Reviewer 1 (Dr. Alexander Tardito Chaudhri):

We thank the reviewer for their kind comments and for taking the time to evaluate the manuscript. Using this feedback, we were able to further improve the paper. Below, we address the suggested clarifications and corrections (text in blue). We would like to note that 300 ensemble members were used to derive our main results.

Specific comments:
These first comments concern experimental configurations that could be commented on and would strengthen these work, and possible substantial conclusions that may have been overlooked in this article:

- Paulot et al. (2024) and Tardito Chaudhri and Stevenson (2025) both identify constraints in hydrogen uptake from its observed seasonality. Presumably your setup is ideal to explore this. I see in Figs E5 g and h that the posterior uptake schemes require distinct seasonality compared to the assumptions based on physical parameterisations. What do your results suggest about the seasonality of soil uptake?

We agree with the reviewer that this could be expanded on. Therefore, we have performed additional analyses to quantify the statistics of our inferred soil uptake and have added the results to Section 3.2.5. In addition, we have performed a comparison of our results with recent work of Stone et al. (2026) and Tardito Chaudhri and Stevenson (2025), which we have added to the Appendix. In short, we found very good agreement with Stone et al., 2026, particularly in the Northern Hemisphere, and relatively poor agreement with Tardito Chaudhri and Stevenson (2025), which we attribute to the large differences in methodology between these studies.

- You find faster soil sinks in South America and Russia than the synthesis (Ouyang et al. 2025) – how sensitive are these to the choice of scaled ocean H2 emission from nitrogen fixing?

We appreciate the hypothesis, but based on our many simulations we expect that other methodological differences are more important (e.g. using top-down instead of bottom-up methods, choice of biomass burning dataset) in explaining these differences. We note that our oceanic emissions (5.0 Tg/yr) are very comparable to the Ouyang et al. 2025 ensemble mean (4.9 Tg/yr) and substantial changes would be needed to induce even faster soil sinks.

- The time evolution of posterior fluxes from non-isoprene photochemical source (Fig 7d) and the soil sink (Fig 7e) are confusing. These require a ~10% increase in the non-isoprene source in 2016-2024 versus 2008-2016, and a 0.23 Tg/yr decrease in soil uptake over the analysis period. The priors had <5% increase in non-isoprene source and an increase in soil uptake.

We regret it if these results were confusing, as we note them as one of the interesting outcomes of our efforts. The temporal evolution in posterior fluxes was needed to explain the observed positive H_2 growth rate after 2016 (particularly evident at SPO; see Fig. 4), which the priors failed to capture despite being forced by increasing CH_4 levels. An increasing soil sink as suggested by the priors would not fit this growth rate, and the network of H_2 observations finds that an increase in non-photochemical sources can only overcome part of the residuals: soil uptake must have responded more strongly to environmental limitations than suggested by either of the deposition models. This outcome shows the additional information gained from incorporating spatiotemporally explicit atmospheric constraints and would likely not have emerged from a simpler global mass-balance for annual mean source and sink magnitudes. We will update Section 3.2.4 to make the findings above more clear.

- Please can you give an explanation for how credible the posterior non-isoprene source variability is (fig.7d)? From line 209, does your choice of the low state vector uncertainty in the methane source imply that this should be well constrained already?

The assignment of a relatively low state vector uncertainty (i.e. in percentage terms) was indeed a conscious choice and stems from the well-known and prescribed CH_4 levels that partly control H_2-production, combined with our spatial OH-fields and formaldehyde levels having been scrutinized in this work and earlier work. This gives relatively more faith in its large-scale spatial pattern and trend than other terms in the budget (deposition, isoprene-derived H_2). We also note that the inferred variability is modest relative to the overall magnitude of the flux, with the second (2008-2016) and third (2016-2024) period mean non-isoprene source amounting to 25 Tg yr^{-1} and 27 Tg yr^{-1}, respectively. We consider it plausible that changes of this magnitude (i.e. 2 Tg/yr; less than 10%) could occur over the considered periods, for example due to variations in methane oxidation rates linked to OH variability, or variations in non-methane, non-isoprene VOC emissions. Nevertheless, we acknowledge in Section 4.1 that aliasing may also (partially) explain our inferred trend in the non-isoprene source.

- The prior and intuition (specific comment for Line 317 included) would suggest that the soil uptake should have increased over the period. Given that value chain emissions estimates of up to 20% are

provided by both Trapani (2025) and Esquivel-Elizondo et al. (2023, Front. Energy Res), and the range of bottom-up and top-down studies (Table 2) that have higher H2 fossil fuel source – would it be feasible to explore configurations with a varied industrial source to estimate how likely changes here are to explain the trend? The approach of Paulot et al. 2024 ACP also includes industrial H2 in a revision CEDS v20210421. A substantial result might explore whether the recent trend is better described by an unaccounted H2 emission, or a downturn in the soil uptake rate.

