Response to reviewers on "Quantifying and addressing the uncertainties in tropospheric ozone and OH in a global chemistry transport model" by O. Wild and E. Ryan (egusphere-2025-4534)

We thank both reviewers for their comments. Reviewer comments are in black, our responses are in blue, and additions to the paper are shown in green. Line numbers used for changes in the text refer to the original submitted version of the manuscript to retain consistency with those used by the reviewers.

Response to Reviewer 1:

This thoughtful and novel study by Wild and Ryan centers on using a chemistry-transport model to understand how the uncertainties in our current knowledge of the physics and chemistry in the model translate into uncertainties in the model generated tropospheric O_3 and OH (the major sink of CH_4). The current state of understanding is readily seen in the recent model intercomparison projects (MIPs), where the budgets for O_3 and CH_4 show wide spreads. So, yes, this is an important area of study for both air quality and climate.

The authors perform an intense series of objective sensitivity simulations with the CTM involving 60 some identified uncertainties and reduce this to 36 for more serious study. This is the most thorough such analysis that this reviewer has seen. The authors then use "Gaussian process emulators" to further study the full range of coupled uncertainties in order to map out cost functions for matching some key large-scale observational constrains (surface and 500 hPa O_3, total tropospheric O_3 burden, CH_4 lifetime).

This work is a valuable contribution to the community's understanding of the how and why our models do not always give us what we want. The only major problem is the denseness of the manuscript and the difficulty in keeping track of the specific uncertainties. I have the following suggestions for minor rewrites.

We thank the reviewer for their positive comments here and address specific concerns below.

Starting with Table 1. This table provides the core data and assumptions used here. The problem is that the reader has little chance of remembering all the 3 letter abbreviations later in the paper. Can the authors come up with a reasonable, abbreviated table for the 36 key uncertainties that could be included in the legend of some of the later figures? Shorthand examples:

fir	BB all
iso	biogenic C_5H_8
hna	Henry's HNO_3
xfm	X-section HCHO

In our original submission we wrestled with the challenge of making the individual parameters both distinct and interpretable while also sufficiently concise to include in figures and tables. The parameters are defined in full detail in Table 1, and we have included shorthand definitions in Figure 1 and Table S1. We have attempted to keep the ordering consistent wherever possible colour-coding the different families of parameters in Figs 2, 7 and 8, and labelling parameters in the caption in Figure 3. We have now revised the captions of Figs 2, 3, 7 and 8 to make the parameters shown easier to interpret (see responses to specific comments below).

Some of the paragraphs are very dense and difficult to follow. Would it be possible to segment these, maybe with bulleted points to give us the sequence of the logic?

We apologize if the reviewer finds parts of the text difficult to follow and have attempted to identify and break up longer paragraphs. We are reluctant to punctuate the discussion with bullet points, as this breaks the flow of the text, but have attempted to make the sequence of argument clearer.

L5: "even when we all agree on the physical or chemical rate constants" Yes, indeed, the diversity is distressing.

L39: But these studies did not look at photochemistry since they did not do cross sections!

The Newsome and Evans study cited considered photolytic reactions, but the reviewer is correct to point out that the uncertainties were not based on the cross sections. We have removed "photo" here to reflect this point but choose not to add an additional statement on the novelty of our study at this point in the discussion as it would break the flow of the argument.

L112: Yes, it is good you are dealing with uncertainties in the numbers we often assume are fixed.

Thank you.

L160: Thanks for doing the cross sections.

L163: is it good to see a clear definition of you uncertainty ranges. You might also note than when you report an uncertainty factor, it is boeh / and x, effectively you are treating the internal range as lon-normal??

Yes, the internal range is effectively log-normal for some variables; this is how the uncertainty in kinetic parameters is conventionally expressed, and it is also appropriate for parameters where the uncertainty used is large (e.g. for deposition velocities).

L172 of the 2-sigma (95% range) uncertainty. Your uncertainty prob distrib is not stopped at 2 sigma.

