Response to the first 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.

The manuscript 'An inter-comparison of tropospheric ozone reanalysis products from CAMS, CAMS-Interim, TCR-1 and TCR-2' presents a description and extensive evaluation of tropospheric ozone from four recent global chemical reanalyses: CAMS-iRean, CAMS-Rean, TCR1 and TCR2. The study performs very detailed comparisons between the reanalyses and independent observations of surface, profile and column ozone and assesses the relative performance of the reanalyses. This includes some very nice analyses of specific aspects of the reanalyses such as the representation of the diurnal ozone cycle. I am really impressed by the amount of work that went into this study and I applaud the authors for their very thorough and well-organized analysis. This type of paper is not easy to write but it is very important for the scientific community, especially as the understanding of the importance of chemical reanalyses is growing.
The paper is well thought out, well-written and well-organized. It is certainly worth of prompt publication. I have only several minor suggestions for edits and some technical corrections.

Minor general comments

1. Tables and the discussion of reanalyses' performance: it would be really good to have the RMSE values shown in percent in addition to absolute values (ppbv, DU). Having absolute values alone makes it difficult to judge if the RMSE is large or not. I appreciate that sometimes, particularly when the mean ozone is low, large percent values may be misleading but that shouldn't be an issue if both, absolute and relative RMSE is shown. In my specific comments I point to some places where having percentages would be particularly useful but it would really be best to have them in all the tables.

We now add barplots with normalized mean bias and normalized standard deviation (instead of RMSE, as per a comment from the second reviewer) in the supplementary material.

2. It would be helpful to have a schematic figure, similar to Davis et al. 2017 Fig. 1, showing the ozone observations assimilated by each reanalysis, indicating whether a bias correction was applied or not, and, as an added benefit, showing time periods of the reanalyses.

We have checked the Fig 1. in Davis et al. (2017), but find it unpractical to introduce a similar figure for our purposes, as this implies considerable overlap with Tables 2-4. Also not only information on the satellite instrument is important, but also the version specification, which implies that the figure cannot replace the existing tables. We now introduce a separate section in the manuscript to discuss any changes in the observing systems.

3. The authors often use the word 'model' as synonymous with 'reanalysis', e.g. L273, L307, L316, L318, and in many other places. I suggest limiting the use of this term to the instances where you are really talking about a model (e.g. 'model levels' or *‘chemistry model’, etc.)*

The reviewer is fully correct. We have checked the manuscript, and replaced ‘model’ with ‘analysis’ or similar, where appropriate.

4. Section 5 contains detailed discussions of reanalyses comparisons against multiple data sets. It’s easy to lose the big picture in all these details. It would be really helpful to include 2-3 sentence summary highlighting the key results at the end of each subsection as it is already done in Subsection 5.3.

We now introduce a summary section at the end of Sec. 4.3, concluding the evaluation against ozone sondes. Likewise we now introduce summary statements at the end of Sec 5.2 an Sec 6:

End of sec. 4.3:
In conclusion, evaluation against ozone sondes has revealed the following:
- The updated reanalyses show on average improved performance compared to the predecessor versions, but with some notable exceptions, such as an increased positive bias over the Antarctic in CAMS-Rean versus CAMS-iRean. Over the Antarctic the TCR-2 strongly improved upon TCR-1, despite the lack of direct observational constraints.
- For individual regions or conditions CAMS Reanalysis and TCR-2 show different performance, but averaged for all regions of similar quality. Best performance, in terms of mean bias, standard deviation and correlation, for the updated reanalyses is obtained for the Western Europe, Eastern US and SH mid latitude regions (both normalized mean bias and standard deviation below 8% at 850 and 650 hPa). Relatively worst performance is found for the Antarctic region, with normalized standard deviation up to 18%. This is likely associated to the fewer observational constraints in the polar regions compared to the other regions.
- In terms of temporal consistency, the CAMS Reanalyses show degraded performance over the polar regions during 2003 and 2004, due to lower quality MIPAS and SCIAMACHY data usage. CAMS-iREAN also shows a change in performance statistics in the polar regions from 2014 onwards, associated to a changes in the MLS retrieval product versions. Furthermore, both CAMS-Rean and CAMS-iRean are affected by the change in the SBUV/2 product versions in 2013.
With the reduced data-availability from TES from 2010 onwards the TCR tropospheric ozone products show changes in their performances. Remarkably, TCR-1 and TCR-2 show overall slight improvements from 2010 onwards. This is marked by reduced positive biases in the lower troposphere over NH-mid-latitude regions and may be attributed to biases in the TES retrieval product, combined with changes in the OMI product, see also Sec. 2.5. Additional Observing System Experiments (OSEs) are needed to identify the relative roles of individual assimilated measurements on the changes in reanalysis bias.
end of sec 5.2:
In summary, CAMS-Rean shows the best ability to capture the regional mean surface ozone and its variability, while particularly TCR-2 (and to lesser extent also TCR-1) shows positive biases and reduced correlations. Particularly good performance is seen over the western US (R=0.95, MB=-0.2), while over east, and particularly southeast, Asia the performance is poorest.

end of sec.6 :
In summary, all reanalyses capture the synoptic to diurnal variability, as illustrated by the assessment of the heatwave event in July 2006. Still there are considerable differences in

performance, depending on the reanalysis, region and season. While CAMS-iRean and CAMS-Rean perform mostly similar, for TCR-2 a considerable improvement was found compared to TCR-1. Overall better temporal correlations are obtained for the summer period compared to winter, and also for Western Europe compared to the Mediterranean region. Further improvements can be obtained by a better description of surface processes, including emissions and deposition, together with higher spatial resolution modelling.

