

We thank the two referees' useful and constructive comments and detail our responses as below. We have revised the manuscript substantially, taking into account all comments, and tried to address all issues raised; we hope that the manuscript has been greatly improved in both its scientific value and the presentation. Significant changes include the inclusion of detailed CO and HCHO budgets, allowing the model differences and the differences in the model response to changes in biogenic emissions to now be more quantitatively addressed. While we aim to address every comment, there are some limitations that we cannot fully resolve at this stage, but we will try to follow up in a future study. In our revised manuscript, the modified text is shown in bold, and the new/modified tables and figures are highlighted. The referees' comments are copied below, with our responses in bold.

Anonymous Referee #1

The purpose of this manuscript is to investigate the causes of model biases for CO and formaldehyde in the southern hemisphere. Overall, this is a well written and very thorough manuscript, but I do have one major concern. I would like to see more discussion on the chemistry that, I believe, will significantly improve the presentation and overall impact of this manuscript.

We greatly appreciate the referee's very thoughtful comments and have revised the manuscript substantially to include chemical production and loss terms for CO and HCHO and to illustrate the differences of these terms among models and between the two simulations carried out in this study. (Section 5)

1) It would be helpful to have tables clearly showing the southern hemispheric budgets for CO and formaldehyde for each model. A graphic showing how the budgets change monthly may be instructive.

We have included the detailed CO and HCHO budget terms globally and for the southern hemisphere (divided into three sub-regions), where the model data are available, and these are shown in three new tables (Tables 4-6) in the revised version. We also include a new plot showing the monthly budgets from all available model output (Fig. 20).

2) In the same vein, it would be helpful to show the major contributors to the chemical production and loss terms for CO and formaldehyde for each model and how they vary in time. How do the four models' Henry's Law constants for formaldehyde compare?

My recollection is that formaldehyde is not highly soluble, so there could be some sensitivity to the choice of Henry's Law constant. Could one model do two simulations, one each using the high and low of the uncertainty range?

As stated above, we have included new tables showing detailed CO and HCHO budgets. HCHO is moderately soluble therefore wet deposition of HCHO is an important loss process. Very similar effective Henry's Law Coefficients are applied in each model although TM5 takes into account hydration once HCHO enters the solution, resulting in an increased effective solubility. From the available HCHO budget terms in the new Table 5, wet deposition accounts for about 10% of total loss in TM5 and about 3% in NIWA-UKCA, which we now state in the text. The absolute differences in wet deposition of HCHO are small compared to other loss terms, i.e. oxidation by OH and HCHO photolysis. Adopting the TM5 approach in other models, or even one, is not a trivial task at this stage so we leave this to a possible future investigation. We agree with the referee that there is room for differences in this respect between the models. However, such differences are not big enough to explain the significantly low modelled HCHO in the SH compared to the column measurements. (Section 5, page 29)

3) There is considerable measurement uncertainty associated with the isoprene oxidation scheme. You say that looking into this issue is “beyond the scope in this study” and cite Archibald et al. (2010). I recommend that you at least provide a paragraph or two discussing the current uncertainties in individual reactions reported in the literature and hypothesize how they may or may not be relevant to your model exercise.

We have now considerably expanded the discussion of the impact of differences in the isoprene oxidation schemes. Considering the complexity of these schemes, we are unable to fully answer the question of how individual differences in these schemes affect HCHO and CO production rates. (Section 5; Page 35 Section 7)

Minor Comments

There are too many typos and grammar issues.

We have gone through extensive proofreading of the revised version.

Why don't you use MOPITT CO data to constrain the models?

We have now shown comparisons between MOPITT CO and modelled CO for January and September 2005 in Fig. 8 and their differences in Fig. 9, and added some discussion in Section 3.2.

Anonymous Referee #2

General comments

This paper presents the results of a modelling study simulating carbon monoxide and formaldehyde concentrations in the Southern Hemisphere, using two different biogenic emissions estimates and four atmospheric chemistry and transport models, over a period of 5 years (2004-8). Model performance was evaluated through inter-model comparisons as well as observational data from surface stations. This is a timely and well-structured study. The use of atmospheric columns of a variety of marker compounds to constrain biogenic emissions estimates and elucidate the skill of atmospheric chemistry models in simulating the oxidation and transport of these species and their products is an important method in regions of sparse observations (ironically those with the highest emissions). The inclusion of simulations with artificial tracers to disentangle the contributions of transport, primary emissions and secondary production of CO is particularly nice. Given recent work indicating the influence of biogenic precursors and their climatically and air quality active products, such an evaluation is urgently needed. This study marks an important first step in this process.

Overall, the study is well presented, described and discussed. I do have concerns over the level of detail provided regarding certain aspects of the work, and these are outlined below. In particular, I feel that the analysis is not of sufficient depth, especially with regard to an attribution of the observed differences to the chemistry mechanisms included in the four atmospheric chemistry models used. There are also sections where the English is poor, particularly earlier in the paper, and the manuscript is in need of a thorough editing.

