Response to reviewers on "Global sensitivity analysis of chemistry-climate model budgets of tropospheric ozone and OH: Exploring model diversity" by Wild et al.

Response to Reviewer 1:

The manuscript is very well written and the results and conclusions are clearly laid out. The findings of the importance of the effect of water vapour for tropospheric OH and ozone and that one of the three models shows a radically different sensitivity for OH are certainly of great interest to the global chemistry modelling community as the causes of the diversity across models is a long-standing problem.

We thank you for your very positive comments here, and address specific concerns below.

The only significant comment I would have is that given the somewhat arbitrary specifications for some of the ranges over which the different factors are varied, it is very difficult to put these results into context. It is shown that varying water vapour by ±50% is the dominant factor affecting ozone and OH, but how does the range of ±50% compare to the actual variability across models? While emissions, including lightning NOx, have well defined and regularly discussed ranges across models and other processes such as boundary layer mixing are very poorly constrained, a range for water vapour across models should be more easily quantified and given the importance of the process found here it should be more clearly justified. From a quick look at Figure 7 of Lamarque et al. (Geosci. Model Dev., 6, 179206, 2013) I think a variation of ±50% might be too large.

The parameter ranges we are working with are intentionally large to capture the full range of uncertainty in each variable, and this provides us with response surfaces that fully encompass the likely uncertainty. However, we agree that realistic, consistently-defined uncertainty ranges are needed when comparing model sensitivity across different variables. These are more difficult to determine for model processes than for emissions as the reviewer notes. The likely uncertainty in water vapour across models is of the order of ±20% from the Lamarque et al. 2013 study (which corresponds to temperature biases of about 2.5 K) and we implicitly acknowledged this in our presentation of water vapour in Fig. 5, which the reviewer noted.

To address this issue, which the other reviewers also raise, we have repeated the sensitivity analysis over a more restricted and directly comparable range of variability for each parameter. We maintain the existing ranges for emissions, which are well grounded in the literature, reduce the deposition ranges from ±80% to ±60%, although these remain highly uncertain, and reduce the range for water vapour to ±20%. We reduce the range for cloud optical depth to a factor of 3 based on Klein et al. (2013), and restrict the range for boundary layer mixing to a factor of 10, which still encompasses both very weak and instantaneous mixing, both of which are simple options that have been adopted in previous model studies. The new sensitivities are now presented in the paper; while these differ from those over the full ranges originally presented, with greater importance for emissions at the expense of humidity, our key messages about the importance of water vapour and about the differing responses in OH across the models remain unchanged, see Figure 1 below.
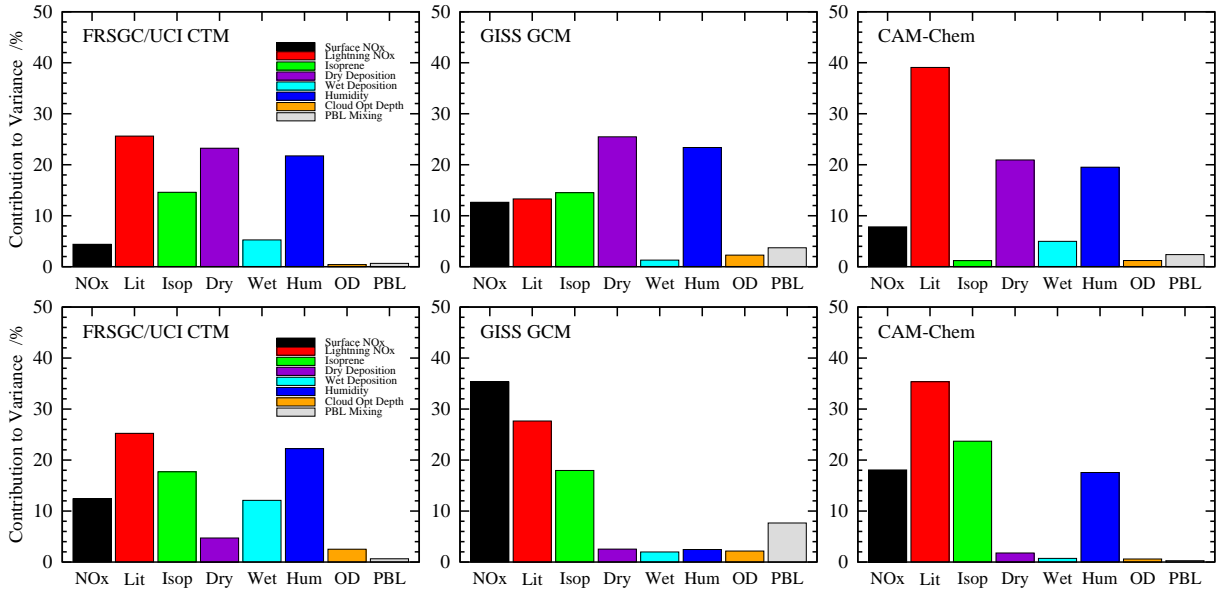
Fig 1: Contributions of each variable to the total variance in the simulated tropospheric O$_3$ burden (top row) and annual mean CH$_4$ chemical lifetime (bottom row) in each model over the revised sensitivity range.

To avoid complicating the paper with two sets of parameter ranges, one over which the models were run, and one for the sensitivity analysis, we have adjusted the text to focus on the new ranges, but include the results from the full ranges in supplementary material. We highlight that this is an exploratory study demonstrating the value of the approaches adopted here, and that the contribution of each parameter to the overall sensitivity is dependent not only on the ranges selected but also on which parameters are considered. A much wider choice of parameters is needed for a more in-depth study with greater explanatory power, and this is noted in the paper.

Lines 32: While I understand the need for brevity here, I do find the statement that changes in surface O$_3$ since the preindustrial era are systematically underestimated (Stevenson et al., 2013) to be a bit of an over-simplification of the situation. Recent work under TOAR (Tarasick et al., under review) and work with oxygen isotopes (Yeung et al, Nature, doi:10.1038/ s41586-019-1277-1, 2019) have found new reasons why the increase in tropospheric ozone may not be as large as shown by the early surface measurements. I am not suggesting an exhaustive discussion here, but just some acknowledgement that there is uncertainty.

These new papers highlight the uncertainty in preindustrial ozone, and the reviewer is correct to point this out. We have toned down this statement to reflect this uncertainty, and now refer the reader to the Tarasick et al. paper which discusses this in detail. We add "although there is continuing uncertainty surrounding preindustrial levels (Tarasick et al., 2019)."

Line 109–110: How do you approach increasing wet deposition rates for a species such as HNO$_3$?

Wet deposition rates are altered by scaling the removal rate for each soluble tracer affected by washout and rainout each model timestep. This is a crude but simple approach that can be implemented consistently across models. While it has the potential to create instability

where removal rates are greatly increased, the scalings used here are relatively small. A more in-depth exploration would involve varying cloud water content, rainfall and the solubility of different species independently. We have rephrased the text to make this clearer, stating "we assume an uncertainty in removal rates of ±60%..."

Lines 176–177: In both Figures 5 and 6 the very different response of OH in GISS to humidity is highlighted, but because humidity is only displayed as a percentage change relative to the baseline it is not possible to ascertain whether the behaviour of GISS is due to some fundamentally different response of the chemical scheme to water vapour or whether the response is due to the models being in a different part of the parameter space. Given the importance of water vapour, is it possible to provide some absolute comparison of water vapour across the three models? Ill also note that while the full range of NOx emissions are plotted for these two figures, water vapour is only plotted as ±20% while the range of the sensitivity experiments is ±50%.

The underlying humidities in the models are relatively similar, and the sensitivities for ozone (Fig 6b) are also similar, suggesting that the differences for OH arise from chemical processes rather than from humidity. However, we appreciate that different model base states may be one reason for differing sensitivities, and we have therefore included a comparison of key variables between the models in supplementary material, as suggested. This shows a high degree of similarity between the models, but points to differences in NOx and PAN that we discuss in the paper. The mean mass of water vapour in the GISS GCM runs (1.21×10^{16} kg) is about 4% less than than in the FRSGC/UCI CTM runs (1.26×10^{16} kg), but this is well within the uncertainty range we explore. We now include these numbers in the supplement, and refer to the differences in our discussion in Section 4 of the paper. As noted above, our restriction to a ±20% variation for water vapour tacitly acknowledges this more realistic assessment of the likely uncertainty in this variable, and we now use ±20% throughout the study.

Lines 198–202: In discussing Figure 7 and the different response of GISS, there is speculation that GISS may have a very different treatment of NOy chemistry because of the more widespread sensitivity to surface NOx and isoprene emission, particularly in the tropics. The quantity plotted here is the percent variance of CH_4 loss that can be assigned to each of the different factors. Since water vapour in the GISS model explains almost none of the variance in the tropics, doesnt the variance have to be assigned somewhere else. Is, in some absolute measure, methane loss in the GISS model more sensitive to NOx and isoprene emissions than the other two models or does it only assign more of the response to NOx and isoprene emissions because of the negligible contribution from water vapour?

This is a valuable point, and the reviewer is correct to note that the absolute variance is also of interest. We have recalculated the contributions using the new parameter ranges, and now present the absolute variances in Figure 7, placing the relative contributions (more formally, the sensitivity indices) in the supplementary material. It is much clearer from this analysis that the sensitivity to NOx and isoprene emissions in the GISS GCM is comparable to that in the other models, but that the sensitivity to humidity is substantially less. We have adjusted the discussion in the paper to reflect this point, and show the revised figure as Figure 2 below.
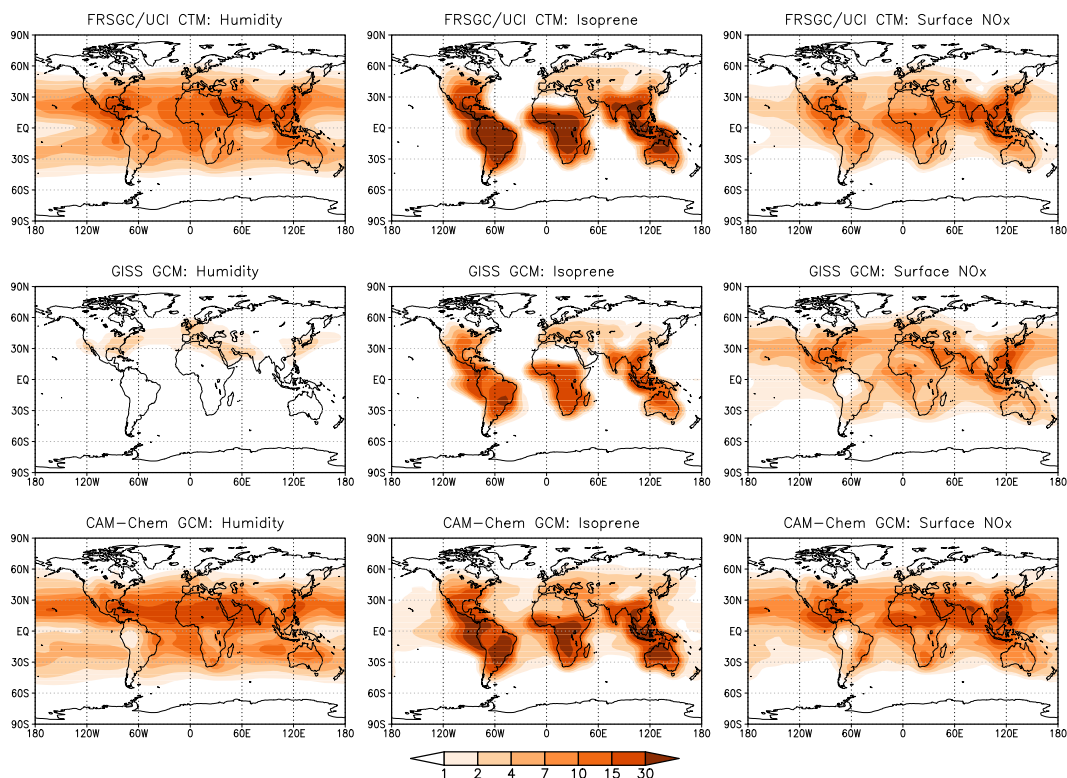
Fig 2: Contributions of each variable to the absolute variance in the simulated tropospheric O$_3$ burden (top row) and annual mean CH$_4$ chemical lifetime (bottom row) in each model over the revised sensitivity range.

