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 NOx 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 NOy 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 N_2O_5 chemical production (Cariolle et al., 2017), reducing in turn the NO_x lifetime because of a larger net N_2O_5 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 NOx 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 NOx 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 NOx)? If so, please clearly indicate this. Otherwise, you should explain which other processes might be involved. For instance, is there aerosol nitrate production, removing NOy 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 NOx through the heterogeneous reaction of N_2O_5 to HNO_3.**"

2.2 The last paragraph of Section 4 (Assessment of inter-model differences) tries to relate the differences in NOx-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 CH_2O 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. CH_2O 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 CH_2O degradation is responsible for about half the source of CO.
However, the amount of modeled chemical CH_2O production is indeed depending on the chemistry mechanism. Note additionally that the OH impacts both on CH_2O production, and its loss. Therefore, evaluation of the CH_2O 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 CH_3OOH, as further discussed in a reviewer comment below. We now include the following comment:

Considering the short lifetimes for CH_2O (a few hours in daytime) and also CH_3OOH, 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 C_2H_4, which is correlated to the comparatively low magnitude of OH, which explains the longer lifetime for C_2H_4.

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 NOx 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 NOx 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 NO_2 are mainly the use of a different numerical solver and differences in the efficiency assumed for N_2O_5 heterogeneous reactions (see Sec. 2.1.4).

** CBA has less NOx, 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 NOx).
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 NOx 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 NOx lifetimes and nitrate aerosol formation"
Lines 653-655 (in the conclusions): "For instance IFS(MOCAGE) tends to predict significantly higher NOx 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 NOx 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 HNO_3 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 HNO_3 production efficiency in CBA and MOZ, compared to MOC, which currently misses the N_2O_5 heterogeneous loss reaction. As this process mostly takes place near the surface, any HNO_3 production here consequently may lead to a more efficient dry (and also wet) deposition of HNO_3, as compared to HNO_3 produced by NO_2 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 NOx 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 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 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 NOx, are prescribed. The only exception is lightning NOx 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 NO_2 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 R^2 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 O_3), 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:

CH_3OOH is a main organic peroxide acting as a temporary reservoir of oxidizing radicals, Zhang et al. (2012). It is mainly formed through reaction of CH_3O_2 + HO_2, which are both produced in the oxidation process of many hydrocarbons. The CH_3OOH lifetime of globally about one day is mainly governed by its reaction with OH, and photolysis. Figure 13 presents an evaluation for CH_3OOH for the same sites are presented for CH_2O 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 CH_2O (a few hours in daytime) and also CH_3OOH, 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 ontropospheric 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 O3 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 (R2 >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 NOx emissions from multi-constituent satellite data assimilation, Atmos. Chem. Phys., 17, 807-837, https://doi.org/10.5194/acp-17-807-2017, 2017.

Response to the second Reviewer

We thank the referee for his/her positive review and for the provision of 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.

This manuscript evaluates tropospheric composition simulations in the CAMS modelling system and quantifies uncertainties related to different chemical schemes. It is well structured and written and illustrates original and interesting results for the . CAMS modelling system. I suggest acceptance of the manuscript for publication after taking into consideration the following comments.

Main Comments

1) I guess that the simulations were carried for the year 2011 but I think the authors should describe in Section 2.3 which was the time period that the simulations were carried out.

Model simulations for 1 July 2010 to 1 January 2012 have been carried out. We now include such a statement explicitly:

To allow for sufficient model spinup, **the model versions are initialized for 1 July 2010 and ran through until 1 January 2012**.
The first 6 months of the simulation are considered as spin-up and therefore not evaluated.

2) The authors mention that the averaging of large number of measurements over space and time partly solves the problem of interannual variability (lines 273-275 in page 11). Can this dataset of Emmons et al. (2000) be representative to compare with the CAMS simulations for the year 2011? I understand the uniqueness of this dataset but could the authors clarify this issue and discuss the uncertainties and the weaknesses of this comparison?

Indeed the referee is fully correct with his analysis. It is also true that for the *total* anthropogenic VOCs emissions, the changes between the year 1990 and 2011 are of the order of 14%, following Emissions Database for Global Atmospheric Research (EDGARv4.3.2 database). Therefore obviously caution has to be taken when analyzing the comparison, but we believe the aircraft dataset comparison to be still a valid methodology despite the large temporal difference between observations and modelled data because of the following two reasons:
We expect the impact of this change to be lower at background locations or outflow regions, as included in the comparison of presented in this manuscript, only partly affected by anthropogenic emissions, while biogenic emissions are expected to remain largely unchanged. Also the variability and measurements uncertainties present in the observations are larger than 14%, implying that we can still consider these observations representative, especially because they are averages over larger regions in space and time. To make this clear, we now write in the manuscript:

For the total anthropogenic VOCs emissions the changes between the year 1990 and 2011 are of the order of 14%, following Emissions Database for Global Atmospheric Research (EDGARv4.3.2 database). Nevertheless, the evaluations presented here are all sampling background locations or outflow regions, and are hence only partly affected by such changes in anthropogenic emissions. Also the variability as well as measurement

uncertainties present in the observations are larger than 14%, implying that we can still consider these observations representative.

3) I would suggest the authors to provide a short description of the method used to calculate the weighted values of bias and correlation in Table 4.

The method has been extensively described in Jöckel et al. (2006), where the mathematical derivation is also explained in the appendix. As the mathematical description of the method would be too tedious, we have added the following short description to the manuscript, referring to Jöckel et al (2006) for detailed information. We now write

As explained in further detail by Jöckel et al. (2006), with this approach, the measurement locations with high variability have less weight, whereas more weight is given to stable, homogeneous conditions. This allows us to compare values that are more representative for the average conditions and to eliminate specific episodes that cannot be expected to be reproduced by the model.

4) The authors write in line 375 (page 15) that " CBA is the only model version to deliver a satisfactory bias" . Is this a robust conclusion? What is statistically satisfactory? Looking Table 4 I see that in some species CBA bias is smaller than in other schemes, in some other species the biases are comparable and in other species the CBA bias is worse.

Note that this specific comment on this line only referred to SO_2, so the reviewer is correct that biases in CBA are not always the best. In general, whether a bias is *satisfactory* small will depend on the application area, which is indeed not detailed in present work. Instead, in Sec. 5 we defined a weighted bias, which relates the bias to the standard deviations in the model and observations. This value should be between [-1,1] to deliver satisfactory results.
Therefore we choose to re-formulate the specific sentence here to:

For SO_2 CBA is the only model version to deliver a weighted bias that is larger than -1.

5) In the evaluation of ozone the authors conclude that "overall, the evaluation at individual station provides reasonable agreement between model simulations and sondes". How these evaluation results compare with other ozone evaluation studies which were based on MACC and CAMS products (e.g. Inness et al., 2015; Katragkou et al., 2015; Akritidis et al., 2018). I think this conclusion could be also supported by these studies.

So far we didn't show evaluation results from other CAMS products as this is beyond the scope of current manuscript. Furthermore, one should realize that chemistry versions and model configurations as adopted here are to some extent different compared to those use in important MACC/CAMS products such as the reanalyses.
However, to aid the interpretation of the model quality, i.e. to put the current model performance into perspective, we now include an assessment of the CAMS Interim Reanalysis (CAMSiRA, Flemming et al., 2017) in the evaluation against ozone sondes. We choose only to show figures for the annual average, zonal average bias and RMSE at various altitude ranges, to give a general indication of our model performance relative to that of CAMSiRA. Indeed, this evaluation shows that biases and RMSE are within the range of those of CAMSiRA, with the free running model versions of equal (or better) performance towards the boundary layer, and CAMSiRA generally better in the free troposphere. For further details about the configuration

and performance the reader is referred to Flemming et al. (2017) and Inness et al. (2019). We now write:

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 O_3 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).**

6) In lines 448-449 (page 20) the authors write "Approximately half of the CO burden is directly emitted, and the rest formed through degradation of methane and other VOC's". Please add a relevant reference.

We now add Hooghiemstra et al. (ACP 2010) as reference for this statement, as they provide a detailed evaluation of a-priori and optimized budgets for global CO production.

7) On how many data points (and years) the temporal correlations shown in Figure 10 are based?

The temporal correlation presented in Figure 10 is based on twelve points (the monthly means) per station, and was evaluated for the year 2011. Therefore this figure shows an evaluation of the model ability to represent the seasonal cycle, as discussed in the manuscript.

8) In lines 526-528 (page 26) it is written "The vertical profiles (see Figure 13) are strongly biased (e.g., SONEX, Newfoundland and PEM-Tropics-A, Tahiti), with positive biases occurring at the surface and negative in the free troposphere."
Could this result also related to inadequate outflow from the atmospheric boundary layer (ABL) to the free troposphere (FT) and hence to model weakness in ABL-FT exchange?

Thank you for this interesting suggestion. Although such processes of vertical mixing could certainly contribute to uncertainties in the vertical distribution of C2H4, this would then also affect any of the other chemical tracers, as well as meteorological variables (e.g. humidity, temperature), which generally do not show indication of this type of issues. Therefore so far we have no indication that such an uncertainty is driving the discrepancy in the modeled vs observed C_2H_4 profiles, but rather believe that our emissions and chemistry contain larger uncertainties, as currently stated on the manuscript.

9) The authors refer to correlation R (that span from -1 to 1) but showing R2 which practically describes the explained variance. Although this is not crucial in the discussion it could propagate a misunderstanding on these statistical parameters when the article is published. I would suggest to modify this accordingly.

The reviewer is correct that there can be some confusion between the use of correlation in terms of R and R^2. Throughout the text we make sure to refer to R^2 when providing quantitative reference to the correlation, and at the start of Sec. 5 we now explicitly write:

“Table 4 summarizes the comparison of the various model results with aircraft measurements, described in Sec. 3.1, in terms of biases and correlation, **in terms of explained variance (R^2)**, …”

While or the table header we now write:

“Table 4. Summary of the Bias and correlation coefficients **(in terms of explained variance, R^2)** …”

10) Generally, I think that the discussion in model difference is rather technical and I would suggest the authors to discuss also the possible scientific reasons for discrepancies among the simulations with different chemical schemes for the different chemical species.

Indeed this manuscript has a largely technical focus, explaining the current state of the modeling system, its validation, and its general ability to quantify uncertainties due to model chemistry. Any more scientific reasons to explain discrepancies among simulations inherently require additional sensitivity studies, which is beyond the scope of this work. Having said this, we now do pay more attention to differences in model performance related to differences in their configurations, as also requested by the other reviewer, see particularly our responses to his ´Main Comments´. We hope this addresses the concerns raised here.

Minor comments
page 13, line 348: should rather be "relative shorter" instead of "relative short"

Thank you, we changed this accordingly.

Response to Executive Editor comment by Astrid Kerkweg.

Please find below our response. The editor comment is in italic. Our response is in regular font, and changes to the manuscript are given in bold.

please note, that if only one model is concerned, the title of a GMD manuscript should state the model name (or its acronym) and a version number. These are always important to know even in the case of an evaluation, as different versions might perform differently for the same evaluation procedure. Therefore please change the title of your manuscript accordingly upon revision; e.g., "Quantifying uncertainties due to chemistry modeling – evaluation of tropospheric composition simulations in the CAMS model (version x.y)"

Thank you for pointing this out. Note that here we evaluate results from three chemistry model versions in the CAMS model, which so far all follow a different versioning scheme. Therefore it is beyond the scope of the title to provide exact details of the version, but their communality is the IFS version adopted here, which is CY43R1, which is a good proxy for the chemistry version adopted. We will change the title accordingly to:

"Quantifying uncertainties due to chemistry modeling - evaluation of tropospheric composition simulations in the CAMS model (Cycle 43R1)"

We now furthermore refer to this cycle in the first sentence of the conclusions section.

Additionally, please note that evaluation papers also need to include a code availability section, telling the reader how to access the exact code version of the evaluated model or providing profound reasons why the code can not be accessed. Furthermore, please provide reasons, why the data is not freely available.

The editor is correct that we should be more explicit about code availability. Full data public availability is beyond reach, considering the large volume of data produced for these experiments. Data is fully archived on the ECMWF Archiving system (MARS) and selections will be made available freely to interested readers by contacting the first author. Also we provide details on accessibility of model code. According to this, we now change this section to:

The source code of the chemistry modules are integrated into ECWMF's IFS code, which is only available subject to a license agreement with ECMWF. The IFS code without modules for assimilation and chemistry can be obtained for educational and academic purposes as part of the openIFS release (https://confluence.ecmwf.int/display/OIFS). A detailed documentation of the IFS code is available from (https://www.ecmwf.int/en/forecasts/documentation-and-support/changes-ecmwf-model/ifs-documentation). The CB05 chemistry module of IFS was originally developed in the TM5 chemistry transport model. Readers interested in the TM5 code can contact the TM5 developers (http://tm5.sourceforge.net). The BASCOE stratospheric chemistry module can be freely obtained from the BASCOE developers (http://bascoe.oma.be). The MOCAGE chemistry module of IFS is developed at Météo-France on the basis of the

MOCAGE chemistry-transport model, http://www.umr-cnrm.fr/spip.php?article128. The MOZART code can be obtained through contacting their developers via https://www2.acom.ucar.edu/gcm/mozart. The MOZART and CB05BASCOE chemistry schemes are also freely available through the Sander et al. (2019) publication.
The model simulation datasets used in this work are archived on ECMWF archiving system (MARS) under the respective experiment IDs listed in Table 3. Readers with no access to this system can freely obtain these datasets from the corresponding author upon request.

Quantifying uncertainties due to chemistry modeling - evaluation of tropospheric composition simulations in the CAMS model (Cycle 43R1)

Vincent Huijnen[1], Andrea Pozzer[2], Joaquim Arteta[3], Guy Brasseur[4], Idir Bouarar[4], Simon Chabrillat[6], Yves Christophe[6], Thierno Doumbia[3], Johannes Flemming[5], Jonathan Guth[3], Béatrice Josse[3], Vlassis A. Karydis[2,7],Virginie Marécal[3], Sophie Pelletier[3]

[1]Royal Netherlands Meteorological Institute, De Bilt, The Netherlands
[2]Max Planck Institute for Chemistry, Mainz, Germany
[3]Centre National de Recherches Météorologiques Université de Toulouse, Météo-France, CNRS, Toulouse, France
[4]Max Planck Institute for Meteorology, Hamburg, Germany
[5]European Centre for Medium Range Weather Forecast, Reading, RG2 9AX, UK
[6]Royal Belgian Institute for Space Aeronomy, BIRA-IASB, 1080 Brussels, Belgium
[7]Forschungszentrum Jülich, Inst Energy & Climate Res IEK-8, D-52425 Jülich, Germany

Correspondence to: (Vincent.Huijnen@knmi.nl)

Abstract.

We report on an evaluation of tropospheric ozone and its precursor gases in three atmospheric chemistry versions as implemented in ECMWF's Integrated Forecasting System (IFS), referred to as IFS(CB05BASCOE), IFS(MOZART) and IFS(MOCAGE). While the model versions were forced with the same overall meteorology, emissions, transport and deposition schemes, they vary largely in their parameterizations describing atmospheric chemistry, including the organics degradation, heterogeneous chemistry and photolysis, as well as chemical solver. The model results from the three chemistry versions are compared against a range of aircraft field campaigns, surface observations, ozone sondes and satellite observations, which provides quantification of the overall model uncertainty driven by the chemistry parameterizations. We find that they produce similar patterns and magnitudes for carbon monoxide (CO) and ozone (O_3), as well as a range of non-methane hydrocarbons (NMHCs), with averaged differences for O_3 (CO) within 10% (20%) throughout the troposphere. Most of the divergence in the magnitude of CO and NMHCs can be explained by differences in OH concentrations, which can reach up to 50% particularly at high latitudes. Also comparatively large discrepancies between model versions exist for NO_2, SO_2 and HNO_3, which are strongly influenced by secondary chemical production and loss. Other~~,~~ common biases in CO and NMHCs are mainly attributed to uncertainties in their emissions. This configuration of having various chemistry versions within IFS provides a quantification of uncertainties induced by chemistry modeling in the main CAMS global trace gas products beyond those that are constrained by data-assimilation.

1. Introduction

The analysis and forecasting capabilities of trace gases are key objectives of the European Copernicus Atmosphere Monitoring Service (CAMS), in order to provide operational information on the state of the atmosphere. This service relies on a combination of satellite observations with state-of-the-art atmospheric composition modelling (Flemming et al., 2017). For that purpose, ECMWF's numerical weather prediction (NWP) system, the Integrated Forecasting System (IFS), contains modules for describing atmospheric composition, including aerosols (Morcrette et al., 2009; Benedetti et al., 2009), greenhouse gases (Agustí-Panareda, et al. 2016; Engelen et al., 2009), and reactive gases (Flemming et al., 2015).

Having atmospheric chemistry available within the IFS allows the use of detailed meteorological parameters to drive the fate of constituents, as well as its capabilities to constrain trace gas concentrations through assimilation of satellite retrievals. Furthermore, having atmospheric chemistry as an integral element of the IFS enables to study feedback processes between atmospheric chemistry and other parts of the earth system, such as the impact of ozone in the radiation scheme on temperature and the provision of trace gases as precursors for aerosol.

Other examples where chemistry modules have been implemented in general circulation models (GCM) for NWP applications have been, for instance, GEM-AQ (Kaminski et al., 2008; Struzewska et al., 2015), GEMS-BACH (de Grandpré et al., 2009; Robichaud et al., 2010), the Met Office's Unified Model (Morgenstern et al., 2009, O'Connor et al., 2014), and, on a regional scale, WRF-Chem (Powers et al., 2017).

The chemistry module that is currently used operationally in the CAMS service originates from the chemistry transport model TM5 (Huijnen et al., 2010). The chemistry module is based on a modified version of CB05 tropospheric chemistry (Williams et al., 2013), while stratospheric ozone is modelled using a linear ozone scheme (Cariolle and Deque 1986, Cariolle and Teyssèdre, 2007). This version, referred to as IFS(CB05), is used in a range of applications, such as for the CAMS operational analyses and forecasts of atmospheric composition (http://atmosphere.copernicus.eu), and for the generation of reanalyses: the CAMS Interim Reanalysis (CAMSiRA; Flemming et al., 2017) and the CAMS Reanalysis (Inness et al., 2018). Furthermore, this module is used in modelling studies, e.g., to analyse extreme fire events (Huijnen et al., 2016a; Nechita-Banda et al., 2018), to study the ~~correlation~~relationship between tropospheric composition with ENSO conditions (Inness et al., 2015). It has also contributed to model intercomparison studies such as ~~arctic~~Arctic pollution (Emmons et al., 2015), HTAP (e.g. Huang et al., 2017) and AQMEII (Im et al., 2018).

Other chemistry versions have also been implemented in the IFS, where each version has its choice regarding the gas phase chemical mechanism, computation of photolysis rates, definition of cloud and heterogeneous reactions, and solver specifics. This enables flexibility in the choice of the atmospheric chemistry component in the global CAMS system. A model version which contains the extension of the CB05 scheme with a comprehensive stratospheric chemistry originating from the Belgian Assimilation System for Chemical ObsErvations (BASCOE, Skachko et al., 2016) has been developed (Huijnen et al., 2016b). Furthermore, in predecessors of the current system, the MOZART (Kinnison et al., 2007) and MOCAGE

(Bousserez et al., 2007) chemistry transport models had also been coupled with IFS (Flemming et al., 2009~~) and, afterwards~~). Afterwards, their chemistry modules were technically integrated into the IFS (Flemming et al., 2015). Only recently, three fully functioning systems have been prepared, as are presented here, based on CB05BASCOE, MOZART and MOCAGE chemistry.

Many studies such as HTAP and AQMEII (Galmarini et al., 2017) try to explore the uncertainties of global chemistry modelling through changing emissions. But in such multi-model assessments also meteorological model parametrisations, such as advection, deposition or vertical diffusion vary (e.g. Emmons et al., 2015; Huang et al., 2017; Im et al., 2018). While such a multi-model approach is appropriate to define the overall uncertainty, it makes it hard to isolate the impact of the differences in the chemistry parameterizations. In this work we study the model spread caused by three chemistry modules that are fully independent, in an otherwise identical configuration for describing meteorology, transport, emissions and deposition. This endeavour intends to provide insights in the uncertainty induced purely by the simulation of chemistry and as such complements the many model intercomparison studies that try to explore other sources of uncertainty in global atmospheric modelling.

The central application area of tropospheric chemistry analyses and forecasts in the IFS are to provide a global coverage of the current state of atmospheric composition, along with its long-term trends (Inness et al., 2018). These are intensively used as boundary conditions to regional models (Marécal et al., 2015). Uncertainty information is relevant to CAMS users of global chemistry forecasts, in particular for the trace gases that are not, or poorly, constrained by observations, such as the non-methane hydrocarbons (NMHCs) and reactive nitrogen species. Therefore we focus here not only on the model ability to represent tropospheric ozone (O_3) and carbon monoxide (CO), but also include evaluations of the ~~NMHC's~~NMHCs, nitrogen dioxide (NO_2), nitric acid (HNO_3) and sulfur dioxide (SO_2).

In this study, we rely on various sets of observations. Comparatively dense in-situ observation networks exist to measure surface and tropospheric CO and O_3, which ~~is~~are further expanded by satellite retrievals for CO and NO_2 columns. Observations from aircraft campaigns form a crucial source of information on atmospheric composition, particularly for the ~~NMHC's~~NMHCs, and have been used in the past in various modelling efforts and intercomparison studies (e.g. Pozzer et al., 2007; Emmons et al., 2015). Even though all model versions considered here contain both parameterizations for tropospheric and stratospheric chemistry, we limit ourselves to evaluating differences in the tropospheric composition; evaluation of stratospheric composition is beyond the scope of this work. It is worth noting that each of the versions are constantly developed further over time, which means that particular aspects of the model performance, and as a consequence inter-model spread, are subject to change depending on model version.