We thank the reviewer for the very detailed review and many useful comments. We are encouraged by the positive view of the usefulness this study to the community, especially the use of column measurements in this vast and data sparse region. We have carried out a substantial revision, also in accordance with Reviewer 1's comments, by including detailed CO and HCHO budgets and analyses of differences in chemical production and loss terms in each model where the data are available. We have gone through thorough proof-reading and revision to improve the presentation of the paper.

Specific comments

2. Model simulations and observations

One of my chief concerns with the paper is the level of detail provided regarding the model set-up and simulations. The information provided here for the processes, boundary conditions and driving data for each model is not consistent, making comparison difficult. The authors refer to previous model descriptions too often with regard to key processes. While it is extremely difficult to ensure full consistency between models, the paper would benefit from greater elucidation of some of the inconsistencies that might affect the conclusions drawn from their results.

We will address these comments below.

2.1 Simulations

Biogenic emissions estimates are strongly dependent on land cover (i.e. vegetation distribution), and driving meteorology. Presumably, although the two inventories are generated off-line, it is done with the “default” input fields for MEGAN-CLM and LPJGUESS.

The authors should state clearly here the input data used, as well as the spatial and temporal resolution of the emissions inventories generated, as this is a clear source of inconsistency with the chemistry models that are then driven with these emissions. For some of the model set-ups, the emissions will be more compatible with the land surface and meteorology driving data for the chemistry and transport than others. Differences in land cover, for example, will also affect deposition of the primary emissions and their oxidation products; differences in meteorology will affect chemical production and loss rates. It is possible that some of the observed differences between simulated concentrations of oxidation products are a result of these inconsistencies.

We agree that inconsistencies exist in meteorology between the fields used to generate the biogenic emissions and those used to drive the chemical transport models (CTMs) and chemistry-climate models (CCMs). CLM-MEGANv2.1 emissions were generated offline using the Community Land Model (reference given in the text), forced by meteorological reanalyses for each year (CRUNCEP; <http://dods.extra.cea.fr/data/p529viov/cruncep/readme.htm>). LPJ-GUESS was generated using the CRU TS 3.1 climate data (eccad.sedoo.fr). It would be impractical, within the constraints of this study, to harmonize the driving meteorology in global models, especially in multimodel intercomparison studies given the complex nature of such models. It is precisely the purpose of this study to minimize the differences from the dependence of biogenic emissions on the driving meteorology that we chose to use prescribed monthly mean biogenic emissions in the model rather than calculate the isoprene emissions online (e.g. GEOS-Chem and CAM-chem both have the option to calculate isoprene emissions interactively). Also we speculate that differences in the driving meteorology might be a smaller factor in defining the biogenic emissions than differences in assumptions about biological factors. (Page 7 Section 2.1)

The use of model specific natural emissions of ocean CO is inconsistent with the use of specified biogenic emissions. Given that one aim of the study is to assess the impact of different biogenic emissions on chemistry transport model output and that anthropogenic emissions are the same for each model, surely this would have been a simple difference to eliminate.

We use the same emission dataset for ocean CO and we have clarified this in the revised version. (Page 7 Section 2.1)

Furthermore, in the descriptions of the GEOS-Chem and CAM-chem model simulations, the authors state that diurnal variability was imposed on the biogenic emissions inventories. Was this also done for NIWA-UKCA and TM5? And why? What was the temporal resolution of the biogenic emissions calculated by MEGAN and LPJ?

Yes, the diurnal variability of isoprene emissions are also imposed in NIWA-UKCA and TM5. We use monthly mean CLM-MEGANv2.1 and LPJ-GUESS emissions so we impose the diurnal variation on this emission dataset. This is done routinely in global models when using prescribed monthly mean isoprene emissions. Relevant text has been added. (Section 2.2)

Likewise the use of model specific lightning NO_x emissions is inconsistent with the approach of harmonizing emissions used to drive the four chemistry models. NO_x levels are critically important to the atmospheric oxidation of volatile organic compounds.

Again this would have been a simple source of discrepancies between models to eliminate from the study. For both lightning NO_x and ocean CO emissions, I would like to see further justification of the use of model specific inventories. How substantial are the differences between the models, and how sensitive are the models to these inputs?

We use the same ocean CO emissions (now clarified). Yearly lightning NO_x emissions are now shown in Table 1, organised from the lowest to the highest used in the models; we can say that there are no substantial inconsistencies among the models with values in CAM-chem (3.3-3.6 TgN yr⁻¹) being moderately lower than from other models (5.8-6.5 TgN yr⁻¹). Most tropospheric chemistry models now use an interactive lightning-NO_x scheme so making these emissions non-interactive would have required significant additional programming in some of the models, with limited impact on our work which focuses on HCHO and CO.

2.2 Models

The details provided about the four atmospheric chemistry and transport models used in this study are not consistent. In addition to the temporal and spatial resolution of the model and the driving meteorology, other aspects of particular importance to VOC oxidation and atmospheric concentrations are imposed methane concentrations, the dry deposition scheme and the treatment of partitioning of gas-phase condensable species to the aerosol phase. The level of detail of the chemistry mechanism is also critical.