This is of course correct, and there is a small probability of values lying outside the 2-sigma ranges defined. We have adjusted the text here to make this clearer.

Line 172: replace "uncertainty" with "2-sigma uncertainty"

For the global sensitivity analysis, we have chosen to assume a uniform sampling distribution across a bounded range defined by the 2-sigma uncertainty as this is conceptually simple. Applying normal or log-normal distributions would have allowed us to incorporate the 5% of values outside the 2-sigma range but would have led to a narrowing in the output uncertainty which might suggest over-confidence. It would be valuable to explore the sensitivity of the results to the choice of sampling distribution in future work. We have added a sentence in the global sensitivity analysis section to make our assumption of a uniform distribution clearer.

Line 225: "We assume a uniform distribution of parameter uncertainty rather than a normal/log-normal distribution to obtain a broad estimate of model uncertainty and avoid heavy emphasis on prior values."

Fig. 1: I really like Fig. 1 and would expect near perfect symmetry in the up/down perturbations, but items like VOC emissions are asymmetric, presumably because you did log normal with a large factor. You should comment on that.

If responses were linear over the ranges considered here, then we might expect symmetry, but in many cases they are not, so some asymmetry is natural. Strong nonlinearities may give more symmetric responses with log normal input uncertainty than with normal input uncertainty, although there aren't clear examples of this here. However, we have now added a comment on the asymmetry as suggested.

Line 200: "although we note that some asymmetry may be expected where we have assumed a logarithmic uncertainty range"

L255: I am a bit worried about the apportionments based on log-normal ranges/

The variance-based decomposition is performed on the normalised uncertainty ranges, so all input ranges are transformed to a linear 0-1 scale, taking account of the assumed parameter scaling that was applied.

Line 255: "Both linear and logarithmic parameter scalings are accommodated by transforming the input ranges to a linear 0 to 1 scale for each parameter."

Fig 2: This really needs a legend with a somewhat expanded explanation for the 3 letter designations.

All 36 parameters are shown in this figure, and repeating a full list of definitions would make the figure difficult to read. We have therefore clarified the caption here, explained the meaning of the colours (which makes process identification easier) and referred the reader to Table 1 for the full set of parameter definitions.

In Fig 2 caption: "Parameters are grouped and ranked by type of process (different colours) and full definitions are given in Table 1."

Fig 3: ditto

Only 10 parameters are shown in this figure, and their definitions are now provided in full in the figure caption.

In Fig 3 caption: "Parameters shown include the stratospheric O_3 column (str), dry deposition to forest, grassland and water/bare ground (ddf, ddg, ddw), emissions from fire (fir) and anthropogenic and biogenic sources of VOC (voc, iso) and chemical rate constants for NO_2+OH (kn2), HO_2+O_3 (kh2) and $O(^1D)$+H_2O (k1d)."

Fig.7: Ditto. This paddle figure is challenging for me, but OK

We have amended the caption to make process identification clearer, and we have altered the colour scale on Figure 2 so that both figures use the same colour scheme.

Fig 7 caption: "Parameters are coloured by process type following the same scheme as Fig 2…"

L325ff: You are worried about the vertical resolution in comparing surface O3, but I think the observations are MDA8 or daytime and so a thin nocturnal BL is not a problem. I would be more worried about the horizontal resolution because you cannot resolve urban centers.

We are using 24-hr mean observations, and hence nighttime biases remain a problem, although isolating MDA8 or daytime ozone could partly circumvent this issue. However, horizontal averaging over polluted and remote environments within a grid box is also an important factor, as the reviewer notes, and we have added a sentence to acknowledge this here.