The paper is structured as follows. Section 2 provides a description of the various chemistry schemes implemented in IFS. Section 3 provides an overview of the observational datasets used for model evaluation, while in Section 4 a basic assessment of model differences for tracers playing a key role in tropospheric ozone is provided. Section 5 contains the evaluation against observations of a full year simulation with the three atmospheric chemistry versions of IFS with focus on tropospheric chemistry. The paper is concluded with a summary and an outlook in Section 6, where also the recent model evolution in the various versions is briefly described.

2. Model description

2.1. Chemical mechanisms

The three chemistry schemes implemented in the IFS are described in more detail in the following subsections. A brief analysis of elemental differences is given in Sec. 2.1.4

2.1.1. IFS(CB05BASCOE)

For IFS(CB05BASCOE), a merging approach has been developed where the tropospheric and the stratospheric chemistry schemes are used side-by-side within IFS (Huijnen et al.,2016b). The tropospheric chemistry in the IFS is based on a modified version of the CB05 mechanism (Yarwood et al., 2005). It adopts a lumping approach for organic species by defining a separate tracer species for specific types of functional groups. Modifications and extensions to this include an explicit treatment of C1 to C3 species as described in Williams et al., (2013), and SO_2, di-methyl sulphide (DMS), methyl sulphonic acid (MSA) and ammonia (NH_3) (Huijnen et al., 2010). Gas-aerosol partitioning of nitrate and ammonium is calculated using the Equilibrium Simplified Aerosol Model (EQSAM, Metzger et al., 2002). Heterogeneous reactions and photolysis rates in the troposphere depend on cloud droplets and the CAMS aerosol fields. The reaction rates for the troposphere follow the recommendations given in either JPL evaluation 17 (Sander et al., 2011) or Atkinson et al. (2006). The modified band approach (MBA) is adopted for the online computation of photolysis rates in the troposphere (Williams et al., 2012) and uses 7 absorption bands across the spectral range 202 − 695 nm, accounting for cloud and aerosol optical properties. At instances of large solar zenith angles (71-85°) a different set of band intervals is used. The complete chemical mechanism as applied for the troposphere is referred to as 'tc01a', and is extensively documented in Flemming et al. (2015).

For the modelling of atmospheric composition above the tropopause, the chemical scheme and the parameterization for Polar Stratospheric Clouds (PSC) have been taken over from the BASCOE system (Huijnen et al., 2016b), version 'sb14a'. Lookup tables of photolysis rates were computed offline by the TUV package (Madronich and Flocke, 1999) as a function of log-pressure altitude, ozone overhead column and solar zenith angle. Gas-phase and heterogeneous reaction rates are taken from JPL evaluation 17 (Sander et al., 2011) and JPL evaluation 13 (Sander et al., 2000), respectively.

Both for solving the tropospheric and stratospheric reaction mechanism we use KPP-based four stages, 3^{rd} order Rosenbrock solvers (Sandu and Sander, 2006). Photolysis rates for reactions occurring both in the troposphere and stratosphere are merged at the interface, in order to ensure a smooth transition between the two schemes. To distinguish between the tropospheric and stratospheric regime, we use a chemical definition of the tropopause level, where tropospheric grid cells are defined at O_3<200 ppb and CO>40 ppb, for P > 40 hPa. With this definition the associated tropopause pressure ranges in practice approximately between 270 and 50 hPa globally, with the lowest tropopause pressure naturally in the tropics.

2.1.2. IFS(MOCAGE)

The MOCAGE chemical scheme (Bousserez et al, 2007, Lacressonnière et al. 2012) is a merge of reactions of the tropospheric RACM (Regional Atmospheric Chemistry Mechanism) scheme (Stockwell et al., 1997) with the reactions relevant to the stratospheric chemistry of REPROBUS (REactive Processes Ruling the Ozone BUdget in the Stratosphere) (Lefèvre et al., 1994, Lefèvre et al. 1998). It uses a lumping approach for organic trace gas species. The MOCAGE chemistry has been extended, in particular by the inclusion of the sulphur cycle in the troposphere (Ménégoz et al. 2009) and peroxyacetyl nitrate (PAN) photolysis.

The RACMOBUS (RACM-REPROBUS) chemistry scheme implemented in IFS uses 115 species in total, including long-lived and short-lived species, family groups and a ~~polar stratospheric clouds (~~PSC~~)~~ tracer. A total of 326 thermal reactions and 53 photolysis reactions are considered to model both tropospheric and stratospheric gaseous chemistry. Nine heterogeneous reactions are taken into account for the stratosphere and 2 for the ~~troposphere for the~~ aqueous oxidation reaction of sulfur dioxide into sulfuric acid in the troposphere (Lacressonnière et al., 2012). For photolysis rates, a lookup table of photolysis rates was computed offline by the TUV package (Madronich and Flocke, 1997, version 5.3.1) as a function of solar zenith angle, ozone column above each cell, altitude and surface albedo.

2.1.3. IFS(MOZART)

The atmospheric chemistry in IFS(MOZART) is based on the MOZART-3 mechanism (Kinnison et al., 2007) and includes additional species and reactions from MOZART-4 (Emmons et al., 2010) and further updates from the Community Atmosphere Model with interactive chemistry, referred to as CAM4-chem (Lamarque et al., 2012; Tilmes et al., 2016). As for IFS(CB05BASCOE), the heterogeneous reactions in the troposphere are parameterized based on aerosol surface area density (SAD) which is derived using the CAMS aerosol fields. IFS(MOZART) contains a parameterization for the gas-aerosol partitioning of nitrate and ammonium (Emmons et al., 2010). The heterogeneous chemistry in the stratosphere accounts for heterogeneous processes on liquid sulfate aerosols and polar stratospheric clouds following the approach of Considine et al. (2000).

The photolysis frequencies in wavelengths from 200 to 750 nm are calculated from a look-up table, based on the 4-stream version of the Stratosphere, Troposphere, Ultraviolet (STUV) radiative transfer model (Madronich et al., 1989). For wavelengths from 120 nm to 200 nm, the wavelength-dependent cross sections and quantum yields are specified and the transmission function is calculated explicitly for each wavelength interval. In the case of J(NO) and $J(O_2)$, detailed photolysis parameterizations are included online. The current IFS(MOZART) version includes the influence of clouds on photolysis rates which is parameterized according to Madronich (1987). However, currently it does not account for the impact of aerosols. A detailed description of the parametrization of photolysis frequencies, absorption cross sections, and quantum yields is given in Kinnison et al. (2007).

2.1.4. Key differences in chemistry modules

An overview of the most important differences in the three chemistry modules described above is given in Table 1. First, there are large differences in the choices made to compile the tropospheric chemistry mechanism. IFS(MOZART) describes the degradation of organic carbon types C1, C2, C3, C4, C5, C7 and C10, together with lumped aromatics, while IFS(CB05BASCOE) only describes explicit degradation up to C3, with the same reactions as present in IFS(MOZART). Instead, emissions and degradation of higher ~~VOC's~~volatile organic compounds (VOCs) in IFS(CB05BASCOE) are lumped to a few tracers. Furthermore, the parameterization of the isoprene and terpenes degradation is simpler in IFS(CB05BASCOE) than in IFS(MOZART). Aromatics are currently not described in IFS(CB05BASCOE), while they are accounted for with simple approaches in IFS(MOZART).

IFS(MOCAGE) describes many more lumped organic species than IFS(CB05BASCOE) and IFS(MOZART), also accounting for the more complex organics beyond C3. Furthermore, IFS(MOCAGE) uses a rather different lumping approach and contains more complexity for different terpene components, and also including aromatics. Such differences are bound to impact the effective degradation of VOCs, and thus ozone production efficiency and oxidation capacity, e.g. Sander et al. (2018).

With respect to the inorganic chemistry, the schemes are mostly similar. Still, IFS(MOCAGE) includes HONO chemistry, which is missing in both IFS(CB05BASCOE) and IFS(MOZART) implementations. Gas-phase sulfur chemistry is mostly similar between IFS(CB05BASCOE) and IFS(MOZART), while IFS(MOCAGE) has some more complexity through considering reactions involving DMSO and H_2S. 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 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. So far two-way coupling of secondary aerosol formation was not available in any of the current model versions.

Regarding the treatment of photolysis in the troposphere, IFS(CB05BASCOE) applies a modified band approach, where for 7 wavelengths the photolysis rates are computed online, taking into account the scattering and absorption properties of gases, (overhead ozone and oxygen), clouds and aerosol. IFS(MOCAGE) adopts a lookup-table approach, accounting for overhead ozone column, solar zenith angle, surface albedo and altitude, providing photolysis rates for clear-sky conditions. The impact of cloudiness on photolysis rates is applied online in IFS during the simulation using the parameterisation proposed by Brasseur et al. (1998). IFS(MOZART) applies the lookup-table approach from MOZART-3 (Kinnison et al., 2007), considering overhead ozone column and cloud scattering effects on photolysis rates. Despite such larger differences, an intercomparison of an instantaneous field of photolysis rates showed similar average profiles, with spread in magnitudes in the range of 5% in the tropical free troposphere for important photolysis rates like jO_3, jNO_2, $jHNO_3$. Locally differences are larger, associated, amongst others, to different cloud treatment (Hall et al., 2018).

As for the stratospheric chemistry, IFS(CB05BASCOE) contains the largest complexity of the three model versions containing both more species and reactions compared to the other mechanisms.

Different methods are used to solve the reaction mechanism. IFS(CB05BASCOE) applies the Rosenbrock solver, IFS(MOCAGE) here applies a first-order semi-implicit solver with fixed time steps, and IFS(MOZART) applies the explicit Euler method for species with long lifetimes (e.g. N_2O) and an implicit backward Euler solver for other trace gases with short lifetimes. Experiments using different solvers for both IFS(CB05BASCOE) and IFS(MOCAGE) have revealed significant differences, with decreases in tropospheric ozone in the order of up to 20% regionally when replacing a semi-implicit solver with the Rosenbrock solver. These differences are mostly traced to ~~differences~~an increase in the N_2O_5 ~~chemistry~~chemical production (Cariolle et al., 2017), ~~affecting~~reducing in turn the NO_x lifetime ~~and hence the~~because of a larger net N_2O_5 loss on aerosol. This in turn leads to a reduced chemical ozone production efficiency.

Table 1. Specification of elemental aspects describing the three chemistry versions.

	IFS(CB05BASCOE)	IFS(MOCAGE)	IFS(MOZART)
Tropospheric chemistry	Carbon Bond	RACM	CAM4-Chem
Stratospheric chemistry	BASCOE	REPROBUS	MOZART3
Number of species	99	115	115
Number of thermal reactions	219	326	266
Number of photolysis rates	60	53	51
Complexity of organic chemistry	Explicit degradation pathways up to C3	Detailed lumping approach	Explicit degradation pathways up to C10
Complexity of	No HONO	More extended, incl.	Similar to CB05BASCOE

inorganic chemistry		HONO	
Aerosol interaction in troposphere	HO_2 and N_2O_5 heterogeneous reactions, aerosol impact on photolysis	HO_2 and N_2O_5 heterogeneous reactions	None
Photolysis parameterization	Modified Band (trop) LUT (strat)	LUT	LUT (trop) Explicit transmission function (strat)
~~Number of heterogeneous reactions~~	~~10~~	~~11~~	~~11~~
Solver	3rd order Rosenbrock	1st order semi-implicit	Explicit forward and implicit backward Euler

2.2. Emission, deposition and surface boundary conditions

The actual emission totals used in the simulation for 2011 from anthropogenic, biogenic and natural sources, biomass burning as well as lightning NO are given in Table 2. MACCity emissions are used to prescribe the anthropogenic emissions (Granier et al., 2011), where wintertime CO traffic emissions have been scaled up according to Stein et al. (2014). Aircraft NO emissions are 1.8 Tg NO yr^{-1} , following Lamarque et al. (2010). Lightning NO emissions are parameterized as described in Flemming et al. (2015).

Monthly specific biogenic emissions originating from the MEGAN-MACC inventory (Sindelarova et al., 2014) are adopted, complemented with POET-based oceanic emissions (Granier et al., 2005).

Daily biomass burning emissions are taken from the Global Fire Assimilation System (GFAS) version 1.2, which is based on satellite retrievals of fire radiative power (Kaiser et al., 2012).

~~The actual emission totals used in the simulation for 2011 from anthropogenic, biogenic and natural sources, biomass burning as well as lightning NO are given in Table 2.~~

~~Dry deposition velocities in the current configuration were provided as monthly mean values from a simulation using the approach discussed in Michou et al. (2004). To account for the diurnal variation in deposition velocities, a cosine function of the solar zenith angle is adopted with a ±50% variation. Wet scavenging, including in-cloud and below cloud scavenging as well as re-evaporation is treated following Jacob et al. (2000). The reader is referred to Flemming et al., (2015) for further details on dry and wet deposition parameterization.~~

As described above, the chemistry mechanisms vary particularly in their description of the VOC degradation, with the most explicit treatment described in IFS(MOZ), while IFS(MOCAGE) and IFS(CB05BASCOE) rely on a more extended lumping approach. This has consequences for the ~~description~~partitioning of the various emissions. Still, we have ensured that the total of VOC and aromatics emissions in terms of Tg carbon are essentially the same for the three chemistry schemes.

For CB05BASCOE, the emissions of ~~Parafins~~ ‘parafins’ (toluene and higher alkane emissions), ~~Olefins~~ ‘olefins’ (butenes and higher alkenes), and ~~Aldehydes~~ ‘aldehydes’ (acetaldehyde and other aldehydes) have been prescribed. Likewise, MOZART applies emissions of BIGALK (Butanes and higher alkanes) and BIGENE (Butenes and higher alkenes). MOCAGE adopts tracers HC3, HC5, and HC8, over which emissions of ~~acetylene~~ethyne, propane, butanes and higher alkanes, esters, methanol and other alcohols are distributed, whereas DIEN contains butenes and higher alkenes emissions.

As for the aromatics, IFS(CB05BASCOE) disregards those, but includes toluene carbon emissions as part of the ~~Parafins~~parafins. IFS(MOZART) treats additionally a TOLUENE tracer, while IFS(MOCAGE) contains two types of aromatics, designated TOL and XYL. These aromatic emissions are composed from toluene, trimethyl-benzene, xylene and other aromatics.

Dry deposition velocities in the current configuration were provided as monthly mean values from a simulation using the approach discussed in Michou et al. (2004). To account for the diurnal variation in deposition velocities, a cosine function of the solar zenith angle is adopted with a ±50% variation. Wet scavenging, including in-cloud and below cloud scavenging as well as re-evaporation is treated following Jacob et al. (2000). The reader is referred to Flemming et al., (2015) for further details on dry and wet deposition parameterization.

Methane~~,~~ (CH$_4$), N$_2$O and a selection of ~~CFC's~~chlorofluorocarbons (CFCs) are prescribed at the surface as boundary conditions. While for N$_2$O and CFC currently annually and zonally fixed values are assumed (Huijnen et al., 2016b), for CH$_4$ zonally and seasonally varying surface concentrations are adopted based on a climatology derived from NOAA flask observations ranging from 2003 to 2014.

Table 2. Specification of annual emission totals from anthropogenic, biogenic and natural sources and biomass burning for 2011, in Tg species, for three chemistry versions.

Species	*Anthropogenic*	*Biogenic+oceanic*	*Biomass burning*
CO	*602*	*91+20*	*326*

NO[a]	71.2+1.8 AC	11.3+9.2 LiNO	8.8
HCHO	3.4	4.8	4.8
CH_3OH	2.2	127	6.7
C_2H_6	3.3	0.3+1.0	2.2
C_2H_5OH	2.2	19.3	0.
C_2H_4	7.6	30+1.4	3.9
C_3H_8	4.0	1.3	1.2
C_3H_6	3.5	15.2+1.5	2.3
CH_3CHO and higher aldehydes	1.3	23.5	3.8
CH_3COCH_3	1.4	38	1.8
~~Butanes~~butanes and higher alkanes	35.	0.1	2.
~~Butenes~~butenes and higher alkenes	4.7	3.1	1.6
C_5H_8		593	
~~Terpenes~~terpenes		95	
SO_2	97	13	1.
DMS		38	0.2
NH_3	43	2+8	6.5

[a] Anthropogenic surface NO emissions (Tg NO) are split according to 90% NO and 10% NO_2 emissions. Additionally, they contain a contribution of 1.8 Tg NO aircraft emissions and 9.2 Tg NO lightning emissions (LiNO).

2.3. Model configuration and meteorology

The IFS model versions evaluated here were implemented in IFS cycle 43R1, and are run on a T255 horizontal resolution (~ 0.7 degree) with 60 model levels in the vertical up to 0.1 hPa, all excluding chemical data assimilation. The naming conventions and experiment ~~ID's~~IDs for the three model runs are specified in Table 3. For brevity we refer to the model runs as ‘CBA’, ‘MOC’ and ‘MOZ’, respectively. A 30 minutes time stepping for the dynamics is applied while meteorology is nudged towards ERA-Interim. To allow for sufficient model spinup, the model versions are initialized for 1 July 2010 and ran through until 1 January 2012. The initial condition (IC) fields have been generated for ~~1 July 2010~~this date, using as much as possible realistic and consistent fields. For this purpose, tropospheric CO, O_3 from the CAMS-Interim reanalysis (Flemming et al., 2017) have been combined with ~~VOC’s~~ VOCs from its control run. ~~CFC’s~~CFCs, halogens and other tracers relevant for stratospheric composition originate from the BASCOE reanalysis v05.06, (Skachko et al., 2016), and have been merged for altitudes below tropopause with model fields from Huijnen et al. (2016b), all specified for 1 July 2010. For MOZ and MOC, these IC fields have been completed for a few missing ~~VOC’s~~ VOCs and ~~CFC’s~~ CFCs using separate MOZART and MOCAGE climatologies, respectively. The first 6 months of the simulation are considered as spin-up and therefore not evaluated.

For the evaluation, the model was sampled in the troposphere and lower stratosphere (i.e. the lowest 40 model levels) every three hours, to have a full coverage of the daily cycle. 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.

Table 3. Specifications of the experiments ~~in~~ evaluated.

name	Short name	expID	Color-coding
IFS(CB05BASCOE)	CBA	a028	**red**
IFS(MOCAGE)	MOC	b0l8	**blue**
IFS(MOZART)	MOZ	b0w3	**Green**

3. Observational datasets

3.1 Aircraft Measurements

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. Although these measurements cover only limited time periods, they provide valuable information about the vertical distribution of the analyzed trace gases. The database is formed by data from a number of aircraft campaigns that took place during 1990-2001, gridded onto global maps, forming data composites of chemical species important for tropospheric ozone photochemistry. These are used to create observation-based climatologies (Emmons et al. 2000). Here we use measurements from ozone, CO, CH_2O, C_2H_6, C_2H_4, methyl hydroperoxide (CH_3OOH), NO_2, nitric acid (HNO_3), and sulphur dioxide (SO_2). Note that the field campaigns used in this evaluation have been extended including also data observed after the year 2000, such as the TOPSE and TRACE-P campaigns. The geographical distribution of the aircraft campaigns and their area coverage are shown in Figure 1.

Although the specific field campaign data is in theory representative for the specific year, the averaging of large number of measurements over space and time partly solves the problem of interannual variability, and therefore these data can be considered as a climatology. Pozzer et al. (2009) showed that the correlation between model results and these observations would vary less than 5% if model results 5 years apart were used. For the total anthropogenic VOCs emissions the changes between the year 1990 and 2011 are of the order of 14%, following Emissions Database for Global Atmospheric Research (EDGARv4.3.2 database). Nevertheless, the evaluations presented here are all sampling background locations or outflow regions, and are hence only partly affected by such changes in anthropogenic emissions. Also the variability as well as measurement uncertainties present in the observations are larger than 14%, implying that we can still consider these observations representative. Finally these data summaries are useful for providing a picture of the global distributions of NMHCs and nitrogen-containing trace gases.

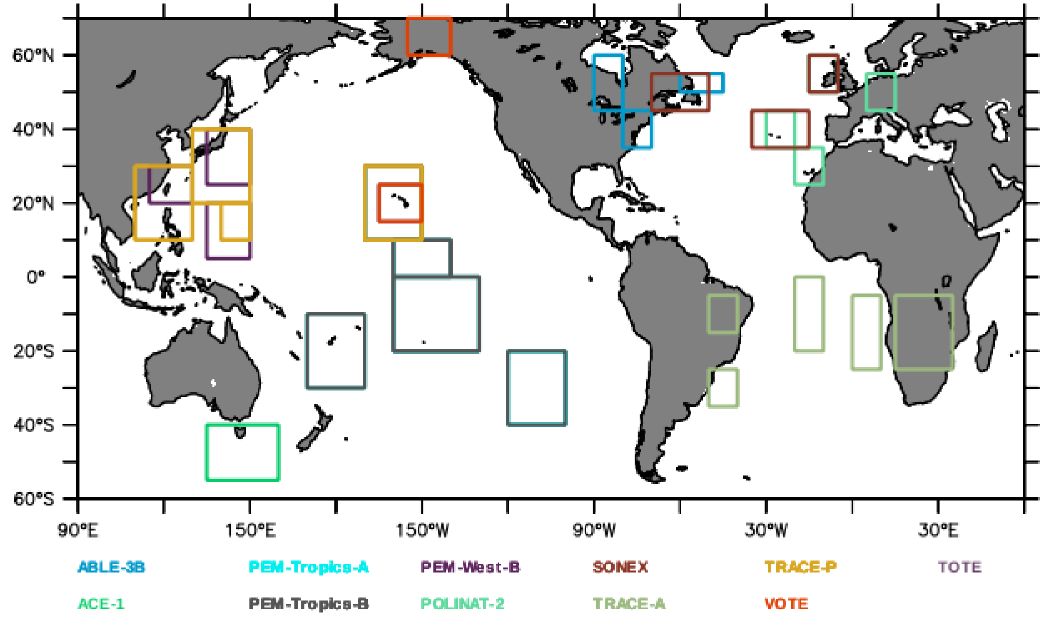


Figure 1: Geographical distribution of the aircraft campaigns presented by Emmons et al. (2000). Each field campaign is represented by a different color. Further information on the campaigns is found in Emmons et al. (2000).