At the very least, the authors should specifically (i.e. not simply by referring the reader to previous model descriptions) provide information of all of these in either the text or Table 2 (or both). Details of the chemistry mechanisms should include the number of chemical species (transported, reactive, fixed) and reactions (photolysis and thermal) in the full mechanism, and in the isoprene and monoterpene oxidation schemes.

Furthermore, details of relevant compound lumping (e.g. are methacrolein and methyl vinyl ketone treated separately or lumped as a single species) should be provided, particularly with regard to species within the biogenic chemical degradation paths, and related species such as methanol and acetone. I would also like to see the authors elucidate further on the treatment of organic nitrates, PAN, peroxy radicals and (in the case of isoprene) epoxides. Monoterpene chemistry, in particular, varies considerably between atmospheric chemistry models and a full description of the schemes used within each model would be useful in the form of an Appendix or as Supplementary Information.

We appreciate the reviewer's suggestions in giving full model details, especially regarding chemical mechanisms used in the models. We now include a table and added text showing many of the relevant details of the four models, including reference to the isoprene oxidation schemes used. Further detail of chemical mechanisms is difficult to include given the complexities of the model physical and dynamical processes and the large number of chemical species and reactions (e.g. hundreds of chemical reactions). Such a comprehensive description of the models is beyond the scope of this paper and would also not necessarily help in understanding the causes of differences in model behaviour. All four models have been clearly and comprehensively described in various publications which we cite in this paper.

Our focus here is not mainly on comparing the differences in mechanisms but the effect of applying different mechanisms; such effects are reflected in the budgets of the various production and loss terms which we have added in the revised version (Tables 4 and 5, and the new Section 5 of the text). Uncertainties arise from the different chemical mechanisms

implemented in the model as well as the physical processes, and all processes interact to produce the differences in modelled chemical species. Short of any targeted investigation, it is not straightforward to identify in detail which process/processes might be responsible for the differences in the modelled species. The detailed differences in chemical mechanisms are best investigated in a box model where one can control the variables, as we now state in the text. Such an approach has been exemplified in the work of Archibald et al. (2010), cited in our paper, which includes all four models used in this study. We do not think that elucidating the different treatments of organic nitrates, PAN, peroxy radicals and epoxides here will be able to explain the differences in modelled CO and HCHO alone. As stated above, we have revised the paper and added detailed budgeting terms for CO and HCHO and this allows a more complete description of the model differences in chemistry. (Section 5)

2.3 Observations

Please give details of the temporal resolution of the raw measurement data so that this can also be clearly compared to that of the models.

Column data are in daily mean format where there were measurements on that day. Indeed, we compare the modelled daily mean output to that measured on the same day when the measurements were carried out. We take monthly mean observed surface CO data and compare with the monthly mean model data. This has been more clearly stated in the revised version.

Sections 3-6

My other main concern is the lack of detailed consideration and discussion of the differences in chemistry between the models and the impact this has on the simulated HCHO and CO concentrations. It seems that it should be straight-forward enough to provide details of the main chemical sources and sinks of HCHO and CO in the various chemistry schemes, and the relative production and loss rates to each. How do these differ between low and high NO_x environments in the models?

This lack of detail makes these sections extremely unsatisfying. The authors speculate on numerous possible causes for the differences between modelled and measured HCHO, but do not explore any of these further. I recommend that they select the most likely causes and perform some sensitivity tests to establish whether it is indeed a contributor to the persistent under-estimation of HCHO in these models.

We have now included the detailed budget terms for CO and HCHO in global, SH and 3 sub-regions of the SH, as well as a new section (Section 5) detailing these terms. We have also deleted some of the speculations as to what may cause the low bias of modelled HCHO other than citing the previous studies, and we rather leave this (finding the cause) to a future study. Unfortunately we do not have sufficient resources at the moment to further explore the various causes and to carry out further sensitivity studies. We also suggest that more targeted observations might lead to the cause of low modelled HCHO in the remote SH. (Section 3.4)

3. Comparison between models and observations

Again, there is a lack of consistency, this time between sites. While it is unavoidable that Darwin has only total column data, while partial columns are preferable for this application, the authors need to consider the likely effect that this has on the comparisons they are making and the conclusions that can be drawn. Ideally, I would like to see the authors show comparisons of total columns for another site to assess how model performance differs between total and partial column values. It could be expected that the use of total columns might tend to smooth differences between models as they have greater spatial coverage.

We can see the reviewer's point of view. But two of the four models do not have detailed stratospheric chemistry and one model has no saved output from the stratosphere so it is impossible to assess how realistic modelled total columns would be compared to the observed

FTIR total columns. The observed total columns of CO include the CO production from the mesosphere which accounts for up to ~10% in the high SH latitudes in the spring when the polar vortex, which is enriched in CO of upper-atmosphere origin, makes up a substantial part of the total air column. That is why we rather use observed partial columns (0-12km) for comparison. In the case of Darwin, the contribution of CO from the upper atmosphere is minimal given its latitudinal location. We believe that this is the best possible approach we can apply in comparing modelled and FTIR columns of CO. We have added some statements in the revised version.