Line 326: “The model is unable to represent the range of clean and polluted environments that may lie within a grid box. In addition, the lowest model layer is nearly 100 m deep, and this precludes representation of sub-grid near-surface ozone…”

L341ff: This is an interesting statement about structural errors. We all certainly have them, But, do they apply in a similar way to all models/

It is likely that some structural errors are common across models, reflecting processes that all models are missing, and that others are not, reflecting different approaches to representing specific processes. We aim to explore this in future studies, and have included a few sentences on this at the end of Section 7.

Fig 7: What is the color code here? More important, why does everything seem to be pushed away from the prior?

The colour coding is by process type and is intended to match that in Figure 2, to aid interpretation (the colour coding wasn’t identical, and this has now been corrected). A comment has been added to the caption to indicate this. The pushing of many processes away from the prior indicates that the prior is not consistent with the observations, although as pointed out in the text this is largely due to the influence of structural errors at this stage; when these are addressed we would expect most processes to lie much closer to the prior. We were missing an overview statement noting substantial deviations from the prior and have now added a sentence to address this.

Fig 7 caption: “Parameters are coloured by process type following the same scheme as Fig 2…”

Adjustments to line 395: “The posterior distributions for some of the key parameters lie at the extremes of the specified uncertainty ranges, far from the prior values, as a consequence of trying to fit observations outside the uncertainty space…”

L411: Nice case study, but too bad it did not work out better.

Agreed, although demonstrating the existence of major biases is in itself very useful!

Fig 8: Again, why is everything (except ddf & ddg) pushed far off the prior? Some discussion.

We have amended the discussion here to note that posterior parameter values still lie far from prior estimates for some parameters, and the subsequent discussion interprets this.

Line 449: “The posterior probability distributions for the constrained parameters remain similar to those for the uncorrected model, with some parameters still far from prior estimates, although the ranges are somewhat smaller and fewer parameters lie towards the extremes of their uncertainty ranges.”

L476ff: Why not do trop O3 burden as NH and SH, since there are distinct differences with separate causes presumably

Yes, this would provide additional information and could be worth exploring in future. However, it would be more meaningful to explore this once the structural errors identified here have been suitably addressed.

L482: Note that errors in J-NO2 would do the same thing?

Uncertainties in J(NO_2) are certainly important, affecting the NO/NO_2 ratio in particular, but we find that as the key chemical sink of NOx, the NO_2+OH is of greater importance, and also has a larger associated uncertainty than many of the other kinetic parameters.

L492: Great, major conclusion!

Thanks.

Response to Reviewer 2:

The manuscript demonstrates the potential for an uncertainty quantification framework built on Gaussian process emulation to identify specific processes most in need of constraining. An impressively large amount of work underlies the results, first identifying key sensitivities with brute force model simulations (60 processes), then developing and training emulators from a subset of 36 processes to target global tropospheric ozone, methane lifetime, surface and 500 hPa ozone sampled at 60 10x10 degree gridded observations. While the processes emphasized here have been pointed out in prior work as driving differences across models or between observations and models, the novelty is the ranking among individual parameters within the categories of emissions, deposition, meteorology, photolysis, and chemistry. The manuscript will be a useful contribution though some clarification noted below would help readers better interpret the findings. Importantly, the authors are providing the perturbation simulations and emulator codes in an archive, which makes them useful to benchmark future work.

Many thanks for your positive assessment of our study. We hope that both the ranking and the approaches used will be of benefit to the broader community.

General comments.

It's tempting to interpret the dominant sensitivities as causal, but is it possible that inter-correlations among the 36 tested processes in the training confound a clean interpretation? What assumptions are made regarding temporal and spatial correlations especially considering variations in ozone lifetime in space and time? Is there an assumption that ozone is locally controlled, and if so, where is that likely to work best?

The reviewer is correct to point out that not all the processes considered here are independent, and that strong correlation between different processes could confound the results. However, the variance-based decomposition we use to calculate the sensitivity indices provides quantification of the interactions between different processes, and these are found to be small for the global quantities considered here and are thus neglected (we state this at Line 255: The interaction between parameters can also be quantified through this approach, but these are relatively small for the metrics we consider here and we choose to neglect them for simplicity). In addition, the dominant sensitivities from the emulation match well with those from the one-at-a-time model studies, which are causal, lending strong confidence in our results.

