We would like to thank the reviewer for their comprehensive comments. They have helped greatly in improving the paper. Reviewer comments are in black. Author responses are in red. Author response line numbers correspond to the tracked changes document.

**Reviewer 1**

**Summary**

This work looks to constrain the variation in the $H_2$ budget by using 3 box models to represent northern, tropics and southern regions. They compare observations and a forward box model to a new Bayesian inversion model, which optimises OH, HFC-152a emissions, $H_2$ emissions, $H_2$ chemical production, and the $H_2$ soil sink. HCF-152a observations are used alongside $H_2$ observations to constrain the seasonality of OH, and therefore improve the OH estimate, which feeds into the $H_2$ budget. Further to this, they use the optimised OH to estimate HCHO production, which in turn is used to calculate $H_2$ chemical production. They run four scenarios with different uncertainty ranges for the $H_2$ emission and chemical production. In the final uncertainty scenario, they estimate HCHO photolysis from the retrieved OH using a pseudo-linear relationship.

The authors found that the retrieved model performed better in the northern hemisphere and were able to create a seasonal variability of soil deposition which was in agreement with other GCMs with no prior assumptions for the soil sink seasonality. They found a larger uncertainty of soil deposition in the tropical region with reduced seasonality, and there was a minimal impact on soil deposition in southern region as expected due to limited land. While they are unable to provide absolute values for $H_2$ sources and sinks due to the high uncertainty in HFC-152a emissions, they are able to determine the seasonal cycle and amplitude of these processes, which can be used to compare against other interactive hydrogen schemes.

**General Comments**

This work is a novel approach to constraining OH by using both $H_2$ and HFC-152a observations. The retrieved seasonality of the soil sink, particularly in the north box, is of great interest, as no physical assumptions of the soil uptake are made.

As the authors point out, they are not able to retrieve absolute values due to not being able to constrain absolute values of OH. As a result of this, I think the title of the paper could be misleading as this isn't an absolute constraining of the hydrogen sinks, but rather their seasonality and amplitude.

I think some further analysis could be drawn out from these results. I've given some more details in the comments below, but can the authors e.g. provide an intercomparison between boxes to establish percentage contributions of different sinks relative to other boxes? This would give a useful weight to how important sources/sinks of H2 are in different regions. Also, I think there could be more of a comparison with other literature, and authors could expand on what is meant when there results are in agreement with other studies.

It would help if the authors made more use of their figures and
referenced them throughout the text for clarity (see technical comments)

Thank you for your comments.

**Technical Comments**

68 : Can you provide more information on the destruction of HFC-152a (e.g. a percentage estimate of OH destruction)

Line 72. We have amended the following sentence to include an estimated percentage of destruction by OH: “The main loss pathway of HFC-152a is through OH oxidation, with a short OH lifetime of 1.55 years” to “The main loss pathway of HFC-152a, accounting for over 99% of the loss, is through OH oxidation, with a short OH lifetime of 1.55 years”.

70-74 : It’s difficult to follow why the authors cannot retrieve absolute values. They later state in line 117 that they scale the OH tropospheric concentration to 1e6 mole cm-3. Please can this sentence be rephrased/broken up to describe this clearer.

Presenting the results as absolute values will unfortunately compound the errors of the overall $H_2$ budget because the retrieval of the absolute values of OH are limited by poorly constrained emissions of HFC-152a. This is described on line 74. For example, emissions of MCF are relatively well known, which previously allowed for a more accurate retrieval of absolute OH. We do still retrieve absolute values hence our prior OH estimate of 1e6, however, we have just chosen to present seasonal anomalies as the seasonality of OH is much better constrained by HFC-152a emissions. We have cleaned up the description to try to make this clearer.

Line 77. Added in: “the yearly range and phase of OH to be retrieved in this study, presented as…”

Line 89. Added in: “presented as seasonal anomalies”

71 : Could you briefly explain the impact of HFC-152a emissions on constraining first order loss via OH (i.e. [HFC-152a ] = Production / k[OH]) to make the reasoning clearer.

We have adjusted some wording to make it clear that emissions are the only source and that since OH is virtually the only loss, uncertainty in the emissions will be reflected in retrieval of OH.

Line 74: Added in “. Anthropogenic emissions are the only source of HFC-152a and…”

Line 75: Added in: “Since OH is virtually the only sink, the uncertainties in emissions will…”

72 : In general, could you provide more information on HFC-152a? Is there expected to be a N-S hemisphere gradient given the short lifetime? What's the global surface average condition and is much of a seasonal variation?

