

Response to Reviewer comments by Carlos Ordóñez.

First, we would like to thank the reviewer, Carlos Ordóñez, for his efforts to critically read the manuscript, and provide us with many useful comments and suggestions. Below we answer them to our best ability. The reviewer comments are in italic. Our responses are in regular font, and changes to the manuscript are given in bold.

The manuscript reports an evaluation of tropospheric reactive gases, including ozone and precursors, in three chemistry versions of the IFS: CB05BASCOE (CBA), MOZART (MOZ) and MOCAGE (MOC). The three configurations use the same meteorology, emissions, transport and deposition schemes. Therefore, the differences in the model results should be basically due to differences in the chemical schemes, namely gas phase chemistry, heterogeneous chemistry, photolysis and chemical solver.

Overall, the manuscript is well-written, the presentation quality is good and the authors give credit to related work. As motivated in the introduction, this work documents the diversity in global model simulations caused by chemistry, unlike many other papers which focus on the impact of emissions or other model parameterizations. This is a useful exercise that falls very well within the scope of the journal.

My only concern is that I miss some in depth discussions about the potential reasons for the differences in the performance of the three model configurations considering what is different in them (i.e. the chemistry). The “Main comments” below include some suggestions on how to improve this. These are not really major comments. I hope that some of them are useful, but I am aware of the complexity involved in comparing different chemistry schemes. There are also some minor and technical corrections which need attention. I will fully support the publication of the manuscript in GMD once these issues have been addressed.

Main comments

(1) Section 2.1.4 (Key differences in chemistry modules, including Table 1) provides a relatively detailed overview of the differences in the three chemical mechanisms. All this information is very valuable but not completely helpful. A simple summary comparing the complexity of the different processes in the three configurations could be useful to get a broader picture of the main differences. For instance:

1.1. Complexity of organic chemistry: MOCAGE (more extended lumping) ~ MOZART (more explicit) > CB05BASCOE

1.2. Complexity of tropospheric inorganic chemistry: MOCAGE (e.g. includes HONO) > MOZART ~ CB05BASCOE

1.3. Complexity of tropospheric heterogeneous chemistry (HO₂ & N₂O₅): CB05BASCOE ~ MOZART > MOCAGE.

Table 1 indicates that MOCAGE has as many heterogeneous reactions as the other schemes, but that information is not very helpful. I assume those reactions are only relevant for the conversion of SO₂ to H₂SO₄ in the troposphere (lines 139-140) and for stratospheric chemistry, because on lines 179-180 the authors indicate that “Heterogeneous reactions of HO₂ and N₂O₅ on aerosol ... not in the IFS(MOCAGE) version considered here”.

We agree with the reviewer that this information is useful. A concern has been how to provide such information in a compact way. We now introduce entries to describe organic and inorganic chemistry, as well as level of complexity regarding aerosol interaction (heterogeneous chemistry and photolysis). The number of heterogeneous reactions is now skipped, as this indeed refers essentially to the stratospheric chemistry, which is not subject of this manuscript.

1.4. *Is there a two-way coupling between MOZART / CB05BASCOE oxidants and CAMS aerosol fields? In other words, are the oxidants passed to the aerosol module to produce aerosols, depleted in that processes and then passed back to the chemistry schemes? Or does the coupling only consist of using CAMS aerosols for heterogeneous chemistry in the two chemistry modules? This information is not clear in the main text and is relevant e.g. to evaluate differences in NO_x and HNO₃.*

None of the current versions contain explicit parameterizations for two-way coupling with the CAMS aerosol fields; this is actually covered in more recent versions of the modeling system. Aerosol fields are currently only used for defining the surface area density needed in the heterogeneous chemistry. We now write

Also two-way coupling of secondary aerosol formation was not available in the current model version.

1.5. *Complexity of photolysis: Is it more complex in CB05BASCOE because it includes the impact of aerosols? Anyway, if there is no easy way to compare complexity of photolysis among the three configurations one can leave this out.*

We now specify the aerosol impact on photolysis. Other aspects are more difficult to specify in the table.

1.6 *Complexity of stratospheric chemistry: CB05BASCOE > MOZART ~ MOCAGE*

We choose to focus on tropospheric chemistry aspects mainly in the table, so we prefer to leave out a comment on this.

1.7 *Chemical solver: CB05BASCOE uses Rosenbrock (differences in NO_y chemistry and up to ~20% less O₃ regionally). It would be interesting to know if the “differences in the N₂O₅ chemistry” indicated by the authors (line 199) often result in smaller amounts of the species evaluated here (NO₂ and HNO₃). If not clear, please leave this out.*

This is correct: The introduction of the more accurate Rosenbrock solver was found to lead to a reduction in NO₂ and HNO₃, associated to a faster N₂O₅ production, together with an equally stronger N₂O₅ heterogeneous loss on aerosol particles. However, as this process is indeed connected to application of heterogeneous chemistry, this effect is only relevant to CB05BASCOE (and MOZART), and not to MOCAGE chemistry. We now describe these processes more explicitly:

These differences are mostly traced to an increase in the N₂O₅ chemical production (Cariolle et al., 2017), reducing in turn the NO_x lifetime because of a larger net N₂O₅ loss on aerosol. This in turn leads to a reduced chemical ozone production efficiency.

Summarising, although I understand how hard it is to compare different chemistry schemes, a new table or some bullet points similar to those indicated above could be used to better see the main differences in the three model configurations. Afterwards, one could elaborate on the potential implications for the representation of the main tracers evaluated here (see next comment).

We expanded the table and its description in response of the reviewer comments as described above.

(2) Some additional comments in different parts of the text could be helpful to try to understand the model differences. As mentioned above, I do not expect the authors to be able to explain absolutely everything, but hope that some of these comments will be useful and easy to include in the text:

2.1 Section 4 (Assessment of inter-model differences). The comparison of NO_x annual mixing ratios (Figure 2, lines 330-333) could easily be improved:

“MOZ and CBA are overall similar, but MOC is showing higher values over the NH high-latitudes (>60°N) and also at altitudes below 900 hPa in the tropics”. First, a minor correction: I clearly see differences in the lower to mid-troposphere from the tropics to the high latitudes, i.e. not only for latitudes >60°N and below 900 hPa in the tropics.