Response to Reviewer 2:

1. How were the key variables selected for the Gaussian process elimination? The selection of variables should be justified. For example, Holmes et al. (2012) (No, I am not an author on this paper) found that OH is most sensitive to the ozone column (at least when examining interannual variability). This was not included in the present study. This choice needs to be justified.

This is an exploratory study and we wanted to investigate a range of different inputs and model processes, but were limited to eight variables for computational expediency and to permit application across unrelated models. The variables we selected were based on one-at-a-time sensitivity studies presented in Wild (2007), as explained in paragraph 4 of Section 2 (lines 83–95), where the impacts on ozone and OH were demonstrated to be substantial. There are many other important variables, including the ozone column, that we were not able to consider. Now that we have established the value of the approach in comparing sensitivities across models, we will explore a more complete range of variables over a larger parameter space. We highlight the necessarily restricted scope of this study in the text, and have now revised it to acknowledge the importance of other variables more clearly. In Section 2 we now state "while they do not encompass all sources of uncertainty, which also include photochemical, transport and radiation processes, they are chosen to represent key uncertainties while ensuring that the study remains computationally tractable". We also explicitly note in Section 4 that we have not explored chemical processes or ozone column

4

in this study.

2. The selected range of the sensitivity variables is very important for the paper particularly when comparing the sensitivity of one variable against another. The paper states the ranges are loosely based on studies in the literature, but do not give the studies. To me, at least it is not believable that cloud optical depth varies could vary by 100, or boundary layer mixing by a factor of 10,000. Of more concern is that humidity varies globally by 50%. This is in fact very large and colors the results and conclusions throughout the paper. In contrast, Holmes et al (2012) gives a variation of 3% in humidity. While the variability might be calculated differently in the two studies, a value of 50% seems huge. Would the range in variability be better quantified by comparing across the models? I also do not believe it is reasonable to take the variability from the smallest and largest values in the literature as this does not likely capture the likely error in model simulations. In any case, the sensitivity and ranges in variability need to be better quantified as this impacts much of the interpretation in the paper. An arbitrary specification of the range in these parameters would seem to imply that the resulting comparison of the different sensitivities is also arbitrary.

We selected large ranges to emulate to ensure that the resulting model response surfaces capture the full range of uncertainty possible for each variable. However, we agree that realistic, consistently-defined uncertainty ranges are needed when comparing model sensitivity across different variables, as the reviewer notes. In particular, the range for humidity is indeed large, and climate model intercomparisons suggest an uncertainty in water vapour of the order of $\pm 20\%$ (see Lamarque et al., 2013). To address the reviewer's concerns, we have repeated the sensitivity analysis using restricted ranges for the variables that are more closely aligned with previous assessments in the literature, and we cite the studies concerned. We maintain the existing ranges for emissions, which are well grounded in the literature, reduce the deposition ranges from $\pm 80\%$ to $\pm 60\%$, and reduce the range for water vapour to $\pm 20\%$. We reduce the range for cloud optical depth to a factor of 3 based on Klein et al. (2013), and restrict the range for boundary layer mixing unchanged to a factor of 10, which still encompasses both very weak and instantaneous mixing, both of which are simple options that have been adopted in previous model studies. The new sensitivities reflect a greater importance for emissions at the expense of humidity, as expected, but our key messages about the importance of water vapour and about the differing responses in OH across the models remain unchanged.

3. Comparing the sensitivities across models only makes sense if the model forcing is similar. (For example, the sensitivity of any one model might be very different when comparing between present day conditions and pre-industrial conditions). The authors need to show that the tropospheric forcing in the three models is roughly the same (CH_4, NOx emissions, tropical ozone column etc). In particular I am concerned about the tropical ozone column and perhaps more importantly the tropical photolysis rates. Might this explain the large difference in the methane oxidation rates between the different models? At any rate quantifying the mean differences (at least in a supplement) seems important in better understanding the results.

We standardised the model simulations as much as possible without altering their underlying configuration, constraining the magnitudes of annual surface NOx emissions, lightning NO, and isoprene emissions as outlined in the paper (lines 105–107) and using consistent fields of CH_4 (1760 ppb) under present-day climate. This mimics the conditions that are typically

applied in model intercomparisons without greatly perturbing the characteristic behaviour of the models. However, differences between tropical ozone columns, and many other variables, are indeed likely, as they are in model intercomparison studies. As suggested, we have therefore included a direct comparison of baseline model conditions in a supplement to provide the reader with greater insight into the underlying differences between the models. In general there is a broad consistency between them, although surface NOx and column NO_2 are slightly lower and PAN higher in the GISS GCM. The annual global average total column ozone is somewhat less in the GISS GCM (294 vs 321 DU in FRSGC/UCI CTM), but this is dominated by smaller columns at mid and high latitudes, and the tropical columns are very similar (263 vs 265 DU). We now refer to these underlying differences in the discussion, and refer the reader to the table and figures in the supplement for further details. We show examples of this comparison for ozone column and humidity in Figure 3 below.
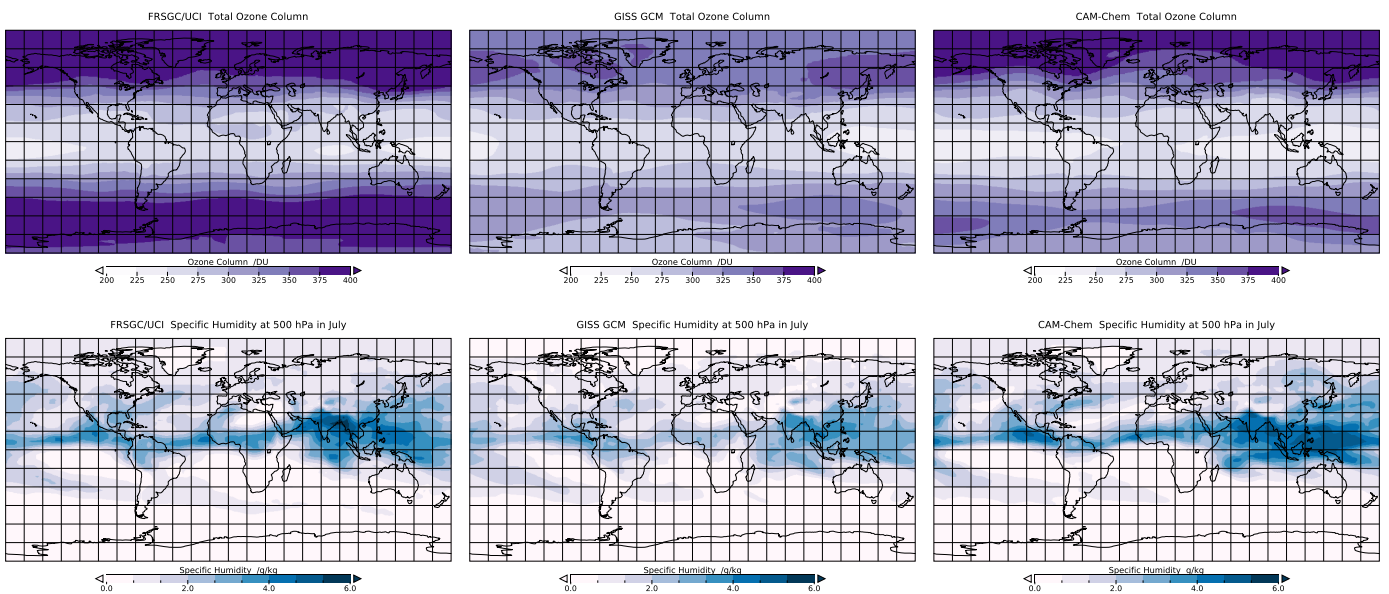
Fig 3: Annual mean total column ozone from the FRSGC/UCI CTM, GISS GCM and CAM-Chem (top row) and specific humidity at 500 hPa in July from each model (bottom row).

4. The paper does not really pinpoint some of the basic causes of the discrepancies between the models, in particular the different lifetimes of methane. As stated in the previous comment some of the mean fields in the models should be given, for example the global burden of CH_4, atmospheric water vapor (perhaps with latitude?), etc. While the sensitivity tests between the different models is revealing evaluating the mean difference between the GISS-GCM and the other models may also be revealing. In a number of locations the authors hypothesize that the difference is due to differences in humidity (lines 177-178, Pages 8 and Page9) or due to differences in the formation/decomposition of PAN. Why dont the authors check? Im not suggesting a lengthy analysis, but an inspection and comparison of some of the mean fields should be sufficient to check some of these hypothesis and possibly reveal some of the key model differences.

A deeper analysis of a much wider range of variables is needed to really pinpoint the origin of the different behaviour observed, including chemical processes in particular, as noted in the paper. However, the reviewer makes a valuable point that demonstrating the mean state of

the different models would support the analysis and is relatively straightforward. To address this, we have included surface and mid-tropospheric distributions of a range of trace gases, along with annual mean columns of ozone, NO_2 and CO in the supplementary material, and refer to these in the text. With the exception of differences in NOx and PAN, the models are reasonably consistent, suggesting that a deeper analysis of chemistry and transport processes is needed to diagnose the different responses in OH.

Minor Comments

1. Please state in the caption to Figure 1 that these results are over different years with different emissions and meteorologies. Also please reference Table 1 in the Figure caption where the references to the points can be found. Finally, please state what the dots refer to in the right hand panel.

The caption has been amended as requested to state "Note that results in the left panel represent differing emissions and meteorological years (study details are given in Table 1)". The dots in the right hand panel are simulations at intermediate points across the ranges specified in the text, but these are not important for understanding the overall responses and they have thus been removed.

2. The methodology behind the Gaussian process elimination should be explained in more detail in methodology section. While I dont expect the authors to go into detail, neither should it be necessary to reference the referenced papers to understand this analysis.

We have now added a brief introduction to Gaussian process emulation, as requested, in Section 2. The section now reads: "To explore the sensitivity of tropospheric budgets to uncertainty in several processes at once, we perform a global sensitivity analysis using Gaussian process emulation, following the approach of Lee et al. (2011). An emulator is a simple statistical model that reproduces the relationships between the inputs and outputs of a more complex model, in this case an atmospheric chemistry model. The much shorter run time of the emulator allows the model parameter uncertainty space to be explored fully through Monte Carlo approaches that would not be feasible with the complex atmospheric model. A Gaussian process is a multivariate normal distribution applied to a function, and we use this non-parametric approach to fit the model input-output relationships as it is well-tested, efficient and relatively easy to implement (O'Hagan, 2006; Lee et al., 2011; Ryan et al., 2018). This allows us to reproduce the nonlinear model response across a multidimensional parameter space based on a small ensemble of model training runs at points representing a combination of inputs that are optimally chosen to fill the space."

3. Does the sensitivity to boundary layer mixing simply involve changing the vertical diffusion coefficient over the boundary layer depth, or something else?

Yes, this involves scaling the vertical diffusion coefficient through the depth of the diagnosed boundary layer. The text describing this in Section 2 has been revised to make this clearer: "by scaling the effective vertical diffusion coefficient through the depth of the boundary layer".

4. The paper states: The models differ in their sources of meteorology, but certainly the sources of meteorology are important. Are the runs using specified dynamics or GCM gener-

ated meteorology? This might be quite important to how well the relative humidity is specified. In addition, using a single year might also introduce significant variability between the models. Please say something more specific about the meteorology used as well as comment on the possible importance of the interannual variability. My guess is the latter might introduce significant variability between the models.

Differences in meteorology are likely to be important, but it is impractical to prescribe a particular source of meteorology given the range of models used. We aimed for consistency with 2001 conditions, but substantial differences are likely to remain. Offline meteorological fields from the ECMWF IFS for 2001 were used for the FRSGC/UCI CTM, the GISS GCM used observed sea-surface temperatures and was nudged to NCEP reanalysis fields, and CAM-Chem was run in GCM model. This information was omitted from the description in Section 2, and has now been added: "Offline meteorological fields for 2001 from the European Centre for Medium-Range Weather Forecasts Integrated Forecast System (ECMWF IFS) were used for the FRSGC/UCI CTM. The GISS GCM used observed sea-surface temperatures and was nudged to National Centers for Environmental Prediction (NCEP) reanalysis fields (Kalnay et al., 1996), while CAM-Chem was run in GCM mode following the Chemistry Climate Model Initiative (CCMI) REF-C1 protocol (Eyring et al., 2013)."

