Response to reviewers

Dear Dr. Maria Kanakidou,

Thank you for the opportunity to submit the response to the comments raised by the reviewers. We are grateful to you and Dr. Maarten Krol and Dr. Alex Archibald for the valuable feedback that has served to improve the article. We have made changes to the article to incorporate most of the suggestions. Please find the point-by-point response to their comments/concerns below.

Thanks,

Srinath Krishnan.

Submitted on 03 Oct 2025

A multi-model approach to constrain the atmospheric hydrogen budget

Srinath Krishnan, Ragnhild Bieltvedt Skeie, Øivind Hodnebrog, Gunnar Myhre, Maria Sand, Marit Sandstad, Hannah Bryant, Didier A. Hauglustaine, Fabien Paulot, Michael Prather, and David Stevenson

Handling editor: Maria Kanakidou, mariak@uoc.gr

Reviewer 1

- **RC1**: ['Comment on egusphere-2025-4898'](https://editor.copernicus.org/#RC1), Maarten Krol, 20 Nov 2025 [reply](https://editor.copernicus.org/index.php?_mdl=msover_md&_jrl=778&_lcm=oc116lcm117t&_acm=open&_ms=134044&p=300508&salt=185993206733478121)

This paper deals with the global hydrogen budget. An important subject acknowledging potential leakages in large scale deployment of hydrogen in the energy transition. First, the paper presents an evaluation of a multi-model ensemble that simulated the global hydrogen budget. Second, a box model is tuned on the global models, and the global budget is constrained by the observed hydrogen isotopic composition. The authors claim to have tightened the atmospheric hydrogen production, potential geological sources, and the soil sink.

Although the subject and the employed methods are interesting, I find the paper rather messy (e.g. typos, overall structure) in its current state. Moreover, the overall claim that the geological sources are constrained to < 9 Tg H2/yr is not very well substantiated. Below and in the attached annotated pdf file I provide more detailed comments.

Global model evaluation

The different parts in the manuscript are not very well connected.

The paper starts with an evaluation of the global models, showing comparisons to CHCO, NO_2 and CO satellite data. The paper remains rather vague here.

Thank you for the suggestion. A comprehensive evaluation of the models with satellite data is limited by the lack of temporal overlap between the two datasets, and because the model data is monthly mean output versus satellite datasets that reflect overpass times. This is now reflected in the results section, where there is less emphasis on the qualitative comparison of the ranges in atmospheric production and loss terms observed in the ACMs with the satellite retrievals and ATom datasets (Figures 2-4 has been moved to the appendix in addition with new figures comparing them in percentages – Figs. A1-A8). We have restructured the paper to emphasize the point of the article is the evaluation of the different terms used in the box model.

I read that the model is forced to H2 (and CH4) surface observations, and that the soil sink is tuned to reproduce reasonable hydrogen concentrations. This seems a double constraint, and it remains unclear how that impacted the H2 budget. Moreover, no detailed H2 budget terms are presented here (used in the box model?), and the model evaluation is limited, showing figures with many panels that are not very informative.

That is correct and is because the different ACMs were not run with free-running H_2 and CH_4 emissions. We have now added a joint analysis of the different source and sink terms (Figure 7) to explore the parameter space of these budget terms.

We have also added another table (Table 2) where the budget terms that are used in the box model (from the different Atmospheric Chemistry Models) are listed. Further, we have restructured the manuscript such that the design of the box model and its application form the focus of the manuscript, and the satellite comparison is used as a qualitative analysis of the spatial comparisons (and moved to the appendix). We agree that the satellite comparison is qualitative at best because of the different time frames and the lack of 3-hourly model output saved (for the use of the averaging kernels).

Moreover, in comparing to NO2 (and CHCO) satellite products, Averaging Kernels (AK are important, as well as co-sampling. The authors acknowledge this (and use 3 hr output still without AK). According to me, this evaluation does not add much current knowledge (and the results presented in Sand et al. (2024)). If anything, the model results could be used to estimate the uncertainty in the OH sink and other uncertainties in the H2 budget.

We agree that averaging kernels are very important for model-satellite comparisons. However, there is no temporal overlap between our three-hourly run and the satellite output, which limits our ability to make meaningful comparisons with the kernel. We have moved Figures 2-4 to the appendix. The spread in the model values (from Sand et al., 2023) in atmospheric lifetime is the different OH sink values for the various models and we conduct a joint analysis of all the terms in Figure 7.

Box model

The box modelling is an interesting way to constrain the H2 budget. But the way the box model is presented and used needs improvement.

First, it seems that the starting condition in the box model is not consistent, specifically for the "lifetime 3" case (which should be lifetime 2, but there are many of these mistakes).