Specific comments and technical corrections

L35 climate-change à climate change

Changed.

L34-36. This sentence conflates two different things: (1) the importance of ozone forcing for climate and (2) a lack of impact of improved ozone representation on long-term weather forecasts. I suggest splitting it into two sentences.

Thanks for this suggestion. We now write:
Owing to its radiative effects, tropospheric ozone is an important driver in climate change (Checa-Garcia et al., 2018). Also it may affect long-range weather forecasts, even if in evaluations no improvement has been detected so far (Cheung et al., 2014).

L38. This deserves more references than just the two that are provided.

This is correct. We now have added references to Monks et al. (2015), Huang et al., (2017), and Hsu and Prather (2009).

L120-121 'to evaluate their fitness for purpose for the various types of application described above'. This sounds a little awkward. Please, consider rephrasing.

We now write:
To assess the quality of these reanalysis products, with attention for the various potential types of application described above, this study evaluates tropospheric ozone…

L158. Was there any kind of bias correction applied to these ozone data, as in CAMS-REAN? Maybe I missed that information. As I stated in my general comments a figure summarizing all these data types and how they're used in each reanalysis would be useful. This information could also be added to tables 2, 3 and 4.

We now explicitly mention the bias correction settings in any of the reanalyses. The settings for CAMS-iRean and CAMS-Rean are identical (variational bias correction for OMI, SCIAMACHY, GOME-2, and anchoring for SBUV/2, MLS and MIPAS), while in TCR-1 and TCR-2 all observations were used without bias correction.

Table 2. For the profile data types it would be helpful to include the vertical ranges or at

least the lowest levels assimilated.

Profile data from MIPAS and MLS instruments in the ranges 0.1 -150 hPa (MIPAS) and 0.1 -147hPa (MLS) is used. For SBUV and GOME (ERS-2) the vertical resolution is very low, implying that they can effectively be considered as total column retrievals. We now include such a comment in the manuscript.

L200. Why couldn't they be filtered out?

These OMI row anomalies could not be filtered out because at the time this information was not available in the BUFR data which are used as input to the IFS data assimilation system. This had unfortunately not been noticed before running the reanalysis. This information will be taken into account in any future reanalysis. We now write:

-Different behaviour of OMI data between 2009 and 2012, associated to a deterioration in the OMI row anomalies (Schenkeveld et al., 2017) which unfortunately have not be filtered out in the CAMS assimilation procedure;

L231 & L247.Livesey et al., 2011 is not in the reference list. If this is the MLS version 4.2 data quality and description document then its latest version is from 2018 (https://mls.jpl.nasa.gov/data/v4-2_data_quality_document.pdf). Is the MLS data quality screening based on some earlier guidelines? Note that v4.2 didn't exist in 2011.

The reviewer is correct: Version 4.2 was actually used in both TCR-1 and TCR-2. The v3.3 MLS data was used in a predecessor of TCR-1, not assessed in our manuscript. MLS data quality screening is also based on the v4.2 guidelines. We have updated the manuscript, and reference on this.

L245-251. TES should also be mentioned here for completeness.

We now include such a sentence, thank you for this suggestion.

Table 4. According to the table TCR-1 uses MLS v3.3. It's version 4.2 in the text (L230).

The reviewer is correct: it should have been version 4.2, as was already mentioned in the text. This is now also updated in the table.

*L273. Data has been collocated à data **have** been collocated*

Updated, thank you.

L299 'any of the reanalysis model resolutions is considered too coarse' please correct the grammar

changed into:
…because none of the reanalysis model resolutions is considered sufficient to resolve …

L310. What's the frequency of EMEP data?

EMEP provides hourly observations. For our evaluation we use a reference three-hourly time frequency. We now clarify these time frequency aspects specifically in Sec 3.3

L325. I think 'multiannual' is one word. At least, please be consistent; 'multi-annual' it's hyphenated a few lines below.

We now consistently write '**multiannual**'

Figure 2 appears to be repeated or at least I can't discern any difference between the top two and bottom two rows. In addition, please, explain in the caption what 'Season: AYR' means ('all year'?) or remove it from the legend.

The reviewer is correct about the duplication, we apologize for this. The reference to 'Season: AYR' (referring to full multiannual averaging as compared to seasonal averaging) is removed from the figures.

L325-342. What about the large discrepancy between the sondes and all the reanalysis near the surface at NH subtropics and, to a lesser extent, the tropics?

The near-surface discrepancy for the NH-subtropical region can mostly be attributed to positive biases in any of the reanalyses against the Hong Kong (114.2° E, 22.3° N) sonde observations, see also Figure R1 below. O_3 at the Hilo (155° W, 19.4° N) and Naha (127.7° E, 26.2° N) stations perform much better at these low altitudes. Likewise, for the tropical region a large bias could be attributed to the Kuala Lumpur, Malaysia station (101.3°E, 2.7°N). But due to the sparseness of the observations in these regions it remains difficult to derive general conclusions.

