost stratosphere. TOMCAT grid cells are chosen to be closest in space and~~in~~ time to the observations, but unresolved mixing processes ~~sometimes~~ can allow more fractionated/older air to be present in the observations than can be reproduced by TOMCAT. ~~Note that these~~These deviations from the 1:1-line are correlated between δ(Ar/N₂) and AoA, and therefore will partially cancel ~~when~~in ~~evaluating~~ the AoA-δ(Ar/N₂) ~~GS~~ relationship (see Fig. 5 in main paper).
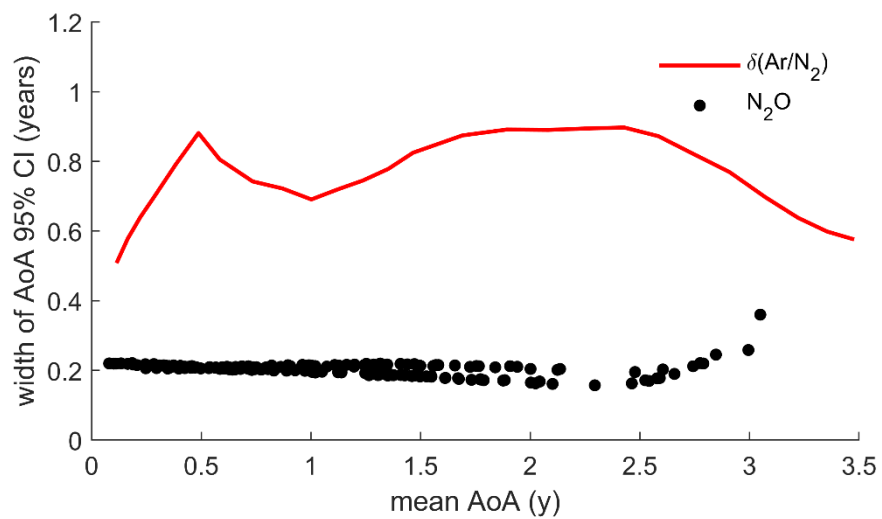
5

Figure S5. Width of the 95% confidence interval for AoA estimates from N_2O based on the Markov chain Monte Carlo algorithm (points) and from $\delta(Ar/N_2)$ using the AoA-$\delta(Ar/N_2)$ relationship in TOMCAT (red line).

60