While interactions are shown not to be important for the metrics we consider here, there may be local metrics or environments in which they are important, and the interaction terms therefore need to be assessed on a case-by-case basis when planning emulation studies, and cannot be blindly neglected.

Temporal and spatial correlations do not require special consideration here and there is no assumption of local control, as these relationships are inherent in the model simulations.

What are the ‘attractive statistical properties’ in line 79?

The attractive statistical properties referred to here are that GP emulation provides not only an estimate of the output value at any point in the parameter space, but also an estimate of its associated uncertainty, giving an indication of its reliability. This emulator uncertainty is illustrated in Figure S1 which shows the emulator validation. The benefits of the approach are described in more detail in the two references cited here, but we have added a statement to clarify what is meant here.

Line 79: “e.g., allowing direct assessment of the uncertainty in the emulated output”

Transport seems like a critical missing piece.

Transport remains a missing piece that we are unable to address cleanly in an offline model framework (although we adopt a simple efficiency scaling to assess the importance of convection, and this could also be crudely applied to advection processes). We already note this in Lines 70, 142, 402 and 510. We will explore approaches to address this in future studies but have clarified the text in the conclusions to make the neglect of transport processes clearer.

Line 510: “We have also been unable to explore the impact of uncertainty in meteorological transport processes, which are external to our offline model, or in interactions with….”

Is the main point that the approach is generalizable but the findings model-specific, or are sensitivities identified here applicable to the real atmosphere or other models? For example, the study design maps out which errors would matter most in which region given the base state distribution. So if that base state distribution were to change dramatically, so too would the rankings identified here?

While the sensitivities outlined here relate to one model, we expect very similar results from other models where the base state lies relatively close to that observed. Our results for specific parameters are very similar to those from previous studies where these are available (e.g., NO_2+OH from Newsome and Evans 2017, cited on line 184). However, the sensitivities are a product of the model formulation, so different parameterizations for individual processes may lead to different sensitivities. We demonstrated this in an earlier study (Wild et al., 2020) where we identified and explained differing model responses in tropospheric OH. We aim to perform similar ensemble simulations using other models in future to explore this further, but we have added an additional statement to note that we expect similar results with other models.

Addition at line 487: “While we expect broadly similar sensitivity to uncertainty in these processes across models….”

Specific Comments.

How are NOx and HOx reservoirs handled by the emulator? Murray et al. (2021) showed strong cross-model sensitivity of global OH to the treatment of NOy and how much emitted carbon is oxidized all the way to CO2, which is understandably beyond the processes studied here and discussed in lines 485-489 but could be acknowledged. Where does this model mechanism fall in terms of chemical complexity? In particular, understanding of isoprene chemistry has advanced rapidly in recent years (e.g., Bates and Jacob, 2019); are those updates in this model?

The emulator reproduces the model output without considering the model formulation, so it is not aware of details of the chemistry treatment. The chemistry scheme used in this model is relatively simple, and while it captures global gas-phase oxidant chemistry reasonably well, it does not represent complex VOC chemistry, halogen chemistry, aerosol interactions, and many other processes that may be important in specific environments. This simplification is a structural feature of the model, and weaknesses in the chemistry constitute structural uncertainty. Observational constraint will help us to identify these biases to inform model development, and we are currently working on refinements to the chemistry (halogens, VOC) in light of our results.

Addition at line 62: "The model includes explicit treatment of HOx/Ox/NOx chemistry and CH_4 oxidation and has a lumped chemistry for representative alkane, alkene and aromatic species and isoprene."

Line 44. Observational constraints are being used to investigate tropospheric oxidation processes, most notably around field campaigns and satellite data. Some examples: Nicely et al., 2016; Bourgeois et al.,2021; Pimlott et al., 2022; Anderson et al., 2023; Baublitz et al., 2023; Guo et al., 2023, Mirrezaei et al., 2025 and others.