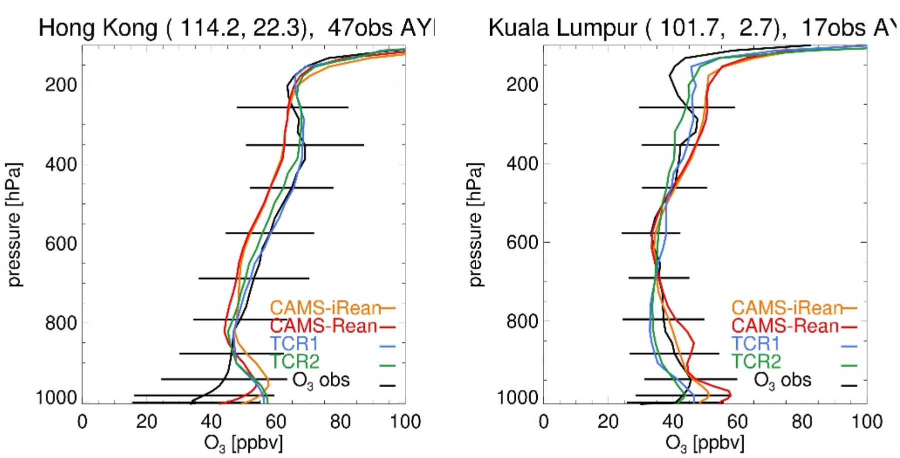


Figure R1. Evaluation of multi-annual mean ozone from all reanalyses sampled at the Hong Kong (left) and Kuala Lumpur (right) stations.

In the manuscript we now write:

In the NH subtropics and the tropics regions the reanalyses show some larger deviation against sonde observations at lower altitudes, which was traced to comparatively large biases at the Hong Kong and Kuala Lumpur stations. Note that in these regions the ozonesonde network is sparse, while the spatial and temporal variability of ozone is large, which limits our understanding of the generalized reanalysis performance (Miyazaki and Bowman, 2017).

L341-346. The comparison with ACCMIP would be easier to see if the biases shown here *were given as percentages in addition to absolute values.*

We now also report on normalized biases (and standard deviation) in new figures in the Supplementary Material. We include a statement on the maximum normalized (absolute) mean bias being below 10%.

L370. Could you briefly justify the use of an 'ozonopause' rather than more commonly used lapse rate or dynamical definitions of the tropopause? In addition, because of the high vertical resolution of ozonesondes they're likely to attain 150-ppbv threshold at very different (and somewhat random) altitudes than the reanalyses. How does that impact these comparisons?

In line with the comment from reviewer #2 we now use a more clear definition of the tropospheric column. We now compute this as the partial column from the surface to 300 hPa. Indeed, this helps to intercompare the reanalyses, as alternatively the altitude of the tropopause level changes between reanalysis.

At the start of Sec 4.2 we now write:
Collocated partial columns from the surface up to 300 hPa, hereafter for brevity referred to as 'tropospheric columns', have been compared to partial columns derived from the sonde observations.
All figures and reporting on error statistics has been updated accordingly.
L373. SH midlatitudes also look messy, especially TCR reanalyses. The absolute RMSE may be less than at high latitudes but relative to the mean column it looks quite high. Here and elsewhere it might be helpful to provide percent values for the mean biases and RMSE.

Following the reviewer's recommendation we now also compute the normalized biases. They indeed indicate difficulties over the SH mid latitudes, although smaller compared to the high latitudes. We now write:

Over the SH mid latitudes the reanalyses show similar features as over the Antarctic, with normalized mean biases within -1DU (-5%, CAMS-iREAN) and 1.5 DU (+10%, TCR-1). The normalized standard deviations over the SH mid latitudes are within 7%, marking a considerably better ability to capture temporal variability than over the Antarctic.

*L390. 'These figures'. It's **one** figure (multiple panels)!*

Changed to '**the panels in this figure'**, thank you.

Figure 4. The caption says that ozonesondes are shown in black but since what's shown is biases w.r.t. the sondes the latter are not really shown at all, are they? I suggest deleting that sentence. Also, please state that numbers of observations are shown as gray dashed lines, even if it's obvious from the previous figure. As a side note, I'm not against multipanel figures but I don't think I've seen one with 21 panels before.

The reviewer is correct: the sentence is deleted now, and explanation of the gray dashed line is included instead.

L401. Why is MIPAS relevant to the troposphere? Is that an indirect impact of assimilating total ozone with stratospheric ozone constrained by MIPAS data? The same question applies to line 451-453 (Antarctic ozone).

The reviewer is correct. To explain this better, in Sec 2.1 the following sentence is included:

Profile observations from limb instruments (MIPAS and MLS) are used to constrain the stratospheric contribution of the total column. In combination with the assimilated total column retrievals this implies that also the tropospheric part is constrained (Inness et al., 2013).

And in sec. 4.3, when discussing the impact of MLS:

Combined with total column retrievals, assimilation of such stratospheric profiles has been shown to also affect the tropospheric contribution (Inness et al., 2013).

L430-433. Any idea what happens around 2010-2011 that causes this improvement over Japan?

It is very difficult to attribute the change in bias statistics over Japan around 2010 for the four reanalyses. Aspects that play a role are following:

- The ozone observations at 650 hPa show relatively large annual mean values, during 2010 and 2012, see Figure S1 in the (original) supplementary material. This may be associated to the increased NOx emissions from China in the preceding decade (e.g. Verstraeten et al., 2015), which show a maximum during 2011 – 2014 (van der A et al., 2017). Note that in the CAMS reanalyses NO_x emissions are not optimized in the data-assimilation procedure, although NO_2 tropospheric columns have been assimilated. Instead an annual trend is assumed in the MACCity based emissions.
- The TCR-based reanalyses show a significant change in their characteristics after 2010 due to a reduction in TES retrievals, which stopped completely after June 2011.
- Both the TCR and CAMS reanalyses are affected by the row anomaly issue in the OMI O_3 (relevant to CAMS-REAN and CAMS-iREAN, particularly during 2009-2012, Inness et al., 2017) and NO_2 retrieval products (relevant to all reanalyses).