Interannual meteorological variability is likely to contribute some variability in ozone and OH (Wild 2007 found variations of ozone burden of about 2% and of methane lifetime of about 5% over three years), but this is likely to be less that that between entirely different sources of meteorology, where different underlying biases may be present. We have revised the text in Section 4 to acknowledge the importance of the meteorological fields, and to note that these are one component of the structural uncertainty associated with the use of different models. In particular, we now note: "Underlying differences in meteorological fields governing vertical transport processes such as convection are also likely to be important in this region."

5. Page 5, line 15. but we scale. It is not altogether clear how this scaling is used. Please specify. Given the non-linear response to NOx, does a linear scaling make sense?

We apply a simple scaling to the emissions to ensure that the total magnitude is the same in each model. There are differences in the location and timing of emissions, particularly for lightning NO and biogenic isoprene where emission are determined interactively based on model meteorology and vegetation schemes. It is impractical to harmonize these between models, and the differences are effectively structural, as noted in the text. The differences are relatively small, as is evident from the surface distributions of NOx and isoprene that we have now included in the supplement.

6. Page 7, line 139-141 discrepancies highlight that uncertainty in chemistry and transport processes not considered here may play a substantial role in governing the CH_4 lifetime. Im not sure I understand this conclusion. It seems it might simply mean, for example, that the variability in water vapor is considerably over-estimated.

Reducing the uncertainty in water vapour narrows the distribution, but does not bring the mean lifetime closer to the observed range. The discrepancy between the model and observed values must therefore be due to uncertainty in processes and variables that have not been considered in this study. Chemistry and transport processes are likely to be important here,

along with other key variables such as the total column ozone. The text has been revised to make this point clearer.

7. Page 7, line 149 it is notable that humidity has not been prescribed in previous model intercomparison studies. Possibly, but again the large response in water vapor is dependent on the large variation in water vapor.

The contribution of water vapour is smaller now that the uncertainty range has been reduced to 20%, but it still makes a sizable contribution to the overall uncertainty, as noted in the text here.

8. Page 8. The explanation in the future responses of the different models is nice and given in several places in the paper. Personally, I think this might belong better in the conclusion section.

We now refer to this in the conclusions, although we do not include the quantitative aspect here to avoid any further repetition. We now explicitly state "... can provide substantial new insight into the differing responses of models under different emission and climate scenarios."

9. Page 9, line 180 suggesting a saturation in OH formation in this model. How does a saturation occur?

It is unclear why OH is relatively insensitive to water vapour in this model, and it would certainly be useful to explore this further with more targetted studies. Any additional formation from higher water vapour must be balanced by additional removal, but the pathway for this remains unclear. The statement here was intended to indicate that net formation had reached a plateau, but this has now been rephrased to make it clearer: "suggesting either little additional formation of OH or a matching OH sink in this model".

10. More generally, I am curious about the conceptual difference between the sensitivity to forcing parameters (NOX emissions, stratospheric ozone column) and internal model parameters. To what extent does it make sense to distinguish between the sensitivity between these type of parameters?

The approaches used here can be applied to uncertainty in model inputs and internal parameters, and there is no conceptual difference. We make the distinction in the paper to highlight that they can be applied to both, even though previous applications in other contexts have tended to focus on one or the other.

Response to Reviewer 3:

This study is an important contribution towards understanding the reasons for model diversity in the evolution of tropospheric ozone and methane lifetime. However, I found that the approach applied here is inadequate to truly understand the reasons for diversity in these non-linear quantities. My main concern (similar to the other two reviewers) is that the sensitivities calculated for each model would depend on the control simulation given the non-linear chemistry of ozone and methane. If the models differ in the forcings (meteorology and emissions) to begin with then how do we know that the calculated sensitivity is not due to

the initial state. Also, I found the description of the Gaussian emulation approach applied here rather limited to appreciate its usefulness for understanding the reasons for diversity in model response.

We thank the reviewer for these comments, which parallel those of the other reviewers, and we have addressed the principal concerns here in our comments above. We standardised the models in terms of the magnitude of key ozone precursor emissions and in methane burden, and the baseline results of the models (which we have now included in a supplement) are broadly similar. We have revised the text to include a brief introduction to Gaussian process emulation so that readers unfamiliar with it can appreciate the approach, particularly for representing non-linear responses, which present no problem.

L23–25: Clarify if this is referring to controls for climate or air pollution. Controls on NOx emissions in the US (e.g., Clean Air Act) and Europe (e.g., LRTAP) have indeed brought down surface ozone.

As pollutants and greenhouse gases these short-lived gases are attractive targets for both types of control. However, selecting appropriate controls is challenging; while emission controls have reduced surface ozone in the US, similar controls have led to increases in urban ozone in China. The Science paper cited here targets climate change, air quality and food security, and all three areas are of relevance here.

L27: Also another ACCMIP paper (Naik et al., 2013) and CCMI models (Zhao et al., 2019 https://www.atmos-chem-phys-discuss.net/acp-2019-281/)

The Naik et al. paper is very useful, but covers the same ACCMIP models described in the study of Voulgarakis et al. that we cite. The Zhao et al., study includes valuable new data from CCMI studies, but methane lifetimes are only available from a limited subset of the models, most of which also contributed to the ACCMIP studies.

L32: There are large uncertainties in PI estimates of surface ozone as discussed by Tarasick et al. (2019) https://www.elementascience.org/articles/10.1525/elementa.376/

As noted above in response to reviewer 1, we have rephrased this statement to acknowledge the uncertainties involved and have cited the Tarasick paper which highlights the magnitude of these uncertainties very nicely. We add "although there is continuing uncertainty surrounding preindustrial levels (Tarasick et al., 2019)."

L58: Observational estimates of global ozone are now available from satellites (Gaudel et al. 2019 https://www.elementascience.org/articles/10.1525/elementa.291/). How do the model estimates discussed here compare with satellite estimates?

Satellite data provides a useful additional constraint on the global ozone burden, but most assessments are limited to the 60^oS–60^oN latitude range. Only two IASI retrievals are available for the full globe (333 and 345 Tg), although these exclude regions in polar night. The TOST ozonesonde product provides a global burden of 337 Tg, but this is very similar to the ozonesonde-based estimates we already cite. We add an additional sentence acknowledging these additional estimates: "Recent satellite and ozonesonde-based estimates of the global burden range from 333–345 Tg (Gaudel et al., 2019)."

L100–104: It would help to know how different the base state is in the models. What is the ozone burden, prescribed methane concentration, methane lifetime, surface and lightning NOx emissions, biogenic emissions, wet and dry deposition rates for all species, atmospheric humidity, cloud optical depth, and boundary layer height in the base simulation for all the three models?

We have now added a direct comparison of the base state of the models in the new supplementary material, and provide a very brief summary in the text. As noted in the paper, the magnitude of surface and lightning NO emissions and biogenic emissions are already constrained, as is the methane concentration, but we have included surface and mid-tropospheric distributions of a range of trace gases, along with annual mean columns of ozone, NO_2 and CO in the supplementary material to provide a simple characterization of the models.

L114–115: How are the emulators built for a non-linear system such the O_3-NOx-CH_4 chemistry? Some description is needed to make the design of emulators transparent for the purpose of this figure.

The ability to fully reproduce non-linear behaviour is one of the principal advantages of Gaussian process emulation, and we now note this in the text. We have added a few sentences to provide more details on the method used to generate the emulators: "An emulator is a simple statistical model that reproduces the relationships between the inputs and outputs of a more complex model, in this case an atmospheric chemistry model. The much shorter run time of the emulator allows the model parameter uncertainty space to be explored fully through Monte Carlo approaches that would not be feasible with the complex atmospheric model. A Gaussian process is a multivariate normal distribution applied to a function, and we use this non-parametric approach to fit the model input-output relationships as it is well-tested, efficient and relatively easy to implement (O'Hagan, 2006; Lee et al., 2011; Ryan et al., 2018). This allows us to reproduce the nonlinear model response across a multidimensional parameter space based on a small ensemble of model training runs at points representing a combination of inputs that are optimally chosen to fill the space."

L142–143: It would be helpful to provide an equation to explain how sensitivity for each variable is determined. As it stands, the process appears too opaque to me.

Variance decomposition is performed through calculation of the sensitivity indices using the Sobol approach as described by Saltelli 2002. The mathematical foundation for this is fully described in Ryan et al., 2018, which was written to fully document and evaluate the approaches and their application to the models used here, and the code that was used is provided at https://doi.org/10.5281/zenodo.1038667 We have added additional text here to provide a clearer description of the approach taken, and refer the reader to this earlier paper for the numerical theory, notation and equations which underpin it.

Saltelli, A.: Making best use of model evaluations to compute sensitivity indices, Comput. Phys. Commun., 145, 280297, 2002.

L151: How different is humidity across the three models for the base run? Is it possible that the three models show large sensitivity of ozone to humidity because such a large (±50%) perturbation is used? How do the sensitivities for ozone calculated here compare with those

calculated by Revell et al. (2018) https://www.atmos-chem-phys.net/18/16155/2018/acp-18-16155-2018.pdf?

We have added a simple comparison of summertime humidity distributions in the supplement. The patterns of humidity look very similar in the models, although the magnitudes are slightly less in the GISS GCM. The mean mass of water vapour in the GISS GCM runs (1.21×10^{16} kg) is about 4% less than than in the FRSGC/UCI CTM runs (1.26×10^{16} kg) on an annual basis, and we now refer to this difference in our discussion in Section 4 and include the numbers in a table in the supplement Revell et al. 2018 did not explicitly consider sensitivity of ozone burden to humidity, but this is a valuable paper and we cite it as another example of use of emulation to explore model sensitivity.

L167–169: and four models showing decreased lifetime - is the implication here that these four models may have greater sensitivity to humidity and therefore show declining methane lifetime? If so, how do we know that these models are like CAM-chem and FRSGC/UCI CTM in their sensitivities?

Greater future humidity would suggest greater OH formation and hence reduced methane lifetime, although this is balanced against reduced secondary OH formation if NOx emissions are reduced. The point we make here is that the relative sensitivity to water vapour and NOx emissions is important, and that the different model behaviours are very likely associated with differing sensitivities. We cannot explain the behaviour of these other ACCMIP models without further analysis, but suggest that exploring sensitivities with them would be very revealing.

L192–194: The chemical loss of methane also depends on the concentration of methane in the models. Are they the same across the models?

We constrain methane to a fixed concentration of 1760 ppb suitable for 2001 conditions, so that methane is the same in all models. We neglected to highlight this in the text, and have now amended the discussion in Section 2 to include this: "In each model we constrain methane to a fixed mixing ratio of 1760 ppb suitable for 2001 conditions."

L198: How different are the model chemical mechanisms implemented in the models? I would imagine the differences in sensitivities due to NOx are due to the implemented chemical mechanisms.

The chemical mechanisms in the models differ, and this is one set of processes that we were not able to explore (as noted on lines 203–204), as it requires consideration of a much larger number of variables than we were able accommodate (currently just eight). Previous studies have explored the contribution of chemical processes in a single model (e.g., Newsome and Evans, 2017, as cited in the paper), and further exploration of this across different models using the approaches described here would be a valuable extension of this study, as we note in the Conclusions.