3.2. ~~In-situ observations~~Near-surface CO and ozone sondes

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 ~~Reactive~~Greenhouse Gases (~~WDCRG~~WDCGG), the data repository and archive for ~~reactive~~greenhouse and related gases of the World Meteorological Organisation's (WMO) Global Atmosphere Watch (GAW) programme. The uncertainty of the CO observations is estimated to be in the order of 1–3 ppm (Novelli et al., 2003).

Tropospheric ozone was evaluated using sonde measurement data available from the World Ozone and Ultraviolet Radiation Data Center (WOUDC, http://woudc.org), further expanded with observations from the Network for the Detection of Atmospheric Composition Change (NDACC) network. About 50 individual stations covering various worldwide regions are taken into account for the evaluation over the Arctic, ~~Northern Hemisphere~~northern hemisphere mid-latitudes, tropics, ~~Southern Hemisphere~~southern hemisphere mid-latitudes and Antarctic. 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. The precision of ozone sonde observations in the troposphere is on the order of -7 to 17% (Komhyr et al., 1995; Steinbrecht et al., 1998), while larger errors are found in the presence of steep gradients and where the ozone amount is low.

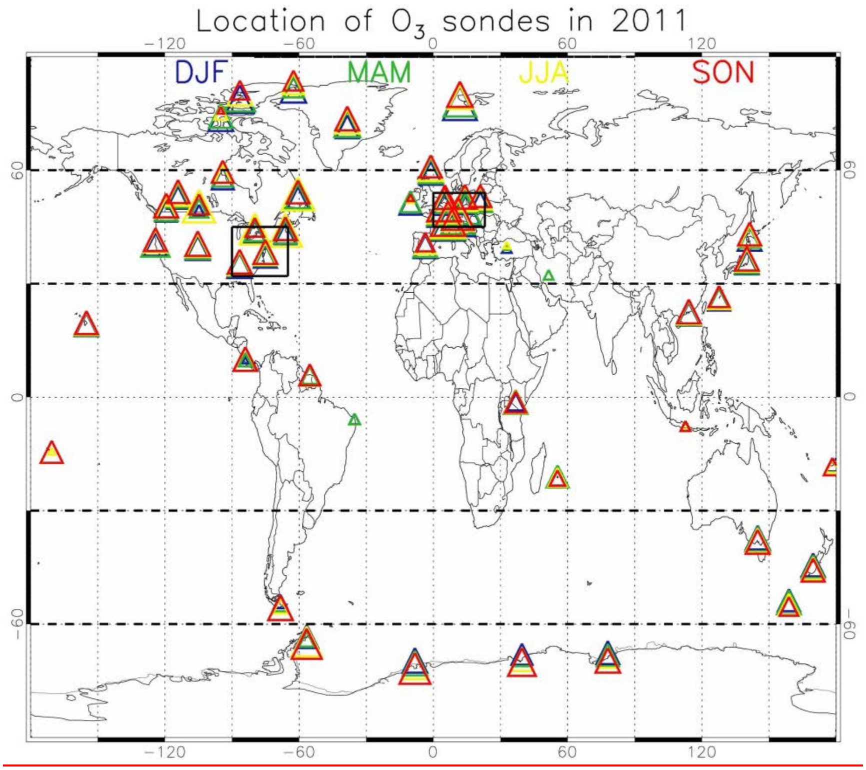


Figure 2: Geographical distribution of the ozone sondes during 2011 used for evaluation, colored for the various seasons. The size of the triangles provides information of the relative amount of observations available for each of the seasons and locations, compared to the other locations. Also the geographical aggregation for the five latitude bands presented in Figures 5 and 7, as well as the Western Europe and Eastern US regions is given.

3.3. Satellite observations

MOPITT (Measurements of Pollution in the Troposphere) V7 CO column observations (Deeter et al., 2017) are used to evaluate the CO total columns. The MOPITT instrument is a multi-channel Thermal InfraRed (TIR) and Near InfraRed (NIR) instrument operating onboard of the Terra satellite. The total column CO product is based on the integral of the retrieved CO volume mixing ratio profile. A CAM4-chem (Lamarque et al., 2012) based climatology is used to provide the MOPITT *a priori* profiles. For our study we use the TIR-derived CO total column observations, which are provided both over the oceans and over land. Highest CO sensitivities of these MOPITT TIR measurements are in the middle troposphere, around 500 hPa. Sensitivity to the lower troposphere depends on the thermal contrast between the land and lower atmosphere, which is higher during the day than in the night. Therefore, in our study we only use daytime MOPITT TIR observations. Standard deviation of the error in individual pixels for the MOPITT V7-TIR product evaluated against NOAA flask measurements is reported as 0.13×10^{18} molec cm^{-2} (Deeter et al., 2017), i.e. in the order of 10% of the observation value. Daily mean model CO columns have been gridded to a 1° x 1° spatial resolution, and for our analysis we applied the MOPITT averaging kernels to the logarithm of the mixing ratio profiles, following Deeter et al. (2012).

OMI retrievals of tropospheric NO_2 were taken from the QA4ECV dataset (Boersma et al., 2017). For this evaluation the 3-hourly model output of NO_2 was interpolated in time to local overpass of the satellite (13:30h), while pixels with satellite-observed radiance fraction originating from clouds greater than 50% were filtered out. The averaging kernels of the retrievals are taken into account, hence making the evaluation independent of the a priori NO_2 profiles used in the retrieval algorithm. Note that by using the averaging kernels the model levels in the free troposphere are given relatively greater weight in the column calculation, which means that errors in the shape of the NO_2 profile can contribute to biases in the total column.

4. Assessment of inter-model differences

In this section we provide a basic assessment of magnitude and differences in annual and zonal mean concentration fields between the three chemistry versions for a few essential tracers: O_3, CO, NO_x (=NO+NO_2) and OH. This provides a first insight in the correspondences and differences between chemistry modules and will help to interpret more quantitative differences seen in the evaluation against observations.

The annual zonal mean O_3 mixing ratios (Figure 2, top) show very similar patterns, with overall low values over the southern hemisphere (SH) and the highest over the Northern Hemisphere (NH) mid-latitudes, associated ~~to~~with the dominating emission patterns. Differences between chemistry versions are in the order of 10%, with MOC showing comparatively the lowest values over the tropical free troposphere and MOZ the highest~~,~~ over the NH extra-tropics. Differences in tropospheric ozone between model versions are remarkably small on a global scale.

Likewise, annual zonal mean CO mixing ratios show highest values associated ~~to~~with pollution regions in the tropics and over the NH. ~~Highest~~The highest values are obtained with CBA, and lowest with MOC, with differences ranging between 10 and 20~~%, suggesting mainly~~%. 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.

Zonal mean NO_x mixing rations, a tracer playing a crucial role in ozone formation, show overall the highest values for MOC and lowest for CBA. MOZ and CBA are overall similar, but MOC is showing higher values ~~over~~in the lower and mid-troposphere in the tropics and up to the NH high-latitudes~~ (>60°N) and also at altitudes below 900 hPa in the tropics~~. 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 N_2O_5 to HNO_3. Additionally, Cariolle et al. (2017) showed limitations of the Semi-Implicit method as used in MOC for resolving NO_x chemistry. Both elements likely contribute to significantly larger tropospheric NO_x lifetimes in MOC compared to CBA and MOZ. In contrast, the NO_x lifetime in IFS(CB05BASCOE) scheme is comparatively short, which is associated ~~to~~with a diagnosed relatively efficient organic nitrate production term from the reaction of NO_x with ~~VOC's~~VOCs in the modified CB05 mechanism compared to other mechanisms, as assessed in a box-modeling configuration (Sander et al., 2018).

Figure 2 also shows the annual, zonal mean concentrations of OH. Overall, the magnitude of OH is largest for MOC and lowest for CBA, with MOZ in between. The largest differences in absolute terms are found in the tropics, where the concentrations are highest. Nevertheless, in relative terms the largest differences are found in the extra-tropics, particularly over the SH, as can be seen from Figure 3. This figure shows the temporal evolution of the difference between MOC and MOZ simulated daily average OH at 600 hPa. This shows that differences can be up to 50% in daily averages, in particular over the extra-tropics where the absolute values are lower compared to those in the tropics.

Tropospheric NO_x in MOC is comparatively high, suggesting relatively efficient O_3 and OH production. On the other hand, the photolysis rates of tropospheric ozone, responsible for the primary production of OH, are very similar (not shown). Therefore the ozone production in MOC must be counter-balanced by a relatively large loss through reaction with OH and HO_2 (which are the other major loss terms in the ozone cycle), suggesting a relatively short tropospheric O_3 lifetime. ~~Such~~An assessment of the ozone chemical production and loss terms is beyond the scope of this work. But such differences in oxidation capacity naturally have important implications for understanding differences in the performance of NMHCs, as discussed in the next sections.

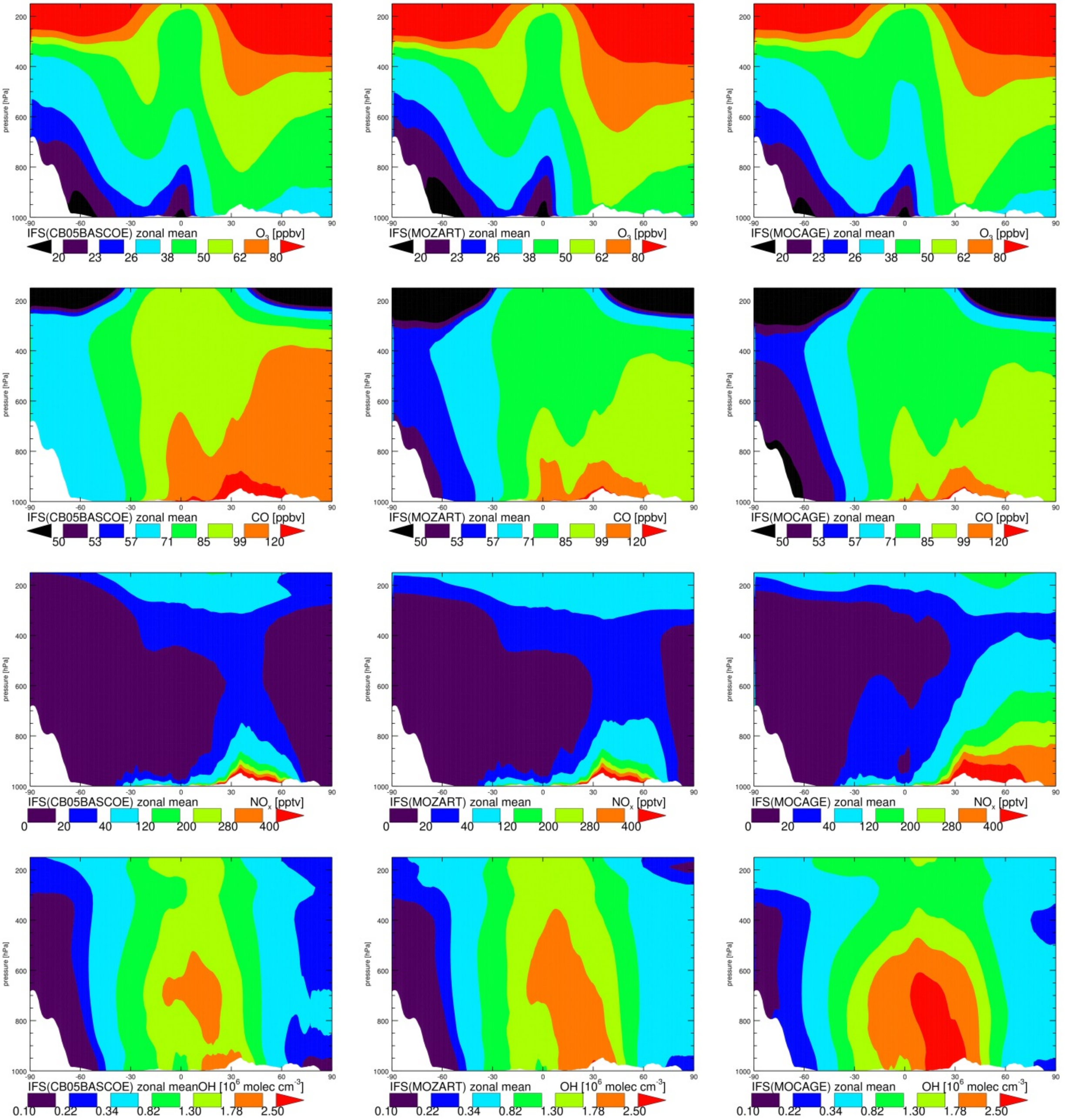


Figure 3: Zonal, annual mean O_3, CO, NO_x, mixing ratios and OH concentrations in CBA (left), MOZ (middle) and MOC (right).

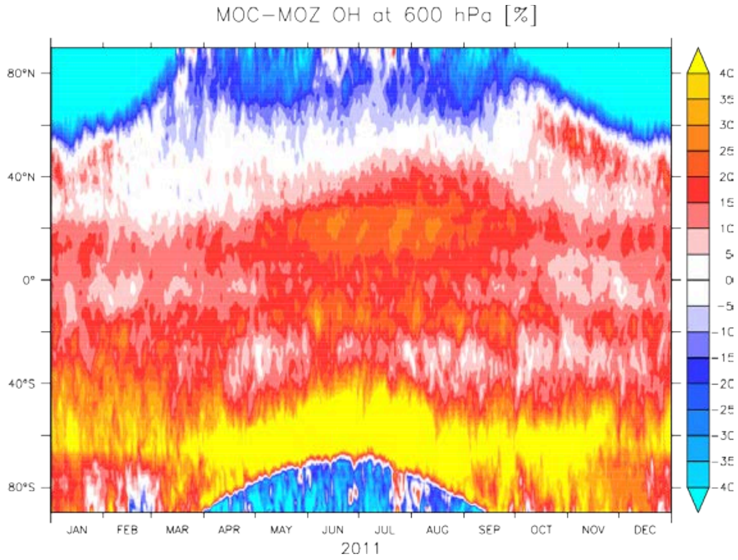


Figure 4: Relative differences (in %) of OH daily averaged mixing ratios of simulation MOC with respect to MOZ at 600 hPa.

5. Evaluation against observations

In this section we evaluate the model simulations against a range of observations, including ozone sondes, aircraft measurements, and satellite observations from carbon monoxide and nitrogen dioxide.

Table 4 summarizes the comparison of the various model results with aircraft measurements, described in Sec. 3.1, in terms of biases and correlation, in terms of explained variance (R^2), both unweighted and weighted with uncertainties which are approximated by the root mean square of model variability and measurement variability~~, where~~. Here model variability is represented by the standard deviation from the averaged output values and measurement variability is by combination of instrumental errors and standard deviation~~(~~. As explained in further detail by Jöckel et al~~.,~~ . (2006~~). With~~), with this approach, the measurement locations with high variability have less weight, whereas more weight is given to stable, homogeneous conditions. This allows us to compare values that are more representative for the average conditions and to eliminate specific episodes that cannot be expected to be reproduced by the model. ~~According~~For this reason the weighted correlations are also generally expected to be higher than the normal correlations.

Also according to this analysis, the discrepancies between model results and measurements are smaller than the uncertainties~~,~~ if the absolute value of the weighted bias (i.e., in units of the normalised standard deviation, Table 4) for a specific tracer is less than one. A high weighted correlation in combination with a weighted bias between [-1,1] indicates that the model is able to reproduce the observed mixing ratios on average. This holds for all versions for CO, O_3, CH_2O, NO_2, and HNO_3, while model versions have more difficulties with CH_3OOH. For SO_2 CBA is the only model version to deliver a ~~satisfactory~~weighted bias that is larger than -1. For C_2H_4 and C_2H_6 none of the versions are able to match the observations to an acceptable degree. 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. The inability of the model versions to reproduce the observed magnitude of C_2H_6 and the vertical distribution of C_2H_4, as indicated by the relatively low correlation with all aircraft measurements included in the database, requires a more detailed analysis. This is investigated in more detail in the next sections.

Table 4. Summary of the Bias and correlation coefficients (~~R~~in terms of explained variance, R^2) of model results versus all available aircraft observations, also weighted with relative uncertainties. Bias = model results minus observations. Bias^a ~~and R^{2a}~~ is given in pmol/mol, (~~except~~nmol/mol for CO and O_3~~).~~), Bias^b ~~and R^{2b} are given~~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).

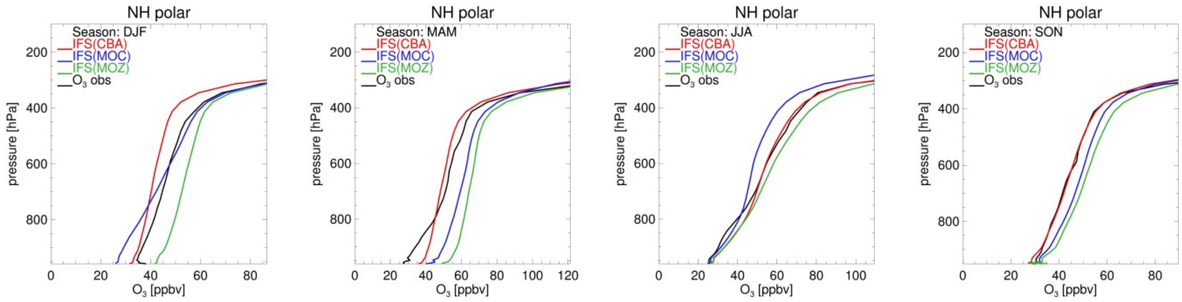
		CBA				MOC				MOZ			
Tracer	N. obs	Bias^a	Bias^b	R^{2a}	R^{2b}	Bias^a	Bias^b	R^{2a}	R^{2b}	Bias^a	Bias^b	R^{2a}	R^{2b}
O_3*	506	10.6	0.32	0.57	0.60	10.1	0.40	0.59	0.65	15.9	0.71	0.58	0.71
CO*	457	-2.11	0.35	0.22	0.88	-14.7	-0.43	0.21	0.86	-14.1	-0.38	0.21	0.89
CH_2O	213	-13.7	-0.11	0.63	0.76	20.1	0.31	0.67	0.72	24.3	0.26	0.70	0.80
CH_3OOH	366	-46.5	-0.47	0.58	0.93	51.4	0.15	0.69	0.88	-114	-0.92	0.74	0.96
C_2H_4	454	-6.28	-4.80	0.58	0.39	-5.35	-2.78	0.54	0.03	-4.02	-13.8	0.54	0.06
C_2H_6	473	-505	-3.18	0.50	0.81	-562	-3.90	0.44	0.77	-524	-3.50	0.46	0.79
NO_2	264	6.09	0.24	0.34	0.98	49.9	0.39	0.27	0.98	8.89	-0.24	0.33	0.99
HNO_3	416	-45.3	-0.32	0.40	0.86	-14.3	-0.12	0.38	0.83	-49.7	-0.34	0.43	0.90
SO_2	350	-17.0	-0.63	0.18	0.87	-48.7	-2.25	0.16	0.95	-31.2	-1.20	0.49	0.88

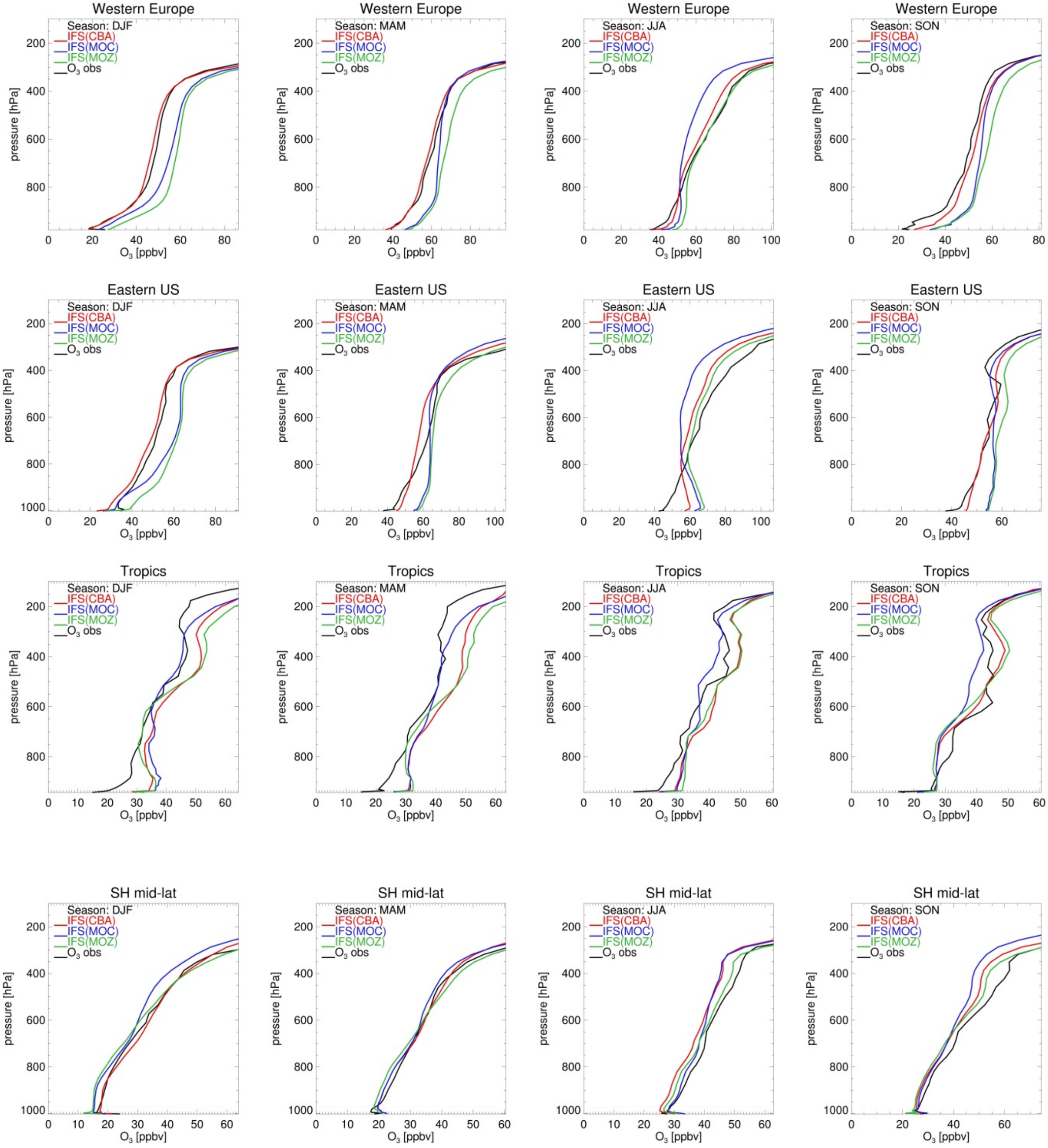
~~* Bias^a for CO and O_3 is given in units nmol/mol.~~

5.1. Ozone (O_3)

Figure 4 ~~compare~~compares tropospheric O_3 profiles simulated by the three model versions with ozone sonde observations for six different regions over the four seasons.~~ Figure 5 shows annually averaged model biases for various latitude bands and for altitude ranges 900-700hPa, 700-500hPa and 500-300hPa (Figure 5).~~ Overall the three chemistry versions deliver similar performance, reproducing the regionally averaged variability in O_3 observations, with various biases depending on the season, region and altitude range. Typically, the model versions tend to simulate lower O_3 mixing ratios in the SH mid and high-latitudes compared to sonde observations, and higher in the tropics. Over the Arctic, Western Europe, Eastern US and Tropics, MOZ simulates too high O_3 concentrations at all altitudes and for all seasons except in June-July-August (JJA), with average positive biases ranged from 1 to 12 ppbv in the free troposphere. 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 O_3 concentrations particularly in the NH.