It is unfortunately beyond the scope of this work to assess the partial contributions of these effects. To provide more clarity, in the manuscript we now write:

The changes in performance statistics for all reanalyses likely have multiple causes. This includes trends in the observed ozone (Verstraeten et al., 2015), associated to changes in Chinese precursor NO_x emissions (e.g. van der A et al., 2017). Also changes in the observing system are important to consider, particularly the reduction of assimilated TES measurements in TCR from 2010 onwards, and the row anomaly issues affecting assimilated OMI O_3 and NO_2, see also Sec. 2.5.

L445-446. The CAMS reanalyses show some large departures before 2005, especially at 382 hPa. Can you comment on that?

We attribute this to similar causes as identified for the NH polar region, namely the use of early SCIAMACHY and NRT MIPAS O3 retrievals, which are of poorer quality than the OMI MLS observations which have been used from August 2004 onwards, and reprocessed MIPAS data used from January 2005 onwards. We now write accordingly:

Furthermore, CAMS-iRean and CAMS-Rean suffer from relatively large negative biases before 2005, particularly at 382 hPa. This is attributed to similar causes as have been discussed for the Arctic region.

L507. But it's not exactly the same period, is it? Figure 6 shows aggregated data from 2005 to 2012 and S3 is extended through 2014.

That is correct, therefore we provide the exact time range in all the table and figure legends. We now also specify this additionally in this particular sentence.

Figure 8. The caption says 'left' and 'right'. It should be 'top' and 'bottom'. Alternatively, the panels could be labelled.

We now change to 'top'/'bottom', thank you.

L652. 'Figure 11'. I think it should be '12'.

The reviewer is correct, this is now changed, thank you.

L727. Here and elsewhere, please provide percentages in addition to absolute values. How large (small) is 6 ppbv in this case?

We now include such assessment here. We write:

Normalized to local mean O_3 from the CAMS Reanalysis, the standard deviation values at 850 hPa reach 20% over Australia and up to 50% over South America and Central Africa. At 650 hPa these maximum ratios decrease to approx. 10% (Australia) and 20% (South America and Central Africa).

L806. Could you expand on this? It would be very helpful to include a paragraph with specific recommendations for the users: What kind of studies are these reanalyses good for? Which reanalyses are recommended for a particular type of study and which ones are less reliable? Are there any types of problems for which these reanalyses are not useful? This is partially addressed in the second to last paragraph where the authors delineate some issues related to trend and long-term variability studies using reanalyses but I think this type of discussion could be expanded to other areas.

The reviewer is correct that such suggestions could be useful. We now include the following sentences:

The well-characterized, small mean bias in tropospheric columns in these reanalyses suggest that they can be used to provide a climatology of present-day tropospheric ozone. This may serve as a reference for the present-day contribution of tropospheric ozone to the radiation budget, or may provide a climatology for a-priori ozone profiles as required for satellite retrieval products (e.g., Fu et al., 2018). The ability of the CAMS Reanalysis to capture the variability of (near-)surface ozone on multiple time scales, and for many regions over the globe, indicates it is fit for use as boundary conditions for hindcasts of regional air quality models.

L 810. Do you really mean 'any' models or is it 'many' models?

We refer to the model configurations discussed in our evaluation. We now rewrite to:
The relatively coarse horizontal resolution in any of the global reanalysis configurations could also cause significant errors at urban sites.

References

Davis, S. M., Hegglin, M. I., Fujiwara, M., Dragani, R., Harada, Y., Kobayashi, C. et al. (2017). Assessment of upper tropospheric and stratospheric water vapor and ozone in reanalyses as part of S-RIP,. Atmos. Chem. Phys., 17, 12743-12778, https://doi.org/10.5194/acp-17-12743-2017

Verstraeten, W. W., Neu, J. L., Williams, J. E., Bowman, K. W., Worden, J. R., and Boersma, K. F.: Rapid increases in tropospheric ozone production and export from China, Nat. Geosci., 8, 690–695, doi:10.1038/ngeo2493, 2015.

van der A, R. J., Mijling, B., Ding, J., Koukouli, M. E., Liu, F., Li, Q., Mao, H., and Theys, N.: Cleaning up the air: effectiveness of air quality policy for SO2 and NOx emissions in China, Atmos. Chem. Phys., 17, 1775–1789, https://doi.org/10.5194/acp-17-1775-2017, 2017.

Response to the second reviewer

We thank the referee for his/her efforts to provide this critical review, which contain many useful comments and suggestions. Below we answer them to our best ability. This has substantially helped to improve the manuscript. The reviewer comments are in italic. Our responses are in regular font, and changes to the manuscript are given in bold.

This paper intercompare four tropospheric ozone reanalyses against independent observations. Each reanalysis and the independent observations are relatively well described. The intercomparison is done between 2003 and 2017 over a large number of diagnostics covering different situation of tropospheric ozone chemistry. There are nevertheless many shortcomings in this manuscript. First, the four reanalyses are not independent (two – CAMS-iREAN and TCR-1 – are the ancestor of the two letters – CAMS-REAN and TCR-2) which is confusing. Moreover, TCR-1 seems to have changed since its published paper (Miyazaki et al., 2015) which is even more confusing. There is a lot of discussion on the impact of change in the observing system during the reanalyses but these are not clearly shown. Finally, the overall presentation is poor – figures and text – which make the paper difficult to be recommended for publications after minor revision. Here below are my detailed comments on the paper where I provide direction for improving the manuscript.

We thank the reviewer for this summary of his/her main concerns. We address them below responding to the major comments. As consequence of this review, we have substantially revised the manuscript, which can hopefully be appreciated by the reviewer.