We have hopefully corrected these mistakes and have added a table where we list the values used in the box model for the different cases. The TAR scheme (lifetime 2) calculates the OH sink lifetime using the formula in Ehhalt et al., 2001:

δ ln(tropospheric OH) = − 0.32 δ ln(CH_4) + 0.0042 δ (e − NOx) − 1.05e− 4 δ (e − CO) − 3.15e− 4 δ (e −VOC), where CH_4 is in ppb and the (e-X) is in Tg of species X per year.

The initial jump in H_2 concentrations is because the model is initialized with pre-industrial values from the OsloCTM3 pre-industrial run (other than OH-sink lifetime). The OH-sink lifetime is calculated using the TAR formula and scaled to 2010 values. The longer lifetimes may also be because the box model setup is designed for 2010 and there are large uncertainties with the terms pre-1970s. As shown in Figure A11, the OH-sink lifetime increases, leading to an increase in H_2 concentrations, which takes 8-10 years for the adjustment.

Second, it is unclear how the H2 budgets differ in the models (needs to be part of the global model evaluation). More detailed information is needed why the isotopic composition differs in the ACM-based box models. Some reasons are given, but this is actually a good way to give the ACM results a decent place in the manuscript.

Thanks for the suggestion. We have added a table (Table 2) where we list the values for the different budget terms in the different models and in the literature study that is used in the figure. And we have added text on how the H_2 budgets differ in the global model evaluation:

"All the models' box model simulation results show a narrow range of atmospheric H_2 concentrations (520-540 ppb; by design) and a wider range for isotopic values (20-210‰), with WACCM at the lower-end and UKCA at the higher-end. The enriched isotopic value in WACCM is due to the combination of high emissions (anthropogenic and biomass burning sources), low atmospheric H_2 production and high soil sink. The more negative value for isotopic composition in UKCA is due to low emissions and low soil sink in UKCA. OsloCTM3 has higher atmospheric production, but this is isotopically balanced out by higher soil sink."

Third, large geological sources are considered unlikely, but I think this is an overstatement of the ability of the box model. Figure 9 is rather misleading, because this show the geological result of the -1000 ‰ signature (at least if I trust Figure A5). Furthermore, Figure 10 suggests that the impact of geological emissions on the isotopic composition is opposite to the effect of OH oxidation. This would imply that a stronger OH sink could compensate for the geological emissions. However, the OH sink is not included in the re-tuning of the model. The soil sink is included, but the soil sink does not enrich the atmosphere. I would suggest allowing the OH-sink to vary at least over the ACM results (you write: _There is a larger diversity in OH indicating more uncertainty and a bigger spread for atmospheric losses)._

Thanks for the input. We have added a new joint analysis (Figure 7) where we explore the diversity range of the different sources and sinks. And we agree that there should be another case with modified OH sinks. We have now added a case where the maximum OH loss in the model is incorporated in the box model as an additional case (Figure 4).

Other issues

The box model should account for the enrichment of H2 due to inflow of stratospheric air. The authors mention that this can account for a 29-37‰ offset in δD. This is not particularly large but introduces a bias. Overall, I earlier pointed to the work of Pieterse et al., to which the paper refers to now. However, I would expect some further discussion, e.g. about the used isotopic values (Figure A4 looks quite different).

Given that the isotopic value of the stratospheric hydrogen itself is related to the δD of CH_4 calculated within the model in Pieterse et al. (2011), here we assess the impact of the stratospheric contribution by adding the offset to the final atmospheric δD values in the new uncertainty analysis figure and discussion that is added to the paper. We have also modified Figure A4 (now Figure A9) to include the impact of stratospheric enrichment and included the following text:

"For example, Fig. A4 shows the resulting budgets if isotopic values and fractionation factors for the different sources and sinks are taken from (Pieterse et al., 2011) rather than (Price et al., 2007) without and with stratospheric enrichment of 29‰. The mean changes for the models and literature values are 7.8‰ and 24.56‰ without stratospheric enrichment and 20.61‰ and 49.61‰ with the stratospheric enrichment respectively. This difference is driven by the Pieterse et al. (2011) having isotopic compositions of -260‰ and 116‰ for biomass burning and atmospheric production of CH_4 instead of -290‰ and 162‰ in Price et al. (2007)."

Other comments:

We have also modified the manuscript with the other clarification and the grammatical changes in the pdf document.

Reviewer 2

- **RC2**: ['Comment on egusphere-2025-4898'](https://editor.copernicus.org/#RC2), Alexander Archibald, 09 Dec 2025   [reply](https://editor.copernicus.org/index.php?_mdl=msover_md&_jrl=778&_lcm=oc116lcm117t&_acm=open&_ms=134044&p=301879&salt=18889040105587718)

Krishnan et al report on a combined analysis of box and global model simulations with the overall aim of quantifying and reducing uncertainty in the tropospheric budget terms of hydrogen. This is an important topic and their use of isotopes is nice. However, the paper as it currently stands does little in i) critically evaluating the models that we are using to define the budget terms ii) systematically exploring uncertainty in these terms. As a result the authors are unable to significantly constrain any of the budget terms -- albeit that they report that the geological source of H2 must be much lower than 20 Tg/yr (which I don't doubt).