We have included an additional introductory figure that helps introduce HFC-152a and $H_2$. Using this figure (Fig. R1 below), we provide information on the seasonal cycle of HFC-152a, the hemispherical gradient, and on the differences in seasonality between HFC-152a and $H_2$. Added in the following discussion in the introduction and Section 3.1

Line 78-85. Added: "$H_2$ has observably different seasonality compared to HFC-152a in both the NH and SH. Therefore, the OH seasonal range and phase information obtained from HFC-152a can be used to help constrain the yearly ranges (maximum – minimum) and phase of the $H_2$ soil sink. This is highlighted in Fig. 1. At the Advanced Global Atmospheric Gases Experiment (AGAGE) (Prinn et al., 2025, 2018) site Mace Head (MHD, Ireland, 53.3° N) (Fig. 1a). HFC-152a exhibits a seasonal peak 1-2 months earlier than $H_2$, suggesting that the soil sink is likely has a different phase than OH and $CH_2O$ and is the dominant driver of the seasonality with a phase difference of 1-2 months. At the AGAGE site Kennaook/Cape Grim (CGO, Australia, 40.7° S) (Fig. 1b), HFC-152a and $H_2$ are completely out of phase, inferring that OH and $CH_2O$ are the dominant drivers of $H_2$ seasonality."

Line 143-144: Added in: "Figure 1 shows a steep gradient in HFC-152a volume mixing ratio between the NH and the SH. Therefore, our choice of SMO for the Tropics box will likely be an underestimation of the tropical average."

Line 291-294. Added: "In the North box, the forward model mole fractions of $H_2$ show very different seasonality compared to the observations (Fig. 3d). Since the initial forward model has a constant soil sink, it does not affect the seasonal cycle. Therefore, OH, chemical production, and $H_2$ emissions control the modeled seasonality, causing the large difference in phase and suggesting that the soil sink is likely the dominant driver of the observed seasonality (also see Fig. 1)."

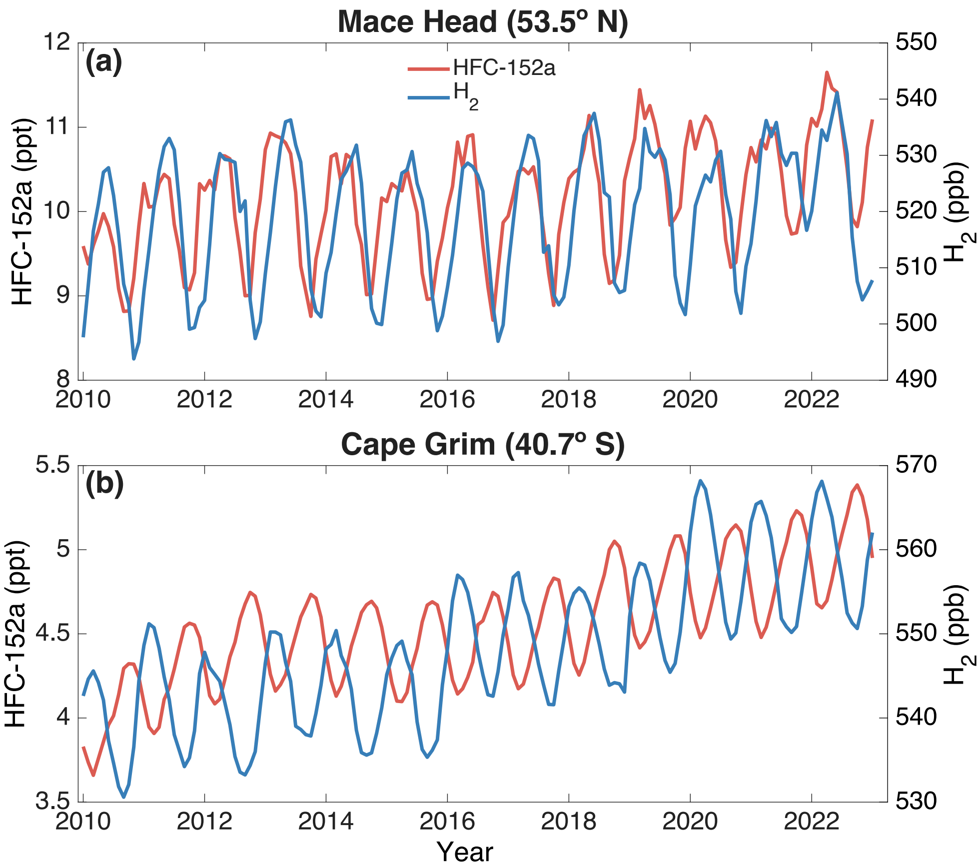


Figure R1. New Figure 1. Comparison of HFC-152a and $H_2$ AGAGE times series at (a) Mace Head and (b) Cape Grim highlighting the different seasonal cycles and hemispheric gradients.