We appreciate the reviewer's suggestion to explore whether changes in assumed industrial H_2 emissions could explain the inferred trend in the soil sink. We agree that such sensitivity experiments can in principle be constructed within the inversion framework, and that sufficiently large and temporally varying industrial losses could redistribute part of the inferred trend. However, as discussed in Section 4.1, the magnitude of such losses across the value chain required to explain the signal would be substantially higher than current best estimates. Bottom-up assessments (e.g. Trapani et al., 2025; Ouyang et al., 2025) generally converge on leakage rates around 1 ± 0.5%, and this is broadly consistent with recent measurement-based studies (e.g. Westra et al., 2024; Roudot et al., 2025). In contrast, values approaching ~20% represent upper-bound or worst-case assumptions in theory based (i.e. non-empirical) studies, and lack real-word observational evidence. Finally, we regret that no published dataset of industrial H_2 emissions from leakage in the value-chain is available to conduct the proposed experiments.

Specific comments line-by-line:

- Line 12, 296, 360: El Nino ~ wet western but dry equatorial-eastern (looking at Cai et al. 2020)?

- L12: Thank you for the comment. Since this refers specifically to the tropical South America RECCAP2 region, we think the current wording is clear and can remain unchanged.
- L296: We agree and will make this statement more specific by adding "in the Amazon".
- L360: Thank you for the suggestion. Since this refers to a specific region (Costa Rica; Aronson et al., 2019), we think the current wording is sufficiently clear.

- Line 22 and 35: It would be helpful to point to the existing constraints for these summarised in Table 2.

Good suggestion, we will add this to the revised manuscript.

- Patterson et al. 2025 also used new observations to find constraints on how sources and deposition might have evolved in the last century.

We agree with the reviewer that Patterson et al. (2025) present an interesting study. However, we do not see a direct connection to the scope of our introduction, as their analysis focuses on changes in hydrogen sources and sinks over the 20th century, whereas our study targets the 2003–2023 period. As such, the two studies are not directly comparable in terms of temporal coverage. In addition, Patterson et al. (2025) do not provide a global budget with mean annual fluxes, so we cannot add them to Table 2.

- Line 24: The distribution of ocean and soil emissions might be poorly constrained. Does the calculation with seawater CO concentration match with REVISED Paulot et al. 2024 ACP supplementary figure 3 where SH ocean H2 emission is reduced versus GFED4s?

Thank you and yes, because we also scale using the oceanic CO emissions from Conte et al (2019) used by Paulot et al (2024).

- Line 40: Northern Hemisphere subtropics + midlatitudes expected to have comparable or higher uptake (Paulot et al. 2024 ACP and Tardito Chaudhri and Stevenson 2025 ACP).

We thank the reviewer for pointing this out and propose changing L40 to: "This introduces a representativeness bias, as tropical regions, which are expected to show high soil H_2 uptake (Tardito Chaudhri and Stevenson, 2025), are not represented in the underlying observations.". This wording is consistent with Figure 2a from Tardito Chaudhri and Stevenson (2025), as well as our prior Sanderson and Bertagni fluxes (Figure 2 of our manuscript).

- Line 96: Here you take the pattern of ocean emissions from CO, then introduce a scaling. My understanding is that we don't necessarily have good constraints on ocean H2 emissions, and inventories for these add a source mostly south of the 25N land bisector (line 189). Increasing this source requires increased uptake south of this divide – previous comment, the revision in Paulot et al. reduces higher-latitude SH ocean emissions to meet observations. Considering this, how robust is your South America posterior uptake to this choice of scaling?

We expect that this dependency is comparitively small given the size of oceanic flux relative to the other sources. Furthermore, our approach is in line with REVISED Paulot et al. 2024, since we also use the Conte Oceanic CO emissions for the distribution of our H_2 oceanic flux and our magnitude is nearly identical (5.0 vs 4.9 Tg/yr).

- Line 171 - 184: A linearisation is made to the uptake scheme, a posterior scheme is then found and the average deviation from linear uptake is then presented. Is it worth checking there aren't particular regions where this linearisation has performed better and worse? This would reassure the reader that your conclusions are sound at the particular locations with the biggest changes in updated fluxes.

The reviewer raises an interesting point here. To investigate this, we plotted the seasonal mean posterior/prior surface H_2 mole fraction ratio for the Sanderson and Bertagni inversions over the 2003 to 2013 period (Figures A and B below). We find that posterior/prior differences do exist and vary somewhat between seasons. Ratios below 1 are found over Central Africa, South America, and Southeast Asia, while ratios above 1 occur over India and Russia during some seasons. This indicates that, after optimisation, the inferred H_2 fluxes are no longer associated with exactly the same surface mole fraction field used in the linearisation. These differences are generally quite modest: for approximately 90% of the surface grid cells, the posterior/prior differences remain well below 10%.

It is also important to note that the inversion optimises the H_2 deposition flux. This flux is determined by the product of the deposition velocity and the surface H_2 mole fraction. Therefore, if the posterior mole fractions were used instead, the resulting differences would simply be projected onto the deposition velocity needed to obtain the same flux. Since the quantity being optimised is the flux rather than the individual components, we do not expect that identifying regions where the linearisation performs relatively better or worse would change the inferred fluxes or the conclusions of this study.