MOC shows positive biases over the NH mid latitudes during winter and spring and negative biases during Arctic winter in the lower troposphere (<700hPa) as well as in the 700-300hPa range in summer. CBA simulates O_3 mixing ratios that are generally in close agreement with observations over the Arctic and NH mid-latitudes, but negative biases up to 10 ppbv are obtained in the Arctic upper troposphere (500-300hPa) during winter time (Figure 4, top panel). All three model versions are consistently too high close to the surface (> 800hPa) over the tropics for all seasons, but particularly during December-January-February (DJF). Over the Antarctic and, to a lesser extent, the ~~Southern Hemisphere (~~SH~~)~~ mid-latitudes all three model versions underestimate O_3, with negative biases up to 10 ppbv for a large part of the year. However, it should be noted that in the SH regions this evaluation is less representative because there are very few observations.




Western Europe
Season: DJF
Season: MAM
Season: JJA
Season: SON
IFS(CBA)
IFS(MOC)
IFS(MOZ)
O_3 obs
pressure [hPa]
O_3 [ppbv]
Eastern US
Tropics
SH mid-lat

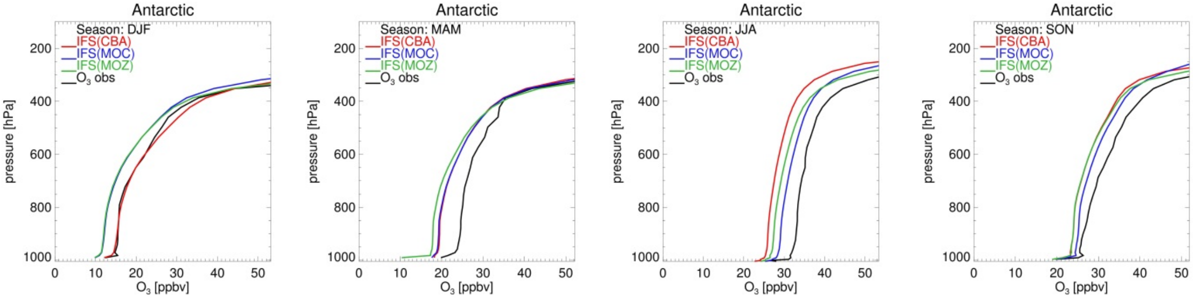


Figure 5: Tropospheric ozone profiles ~~of volume mixing ratios (ppbv) against by~~ from model versions CBA (red), MOC (blue) and MOZ (green) against sondes (black) in volume mixing ratios (ppbv) over six different regions: from top row to bottom row, NH-Polar [90°N-60°N], Western Europe [45°N-54°N; 0°E-23°E], Eastern US [32°N-45°N; 90°W-65°W], Tropics [30°N-30°S], SH mid-latitudes [30°S-60°S] and Antarctic [60°S-90°S], averaged over four seasons (from left to right: December-January-February, March-April-May, June-July-August, September-October-November).

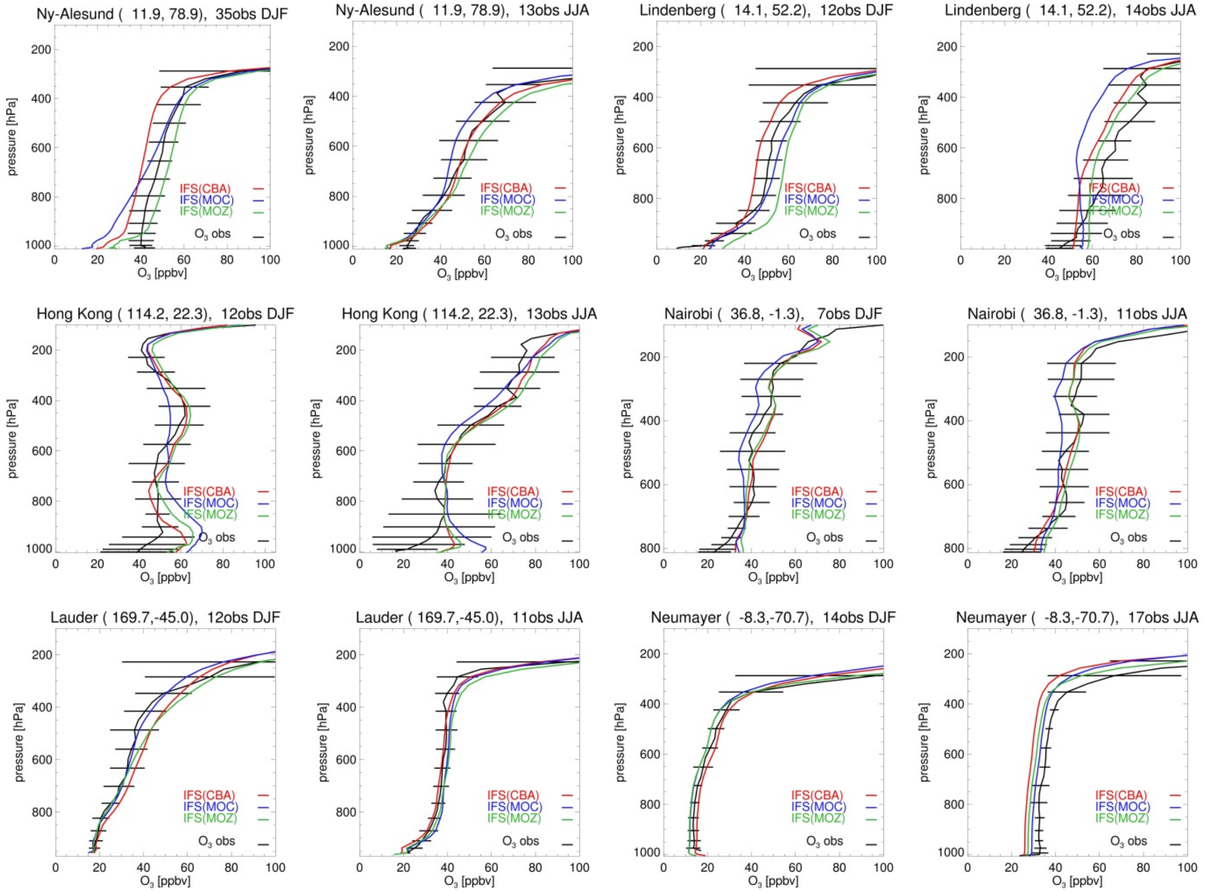


Figure 6: Mean tropospheric ozone profiles from model versions CBA (red), MOC (blue) and MOZ (green) against sondes (black) in volume mixing ratios (ppbv) during DJF and JJA at selected individual stations. Error bars represent the 1σ spread in the seasonal mean observations.

Figure 5 shows an evaluation of O_3 profiles against sondes at selected individual WOUDC sites representative of the Arctic (Ny-Alesund), NH mid-latitudes (Lindenberg), Tropics (Hong Kong, Nairobi), SH mid-latitudes (Lauder) and Antarctic (Neumayer) for DJF and JJA seasons in 2011. We note generally similar biases as compared to those for the regional averages, even though local conditions play a larger role explaining the different performance statistics for these stations. Overall, the evaluation at individual station provides reasonable agreement between model simulations and sondes.

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 O_3 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 N_2O_5 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).

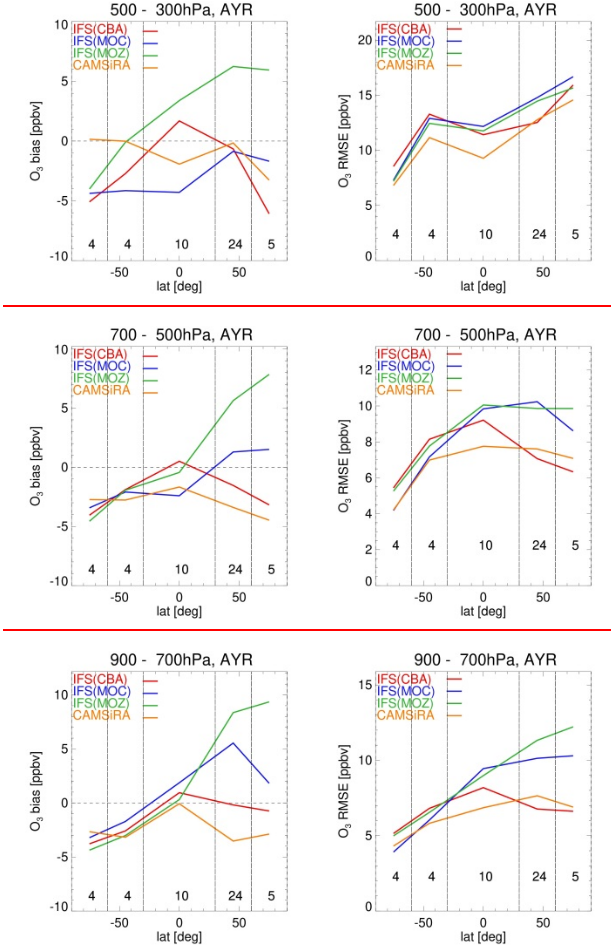


Figure 7: Mean of all model biases (left) and RMSE (right) values against ozone sondes as function of latitude for various pressure ranges (top row: 300-500hPa; middle row: 500-700hPa; bottom row: 700-900hPa), averaged over the full year. Same color codes as in the previous figure. The numbers in each latitude range indicate the amount of stations that contribute to these statistics. For reference, also the corresponding results from the CAMS-Interim Reanalysis (CAMSiRA) are given in orange.

5.2. Carbon Monoxide (CO)

Carbon monoxide is a key tracer for tropospheric chemistry, as a marker of biomass burning and anthropogenic pollution, and provides the most important sink for OH. Approximately half of the CO burden is directly emitted, and the rest formed through degradation of ~~methane~~CH_4 and other ~~VOC's.~~ VOCs (Hooghiemstra et al., 2011). Hence, a correct simulation of this tracer is very important for studies of atmospheric oxidants. Considering the use of the same emissions and CH_4 surface conditions, differences in CO concentrations are essentially caused by differences in chemistry.

Figures 7 and 8 show the monthly mean evaluation against MOPITT total CO columns for April and August 2011. Whereas generally the model versions show good agreement with the observations in terms of their spatial patterns, persistent seasonal biases remain, such as the ~~low~~negative bias over the NH during April (further analysed in, e.g., Shindell et al., 2006; Stein et al., 2014), as well as a ~~low~~negative bias over Eurasia during August. For all three chemistry versions the patterns of enhanced CO in the tropics, associated ~~to~~with biomass burning, are generally well captured, as well as the magnitude of CO columns over the SH. Looking at differences between model versions, CBA shows overall the highest magnitudes, implying a smaller negative bias over the NH particularly during April, while this simultaneously results ~~into~~in an emerging positive bias in the tropics.

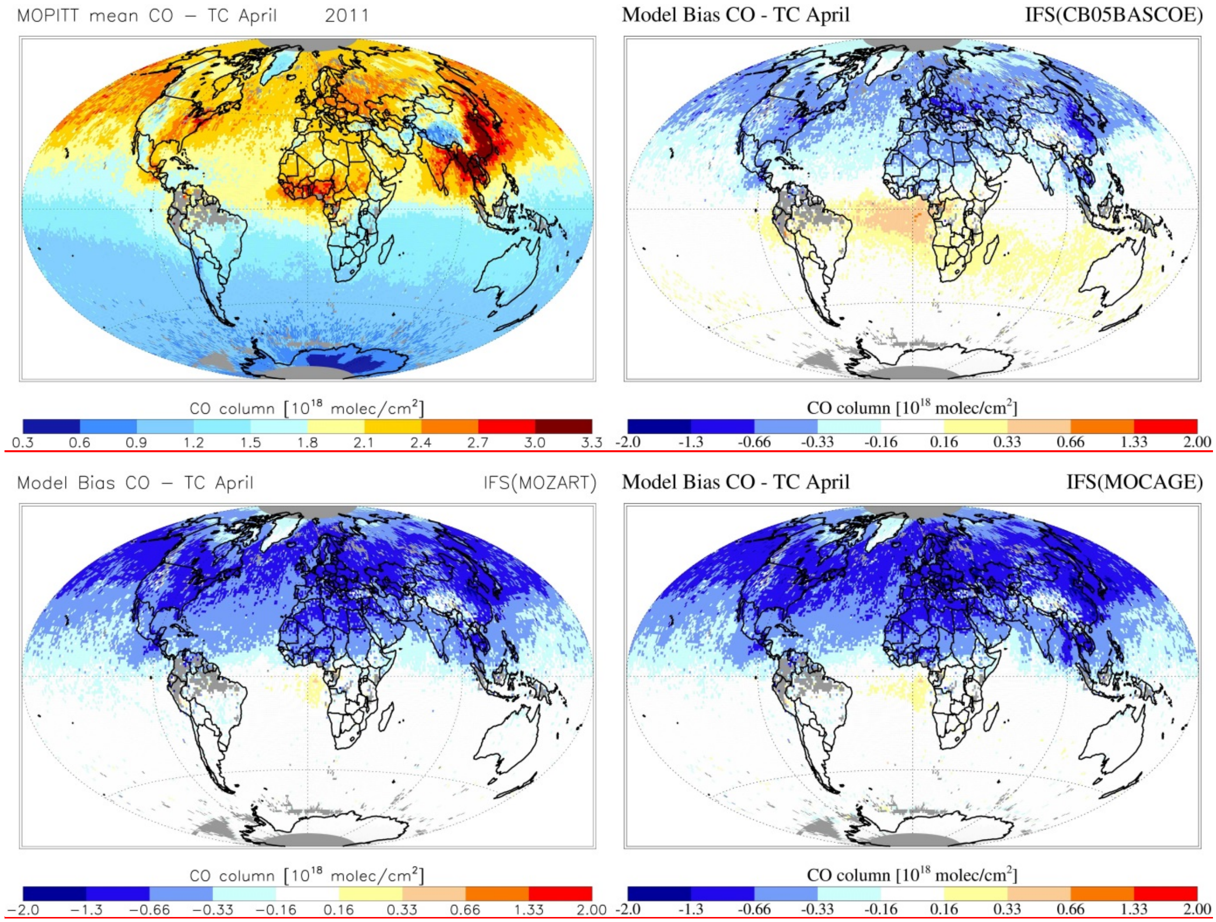


Figure 8: MOPITT CO total column retrieval for April 2011 (top left) and simulated by IFS(CBA) (top right), IFS(MOZ) (bottom left) and IFS(MOC) (bottom right).

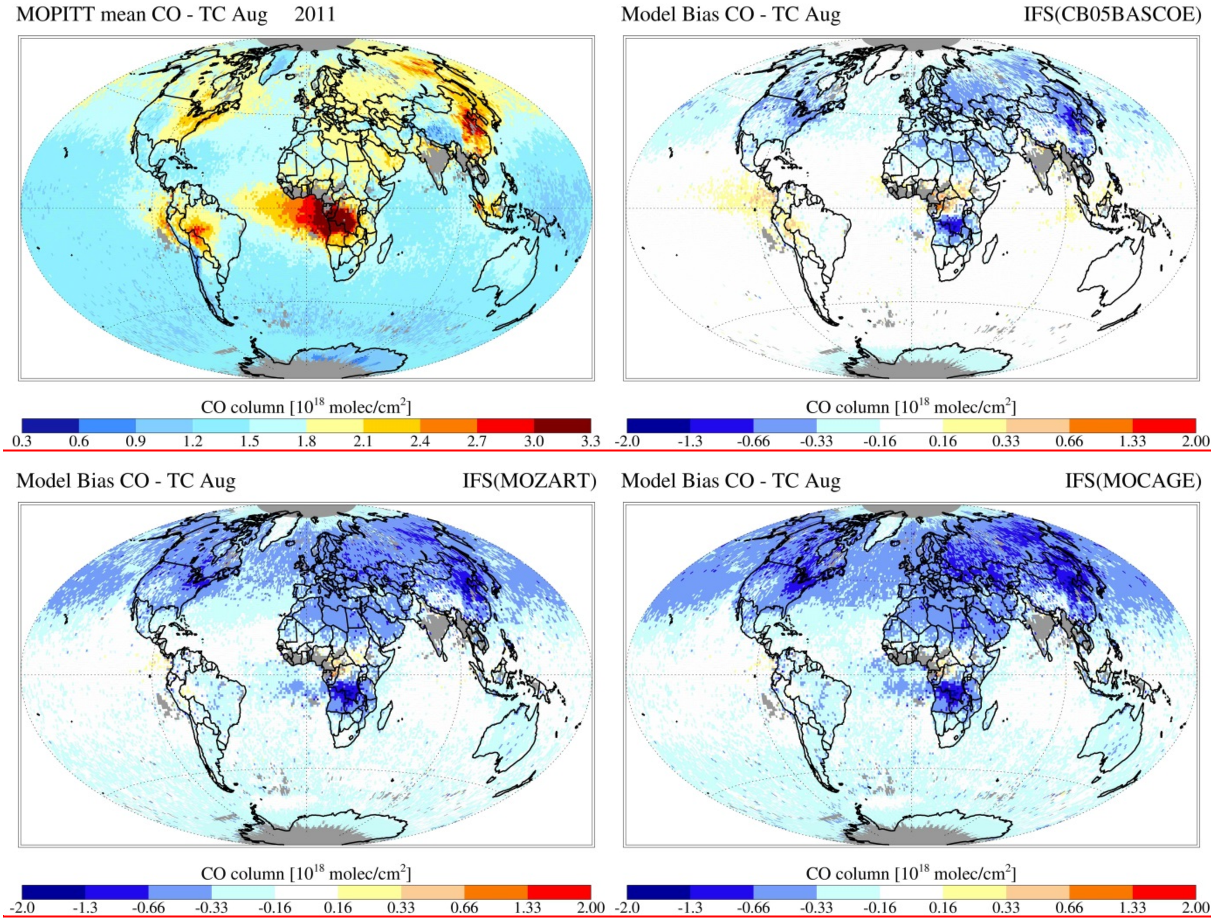


Figure 9: MOPITT CO total column retrieval for August 2011 (top left) and simulated by IFS(CBA) (top right), IFS(MOZ) (bottom left) and IFS(MOC) (bottom right).