Section 2.1 : It would be easier to follow if the authors introduced here they were running both a forward model and an inversion model (and that the latter is new) or made this more clear in lines 76-78.

Line 87: Added in “forward model” and “ and the inversion model” to make it clearer we are using both a forward and inversion model.

Also, see next comment response.

83-4 : I know the refers to Chen et al. 2024 for more description on the box model, but I think the paper would greatly benefit from a more detail on the box model e.g. stating which species are included, how OH is parameterised in the forward model.

We have adjusted the following to clarify what the box model is designed to do. The reference to Chen et al. was misplaced. That paper only uses a single box, with subsequent unpublished work having 3 boxes. Therefore, we have updated the text to better describe the 3 boxes we are using and to emphasize we are only solving for $H_2$ and HFC-152a:

Line 113-115. Added "No short-lived species are explicitly calculated. Therefore, the forward model uses prior OH values and HFC-152a emissions to calculate HFC-152a, and uses prior OH values, $H_2$ emissions, $CH_2O$ photolysis, and $H_2$ soil sink values to calculate $H_2$."

86 : What modifications were made to the box model boundaries?

We have adjusted the following sentence to provide more information:

Line 110-113. Modified "Transport between the boxes is based on diffusive transport terms from (Cunnold et al., 1994), with modifications to account for different box boundaries and to ensure a realistic North-South gradient constrained by observations of the long-lived tracer…" to "Transport between the boxes is based on diffusive transport terms from (Cunnold et al., 1994), with modifications to account for the different box boundaries used in that study (0–30° N/S, 30°–60° N/S, 60°–90° N/S) by averaging the closest coincident terms. The values are then adjusted to ensure a realistic North-South gradient constrained by observations of the long-lived tracer…"

89 : Authors say they use monthly temperatures and are running on a daily timestep. From S1, it looks like these are static values, but I think it would help to state this in the main text too if this is the case.

Thanks for pointing this out. They are not static values but monthly average times series over 2010–2022. We have adjusted the sentence to make this clearer:

Line 115. Modified "The box model steps daily in time, and gas phase oxidation of HFC-152a and $H_2$ are calculated using an ERA5 monthly temperatures" to "The box model steps daily in time, and gas phase oxidation of HFC-152a and $H_2$ are calculated using an ERA5 monthly temperature time series over 2010–2022,"

Line 117. Added "time series average".

100 : It is not clear to me why an offset of 20% was applied to SMO, given that the observations from RPB were obscured by extra- tropical air. Could the authors expand on this?

Thanks, we are missing some important justification here. We have added in further information.

Line 143: "Figure 1 shows a steep gradient in volume mixing ratio between the NH and SH. Therefore, our choice of SMO for the Tropics box will likely be an underestimation of the tropical average. Indeed, a second Tropical location at Ragged Point (RPB, Barbados, 13.2° N) has higher mole fractions than SMO (not shown). However, this site was not used as it has more frequent intrusions of extra-tropical air. Instead, to account for the underestimation, a 20 % increase offset has been applied to the SMO data."

108 : Please could the authors show a standard deviation or range for the x3 H2 time series in Fig S1b

This is a good suggestion, and we have now included the range of NOAA measurements within each box in Figure S1a,b, and c. We have discussed this range within the text as described below

Line 153: "In the North box, there is a large spread in the range of NOAA values. However, there is good agreement of the averages of the two datasets."