Major comments.
There are several aspects of the study that should be revised before the paper be accepted in GMD which are listed below:
1. The paper uses four reanalyses which are by far not independent. CAMS-REAN has been built above CAMS-iREAN in order to solve some of its shortcomings. This is the same for TCR-2 vs TCR-1. For me, the authors need to refocus the study by comparing only CAMS-REAN and TCR-2. If they want to compare CAMS-REAN and CAMS-iREAN, this should be done in a separate section. For TCRs, such a section is necessary since no publication have done a dedicated comparison as it is the case for CAMS in Inness et al. (2019).

We acknowledge that the four reanalyses are not equally independent, which is clearly reflected in the naming of the products. We also agree that the newer reanalyses can overall be considered as improvements with respect to the predecessor versions, as we also conclude in the manuscript.
The reviewer is correct that Inness et al. (2019) has presented some evaluations of tropospheric ozone, intercomparing the CAMS reanalysis with the CAMS Interim Reanalysis. Nevertheless, Inness et al. (2019) covers much more aspects of the composition reanalysis, at the expense of level of detail of the evaluation of tropospheric ozone. Therefore we believe that providing this evaluation is still useful.

Furthermore, we believe it is fully meaningful to compare the reanalysis performance between the different versions of chemical reanalyses produced using similar frameworks (TCR-1 vs TCR-2 and CAMS-iRean vs CAMS-Rean). This allows us to demonstrate the impact of updating

the data assimilation configurations on the performance of the reanalyses. It also provides information whether the recent reanalyses have got closer, in any of the aspects analyzed in this manuscript. These can be expected to provide important information on future developments of chemical reanalysis. As seen in the manuscript, strong statements were already made on the CAMS-Rean and TCR-2 comparisons.
To clarify this aspect, we now write in the revised manuscript, in the introduction:

Even though these four reanalysis products are not equally independent, each of their configurations show substantial differences which are bound to impact the performance of the reanalysis products. This intercomparison aims to reveal to what extend the reanalysis products agree, depending on region and time periods.

2. There is a large confusion between TRC-1 (Miyazaki et al., 2015, available here https://ebcrpa.jamstec.go.jp/ miyazaki/tcr) and the version used in this paper. First, two different names should be used for these two different products. TCR-1 being already used, I suggest TCR-M (for MIROC) or anything that would clarify the confusion. But TCR-M seems closer to TCR-2 than TCR-1, except for the model spatial resolution. Moreover, on the TCR-1 webpage, it seems that surface NOx has been updated from Miyazaki et al. (2015) so it is difficult to know what is really TRC-M. In the revised paper, and in the section comparing TCRs reanalyses as suggested above, the authors should compare TCR-2 and TCR-1, not TCR-2 and TCR-M.

Thank you for these suggestions. We agree that there have been some confusions. To solve the problem, (1) the TCR-1 website (https://ebcrpa.jamstec.go.jp/~miyazaki/tcr) has been updated. Now the original TCR-1 data using CHASER model (Miyazaki et al., 2015), as well as the updated version, as used in this manuscript, using the MIROC-Chem model (Miyazaki et al., 2017; Miyazaki and Bowman, 2017) are both provided on the TCR-1 website. So, now any reader can access both versions. Because the data assimilation settings are similar except for the forecast model, both versions are considered to be kinds of TCR-1. More detailed statements about these TCR-1 products are given in the revised manuscript to avoid any confusions. At the start of sec. 2.3 where we now write:

A revised version of the TCR-1 data is used in this study. A major update from the original TCR-1 system (Miyazaki et al., 2015) to the system used here (Miyazaki et al., 2017; Miyazaki and Bowman, 2017) is the replacement of the forecast model from CHASER (Sudo et al., 2002) to MIROC-Chem (Watanabe et al., 2011), which caused substantial changes in the a priori field and thus the data assimilation results of various species.

3. The paper lack a dedicated section on the changes in the observing systems and its impact on the reanalyses which is largely commented throughout the paper. How does the time series of the Obervation-minus-Forecast statistics affected by these changes? Or the χ^2-test, or the spread of the ensemble for EnKF systems, or the size of the analysis increments, or the number of relevant observations, or the comparison with a control run... This is essential for the users to know what they could expect – and what they can't – from these products.

In response, changes in the observing systems indeed appear crucial to explain the behavior of the time series. The use of various satellite data streams is already mentioned in the manuscript, particularly Tables 2, 3 and 4. For detailed information of the assimilation statistics the reader is referred to Inness et al (2019) and Miyazaki et al. (2015), which we do not intend to repeat here. Nevertheless, we now provide a new, dedicated section to discuss issues associated to the temporal consistency of the observing systems (sec 2.5), where we summarize the main issues with respect to the CAMS and TCR reanalyses. This now also includes references to the first guess and analysis departures relevant to the CAMS reanalyses, and reference to χ^2 analysis relevant to TCR.

Furthermore, in the evaluation section we are now more specific as to which change we refer to, where 'changes of the observing system' are mentioned as a cause of artifacts.

Regarding the use of the assimilated observations, the paper discuss ozone reanalyses in the polar region where TCRs are poorly constrained (no TES observations poleward 72 deg). What is not said in the paper is that CAMS reanalyses are probably not well constrained as well in the winter poles since the assimilated ozone column are from UV sensors which are blind during the polar night. In the revised manuscript, I suggest removing all the discussion related to the polar regions (thus removing these regions also from the figures).