Figure A: Posterior/prior surface mole fraction ratio for Sanderson inversion (seasonal means calculated for 2003-2013 period).

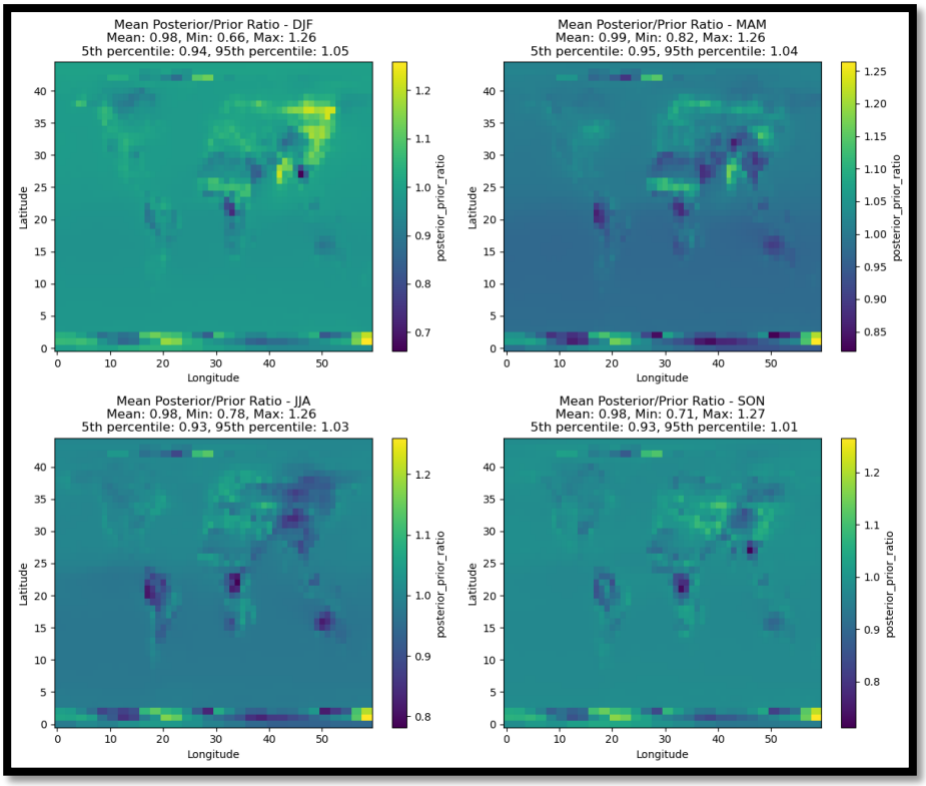
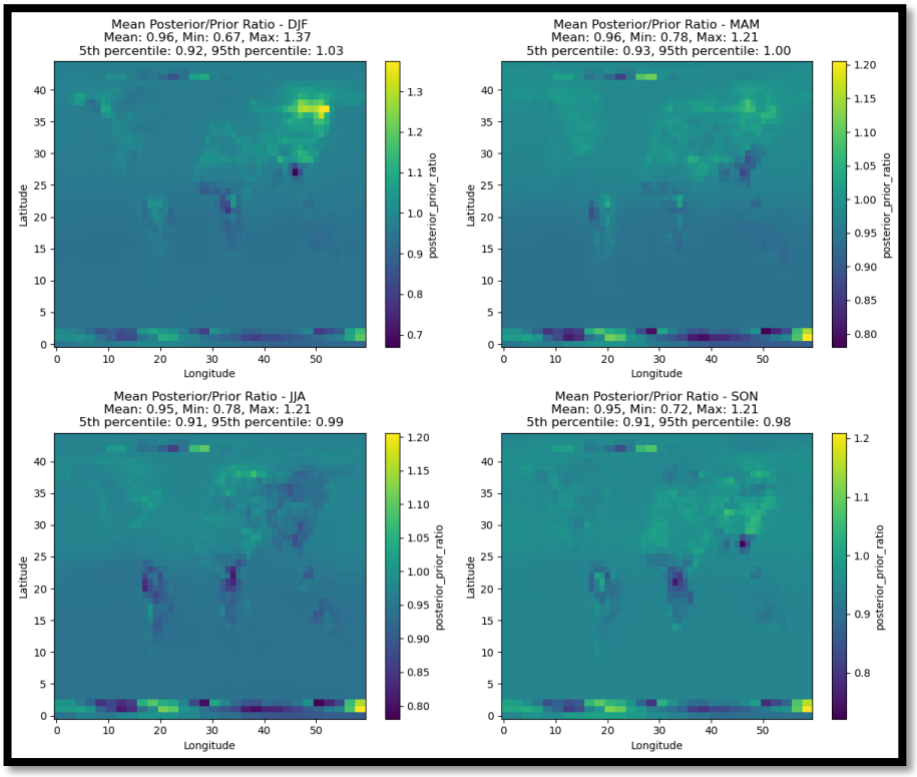


Figure B: Posterior/prior surface mole fraction ratio for Bertagni inversion (seasonal means calculated for 2003-2013 period).