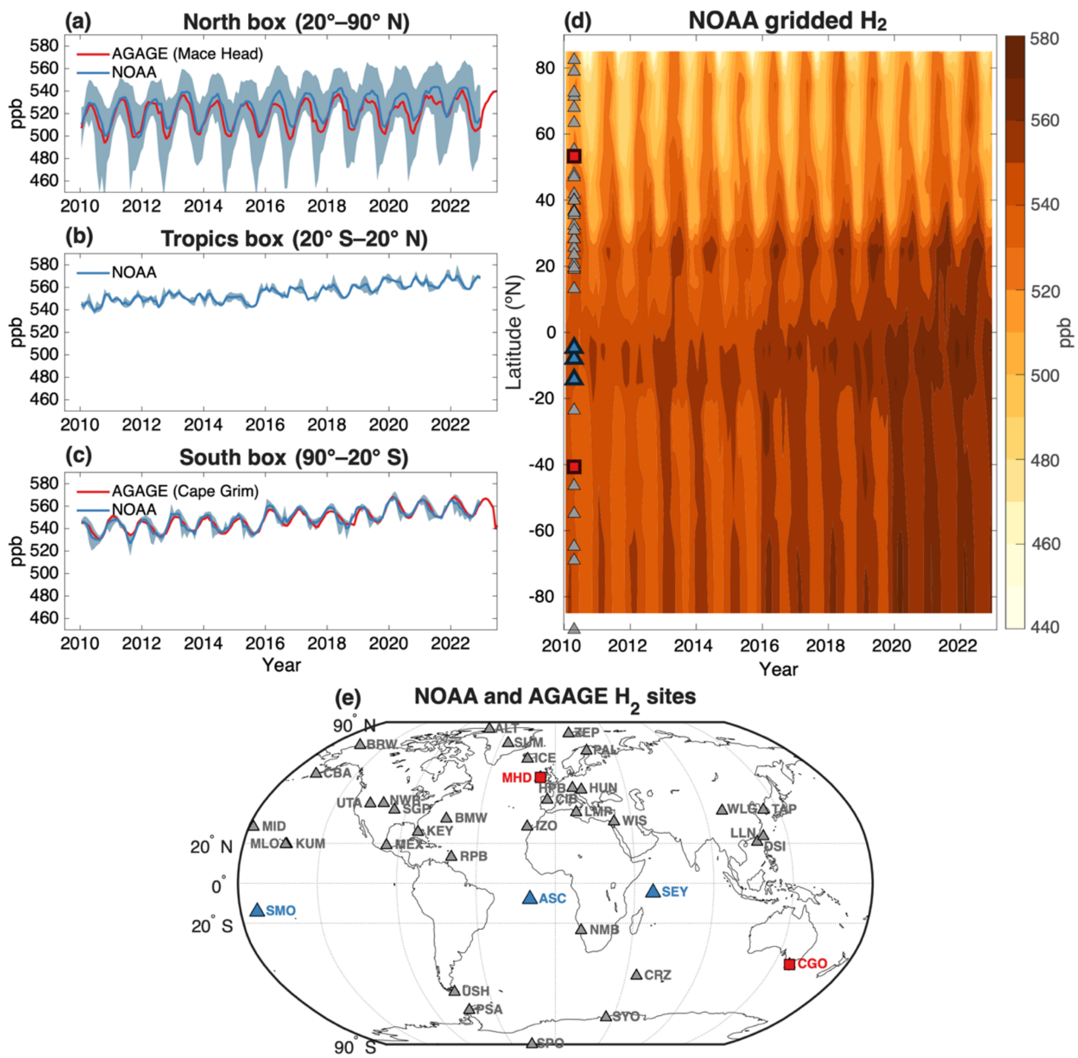


Figure R2. New Figure S1.

116 : It's a little confusing what is being derived from HFC-152a for OH and what is being taken from the literature. In line 116, the authors state that the prior information for the seasonal amplitude of OH is taken from Spivakovsky et al, but in line 94, the authors state the OH seasonal cycle is derived from HFC-152a measurements. Do they mean that only the amplitude is taken from literature and the monthly cycle (i.e. when in the year this amplitude occurs) is derived from HFC-152a? Or simply that the prior OH seasonality is taken from the literature and then optimised by HFC-152a?

The prior is taken from the literature and optimized by HFC-152a. We have adjusted the following sentence to be clearer:

Line 137: Changed: "HFC-152a is used in the model to obtain the monthly resolved OH concentration" to "HFC-152a is used in the retrieval to optimize monthly resolved OH concentrations"

139 : Can you give a (very) brief justification for using up to 300hPa in the N and S box, and 200hPa in the tropics

These values were chosen to make sure that only tropospheric formaldehyde photolysis values were used while still extending high enough to capture more than 90% of the integrated loss (as shown in in Liang et al (2017) for methane oxidation, and will be very similar for $CH_2O$).

Line 199-200: Added in "These pressure values are chosen to ensure only tropospheric values are used while still maintaining over 90% of column integrated loss (e.g. Liang et al., 2017)"

147 : It is unclear what the authors mean by "method 1" here

We have changed the following to make it clearer that we are using two methods of calculating $CH_2O$ photolysis:

Line 197: Changed "in two ways" to "using two methods".