The CAMS reanalyses do not use O_3 total columns observations at solar elevation below 6° (Inness et al., 2019), which indeed implies that the CAMS reanalyses are not directly constrained during polar winters. Limb observations are used over a wider range of conditions, putting some constraints on tropospheric ozone as well. Therefore, as also suggested by the reviewer in a specific comment below, we move our comment on the TCR to Sec 2.3, and additionally we now include a comment in Sec 2.1 specifically on the CAMS systems:

Note that no total columns are assimilated for solar elevations less than 6°, hence excluding polar winters.

Nevertheless, we do not agree with the reviewer that any evaluation during polar conditions should be removed. Figure 4 of the original manuscript (time series of biases) in fact show that the tropospheric ozone during conditions where direct observations are absent are still influenced from satellite observations, as the biases are actually affected by changes in the observing system (e.g. the use of early SCIAMACHY and MIPAS retrievals during 2003). Also we believe it is worth evaluating the quality of the reanalyses for such conditions for any potential users. Although not perfect, the evaluation statistics still shows mostly acceptable values (with exception of TCR-1 over the Antarctic, and CAMS reanalyses before 2005), which could make this a useful product within its uncertainties.

4. The figures need to be improved. The resolution of all the figures are too small. Many readers, like me, will try to zoom into them in the PDF document, which is not possible with their current resolution. Please, increase them. For the line plots, add a grid in the background of the figure. In general, the fonts are too small, they must be increased, as well as the line width. The legends are not always complete, please, describe everything shown in the figure. E.g. in Fig. 4, what is the dashed line referring to the left y-axis (which I cannot read due to the small size of the fonts)? You must also write what is shown when biases are plotted: obs-reanalyses or reanalyses-obs. If normalized differences, what is the norm? In Fig. 5, the colour levels in the bias are not very well chosen because it appears that all of the reanalyses seems to be highly biased. Why not using a constant colorbar with large steps showing only relevant differences? To extract major signal from the time series, I am suggesting plotting moving *average allowing to detect the major differences between the observations and the reanalyses. Also, their readability will be improved by plotting the values of CAMS-REAN and TCR-2 only.*

We apologize for the quality of the figures in the manuscript published in GMDD, which was indeed generally not sufficient. We will ensure figures with better quality for the revised manuscript.
Likewise to Figure 3, the gray dashed line in Figure 4 refers to the number of stations that contribute to the statistics (right vertical axis). This is now included in the legend.
Biases are always defined as 'reanalysis-observation', which is the most obvious for this type of validation activity. A corresponding sentence has been introduced in the manuscript at the start of Sec. 4.1, as well as in label of the new Figure 3.
Normalization is done with respect to observations, as now included in the legend of new Figure 7. The color levels were chosen non-linear on purpose, as we believe the order of magnitude in bias values is the most relevant information, particularly in this type of figures showing bias on a global scale. Nevertheless, we simplified and optimized the color scale such that the relevant information is more easily visible from the figures. The legends in Figure 9 in the revised manuscript have been increased.

5. Many aspects of the conclusions and in the abstract are not shown in the paper, e.g. the impact of the change in the observing system or the differences between the forecast models. On the other hand, the performance of the reanalyses in different tropospheric layer, conditions and seasons – which what this paper discusses – is almost ignored. In the conclusions it should make clear of what are the findings of this paper and what are subjects for future research.

We agree with the reviewer that the abstract and conclusions can be improved to better reflect the findings of this work. In response, we have revisited the conclusions by reporting quantitatively on the biases in tropospheric columns, and on important changes in the observing systems throughout the reanalyses, affecting the long-term consistency:

For instance, averaged over the NH mid latitude region the mean bias in tropospheric ozone columns (surface to 300 hPa) is -0.3 DU (corresponding to approx. 1% of observed tropospheric column) for CAMS-Rean, which was 0.8 DU (3%) in CAMS-iRean.
(…)
Similar to the CAMS reanalyses, for the NH mid latitudes the mean bias in tropospheric columns against ozone sondes improved from 1.8 DU (7%) in TCR-1 to 0.8 DU (3%) in TCR-2.
(..)
Also changes in the NO_2 observing system, including the OMI row anomaly after December 2009 and the limited temporal coverage of SCIAMACHY and GOME-2, are considered to affect long-term consistency. These results indicate the requirements for additional observational information and/or stronger inflation of the forecast error covariance for measuring the long-term analysis spread corresponding to actual analysis uncertainty.

In the abstract we have added the following sentence, to identify the quality of the latest reanalysis products:
For instance, for the NH mid latitudes the tropospheric ozone columns (surface to 300 hPa) from the updated reanalyses show mean biases to within 0.8 DU (3% relative to the observed column) with respect to the ozonesonde observations.

6. The writing lack of clarity. For example, I do not understand the first sentence of the introduction. A careful reread of the paper is necessary to improve its readability. See some example in the specific comments.

We have improved the formulations throughout the manuscript, particularly at the sentences identified by the reviewer, and the conclusions section. Thank you for addressing this.

Other general comments

1. Tables 5-9 provides a summary of the performances of each reanalysis compared to independent observations. This information is important and the values in the tables are mentioned throughout the paper. I have two major concerns with these tables. First, extracting the comparison between the reanalyses is difficult and I suggest replacing the tables by bar-plots. Second, I suggest replacing the RMS with the standard deviation of the difference. The RMS combines a measure of the bias and the variability of the difference. Since the bias is already provided, the standard deviation will tell us by how much the differences are distributed around the bias. For these figures, TCR-1 and CAMS-iREAN could be compared with their updates versions.