The statement made in the text was intended to refer to formal statistical constraint through observation-based methods rather than to broader use of observations to help tune or adjust models. However, we did not mean to downplay the valuable work done in this field and have adjusted the sentence to acknowledge this.

Line 42: Replace "Observational constraints" with "Formal statistical constraints".

Lines 62-65. Is the sensitivity of CH4 to H2 at present levels correct, and if so, is there a soil sink for H2 in the model?

In these studies, we have used fixed tropospheric distributions of CH_4 and H_2 rather than treating them as tracers and incorporating emissions and deposition (as noted on line 129), and thus no soil sink is needed. The previous version of this model neglected H_2 chemistry entirely; including it here increases the chemical lifetime of CH_4 by about 0.5 years, and this sensitivity agrees well with the findings of previous studies (e.g., Warwick et al., 2023).

Section starting line 72. How long was the model spun up? That comes later but might fit here.

The model was spin up for 8 months from a previous steady-state run, which is sufficient for simulations using fixed CH_4 distributions. This is stated on line 227, where details of the model simulations are first described, and we feel that the information belongs better there.

Table 1. Consider adding the base case values for parameters to make it easier for readers to contextualize the findings. Is isoprene the only non-anthropogenic VOC emitted?

Adding values for all parameters would make the table very busy and would be difficult for variable quantities and meteorological processes. However, base case values would be useful for parameters such as emissions, so these have now been incorporated in the accompanying text in Section 3.

Line 127: “To aid reproducibility with other models we have used standardised base emissions of 30 TgN/yr for surface anthropogenic NOx emissions, 600 Tg/yr for surface anthropogenic CO emissions, 500 TgC/yr for biogenic isoprene emissions, and 5 TgN/yr each for lightning and soil NOx emissions.”

Isoprene is the only non-anthropogenic VOC emitted here, and we have now added a statement to this effect in the paper.

Line 122: “biogenic emissions of VOC (represented here by isoprene)…”

Line 138. Deposition is likely a structural issue; for example see a community review paper suggesting most global models do not include dry deposition processes consistent with latest understanding from the field and lab (Clifton et al., 2020).

Dry deposition processes remain rather poorly understood, and this is why we have adopted a relatively large uncertainty on the three parameters considered here. We have now added a reference to the Clifton paper here which nicely summarises the uncertainties involved.

Lines 237-238. Recommended guidance here?

Changes in CH_4 loss behave more linearly than changes in CH_4 lifetime, and thus it would be better to emulate the loss rate; we have now added text to indicate this. However, this behaviour may differ if a different numerical kernel is used for emulation, so the guidance may not be entirely general.

Line 237: add “suggesting that this would be a better approach.”

Line 250. ACCMIP models all used the same anthropogenic emissions

Anthropogenic emissions make a relatively small contribution to the total uncertainty, but we have added a statement so that the reader is aware of this.

Addition to line 251: “although we note that all models used the same anthropogenic emissions in the latter studies.”

Line 264. Does the larger error in the ozone burden relative to surface reflect error accumulation and propagation given the longer ozone lifetime away from the surface?

This difference is more likely to reflect the impact of uncertainty in additional processes, particularly deposition, that more strongly affect surface concentrations (see Fig 2). This is seen in the new Fig. S2 (see response to comment below) and accompanying new text at line 284.