We agree with the reviewer that this is actually better, we modify the text accordingly.

“This is likely related to the fact that in this version of IFS(MOCAGE) the coupling with the aerosol module has not yet been established, contrary to CBA and MOZ”. Do you mean that the lack of heterogeneous hydrolysis of N₂O₅ in MOC will probably raise the NO_x concentrations (because that reaction produces the reservoir species HNO₃, which is efficiently removed from the atmosphere by dry and wet deposition, resulting in a removal of NO_x)? If so, please clearly indicate this. Otherwise, you should explain which other processes might be involved. For instance, is there aerosol nitrate production, removing NO_y from the gas phase, in CBA and MOZ?

Yes, here we indeed exactly refer to the differences in heterogeneous reaction approaches, as we now more explicitly describe in the table. We now write this explicitly:

“This is likely related to the fact that in this version of IFS(MOCAGE) the coupling with the aerosol module has not yet been established, contrary to CBA and MOZ, **implying a missing sink of NO_x through the heterogeneous reaction of N₂O₅ to HNO₃.**”

2.2 The last paragraph of Section 4 (Assessment of inter-model differences) tries to relate the differences in NO_x-O₃-OH in the models. Overall, the authors point to faster chemical production of O₃ in MOCAGE, counterbalanced by fast loss by reaction with OH and HO₂. They also suggest a relatively short lifetime of O₃ in MOCAGE and conclude the paragraph with “Such differences in oxidation capacity naturally have important implications for understanding differences in the performance of NMHCs, as discussed in the next sections”. A couple of questions:

* The authors are probably right, but does the C-IFS have any diagnostics that can be used to quantitatively compare budgets/lifetimes of O₃ due to different processes in the three configurations? That would be useful.

Unfortunately such diagnostics, keeping track of chemical ozone and carbon monoxide production and loss budgets, are only available for CB05BASCOE chemistry, and not for the other chemistry schemes, hence we cannot easily intercompare them here. It is a good suggestion to actually implement such diagnostics for the other chemistry versions as well. We now include the following comment:

An assessment of the ozone chemical production and loss terms is beyond the scope of this work.

** That paragraph should also include a comment for CO. A few lines above (lines 326-328), the authors compare CO in the three configurations (CBA >> MOZ > MOC), “suggesting mainly differences in oxidizing capacity”. I think they could be more specific. I am aware that there is no straightforward comparison of the CO and OH levels in the models, because different VOC degradation schemes will impact the CO mixing ratios (as correctly indicated on first paragraph of Section 5.2). Nevertheless, the authors could relate the higher (lower) CO levels in CBA (MOC) to the lower (higher) OH levels seen for the same models in Figure 2. I will come back to this below (comment on Section 5.2).*

The reviewer is correct that a large portion of the spread in CO is likely caused by differences in OH, considering that CO and precursor emissions are mostly identical. Unfortunately we do not track the CO chemical production and loss budgets, so we cannot quantify this easily. Still, in the section describing CO we now point explicitly to the impact of OH, and refer to its discussion:

As CO and precursor emissions are essentially identical, this is likely caused by differences in oxidizing capacity which is governed by OH abundance, as described below.

2.3. Section 5.1 provides a detailed evaluation of ozone. However, after so many comparisons, it is a bit hard to conclude what we have learnt. A summary of the model performance, even if the reasons for the model differences need to be better investigated, might help.

For instance, I know there are some specific comments for MOZ in the first paragraph of the section (e.g. “MOZ simulates too high O3 concentrations ... Here it is worth mentioning that recent updates to reaction probabilities and aerosol radius assumptions in the heterogeneous chemistry module in IFS(MOZART) significantly improved O3 concentrations particularly in the NH”), but a summary mentioning that at the end would be useful.

Some conclusions might be that modelled ozone is often within the uncertainty of the selected sondes (Figure 6), but that MOZ tends to yield higher O3 levels everywhere in the mid-to-upper troposphere (500-300 hPa) and everywhere in the northern hemisphere (Figure 5). Somewhere one should also mention that this configuration yields the highest biases (15.9 ppb), but at the same time the best weighted correlations, when compared to aircraft data (Table 4). Then, one could say again that the biases have partly been corrected in recent versions with improved heterogeneous chemistry.

Finally, the authors could try to repeat Figure 5 but showing accumulated errors (e.g. RMSE) instead of biases. If that provides any further information to document the performance of the three configurations then it should be included.

We now include figures on the RMSE within Figure 5, and link this to a summary section to the discussion of the ozone evaluation. For this purpose we have slightly re-ordered the section, to describe the findings from Figure 5 (new Figure 7) at the end of the section. Also, as per the second reviewer comments, we include an evaluation against the CAMS Interim Reanalysis, to put this evaluation into further perspective. We now write:

Evaluation against the aircraft climatology as provided in Table 4 shows on average a positive bias in the range 10 (CBA and MOC) to 16 (MOZ) ppbv, while the correlation statistics shows generally acceptable values ($R^2 > 0.57$), giving overall confidence in the model ability to describe ozone variability. Figure 6 shows annually averaged model biases and root mean square errors (RMSE) for various latitude bands and for altitude ranges 900-700hPa, 700-500hPa and 500-300hPa against WOUDC sondes. In this

evaluation we also present data from the CAMS Interim Reanalysis (CAMSiRA) for the year 2011, to put the current model evaluation into perspective. This summary analysis shows averaged biases within ± 10 ppbv, which is also in line with the O₃ bias statistics against the aircraft climatology. At lower altitudes the model biases are mostly equal or better than those from CAMSiRA, while above 500 hPa CAMSiRA delivers mostly smaller biases thanks to the assimilation of satellite ozone observations. The RMSE shows a larger spread in the lower troposphere of the NH, while at higher altitudes, above 500 hPa the overall magnitude of the RMSE for the three chemistry versions converges to values ranging from 10 to 16 ppbv, depending on the latitude. Here the CAMSiRA shows overall better performance, mainly for the tropics and SH, while over the NH its performance is similar to IFS(CBA). This evaluation summarizes common discrepancies between model versions and observations, such as the negative bias over the Antarctic and positive bias below 700 hPa for tropical stations (see also Figure 4), suggesting biases in common parameterizations such as transport, emissions and deposition. The largest discrepancies between model versions have been detected at northern mid- and high latitudes below 500 hPa, with significantly higher values for RMSE for MOC and MOZ compared to CBA. A comparatively large positive bias for MOZ was detected, which has been linked to an under-estimate of the N₂O₅ heterogeneous loss efficiency. The differences between MOC and CBA can likely be explained by similar aspects are likely as important to explain differences with respect to the performance of IFS(MOCAGE).

2.4. Section 5.2 (Carbon monoxide).

The authors include a useful introduction about all the processes impacting CO levels (emissions and VOC degradation) and then comment several plots comparing models with different datasets (MOPITT, surface, aircraft profiles). Overall, I am rather happy with this section, but again it would be nice to finish it with a paragraph summarising what we have learnt from these comparisons.

For instance, from Figures 7-9 (MOPITT & surface CO) one can see that CBA is the configuration with the highest CO levels, yielding the best comparisons (lowest negative biases) in the northern hemisphere but positive biases in the southern hemisphere (Figure 9). The authors should mention that this may partly be related to the fact that this scheme is the one with the lowest OH levels (Figure 2). I would say this is one of the main outcomes from these comparisons.

Towards the end of this section the authors could indicate whether modelled CO is within a given range for the three schemes (they mention 10% only for the comparisons with aircraft observations, but in the abstract and conclusions they indicate 20%). That range (expressed as a percentage or ppb) could be compared to the overall model error for this species (when compared e.g. with surface observations) to get an idea about the fraction of that error that could be attributed to differences in the chemistry. Here or in the "Conclusions" section one could even point to the need of producing some model diagnostics in the future to understand to what extent the differences in the atmospheric mixing ratios of CO arise from differences in CO degradation (by reaction with OH) or differences in CO production (from VOC degradation) in the chemistry schemes.

In response to the reviewer we now make the relevance of OH for explaining the differences in CO clearer. The discrepancy in reporting of the 10% to 20% spread in CO at different parts in the manuscript is because of the different means to quantify this. The 20% number originates from assessment of the zonal mean CO fields (Sect. 4), while the 10% specifically relates to the model CO profiles presented in the evaluation against specific aircraft observations (Sect. 5.2). Closer assessment of the evaluation of vertical profiles shows that this is rather in the range 10-20%, depending on the location. We now modify the text accordingly.

2.5. Section 5.3 (formaldehyde). It is nice that the authors use weighted biases to show that the simulations are within the range of uncertainty of the few observations available. The correlations are also quite decent. However, this section is rather short and could probably benefit from some improvements and/or discussions.

Lines 506-508 say: "Formaldehyde is important as the most ubiquitous carbonyl compounds in the atmosphere (Fortems-Cheiney et al., 2012). It is mainly formed through the oxidation of methane, isoprene and other VOC's such as methanol (Jacob et al., 2005), while its oxidation and photolysis is responsible for about half of the source of CO in the atmosphere".

First, a couple of minor comments: You should mention that HCHO is not only secondarily formed but also directly emitted. In the first sentence, do you mean "one the most ubiquitous carbonyl compounds"?

I am not sure that the end of the second sentence is right. There the authors mention that the degradation of HCHO is half the source of CO in the atmosphere. Somewhere else (lines 448-449) they mention that "Approximately half of the CO burden is directly emitted, and the rest formed through degradation of methane and other VOCs". Are the authors implying that most of the CO formed (not emitted) in the atmosphere comes from HCHO? Isn't that too much? Before making that statement, the authors should check some references or compare the production terms of CO in the chemistry schemes. If HCHO is so important in terms of CO production, could one try to relate the model performance for both species? For instance, it is possible to relate the fact that CBA is the simulation with the highest CO mixing ratios (which I previously attributed to the low OH in that model) to the negative HCHO biases found for the same scheme in this section?

In response to the reviewer, first of all, we should indeed have written "**one of the most ubiquitous carbonyl compounds**"; we do this now.

The reviewer is in principle correct that CH₂O is also emitted, with an amount of ~13 Tg yr⁻¹, see Table 2 in the manuscript. However, these direct emissions are negligible compared to the secondary production through VOC oxidation: from the chemistry budget analysis as available in the CB05BASCOE scheme this amount is estimated to be ~ 1550 Tg yr⁻¹, with contributions indeed from methane, isoprene, methanol, and many other VOC's. CH₂O degradation through photolysis and oxidation in turn corresponds to approx. 1300 Tg CO production. The remainder of secondary CO production from other VOC oxidation is estimated about 250 Tg, while direct CO emissions are 1040 Tg, which explains our statement that CH₂O degradation is responsible for about half the source of CO.

However, the amount of modeled chemical CH₂O production is indeed depending on the chemistry mechanism. Note additionally that the OH impacts both on CH₂O production, and its loss. Therefore, evaluation of the CH₂O in order to understand CO biases, is not trivial. As any explanation is essentially speculation we refrain from further discussion. Indeed, this requires information on the chemical production and loss budget terms, which should be made available for all chemistry versions. The same argument holds for CH₃OOH, as further discussed in a reviewer comment below. We now include the following comment:

Considering the short lifetimes for CH₂O (a few hours in daytime) and also CH₃OOH, and the large dependence of their abundances on details of the VOC degradation scheme which vary across the chemistry versions presented here, it is beyond the scope of this manuscript to explain these differences. This would require a detailed assessment of the respective production and loss budgets which are currently not available.