4. Model differences in chemistry and transport

Again there is an unnecessary inconsistency in approach. Previous evaluations of model performance had used partial (or total in the case of Darwin) columns. Here the authors choose to use tropospheric columns, with total heights set by the model specific tropopause.

For observed FTIR columns, we are restricted to the columns below 12km (for sites other than Darwin) due to the nature of the retrieved CO data, so we have to compare model results and observations in a like-for-like manner. However, when we compare the model differences we are not restricted to the region below 12km. For global (and SH) comparisons, a 12 km cut-off would not be a good choice as some model levels are the stratosphere (polar regions) and some tropospheric levels will be missing in the tropics which make significant contributions to the tropospheric columns. Likewise, for tropospheric budget calculations we define the chemical tropopause in the model using their modelled ozone as a marker as commonly done in multi-model comparison studies (e.g. Stevenson et al. 2006). Again we believe that this approach would not introduce substantial inconsistencies among the models, given the sharp gradients of the modelled ozone in the tropopause region.

Technical corrections

Title - The authors do not, to my mind, currently address model uncertainties in a quantitate enough way to justify including this in the title.

We have now changed the title to “Multi-model simulation of CO and HCHO in the Southern Hemisphere: Comparisons with observations and the impact of biogenic emissions”.

Abstract – the Abstract contains far too many undefined acronyms (e.g. NOAA GMD, NDACC, etc.). Please either spell these out in full or define them here.

Done.

1. Introduction – the English is particularly difficult to understand in this section of the paper. I suggest the authors read it through and edit it appropriately. Here, and throughout the paper, abbreviations and acronyms (particularly for the models) are used without having been previously defined (e.g. MEGAN).

Text has been modified accordingly.

The section on p2618, L4-10, seems rather contradictory. The authors state that global models are suitable tools for investigations such as these, and follow this up with the assertion that “For instance, global atmospheric models systematically overestimate observed : : :”, which seems to indicate their lack of suitability. Perhaps some rearrangement would help – the authors are right to highlight the shortcomings of the tools they are using, and to identify them as suitable.

Revised

P2618, L17 – are the authors aware of any atmospheric chemistry models that are still using a value of $_{200} \text{Tg y}^{-1}$ for isoprene emissions? I was under the impression that as model chemistry had improved, this value had not been used since TAR.

We stated that this low limit was used in Stevenson et al. (2006) so it is not very recent.

P2618, L24-29 – I'm not sure that I understand the point the authors are making here, regarding the Marais et al. study and the sparsity of direct observations. Again, rephrasing would be helpful.

Revised.

P2619, L1-4 – Again, this section appears to contradict the previous section. Further details of the model set-up used would perhaps help. Were biogenic emissions included in this study? If so, were they also generated by MEGAN?

The IMAGESv used in Vigoroux adopts isoprene emissions from Muller et al. 2008 and was generated based on MEGAN using ECMWF meteorology. The amount of yearly isoprene emissions are comparable to that of LPJ-GUESS used here. In Jones et al (2009) that we cited here the box modelling was carried out assuming initial isoprene concentrations and therefore no isoprene emission applied. The purpose of citing these two studies are that modelled HCHO in the remote SH has been underestimated by previous studies. We have moved the citations to Section 3.3, and modified the text

P2619, L14 – does the “They” used here refer again to the Morgenstern et al., study

Yes. Clarified.

2. Model simulations and observations

P2621, L2 – Please state the % difference in monoterpene emissions as has been done for isoprene.

Added in page 7.

P2622, L10 – Figure 2 does not show differences; it shows the absolute values from two models side-by-side.

Text corrected.

P2623, L13 – Please provide further information here (or elsewhere in this paper) of the simple monoterpene oxidation scheme used.

Added in page 9.

P2623, L13-4 – Describe further how this gradient is implemented.

Two different constants are applied for the NH and SH respectively. Clarified. Page 9.

P2624, L19-21 – How does this assumed yield of CO from monoterpenes compare with yields in the other chemistry schemes? What is the assumed timescale of this conversion, and again how does this compare?

This is a long-standing assumption in GEOS-Chem, adopted due to the absence of explicit monoterpene oxidation and justified in the references given in the text. We do not have sufficient diagnostics to calculate the specific yields of CO from monoterpenes from other models for this study, therefore cannot compare them in this paper.

P2628, L20 – The sites are not shown anywhere in the paper. However, I would very much like to see them shown on a map to better understand their locations and the differences between them.

A new figure (Fig. 3) has been added to show all locations of ground-based measurement sites used in this study.

3. Comparison between models and observations

P2628, L17-21 – Could the differences in direct CO emissions from the ocean (different in the various models) also be a factor?

No. They are the same among models, and this has now been clarified in the text.

P2628 – Please include a direct comparison between modelled and FTIR CO columns for LPJ simulations, either as an addition to Figure 3 or as a new figure.

Now new Fig. 5.

P2629, L2-3 – Were the emissions inventories generated from the same meteorological driving data (in which case, the model columns might be expected to show similar seasonal and inter-annual variations)? Or, if not, please could the authors indicate how seasonal and inter-annual differences in relevant meteorology compare between the different models used here.