I would be happy to review a revised paper but I would suggest that the authors focus on the following two major areas in a re-submission:

1) more critical analysis of the models against observations. Presently the model-obs comparison is rather qualitative. From my understanding of the data, the models and observations are all simulating (observing) different time periods. How can this alias the interpretation? What can be done to minimise this?

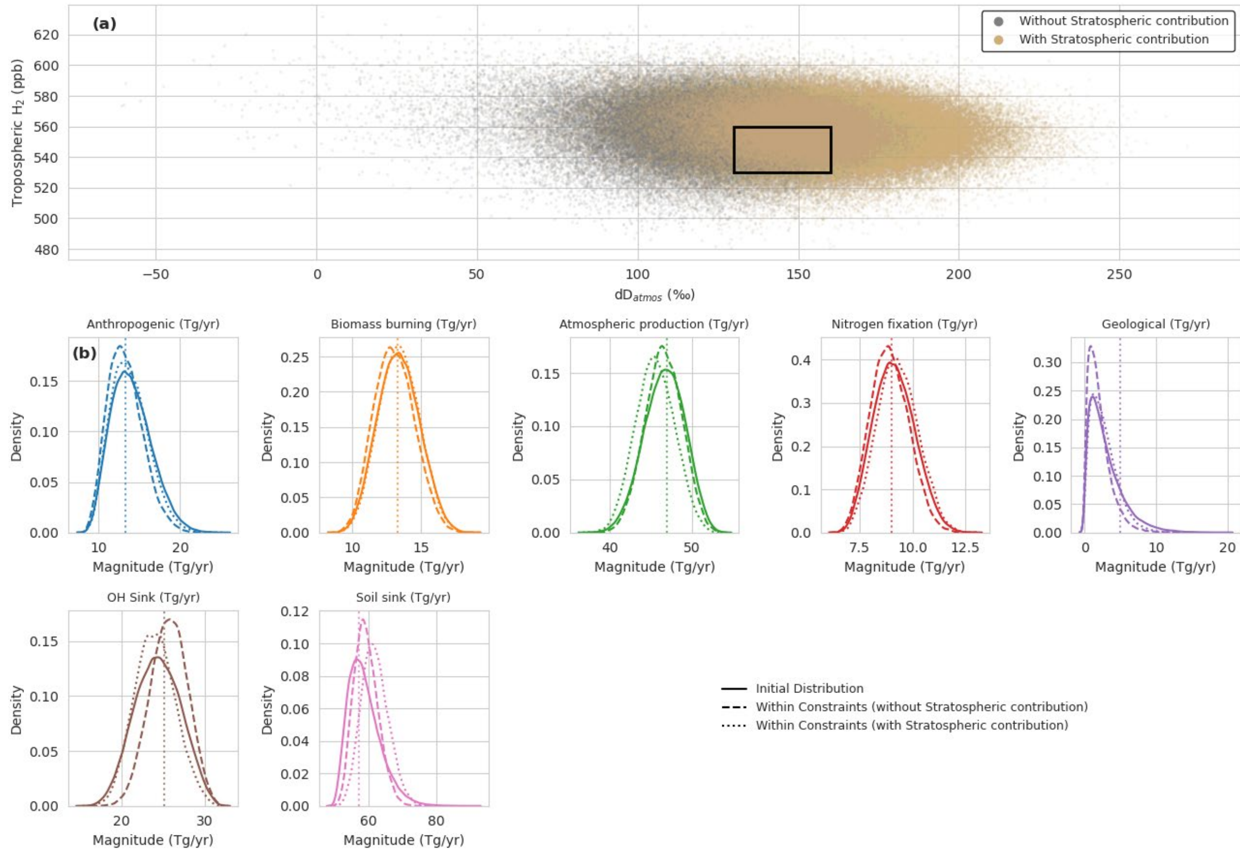
Thanks for the suggestion. Unfortunately, the model output that we present for monthly mean output for simulations run for 2010 versus the satellite retrievals are 2018-2019. We have rearranged the results sections to highlight the inter-model comparison and focus less on the model-satellite comparison to underline that it is a qualitative analysis (while we have kept the Figure 1 spatial figure in the main manuscript, we have moved Figures 2-4 to the appendix).

2) we have large uncertainties in a number of terms within the hydrogen budget and there are more elegant ways to explore this joint uncertainty than a series of case studies focusing on high and low end member combinations. It would be interesting to see your box model results if you were to span the range of uncertainties. You could even weight some of these sources of uncertainty by using the results shown in Figure 11, for example.