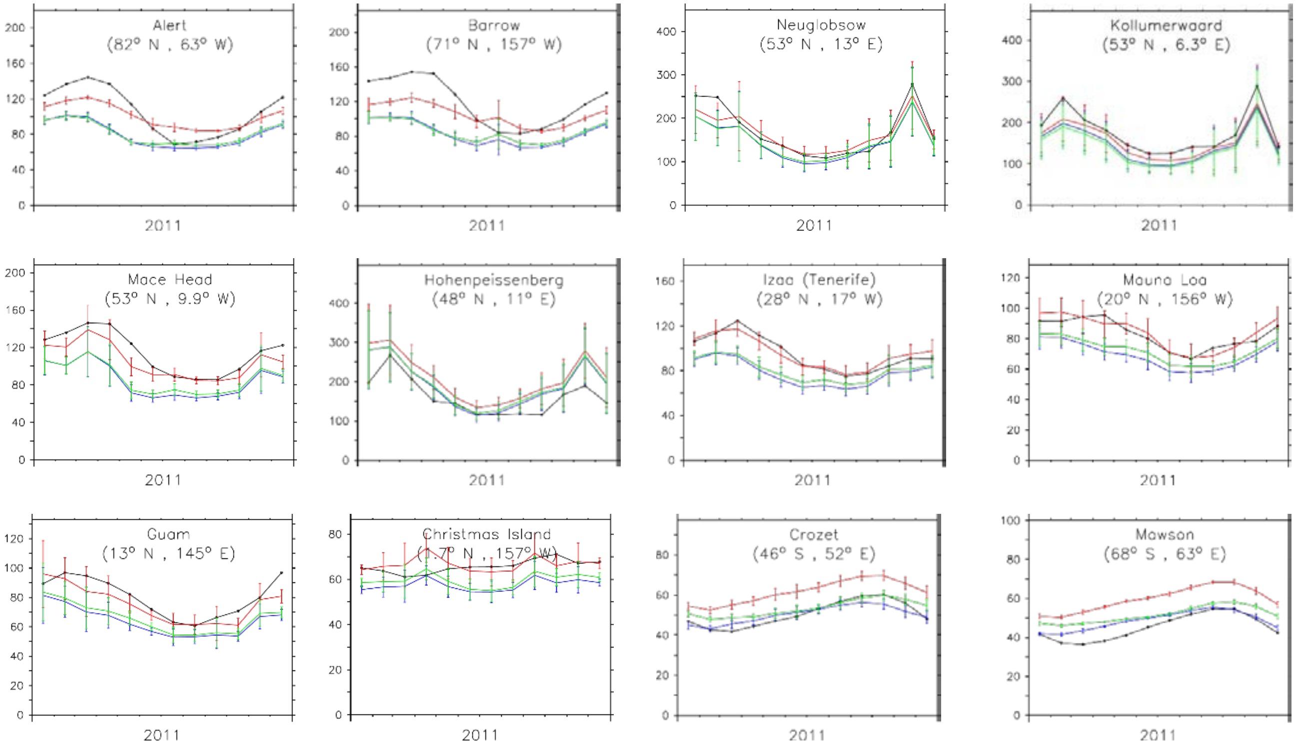


Figure 10: Comparison of CO mixing ratios (ppbv) at the surface as simulated (red, blue and green are model results from CBA, MOC and MOZ, respectively) and observed (black) ~~CO mixing ratios in nmol/mol at the surface.~~at twelve stations sorted by decreasing latitudes. The bars represent one-standard deviation of the monthly average for the location of the station.

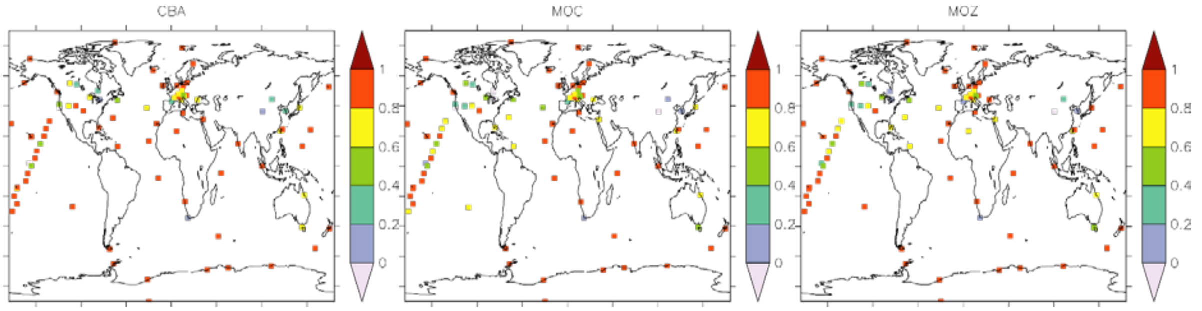


Figure 11: Temporal correlation (R^2) between monthly mean surface CO ~~by~~as derived from observations (GAW network) and ~~by the~~ model simulations (left: CBA, middle: MOC, right: MOZ).

In Figure 9 the annual cycle at selected GAW stations are shown while in Figure 10 additionally shows the corresponding temporal correlation between the simulated monthly mean CO for all stations. Even though the ~~phae~~phase and amplitude of the annual cycle are well reproduced by the model versions at several locations (e.g., Mauna Loa, Hawaii), the concentrations tend to be overestimated in the ~~Southern Hemisphere~~southern hemisphere, particularly by CBA, and to a lesser extent by the other chemistry versions, and underestimated over the remote Northern Hemisphere. This points to sensitivities due to the applied chemistry scheme mainly associated ~~to~~with differences in OH, which is lowest in CBA and highest in MOC (see also Sec. 4). A possible over-estimation of CO over the tropics and southern hemisphere could relate to uncertainties in the biogenic emissions (Sindelarova et al., 2014).

The correlations (in terms of R^2) of monthly mean time series against GAW stations are mostly above 0.8. Particularly over Antarctica the correlation is very high with $R^2 \approx 0.9$, indicating that indeed the main processes controlling the CO abundance are well represented by the model. Nevertheless, at locations between 40°N and 60°N the correlation is lower. These regions are strongly influenced by local chemistry and emissions, including industry and biomass burning. Clearly, the seasonal cycle is not optimally reproduced in Northern America (Canada regions) by any of the three chemistry versions, indicating that uncertainties in regional emissions, such as boreal biomass burning, could be responsible for these disagreements.

Compared to aircraft observations (see Figure 11), the three model versions produce similar CO mixing ratio vertical profiles, with differences among them typically within the range of 10~~%.~~-20%, depending on the location. The biomass burning plumes are reproduced consistently (see Figure 11, TRACE-A, West Africa coast), and all three models compare well with observations both for background conditions in the Northern Hemisphere (SONEX, Ireland) and highly polluted condition (PEM-West-B, China Coast).

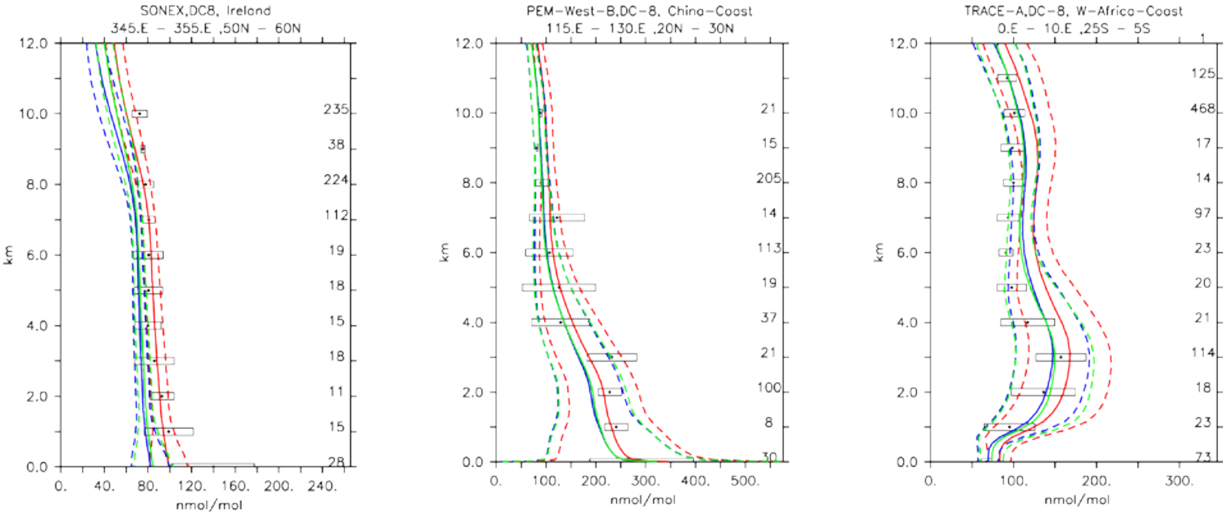


Figure 12: Comparison of simulated CO vertical profiles by using the CBA (red solid line), MOC (blue solid line) and MOZ (green solid line) chemistry versions against aircraft data (black dots). Also shown are the modeled (dashed lines) and

measured (black rectangular) standard deviations. The numbers on the right vertical axis indicate the number of available measurements.

5.3. Formaldehyde (CH_2O) and methyl hydroperoxide (CH_3OOH)

Formaldehyde is important as one of 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~~ VOCs 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. A good agreement of the simulations with the observations can be seen from Figure 12, where the vertical profile from selected aircraft observations and model simulations are shown. Also from Table 4 it is clear that all the three model versions do reproduce formaldehyde accurately. The weighted bias always well below 1 standard deviation unit (i.e. -0.11, 0.31 and 0.26 for CBA, MOC and MOZ, respectively), indicating that the simulations are well within the statistical uncertainties.

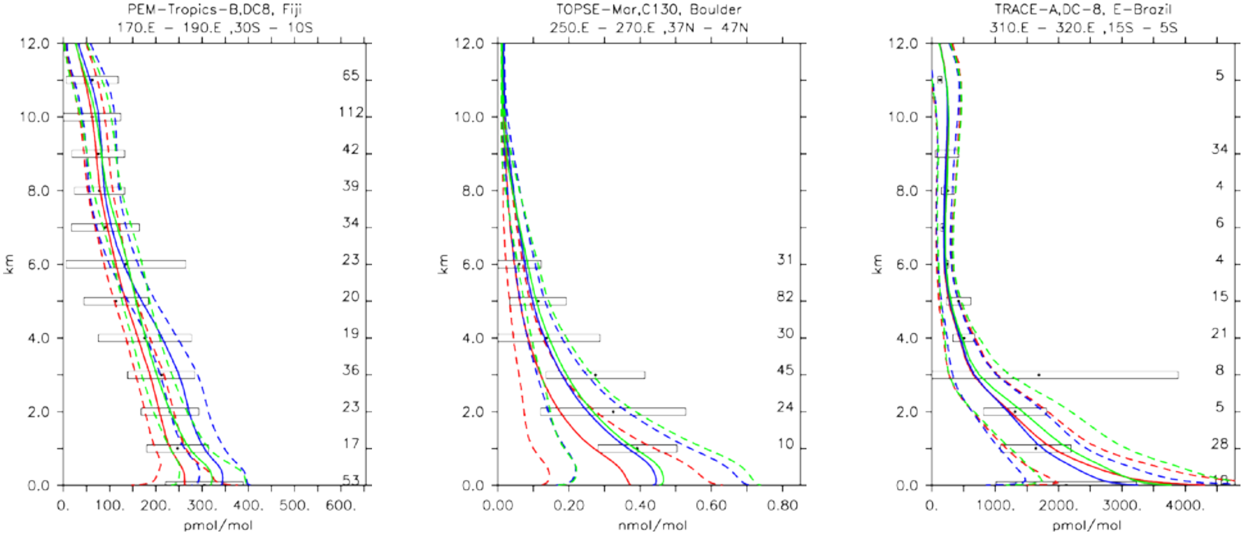


Figure 13: Comparison of simulated CH_2O vertical profiles by using the CBA (red), MOC (blue) and MOZ (green) chemistry versions against aircraft data (black~~), see also caption of~~). Line styles and symbols have the same meaning as in Figure ~~11~~12.

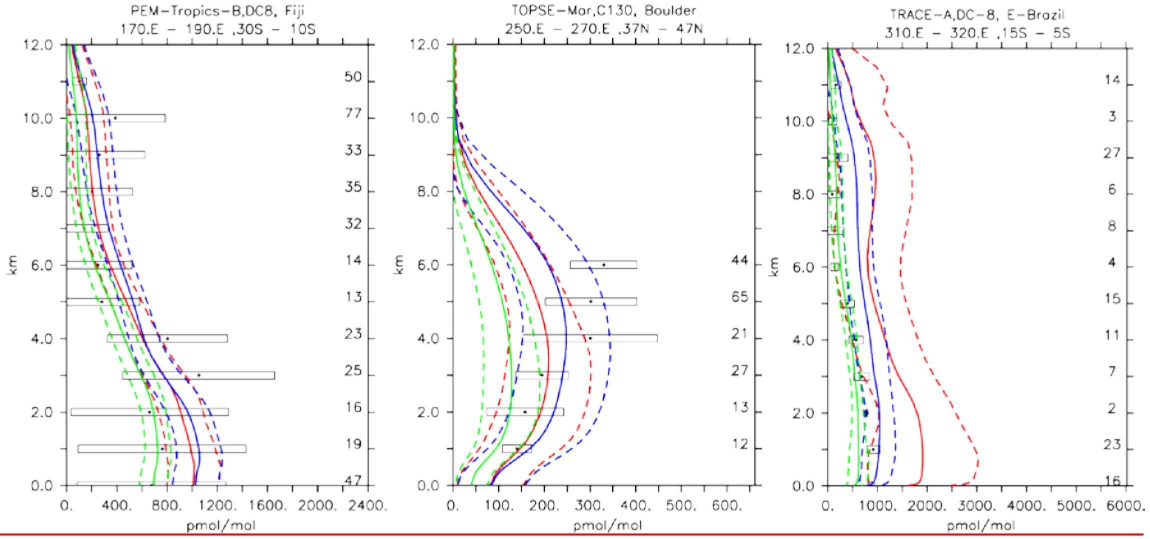


Figure 14: Comparison of simulated CH_3OOH vertical profiles by using the CBA (red), MOC (blue) and MOZ (green) chemistry versions against aircraft data (black). Line styles and symbols have the same meaning as in Figure 11.

CH_3OOH is a main organic peroxide acting as a temporary reservoir of oxidizing radicals, Zhang et al. (2012). It is mainly formed through reaction of CH_3O_2 + HO_2, which are both produced in the oxidation process of many hydrocarbons. The CH_3OOH lifetime of globally about one day is mainly governed by its reaction with OH, and photolysis. Figure 13 presents an evaluation for CH_3OOH for the same sites are presented for CH_2O 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 CH_2O (a few hours in daytime) and also CH_3OOH, 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.

5.4. Ethene (C_2H_4)

Ethene is the smallest alkene which is primarily emitted from biogenic sources. In our configuration, biogenic C_2H_4 emissions are 30 Tg yr^{-1}, which appears at the upper end of such emission estimates as reported by Toon et al. (2018). The rest of the emissions are attributed to incomplete combustion from biomass burning or anthropogenic sources.

The three chemical mechanisms produce mostly very similar mixing ratios of C_2H_4. Nevertheless, as indicated by the bias (Table 4), which ranges between -2 and -14 in standard deviation units, as well as the weighted correlations, the model versions have difficulties in simulating C_2H_4. ~~The~~Even though this evaluation should only be considered in a climatological sense, the vertical profiles (see Figure 13) are strongly biased (e.g., SONEX, Newfoundland and PEM-Tropics-A, Tahiti), with positive biases occurring at the surface and negative in the free troposphere. In remote regions and at higher altitudes, where the direct influence of emissions is lower, the model is at the lower end of the range of observations, with frequent underestimates (see Figure 13 PEM-Tropics-A, Christmas Island). This was already observed in other studies (e.g. Pozzer et al. 2007), implying that the chemistry of this tracer is not well understood. As the underestimation appears to be ubiquitously distributed this suggests that C_2H_4 decomposition is too strong, or that the model versions miss some chemical production ~~term~~terms (e.g., Sander et al., 2018).

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. CBA indeed shows the largest values for C_2H_4, which is explained by the comparatively low abundance of OH in this model version.

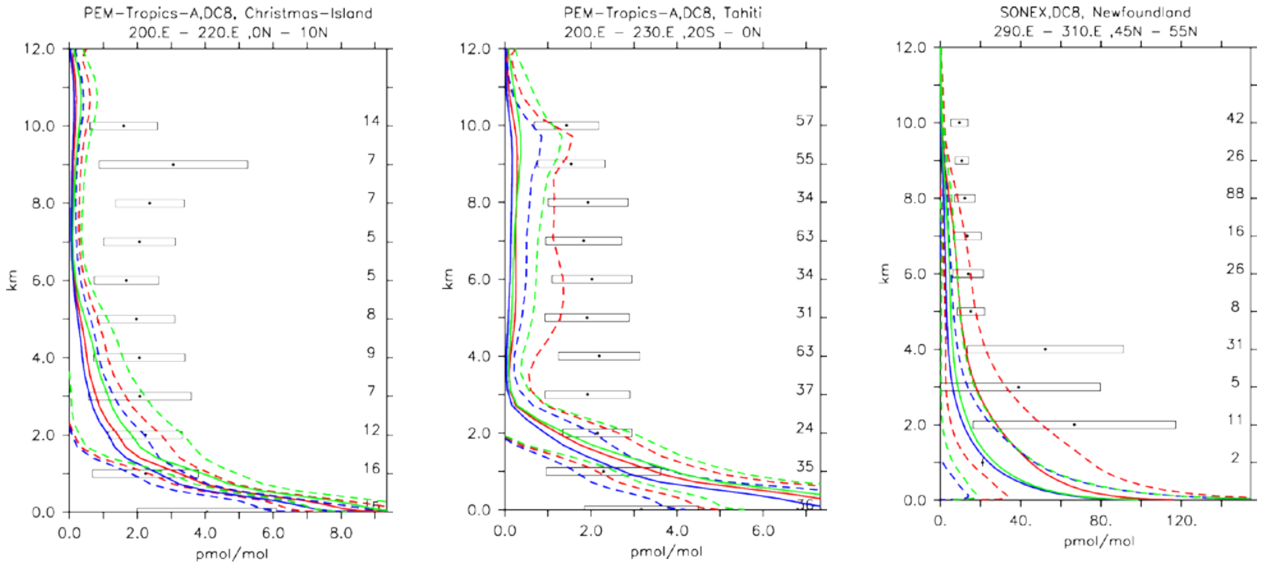


Figure 15: Comparison of simulated C_2H_4 vertical profiles by using the CBA (red), MOC (blue) and MOZ (green) chemistry versions against aircraft data (black~~), see also caption of~~). Line styles and symbols have the same meaning as in Figure ~~11~~12.

5.5. Ethane (C_2H_6)

Ethane (C_2H_6) is the lightest trace gas of the family of alkanes and has an atmospheric lifetime of about two months. Ethane emissions are primarily of anthropogenic nature, and have seen a relatively strong decrease since the 1980s (Aydin et al., ~~2014~~2011). Nevertheless, since 2009 an increase in C_2H_6 concentrations has been observed, believed to be associated ~~to~~with recent increases in CH_4 fossil fuel extraction activities (Hausmann et al., 2016, Monks et al., 2018).

Compared to aircraft observations, all three model versions significantly underestimate the C_2H_6 observed mixing ratios at all locations and ubiquitously (see Figure 14). A particularly strong underestimation is found in the Northern Hemisphere, where most of the observations are located (e.g. the SONEX campaign over Ireland). A strong negative bias was also reported in the overall statistics (Table 4), even though, contrarily to C_2H_4, the weighted correlation showed acceptable values for all versions ($R^2>0.7$). These findings can well be explained by an underestimation of the MACCity-based C_2H_6 emissions, which are at least a factor two lower than the corresponding estimates of 12-17 Tg yr^{-1} reported in the literature (Monks et al., 2018, Aydin et al., ~~2014~~2011, Emmons et al., 2015; and Folberth et al., 2006). On the other hand, the comparison with the TRACE-A field campaign, which covered long-range transport of biomass burning plumes, shows a reasonable agreement in the lower troposphere (1-4 km), i.e. at the location of the biomass plume, suggesting appropriate biomass burning emissions. Still a considerable underestimation is present in the upper troposphere, probably due to the missing background concentration.

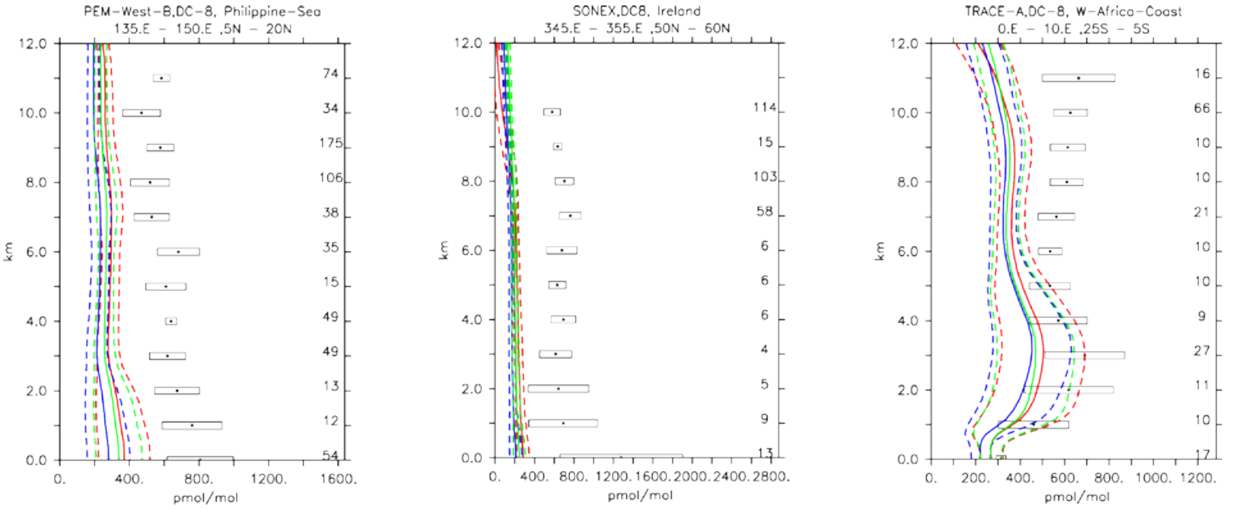


Figure 16: Comparison of simulated C_2H_6 vertical profiles by using the CBA (red), MOC (blue) and MOZ (green) chemistry versions against aircraft data (black~~), see also caption of~~). Line styles and symbols have the same meaning as in Figure ~~11~~12.