2.6. Section 5.4 (Ethene, C₂H₄).

The authors have mentioned the main sources of this species and have provided some insights into the reasons for the model performance. Most of this section looks good to me, but I have some comments:

* The last paragraph says: “Furthermore, interesting is the comparatively large difference present between the simulations at high latitudes (e.g. SONEX, Newfoundland), where the largest relative differences in modelled OH have been found, (see also Sec. 4), illustrating the importance of OH for explaining inter-model differences”. Could one explicitly mention that the higher C₂H₄ mixing ratios in CBA (red in Figure 13, right) compared to the other two schemes are probably related to the lower OH levels in the same scheme?

We now explicitly mention this correlation between the relatively high C₂H₄ in relation to the low OH:

CBA indeed shows the largest values for C₂H₄, which is correlated to the comparatively low magnitude of OH, which explains the longer lifetime for C₂H₄.

On the other hand, I am surprised that CBA is the scheme with the largest negative biases for C₂H₄ in Table 4 (-6.28 ppt). It does not look like that from Figure 13. Is that because the figure only shows data for some specific campaigns and locations? If so, my previous comment might not be valid.

In fact the reviewer has spotted a typo when generating Table 4. The reported overall model bias should not have been negative but positive by 6.28 ppt for CBA. This is corrected now, which makes the evaluation more consistent. Indeed, the reviewer comment still holds that here we show figures for a few locations; for other sites the magnitude of the biases vary to some extent.

* Lines 531-532: “As the underestimation appears to be ubiquitously distributed this suggests that C₂H₄ decomposition is too strong, or that the model versions miss chemical production term (e.g., Sander et al., 2018)”. This conclusion looks fine, but they should change “miss chemical production term” to “miss some chemical production terms”.

The reviewer is correct, we change accordingly.

Somewhere in this section it would be worthwhile to briefly mention (without all the explanations given below) that one should not expect to get this species right because model and observations are not really collocated in time. I understand that model data are monthly averages (considering both daytime and night-time data) while observations were probably collected at daylight. In addition, model and aircraft observations are from different years. All this is relevant if we keep in mind all the processes affecting the mixing ratios of this species.

The reviewer is correct that no exact correspondence should be expected for the evaluation against the aircraft observations, particularly considering the different model year. This has been discussed in Sec. 3.1. To better explain the procedure for the model evaluation, we now add in Sec. 3.1 the following comment:

Aircraft measurements of trace gas composition from a database produced by Emmons et al. (2000) were used for evaluation **of distributions of collocated monthly mean modelled fields.**

While in Sec. 5.4 we add the comment:

Even though this evaluation should only be considered in a climatological sense, ...

Many of those processes have already been mentioned by the authors: terrestrial biogenic emissions (probably with strong diurnal cycle), which should be much larger in Newfoundland (Figure 13 right, high C₂H₄ mixing ratios) than in Christmas Islands and Tahiti (oceanic environment, Figure 13 left and centre, low C₂H₄); diurnal cycle of anthropogenic emissions (again source probably much higher in Newfoundland); biomass burning sources (with strong interannual variability); upward lifting and subsequent horizontal transport of plumes (not sure if seen on observations from Figure 13 left for Christmas Islands), and relatively short lifetime, mainly controlled by its reaction with OH (higher levels at daylight than at night-time).

The reviewer is correct, but for reasons of brevity we prefer not to include such a listing here.

2.7 Section 5.6 (Nitrogen dioxide, NO₂).

Overall, this section is fine too, but I miss more detailed explanations and some integration with the rest of the text towards the end of the section. In addition, providing maps with biases would be helpful to see the differences between the model configurations and OMI (this is also indicated in a minor comment).

We now include bias plots instead of actual model fields for tropospheric NO₂ columns, to better assess the differences with respect to the retrievals.

For instance, around lines 582-586, the authors mention the lower NO₂ levels for CBA than for the other two models, over polluted and clean regions in the northern hemisphere. Then, around the same lines they say: "This all suggests a relatively short NO_x lifetime in CBA compared to MOZ and MOC, which in turn helps to explain the lower O₃ over the NH-mid latitude regions as modelled with CBA (see Figure 5)". A couple of comments on this:

** In Section 2.1.4 the authors mentioned some implications of the use of the Rosenbrock solver (in CBA) for NO_x and O₃, but I think that is hardly mentioned again in the rest of the text. Shouldn't one mention it here?*

The reviewer is correct that the use of different solvers is an important aspect to explain the differences, particularly for IFS(MOCAGE) but it is not the only explanation. Also the N₂O₅ heterogeneous chemistry, missing in this version of IFS(MOCAGE) and not optimized in IFS(MOZART) contributes to various degree. To point this out, we now include the following comment:

The causes of these differences in modelled NO₂ are mainly the use of a different numerical solver and differences in the efficiency assumed for N₂O₅ heterogeneous reactions (see Sec. 2.1.4).

** CBA has less NO_x, O₃ and OH, but more CO, than the other model configurations. Not sure if all these facts are related. Shouldn't this be mentioned somewhere?*

These aspects are certainly related. This has been discussed in Sec. 4, describing an overview of the main differences. We prefer not to describe such correlations here again – the subsections in Sec. 5 are mostly focused on the quality of individual species.

2.8. HNO₃ (also related to NO_x).

There is a good introduction relevant to HNO₃ around lines 601-604: “It has a very high solubility and therefore tends to be scavenged by precipitation very efficiently, providing an effective sink for the NO_x family. Furthermore, it can act as a precursor for nitrate aerosols (Bian et al., 2017). HNO₃ concentrations are therefore expected to show amongst the largest variation between the simulations, as the production and the sink terms can largely differ due to uncertainties in the parameterizations”. However, later on, the authors are not very specific about the potential reasons for the model differences they find. They just add some general sentences like these ones:

Lines 612-613 (for HNO₃): “The discrepancies between the model versions can be mainly attributed to differences in NO_x lifetimes and nitrate aerosol formation”

Lines 653-655 (in the conclusions): “For instance IFS(MOCAGE) tends to predict significantly higher NO_x and HNO₃ concentrations in the lower troposphere compared to the other two chemistry versions”.