These are good suggestions, thank you. We now compute the unbiased standard deviation, and provide the information in terms of bar-plots, see new Figures 3 and 5. We note that the information on the standard deviation now closely relates to the correlation analysis.

2. Also regarding differences, how are them calculated: obs-rean or the opposite? When normalized, what is the norm?

All biases are computed as 'rean-obs'. The normalization is always done with respect to the observations. We now include such comm
Biases are always defined as 'reanalysis-observation'. A corresponding sentence has been introduced in the manuscript at the start of Sec. 4.1, as well as in label of Figure 3. Normalization is done with respect to observations, as now included in the legend of new Figure 7, and at the start of Sec 4.1:

Corresponding mean biases […] are given in **Figure 3, where the bias is defined as the reanalysis-observation, throughout this work. The normalized values, as scaled with the mean of the observations, are given in Figure S1 in the Supplementary Material.**

3. In Figure 3, the authors define the tropopause in each product as the altitude where ozone exceeds 150 ppbv which means that the altitude of the tropopause change from a product to the other. I suggest taking a surface pressure as defining the upper level of the free troposphere, e.g. 200 or 300 hPa. By using 300 hPa, they will be able to remove Fig. 12, which I suggest.

The definition of the top altitude defining the troposphere indeed deserves some further consideration. The argument for choosing the 150 ppbv level is that in this way the tropospheric columns, as predicted by the reanalyses, and as observed from the ozone soundings, can most clearly be intercompared. But this indeed does not correct for any discrepancies in the altitude

of the chemical tropopause level between the reanalyses, and hence the actual partial columns within a pressure range can give a different values. This is particularly relevant for conditions where the reanalysis shows a significant under-estimation of the tropopause altitude, which would not be penalized. Indeed, using this metric, as a most remarkable change the TCR-1 performance over the Antarctic now shows decreased performance with mean bias of 2.6 DU instead of 2.1 DU.
Therefore we agree now to evaluate the O_3 PC from surface to 300 hPa. Also in the time series plots (new Figure 4) the 300 hPa level is now used. Differences in performance quality for the other reanalyses, and for regions are overall similar, so this does not affect our conclusions.

The key difference of (old) Figure 12 with respect to Figure 4 is that in Figure 12 the tropospheric ozone is *not* sampled at the locations of the observations, but assessed for the whole latitude band. Particularly for the tropics, but also for the Antarctic region this makes a large difference, relevant for the interpretation, which is otherwise not highlighted. Nevertheless, considering the length of the manuscript, together with the limited additional value, we agree to move this figure to the Supplementary Material and only briefly refer to it.

Also, why showing the number of stations and not the number of soundings?

We choose to present the number of stations in the figure, as we believe this quantity is most suitable for representing any changes in the evaluation configuration relevant to explain potential jumps in the reanalysis performance. Changes in the number of actual observations for different month would not reflect this, but would instead give a better indication of the robustness of the evaluation. Please note that in Figure 1 the number of observations per station that is contributing to the statistics has been indicated already.

4. Regarding the use of the observations and in addition to my major comment above, the Tables 2 and 3 need to be revised.
(a) As far as I know, there is only one CCI product for SCIAMACHY/GOME-2 TC and MIPAS profiles. I thus recommend to remove "(BIRA)" and "(KIT)".

The reviewer is correct, we now remove this in Table 2.

(b) What version of SCIAMACHY CCI is used? Same for MIPAS CCI, and GOME profiles? (I understand that NRT products have version changing during the time but this should not be the case for scientific – or offline – products.)

The ERS GOME profiles used in CAMS-iRean are a version provided by the Rutherford Appleton Laboratory (RAL) that was also used previously in ERA-40, Munro et al. (1998). The MIPAS, GOME-2 and SCIAMACHY CCI data were obtained from http://cci.esa.int/ozone. To be more precise, the CAMS reanalyses used the HARMOZ_MIPAS/fv0004, TC_GOME2-A/B fv0100 and fc0300, and TC_SCIAMACHY/fv0300 data.
We now specify these version numbers in Tables 2 and 3.

(c) Also, does CAMS-iREAN and CAMS-REAN both assimilated MIPAS ESA NRT and CCI profiles? Which seems to use twice the profiles of the same instruments? Please, clarify. I am also surprised to see that CAMS use MIPAS NRT, a product older than 15 years and which was reprocessed by ESA several times (the ESA offline v7 is now the latest validated version).

The MIPAS NRT data were only assimilated for the period between January 2003 and February 2004, because no reprocessed CCI MIPAS data were available from the HARMOZ_MIPAS/fv0004 product for dates before 2005. For future reanalyses this dataset should be revisited to resolve this inconsistency.

(d) You also mention MLS V3.4 which does not exist (at least for the offline products) – this is it either V3.3 or V4.2 (or shortly V3 or V4).

We should clarify that the CAMS-interim reanalysis was using the V3.4 from January 2013 onwards, i.e. not the offline product. Note that V3.4 is documented in https://mls.jpl.nasa.gov/data/v3_data_quality_document.pdf . We now add this link in the manuscript.

(e) I would also add the reference to each dataset in an additional column.

We acknowledge that including references helps traceability, and also gives proper credit to the retrieval providers, if not given yet in the text. We now include full references in the tables.

(f) The MLS version used in TCR-1 and TCR-2 are not clear. Version 4 is mentioned in the text while Table 4 mention version 3. Please clarify. Also use the appropriate MLS data quality document when referencing a version.

The reviewer is correct: this should have been version 4.2 both for TCR-1 and TCR-2. This is now updated. We now also refer to Livesey et al. (2018) rather than Livesey et al. (2011).