Global sensitivity analysis of chemistry-climate model budgets of tropospheric ozone and OH: Exploring model diversity

Oliver Wild[1], Apostolos Voulgarakis[2], Fiona O'Connor[3], Jean-François Lamarque[4], Edmund M. Ryan[1,5], and Lindsay Lee[6,7]

[1]Lancaster Environment Centre, Lancaster University, Lancaster, UK
[2]Department of Physics, Imperial College, London, UK
[3]Met Office Hadley Centre, Exeter, UK
[4]National Center for Atmospheric Research, Boulder, CO, USA
[5]Now at: School of Mathematics, University of Manchester, Manchester, UK
[6]School of Earth and Environment, University of Leeds, Leeds, UK
[7]Now at: Department of Engineering and Mathematics, Sheffield Hallam University, Sheffield, UK

Correspondence: Oliver Wild (o.wild@lancaster.ac.uk)

Abstract. Projections of future atmospheric composition change and its impacts on air quality and climate depend heavily on chemistry-climate models that allow us to investigate the effects of changing emissions and meteorology. These models are imperfect as they rely on our understanding of the chemical, physical and dynamical processes governing atmospheric composition, on the approximations needed to represent these numerically, and on the limitations of the observations required to constrain them. Model intercomparison studies show substantial diversity in results that reflect underlying uncertainties, but little progress has been made in explaining the causes of this or in identifying the weaknesses in process understanding or representation that could lead to improved models and to better scientific understanding. Global sensitivity analysis provides a valuable method of identifying and quantifying the main causes of diversity in current models. For the first time, we apply Gaussian process emulation with three independent global chemistry transport models to quantify the sensitivity of ozone and hydroxyl radicals (OH) to important climate-relevant variables, poorly-characterized processes and uncertain emissions. We show a clear sensitivity of tropospheric ozone to atmospheric humidity and precursor emissions which is similar for the models, but find large differences between models for methane lifetime, highlighting substantial differences in the sensitivity of OH to primary and secondary production. This approach allows us to identify key areas where model improvements are required while providing valuable new insight into the processes driving tropospheric composition change.

15 *Copyright statement.* TEXT

1 Introduction

Atmospheric photochemistry and transport processes play important roles in the Earth system by controlling the impact of natural and anthropogenic trace gas emissions on air quality and global climate. Methane (CH_4) and ozone (O_3) are the

second and third most important greenhouse gases contributing to climate change since the preindustrial era (IPCC, 2013).

20 The atmospheric abundance of both gases has increased substantially due to anthropogenic activity, and their fates are strongly coupled through the short-lived hydroxyl (OH) radical. CH_4 is an O_3 precursor and O_3 is a major source of OH, which controls the oxidation of CH_4 and many other trace gases. At the surface O_3 contributes to poor air quality and is damaging to human health, crop yields and natural ecosystems (Monks et al., 2015). The relatively short lifetime of these gases makes them attractive targets for emission controls (Shindell et al., 2012), but scientific uncertainties associated with the processes

25 that govern their abundance and distribution has hindered implementation of effective control policies.

Current global chemistry-climate models representing the co-evolution of atmospheric O_3 and CH_4 show differences in CH_4 lifetime of almost a factor of two (Wild, 2007; Voulgarakis et al., 2013). This prevents them from simulating the observed atmospheric build-up of CH_4 correctly or attributing its causes reliably, and leads to substantial uncertainty in the impact of future emission changes on global climate (Stevenson et al., 2013; IPCC, 2013). The underlying cause is differences in

30 OH, which depends on humidity, sunlight, O_3, and on a wide range of chemical and dynamical processes. For O_3, on the other hand, the abundance, seasonality and spatial variation are represented relatively well in models under present-day conditions, but observed changes in surface O_3 since the preindustrial era are ~~systematically underestimated (Stevenson et al., 2013)~~ thought to be underestimated (Stevenson et al., 2013), although there is continuing uncertainty surrounding preindustrial levels (Tarasick et al., 2019). Models have difficulty reproducing recent observed trends in surface O_3 driven by changes in precursor

35 emissions, natural sources, stratospheric influx and transport patterns (Parrish et al., 2014). This is a major concern because changes in the tropospheric abundance of O_3 influence our assessment of radiative forcing and also attainment of air quality objectives on regional and urban scales (e.g., Akimoto, 2003). These discrepancies suggest that there are major weaknesses in our fundamental understanding of the chemical, dynamical, and emission processes controlling the distribution, interaction and fate of O_3, CH_4 and OH, or in how these processes are represented in global chemistry and climate models.

40 Global sensitivity analysis provides a valuable approach to determine the major drivers of model behaviour, and has been applied to atmospheric chemistry schemes to explore uncertainties in tropospheric O_3 (Derwent and Murrells, 2013; Christian et al., 2017; Ridley et al., 2017; Newsome and Evans, 2017). These studies have typically used Monte Carlo-based ensemble approaches for simple models (e.g., Ridley et al., 2017) or structured random-sampling approaches for more computationally intensive models (e.g., Christian et al., 2017), and have focussed on sensitivities in a single model framework. In this study

45 we demonstrate the use of Gaussian process emulation for global sensitivity analysis, applied previously to models of aerosol processes (Lee et al., 2011, 2013) and air quality (Beddows et al., 2017; Aleksankina et al., 2019), and apply it to explore the sensitivity of global tropospheric O_3 and CH_4 lifetime to uncertainty in key model processes and inputs. We investigate how the sensitivities differ across three independent chemistry-transport models, and demonstrate how this approach may be used to explore the diversity in model responses and to identify where model results differ.
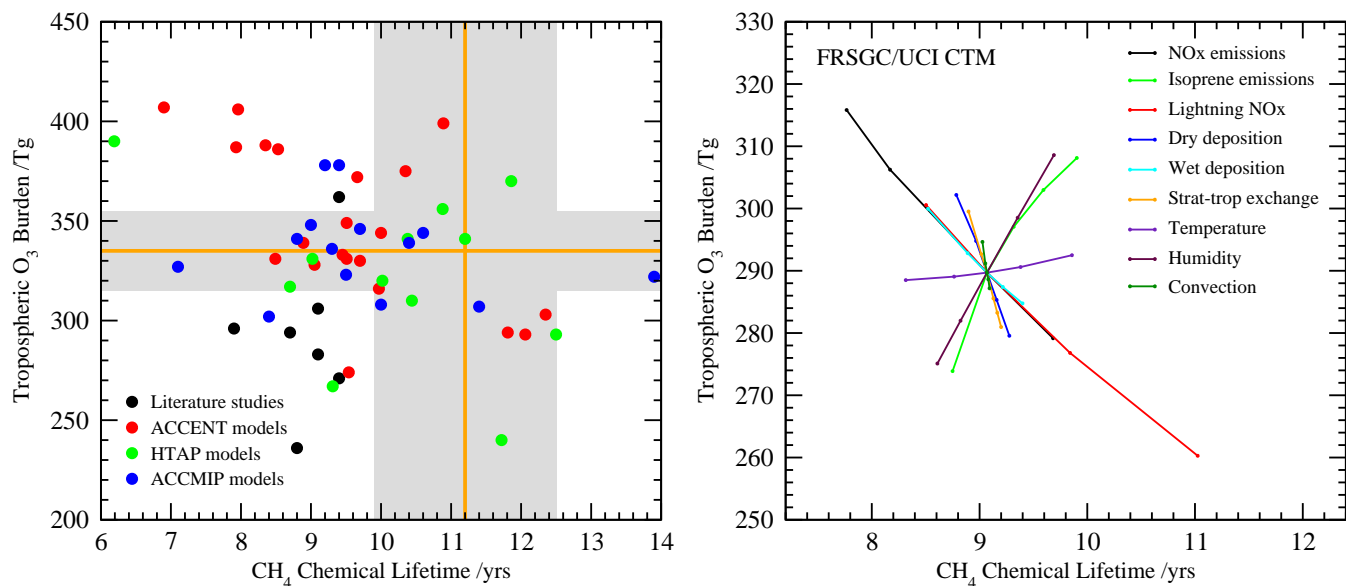
Figure 1. Tropospheric oxidant budgets from previous published studies and model intercomparisons (left panel, **a**), along with measurement-based estimates of the tropospheric O_3 burden and CH_4 lifetime (shaded regions). The right panel (**b**) shows results from one-at-a-time sensitivity studies with a single model revealing the extent to which individual processes can influence the budgets (see Wild (2007) for details). Note that results in the left panel represent differing emissions and meteorological years (study details are given in Table 1), and that the right panel covers only part of the parameter space shown in the left panel.

2 Approach

50

We consider here two important global diagnostics of model performance, the tropospheric O_3 burden and the chemical lifetime of CH_4 in the troposphere. The tropospheric O_3 burden is the annual mean mass of O_3 below the tropopause, defined here by the 150 ppb isopleth of monthly mean O_3. The chemical lifetime of CH_4 reflects the lifetime of CH_4 to removal by OH in the troposphere, and provides a useful proxy for global tropospheric oxidizing capacity. Global model studies in the literature

55 and previous model intercomparisons show a large diversity in modelled budgets (see Fig. 1), where the range in O_3 burden and CH_4 lifetime both span about a factor of two. There is no clear relationship between the budget terms on an annual basis, highlighting the relatively complex relationship between tropospheric O_3 and OH that reflects physical and dynamical processes as well as photochemistry.

Observation-based determination of these global quantities is difficult. However, assessment of three global O_3 climatologies

60 derived from ozonesonde measurements over the 1980s and 1990s indicates an annual mean tropospheric O_3 burden of 327–344 Tg when applying the same 150 ppb isopleth definition of the tropopause used in model analysis (Wild, 2007), suggesting a burden of about 335±20 Tg. Recent satellite and ozonesonde-based estimates of the global burden range from 333–345 Tg (Gaudel et al., 2018). Ensemble mean O_3 burdens from recent model intercomparisons lie close to this: 344±39 Tg from

Table 1. Global tropospheric metrics from previous model studies

Studies	Number	O$_3$ burden	CH$_4$ lifetime	References
Early literature studies	33 studies	307$\pm$38 Tg		Wild (2007)
ACCENT intercomparison	21 models	344$\pm$39 Tg	9.6$\pm$1.4 yr	Stevenson et al. (2006)
HTAP intercomparison	12 models	328$\pm$41 Tg	10.2$\pm$1.7 yr	Fiore et al. (2009)
ACCMIP intercomparison	14 models	337$\pm$23 Tg	9.8$\pm$1.6 yr	Young et al. (2013); Voulgarakis et al. (2013)
Observational estimates		335$\pm$20 Tg	11.2$\pm$1.3 yr	Wild (2007); Prather et al. (2012)

ACCENT (Stevenson et al., 2006), 328$\pm$41 Tg from HTAP (Fiore et al., 2009) and 337$\pm$23 Tg from ACCMIP (Young et al., 2013), see Table 1, but about half of published studies lie outside the observationally-constrained range (see Fig. 1). A thorough observation-based sensitivity analysis of the factors contributing to CH$_4$ removal gave a whole-atmosphere lifetime of 9.1$\pm$0.9 yr, and a corresponding CH$_4$ chemical lifetime of 11.2$\pm$1.3 yr (Prather et al., 2012). The latter is substantially longer than that derived from model intercomparisons: 9.6$\pm$1.4 yr from ACCENT (Stevenson et al., 2006), 10.2$\pm$1.7 yr from HTAP (Fiore et al., 2009) and 9.8$\pm$1.6 yr from ACCMIP (Voulgarakis et al., 2013), and two thirds of the model studies shown in Fig. 1 lie outside this range. However, it is difficult to judge the validity of existing model results without a clearer idea of the uncertainties involved and how they contribute to the corresponding biases.

The sensitivity of the budget terms to individual processes has been explored in previous studies using the Frontier Research System for Global Change version of the University of California Irvine Chemical Transport Model (FRSGC/UCI CTM) in Wild (2007). One-at-a-time sensitivity runs were performed varying surface NO$_x$ emissions (30–60 TgN yr^{-1}), isoprene emissions (0–650 TgC yr^{-1}), lightning NO$_x$ emissions (0–7.5 TgN yr^{-1}), convective lifting, stratospheric influx and deposition processes (all $\pm$50%), temperature ($\pm$5^o C) and humidity ($\pm$20%), and results are summarised in Fig. 1. This study highlighted the responses of a single model to particular processes, but the variations spanned relatively little of the parameter space defined by previous model studies, suggesting that substantial additional uncertainties were not accounted for here, including process interactions, neglected processes, and structural differences between models.