5.6. Nitrogen Dioxide (NO_2)

Nitrogen dioxide is a trace gas difficult to compare with in-situ observations, due to its photochemical balance with ~~nitrogen~~nitric oxide. ~~Considering its~~Nitrogen dioxide shows a strong diurnal cycle, mainly due to the fast photolysis rate~~,~~ ~~here~~. Here only daytime values have been used to construct the model averages, because the observations from the various field campaigns were equally conducted in daylight conditions. Figure 15 shows the strong variability in daytime NO_2 values, both in the measurements and in the simulations. In general the MOC simulation shows the highest concentration of NO_2 in different locations, particularly over source regions (see Figure 15, TRACE-P, Japan and TOPSE-Feb, Boulder), with MOZ and CBA being more similar. This is in line with the analysis given in Sec. 4. Outside the source regions the secondary processes (such as its equilibrium with HNO_3, see also next section) have larger influences, hence the model and observation profiles of NO_2 show even stronger variability and larger differences (see Figure 15, TOPSE-May, Thule). Still, in general all the chemical mechanisms are able to reproduce NO_2 within 1 standard deviation (see Table 4), even though the unweighted mean bias for MOC is significantly higher than for CBA and MOZ.

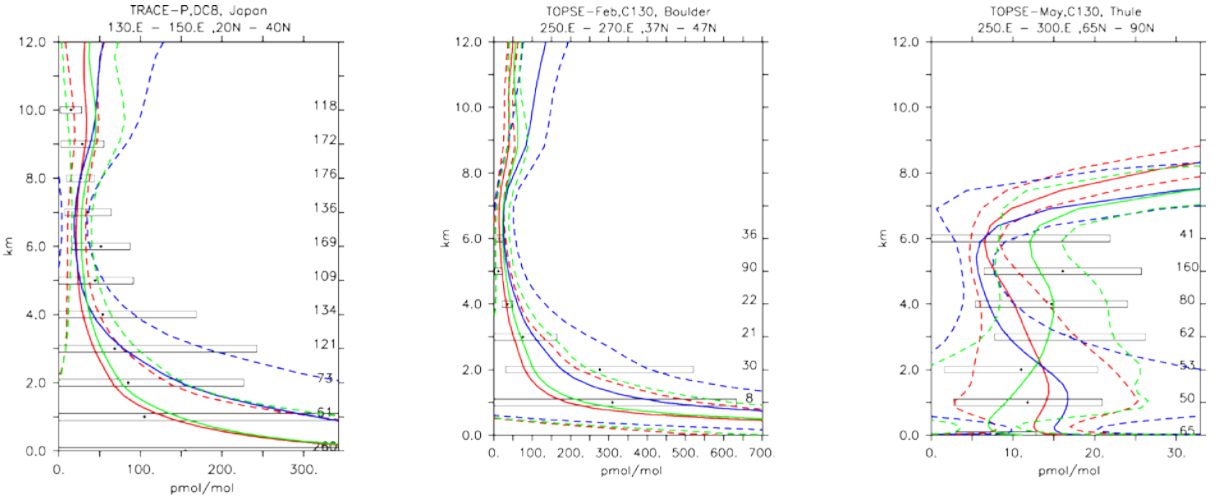


Figure 17: Comparison of daytime NO_2 vertical profiles simulated by CBA (red), MOC (blue) and MOZ (green) chemistry versions against aircraft data (black~~), see also caption of~~). Line styles and symbols have the same meaning as in Figure ~~11~~12.

Figures 16 and 17 evaluate tropospheric NO_2 using the OMI satellite observations. The simulations deliver generally appropriate distributions with a correct extent of the regions with high pollution, as largely dictated by the emission patterns. Nevertheless, a general underestimation of NO_2 over West Africa in April, and Central Africa and South America in August is found, suggesting uncertainties associated ~~to~~with the modelling of biomass burning emissions.

Another interesting finding is a relatively strong negative bias over the Eurasian and North American continents in April for CBA, stronger than modelled in MOZ and MOC. In contrast, particularly MOC, but also MOZ over-estimates NO_2 over the comparatively clean North Atlantic and North Pacific oceans in April. This all suggests a relatively short NO_x lifetime in CBA compared to MOZ and MOC, which in turn helps to explain the lower O_3 over the NH-mid latitude regions as modelled with CBA (see Figure 5). The causes of these differences in modelled NO_2 are mainly the use of a different numerical solver and differences in the efficiency assumed for N_2O_5 heterogeneous reactions (see Sec. 2.1.4). In August the differences in tropospheric NO_2 between the three model versions are smaller than in April.

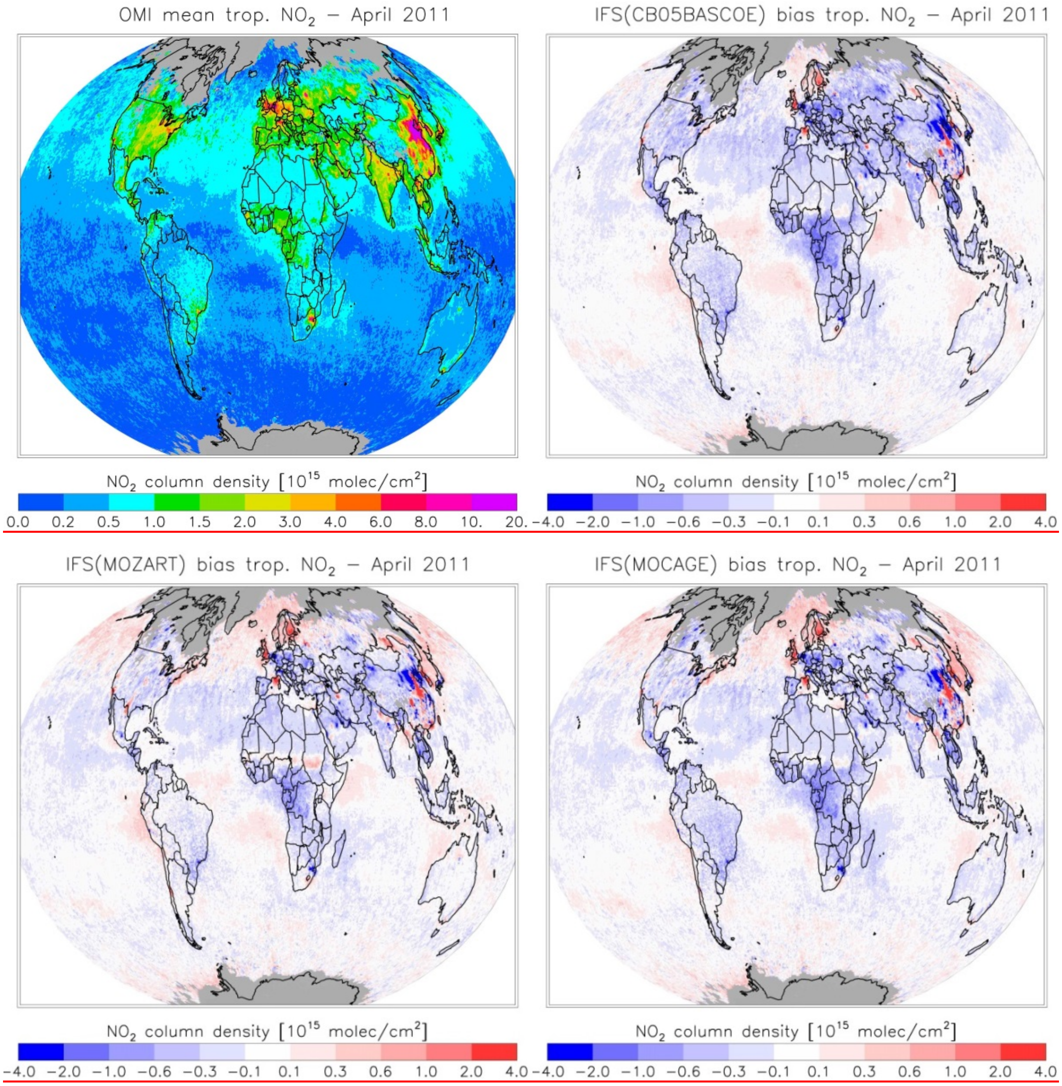


Figure 18~~:~~: Monthly mean Tropospheric NO_2 columns from OMI satellite retrievals from the QA4ECV product for April 2011, along with the corresponding collocated model ~~values~~biases.

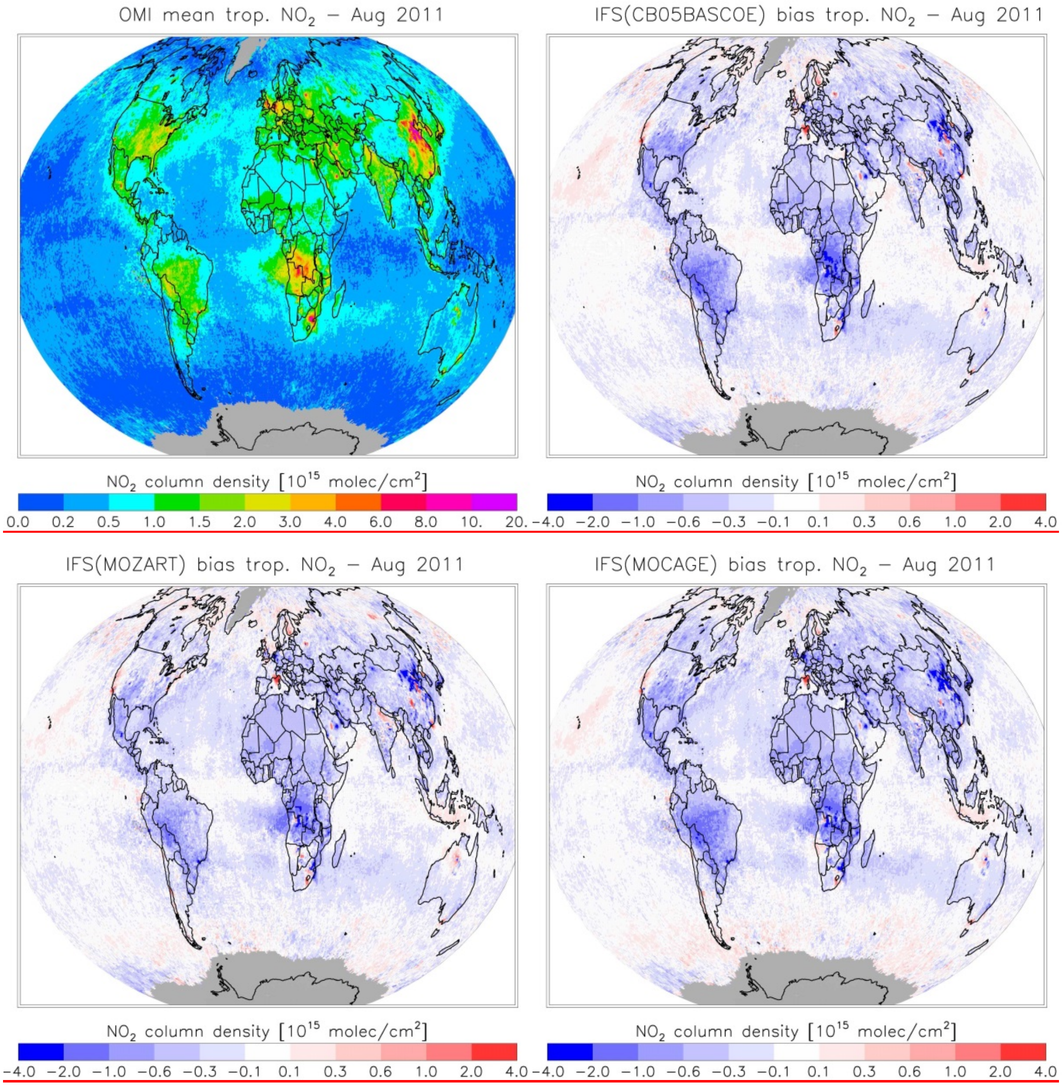


Figure 19~~:~~: Monthly mean Tropospheric NO_2 columns from OMI satellite retrievals from the QA4ECV product for August 2011, ~~and the~~along with the corresponding collocated model biases~~ with respect to this~~.

5.7. Nitric Acid (HNO_3)

Compared to several of the trace gases previously analysed, ~~Nitric~~nitric acid is not primary emitted but is purely photochemically formed in the atmosphere. 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). HNO_3 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. In Figure 18, the model results are compared with selected aircraft measurements. Although all three models tend to reproduce HNO_3 in a statistically similar way, over the lower troposphere and up to 2 km height MOC tends to result in higher HNO_3 concentrations compared to the other two chemical mechanisms and measurements. This is also reflected by overall the lowest negative biases in Table 4. While MOC performs better at higher altitudes, in a biomass burning plume (e.g. TRACE-A, Figure 18), it also overestimates the production of HNO_3 or underestimates its sinks. Over polluted regions (Figure 18, TRACE-P, Japan), all models tend to perform well but in remote areas (Figure 18, TOPSE, Churchill) the discrepancies between the models increase with MOC delivering twice more HNO_3 than the other two model versions. Nevertheless, as the variability of the observations is very large, all the model versions still fall within the range of uncertainties of the observations. 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.

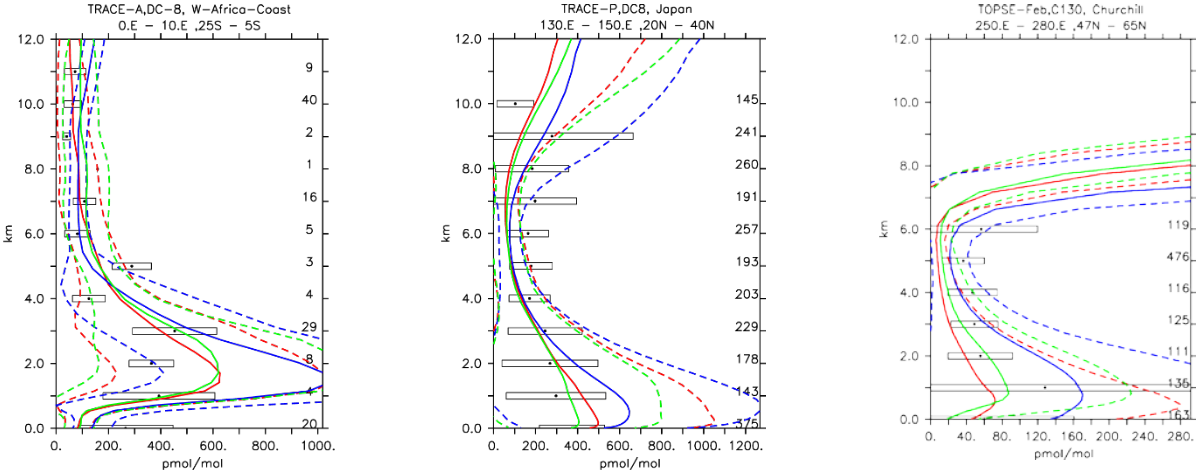


Figure 20: Comparison of simulated HNO_3 vertical profiles by using the CBA (red), MOC (blue) and MOZ (green) chemistry versions against aircraft data (black~~), see also caption of~~). Line styles and symbols have the same meaning as in Figure ~~11~~12.

5.8. Sulfur Dioxide (SO_2)

Similar to HNO_3, SO_2 is also strongly influenced by wet deposition due to its high solubility. Furthermore, SO_2 is primarily emitted and converted to sulfuric acid (H_2SO_4) both by gas phase and aqueous phase oxidation, an essential process for the production of new sulfate aerosol particles. Considering the complexity of the processes that control the SO_2 fate in the atmosphere, a large variability is expected for this tracer. The evaluation of SO_2 shows that among the three chemistry versions, CBA produces always the highest SO_2 mixing ratios, whereas MOC produce the lowest, and MOZ lies always in between. Nevertheless, all three mechanisms tend to underpredict SO_2 mixing ratios (see Table 4) compared to the aircraft observations (see Figure 19). Notwithstanding significant uncertainties regarding SO_2 emissions, the simulated mixing ratios over polluted regions seem to reproduce the observed values (Figure 19, Trace-P, China and Japan). CBA presents the best comparison with aircraft observations, as can be seen in Figure 19 for the TOPSE aircraft measurements. Also from Table 4, only CBA delivers a normalized weighted bias within [-1, 1] for SO_2, while for the other model versions these are below -1 (-2.25 and -1.20 for MOC and MOZ, respectively).

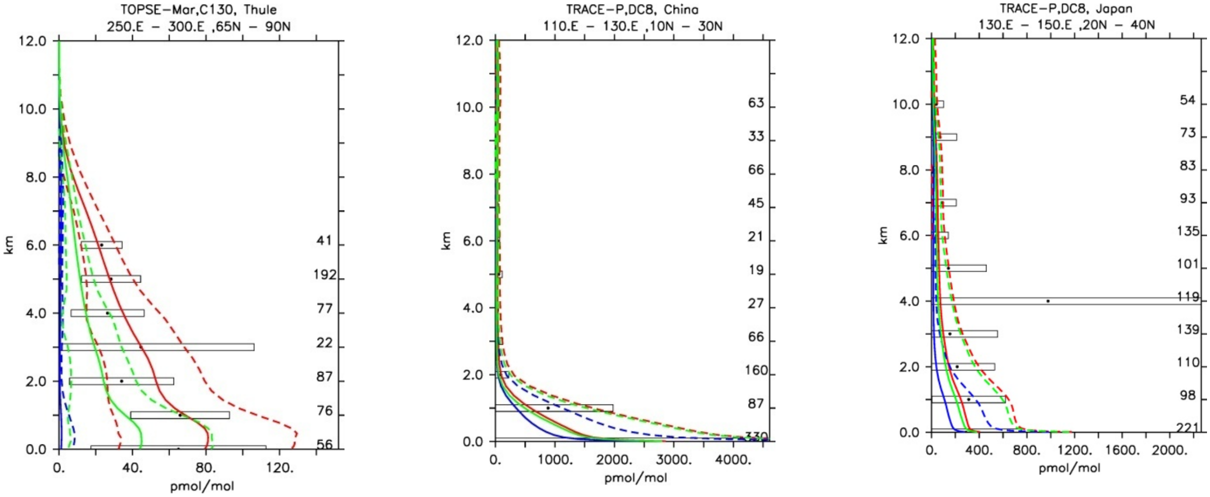


Figure 21: Comparison of simulated SO_2 vertical profiles by using the CBA (red), MOC (blue) and MOZ (green) chemistry versions against aircraft data (black~~), see also caption of~~). Line styles and symbols have the same meaning as in Figure ~~11~~12.

Conclusions

We have reported on an extended evaluation of tropospheric trace gases as modelled in three largely independent chemistry configurations to describe ozone chemistry, as implemented in ECMWF's Integrated Forecasting System~~.~~ of cycle 43R1. These configurations are based on IFS(CB05BASCOE), IFS(MOZART) and IFS(MOCAGE) chemistry versions. While the model versions were forced with the same overall emissions and adopt the same parameterizations for transport and dry and wet deposition, they largely vary in their parameterizations describing atmospheric chemistry. In particular their VOC degradation, treatment of heterogeneous chemistry and photolysis, and the adopted chemical solver vary strongly across model versions. Therefore this evaluation provides a quantification of the overall model uncertainties in the CAMS system for global reactive gases which are due to these chemistry parameterizations, as compared to other common uncertainties such as emissions or transport processes.

Overall the three chemistry versions implemented in the IFS produce similar patterns and magnitudes for CO, O_3, CH_2O, C_2H_4 and C_2H_6. For instance, the averaged differences for O_3 (CO) ~~is~~are within 10% (20%) throughout the troposphere, which is in line with larger model intercomparison studies reported in literature (Emmons et al., 2015; Huang et al., 2017). Except for C_2H_6 and C_2H_4, all these trace gases are also well reproduced by the various model versions, with an uncertainties-weighted bias always well within one standard deviation when compared to aircraft observations. Nevertheless the daily average OH levels may vary by up to 50% between the different simulations, particularly at high latitudes where absolute values are smaller. This may explain the larger model spread seen for C_2H_4. Comparatively large discrepancies between model versions exist for NO_2, SO_2 and HNO_3, because they are strongly influenced by parameterized processes such as photolysis, heterogeneous chemistry and conversion to aerosol through gas-phase and aqueous phase oxidation. For instance IFS(MOCAGE) tends to predict significantly higher NO_x and HNO_3 concentrations in the lower troposphere compared to the other two chemistry versions.