150 : "A constant prior soil sink was used in each box" I assume the authors are referring to H2 here, but it sounds like they are referring to HCHO from wording of the previous sentence.

Line 214. Thank you, we added in $H_2$.

Section 2.3 : Could the authors give more information about the Bayesian Inversion model? Do they assume any prior distributions for their inputs, or are these all based on monthly averaged time series? Are the prior uncertainties inputted as a normal distribution (where e.g. 35% is the standard deviation) or are they the range of inputs?

All priors are monthly averaged time series, and we do assume normal distribution. We have added in the following:

Line 221: changed: “The model optimizes OH and yearly emissions for HFC-152a, as well as the monthly $H_2$ soil sink, emissions, and chemical production rates” to “The model optimizes monthly average time series of OH, the time series of yearly emissions for HFC-152a, as well as the monthly average time series of the $H_2$ soil sink, emissions, and chemical production rates”

Line 250: “Note all uncertainties are expressed as normal distribution standard deviations.”

Table. 1 caption. Added in “normal distribution standard deviation”

172-3 : Can authors clarify what is meant when HCHO photolysis rates are “retrieved with cross-correlation between OH”?

The cross correlation ensures the strong causal correlation between OH and $CH_2O$ is included in the non-diagonal elements of the prior covariance. We have added in the following:

Line 240: “Since there is a strong causal relationship between OH and $CH_2O$ atmospheric concentrations, a zero-lag cross correlation between OH and $CH_2O$ is implemented”

200 : Bring forward the reference to Fig 2b on line 201 to the last sentence of line 200 for clarity

Done, thanks.

201 : Do the authors mean to reference Fig 3c here (instead of 2b) which shows the double OH peak in the tropical region?

We are comparing the observations and forward model here so Figure 2b (now Figure 3b) is accurate. However, we have added in the reviewer’s suggestion as it will also be helpful for the reader.

209 : Reference Fig 2a. It might be worth noting that the posterior standard deviation of these discrepancies is also within the observations.

Line 290. Added in the reference to Fig. 3a and “However, these differences are within the posterior standard deviation”

212 : It is not immediately clear why the forward model would be 9 months out of phase but still simulate a strong seasonal cycle. Can the authors offer an explanation for this?

The forward model uses a constant soil sink, so if the soil sink seasonality is the main driver of the $H_2$ seasonality, the forward model will be out of sync with the observations. The seasonality seen is then due to OH/$CH_2O$ and emissions.

Line 291-294: Added in: "In the North box, the forward model mole fractions of $H_2$ show very different seasonality compared to the observations (Fig. 3d). Since the initial forward model has a constant soil sink, the soil sink does not affect the seasonal cycle. Therefore, OH, chemical production, and $H_2$ emissions control the modeled seasonality, causing the large difference in phase and inferring that the soil sink is likely the dominant driver of the observed seasonality (also see Fig. 1)."

220 : Please refer to the relevant figure here

Thanks, done.

227 : This is a confusing way of describing the plots. Why done write "North box (3a), Tropics box (3b), South box (3e)" etc.?

Agreed, changed.

234 : Could the authors expand on the comparison with Bousquet et al? How 'good' is the agreement? Did they have the same peak? Similar variation/standard deviations?

It is hard to compare directly because of the different boxes used in that study. That paper also doesn't report monthly errors for the individual boxes used. We have expanded the discussion slightly to highlight what we can compare.

Line 314-315. Added "The seasonal range and phase of values are in good agreement with a previous inversion for a 30°–90° N box using MCF (Bousquet et al., 2005)"

236 : What is the OH range in the tropics? Perhaps give the standard deviation as a percentage of this, so it is easier to compare with the North and South boxes (also see next comment)

We agree this will be an improvement. We have added in percentages and the standard deviations for all boxes. Please see next comment response.

231-243 : The reader is only given an OH range and STD for the north, a range for the tropics, and a range for the south box. Given that the authors have specifically chosen not to plot the absolute values in Figure 2, I think it is important to state consistently the OH values and standard deviations in all three regions.

Line 315. Added in: "(~15 % of the range)."

Line 320. Added in: "The posterior standard deviation is ~$0.25 \times 10^5$ (~50 % of the range)."

Line 325. Added in: "with a posterior standard deviation of $2.0 \times 10^5$ molecules $cm^{-3}$ (~15 % of the range)"

260-261: Need to state that there is reasonable agreement in the soil-sink variation between the two methods. Can the authors also state a percentage contribution from the soil sink from the
different regions?