The emissions inventories were generated using similar meteorology (see the previous response). The two emission datasets show similar seasonal variation but do not show close correlation between their inter-annual variations in some source regions (see Figure 1). We do not know if such differences are caused by differences in meteorology and/or other forcing that constrains the models. We are not able to compare the different meteorology used in the models and those for generating the emissions which have been provided using entirely different models. However such differences in the various meteorological forcing adopted in each model should have similar features in terms of seasonal and inter-annual variations in general, because most of the climate forcing used in the models are derived from the observed data. The exception is that NIWA-UKCA is a free-running general circulation model. But this model is constrained by monthly observed sea surface temperatures, and calculated chemical species concentrations do not feed into the dynamics. This is to minimise inconsistencies with other models that are either entirely driven by prescribed meteorology or nudged towards the observed meteorology.

P2629, L5 Was the multi-annual ensemble mean data derived as daily or monthly averages? If monthly, why (given that the earlier comparisons were made with daily data)? If daily, why is daily data not presented in Figure 5?

The multi-annual ensemble mean data were shown as monthly averages in the new Figure 7 so the seasonality is more clearly shown. The monthly data are the average of daily mean data of course. The purpose of showing monthly mean (averaged from daily data) is to show the seasonality of the deviation of modelled data from the observed data. Observed monthly data are also generated from the daily data.

P2629, L17 – Please make clear in the text that these are annual averages (presumably).

Done.

P2629, L18-19 – Why should the magnitude of the model biases have this order? Could the authors speculate here, or refer the reader to the section in which this is explored.

See revised text on page 16.

P2629, L20-22 – Again please speculate on the possible reasons for this seasonal difference between model variability.

See revised text on p16.

P2629, L26 – Is the greater inter-model variability due to the relative scale of the biogenic emissions from each of MEGAN and LPJ? Have the authors tried normalizing the biases by the magnitudes of the emissions inventories?

We have revised the text to indicate that this is “primarily due to the larger response of modelled CO to its higher precursor emissions”. We have not normalized the biases.

P2630, L8 and elsewhere – Please provide % differences as well as absolute.

We have modified the text accordingly. (Page 18)

P2630, L19-25 – It seems that this could be easily checked by selecting only the model data from dates that coincide with the sampling dates at Cape Grim. This would provide a much more rigorous evaluation in any case.

We have now replaced Cape Grim data with the Pacific Ocean data which is more representative of background atmosphere of the SH, and is consistent with comparisons from other sites.

P2630, L26-7 – Please speculate as to why NIWA-UKCA and GEOS-Chem should perform so well when compared against surface observations in spite of the fact their deviations were not always the lowest when assessed against CO columns.

We should clarify that models have systematic low biases compared to the FTIR column CO. Therefore the good agreement between modelled and observed surface CO are not entirely reflected in the column comparison. The modelled column CO are attributed to not only surface distribution but also the vertical distribution which is controlled by the vertical transport of both CO and its precursors. We have noted this point in the revised text. (Page 18 in Section 3.3)

P2631, L5-7 – What are the characteristics of the sites for which LPJ performs better than MEGAN? Why might this be the case?

It is clearly shown that modelled surface CO shows a lower bias from the simulation with LPJ-GUESS than that with CLM-MEGANv2.1 emissions in all cases but this does not mean that, in absolute terms, the MEGAN simulation is better than GUESS in all locations and seasons. If in the case MEGAN has a high bias (e.g. at Ascension Island in JJA), the GUESS simulation is closer to the observed surface CO. We cannot characterise the sites where GUESS performs better as there is no statistical significance due to the limited locations selected.

P2631, L9-13 – See above regarding the anomalies at Cape Grim. The strong positive bias here provides further motivation for removing polluted dates from the analysis and selecting only the sampling dates from the model output data. For the purposes of transparency, the authors could show the current time series (i.e. with all dates included) on the same plot.

We have now replaced the comparison for Cape Grim surface CO with surface CO from the “Pacific Ocean” site, which is more representative of the remote SH. This is because unlike at Cape Grim, in the models we do not only sample air during clean-air days.

P2631, L22-28 – Does Figure 8 show model simulations driven with emissions data from MEGAN or LPJ? Please make it clear here and in the caption of Figure 8.

Thanks, done. They are MEGAN simulations.

P2632, L1 – Again what are the key differences between these sites that might drive that difference in variability.

Lauder is further from the source regions than Wollongong; this indicates that the difference in chemistry among the models dominates the HCHO variation in the more remote location. Relevant statements have been added in the text.

P2632, L6-7 – 10-15% is not a small difference. What is the % underestimation?

Texts have been modified to read as: “This resulted in small overall changes, with ~10-15% increases that occur in some summer months, and the increases were not sufficient to close the gap between the models and the observations.” Note that we assume the maximum daily HCHO and 10-15% increase is the upper limit. We do not have data with higher temporal resolution than daily mean from three models. (Page 19 Section 3.4)

P2632 – Again please show the comparisons between the modelled and observed columns for the chemistry models driven by the second emissions inventory (presumably LPJ again) either as an addition to Figure 8 or as an additional figure.