To explore the sensitivity of tropospheric budgets to uncertainty in several processes at once, we perform a global sensitivity analysis using Gaussian process emulation, following the approach of Lee et al. (2011). An emulator is a simple statistical model that reproduces the relationships between the inputs and outputs of a more complex model, in this case an atmospheric chemistry model. The much shorter run time of the emulator allows the model parameter uncertainty space to be explored fully through Monte Carlo approaches that would not be feasible with the complex atmospheric model. A Gaussian process is a multivariate normal distribution applied to a function, and we use this non-parametric approach to fit the model input-output relationships as it is well-tested, efficient and relatively easy to implement (O'Hagan, 2006; Lee et al., 2011; Ryan et al., 2018). This allows us to reproduce the nonlinear model response across a multidimensional parameter space based on a small ensemble of model training runs at points representing a combination of inputs that are optimally chosen to fill the space. We select eight key variables that influence global oxidant budgets substantially, and that span a range of model inputs (e.g., emissions),

processes (e.g., deposition) and meteorological variables, see Table 2. These are ~~loosely based on the~~ based on our earlier one-at-a-time studies, and while they do not encompass all sources of uncertainty, which also include photochemical, transport and radiation processes, they are chosen to represent key uncertainties while ensuring that the study remains computationally tractable. We select surface emissions of NO_x from natural and anthropogenic sources, the dominant precursor for O_3 in the troposphere; lightning emissions of NO, which are highly uncertain and have a disproportionately large impact on O_3 and OH due to the altitude of the source; and biogenic emissions of isoprene, which dominate global sources of volatile organic compounds. We include dry deposition, which is important for uptake of O_3 and other species at the surface, and wet deposition which is important for removal of soluble precursors. We vary the atmospheric humidity used by the model photochemistry, which plays an important role in O_3 chemistry and OH formation, but leave it untouched for other processes to avoid perturbing model dynamical processes. We vary cloud optical depth, an uncertain variable which has a major influence on photolysis rates in the lower troposphere. Finally, we vary turbulent mixing in the planetary boundary layer (PBL), which has an important role in lifting and dispersing surface oxidants, but which remains poorly constrained.

For each variable, we define a range that encompasses the maximum and minimum likely values that is loosely based on published studies from the literature, and these are presented in Table 2. We assume uncertainty ranges of $\pm 25\%$ for surface NO_x, representing 30–$50\,\mathrm{TgN\,yr^{-1}}$, $\pm 60\%$ for lightning NO (Schumann and Huntrieser, 2007) and $\pm 60\%$ for isoprene emissions (Ashworth et al., 2010). For dry and wet deposition, we assume an uncertainty in removal rates of $\pm 60\%$ that is applied to all species considered. We assume an uncertainty of $\pm 20\%$ for atmospheric water vapour, reflecting the variation across models contributing to the ACCMIP intercomparison (Lamarque et al., 2013), and this is applied in the model photochemistry scheme only. We assume an uncertainty of a factor of three in cloud optical depth based on Klein et al. (2013) and apply this for photolysis calculations only. Boundary layer mixing is perturbed by scaling the effective vertical diffusion coefficient through the depth of the boundary layer so that turbulent mixing of tracers between model layers varies from negligible to almost complete every model time step.

Following Lee et al. (2011), we use maximin Latin hypercube sampling to optimally select 80 points from across the eight-dimensional parameter space. Each point represents a combination of values chosen from the range for each variable, and specifies the values to use for a full model simulation. An additional 24 points are selected to provide an independent test of the validity of the emulators that are built. This defines a set of 104 model simulations to perform. For this study, we use three independent global chemistry-transport models: the FRSGC/UCI CTM (Wild, 2007), the Goddard Institute for Space Studies Global Climate Model, GISS GCM (Shindell et al., 2013), and the Community Atmosphere Model with Chemistry, CAM-Chem (Lamarque et al., 2012). The models differ in their sources of meteorology, but are run for a full year (following 6–12 months spin-up) under conditions that are broadly consistent with 2001 meteorology, a year without strong climate phenomena such as El Niño. Offline meteorological fields for 2001 from the European Centre for Medium-Range Weather Forecasts Integrated Forecast System (ECMWF IFS) were used for the FRSGC/UCI CTM. The GISS GCM used observed sea-surface temperatures and was nudged to National Centers for Environmental Prediction (NCEP) reanalysis fields (Kalnay et al., 1996), while CAM-Chem was run in GCM mode following the Chemistry Climate Model Initiative (CCMI) REF-C1 protocol (Eyring et al., 2013). In each model we constrain methane to a fixed mixing ratio of 1760 ppb suitable for 2001 conditions.

**Table 2.** Variables and uncertainty ranges used in this study

Variables	Range
Surface NOx emissions	30–50 TgN yr^{-1}
Lightning NO emissions	2–8 TgN yr^{-1}
Biogenic isoprene emissions	200–800 TgC yr^{-1}
Dry deposition rates	± ~~80~~60%
Wet deposition rates	± ~~80~~60%
Atmospheric humidity	± ~~50~~20%
Cloud optical depth	× ~~0.1–10~~ 0.33–3.0
Boundary layer mixing	× ~~0.01–100~~ 0.10–10.0

125 Natural and anthropogenic emissions differ somewhat across the models, reflecting different assumptions and online generation of natural emissions, but we scale the magnitude of global annual emissions to 40 TgN yr^{-1} for surface NO$_x$, 5 TgN yr^{-1} for lightning NO and 500 TgC yr^{-1} for isoprene in the control run, accepting that differences in emission distributions ~~in the models contribute to~~ represent a source of structural uncertainty. ~~We assume uncertainty ranges of ±25% for surface and ±60% for lightning NO (Schumann and Huntrieser, 2007) and isoprene emissions (Ashworth et al., 2010).~~ Other variables are

130 scaled according to the factors shown in Table 2 without further standardization between models. ~~Dry and wet deposition rates are scaled for all species considered; atmospheric water vapour is scaled for the model photochemistry scheme only and cloud optical depth based on prescribed or calculated cloud fields is scaled for the model photolysis scheme only. Boundary layer mixing is perturbed through scaling the effective vertical diffusion coefficient so that turbulent mixing of tracers between layers varies from negligible to almost complete every model time step.~~

135 Emulators are then built for each model for each output of interest using the methods described in Lee et al. (2011) and Ryan et al. (2018). We focus here on global annual mean tropospheric O$_3$ burden and CH$_4$ chemical lifetime for simplicity. The emulators are tested through use of the additional 24 validation simulations to evaluate their performance. For the outputs considered here, the model response surfaces are relatively smooth, reflecting the stable behaviour of the global O$_3$ burden and CH$_4$ lifetime, and the emulators fit the validation runs very closely with a correlation coefficient $r > 0.99$ (see Ryan et al.,

140 2018). The emulators reproduce the response of the full model within the variable ranges defined, and can be used in place of the model for intensive analysis such as uncertainty propagation through the use of Monte-Carlo approaches that would not be computationally feasible with the full model. This allows us to define formal error bars for the response of each model, and to carry out global sensitivity analysis by determining the contribution of each variable to the overall variance in modelled O$_3$ burden and CH$_4$ lifetime.
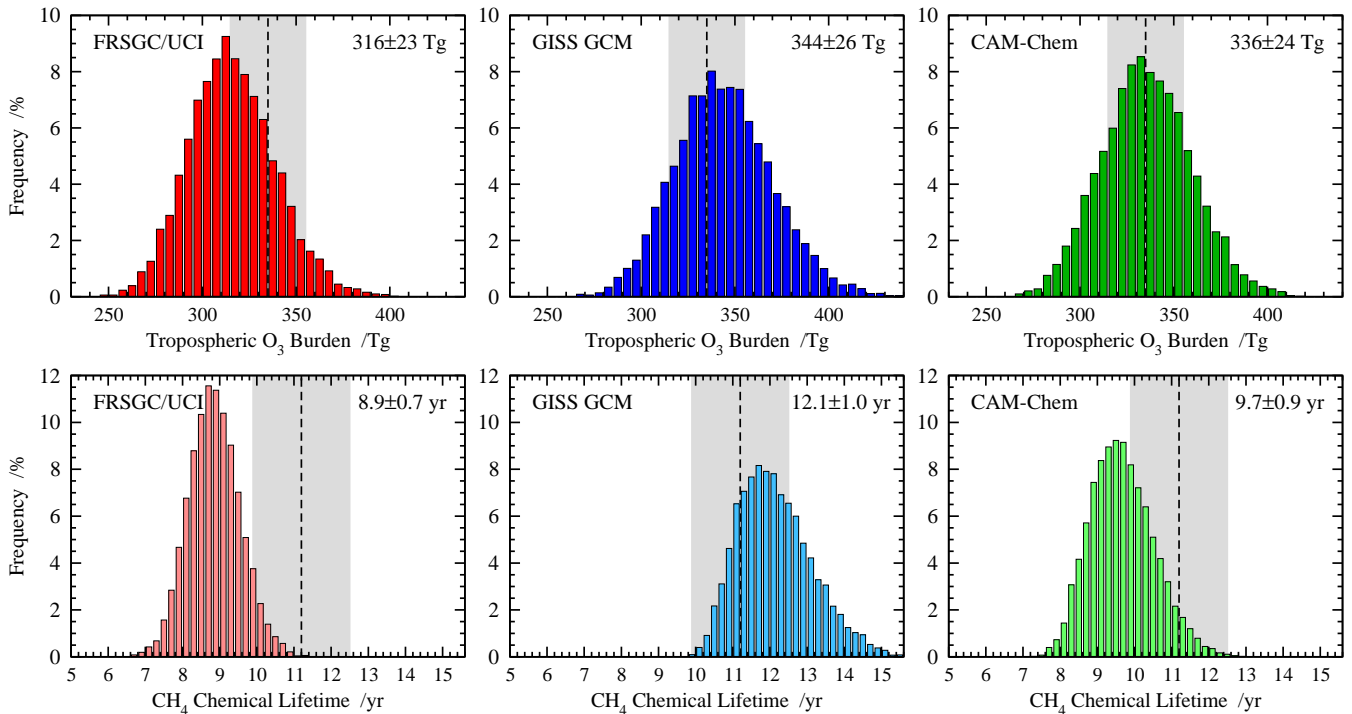
Figure 2. Probability distributions for the global annual mean tropospheric O_3 burden (top row) and tropospheric chemical lifetime of CH_4 (bottom row) for each model. The mean and standard deviation over 10,000 realizations are shown on the upper right of each panel, and observation-based estimates of O_3 burden and CH_4 lifetime are shown shaded.

3   Model responses and contributions to variance

We first use the emulators built for each model to propagate the uncertainty in the selected variables to uncertainty in O_3 burden and CH_4 lifetime. We use a Monte Carlo approach to randomly select 10,000 points from across the response space for each model, sampling uniformly across the full input range of each variable, and use this to generate the probability distribution for each model. Figure 2 shows the distribution in global O_3 burden and CH_4 lifetime from each model. The behaviour of the models is similar, with a normalised standard deviation of ~~12–13~~7–8% for O_3 burden and ~~9–13~~7–9% for CH_4 lifetime, and the distributions are slightly skewed, reflecting the nonlinear response of these budget terms to the governing processes. The 1σ uncertainty in each budget term is comparable in magnitude to that seen between different models in ~~recent model intercomparisons~~ the ACCMIP model intercomparison (see Table 1); while this may be fortuitous, it demonstrates that process uncertainty contributes substantially to model diversity.