Thanks for the suggestion. We have now added an uncertainty analysis where we ran 100,000 box model simulations with randomly picked values for the different sources and sinks and the isotopic compositions to test whether they fall within the observed range. We then check the changes to the shape of the distributions and whether they fall within the "observed ranges" of atmospheric H_2 and δD values when we change these sources and sinks.

We have added the following text and figure:

"In addition, we also performed an uncertainty analysis to evaluate the joint uncertainties using the distributions of all the sources and sinks and their isotopic values (ranging from the most negative and most positive values in literature; Fig. 7). Additionally, we also check the impact of the stratospheric contribution to the analysis, by assuming the enrichment of the final atmospheric δD composition by randomly picking a value between 29-37‰. We use a Monte-Carlo approach by randomly picking 100,000 different combinations of sources and sinks and isotopic values to run the box model and check whether the resulting atmospheric concentration and the isotopic compositions of atmospheric H2 is within the ranges of 530-550ppb and 130-160‰, respectively. Applying the joint certainty analysis, the 97.5th percentile upper bound on the geological sources decreases from 9.4 Tg/yr in the initial distribution to 6.1 Tg/yr and 7.3 Tg/yr without and with a stratospheric contribution, respectively. Relative to the initial distribution, only 88% and 91% of the geological flux distribution overlap with the 95% constrained intervals without and with a stratospheric contribution, ruling out 12% and 10% of the distribution, respectively."



Minor points to address:

Line 19: I disagree that the lack of global-scale knowledge of the abundance of OH is the key point here.

Changed to add soil sink uncertainty.

Line 20: Caps for Atmos. Chem. Models but no definition of the ACM acronym.

Changed.

Line 21: Hyphens used when it should be em dashes.

Changed.

Line 24: Caps used for VOC but VOC not actually defined.

Changed.

Line 25: Isotopic composition of what?

Added Isotopic composition of H_2.

Line 33: Define today.

Changed to "in 2024".

Line 37: Define VOC and NMVOC.

We have changed that.

Line 49: Typo --> release(s)

We have fixed it.

Line 55: Is this O(1D) + H2O -> H2 + O2 source in any models? My reading of Zellner et al. (1980) suggests that the reaction involved is H + OH -> H2 + O, which I think would only happen in the stratosphere.

The suggestion that it is a minor source is from Ehhalt and Rohrer, 2009 (reactions R6, R7a, R7b), who estimate that about (0.6 ± 0.6) Tg H2 yr−1 could be produced via (R7b) in the troposphere, which is the O(1D) reaction.

Line 60: typo 2->20?

Changed

Line 71: Please explain a bit more for the general reader what the positive value of delD means (enriched, depleted etc).

Added "Isotopic fractionation usually occurs during chemical or physical transformations leading to an isotopic enrichment (more positive δD) or depletion (more negative δD) of the product, depending on the process."

Line 74: "produced H2" sounds odd to me. I suggest a change.

Changed to resulting.

Line 75: "(Piet.." -> "Piet.. (see marked up pdf).

Changed.

Line 78: Define ACM

Changed.

Around Line 90: I suggest to add in something about H2 yields here also being uncertain.

Added "The yields of H_2 from these reactions are uncertain."

Line 99: Suggest adding in "Here".

Added.

Line 148: Delete "which varies between models"

Deleted.

Line 150: How are these calculated? And are they used? Later it is implied that this split is not available.

We do not have a spread of the splits (i.e. percentage of H_2 production from NMVOC vs. CH_4 relative to the sum of the two) from the different models. Here we use two splits, 44% from Ehhalt and Rohrer (2009) and 39% from Paulot et al., 2024. We use 44% as the split in all simulations other than the sensitivity test (Figure 3).

Line 162: Is there a reference for this?

Added the references - (Price et al., 2007; Pieterse et al., 2009, 2011).

Line 182: What about uncertainty in J(HCHO->H2+CO)?

That is a good question. We do not account for that uncertainty in our model simulations and that can affect the qualitative comparison. We have added
"Broadly similar values between HCHO in the models and TROPOMI retrievals suggest that the 3-D models represent the production of atmospheric H_2 fairly well, however there are several uncertainties that can affect model simulations that have not been explored in this study, including the uncertainty in the photodissociation of HCHO."

Figure 1: Please plot % difference model-obs and the obs using different colour scales to help readers make clear distinctions of areas of difference. For the caption make clear the difference between the observation time period and the model data time period.

Thanks for the suggestion. We have added Figures A4 and A5 with the percentage differences between the model and satellite data. We have also added the years in the caption.

Figure 2: Would it not make more sense to integrate/sum the data? After all the budget terms are totals so why not do the same here? If not, what are the error bars on these averages?