It does not add much new information to show the comparison between observed and modelled HCHO with LPJ-GUESS emissions because the two simulated HCHO fields are almost indistinguishable, as also shown in Figure 13. We have too many plots already together with some new plots so prefer to leave as is.

P2632, L10 – the multi-annual monthly mean (?)

Yes. Clarified.

P2632, L23-5 – Have the authors tried to scale methanol emissions input into the chemistry models to determine whether this makes a substantial difference to HCHO concentrations?

We have now removed this discussion. We do not yet have observational data of methanol in this region so do not want to speculate at this stage. This could be the subject of future investigations.

P2632, L25 – P2633, L1 – This section is highly confusing and poorly explained. I would strongly urge the authors to include the full set of reactions to which they are referring. The authors jump from a discussion of methanol emissions to a reaction yield of HCHO from CH₃OOH (acetic acid, rather than methanol). Please explain the connection between a reaction commonly included in methane oxidation in chemistry models and the methanol budget. Are the specific reactions referred to here included explicitly in all of the models used in this study?

We have revised the text on this discussion and made some clarifications. The purpose was to cite the previous work on modelling HCHO albeit using a box model. We drew a common conclusion which is that models underestimate HCHO in the remote SH. (Page 20 Section 3.4)

P2633, L11 – What % of global terrestrial isoprene emissions?

Related text deleted

P2633, L21-29 – Please give details of the chemistry included in the IMAGES model. How does it compare to that included in the models used in this study, and how do the conclusions drawn by Vigouroux et al. apply to the results here?

These points have been added/discussed in the revised text. Page 21.

P2634, L1-2 – Why should a reduced oxidative capacity result in an increase in HCHO?

Corrected – should be “lower OH in the model”.

P2634, L9 – “differnces” should read “differences”

Corrected.

P2634, L21 – Please explain here rather than later in the text why January and September have been chosen for this analysis.

Revised. Page 22.

P2635 – Please provide the same figure for CO_OH as for CO_25 (Fig. 11)

Added. Fig. 16.

P2635, L25 – “towars” should read “towards”

Corrected.

P2636, L5 – “reflects” should read “reflects”

Corrected.

P2636, L15-6 – Please explain/speculate here why NIWA-UKCA should show such a small contribution from primary emissions, and relate this to the results/discussion already presented from comparisons of the partial CO and surface CO data.

The relatively small contribution from primary emissions to CO column in NIWA-UKCA is related to the relatively large chemical production compared to other models; this now has been shown in Table 4.

P2636, L21 – “agreement” would read better than “correspondence”

Changed.

P2637, L4 – “deduce” would read better than “derive”

Changed.

P2637, L25-9 – As noted previously, the authors should include full details of the isoprene and monoterpene oxidation schemes of the four chemistry models used in these simulations. Presumably, the authors do not mean that isoprene itself is lumped, but it would be instructive to see which of its products are lumped and how this differs between models.

The paragraph has been removed and is replaced with a section on the CO and HCHO budget analysis. We show that the underlying differences in chemistry are reflected in the rates of chemical production of CO and HCHO presented here, although we cannot identify in more

detail what processes in the chemical mechanisms are responsible. Such investigation would be better studied to a box model in which parameters can be controlled. Page 26-29.

P2637, L25-29 – It would be helpful if the authors were to refer the reader back to the bottom row of panels in Figure 13.

This has now been replaced with new text.

P2638, L1 – Please state in the text that the profiles shown in Figure 14 are again in latitudinal zones.

Added in now Fig.19.

P2638, L1-3 – Why are there no observations included in Figure 14?

This figure only shows model-model comparisons; the model data are averaged over three sub-regions in the SH and are compared. The observational data we present in this study are only for specific locations, therefore are not appropriate for comparison in this figure.

P2638 – Although I appreciate that the authors have chosen to focus on output from the MEGAN simulations, I would suggest that it would be useful to include at least a discussion of the LPJ simulations and how they compare. Are the results in keeping with those from MEGAN (even if of lower magnitude)? Are there any notable differences that might highlight the behaviour of the chemistry mechanisms under different NO_x:VOC ratios?

We have now added detailed CO and HCHO budgets for both simulations (new Section 5). Hence there is now more discussion on LPJ-GUESS results. Substantial revisions have been carried out as a result.

P2638, L8 – Again, inclusion of the detailed isoprene and monoterpene oxidation schemes employed by each of the models would greatly facilitate the discussion here.

What are the relative rate constants for the photolysis of HCHO in each of the models?

What are the production rates of CO and HCHO from the oxidation of the biogenic VOCs in each?

We now show reaction fluxes of HCHO photolysis from TM5 and NIWA-UKCA (Table 5), and find that the photolysis rates of HCHO are very similar in these two models. However we note that wet deposition of HCHO in TM5 is much higher than in NIWA-UKCA (~10% and ~3% relative to total chemical loss of HCHO, respectively). This is now discussed in Section 5.

P2638 – See above comment. It would be useful to see the production rates of CO and HCHO from OH oxidation of biogenic VOCs for each model. Ditto loss rates of HCHO to OH-initiated reactions.