The comparison of the model simulations of NMHCs against a selection of aircraft observations reveals two major issues. First, the evaluation shows that large uncertainties remain in current and widely used emission estimates. For instance, the MACCity ethane emissions are likely under-estimated by at least a factor 2 (Hausmann et al., 2016; Monks et al., 2018) and were shown to lead to significantly lower C_2H_6 concentrations compared to the aircraft observations. Secondly, as has been shown before (Pozzer et al., 2007), the significantly lower C_2H_4 levels at high altitudes compared to measurements, even though C_2H_4 emissions appear in the right order of magnitude, suggest that the C_2H_4 chemistry is not well described. Other issues to constrain tropospheric ozone chemistry, as revealed from this assessment, are the model spread in NO_2, and its biases against observations. To handle the various discrepancies discussed here, several promising updates are being introduced in the three chemistry versions of IFS, specifically:

- Coupling of the heterogeneous reactions in the troposphere with CAMS-aerosol in IFS(MOCAGE),
- Implementations of more accurate solvers for atmospheric chemistry based on Rosenbrock (Sandu and Sander, 2006) or alternatively ASIS (Cariolle et al., 2017) in IFS(MOCAGE),
- Revisions in atmospheric chemistry scheme in IFS(MOZART) by revising assumptions in the heterogeneous chemistry, expending the complexity of the scheme with additional species, detailed aromatic speciation instead of lumped TOLUENE, and updated reaction products following recent developments in CAM-Chem,
- Update to the look up table for photolysis rate determination in IFS(MOZART),
- Updates of the reaction rate coefficients in any of the chemistry schemes to follow latest recommendations from IUPAC or JPL.

An update of the emission inventories is also foreseen for the near future. All these updates should tend to narrow the spread between the three model versions, and bring them closer to observations. This suggests that the present estimates of uncertainties in atmospheric chemistry parameterizations are on the conservative side. Still, the diversity of chemistry versions will be useful to provide a quantification of uncertainties in key CAMS products due to the chemistry module, as compared to other sources of uncertainties.

Author contributions

VH designed the study, contributed in the evaluations against sondes and satellite retrievals and wrote large parts of the manuscript. VH, SC, YC and JF developed the IFS(CB05BASCOE) chemistry module, VM, JA, TD, JG, BJ and SP developed the IFS(MOCAGE) chemistry module, IB and GB contributed to the development of the IFS(MOZART) chemistry module, AP and VK performed the evaluation against aircraft observations, and contributed to the writing.

Data Availability

The source code of the chemistry modules are integrated into ECWMF's IFS code, which is only available subject to a license agreement with ECMWF. The IFS code without modules for assimilation and chemistry can be obtained for educational and academic purposes as part of the openIFS release (https://confluence.ecmwf.int/display/OIFS). A detailed documentation of the IFS code is available from (https://www.ecmwf.int/en/forecasts/documentation-and-support/changes-ecmwf-model/ifs-documentation). The CB05 chemistry module of IFS was originally developed in the TM5 chemistry transport model. Readers interested in the TM5 code can contact the TM5 developers (http://tm5.sourceforge.net). The BASCOE stratospheric chemistry module can be freely obtained from the BASCOE developers (http://bascoe.oma.be). The MOCAGE chemistry module of IFS is developed at Météo-France on the basis of the MOCAGE chemistry-transport model, http://www.umr-cnrm.fr/spip.php?article128. The MOZART code can be obtained through contacting their developers via https://www2.acom.ucar.edu/gcm/mozart. The MOZART and CB05BASCOE chemistry schemes are also freely available through the Sander et al. (2019) publication.

The model simulation ~~data as~~datasets used in this work ~~can be obtained upon request~~are archived on ECMWF archiving system (MARS) under the experiment IDs listed in Table 3. Readers with no access to this system can freely obtain these datasets from the corresponding author upon request.

Acknowledgements

We acknowledge funding from the Copernicus Atmosphere Monitoring Service (CAMS), which is funded by the European Union's Copernicus Programme. We are grateful to the World Ozone and Ultraviolet Radiation Data Centre (WOUDC) for providing ozone sonde observations. We thank the Global Atmospheric Watch programme for the provision of CO surface observations. MOPITT data were obtained from the NASA Langley Research Atmospheric Science Data Centre. We acknowledge the free use of tropospheric NO_2 column data from the OMI sensor from the QA4ECV project.

References

Agustí-Panareda, A., Massart, S., Chevallier, F., Balsamo, G., Boussetta, S., Dutra, E., and Beljaars, A.: A biogenic CO2 flux adjustment scheme for the mitigation of large-scale biases in global atmospheric CO2 analyses and forecasts, Atmos. Chem. Phys., 16, 10399-10418, https://doi.org/10.5194/acp-16-10399-2016, 2016.

Atkinson, R., Baulch, D. L., Cox, R. A., Crowley, J. N., Hampson, R. F., Hynes, R. G., Jenkin, M. E., Rossi, M. J., Troe, J., and IUPAC Subcommittee: Evaluated kinetic and photochemical data for atmospheric chemistry: Volume II – gas phase reactions of organic species, Atmos. Chem. Phys., 6, 3625–4055, doi:10.5194/acp-6-3625-2006, 2006.

Aydin, M., Verhulst, K. R., Saltzman, E. S., Battle, M. O., Montzka, S. A., Blake, D. R., Tang, Q., and Prather, M. J.: Recent decreases in fossil-fuel emissions of ethane and methane derived from firn air, Nature, 476, 198–201, doi:10.1038/nature10352, 2011.

Benedetti, A., Morcrette, J.-J., Boucher, O., Dethof, A., Engelen, R. J., Fisher, M., Flentje, H., Huneeus, N., Jones, L.,Kaiser, J. W., Kinne, S., Mangold, A., Razinger, M.,Simmons, A. J., Suttie, M., and the GEMS-AER team: Aerosol analysis and forecast in the European Centre for Medium-Range Weather Forecasts Integrated Forecast System: 2. Data assimilation, J. Geophys. Res., 114, D13205, doi:10.1029/2008JD011115, 2009.

~~Benedetti, A., Morcrette, J.-J., Boucher, O., Dethof, A., Engelen, R. J., Fisher, M., Flentje, H., Huneeus, N., Jones, L.,Kaiser, J. W., Kinne, S., Mangold, A., Razinger, M.,Simmons, A. J., Suttie, M., and the GEMS-AER team: Aerosol analysis and forecast in the European Centre for Medium-Range Weather Forecasts Integrated Forecast System: 2. Data assimilation, J. Geophys. Res., 114, D13205, doi:10.1029/2008JD011115, 2009.~~

Bian, H. S., Chin, M., Hauglustaine, D. A., Schulz, M., Myhre, G., Bauer, S. E., Lund, M. T., Karydis, V. A., Kucsera, T. L., Pan, X. H., Pozzer, A., Skeie, R. B., Steenrod, S. D., Sudo, K., Tsigaridis, K., Tsimpidi, A. P., and Tsyro, S. G.: Investigation of global particulate nitrate from the AeroCom phase III experiment, Atmospheric Chemistry and Physics, 17, 12911-12940, 10.5194/acp-17-12911-2017, 2017.

Boersma, K. F., Eskes, H., Richter, A., De Smedt, I., Lorente, A., Beirle, S., Van Geffen, J., Peters, E., Van Roozendael, M. and Wagner, T., (2017). QA4ECV NO2 tropospheric and stratospheric vertical column data from OMI (Version 1.1) [Data set]. Royal Netherlands Meteorological Institute (KNMI). http://doi.org/10.21944/qa4ecv-no2-omi-v1.1

Bousserez, N., Attié, J.-L., Peuch, V.-H., Michou, M., and Pfister, G.: Evaluation of the MOCAGE chemistry and transport model during the ICARTT/ITOP experiment, J. Geophys. Res., 112, D10S42, doi:10.1029/2006JD007595, 2007.

Brasseur, G. P., Hauglustaine, D. A., Walters, S., Rasch, P. J., Muller, J.-F., Granier, C., and Tie, X.-X.: MOZART: A global chemical transport model for ozone and related chemical tracers, Part 1: Model description, J. Geophys. Res., 103, 28265–28289, 1998.

Cariolle, D. and Deque, M.:. Southern hemisphere medium-scale waves and total ozone disturbances in a spectral general circulation model, J. Geophys. Res., 91D, 10825–10846, 1986.

Cariolle, D. and Teyssèdre, H.: A revised linear ozone photochemistry parameterization for use in transport and general circulation models: multi-annual simulations, Atmos. Chem. Phys., 7, 2183-2196, doi:10.5194/acp-7-2183-2007, 2007.

Cariolle, D., Moinat, P., Teyssèdre, H., Giraud, L., Josse, B., and Lefèvre, F.: ASIS v1.0: an adaptive solver for the simulation of atmospheric chemistry, Geosci. Model Dev., 10, 1467-1485, https://doi.org/10.5194/gmd-10-1467-2017, 2017.

Considine, D. B., A. R. Douglass, P. S. Connell, D. E. Kinnison, and D. A. Rotman (2000), A polar stratospheric cloud parameterization for the global modeling initiative three-dimensional model and its response to stratospheric aircraft, J. Geophys. Res., 105(D3), 3955–3973.

Deeter, M. N., Worden, H. M., Edwards, D. P., Gille, J. C., and Andrews, A. E.: Evaluation of MOPITT Retrievals of Lower tropospheric Carbon Monoxide over the United States, J. Geophys. Res., 117, D13306, doi:10.1029/2012JD017553, 2012.

Deeter, M. N., Edwards, D. P., Francis, G. L., Gille, J. C., Martínez-Alonso, S., Worden, H. M., and Sweeney, C.: A climate-scale satellite record for carbon monoxide: the MOPITT Version 7 product, Atmos. Meas. Tech., 10, 2533-2555, https://doi.org/10.5194/amt-10-2533-2017, 2017.

de Grandpré, J., Ménard, R., Rochon, Y. J., Charette, C., Chabrillat, S., and Robichaud, A.: Radiative Impact of Ozone on Temperature Predictability in a Coupled Chemistry–Dynamics Data Assimilation System, Mon. Weather Rev., 2009, 137, 679-692.

Emmons, L.K., Hauglustaine, D.A., Müller, J., Carroll, M.A., Brasseur, G.P., Brunner, D., Staehelin, J., Thouret, V. and Marenco, A.: Data composites of airborne observations of tropospheric ozone and its precursors. Journal of Geophysical Research 105 (D16), 20497–20538, doi: 10.1029/2000JD900232, 2000.

Emmons, L. K., Walters, S., Hess, P. G., Lamarque, J.-F., Pfister, G. G., Fillmore, D., Granier, C., Guenther, A., Kinnison, D., Laepple, T., Orlando, J., Tie, X., Tyndall, G., Wiedinmyer, C., Baughcum, S. L., and Kloster, S.: Description and evaluation of the Model for Ozone and Related chemical Tracers, version 4 (MOZART-4), Geosci. Model Dev., 3, 43-67, doi:10.5194/gmd-3-43-2010, 2010.

Emmons, L. K., Arnold, S. R., Monks, S. A., Huijnen, V., Tilmes, S., Law, K. S., Thomas, J. L., Raut, J.-C., Bouarar, I., Turquety, S., Long, Y., Duncan, B., Steenrod, S., Strode, S., Flemming, J., Mao, J., Langner, J., Thompson, A. M., Tarasick, D., Apel, E. C., Blake, D. R., Cohen, R. C., Dibb, J., Diskin, G. S., Fried, A., Hall, S. R., Huey, L. G., Weinheimer, A. J., Wisthaler, A., Mikoviny, T., Nowak, J., Peischl, J., Roberts, J. M., Ryerson, T., Warneke, C., and

Helmig, D.: The POLARCAT Model Intercomparison Project (POLMIP): overview and evaluation with observations, Atmos. Chem. Phys., 15, 6721-6744, https://doi.org/10.5194/acp-15-6721-2015, 2015.

Engelen, R. J., S. Serrar, and F. Chevallier (2009), Four-dimensional data assimilation of atmospheric CO2 using AIRS observations, J. Geophys. Res., 114, D03303, doi: 10.1029/2008JD010739.

Flemming, J., Inness, A., Flentje, H., Huijnen, V., Moinat, P., Schultz, M. G., and Stein, O.: Coupling global chemistry transport models to ECMWF's integrated forecast system, Geosci. Model Dev., 2, 253–265, doi:10.5194/gmd-2-253-2009, 2009.

Flemming, J., Huijnen, V., Arteta, J., Bechtold, P., Beljaars, A., Blechschmidt, A.-M., Diamantakis, M., Engelen, R. J., Gaudel, A., Inness, A., Jones, L., Josse, B., Katragkou, E., Marecal, V., Peuch, V.-H., Richter, A., Schultz, M. G., Stein, O., and Tsikerdekis, A.: Tropospheric chemistry in the Integrated Forecasting System of ECMWF, Geosci. Model Dev., 8, 975-1003, doi:10.5194/gmd-8-975-2015, 2015.

Flemming, J., Benedetti, A., Inness, A., Engelen, R. J., Jones, L., Huijnen, V., Remy, S., Parrington, M., Suttie, M., Bozzo, A., Peuch, V.-H., Akritidis, D., and Katragkou, E.: The CAMS interim Reanalysis of Carbon Monoxide, Ozone and Aerosol for 2003–2015, Atmos. Chem. Phys., 17, 1945-1983, https://doi.org/10.5194/acp-17-1945-2017, 2017.

Folberth, G., Hauglustaine, D., Lathiere, J., and Brocheton, F.: Interactive chemistry in the Laboratoire de Meteorologie Dywww namique general circulation model: model description and impact analysis of biogenic hydrocarbons on tropospheric chemistry, Atmos. Chem. Phys., 6, 2273–2319, 2006, http://www.atmos-chem-phys.net/6/2273/2006/.

Fortems-Cheiney, A., Chevallier, F., Pison, I., Bousquet, P., Saunois, M., Szopa, S., Cressot, C., Kurosu, T. P., Chance, K., and Fried, A.: The formaldehyde budget as seen by a global-scale multi-constraint and multi-species inversion system, Atmos. Chem. Phys., 12, 6699-6721, https://doi.org/10.5194/acp-12-6699-2012, 2012.

Galmarini, S., Koffi, B., Solazzo, E., Keating, T., Hogrefe, C., Schulz, M., Benedictow, A., Griesfeller, J. J., Janssens-Maenhout, G., Carmichael, G., Fu, J., and Dentener, F.: Technical note: Coordination and harmonization of the multi-scale, multi-model activities HTAP2, AQMEII3, and MICS-Asia3: simulations, emission inventories, boundary conditions, and model output formats, Atmos. Chem. Phys., 17, 1543-1555, https://doi.org/10.5194/acp-17-1543-2017, 2017.

Granier, C., J.F. Lamarque, A. Mieville, J.F. Muller, J. Olivier, J. Orlando, J. Peters, G. Petron, G. Tyndall, S. Wallens, POET, a database of surface emissions of ozone precursors. Available from http://eccad.aeris-data.fr/ , 2005.

Granier, C., B. Bessagnet, T. Bond, A. D'Angiola, H.D.v.d. Gon, G.J. Frost, A. Heil, J.W. Kaiser, S. Kinne, Z. Klimont, S. Kloster, J.-F. Lamarque, C. Liousse, T. Masui, F. Meleux, A. Mieville, T. Ohara, J.-C. Raut, K. Riahi, M.G. Schultz, S.J. Smith, A. Thomson, J.v. Aardenne, G.R.v.d. Werf, and D.P.v. Vuuren, Evolution of anthropogenic and biomass burning emissions of air pollutants at global and regional scales during the 1980-2010 period, Climatic Change, 109(1-2), 163-190, doi:10.1007/s10584-011-0154-1, 2011.

Hall, S. R., Ullmann, K., Prather, M. J., Flynn, C. M., Murray, L. T., Fiore, A. M., Correa, G., Strode, S. A., Steenrod, S. D., Lamarque, J.-F., Guth, J., Josse, B., Flemming, J., Huijnen, V., Abraham, N. L., and Archibald, A. T.: Cloud impacts on photochemistry: building a climatology of photolysis rates from the Atmospheric Tomography mission, Atmos. Chem. Phys., 18, 16809-16828, https://doi.org/10.5194/acp-18-16809-2018, 2018.

Hausmann, P., Sussmann, R., and Smale, D.: Contribution of oil and natural gas production to renewed increase in atmospheric methane (2007–2014): top–down estimate from ethane and methane column observations, Atmos. Chem. Phys., 16, 3227-3244, https://doi.org/10.5194/acp-16-3227-2016, 2016.

~~Horowitz, L. W.,~~ Hooghiemstra, P. B., Krol, M. C., Meirink, J. F., Bergamaschi, P., van der Werf, G. R., Novelli, P. C., Aben, I., and Röckmann, T.: Optimizing global CO emission estimates using a four-dimensional variational data assimilation system and surface network observations, Atmos. Chem. Phys., 11, 4705-4723, https://doi.org/10.5194/acp-11-4705-2011, 2011.

~~Walters, S., Mauzerall, D. L., Emmons, L. K., Rasch, P. J., Granier, C., Tie, X., Lamarque, J.-F., Schultz, M. G., Tyndall, G. S., Orlando, J. J., and Brasseur, G. P.: A global simulation of tropospheric ozone and related tracers: Description and evaluation of MOZART, version 2, J. Geophys. Res., 108, 4784, doi:10.1029/2002JD002853, 2003.~~

Huang, M., Carmichael, G. R., Pierce, R. B., Jo, D. S., Park, R. J., Flemming, J., Emmons, L. K., Bowman, K. W., Henze, D. K., Davila, Y., Sudo, K., Jonson, J. E., Tronstad Lund, M., Janssens-Maenhout, G., Dentener, F. J., Keating, T. J., Oetjen, H., and Payne, V. H.: Impact of intercontinental pollution transport on North American ozone air pollution: an HTAP phase 2 multi-model study, Atmos. Chem. Phys., 17, 5721-5750, https://doi.org/10.5194/acp-17-5721-2017, 2017.

Huijnen, V., Williams, J., van Weele, M., van Noije, T., Krol, M., Dentener, F., Segers, A., Houweling, S., Peters, W., de Laat, J., Boersma, F., Bergamaschi, P., van Velthoven, P., Le Sager, P., Eskes, H., Alkemade, F., Scheele, R., Nédélec, P., and Pätz, H.-W.: The global chemistry transport model TM5: description and evaluation of the tropospheric chemistry version 3.0, Geosci. Model Dev., 3, 445-473, doi:10.5194/gmd-3-445-2010.

Huijnen V., Wooster, M. J., Kaiser, J. W. , Gaveau, D.L.A. , Flemming, J. , Parrington, M. Inness, A., Murdiyarso, D., Main, B. and van Weele, M.: Fire carbon emissions over maritime southeast Asia in 2015 largest since 1997, Scientific Reports, 6:26886, DOI: 10.1038/srep26886, 2016a.

Huijnen, V., Flemming, J., Chabrillat, S., Errera, Q., Christophe, Y., Blechschmidt, A.-M., Richter, A., and Eskes, H.: C-IFS-CB05-BASCOE: stratospheric chemistry in the Integrated Forecasting System of ECMWF, Geosci. Model Dev., 9, 3071-3091, doi:10.5194/gmd-9-3071-2016, 2016b.

Im, U., Christensen, J. H., Geels, C., Hansen, K. M., Brandt, J., Solazzo, E., Alyuz, U., Balzarini, A., Baro, R., Bellasio, R., Bianconi, R., Bieser, J., Colette, A., Curci, G., Farrow, A., Flemming, J., Fraser, A., Jimenez-Guerrero, P., Kitwiroon, N., Liu, P., Nopmongcol, U., Palacios-Peña, L., Pirovano, G., Pozzoli, L., Prank, M., Rose, R., Sokhi, R., Tuccella, P., Unal, A., Vivanco, M. G., Yarwood, G., Hogrefe, C., and Galmarini, S.: Influence of anthropogenic emissions and boundary conditions on multi-model simulations of major air pollutants over Europe and North America in the framework of AQMEII3, Atmos. Chem. Phys., 18, 8929-8952, https://doi.org/10.5194/acp-18-8929-2018, 2018.

Inness, A., Benedetti, A., Flemming, J., Huijnen, V., Kaiser, J. W., Parrington, M., and Remy, S.: The ENSO signal in atmospheric composition fields: emission-driven versus dynamically induced changes, Atmos. Chem. Phys., 15, 9083-9097, https://doi.org/10.5194/acp-15-9083-2015, 2015.

Inness, A., Ades, Agusti-Panareda, J. Barré, Dominguez, Engelen, J. Flemming, V. Huijnen, L. Jones, Z. Kipling, S. Massart, M. Parrington, M. Razinger, S. Remy, M. Suttie, et al., The CAMS reanalysis of atmospheric composition. Submitted to ACPD, 2018.

Jacob, D.J. H. Liu, C.Mari, and R.M. Yantosca, Harvard wet deposition scheme for GMI, Harvard University Atmospheric Chemistry Modeling Group, revised March 2000. http://acmg.seas.harvard.edu/geos/wiki_docs/deposition/wetdep.jacob_etal_2000.pdf . Last access: 28 nov 2018.

Jacob, D. J., Field, B. D., Li, Q., Blake, D. R., de Gouw, J., Warneke, C., Hansel, A., Wisthaler, A., Singh, H. B., and Guenther, A.: Global budget of methanol: Constraints from atmospheric observations, J. Geophys. Res., 110, doi:10.1029/2004JD005 172, 2005.

Jöckel, P., Tost, H., Pozzer, A., Brühl, C., Buchholz, J., Ganzeveld, L., Hoor, P., Kerkweg, A., Lawrence, M. G., Sander, R., Steil, B., Stiller, G., Tanarhte, M., Taraborrelli, D., van Aardenne, J., and Lelieveld, J.: The atmospheric chemistry general circulation model ECHAM5/MESSy1: consistent simulation of ozone from the surface to the mesosphere, Atmos. Chem. Phys., 6, 5067-5104, https://doi.org/10.5194/acp-6-5067-2006, 2006.

Kaiser, J. W., Heil, A., Andreae, M. O., Benedetti, A., Chubarova, N., Jones, L., Morcrette, J.-J., Razinger, M., Schultz, M. G., Suttie, M., and van der Werf, G. R.: Biomass burning emissions estimated with a global fire assimilation system based on observed fire radiative power, Biogeosciences, 9, 527-554, doi:10.5194/bg-9-527-2012, 2012.