Unfortunately, we only have monthly mean values for the model output and not monthly totals. We have now added an error bar that corresponds to the standard deviation of the multi-model mean.

Line 205-208: Re-word.

Changed to, “The simulation and the satellite reanalysis datasets are for two different time-periods, so we cannot replicate the specific air mass factors in OsloCTM3. But if the mean calculations are restricted to specific satellite overpass times with the 3-hourly OsloCTM3 simulation, results show much lower values, bringing the model means closer to TROPOMI values, with an underestimation (Fig. A4a).”

Line 209: Unclear what Figure A3 is showing. See later comment.

We have modified the figure (now A8) and only show the NO_2 columns where we compare the annual mean NO_2 values with the NO_2 values calculated using the ratio of overpass NO_2: monthly NO_2 means from the OsloCTM3 simulations where 3-hourly output was saved.

Line 213: Change supplementary to Appendix

Changed.

Figure 3: Too many plots all together. I suggest splitting this up and again focusing on the obs data and deviations from it (%?) using different colour scales. If the multi model mean is important please show that too and also the % standard deviation in the multi model mean so that spatial features of uncertainty can be highlighted.

We have moved the figure to the appendix (Figs A2) and now also include the % standard deviation and the % differences (Figs. A4, A5).

Figure 4a: How useful is the NO2 comparison given the huge impacts of the 3hr vs monthly mean sampling difference shown for OsloCTM3. The difference from sampling is larger than the multi-model range.

That is a good question. Yes, there is an issue in using monthly mean output versus the 3-hourly output, but, unfortunately, there is not much we can do given the model output. We have added, “But comparing the monthly mean model output with satellite-overpass times only allows for a broad qualitative analysis. Future simulations should save the 3-hourly output so that such comparisons can be made.”

Figure 4b: The difference between MOPITT and TROPOMI is maybe 30% (by eye) of the standard deviation of the multi model mean. In other words observational error means we can only constrain a fraction of the model variance, right? This point is not made very clearly.

We have added the following lines regarding MOPITT and TROPOMI:

"The global mean difference between TROPOMI and MOPITT is 0.002 mol CO/m^2 and the multi-model standard deviation is also 0.002 mol CO/m^2. Over the oceans, the difference between the 0.003 mol CO/m^2, while over land it is 0.001 mol CO/m^2. For the models, the standard deviation is 0.003 mol CO/m^2, while over oceans it is 0.001 mol CO/m^2. While the global mean values of the differences between MOPITT and TROPOMI are similar to the multi-model standard deviation, the models have lower values of CO over land, compared to the satellite retrievals."

Line 251 and 253: Make clearer what the control and test cases are by adding the details to a table (maybe Table 2?).

We have updated Table A1.

Figure 6: Make clearer what the fraction (ratio) is/means.

We have clarified it in the text as:
"We run two cases - in the first case (solid lines in Fig. 3), production from NMVOCs accounts for 44% of the total production consistent with the estimate of Ehhalt and Rohrer (2009), while in the second case (dashed lines in Fig. 3), this production from NMVOCs is reduced to 39% following (Paulot et al., 2024)."

Line 267: How did you implement the TAR formula? It shows a delta OH so what is the reference [OH] at t0?

The box model uses the OH-sink lifetime, which is calculated using the TAR formula. The lifetime is then scaled to the atmospheric sink lifetime for 2010. We have added Figure A11 that shows the atmospheric lifetimes used for the three different cases. Longer lifetimes in the early period (pre-1970s) could be due to the box model setup designed for 2010 values, and the uncertainties in the different terms.

Section 3.4: My understanding here for the isotope work is that you are modelling 2010 conditions? Make clear.

We are plotting for 2010 conditions from the ACM. We have added that to the caption and added a table with the budget terms used.

Figure 8: Could a ternary plot help visualise the impact on delDatmos to [H2] and [CH4] which seem to vary in all the studies? Or is there a way to normalise the y-axis

([CH4]/[H2])? With so much scatter it's hard for the reader to draw anything useful from this. Would marginal PDFs help the reader? It seems like the brown circles median is around 100permil and the blue triangle mode is around 150permil (closer to the black cross). So does that mean the newer models are better?

This is a good point. However, we do not have the [CH_4] values for the different literature studies and the relationship is not a linear one. Figure 5 shows small variations in H_2, so this may be an issue, but the analysis has to be limited (at this point). The point of the figure is to simply show the range in isotopic values given the different budget terms in model and previous literature, rather than suggesting that the models have improved over time, given that the models are by design closer to the observed concentrations. We have changed text through the paragraphs to clarify this.

Figure 8 caption: delete "possibly" and the end of the caption: "suggests..."

Changed.

Line 309: Add in "the"?