By "variation" I believe the reviewer is referring to the posterior standard deviations. Therefore, we have now included a comparison of the errors in the discussion. We have also added in a percentage of the total soil sink for each box.

Line 368. Added in "The posterior standard deviations between the two cases are also similar".

Line 397-399. Added in: "The yearly average soil sink contributions from each box are retrieved as 43–46 %, 41–46 %, and 11–13% for the North box, Tropics box, and South box respectively over all cases."

266 : Please state relevant figure

We now refer to the relevant subplots when discussing Figure 4 (now Figure 5).

269-270 : I'm not convinced by this statement of seasonality. The authors say that the range of H2 soil-sink is also the same as the standard deviation, which implies there is no significant seasonality at all. If there is an agreement of a lack of seasonality, they should state this

We have added in the following clarifying sentence

Line 378. Added: "Although, it is important to note the significance is limited in this study as the retrieved posterior standard deviation is of similar magnitude to the range."

273-275 : Similarly here, can the authors comment on the significance of the peak and trough sink soil in the south given that the posterior standard deviation is similar to the range

Line 385. Added: "However, the small range leads to limited significance outside of the posterior standard deviation."

282 : Is it possible to say how much each box contributes to the overall the global soil sink (I imagine this is weighted heavily towards the north, but % contribution would help clarify this)

Because of the uncertainty in our absolute OH values, we did not want to mention the budget, but we have added in the percentage contributions as suggested. See previous comment (Line 260-261).

297 : I think the authors could expand on the reasons why there is more uncertainty for the soil sink phase in the tropics. There's a greater abundance of OH in this region, so it would make sense for this area to be more sensitive to OH. We also wouldn't expect to see as much seasonality in the tropics as in the other two boxes (averaging across ~20 years with various ENSO indices). Is it also possible that there is more uncertainty in the H2 observations for this region (the H2 observations were averaged between several sites)?

Primarily the reason is that the tropics has a large soil sink, large emissions, large chemical production and large OH loss. With all those uncertainties compounding the retrieved uncertainty. We agree it is important to state this in the paper.

Line 421ff. Added: "This is especially the case in the Tropics box due to larger chemical production and OH loss values compared to other boxes and larger $H_2$ emissions compared to the South box."

297 : Are authors referring to all cases here? (And again in line 300)

Fixed

301 : Can the authors expand more on the H2 chemical production. I would also consider moving Fig S7 to the main text and add a short section about this. Fig 4 + S7 : only three error bars are given in these plots (compared to Fig 3). Is there a specific reason why the authors cannot show the other error bars for the remaining months? For Fig S7, please also show error bars for the HCHO derived OH case

We did not include the error bars for each month in these figures to avoid clutter. We don't think it is needed as the three months we have chosen show the range of error over the seasonal cycle. We did not include error bars in the HCHO derived from OH case because in this case HCHO is not retrieved, but is calculated from retrieved OH. We can estimate the error by using the retrieved OH and we have now included that and explained in the caption.

Line 372. Added: "In this study, even though our retrieved $CH_2O$ is directly linked to our retrieval of OH, our prior values are derived from model and reanalysis estimates and is therefore a source of uncertainty."

Table S1 : Do the authors mean days-1? The values in the final two columns seem to show the lifetime transport? Also, are the same transport mixing timescales used in reverse e.g. Tropics → North and Tropics → South?

The units display as $day^{-1}$ in our copy. We will make sure during the revision upload that it is correct. Yes, this is diffusion transport so is the same in both directions.

Fig S2 : What starting conditions are used for the slow and fast OH transport?

The starting conditions are ~30% slower and faster than the values in Table S1. We have added this to the caption for Figure S2.

**Technical Comments**

44: "on the impact of OH on…"

Fixed, thanks.

87 : comma missing after SF6

Fixed, thanks.

91 : remove brackets from citation

Fixed, thanks.

129 : Missing a bracket

We can't find the missing bracket on this line

146 : P0 → P0

Fixed, thanks.

165 : m → m

Fixed, thanks.

170 : "the prior uncertainties values minimized retrieval uncertainties" → "the prior uncertainty values minimized retrieval uncertainties"

Fixed, thanks.

172 : section 2.2 → Section 2.2

Fixed, thanks.

229 : 2 → two

Fixed, thanks.

Table S1 : Difference hyphen type in North-Tropics to South-Tropics

Fixed, thanks.

300 : remove “is”

Fixed, thanks.

316 : remove comma after ‘production’

Fixed, thanks.