- Line 213-221: Another possible error comes from flask observations typically being made at similar times of day at each site, Forster et al. 2012 record ~10ppb variations in hydrogen from diurnal cycles, with notable drops during night-time inversions.

We thank the reviewer for raising this issue. H_2 mole fraction measurements are generally collected during midday or early afternoon, when atmospheric mixing is strongest and transport errors are smaller than during nighttime stable conditions. These conditions make the observations more representative of larger-scale atmospheric signals and reduce the influence of local boundary-layer variability. In our framework, modelled H_2 concentrations are sampled at the actual time of the flask measurements, ensuring a consistent observation–model comparison. Therefore, we believe that the timing of most flask sampling is in fact beneficial for our inversions.

- Line 247-248: Please can this sentence be made clearer? ‘Due to the larger freedom to adjust sources in sinks …’

We have removed this part from Line 247 so it becomes: “Our detailed inversions substantially improve the agreement between simulated and observed H_2 mole fractions used in the inversions”.

- Line 267: Would this suggest that the prior configuration would lead to too-high H2 in the southern hemisphere versus Northern Hemisphere? Would it be clearer to also include the latitudinal distribution that the prior would have achieved and what the posterior achieves?

Yes, Fig 4. Shows that our priors overestimated H_2 mole fractions in the SH. The latitudinal distribution of the prior and posterior soil sink is shown in Fig. E2 and E3.

- This might be beyond scope, but would a configuration with higher anthropogenic H2 emission (more NH) see increased but similar pattern of H2 uptake as the prior?

If we understand the reviewer correctly, they are asking whether a prior with higher NH emissions, would result in a posterior soil sink that more closely resembles the prior. We think that this would be likely, since an increase in sources is probably compensated for by larger sinks. However, within our modelling framework, we consider it unlikely that anthropogenic emissions would need to be substantially increased as our fossil fuel burning source is already large compared to the Ouyang et al. (2025) range.

- Line 317 and Fig 7e: This result is counter-intuitive. If we assume rates of microbial activity increase up to about 36C and that global patterns of soil moisture have not generally changed over 2003-2023 then intuition would suggest that it would be more likely that soil uptake would have increased rather than decreased?

We agree that the result may appear counter-intuitive under simplified assumptions of a lack of trend in soil moisture and the rightfully mentioned expected temperature optimum for microbial activity. However, the premise that global soil moisture patterns have remained broadly unchanged over 2003–2023 is not well supported. Multiple studies report substantial regional-scale trends during this period (e.g., Hirschi et al., 2025; Wang et al., 2021). Importantly, the net effect of such trends on soil uptake is not straightforward, given that both moisture limitation and saturation effects can reduce uptake, and the relevant optima likely vary across regions and systems. As such, we do not find sufficient grounds to rule out that the soil sink has decreased on a global level over this period, especially as our inferred trend is relatively small (0.23 Tg/yr^2 over the 2003-2023 period) in comparison to the overall soil sink magnitude (52.8 Tg/yr).

- Line 325: However, this could be a useful result. Is it easy to estimate HD depletion rates that would be expected for the constraints found in this work versus an emission change explanation?

We agree with the reviewer that HD could help evaluate potential misattribution between source and sink changes, as outlined in Section 4.1. However, implementing and analyzing this would require an amount of work that exceeds the scope of the current project. Even then, we are limited by the lack of recent HD observations over the past 15 years. The available measurements date from 2006–2011 and are the same observations used by Pieterse et al. (2013), meaning that no new measurements are currently available to provide an updated assessment. Efforts to extend our TM5 model to allow their interpretation are currently underway, building on the framework developed by Pieterse et al. (2013).

References provided by the reviewer:
Patterson et al. 2025, https://agupubs.onlinelibrary.wiley.com/doi/epdf/10.1029/2025JD043662
Paulot et al. 2024, ACP, https://acp.copernicus.org/articles/24/4217/2024/
Esquivel-Elizondo et al. 2023, Front. Energy Res, https://www.frontiersin.org/journals/energy-research/articles/10.3389/fenrg.2023.1207208/full
Cai et al 2020, https://www.nature.com/articles/s43017-020-0040-3
Forster et al. 2012, https://www.tandfonline.com/doi/full/10.3402/tellusb.v64i0.17771

References used in replies:
Stone et al. (2026), https://doi.org/10.5194/acp-26-10241-2026
Westra et al. (2024), https://doi.org/10.1038/s41598-024-76373-2
Roudot et al. (2025), http://dx.doi.org/10.2139/ssrn.6041754
Hirschi et al. (2025), https://doi.org/10.5194/hess-29-397-2025
Wang et al. (2021), https://doi.org/10.5194/essd-13-4385-2021

Reviewer 2 (Anonymous Referee #2):

We thank the reviewer for their kind words about our study and for taking the time to critically evaluate it. We were able to make several improvements to the manuscript based on the comments. Below, we provide our proposed clarifications and corrections (text in blue).