For each model, the mean O_3 burden lies within the observational uncertainty range, along with ~~32–38~~44–60% of the distribution. A ~~large~~ substantial proportion of each distribution lies outside the observational range, suggesting that the uncertainty ranges adopted for some of the variables were larger than needed, or that a normal distribution of uncertainty could have been assumed across each range in place of a uniform distribution. For mean CH_4 lifetime, agreement with observations is less

7
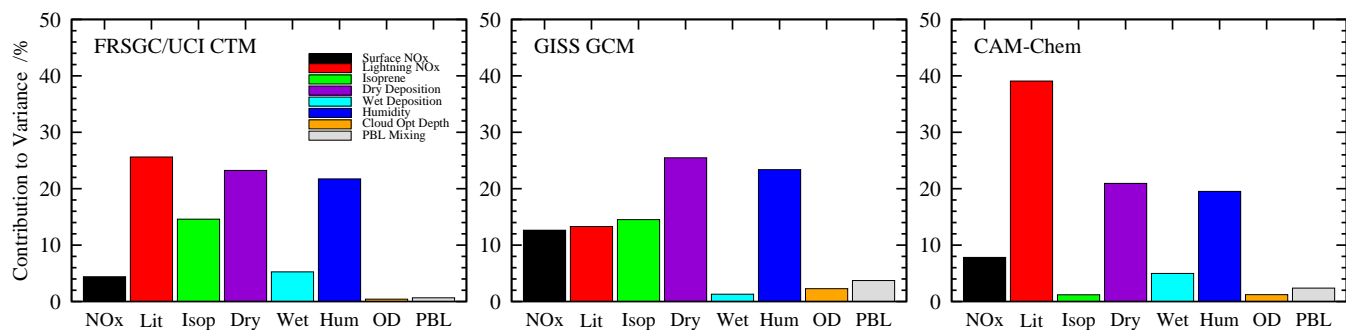
Figure 3. Contributions of each variable to the total variance in the simulated tropospheric O$_3$ burden in each model.

good, with the GISS GCM and CAM-Chem lying at opposite boundaries of the observed range and the FRSGC/UCI CTM lying outside it. For the GISS GCM, ~~53~~63% of the distribution lies inside the observed range, while for the FRSGC/UCI CTM it is only ~~24%. These discrepancies highlight~~ 10%. The discrepancies between the modelled and observed estimates suggest that uncertainty in chemistry and transport processes~~not considered here~~, which have not been considered here, may play a substantial role in governing the CH$_4$ lifetime.

The sensitivity ~~for~~ to each variable is determined by variance decomposition, which quantifies the contribution of each variable to the variance in the model output, and is shown in Fig. 3. This is performed through calculation of the sensitivity indices using the Sobol approach (e.g., Saltelli, 2002), and the mathematical foundation for this is described in Ryan et al. (2018). We neglect the contribution of interactions between variables, which can be identified through this approach but which remain below 4% of the variance for the model responses examined here. For the global O$_3$ burden, the models show relatively similar sensitivities ~~, with atmospheric humidity contributing 55–60~~to atmospheric humidity which contributes 20–23% of the variance in all three models, and ~~to~~ dry deposition processes ~~contributing 15–20~~which contribute 21–25%, see Fig. 3. ~~The models show different~~ However, there are substantial differences in sensitivities to lightning ~~, isoprene emissions and wet deposition~~ ~~, but the contribution of these terms is relatively small. While a large sensitivity to humidity is expected, given the important role of atmospheric water vapour as a chemical sink of~~ NO, which varies from 13% in the GISS GCM to 40% in CAM-Chem, and to isoprene emissions, which are 14% in FRSGC/UCI CTM and GISS GCM but only 1% in CAM-Chem. The consistent sensitivities to humidity and dry deposition are expected, given the important roles that these play as sinks of O$_3$ ~~, it is notable that humidity has not been prescribed in previous model intercomparison studies and is rarely diagnosed except where changes in climate are expected (e.g., Lamarque et al., 2013). It may therefore constitute a substantial fraction of the diversity in tropospheric~~ in the troposphere. A strong sensitivity to lightning NO is also expected given the greater chemical O$_3$ ~~burden seen across these studies.~~ production efficiency of NO$_x$ in the mid- and upper troposphere, but the differing sensitivities between models likely reflect both differences in chemical environment and in lightning source distribution. Similarly, differences in sensitivity to isoprene are likely to reflect differences in the complexity of the photochemical schemes in the models and in the resulting chemical environment in the tropical boundary layer.

8
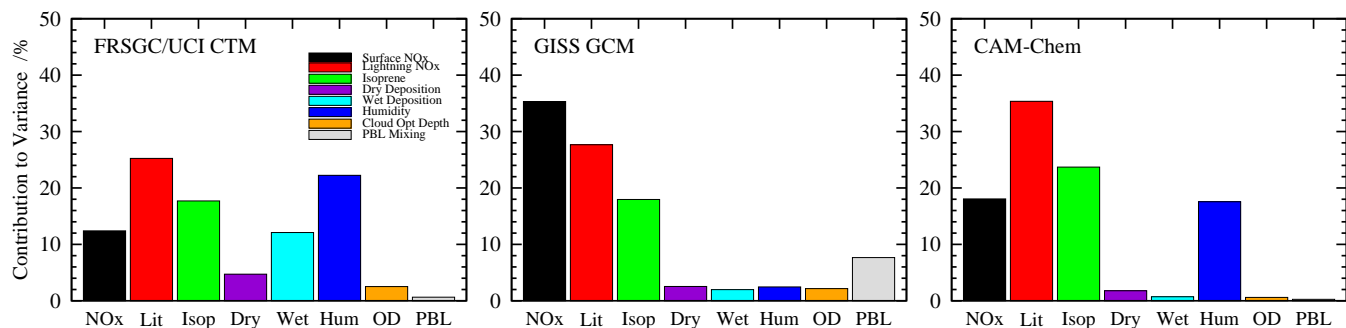
Figure 4. Contributions of each variable to the total variance in the simulated annual mean CH_4 chemical lifetime in each model.

For the tropospheric CH_4 lifetime, the models show ~~markedly~~ notably different sensitivities, with humidity contributing ~~nearly 60~~about 20% of the variance for the FRSGC/UCI CTM and CAM-Chem, but less than ~~5~~3% for the GISS GCM, ~~where~~
185 ~~emissions of~~ see Fig. 4. There is broad consistency between the FRSGC/UCI CTM and CAM-Chem, where uncertainty in lightning NO is the largest contributor and emissions of isoprene and surface NO_x are about 30% and 50% less, respectively, but in the GISS GCM the strongest sensitivity is to surface NO_x ~~and isoprene are responsible for nearly 80% of the variance, see Fig. 4.~~ emissions. It is clear that the factors governing tropospheric OH are ~~very~~ substantially different in the models, highlighting differences in chemical environment and transport patterns that affect the location and magnitude of CH_4 oxidation.
190 Sensitivity to humidity suggests that primary sources of OH ~~dominate~~ through photolysis of O_3 and subsequent reaction of O^1D with water vapour are important. Sensitivity to NO_x emissions ~~suggests that~~ reflects the importance of secondary sources of OH ~~dominate~~ through oxidation of NO, and sensitivity to isoprene highlights the importance of VOC as a source and sink of OH and as a mechanism for locking up and transporting NO_x. Interestingly, the GISS GCM ~~also~~ shows substantial sensitivity to boundary layer mixing, highlighting the importance of transport of fresh emissions from the surface for secondary OH for-
195 mation. The FRSGC/UCI CTM shows some sensitivity to wet deposition, suggesting that scavenging of nitric acid has a direct impact on OH through its influence on the abundance of NO_x.

These differences have important implications for assessment of future composition change. Future scenarios projecting increased emissions of greenhouse gases and reduced emissions of O_3 precursors (e.g., RCPs 4.5, 6.0 and 8.5) are likely to lead to increased future humidity and reduced surface NO_x. The FRSGC/UCI CTM and CAM-Chem would be expected to show a
200 reduction in CH_4 lifetime due to greater OH concentrations associated with higher water vapour, while the GISS GCM would show an increase in CH_4 lifetime due to lower secondary production of OH associated with reduced surface NO_x emissions. Analysis of future changes in CH_4 lifetime for models contributing to the ACCMIP intercomparison suggests that this is indeed the case, with the GISS GCM one of three models showing increased lifetime by 2100 for the RCP6.0 pathway, and four models showing decreased lifetime (Voulgarakis et al., 2013). An understanding of the causes of this differing sensitivity
205 is thus important for explaining the different model responses.

## 4 Investigating model differences

The sensitivity of modelled O_3 burden and CH_4 lifetime to two key variables, humidity and surface NO_x emissions, is shown for the FRSGC/UCI CTM and GISS GCM in Fig. 5. These response surfaces are generated using the emulator for each model assuming that the other six variables are unchanged. While the O_3 burden is slightly higher in the GISS GCM than the FRSGC/UCI CTM, 342 vs 314 Tg in the model control runs, the gradients across the response surfaces are similar in the models. The highest O_3 burdens occur at high NO_x emissions and low humidity, reflecting greater production and reduced loss, respectively. The relative changes in O_3 burden with NO_x emissions and humidity are very similar across all three models, as shown in Fig. 6. The responses for CH_4 lifetime show notably different behaviour, with greater sensitivity to NO_x and much less sensitivity to humidity in the GISS GCM compared to the other models. At high humidities the CH_4 lifetime appears almost insensitive to humidity, suggesting either little additional formation of OH or a matching OH sink in this model. In contrast, the other models show a very similar degree of sensitivity to humidity in both O_3 burden and CH_4 lifetime that ranges from +7% to -5% across the humidity range considered here, see Fig. 6. This suggests a much stronger coupling between O_3 and OH formation, and highlights the greater importance of the primary OH source in these models.

The response surfaces shown here allow us to estimate the impact of changes in future humidity and surface NO_x emissions in the absence of other changes. A reduction in NO_x emissions from 40 to $30\,\mathrm{TgN\,yr^{-1}}$ and increase in humidity of 15%, corresponding loosely to the changes between 2000 and 2050 expected along the RCP8.5 pathway (van Vuuren et al., 2011), would lead to an increase in CH_4 lifetime of 1.3 yr in the GISS GCM (from 11.7 to 13.0 yr), an increase of 0.2 yr in CAM-Chem, and no change in the FRSGC/UCI CTM. While this neglects the influence of other emission and climate changes, particularly the increase in CH_4 concentrations which would extend the lifetime in all models, it demonstrates the very different sensitivities anticipated for different models under future climate scenarios.

To help identify the cause of the differing model responses, we show the contribution of key variables to the variance in the annual mean tropospheric column CH_4 chemical loss rate at each model grid point in Fig. 7. This shows how the contribution of the different processes governing CH_4 removal varies geographically and reveals further differences between the models. For the FRSGC/UCI CTM and CAM-Chem, humidity makes an important contribution to the variance in tropical regions and at mid-latitudes, and makes a smaller contribution at the equator, where the greatest contribution is from lightning NO in all three models. Humidity makes very little contribution to the variance in the GISS GCM, and this principally occurs downwind of major anthropogenic emission regions. The underlying humidities in the models are relatively similar (see distributions presented in the supplement), and the annual mean global atmospheric water burden is also similar, only 4% less in the GISS GCM than in the FRSGC/UCI CTM. Given the similar humidities and similar responses in O_3 burden, this suggests that there are significant differences in
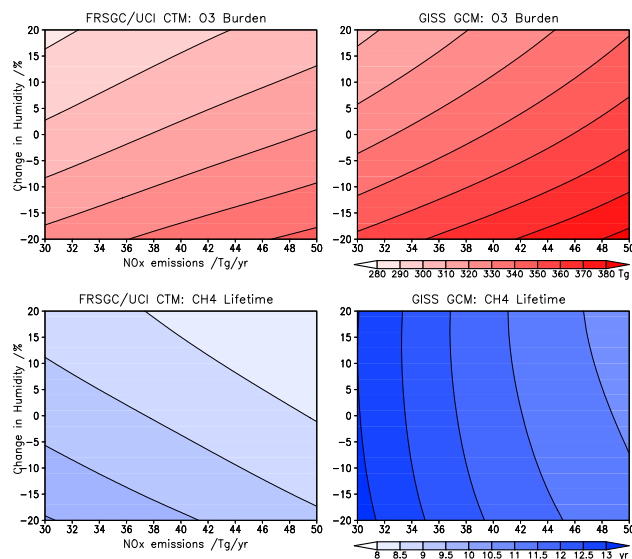
Figure 5. Sensitivity of tropospheric O_3 burden and CH_4 chemical lifetime to changes in surface NO_x emissions and humidity in the FRSGC/UCI CTM and GISS GCM.
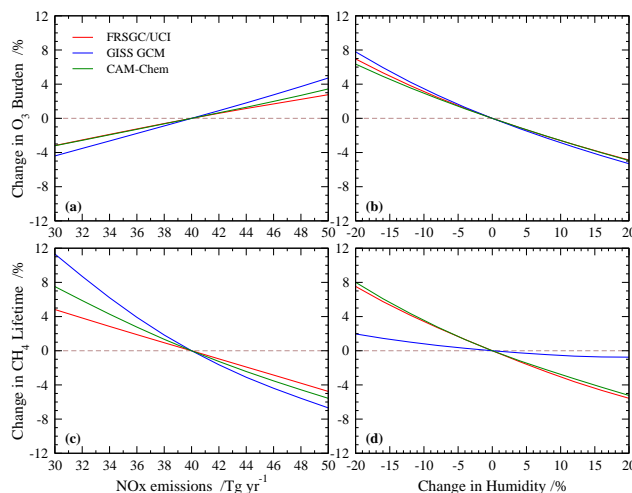
Figure 6. Relative changes in tropospheric O_3 burden **(a, b)** and CH_4 chemical lifetime **(c, d)** to changes in surface NO_x emissions and humidity alone in each model.