Can one relate some of the model differences to the lack of coupling of MOCAGE chemistry with the aerosol module? I assume this is not simple. For instance, the inclusion of heterogeneous loss of N₂O₅ in CBA and MOZ should remove NO_x from the system but form HNO₃, which would not explain the high HNO₃ in MOC. On the other hand, if there is a two-way coupling between the chemistry and the aerosol modules, CBA and MOZ will provide oxidants (e.g. HNO₃) to the aerosol model to form aerosols (e.g. nitrate) and the oxidants will be depleted in that process. This could counteract the production of HNO₃ by heterogeneous reactions. Are all these processes taking place in CBA and MOZ but not in MOC? And is wet deposition done exactly in the same way for the three configurations? Otherwise, it should have an impact on the HNO₃ mixing ratios.

In response to the reviewer, we confirm that interpretation of differences in HNO₃ is not straight forward, exactly because it is a secondary product. We note that dry and wet deposition parameterizations, and loss rates, are identical across the three versions.

Aspects that may play a role are in leading to differences are, once again, a larger HNO₃ production efficiency in CBA and MOZ, compared to MOC, which currently misses the N₂O₅ heterogeneous loss reaction. As this process mostly takes place near the surface, any HNO₃ production here consequently may lead to a more efficient dry (and also wet) deposition of HNO₃, as compared to HNO₃ produced by NO₂ oxidation at more remote locations in the free troposphere.

Also the versions include different parameterizations for secondary aerosol formation (nitrate and ammonium), for which we now provide details in Sec. 2.1.1 for IFS(CB05BASCOE) and 2.1.3 IFS(MOZART). Note that the respective chemistry versions contain separate tracers to describe the nitrate aerosol; they are not yet part of the CAMS aerosol module.

However, as it is speculation as to what are the relative contributions of the various mechanisms to the observed model spread, we limit ourselves by adding the following statement to this section:

“The discrepancies between the model versions can be mainly attributed to differences in NO_x lifetimes, associated to differences in heterogeneous chemistry, and parameterizations for nitrate aerosol formation, as discussed in Sec. 2.1.4.”

While in Sec 2.1.4 we now describe in more detail the differences in nitrate aerosol formation modeling and heterogeneous chemistry:

“Instead, IFS(CB05BASCOE) and IFS(MOZART) contain a treatment of gas-aerosol partitioning for nitrate and ammonium, which is missing in IFS(MOCAGE).

Significant uncertainty remains in the magnitude of heterogeneous reaction probabilities. Heterogeneous reactions of HO₂ and N₂O₅ on aerosol are included in IFS(CB05BASCOE) and IFS(MOZART), **although with different efficiencies**, but not in the IFS(MOCAGE) version considered here. This has only become available in a more recent model version. **Also two-way coupling of secondary aerosol formation was not available in any of the current model versions.**

Minor comments

** line 22: “compared against a range of aircraft field campaigns, ozone sondes and satellite observations”. Surface observations should also be mentioned here (even if only used for CO).*

Done

** Lines 25-26: “Most of the divergence in the magnitude of NMHCs can be explained by differences in OH concentrations”. Why only NMHCs? Shouldn’t one mention CO too?*

The reviewer is correct: CO should be mentioned here as well.

** Chemical names, formulae and abbreviations are not used in a consistent way throughout the text. There are some examples here, but this list is not exhaustive at all (the authors should revise the whole text):*

“PAN” should be changed to “peroxyacetyl nitrate (PAN)” on line 135.

Done

“polar stratospheric clouds (PSC)” needs to be mentioned only once (for B05BASCOE). In the description of the other two configurations the authors can simply use “PSC”.

Done

It should be “VOCs” and not “VOC’s”. The same applies to “NMHC’s” and “CFC’s”. Please, also spell out VOC and CFC.

In particular, special attention is required in Section 2.2 (Emissions). There the authors use capitals for names such as “Olefins”, “Aldehydes”, “Butenes” or “Parafins”, but that is not needed.

The reason for using capitals was as these refer to actual tracers in the CB05 mechanism. Instead, we now put brackets around these tracer names.

The names “methane” and “CH₄” are used in different parts of the text, but the authors never indicate “methane (CH₄)”. All names of chemical species should be given together with their formulae the first time they appear in the text.

Done

They authors use the name “acetylene”, while the IUPAC name (ethyne, C₂H₂) might be more appropriate.

Done

When possible they should indicate both the name and the formula for all the species in Table 2 (otherwise, please give both name and formula in the main text). There should be a reference to this table from the very beginning of section 2.2.

Done

** Section 2.2. Please indicate whether the non-anthropogenic emissions (e.g. biogenic VOCs and NO, NO from lightning) are prescribed or interactive.*

All emissions, including biogenic and soil NO_x, are prescribed. The only exception is lightning NO_x emission, which is parameterized depending on convection. We now make this more explicit.

** The authors have indicated the experiment IDs in Table 3. This and other information they provide will be very useful to ensure traceability. I have also noted that in the “Data Availability” section they say: “The model simulation data as used in this work can be obtained upon request from the corresponding author”. Are the model experiments from Table 3 stored on ECMWF’s servers and accessible to other users? If that is the case, the data availability section could be rewritten.*

Data can be retrieved by other users who have access to ECWMF MARS archiving system, indeed using the experiment ID’s. Users are welcome to download the data from there. However, as it is not fully straight forward to identify the actual composition fields, it is recommended that potential users contact one of the authors and/or the corresponding author, who can provide further support. We provide such details in the Data Availability section now.