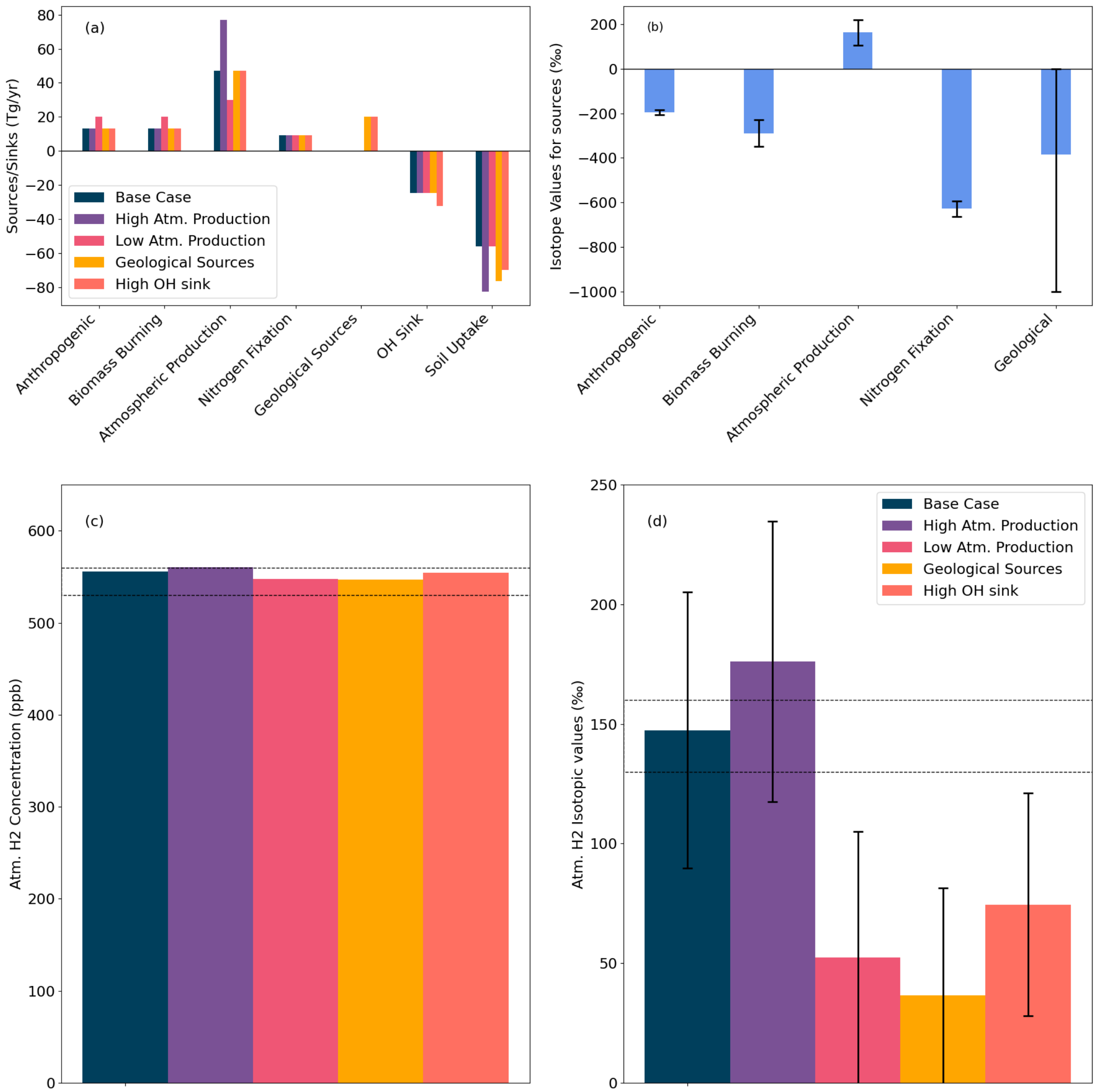
Changed.

Table 2: Make clear the base year.

Added "with steady-state 2010 conditions".

Figure 9: I think in b you can drop "iso" in the labels. In panel d I take it you can't plot a violin plot because you only have four points per bar? The text in the caption for panel d is unclear. Change.

Yes, this is correct about the violin plots. We have added a new case for higher OH sink to the figure and dropped the iso in the labels.



Line 326: By design this should be the case no?

Yes, we have clarified that.

Line 332: Expand on this please. See my major concern and consider a more formal uncertainty assessment.

We have added an assessment and discussion based on Figure 7.

Figure 10: Caption. Change dry dep -> soil sink. Should the y-axis label be "dH2 (ppb)"

Changed.

Aside: please replace ppbv with ppb and on first using make clear you mean the mole fraction (nmol/mol of dry air).

Changed.