240 chemical processes specific to OH. Despite the larger relative importance of surface NO_x emissions ~~are largely limited to the~~ ~~source regions of these gases, indicating that their impacts on OH are relatively localised. For the~~ in the GISS GCM, ~~humidity~~ ~~dominates at mid-latitudes, but contributes little in tropical regions, where primary emissions of~~ the absolute contribution to the variance in the three models is similar. Surface NO_x ~~and isoprene dominate. Of particular interest is the widespread impact from~~

11

~~surface~~ emissions ~~, which contribute~~ have a widespread impact, contributing substantially to CH_4 ~~lifetime~~ removal over remote
ocean regions. The effect of NO_x on OH in these locations suggests that substantial nitrogen is transported to these regions in
the form of reservoir species such as peroxyacetyl nitrate (PAN), and this is supported by the patterns of transport seen in the
isoprene contribution~~, where the effects are not localised to source regions as they are for the other two models. ~~. The greatest
effect of isoprene emissions is localised in the tropical continental source regions due to the relatively short lifetime of isoprene
and its oxidation products, but there are substantial contributions downwind over the oceans, particularly in CAM-Chem and
the GISS GCM. Mid-tropospheric PAN concentrations are much greater in the GISS GCM, and comparison of tropospheric
NO_2 columns suggest that there are higher levels of NO_x over oceanic regions in this model (see supplementary material). It is
therefore likely that differing treatments of NO_y chemistry are one cause of the different model sensitivities. However, a more
detailed exploration of the sensitivity to ~~chemical~~ photochemical processes would be needed to confirm this. Tropospheric
OH is dependent on the total ozone column in the tropics through its effect on photolysis rates, and this may play a role in
model differences, although we note that mean tropical ozone column in the present models is very similar at 258–265 DU
(see Table S2). Underlying differences in meteorological fields governing vertical transport processes such as convection are
also likely to be important in this region. Our analysis provides a valuable guide to locations where model responses are likely
to differ most, such as in tropical oceanic regions, and further investigation of OH sensitivity in these regions should bring
improvements in our understanding of atmospheric processes and in their representation in current global-scale models.

5   Conclusions

We have demonstrated the value of Gaussian process emulation in performing global sensitivity analysis of computationally-
intensive global atmospheric chemistry transport models, and in applying this across a number of models to investigate model
diversity. The approach provides a simple way of exploring the sensitivity of key terms in the tropospheric oxidant budget to
governing processes and inputs, and we show that it can provide substantial new insight into the ~~causes of differences between
models~~ differing responses of models under different emission and climate scenarios.

Our study has highlighted the large sensitivity of the tropospheric O_3 burden to atmospheric water vapour, suggesting that
this variable should be diagnosed or perhaps constrained in future model intercomparisons to permit clearer characterization
of differences in model chemistry. We also find a strong sensitivity to precursor emissions and to dry deposition processes, as
expected. More surprisingly, we find that the drivers of variability in global OH can be very different between models, and this
may contribute to the large diversity in modelled tropospheric CH_4 lifetimes seen in recent model intercomparisons. Given
the importance of atmospheric oxidising capacity for both air quality and climate change, this difference in OH behaviour is a
major cause for concern and is a clear priority for further investigation.

While we have shown the value of emulation approaches for exploring model behaviour much more thoroughly than through
simple one-at-a-time sensitivity studies, this study has been largely exploratory in nature, investigating the effects of a very
limited number of variables. A more detailed global uncertainty analysis is required that considers a wider range of model
processes and inputs and incorporates a more rigorous assessment of uncertainty in each variable. Application of observation-
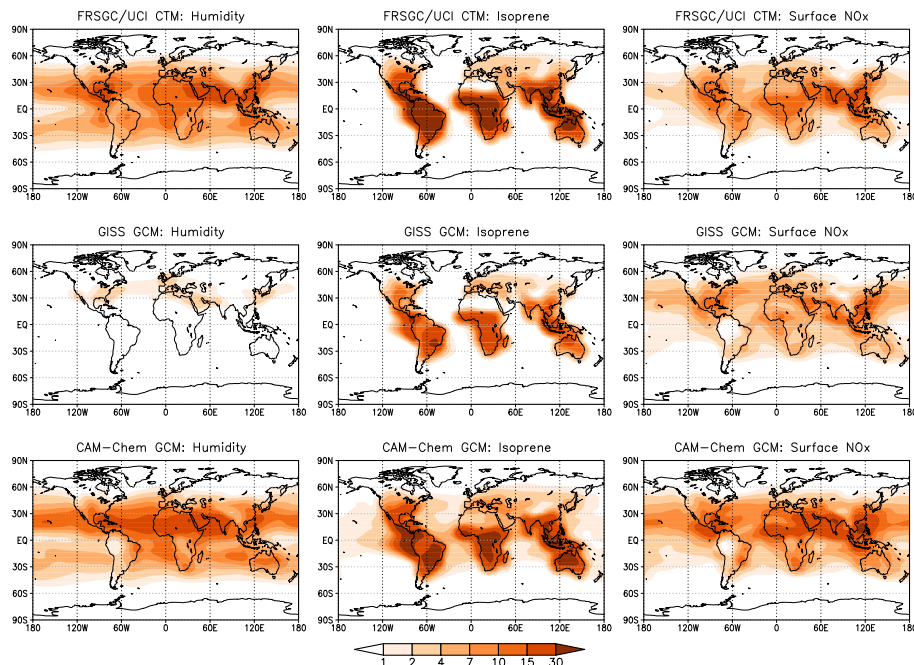
Figure 7. Contributions ~~(in %)~~ to the total variance in the annual tropospheric column CH_4 chemical loss rate (in $\mathrm{mg\,m^{-2}\,yr^{-1}}$) in each model from ~~the dominant variables:~~ humidity, isoprene emissions and surface NO_x emissions. Fractional contributions (sensitivity indices) are presented in Fig. S3 in the supplement.

based constraints is then needed to restrict the size of the response space to calibrate the models and identify specific processes in need of refinement. Applying this approach across different models accommodates the structural uncertainties in model formulation, permitting a more robust assessment of process understanding. This would provide a strong evaluation framework

280 for improving understanding of the physical and chemical processes driving atmospheric composition change and its effects on air quality and climate.

Data availability. The monthly mean output from each model for the ensemble of runs performed in this study will be made available from the CEDA data archive, and can be accessed by request to the corresponding author.

Author contributions. OW, LL, FO and AV designed the study. OW, AV and JL ran model simulations, and ER and LL performed the

285 analysis. OW prepared the manuscript with contributions from all co-authors.

Competing interests. The authors declare that they have no conflict of interest.

13

290

14

References

Akimoto, H.: Global air quality and pollution, Science, 302, 1716–1719, 2003.

Aleksankina, K., Reis, S., Vieno, M., and Heal, M. R.: Advanced methods for uncertainty assessment and global sensitivity analysis of an Eulerian atmospheric chemistry transport model, Atmos. Chem. Phys., 19, 2881–2898, https://doi.org/10.5194/acp-19-2881-2019, 2019.

Ashworth, K., Wild, O., and Hewitt, C. N.: Sensitivity of isoprene emissions estimated using MEGAN to the time resolution of input climate data, Atmos. Chem. Phys., 10, 1193–1201, https://doi.org/10.5194/acp-10-1193-2010, 2010.

Beddows, A. V., Kitwiroon, N., Williams, M. L., and Beevers, S. D.: Emulation and sensitivity analysis of the Community Multiscale Air Quality Model for a UK ozone pollution episode, Environ. Sci. Technol., 51, 6229–6236, https://doi.org/10.1021/acs.est.6b05873, 2017.

Christian, K. E., Brune, W. H., and Mao, J.: Global sensitivity analysis of the GEOS-Chem chemical transport model: ozone and hydrogen oxides during ARCTAS (2008), Atmos. Chem. Phys., 17, 3769–3784, https://doi.org/10.5194/acp-17-3769-2017, 2017.

Derwent, R. G., and Murrells, T. P.: Impact of policy-relevant scenarios on ozone in southern England: Influence of chemical mechanism choice, Atmos. Environ., 72, 89–96, https://doi.org/10.1016/j.atmosenv.2013.02.038, 2013.

Eyring, V., Lamarque, J.-F., Hess, P., Arfeuille, F., Bowman, K., Chipperfield, M. P., Duncan, B., Fiore, A., Gettelman, A., Giorgetta, M. A., Granier, C., Hegglin, M., Kinnison, D., Kunze, M., Langematz, U., Luo, B., Martin, R., Matthes, K., Newman, P. A., Peter, T., Robock, A., Ryerson, T., Saiz-Lopez, A., Salawitch, R., Schultz, M., Shepherd, T. G., Shindell, D., Staehelin, J., Tegtmeier, S., Thomason, L., Tilmes, S., Vernier, J.-P., Waugh, D. W., and Young, P. J.: Overview of IGAC/SPARC Chemistry-Climate Model Initiative (CCMI) community simulations in support of upcoming ozone and climate assessments, SPARC Newsletter, 40, 48–66, 2013a.

Fiore, A. M., Dentener, F. J., Wild, O., Cuvelier, C., Schultz, M. G., Hess, P., Textor, C., Schulz, M., Doherty, R. M., Horowitz, L. W., MacKenzie, I. A., Sanderson, M. G., Shindell, D. T., Stevenson, D. S., Szopa, S., Van Dingenen, R., Zeng, G., Atherton, C., Bergmann, D., Bey, I., Carmichael, G., Collins, W. J., Duncan, B. N., Faluvegi, G., Folberth, G., Gauss, M., Gong, S., Hauglustaine, D., Holloway, T., Isaksen, I. S. A., Jacob, D. J., Jonson, J. E., Kaminski, J. W., Keating, T. J., Lupu, A., Marmer, E., Montanaro, V., Park, R. J., Pitari, G., Pringle, K. J., Pyle, J. A., Schroeder, S., Vivanco, M. G., Wind, P., Wojcik, G., Wu, S. and Zuber, A.: Multimodel estimates of intercontinental source-receptor relationships for ozone pollution, J. Geophys. Res., 114, D04301, https://doi.org/10.1029/2008JD010816, 2009.

Gaudel, A., Cooper, O. R., Ancellet, G., Barret, B., Boynard, A., Burrows, J. P., Clerbaux, C., Coheur, P.-F., Cuesta, J., Cuevas, E., Doniki, S., Dufour, G., Ebojie, F., Foret, G., Garcia, O., Granados Muños, M. J., Hannigan, J. W., Hase, F., Huang, G., Hassler, B., Hurtmans, D., Jaffe, D., Jones, N., Kalabokas, P., Kerridge, B., Kulawik, S. S., Latter, B., Leblanc, T., Le Flochmoën, E., Lin, W., Liu, J., Liu, X., Mahieu, E., McClure-Begley, A., Neu, J. L., Osman, M., Palm, M., Petetin, H., Petropavlovskikh, I., Querel, R., Rahpoe, N., Rozanov, A., Schultz, M. G., Schwab, J., Siddans, R., Smale, D., Steinbacher, M., Tanimoto, H., Tarasick, D. W., Thouret, V., Thompson, A. M., Trickl, T., Weatherhead, E., Wespes, C., Worden, H .M., Vigouroux, C., Xu, X., Zeng, G. and Ziemke, J.: Tropospheric Ozone Assessment Report: Present-day distribution and trends of tropospheric ozone relevant to climate and global atmospheric chemistry model evaluation. Elem Sci Anth, 6(1), 39, http://doi.org/10.1525/elementa.291, 2018.