Specific comments:

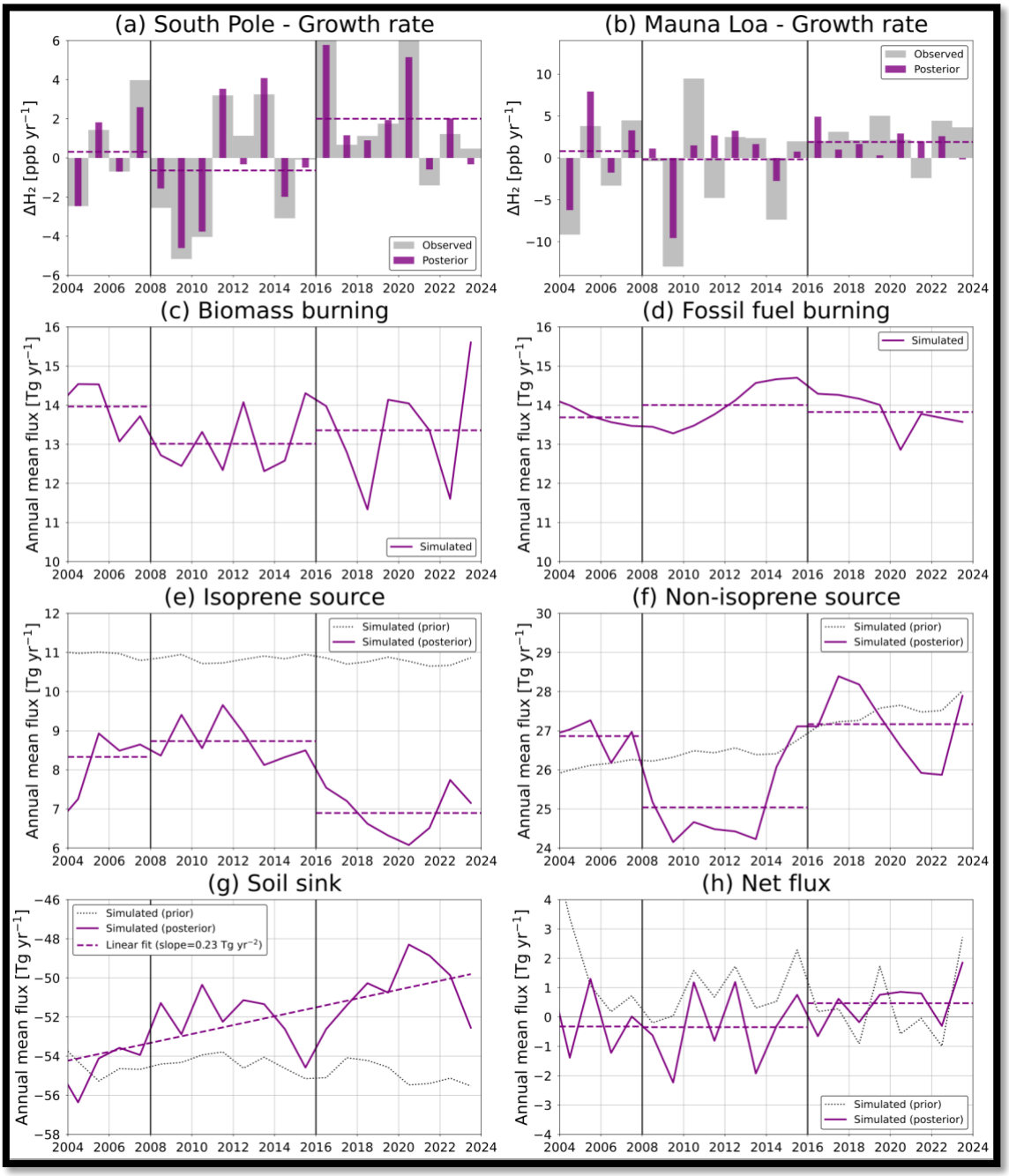
- L53 "…Bousquet et al. (2011) used observations from 1991-2004, despite nearly two decades of additional observations being available since then." This comment seems a bit unfair, since Bousquet et al. could hardly have used observations from the future in 2011! I guess you mean something slightly different.

We agree with the reviewer that this was awkwardly formulated and propose rephrasing L53 to: "Subsequent studies employing three-dimensional chemical transport models yielded budgets more consistent with bottom-up literature, but relied on shorter observational records. For example, Pison et al. (2009) used observations from a single year (2004), Yver et al. (2011) relied on relatively few stations, mostly in Europe, and Bousquet et al. (2011) analysed observations from 1991–2004. Since then, the observational record has expanded considerably, providing a clear opportunity to revisit this problem."