** In Section 2.3 (Model configuration and meteorology) the authors mention that “the model was sampled ... every three hours. These are used to compute monthly to yearly averages. Standard deviations are computed to represent the model variability for a specified range in time and space”. This is fine but, after reading section 3 (Observational datasets), it is not clear how the model output and observations are sampled for many of the comparisons shown in the text. I assume that the data are always collocated in space but not completely collocated in time. For instance, aircraft observations are from different years. Please indicate for all datasets whether they are for the simulated year (i.e. 2011) or for another period. Then indicate whether all comparisons are made using model and observations from the same month (even if the data are from different years).*

We now specify more explicitly the sampling approach for the various observations. For evaluation against the aircraft measurements data for 1990-2001 were used. For this

comparison the model results were co-located spatially and temporally with the observations, although for a different year. Monthly mean surface CO observations, as provided by the World Data Center for Greenhouse Gases (WDCGG) for the year 2011 were used while for the other evaluations (ozone sondes, CO and NO₂ satellite retrievals) we use a different collocation procedure as already described in the manuscript. We now write:

Aircraft measurements of trace gas composition from a database produced by Emmons et al. (2000) were used for evaluation **of distributions of collocated monthly mean modelled fields**.

(...) The database is formed by data from a number of aircraft campaigns **that took place during 1990-2001, ...**

In-situ observations for monthly mean CO for the year 2011 are used **to evaluate monthly mean modelled surface CO fields. Observational data is taken from the World Data Centre for Greenhouse Gases (WDCGG), the data repository and archive for greenhouse and related gases** of the World Meteorological Organisation's (WMO) Global Atmosphere Watch (GAW) programme.

(...)

The 3-hourly output of the three model versions has been collocated to match to the location and launching time of the individual sonde observations during 2011.

** Section 3.1 (Aircraft measurements). The authors used the compilation of Emmons et al. (2000) as well as TOPSE and TRACE-P data. If it is not too complicated, they could also indicate the names of all the campaigns used, the period covered and so on. In Figure 1 they indicate that "Each field campaign is represented by a different colour" but they do not mention the names. Later, some of those names appear on the profile plots (Figure 11, 12, ...). I simply think it would be helpful to know the geographical location / time of the year for the comparisons between model and observation. This would also be great to ensure traceability.*

This is a good suggestion. In Figure 1 we now provide the names of the campaigns in the map presenting the location of the aircraft data, which helps to identify the location of the evaluations.

** Section 3.2 (In situ observations). I wonder myself whether "Near-surface CO and ozonesondes" could be a more appropriate name, because I assume that most of the aircraft measurements in the previous section will also be in situ.*

The reviewer is actually correct. We change this now.

** Table 4 (summary of bias and correlation coefficients of models vs. aircraft observations) is a bit confusing: The authors first mention correlation coefficient (R) but then they indicate R-squared both in the caption and the header. Please be consistent.*

We change this consistently to R² in the header, see also the second reviewer comment

"Bias and R2a is in pmol/mol, (except for CO and O3)". R or R2 cannot have units. In addition, it might be simpler to write "Bias in pmol/mol (nmol/mol for CO and O3)" and remove the note below the table.

We follow the suggestions from the reviewer, and now write in the table header:

Bias^a is given in pmol/mol, (nmol/mol for CO and O₃), Bias^b is in standard deviation units. Likewise, R^{2a} is the normal correlation coefficient, and R^{2b} the correlation coefficient weighted with standard deviations (see text).

The authors show methyl hydroperoxide (CH₃OOH) in the table because there are some aircraft observations, but this species is hardly discussed in the text. Why is that? I think there is only one comment about this species around lines 372-374 (“... while model versions have more difficulties with CH₃OOH”), but it is a little bit out of place there. The negative biases can be high for this species but the correlations are fine.

We now add a small section describing the CH₃OOH evaluation within the section describing CH₂O. In short, the reviewer is correct that the model versions are in reasonable shape to represent the variability in CH₃OOH, reflected by good correlations. Only IFS(MOZART) is somewhat on the low side with a comparatively large negative bias. Unfortunately as with CH₂O, it is very difficult to provide additional interpretation to these model differences, considering that this species, as CH₂O, is relatively short lived (governed by OH and photolysis mainly), while its production rate much depending on details of the VOC degradation scheme:

CH₃OOH is a main organic peroxide acting as a temporary reservoir of oxidizing radicals, Zhang et al. (2012). It is mainly formed through reaction of CH₃O₂ + HO₂, which are both produced in the oxidation process of many hydrocarbons. The CH₃OOH lifetime of globally about one day is mainly governed by its reaction with OH, and photolysis. Figure 13 presents an evaluation for CH₃OOH for the same sites are presented for CH₂O in Figure 12. Mixing ratios are generally reasonably within the range of the observations, as for example over the tropical Pacific over Fiji. A larger spread between model versions, with a strong over-estimate for CBA, is found in the Amazon region over Brazil. As a global average, a comparatively large under-estimate for MOZ and, to a lesser extent also for CBA, was found, see also Table 4. Nevertheless, correlations, especially those weighted with the uncertainties, are overall good, giving general confidence in the modeling.

Considering the short lifetimes for CH₂O (a few hours in daytime) and also CH₃OOH, and the large dependence of their abundances on details of the VOC degradation scheme which vary across the chemistry versions presented here, it is beyond the scope of this manuscript to explain these differences. This would require a detailed assessment of the respective production and loss budgets which are currently not available.

The concept of weighted biases and correlations has been introduced on page 15, but shouldn't the weighted vs. non-weighted values be discussed in some parts of the text? For instance, in the case of C₂H₄, the three simulations have similar values of R (0.54-0.58) but rather different values for weighted R (0.03-0.39); why is that? In other cases, such as for CO and C₂H₆, the values of weighted R are higher than those of R. So the weighting can improve or deteriorate the correlations. Summarising, it might be good to explain why the values of R sometimes change so much, not always in the same direction, after weighting.