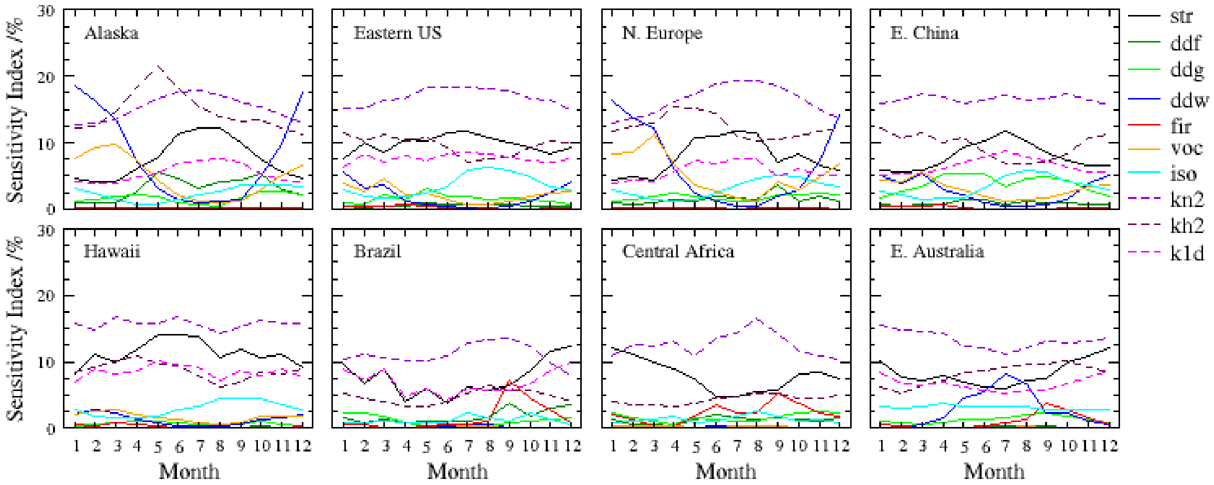
Lines 280-284. It makes sense that uncertainty in a particular process should matter most where that process dominates. Could that be illustrated?

It would be difficult to illustrate this without a separate study to identify which process dominates (i.e., process attribution rather than sensitivity analysis). We argue here that identifying the largest source of uncertainty is more useful than quantifying the current contribution of a specific process.

Figure 3. It would be interesting to include some selected free troposphere locations.

The dominant contributions to uncertainty in the free troposphere are much more similar to each other than those at the surface, and largely reflect the global contributions shown in Fig 2, with the NO2+OH reaction rate dominating. However, we appreciate that this may be of interest to the reader and have added an additional figure to the supplement showing the sensitivity indices at 5 km for the same locations shown in Fig 3 (see below). We have also included accompanying text to the discussion in Section 4.

Additional text at line 284: “In the free troposphere, the difference between locations is smaller and the seasonality is substantially less (see Fig. S2). The contribution of surface processes such as emissions and deposition is less, as expected, and photochemical processes play a larger role. At 500 hPa, the reaction rate for NO_2+OH dominates throughout the year at most locations, typically contributing 15-20% of the total uncertainty.”



“Figure S2: Sensitivity indices (%) for the dominant parameters affecting ozone in the free troposphere at 500 hPa (~5.5 km altitude) for selected locations around the world matching those shown for the surface in Fig. 3. Parameters shown include the stratospheric O_3 column (str), dry deposition to forest, grassland and water/bare ground (ddf, ddg, ddw), emissions from fires (fir) and anthropogenic and biogenic sources of VOC (voc, iso) and chemical rate constants for NO_2+OH (kn2), HO_2+O_3 (kh2) and $O(^1D)$+H_2O (k1d).”

Line 306. Do the pdfs from the independent chains end up overlapping? Do we have to worry about identifying local minima during the training?

Yes, the pdfs from the chains overlap. The chains are attracted to local minima, but the size of the perturbation applied each step is sufficient to prevent them getting stuck (outlined on line 299, but described in detail in Ryan and Wild, 2021). These aspects need to be considered when designing the constraints, but failure of the pdfs to overlap would be clear indication of interference from local mimima, and would be evident in Figs 7 and 8 as multimodal posterior distributions.

Lines 435. Are these scalings provided in the zenodo repository?