IPCC, Climate Change 2013: The Physical Science Basis, Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)], Cambridge University Press, Cambridge, UK, and New York, USA, 1535 pp, https://doi.org/10.1017/CBO9781107415324, 2013.

330   Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, B., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K., Ropelewski, C., Wang, J., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-year reanalysis project, B. Am. Meteorol. Soc., 77, 437–471, https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2, 1996.

Klein, S. A., Zhang, T., Zelinka, M. D., Pincus, R., Boyle, J. and Gleckler, P. J.: Are climate model simulations of clouds improving? An

335   evaluation using the ISCCP simulator, J. Geophys. Res. Atmos., 118, 1329–1342, https://doi.org/10.1002/jgrd.50141, 2013.

Lamarque, J.-F., Emmons, L. K., Hess, P. G., Kinnison, D. E., Tilmes, S., Vitt, F., Heald, C. L., Holland, E. A., Lauritzen, P. H., Neu, J., Orlando, J. J., Rasch, P. J., and Tyndall, G. K.: CAM-chem: description and evaluation of interactive atmospheric chemistry in the Community Earth System Model, Geosci. Model Dev., 5, 369–411, https://doi.org/10.5194/gmd-5-369-2012, 2012.

Lamarque, J.-F., Shindell, D. T., Josse, B., Young, P. J., Cionni, I., Eyring, V., Bergmann, D., Cameron-Smith, P., Collins, W. J., Doherty, R.,

340   Dalsoren, S., Faluvegi, G., Folberth, G., Ghan, S. J., Horowitz, L. W., Lee, Y. H., MacKenzie, I. A., Nagashima, T., Naik, V., Plummer, D., Righi, M., Rumbold, S. T., Schulz, M., Skeie, R. B., Stevenson, D. S., Strode, S., Sudo, K., Szopa, S., Voulgarakis, A., and Zeng, G.: The Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP): overview and description of models, simulations and climate diagnostics, Geosci. Model Dev., 6, 179-206, https://doi.org/10.5194/gmd-6-179-2013, 2013.

Lee, L. A., Carslaw, K. S., Pringle, K. J., Mann, G. W., and Spracklen, D. V.: Emulation of a complex global aerosol model to quantify

345   sensitivity to uncertain parameters, Atmos. Chem. Phys., 11, 12253–12273, https://doi.org/10.5194/acp-11-12253-2011, 2011.

Lee, L. A., Pringle, K. J., Reddington, C. L., Mann, G. W., Stier, P., Spracklen, D. V., Pierce, J. R., and Carslaw, K. S.: The magnitude and causes of uncertainty in global model simulations of cloud condensation nuclei, Atmos. Chem. Phys., 13, 8879–8914, https://doi.org/10.5194/acp-13-8879-2013, 2013.

Monks, P. S., Archibald, A. T., Colette, A., Cooper, O., Coyle, M., Derwent, R., Fowler, D., Granier, C., Law, K. S., Mills, G. E., Steven-

350   son, D. S., Tarasova, O., Thouret, V., von Schneidemesser, E., Sommariva, R., Wild, O., and Williams, M. L.: Tropospheric ozone and its precursors from the urban to the global scale from air quality to short-lived climate forcer, Atmos. Chem. Phys., 15, 8889–8973, https://doi.org/10.5194/acp-15-8889-2015, 2015.

Newsome, B. and Evans, M.: Impact of uncertainties in inorganic chemical rate constants on tropospheric composition and ozone radiative forcing, Atmos. Chem. Phys., 17, 14333–14352, https://doi.org/10.5194/acp-17-14333-2017, 2017.

355   O'Hagan, A.: Bayesian analysis of computer code outputs: A tutorial, Reliability Engineering and System Safety, 91, 1290–1300, https://doi.org/10.1016/j.ress.2005.11.025, 2006.

Parrish, D. D., Lamarque, J.-F., Naik, V., Horowitz, L., Shindell, D. T. , Staehelin, J., Derwent, R., Cooper, O. R., Tanimoto, H., Volz-Thomas, A., Gilge, S., Scheel, H.-E., Steinbacher, M., and Fröhlich, M.: Long-term changes in lower tropospheric baseline ozone concentrations: Comparing chemistry-climate models and observations at northern midlatitudes, J. Geophys. Res. Atmos., 119, 5719–5736,

360   https://doi.org/10.1002/2013JD021435, 2014.

Prather, M. J., Holmes, C. D., and Hsu, J.: Reactive greenhouse gas scenarios: Systematic exploration of uncertainties and the role of atmospheric chemistry, Geophys. Res. Lett., 39, L09803, https://doi.org/10.1029/2012GL051440, 2012.

Ridley, D. A., Cain, M., Methven, J., and Arnold, S. R.: Sensitivity of tropospheric ozone to chemical kinetic uncertainties in air masses influenced by anthropogenic and biomass burning emissions, Geophys. Res. Lett., 44, https://doi.org/10.1002/2017GL073802, 2017.

365   Ryan, E., Wild, O., Voulgarakis, A., and Lee, L.: Fast sensitivity analysis methods for computationally expensive models with multidimensional output, Geosci. Model Devel., 11, https://doi.org/10.5194/gmd-11-3131-2018, 2018.

Saltelli, A.: Making best use of model evaluations to compute sensitivity indices, Comput. Phys. Commun., 145, 280–297, https://doi.org/10.1016/S0010-4655(02)00280-1, 2002.

Schumann, U. and Huntrieser, H.: The global lightning-induced nitrogen oxides source, Atmos. Chem. Phys., 7, 3823–3907,
370   https://doi.org/10.5194/acp-7-3823-2007, 2007.

Shindell, D., Kuylenstierna, J. C. I., Vignati, E., van Dingenen, R., Amann, M., Klimont, Z., Anenberg, S. C., Muller, N., Janssens-Maenhout, G., Raes, F., Schwartz, J., Faluvegi, G., Pozzoli, L., Kupiainen, K., Höglund-Isaksson, L., Emberson, L., Streets, D., Ramanathan, V., Hicks, K., Oanh, N. T. K., Milly, G., Williams, M., Demkine, V., and Fowler, D.: Simultaneously Mitigating Near-Term Climate Change and Improving Human Health and Food Security, Science 335, 6065, 183–189, https://doi.org/10.1126/science.1210026, 2012.

375 Shindell, D. T., Pechony, O., Voulgarakis, A., Faluvegi, G., Nazarenko, L., Lamarque, J.-F., Bowman, K., Milly, G., Kovari, B., Ruedy, R., and Schmidt, G. A.: Interactive ozone and methane chemistry in GISS-E2 historical and future climate simulations, Atmos. Chem. Phys., 13, 2653–2689, https://doi.org/10.5194/acp-13-2653-2013, 2013.

Stevenson, D. S., Dentener, F. J., Schultz, M. G., Ellingsen, K., van Noije, T. P. C., Wild, O., Zeng, G., Amann, M., Atherton, C. S., Bell, N., Bergmann, D. J., Bey, I., Butler, T., Cofala, J., Collins, W. J., Derwent, R. G., Doherty, R. M., Drevet, J., Eskes, H. J., Fiore, A. M., Gauss,
380   M., Hauglustaine, D. A., Horowitz, L. W., Isaksen, I. S. A., Krol, M. C., Lamarque, J.-F., Lawrence, M. G., Montanaro, V., Müller, J.-F., Pitari, G., Prather, M. J., Pyle, J. A., Rast, S., Rodriguez, J. M., Sanderson, M. G., Savage, N. H., Shindell, D. T., Strahan, S. E., Sudo, K. and Szopa, S.: Multimodel ensemble simulations of present-day and near-future tropospheric ozone, J. Geophys. Res., 111, D08301, https://doi.org/10.1029/2005JD006338, 2006.

Stevenson, D. S., Young, P. J., Naik, V., Lamarque, J.-F., Shindell, D. T., Voulgarakis, A., Skeie, R. B., Dalsoren, S. B., Myhre, G., Berntsen,
385   T. K., Folberth, G. A., Rumbold, S. T., Collins, W. J., MacKenzie, I. A., Doherty, R. M., Zeng, G., van Noije, T. P. C., Strunk, A., Bergmann, D., Cameron-Smith, P., Plummer, D. A., Strode, S. A., Horowitz, L., Lee, Y. H., Szopa, S., Sudo, K., Nagashima, T., Josse, B., Cionni, I., Righi, M., Eyring, V., Conley, A., Bowman, K. W., Wild, O., and Archibald, A.: Tropospheric ozone changes, radiative forcing and attribution to emissions in the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP), Atmos. Chem. Phys., 13, 3063–3085, https://doi.org/10.5194/acp-13-3063-2013, 2013.

390 Tarasick, D., Galbally, I. E., Cooper, O. R., Schultz, M. G., Ancellet, G., Leblanc, T., Wallington, T. J., Ziemke, J., Liu, X., Steinbacher, M., Staehelin, J., Vigouroux, C., Hannigan, J. W., Garcia, O., Foret, G., Zanis, P., Weatherhead, E., Petropavlovskikh, I., Worden, H., Osman, M., Liu, J., Chang, K.-L., Gaudel, A., Lin, M., Granados-Muñoz, M., Thompson, A. M., Oltmans, S. J., Cuesta, J., Dufour, G., Thouret, V., Hassler, B., Trickl, T., and Neu, J. L.: Tropospheric Ozone Assessment Report: Tropospheric ozone from 1877 to 2016, observed levels, trends and uncertainties, Elem. Sci. Anth., 7, 39, https://doi.org/10.1525/elementa.376, 2019.

395 van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., Lamarque, J.-F., Matsui, T., Meinshausen, M., Nakicenovic, N., Smith, S. J., and Rose, S. K.: The Representative Concentration Pathways: An overview, Clim. Change, 109, 5—31, https://doi.org/10.1007/s10584-011-0148-z, 2011.

Voulgarakis, A., Naik, V., Lamarque, J.-F., Shindell, D. T., Young, P. J., Prather, M. J., Wild, O., Field, R. D., Bergmann, D., Cameron-Smith, P., Cionni, I., Collins, W. J., Dalsøren, S. B., Doherty, R. M., Eyring, V., Faluvegi, G., Folberth, G. A., Horowitz, L. W., Josse, B.,
400   MacKenzie, I. A., Nagashima, T., Plummer, D. A., Righi, M., Rumbold, S. T., Stevenson, D. S., Strode, S. A., Sudo, K., Szopa, S., and Zeng, G.: Analysis of present day and future OH and methane lifetime in the ACCMIP simulations, Atmos. Chem. Phys., 13, 2563–2587, https://doi.org/10.5194/acp-13-2563-2013, 2013.

Young, P. J., Archibald, A. T., Bowman, K. W., Lamarque, J.-F., Naik, V., Stevenson, D. S., Tilmes, S., Voulgarakis, A., Wild, O., Bergmann, D., Cameron-Smith, P., Cionni, I., Collins, W. J., Dalsøren, S. B., Doherty, R. M., Eyring, V., Faluvegi, G., Horowitz, L. W., Josse, B.,

405   Lee, Y. H., MacKenzie, I. A., Nagashima, T., Plummer, D. A., Righi, M., Rumbold, S. T., Skeie, R. B., Shindell, D. T., Strode, S. A., Sudo, K., Szopa, S., and Zeng, G.: Pre-industrial to end 21st century projections of tropospheric ozone from the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP), Atmos. Chem. Phys., 13, 2063–2090, https://doi.org/10.5194/acp-13-2063-2013, 2013.

       Wild, O.: Modelling the global tropospheric ozone budget: exploring the variability in current models, Atmos. Chem. Phys., 7, 2643–2660, 410   https://doi.org/10.5194/acp-7-2643-2007, 2007.