Generally higher correlations can be expected when accounting for the uncertainties, quantified by the standard deviation in the model and observations. In fact, for cases of high variability in either the model or observations, a poor matching between the observations and model can be expected, which causes poor correlation. In the weighted correlation, as these situations are weighted less, an increase in the value for R^2 is expected. Part of this argumentation is already discussed when introducing this statistics, in Sec 5.

An exception is the C_2H_4 , which does not perform well across the chemistry versions with rather different profile shapes compared to the observations. In this case, the weighting with variability apparently does not help, and in fact only leads to further degradation. This indicates more fundamental problems with describing the spatial and temporal variability for species properly, as further discussed in Sec. 5.4.

We now expand on the discussion of normal and weighted R^2 and biases by writing in Sec. 5:

For this reason the weighted correlations are also generally expected to be higher than the normal correlations.

And:

Remarkably, C_2H_4 is the only trace gas where values for the weighted R^2 are lower than the normal R^2 values, suggesting fundamental problems representing this trace gas properly in any of the chemistry versions.

** Around lines 395-397 the authors say “MOZ simulates too high O3 concentrations ... Here it is worth mentioning that recent updates to reaction probabilities and aerosol radius assumptions in the heterogeneous chemistry module in IFS(MOZART) significantly improved O3 concentrations particularly in the NH”.*

It might be worth to mention whether the assumptions made in the heterogeneous chemistry are very different in MOZART and CB05BASCOE. Are the uptake coefficients very different in both configurations? Are they constant or parameterized? And which are those aerosol radius assumptions, considering that both configurations use the same CAM aerosol fields?

By the way, the changes mentioned for O3 in the NH, if they are partly consequence of updates in the heterogeneous uptake of N2O5, might have some seasonality and not be so relevant all year round (see e.g. Tie et al., 2003, which used a very simple parameterization of that process). Or do the “recent updates to reaction probabilities” refer to reactions in the gas phase?

Tie, X., L. Emmons, L. Horowitz, G. Brasseur, B. Ridley, E. Atlas, C. Stround, P. Hess, A. Klonecki, S. Madronich, R. Talbot, and J. Dibb: Effect of sulfate aerosol on tropospheric NOx and ozone budgets: Model simulations and TOPSE evidence, J. Geophys. Res., 108(D4), 8364, doi:10.1029/2001JD001508, 2003.

The rate of N_2O_5 uptake on aerosols, which depends on aerosol composition and meteorological conditions, remains highly uncertain. In order to investigate the impact of such uncertainties on simulated ambient N_2O_5 and hence on O_3 in MOZ, sensitivity simulations have been recently conducted using reaction probability and aerosol parameters for SO_4 , OC and BC from MOZART-4 (Emmons et al., 2010). The aerosol parameters and reaction probability applied show significant differences as compared to those in CB05BASCOE (e.g. γ values of 0.02 and 0.1 for uptake on SO_4 in CB05BASCOE and MOZART-4, respectively). Both MOZ and

CB05BASCOE use constant values for reaction probabilities and mean particle radius while hygroscopic growth rate factors at different RH were applied according to Chin et al. (2002).

The sensitivity simulations showed significant decreases in N_2O_5 concentrations due to heterogeneous reactions on aerosols and important decreases in NO_x which led to lower O_3 in MOZ particularly in the NH. The changes in O_3 and the overall NO_x budget in the NH are found to be more important in winter and spring time periods, consistent with results of Tie et al. (2003). We now include the following comment in the manuscript, sec. 2.1.4:

Significant uncertainty remains in the magnitude of heterogeneous reaction probabilities. Heterogeneous reactions of HO_2 and N_2O_5 on aerosol are included in IFS(CB05BASCOE) and IFS(MOZART), **although with different efficiencies**, but not in the IFS(MOCAGE) version considered here. This has only become available in a more recent model version. **Also, for instance, a more recent version of IFS(MOZ) with updated values following Emmons et al. (2010) leads to a significantly reduced NO_x lifetime.**

** Figure 4 and Lines 405-406 (comparison with ozonesondes): "However, it should be noted that in the SH regions this evaluation is less representative because there are very few observations".*

I see that the number of sites within the different latitude ranges is shown later (Figure 5), but some additional information about the geographical data coverage for this and other datasets would be useful (e.g. in Section 3, Observational datasets). The authors should also indicate if the ozonesonde dataset is for 2011 or corresponds to the climatology for a longer period. And why not including a map (in a supplement if you prefer) indicating the location of all ozonesonde sites, with a different colour to highlight those shown in Figure 6?

By the way, this sentence in the caption of Figure 4:

"Tropospheric ozone profiles of volume mixing ratios (ppbv) against by model versions CBA (red), MOC (blue) and MOZ (green) against sondes (black) over six different regions" should be rewritten to something like:

"Tropospheric ozone profiles (volume mixing ratios, ppbv) from model versions CBA (red), MOC (blue) and MOZ (green) as well as sondes (black) over six different regions"

Ozone data used here is indeed for the year 2011, as we now specify in Sec 3.2. We now also include a figure presenting the location of ozone sondes used in the evaluation.

Also we have adapted the legend in Figures 4 and 5 according to the reviewer suggestion.

** Figure 5. Please draw a line at 0 to better distinguish positive and negative biases. I would also add a symbol to better identify the mean bias within each latitudinal band.*

We now draw a line at zero, to better identify the positive/negative biases.

** Lines 454-455 (for modelled O_3 vs. MOPITT CO , Figures 7 and 8): "low bias over the NH during April (further analysed in, e.g., Stein et al., 2014), as well as a low bias over Eurasia during August". Need to change "low bias" to "negative bias". In addition, I would say that in August the negative biases occur not only over Eurasia, but over most continental areas of the northern hemisphere.*

The authors could also be a bit more quantitative and provide spatial correlations between modelled and observed CO for each of the maps on Figures 7 and 8. They could even show maps with biases in a supplement if some of the biases I mention are not so clear.