Line 355: Unclear what you mean.

We have removed that.

Line 367: Change "observed" -> "documented". Add "the"

Changed.

Figure 11: Make clear what "model values" refers to. I would also encourage a table being added (SI?) with the sources of data used in generating this plot.

We have added Table 2 with the data in the different ACMs and literature and changed the figure to be clear that "model values" refer to Sand et al., 2023.

Line 382-383: See the marked up typos.

Fixed.

Line 393-394: See the marked up typos.

Fixed.

General comment: When discussing the conclusions and your new constraints it would help to remind the reader of where we started from before this work so we can understand how well constrained we are now.

Added "The ranges for Atmospheric Production and Atmospheric Loss budget terms from published literature range from 37-80 Tg/yr and 15-20 Tg/yr. A recently published global hydrogen budget for 2010-2020 suggests values of 38.4 ± 6.1 Tg/yr and 18.4 ± 2.2 Tg/yr for photochemical production and sink, respectively."

Figure A1: Limit the min of the y-axis to 10 ppt

Changed.

Figure A1 and A2: See comment on pdf. I feel like more should be drawn to the models than the observations so suggest grey symbols for the observations and the models overlaid on top of them. Or, do more detailed analysis of the models and obs.

We have modified the figures, but it is difficult to do a detailed analysis between model and obs. given the different time periods.

Figure A3: I think the key plots are columns a and b, no? What's the point of the other columns?

We have removed the additional columns.

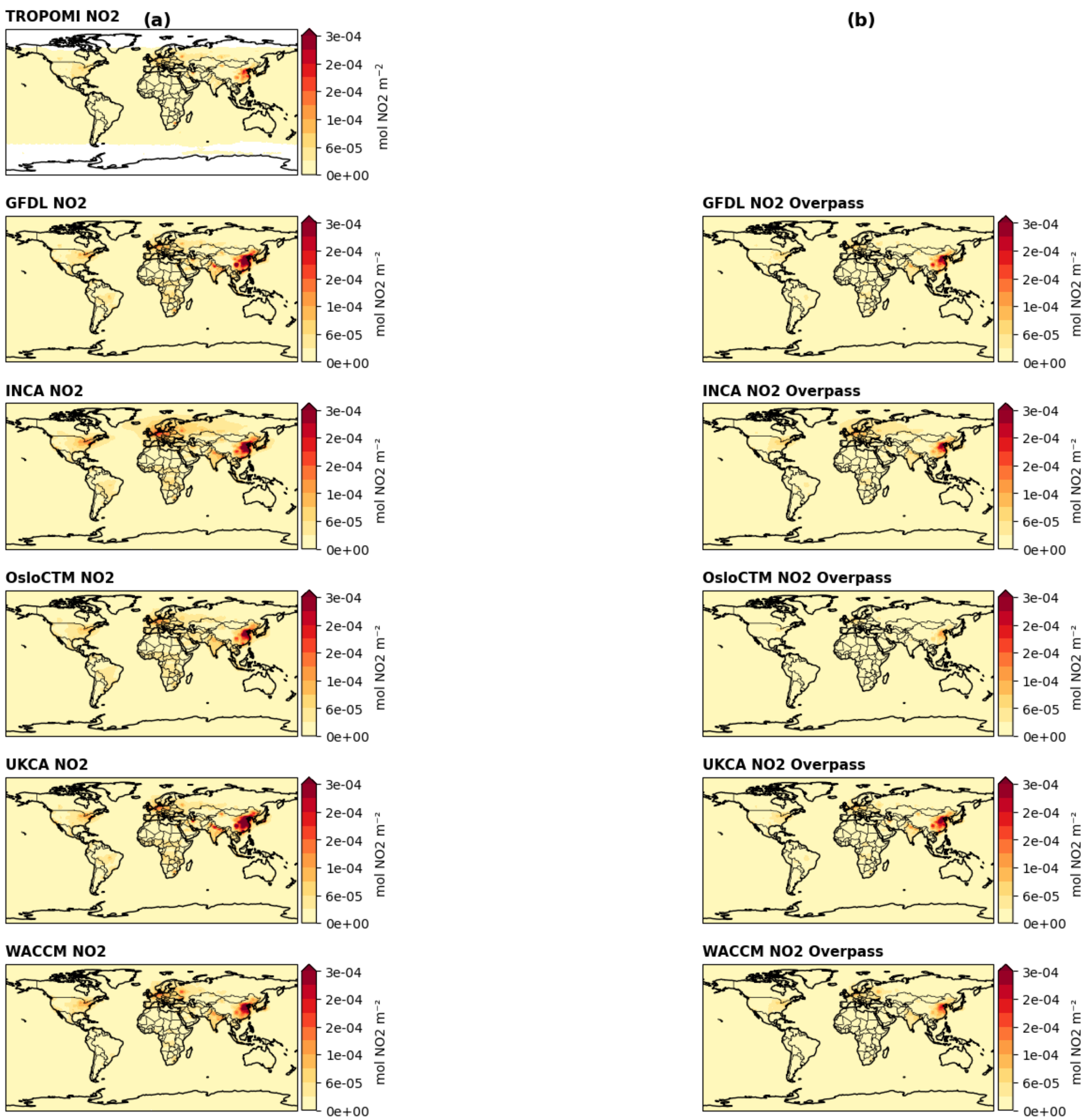


Figure A4: Pretty pointless without helping the reader see how the dD values have changed.