- L92 I am slightly unclear on the fossil fuel related H2 emissions derived from scaling CO emissions. If I understand correctly, your simulations cover 2003-2023 (it would be useful to explicitly state this near the start of Section 2.2) and use CO emissions from CAMS and H2:CO emission ratios from CEDS. So presumably there is a significant trend of these H2 emissions over 2003-2023? Later (e.g., Figure 7), you discuss the growth rate of H2 and relate this to trends in the photochemical sources of H2, the soil sink, and biomass burning emissions. But what about the trend in anthropogenic emissions? This doesn't seem to be documented anywhere in the paper. Don't we also need to know the contribution of that to trends in atmospheric H2? Isn't there also some (quite large?) uncertainty in the magnitude and trend of this source of H2? Or are you assuming this is small and insignificant? If so, say so.

Thank you, we will mention the period again in Section 2.2. The reason we originally did not include fossil fuel emissions in Fig. 7 was that they show no clear long-term trend (the trend is negligible and amounts to 0.003 Tg/yr^2) between 2003-2023. Also, the period means (i.e. the means over 2003-2008, 2008-2016, 2016-2024) differ by less than 0.2 Tg/yr (see Fig. below). However, we agree that this could be clearer. Therefore, we have updated Section 3.2.4 to provide a more comprehensive discussion of the temporal trends in prior and posterior sources in relation to the growth-rate (including the updated Figure below).

Figure A: Observed and simulated (posterior) annual growth rates at SPO (a) and MLO (b). Remaining panels show global annual fluxes for (c) biomass burning, (d) fossil fuel burning, (e) isoprene H_2 source, (f) non-isoprene H_2 source, (g) soil sink, and (h) net sources and sinks. The vertical lines delineate three periods (2003–2008, 2008–2016, 2016–2024) that characterise the observed H_2 growth rate.



- L99 The biogenic source of H2 is also unclear. You say you use the biogenic CO fluxes from MEGAN, scaled to 3 Tg/yr (I guess you mean 3 Tg(H2)/yr?). Is this fixed over the whole period, with no inter-annual variation? Or is it scaled to 3 Tg/yr in the first year, then allowed to vary inter-annually, based on MEGAN? How do you represent the biogenic source of H2 that comes from isoprene (and other biogenic VOCs)? (Is the biogenic CO somehow approximating the source from BVOCs?) Presumably isoprene emissions are included explicitly in the full chemistry runs (L112)? (You say you derive the photochemical H2 source from isoprene oxidation, but don't say anything about isoprene emissions). Similar to the previous comment, why don't you show the trend and variability in the biogenic H2 source in Figure 7? If you are assuming it is small and insignificant, explicitly say so.

The biogenic H_2 land fluxes are prescribed as a repeating monthly climatology scaled to an annual total of 3 Tg/yr according to the biogenic CO land fluxes throughout the study period. As such, they do not show interannual variability or a long-term trend. We agree that this could be stated more explicitly in the manuscript and have updated L99 and L96 (since this was also not explicit for the oceanic H_2 source).

To make this more clear, we propose changing L96-97 to: "Oceanic H_2 emissions are prescribed from an annually repeating monthly climatology based on the oceanic CO flux distribution of Conte et al. (2019), scaled to a global annual emission of 5 Tg yr^{-1} (Price et al., 2007)." And L99-101 to: "Similarly, land H_2 emissions are prescribed from an annually repeating monthly climatology based on biogenic CO land fluxes from MEGAN (Model of Emissions of Gases and Aerosols from Nature; Guenther et al., 2012), scaled to a global annual total of 3 Tg yr^{-1} (Conrad et al., 1989; Conrad, 1996; Pieterse et al., 2011)."

The biogenic H_2 source does not include secondary H_2 production from the oxidation of biogenic VOCs, as this source is accounted for separately in the photochemical production term, which is driven by monthly interannually varying biogenic VOC emissions in the full chemistry simulations. Over our simulation period, global isoprene (C_5H_8) emissions averaged 598.23 Tg C_5H_8 yr^{-1}, ranging from 519.79 to 671.68 Tg C_5H_8 yr^{-1}.

- L129 It is perhaps worth noting that the seven ecosystems in the Sanderson et al. (2003) scheme do not very obviously relate to different soil types, whereas the Bertagni et al. (2021) scheme does this more explicitly.

We thank the reviewer for this suggestion and propose changing L128 to: "This scheme defines seven ecosystem regions representing broad ecosystem categories rather than explicit soil types, and is driven by soil moisture and soil temperature fields from ERA5 (Hersbach et al., 2020)."

- L139 Thanks for referencing the source for the H2 + OH reaction rate (Stuhl and Niki, 1972). But it would also be useful to know if this is also the recommendation from more up-to-date compilations of reaction rates (i.e., IUPAC, JPL), or perhaps they do not document this reaction?

Thanks for bringing this up. This citation was not correct, as we used the reaction rate from the latest JPL version (Chemical Kinetics and Photochemical Data for Use in Atmospheric Studies Evaluation Number 20; reaction B7) in all experiments. We will update this citation accordingly.

- L207 You use "statevector", "state-vector" and "state vector" at various points in the text. You should probably decide which one is correct and consistently use it.

We agree that this is confusing and will only use "statevector" in the revised manuscript.