We now include CO budget from all models and HCHO budget from two models as stated above. The chemical production terms include oxidation of methane and NMVOCs. We are not able to provide oxidation rates from biogenic VOCs here as not all relevant reaction rates have been saved.

P2638, L17-8 – What is the correlation coefficient for each model? Or the ratio of HCHO to OH?

We have not calculated such a correlation coefficient. In the remote SH, HCHO formation is mainly through methane oxidation; therefore, OH and HCHO are correlated.

P2638, L25 – I suggest that the authors could easily produce vertical profiles of H₂O vapour for each model, or H₂O/OH in line with the HO₂/OH profiles in Figure 14.

Water vapour fields are very similar among the models as prescribed meteorological fields are used to drive the models. The differences in OH between the models are likely due to the differences in their photolysis schemes. This has been noted in the revised text. Page 26.

P2637, L28-9 – Again, the authors should calculate the production/loss rates for these reactions to determine which is the cause, or if it is a combination of the two.

Substantial revisions have been carried out based on calculated production /loss terms. Section 5.

P2639, L1-3 – Given the complexities of the chemistry involved, it is hard to see how the authors can draw this conclusion without having conducted a more detailed analysis than is presented here.

We have more quantitative discussion on this in Section 5.

5. Sensitivity of modelled SH CO and HCHO to uncertainties in biogenic emissions

Please could the authors explain why they have (again) introduced an unnecessary inconsistency into their analysis. In the previous section they chose to focus on January and September; here January and July. Please justify fully this decision, or ideally, change one of the sections so that the same months are discussed in both.

This has been explained in the revised text. Page 30.

P2639, L17-19 – The right-hand panels in Figure 15 appear to show increases in the CO columns in LPJ relative to MEGAN in some parts of northern and central Africa. However, the choice of colour scale is poor. Furthermore, as this is a discussion of the columns over the Southern Hemisphere, there is no reason for the plots to extend north of the equator.

Increases of CO in this region reflect the higher isoprene emissions in this region in the LPJ-GUESS inventory. The “meteorological equator” (i.e. the ITCZ) divides the atmosphere into northern and southern-hemisphere air masses but the ITCZ is often north of the geographical equator. We therefore prefer the plot to also cover some areas north of the Equator.

P2640, L2 – “reginos” should read “regions”

Thanks - corrected.

P2640, L4-6 – What is the relative change in biogenic emissions between the models for each season? Does this correlate with the changes in columns? What are the seasonal differences of NO_x emissions in the region?

There are no changes in biogenic emissions between the models as the same emissions are prescribed in each model. The differences in columns here are due to differences in chemistry. Given the same changes in emissions, each models reacts in different degrees to such changes due to the underlining differences in chemistry in each model. We have removed this note to avoid confusion.

P2640, L15 – Please could the authors explain what they mean by spatial variations of the biogenic emissions. Is that variations between the two inventories? Are these the only two regions with marked spatial variations? At what resolution are these variations apparent?

We mean the increase of LPJ-GUESS isoprene emissions in these regions relative to MEGAN inventory. We have modified the text to make it clear. Page 30.

P2640, L23-29 – As these are both oxidation products of biogenic VOCs perhaps the authors could explain more clearly why one should be correlated and the other anticorrelated with OH concentrations. Again, this would be best done via an analysis of production and loss rates. What are the correlation coefficients?

We have replaced “anti-correlated” with “in opposite sign”. We have also modified the text to clearly explain why in remote region models have lower CO with LPJ-GUESS simulation but higher OH and HCHO. As suggested by the reviewer, this has been backed up by the calculated production and loss rates.

P2641, L1-4 – Again, please quantify the relative speeds of the isoprene oxidation processes in each model, and present the mechanisms used in each.

Unfortunately we cannot provide directly the HCHO production from isoprene oxidation as we do not have sufficient diagnostics to quantify it. We assume that isoprene oxidation is a dominate component. Table 5 shows that oxidation rates of NMVOCs in the SH are larger in TM5 than in NIWA-UKCA in both the absolute values and relative to the total chemical production rates.

P2641, L5-14 – Given the limited nature of the discussion of the zonal vertical profiles presented in Figure 18, I would suggest that showing the panels for January and July only would suffice. There is little difference between those for October and January in particular.

We understand the reviewer’s concern but would still like to leave the plot as is showing all four seasons. There are some seasonal differences that are reflected both in differences in two emission inventories and the difference in transport pattern. We have added a sentence regarding seasonality in the revised text. Page 32.

6. Summary and conclusions

P2641, L23 – I would not describe -19.2% as comparing well, certainly not relative to the other locations.

Changed to “reasonable agreement” Page 32.

P2642 – The word “significant” has a specific statistical meaning. As the authors have not performed any tests of significance they should not use it in this context.

We have dropped the word “significant”.

P2642, L9-12 – While the complexities of the system preclude a simple solution to the discrepancies between modeled and observed HCHO, a more detailed analysis of the differences between the chemical mechanisms and the HCHO production and loss rates to the various reactions included in each would at least indicate where future research should be directed.