The scalings are not currently provided in the zenodo repository as they are intended as a quick fix to test the impact of addressing structural errors and are thus of little value to anyone reproducing the study with their own model. We have indicated the general magnitude of the scalings as percentages in the text of Section 7.

Lines 505-510. Is some of this implicitly accounted for by training separately for each month?

We emulate each month independently, but we did not consider different process uncertainties each month, which would require many more parameters. NOx emissions may be overestimated in winter and underestimated in summer, for example; our treatment assumes a single under- or overestimate occurs throughout the year. It would be very interesting to explore this in more targeted future studies.

Line 513 or elsewhere: consider noting that emulation should be a simpler problem for short-lived OH ?

Unfortunately, this is highly unlikely to be the case, as OH is influenced by so many different processes, directly or indirectly, and the relative importance of these processes is likely to vary strongly with chemical and physical environment.

Technical Corrections

Line 161. precursors --> reservoirs ?

Precursors was the wrong word and has now been replaced with "species" here.

Line 230. 1536 (64 locations x 12 months x 2 layers)? Either way, please briefly explain where this value is coming from.

Yes, 64 locations, 12 months, 2 layers, plus 2 global metrics (O_3 burden and CH_4 lifetime). This information has now been added as suggested.

Addition at line 230: "64 locations x 12 months x 2 altitudes + 2 global metrics"

Line 475. refined --> confirmed?

Refined now replaced with "improved".

References

Anderson, D. C., Duncan, B. N., Nicely, J. M., Liu, J., Strode, S. A., and Follette-Cook, M. B.: Technical note: Constraining the hydroxyl (OH) radical in the tropics with satellite observations of its drivers – first steps toward assessing the feasibility of a global observation strategy, Atmos. Chem. Phys., 23, 6319–6338, https://doi.org/10.5194/acp-23-6319-2023, 2023.

Bates, K. H. and Jacob, D. J.: A new model mechanism for atmospheric oxidation of isoprene: global effects on oxidants, nitrogen oxides, organic products, and secondary organic aerosol, Atmos. Chem. Phys., 19, 9613–9640, https://doi.org/10.5194/acp-19-9613-2019, 2019.

Baublitz et al., An observation-based, reduced-form model for oxidation in the remote marine troposphere, Proc. Natl. Acad. Sci. U.S.A. 120 (34) e2209735120, https://doi.org/10.1073/pnas.2209735120 (2023).

Bourgeois et al., Large contribution of biomass burning emissions to ozone throughout the global remote troposphere, Proc. Natl. Acad. Sci. U.S.A. 118 (52) e2109628118, https://doi.org/10.1073/pnas.2109628118 (2021).

Clifton, O. E. et al. (2020). Dry deposition of ozone over land: processes, measurement, and modeling. Reviews of Geophysics, 58, e2019RG000670.

Guo et al., Heterogeneity and chemical reactivity of the remote troposphere defined by aircraft measurements – corrected, Atmos. Chem. Phys., 23, 99–117, https://doi.org/10.5194/acp-23-99-2023, 2023.

Murray et al., Large uncertainties in global hydroxyl projections tied to fate of reactive nitrogen and carbon, Proc. Natl. Acad. Sci. U.S.A. 118 (43) e2115204118, https://doi.org/10.1073/pnas.2115204118 (2021).

Mirrezaei et al. Toward Realistic Prognostic Modeling Of The Methane Chemical Loss. ESS Open Archive . July 19, 2025. DOI: 10.22541/essoar.175288325.58779694/v1

Nicely, J. M., et al. (2016), An observationally constrained evaluation of the oxidative capacity in the tropical western Pacific troposphere, J. Geophys. Res. Atmos., 121, 7461–7488, doi:10.1002/ 2016JD025067.

Pimlott, et al. Investigating the global OH radical distribution using steady-state approximations and satellite data, Atmos. Chem. Phys., 22, 10467–10488, https://doi.org/10.5194/acp-22-10467-2022, 2022.