- L213 Please explicitly define what you mean by "model-data mismatch". Is this the maximum difference between model and observations that you allow in the inversions?

Thank you for asking for this clarification. In our modelling system, the Model-data mismatch (MDM) is the error assigned to each observation to account for measurement errors as well as modelling errors of that observation. Rather than a threshold, MDM sets the relative weight given to each observation in the inversion, where smaller values mean stronger constraint on the posterior flux. We will update Section 2.5.5. to make this more clear.

- L244 The caption for Figure 3 doesn't define the cross and triangle points (you may just want to say "defined in the text").

We are sorry for the confusion, but we describe these icons in the legend.

- L278 The caption for Figure 6 says "Photochemical source…", but the figure only seems to show soil sink results? (Are you referring to Figure E4?) The text discussing Figure 6 (l272-274) also seems to imply the figure shows results about the photochemical source. Please clarify.

We will update the caption to only refer to the soil sink.

- L280 Suggest add "…shows the observed annual H2 growth rate at 2 sites (SPO and MLO)…"

We will follow this suggestion in the revised manuscript.

- L280-285 Suggest say Figure 7a, 7b, 7c, 7d rather than "panel a" etc.

We will follow this suggestion in the revised manuscript.

- L291 Figure 7 – See my earlier comments: what happens to other H2 emissions and the isoprene-related source of H2?

We will update Section 3.2.4. accordingly (see earlier reply on comment about L99).

- L295 Suggest change to "…in most regions, including tropical South America, the soil sink is dominantly diffusion-limited…"

We will follow this suggestion in the revised manuscript.

- L300 Figure 8 Does the Oceanic Nino Index have units of K, or is it dimensionless?

Thanks for pointing this out; In the cited publication (Barnston et al., 1997), units [°C] are used. We will update Figure 8 accordingly.

- L301 Section 4 – suggest add some discussion about trends in anthropogenic H2 emissions (based on earlier comments).

We will update Section 4.1 accordingly.

- L328 In addition to the isotopic signature of fossil fuel emissions changing with engine technology, I suspect the H2:CO ratio of the emissions may also have changed.

Good suggestion, we will update Section 4.1 accordingly (L329).

- L354 There has been some work on the HCHO photolysis rate by Holland et al. (https://meetingorganizer.copernicus.org/EGU24/EGU24-3142.html). I am unsure if there is an associated publication, rather than just a conference abstract.

Thanks for referring to this conference abstract. We could not find any associated publications. However, we did perform an additional analysis in response to the Community Comment of Alexander Archibald on our manuscript (see this response). Based on this comparison with ATom derived J-values (photolysis rate constant), we have updated section 4.2 and Appendix A of the manuscript accordingly.

- P21 Figure A2 caption. Are these data based on all ATom flights (N=35 mentioned in the text, L395)? This seems inconsistent with the 48 ATom "Stations/Flights" noted in Table E1?

Thank you for pointing this out. In fact, our comparison spans 41 ATom flights, we will update the contents and captions of Figure A2 and Table E2 accordingly.

- L404 I understand that the underestimation of HCHO "…is likely related to an overestimation of the photolysis rate of CH2O…" But if this is the case, and HCHO was higher, but the flux through HCHO + h.nu was less, would the H2 source be larger or smaller?

We thank the reviewer for this comment. Please see our response to the Community Comment by Alexander Archibald on this manuscript, where we describe the additional analysis performed in response to his comments. Based on a comparison with ATom-derived J-values (photolysis rate constants), we confirmed an overestimation of J and have updated Sect. 4.2 and Appendix A accordingly.

- L418 I don't really understand what a "5-year value" (or "monthly scaling" (L416)) means here. Do you mean you use a 5-year running mean (with a moving window) (for each month)? What does "multi-year correlations" in the caption for Table B1 refer to? Please clarify.

The "5-year value" does not refer to a 5-year running mean or a moving window. Instead, it refers to the interannual correlation timescale prescribed for the prior flux uncertainties, where observations from neighbouring years influence the optimized fluxes, with observations from years closer in time having a stronger influence than

observations from years further away. The term “multi-year correlations” in the caption of Table B1 refers to these prescribed interannual correlation timescales. We have rewritten Section 2.5.3 and Appendix B to make this fully clear.

- L420 Here you refer to “the isoprene source” – so, related to my earlier comment about biogenic emissions, are you using MEGAN for isoprene emissions? (Unclear in Section 2.3).

Here, our intention was to discuss the photochemical production of H_2 originating from isoprene, not the isoprene emissions itself. We understand the confusion here due to the current phrasing and propose changing “isoprene source” to “isoprene H_2 source” in lines 115, 116, 120, 370, 470.

In the full chemistry run, monthly, interannually varying emissions of the actual species (i.e., isoprene, rather than H_2) were used from CMIP6 for anthropogenic sources and MEGAN for biogenic sources. Global isoprene emissions over the simulation period averaged 598.23 Tg yr^{-1}, ranging from 519.79 to 671.68 Tg yr^{-1}. We will update line 114 to specify that emissions for ‘almost all’ remaining species are taken from CMIP6, since isoprene provides an example of a species for which emissions are instead derived from MEGAN.