As for NO₂, we now present observed CO tropospheric columns together with the model biases, to better diagnose the magnitude of respective biases. We do not believe that the spatial correlations provide added value that helps interpretation of the results. We change 'low bias' to 'negative bias', thank you for spotting this.

By the way, apart from Stein et al. (2014), the authors could also cite this previous paper regarding the low CO bias over the NH:

Shindel et al., Multimodel simulations of carbon monoxide: Comparison with observations and projected near-future changes, 2006, <https://doi.org/10.1029/2006JD007100>.

We now include the Shindel et al., reference here.

** Lines 563-565: This sentence is a bit awkward: "Considering its strong diurnal cycle, due to the fast photolysis rate, here only daytime values have been used to construct the averages, because the observations from the various field campaigns were equally conducted in daylight conditions".*

It would be convenient to split all this information into two separate sentences. Moreover, the authors should mention not only the fast photolysis rate but also the strong diurnal cycle of anthropogenic emissions.

We follow the reviewer suggestion to split the sentence in two. As for the diurnal cycle in emissions: this is not fully straight forward. The current model versions do not yet include such a diurnal cycle for anthropogenic emissions, even though literature suggests ~2x larger emissions during daytime compared to night-time, see, e.g., Miyazaki et al., (2017). Therefore, we choose not to refer explicitly to the diurnal cycle in the emissions, as this may raise confusion at this point. Also the aircraft observations mostly sample more background conditions, less affected by direct emissions.

** Lines 578-582: "Figures 16 and 17 evaluate tropospheric NO₂ using the OMI satellite observations. The simulations deliver generally appropriate distributions ... Another interesting finding is a relatively strong negative bias over ...". It is a bit hard to see the biases in the figures. It would be helpful to show analogous maps with those biases (in a supplement).*

We now present bias maps instead of actual model columns.

** Line 600: "Compared to the trace gases previously analysed, Nitric acid is not primary emitted but is purely photochemically formed". Reformulate this sentence, because O₃ is not primarily emitted. Also, change "Nitric" to "nitric".*

Thank you for spotting this. we now change the sentence to:

Compared to several of the trace gases previously analysed, nitric acid is not primary emitted but is purely photochemically formed in the atmosphere.

Technical corrections

All technical corrections have been processed, thank you for spotting them.

** line 28: Remove comma in "Other, common".*

- * line 56: “correlation between tropospheric composition with ENSO conditions”. Change to “relationship between tropospheric composition and ENSO conditions” or “dependence of tropospheric composition on ENSO conditions”?
- * line 57: arctic --> Arctic
- * lines 65-66. Change “and, afterwards, their chemistry modules technically integrated” to “. Afterwards, their chemistry modules were technically integrated”.
- * line 85: “which is further expanded” --> “which are further expanded”
- * line 130: I think Lacressonnière et al. 2012 is missing in the reference list.
- * lines 139-140: "and 2 for the troposphere for the aqueous oxidation reaction of sulfur dioxide into sulfuric acid"-->"and 2 for the aqueous oxidation reaction of sulfur dioxide into sulfuric acid in the troposphere"
- * Line 186: Brasseur et al. (1998) is missing in the reference list.
- * line 222: “This has consequences for the description of the various emissions”. Do you mean “partitioning” instead of “description”?
- * Line 246: “experiment ID’s” --> “experiment IDs”
- * Table 3. Remove “in” from “Specifications of the experiments in evaluated.
- * Line 324: No need comma before “over the NH extra-tropics”
- * Lines 326, 327 and probably other parts of the text: Please change “highest” to “the highest”
- * The expression “associated to” is used three times on page 13 as well as in other parts of the text. I think the use of “associated with” is preferable.
- * Line 367: Need space before parenthesis in “deviation(Jöckel”.
- * line 389: compare --> compares
- * "Southern Hemisphere (SH)" is spelled out on line 404, but this should have been done before (around line 393).
- * Figure 9: “Izaa” should be changed to “Izaña” (or “Izana” if the “ñ” is problematic) in one of the panels.
- * Figure 10: “monthly mean surface CO by observations (GAW network) and by the model simulations” --> “monthly mean surface CO as derived from observations (GAW network) and model simulations”
- * line 482: “phae” --> “phase”
- * caption of Figures 12, 13, 14, 18 and 19: “Comparison simulated” --> “Comparison of simulated”
- * page 27: The reference “Aydin et al., 2014” is mentioned twice. I believe this paper is from 2011, as indicated in the reference list.
- * lines 549-550: “the correlation showed acceptable values for all versions ($R^2 > 0.7$)”. The authors should explicitly refer to the “weighted correlation” rather than just “correlation”.
- * lines 562-563: “due to its photochemical balance with nitrogen oxide”. Do you refer to NO? If so, please change “nitrogen oxide” to “nitrogen monoxide” or “nitric oxide”.
- * line 569: Need comma or semicolon before “hence”.
- * Caption of Figure 16: “Tropospheric NO₂ from OMI” --> “Tropospheric NO₂ columns from OMI”
- * Caption of Figure 17: Remove “, and the model biases with respect to this”.
- * line 646: Change “is” to “are” in “the averaged differences for O₃(CO) is”
- * Reference list: Benedetti et al. (2009) is included twice.
- * Reference list: I think that Horowitz et al. (2003) is not mentioned in main text.
- * The main text includes these four references: Madronich and Flocke, 1997; Madronich and Flocke, 1999; Madronich et al., 1989; Madronich, 1987. They are somewhat different in the reference list. Please double-check.

References:

Miyazaki, K., Eskes, H., Sudo, K., Boersma, K. F., Bowman, K., and Kanaya, Y.: Decadal changes in global surface NO_x emissions from multi-constituent satellite data assimilation, *Atmos. Chem. Phys.*, 17, 807-837, <https://doi.org/10.5194/acp-17-807-2017>, 2017.