Agreed. However the inter-model differences are much smaller than the differences between the models and the observations, indicating that the cause of such a large discrepancy goes beyond what differences in chemical mechanisms can explain, at least based on our study. We have modified the text accordingly. Page 35-36.

P2642, L17-20 – Please state here in the text the % difference in biogenic emissions between the MEGAN and LPJ simulations.

Added. Page 33.

P2642, L21-2 – “neither” : : : “nor” should read “either” : : : “or” as the sentence is already negative.

Thanks - corrected.

P2642, L25-26 – “reduced” would read better as “reductions in”

Replaced.

L2642, L25-28 – What is the % difference between biogenic emissions in MEGAN and LPJ for these source regions?

Added in the text. Page 33.

P2642, L1-2 – As before, the inclusion of details of the oxidation processes and a more thorough analysis and discussion of how the differences between the models affect the production/loss of HCHO and CO would strengthen this assertion.

We now show production/loss terms and include a discussion of differences in production rates of CO and HCHO from the oxidation of NMVOCs between simulations using these two emission inventories.

P2642, L6 – Without the in-depth analysis of the chemical production and loss rates of HCHO in the various models, the authors are over-stating the case, by saying that they “show”. Their results may suggest this, but they have not conclusively demonstrated causality.

This statement is now based on the detailed budget shown in Table 5.

P2643, L8-12 – To what extent do the models include the latest findings on isoprene/OH in low NO_x environments?

To various degree. For example, GEOS-Chem includes many recent updates on isoprene oxidation mechanism and does include formation of epoxide species in the model which regenerate OH under low NO_x conditions. Epoxides are not included in the other models. This point has been noted in the revised text. Page 35.

P2644, L1-2 – It would be instructive for the authors to perform sensitivity tests with the models using the same methane fields for all. This would allow them to assess the effect of NMVOC chemistry differences alone.

Unfortunately it is not realistic for all models to perform such sensitivity tests. We have now included mean surface methane concentrations used in each model (see Table 2) and hope it will satisfy. The differences in methane among the models are not sufficient to explain the differences in NMVOC chemistry.

P2644, L3 – “distribuitons” should read “distributions”

Corrected.

P2644, L3-18 – This entire section would be strengthened if the authors included the analyses of the biogenic VOC oxidation mechanisms, together with production and loss rates, as suggested previously.

Agreed. We have now quantified the production and loss rates through detailed budget calculations. Text modified accordingly.

P2644, L19-20 – a more detailed analysis of the sources of the differences in modelled CO and HCHO concentrations between the mechanisms, might result in an improvement in these mechanisms, allowing a more robust use of HCHO and CO columns to constrain biogenic emissions and reduce this uncertainty.

Text modified to reflect this comments Page 36.

Table 1 – The text in section 2 states that the models use internally generated emissions of lightning NO_x and ocean CO emissions. These emissions should be clearly stated somewhere either in the text or in this table, which currently implies identical CO emissions across all models.

Lightning NO_x emissions have been added in Table 1. Ocean CO emissions are the same for all models, as is now stated in the revised text.

Table 1 – While C₅H₈ and C₁₀H₁₆ have been identified within the Introduction as isoprene and monoterpenes respectively, these compounds are referred to almost exclusively by name rather than formula in the text. Please use the names here.

Done.

Figure 3 – The title of 3rd panel should read “Wollongong”

Corrected.

Figure 3 – The fourth panel would benefit from being (vertically) larger so that differences between the models can be more clearly seen.

We can see the reviewer’s point of view. However we would like to keep the plotting aspects the same for all panels. The inter-model differences are relatively smaller at this site (Darwin) compared to at other sites.

Figure 5 – Why are the observations plotted as symbols and the model output by lines?
If the observations and model output is equivalent then they should all be depicted in a consistent way (i.e. all as symbols or lines).

The symbols here are not observations but the deviations of model ensemble mean from the observations. We have now changed all plots to lines.

Figure 15 – The choice of colour scale is poor. It is particularly hard to distinguish between the different reds used for positive changes, and for the greens used for changes of -5-10% and -10-15%. Why do the panels extend north of the equator when this paper is focused entirely on the Southern Hemisphere? Ending the plots at 0S would help with the issue of the scale, as well as removing the area in northern Africa where the changes are strongly positive.

We have now changed the colour scale, addressing the reviewer’s concern. We don’t entirely agree with the reviewer’s suggestion regarding the domain of the plots for the reasons we stated before. Therefore we would like to leave the plots as they are to include the domain 0-30N.

Figure 15, Caption – What do the authors mean by “multi-annual and ensemble mean”?
Surely the plot only shows the multi-annual mean for each of the four models?

We have corrected this – thanks for pointing out.

Figure 16 – See the comments regarding the colour scale and geographical extent of Figure 15.

Colour scale has been adjusted in now Fig. 22

Figure 17 – As above.

Colour scale changed in now Fig. 23.

Figure 18 – While I realise the contours are clearly labelled this figure would again benefit from a better choice of colour scale.

Colour scales have been adjusted in now Fig.24.