Kaminski, J. W., Neary, L., Struzewska, J., McConnell, J. C., Lupu, A., Jarosz, J., Toyota, K., Gong, S. L., Côté, J., Liu, X., Chance, K., and Richter, A.: GEM-AQ, an on-line global multiscale chemical weather modelling system: model description and evaluation of gas phase chemistry processes, Atmos. Chem. Phys., 8, 3255-3281, 2008.

Kinnison, D. E., Brasseur, G. P., Walters, S., Garcia, R. R., Marsh, D. R, Sassi, F., Harvey, V. L., Randall, C. E., Emmons, L., Lamarque, J. F., Hess, P., Orlando, J. J., Tie, X. X., Randel, W., Pan, L. L., Gettelman, A., Granier, C., Diehl, T., Niemeier, U. and Simmons, A. J.: Sensitivity of Chemical Tracers to Meteorological Parameters in the MOZART-3 Chemical Transport Model, J. Geophys. Res, 112, D03303, doi:10.1029/2008JD010739, 2007.

Komhyr, W. D., Barnes, R. A., Borthers, G. B., Lathrop, J. A., Kerr, J. B., and Opperman, D. P.: Electrochemical concentration cell ozonesonde performance evaluation during STOIC 1989, J. Geophys. Res., 100, 9231–9244, 1995

Lacressoniere, G.,V.-H. Peuch, J. Arteta, B. Josse,M. Joly, V. Marecal, D. Saint-Martin, M. Deque, and L. Watson, 2013 : How realistic are air quality hindcasts driven by forcings from climate model simulations? Geoscientific Model Development, Volume : 5, Pages : 1565-1587, DOI:10.5194/gmd-5-1565-2012, 2012.

Lamarque, J.-F., Bond, T. C., Eyring, V., Granier, C., Heil, A., Klimont, Z., Lee, D., Liousse, C., Mieville, A., Owen, B., Schultz, M. G., Shindell, D., Smith, S. J., Stehfest, E., Van Aardenne, J., Cooper, O. R., Kainuma, M., Mahowald, N., McConnell, J. R., Naik, V., Riahi, K., and van Vuuren, D. P.: Historical (1850\u20132000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: methodology and application, Atmos. Chem. Phys., 10, 7017-7039, https://doi.org/10.5194/acp-10-7017-2010, 2010.

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

Lefevre, F., G.P. Brasseur, I. Folkins, A.K. Smith, and P. Simon, Chemistry of the 1991-1992 stratospheric winter: three-dimensional model simulations, J. Geophys. Res., 99 (D4), 8183-8195, 1994.

Lefèvre, F., F. Figarol, K. S. Carlsaw and T. Peter: The 1997 ozone hole quantified from three dimensional model simulations, G.eophys. Res. Let., 25, vol 13, 2425-2428, 1998.

Madronich, S~~.:~~., Photodissociation in the atmosphere 1. Actinic flux and the effects of ground reflections and clouds, J. Geophys. Res., 92, 9740–9752, 1987.

Madronich, S~~. (1989),~~., Photodissociation in the atmosphere: 1. Actinic flux and effect of ground reflection and clouds, J. Geophys. Res., 92, 9740–9752~~.~~, 1989

Madronich, S. and Flocke, S.: The Role of Solar Radiation in Atmospheric Chemistry, in: Environmental Photochemistry, edited by Boule, P., vol. 2 / 2L of The Handbook of Environmental Chemistry, pp. 1–26, Springer Berlin Heidelberg, doi:10.1007/978-3-540-69044-3_1, 1999.

Madronich, S. and Flocke, S. J.: Theoretical estimation of biologically effective uv radiation at the Earth's surface, in: Solar Ultraviolet Radiation – Modelling, Measurement and Effects, edited by: Zerefos, C., Springer-Verlag, Berlin, 23–48, 1997.

Marécal, V., Peuch, V.-H., Andersson, C., Andersson, S., Arteta, J., Beekmann, M., Benedictow, A., Bergström, R., Bessagnet, B., Cansado, A., Chéroux, F., Colette, A., Coman, A., Curier, R. L., Denier van der Gon, H. A. C., Drouin, A., Elbern, H., Emili, E., Engelen, R. J., Eskes, H. J., Foret, G., Friese, E., Gauss, M., Giannaros, C., Guth, J., Joly, M., Jaumouillé, E., Josse, B., Kadygrov, N., Kaiser, J. W., Krajsek, K., Kuenen, J., Kumar, U., Liora, N., Lopez, E., Malherbe, L., Martinez, I., Melas, D., Meleux, F., Menut, L., Moinat, P., Morales, T., Parmentier, J., Piacentini, A., Plu, M., Poupkou, A., Queguiner, S., Robertson, L., Rouïl, L., Schaap, M., Segers, A., Sofiev, M., Tarasson, L., Thomas, M., Timmermans, R., Valdebenito, Á., van Velthoven, P., van Versendaal, R., Vira, J., and Ung, A.: A regional air quality forecasting system over Europe: the MACC-II daily ensemble production, Geosci. Model Dev., 8, 2777-2813, https://doi.org/10.5194/gmd-8-2777-2015, 2015.

Ménégoz, M., Salas y Melia, D., Legrand, M., Teyssèdre, H., Michou, M., Peuch, V.-H., Martet, M., Josse, B., and Dombrowski- Etchevers, I.: Equilibrium of sinks and sources of sulphate over Europe: comparison between a six-year simulation and EMEP observations, Atmos. Chem. Phys., 9, 4505–4519, doi:10.5194/acp-9-4505-2009, 2009.

Metzger, S., Dentener, F., Krol, M. C., Jeuken, A., and Lelieveld, J.: Gas/aerosol partitioning 2. Global modeling results, J. Geophys. Res., 107(D16), 4313, doi:10.1029/2001JD001103, 2002.

Michou, M., Laville, P., Serça, D., Fotiadi, A., Bouchou, P., and Peuch, V.-H.: Measured and modeled dry deposition velocities over the ESCOMPTE area, Atmos. Res., 74, 89–116, 2004.

Monks, S. A., Wilson, C., Emmons, L. K., Hannigan, J., Helmig, D., Blake, N. J., & Blakes, D. R. (2018). Using an inverse model to reconcile differences in simulated and observed global ethane concentrations and trends between 2008 and 2014. *Journal of Geophysical Research: Atmospheres*, 123. https://doi.org/10.1029/2017JD028112

Morgenstern, O., Braesicke, P., O'Connor, F. M., Bushell, A. C., Johnson, C. E., Osprey, S. M., and Pyle, J. A.: Evaluation of the new UKCA climate-composition model – Part 1: The stratosphere, Geosci. Model Dev., 2, 43–57, doi:10.5194/gmd-2-43-2009, 2009.

Morcrette, J.-J., Boucher, O., Jones, L., Salmond, D. , Bechtold, P., Beljaars, A., Benedetti, A., Bonet, A., Kaiser, J. W., Razinger, M., Schulz, M. , Serrar, S. , Simmons, A. J., Sofiev, M., Suttie, M., Tompkins, A. M. and Untch, A.: Aerosol analysis and forecast in the ECMWF Integrated Forecast System. Part I: Forward modelling, J. Geophys. Res., 2009.

Nechita-Banda, N., Krol, M., van der Werf, G. R., Kaiser, J. W., Pandey, S., Huijnen, V., Clerbaux, C., Coheur, P., Deeter, M. N., Röckmann T.: Monitoring emissions from the 2015 Indonesian fires using CO satellite data. Phil. Trans. R. Soc. B 2018 373 20170307; DOI: 10.1098/rstb.2017.0307, 2018.

Novelli, P. C., Masarie, K. A., Lang, P. M., Hall, B. D., Myers, R. C., and Elkins, J. W.: Reanalysis of tropospheric CO trends: effects of the 1997–1998 wildfires, J. Geophys. Res., 108, 4464, doi:10.1029/2002JD003031, 2003.

O'Connor, F. M., Johnson, C. E., Morgenstern, O., Abraham, N. L., Braesicke, P., Dalvi, M., Folberth, G. A., Sanderson, M. G., Telford, P. J., Voulgarakis, A., Young, P. J., Zeng, G., Collins, W. J., and Pyle, J. A.: Evaluation of the new UKCA climate-composition model – Part 2: The Troposphere, Geosci. Model Dev., 7, 41-91, doi:10.5194/gmd-7-41-2014, 2014.

Pozzer, A., Jöckel, P., Tost, H., Sander, R., Ganzeveld, L., Kerkweg, A., and Lelieveld, J.: Simulating organic species with the global atmospheric chemistry general circulation model ECHAM5/MESSy1: a comparison of model results with observations, Atmos. Chem. Phys., 7, 2527-2550, https://doi.org/10.5194/acp-7-2527-2007, 2007.

Pozzer, A., Jöckel, P., and Van Aardenne, J.: The influence of the vertical distribution of emissions on tropospheric chemistry, Atmos. Chem. Phys., 9, 9417-9432, https://doi.org/10.5194/acp-9-9417-2009, 2009.

Powers, J.G., J.B. Klemp, W.C. Skamarock, C.A. Davis, J. Dudhia, D.O. Gill, J.L. Coen, D.J. Gochis, R. Ahmadov, S.E. Peckham, G.A. Grell, J. Michalakes, S. Trahan, S.G. Benjamin, C.R. Alexander, G.J. Dimego, W. Wang, C.S. Schwartz, G.S. Romine, Z. Liu, C. Snyder, F. Chen, M.J. Barlage, W. Yu, and M.G. Duda, 2017: The Weather Research and Forecasting Model: Overview, System Efforts, and Future Directions. Bull. Amer. Meteor. Soc., 98, 1717–1737, https://doi.org/10.1175/BAMS-D-15-00308.1

Robichaud, A., Ménard, R., Chabrillat, S., de Grandpré, J., Rochon, Y. J., Yang, Y., and Charette, C.: Impact of energetic particle precipitation on stratospheric polar constituents: an assessment using monitoring and assimilation of operational MIPAS data, Atmos. Chem. Phys., 10, 1739-1757, 2010.

Sander, S. P., Friedl, R. R., DeMore, W. B., Golden, D. M., Kurylo, M. J., Hampson, R. F., et al.. Chemical Kinetics and Photochemical Data for Use in Stratospheric Modeling. Supplement to Evaluation 12: Update of Key Reactions, Evaluation Number 13, JPL Publication 00-03, Jet Propulsion Laboratory, Pasadena, 2000.

Sander, S. P., Abbatt, J. R., Burkholder, J. B., Friedl, R. R., Golden, D. M., Huie, R. E., Kolb, C. E., Kurylo, G., Moortgat, K., Orkin, V. L. and Wine, P. H.: Chemical kinetics and Photochemical Data for Use in Atmospheric studies, Evaluation No.17, JPL Publication 10-6, Jet Propulsion Laboratory, Pasadena, 2011.

Sander, R., Baumgaertner, A., Cabrera, D., Frank, F., Grooß, J.-U., Gromov, S., Harder, H., Huijnen, V., Jöckel, P., Karydis, V. A., Niemeyer, K. E., Pozzer, A., Riede, H., Schultz, M. G., Taraborrelli, D., and Tauer, S.: The atmospheric chemistry box model CAABA/MECCA-4.0gmdd, Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-201, in review, 2018.

Sandu, A. and Sander, R.: Technical note: Simulating chemical systems in Fortran90 and Matlab with the Kinetic PreProcessor KPP-2.1, Atmos. Chem. Phys., 6, 187-195, doi:10.5194/acp-6-187-2006, 2006.

Shindell, D. T., Faluvegi, G., Stevenson, D. S., Krol, M. C., Emmons, L. K., Lamarque, J.-F., Petron, G., Dentener, F. J., Ellingsen, K., Schultz, M. G., Wild, O., Amann, M., Atherton, C. S., Bergmann, D. J., Bey, I., Butler, T., Cofala, J., Collins, W. J., Derwent, R. G., Doherty, R. M., Drevet, J., Eskes, H. J., Fiore, A. M., Gauss, M., Hauglustaine, D. A., Horowitz, L.W., Isaksen, I. S. A., Lawrence, M. G., Montanaro, V., Muller, J. F., Pitari, G., Prather, M. J., Pyle, J. A., Rast, S., Rodriguez, J. M., Sanderson, M. G., Savage, N. H., Strahan, S. E., Sudo, K., Szopa, S., Unger, N., van Noije, T. P. C., and Zeng, G.: Multimodel simulations of carbon monoxide: Comparison with observations and projected near-future changes, J. Geophys. Res., 111, D19306, doi:10.1029/2006JD007100, 2006.

Skachko, S., Ménard, R., Errera, Q., Christophe, Y., and Chabrillat, S.: EnKF and 4D-Var data assimilation with chemical transport model BASCOE (version 05.06), Geosci. Model Dev., 9, 2893-2908, https://doi.org/10.5194/gmd-9-2893-2016, 2016.

Stein, O., Schultz, M. G., Bouarar, I., Clark, H., Huijnen, V., Gaudel, A., George, M., and Clerbaux, C.: On the wintertime low bias of Northern Hemisphere carbon monoxide found in global model simulations, Atmos. Chem. Phys., 14, 9295-9316, https://doi.org/10.5194/acp-14-9295-2014, 2014.

Steinbrecht, W., Shwartz, R., and Claude, H.: New pump correction for the Brewer-Mast ozone sonde: Determination from experiment and instrument intercomparisons, J. Atmos. Ocean. Tech.15, 144–156, 1998.

Stockwell, W. R., Kirchner, Kuhn, F., M., and Seefeld, S.: A new mechanism for regional atmospheric chemistry modeling, J. Geophys. Res., 102(D22), 25847–25879, doi: 10.1029/97JD00849, 1997.

Sindelarova, K., Granier, C., Bouarar, I., Guenther, A., Tilmes, S., Stavrakou, T., Müller, J.-F., Kuhn, U., Stefani, P., and Knorr, W.: Global data set of biogenic VOC emissions calculated by the MEGAN model over the last 30 years, Atmos. Chem. Phys., 14, 9317-9341, https://doi.org/10.5194/acp-14-9317-2014, 2014.

Struzewska, J., Zdunek, M., Kaminski, J. W., Łobocki, L., Porebska, M., Jefimow, M., and Gawuc, L.: Evaluation of the GEM-AQ model in the context of the AQMEII Phase 1 project, Atmos. Chem. Phys., 15, 3971-3990, https://doi.org/10.5194/acp-15-3971-2015, 2015.

Tilmes, S., Lamarque, J.-F., Emmons, L. K., Kinnison, D. E., Marsh, D., Garcia, R. R., Smith, A. K., Neely, R. R., Conley, A., Vitt, F., Val Martin, M., Tanimoto, H., Simpson, I., Blake, D. R., and Blake, N.: Representation of the Community Earth System Model (CESM1) CAM4-chem within the Chemistry-Climate Model Initiative (CCMI), Geosci. Model Dev., 9, 1853-1890, doi:10.5194/gmd-9-1853-2016, 2016.

Toon, G. C., Blavier, J.-F. L., and Sung, K.: Measurements of atmospheric ethene by solar absorption FTIR spectrometry, Atmos. Chem. Phys., 18, 5075-5088, https://doi.org/10.5194/acp-18-5075-2018, 2018.

Williams, J. E., Strunk, A., Huijnen, V., and van Weele, M.: The application of the Modified Band Approach for the calculation of on-line photodissociation rate constants in TM5: implications for oxidative capacity, Geosci. Model Dev., 5, 15-35, doi:10.5194/gmd-5-15-2012, 2012.

Williams, J. E., van Velthoven, P. F. J., and Brenninkmeijer, C. A. M.: Quantifying the uncertainty in simulating global tropospheric composition due to the variability in global emission estimates of Biogenic Volatile Organic Compounds, Atmos. Chem. Phys., 13, 2857-2891, doi:10.5194/acp-13-2857-2013, 2013.

Yarwood, G., Rao, S., Yocke, M., and Whitten, G.: Updates to the carbon bond chemical mechanism: CB05. Final report to the US EPA, EPA Report Number: RT-0400675, available at: www.camx.com, last access: 28 Nov. 2018, 2005.

Zhang, X., He, S. Z., Chen, Z. M., Zhao, Y., and Hua, W.: Methyl hydroperoxide (CH3OOH) in urban, suburban and rural atmosphere: ambient concentration, budget, and contribution to the atmospheric oxidizing capacity, Atmos. Chem. Phys., 12, 8951-8962, https://doi.org/10.5194/acp-12-8951-2012, 2012.

List of Figures

Figure 1: Geographical distribution of the aircraft campaigns presented by Emmons et al. (2000). Each field campaign is represented by a different color. Further information on the campaigns is found in Emmons et al. (2000). 12

Figure 2: Geographical distribution of the ozone sondes during 2011 used for evaluation, colored for the various seasons. The size of the triangles provides information of the relative amount of observations available for each of the seasons and locations, compared to the other locations. Also the geographical aggregation for the five latitude bands presented in Figures 5 and 7, as well as the Western Europe and Eastern US regions is given. 14

Figure 3: Zonal, annual mean O_3, CO, NO_x, mixing ratios and OH concentrations in CBA (left), MOZ (middle) and MOC (right). 16

Figure 4: Relative differences (in %) of OH daily averaged mixing ratios of simulation MOC with respect to MOZ at 600 hPa. 17

Figure 5: Tropospheric ozone profiles from model versions CBA (red), MOC (blue) and MOZ (green) against sondes (black) in volume mixing ratios (ppbv) over six different regions: from top row to bottom row, NH-Polar [90°N-60°N], Western Europe [45°N-54°N; 0°E-23°E], Eastern US [32°N-45°N; 90°W-65°W], Tropics [30°N-30°S], SH mid-latitudes [30°S-60°S] and Antarctic [60°S-90°S], averaged over four seasons (from left to right: December-January-February, March-April-May, June-July-August, September-October-November). 21

Figure 6: Mean tropospheric ozone profiles from model versions CBA (red), MOC (blue) and MOZ (green) against sondes (black) in volume mixing ratios (ppbv) during DJF and JJA at selected individual stations. Error bars represent the 1σ spread in the seasonal mean observations. 21

Figure 7: Mean of all model biases (left) and RMSE (right) values against ozone sondes as function of latitude for various pressure ranges (top row: 300-500hPa; middle row: 500-700hPa; bottom row: 700-900hPa), averaged over the full year. Same color codes as in the previous figure. The numbers in each latitude range indicate the amount of stations that contribute to these statistics. For reference, also the corresponding results from the CAMS-Interim Reanalysis (CAMSiRA) are given in orange 23

Figure 8: MOPITT CO total column retrieval for April 2011 (top left) and simulated by IFS(CBA) (top right), IFS(MOZ) (bottom left) and IFS(MOC) (bottom right). 24

Figure 9: MOPITT CO total column retrieval for August 2011 (top left) and simulated by IFS(CBA) (top right), IFS(MOZ) (bottom left) and IFS(MOC) (bottom right). 25

Figure 10: Comparison of CO mixing ratios (ppbv) at the surface as simulated (red, blue and green are model results from CBA, MOC and MOZ, respectively) and observed (black) at twelve stations sorted by decreasing latitudes. The bars represent one-standard deviation of the monthly average for the location of the station. 26

Figure 11: Temporal correlation (R^2) between monthly mean surface CO as derived from observations (GAW network) and model simulations (left: CBA, middle: MOC, right: MOZ). 26

Figure 12: Comparison of simulated CO vertical profiles by using the CBA (red solid line), MOC (blue solid line) and MOZ (green solid line) chemistry versions against aircraft data (black dots). Also shown are the modeled (dashed lines) and measured (black rectangular) standard deviations. The numbers on the right vertical axis indicate the number of available measurements. 27

Figure 13: Comparison of simulated CH_2O vertical profiles by using the CBA (red), MOC (blue) and MOZ (green) chemistry versions against aircraft data (black). Line styles and symbols have the same meaning as in Figure 12. 28

Figure 14: Comparison of simulated CH_3OOH vertical profiles by using the CBA (red), MOC (blue) and MOZ (green) chemistry versions against aircraft data (black). Line styles and symbols have the same meaning as in Figure 11. 29

Figure 15: Comparison of simulated C_2H_4 vertical profiles by using the CBA (red), MOC (blue) and MOZ (green) chemistry versions against aircraft data (black). Line styles and symbols have the same meaning as in Figure 12. 31

Figure 16: Comparison of simulated C_2H_6 vertical profiles by using the CBA (red), MOC (blue) and MOZ (green) chemistry versions against aircraft data (black). Line styles and symbols have the same meaning as in Figure 12. 32

Figure 17: Comparison of daytime NO_2 vertical profiles simulated by CBA (red), MOC (blue) and MOZ (green) chemistry versions against aircraft data (black). Line styles and symbols have the same meaning as in Figure 12. 32

Figure 18: Monthly mean Tropospheric NO_2 columns from OMI satellite retrievals from the QA4ECV product for April 2011, along with the corresponding collocated model biases. 34

Figure 19: Monthly mean Tropospheric NO_2 columns from OMI satellite retrievals from the QA4ECV product for August 2011, along with the corresponding collocated model biases. 35

Figure 20: Comparison of simulated HNO_3 vertical profiles by using the CBA (red), MOC (blue) and MOZ (green) chemistry versions against aircraft data (black). Line styles and symbols have the same meaning as in Figure 12. 36

Figure 21: Comparison of simulated SO_2 vertical profiles by using the CBA (red), MOC (blue) and MOZ (green) chemistry versions against aircraft data (black). Line styles and symbols have the same meaning as in Figure 12. 37