We have modified the figure to show arrows depicting how the isotopic values change if fractionation factors from Pieterse et al. (2011) is used instead of Price et al. (2007).

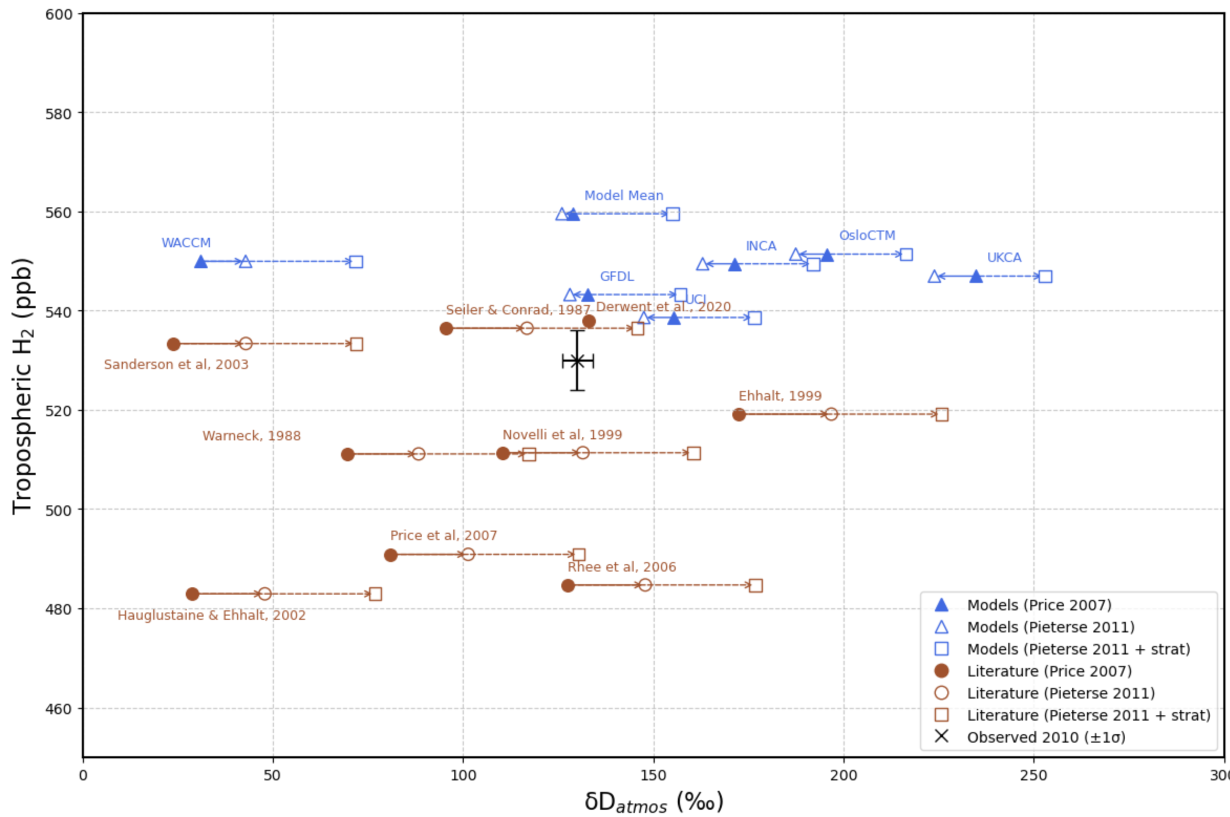


Figure A9: Similar to Fig. 8, but using isotopic compositions and fractionation factors from (Pieterse et al., 2011) and 29‰ as stratospheric contribution.

References:

Zellner, R.; Wagner, G.; Himme, B. H2 Formation in the Reaction of O(1D) with H2O. J. Phys. Chem., 84, 1980.

Citation: https://doi.org/10.5194/egusphere-2025-4898-RC2

[](https://editor.copernicus.org/index.php?_mdl=msover_md&_jrl=778&_lcm=oc108lcm109w&_acm=get_comm_sup_file&_ms=134044&c=301879&salt=1891539917157104131 6 "Supplement download")