- L424 I am surprised the inversions sometimes generate “negative photochemical sources”. Should the inversion be allowed to do this, since that would be physically impossible? Is it not feasible to force the inversion to not do this? (On the other hand, a “positive soil sink” – i.e., a soil source, is not out of the question – occasionally chamber measurements of soils show a source rather than a sink, although, again, if you feel this is physically unreasonable, I don’t understand why you can’t force your inversion to not allow this).

We thank the reviewer for this comment. We agree that a positive soil flux (i.e., a soil source) is physically possible, as reported occasionally in chamber measurements, whereas negative photochemical production is unphysical. The occurrence of small negative photochemical H_2 sources is a statistically correct solution to the Bayesian problem solved in our EnKF inversion framework. Such local negative values are often compensated by nearby positive ones that balance the regional photochemical budget to a source. This phenomenon is often referred to as “dipoles” in the solution which have turned out to be very hard to fully prevent, also in other systems. A hard positivity constraint could be imposed within our system but comes with other disadvantages (non-linear solution, different distribution of the covariances). We have confirmed that the occasional occurrence of a dipole did not affect our aggregated solutions. We therefore chose to mitigate this minor issue by evaluating the sensitivity of the inversion setup and uncertainty assumptions through the experiments presented in Appendix B.

- P24 Table B2 footnotes – all the numbers should have units of ppb, i.e., 1.5 ppb MDM, etc.

Thank you, we will update the footnotes accordingly.

- P25 Figure B3 caption – what does “innovation” chi-squared mean? Please clarify.

The innovation chi-squared (χ^2_{ino}) is the ratio between the residual (i.e. observed mole fraction – prior simulated mole fraction) and the sum of the statevector (P) and observational uncertainties (R). A chi-squared of 1 means the residuals and uncertainties are perfectly balanced, while a higher chi-squared value indicates that uncertainties are underestimated, and vice versa. The posterior chi-squared (χ^2_{post}) is similar, only the posterior simulated mole fraction is used in the calculation of the residual and the statevector uncertainty is no longer in the denominator.

$$\chi^2_{ino} = \frac{(y - H(x_{prior}))^2}{(HPH^T + R)}$$

$$\chi^2_{post} = \frac{(y - H(x_{posterior}))^2}{(R)}$$

For further explanation of these metrics, please see:
Tarantola (2005), https://doi.org/10.1137/1.9780898717921

- P29 Figure E1 – suggest add a line at 25N to separate the two regions.

Nice suggestion, we will update Fig. E1 accordingly.

- P30/31 Figure E2/E3. These maps do not show global mean values! (They vary spatially!)

We thank the reviewer for pointing this out. We will change “Global mean fluxes (2003–2023)” to “Annual mean fluxes (2003–2023)” in these captions.

- P32 Figure E4. Please clarify that “non-isoprene source” and “isoprene source” refer to photochemical sources of H2, and not H2 emissions. (I assume).

This assumption is correct, as explained in L115 and L116.

- P34 Table E1. See earlier comment – Are there 35 or 48 ATom flights?

Thank you for noting this. There are in fact 41 ATom flights, we will update Table A1 and E1 and their captions accordingly.

- L498 neither -> none?

We will update the text accordingly.

- L224 The choice of the four observational metrics is not well justified and feels a bit arbitrary. The choices are not unreasonable, but could you better justify: Why only five stations were chosen for the mean RMSE metric (and why those five?); Why MLO for the growth rate (why not more/all stations?); Why not all/more high-latitude sites for the interpolar difference? And Why only use Mace Head for the seasonal amplitude – why not more/all sites?
- I also wasn’t entirely clear on how these metrics were used to identify the best-performing configurations in Figure 3. Is this a quantitative process? E.g., is each metric weighted equally? Given the relatively arbitrary choice of metrics, this process felt only semi-quantitative at best.
- Maybe the point is these metrics are only used to loosely identify global scalings that are refined in the detailed inversions, so they don't need to be very rigorous. Is that the correct way to think about it? This process could perhaps be more clearly explained.

We acknowledge that the selection of individual stations and metrics is an explicit choice in the setup of our study. However, these choices were made a priori, without knowledge of the model results, and were based on the availability of long observational records, the representativeness of the selected sites for their respective regions, and their relevance for evaluating key features of atmospheric H_2 variability. The metrics were not intended as a formal quantitative optimization framework, but rather as diagnostics to identify plausible global scaling factors before performing the detailed inversions.

- l247 Should this read "Due to the larger freedom to adjust sources and sinks..."?

Thank you for noting this, it should indeed have been ‘and’ here. We have removed this part of the sentence altogether in response to a comment from the other reviewer, and will change L247 to:
“Our detailed inversions substantially improve the agreement between simulated and observed H_2 mole fractions used in the inversions.”

References used in replies:
Tarantola (2005), https://doi.org/10.1137/1.9780898717921