5. The terminology of "error statistics" is misused in the paper. It is generally applied to the error statistics in the DA system (i.e. B and R matrix and model error if any). In the case of this study, it is applied to the differences between the reanalyses and the observations so I would use the "observation-minus-analysis" statistics instead.

Thank you for this comment. Our use of the wording 'error-statistics' is meant rather general, but may indeed be confusing in this context. We believe "observation-minus-analysis statististics" is also not appropriate, as this generally refers to the error statistics of any reanalysis against observations that are actually assimilated. Instead, we now change 'error-statistics' into 'reanalysis performance statistics'

6. The authors use the inter-annual variability (IAV) and elsewhere deseasonalized anomaly, which seems to reflect to the same quantity. Could they clarify and use only one of those terminology?

In our manuscript we analyze the inter-annual variability (IAV) of monthly mean variables. For this purpose we compute and assess the deseasonalized anomaly, by subtracting the multi-year average monthly mean concentrations from their instantaneous values, similar to what is for instance presented in Davis et al., (2017). To prevent confusion we now make a more strict difference in our referencing to IAV (which refer to variability in the absolute values), and anomalies with respect to the mean value.

7. I prefer the acronyms CIRA and CAMSRA, it is much easier when speaking than CAMS-iREAN and CAMS-REAN.

We agree that the definition of these acronyms is a little subjective, and CIRA and CAMSRA may be easier to read and pronounce. Nevertheless, the use of CAMS-iRean and CAMS-Rean was chosen to stress its common assimilation framework, in analogy to TCR-1 and TCR-2. Therefore we choose to stick to these acronyms. There have been some inconsistencies between use of capitals or not, this is now also resolved.

8. Many acronyms are undefined and should be

We went through the manuscript and now consistently defined acronyms at first appearance.

Specific comments
L13-16: "Global tropospheric ozone reanalyses constructed using different state-ofthe-art satellite data assimilation systems, prepared as part of the Copernicus Atmosphere Monitoring Service (CAMS-iRean and CAMS-Rean) as well as two fully independent Tropospheric Chemistry Reanalyses (TCR-1 and TCR-2), have been intercompared and evaluated for the past decade." This is not true. CAMS-iREAN and TRC-1 are not constructed using state-of-the-art satellite data assimilation systems since these systems have been updated for CAMS-REAN and TCR-2.

We do not agree with the reviewer on this point, arguing that the data assimilation systems used either for CAMS and TCR have not fundamentally changed between the predecessor and their latest versions. The reviewer is correct that the resulting reanalyses, which depend on more aspects than the data-assimilation system (forward model configuration, model resolution, etc) cannot equally be referred to as 'state-of-the-art', but we also do not claim that. The second sentence in the abstract ("the updated reanalyses generally show substantially improved agreements..") indeed clarifies that the latest versions should be considered 'state-of-the-art'.

L18-20: "The improved performance can be attributed to a mixture of various upgrades..." This is not shown in the paper.

The reviewer is correct that we are not able to pinpoint exactly the cause of the improved performance, as that requires dedicated sensitivity experiments. Nevertheless, the improvements seen for the updated reanalyses must be a consequence of their different configuration, both in data-assimilation and forecast model, as specified in particular in Sec. 2 Therefore we now rewrite this statement as:

"The improved performance can **likely** be attributed to…"

L21-23: "Meanwhile, significant temporal changes in the reanalysis quality in all the systems can be attributed to discontinuities in the observing systems." Idem, this is not shown in the paper.

We now provide a specific section (Sec. 2.5) where we summarize the changes in time in the observing system, and also throughout the various evaluations we refer to specific changes. Therefore we consider this to be shown by our evaluations.

L22-24: "To improve the temporal consistency, a careful assessment of changes in the assimilation configuration, such as a detailed assessment of biases between various retrieval products, is needed." Which is what this paper should have been shown.

Here we do not fully agree with the reviewer. This paper is meant as an a-posteriori evaluation of the reanalysis products, and it is beyond the scope of this work to analyze biases between retrieval products. This has is in part been addressed in Inness et al (2019), see their Sec. 3.2, and Figure 6, as well as Figures S1-S3 in their Supplementary Material. Nevertheless, the posteriori evaluation shown in our work indicates various other jumps which cannot be explained from changes in foreward model configuration, and hence implies biases between retrieval products. Likewise for TCR, changes in performance are detected which have already been briefly addressed in Miyazaki et al (2015), and hence do not need analysis here. The recommendation written in our abstract addresses the identified issue of biases between retrieval products, which needs to be addressed in future reanalysis configurations to obtain an improved consistency over time in tropospheric ozone reanalyses.

L24-26: "Even though the assimilation of multi-species data influences the representation of the trace gases in all the systems and also the precursors' emissions in the TCR reanalyses, the influence of persistent model errors remains a concern, especially for the lower troposphere." Again, this is not shown in the paper.

The reviewer is correct that we do not assess the impact of model errors in the scope of this work, but only make various references to its potential impact. Therefore we agree to remove this sentence from the abstract. We still believe there is sufficient evidence that part of the discrepancies seen in the observations are due to biases in model parameterizations, which would justify the last sentence of the abstract, discussing potentials for improvement.

L31-32: "The global distribution of present-day tropospheric ozone..." I don't understand this sentence, please, rephrase.

Thank you for your fair comment. We have rewritten, and thereby simplified, the formulation of this sentence into:

Both human activity and natural processes influence the global distribution of present-day tropospheric ozone, together with its interannual variability and trends.

L41: "...tropospheric ozone, but are generally..." => "...tropospheric ozone, which is generally..."

Changed

L45: "Tropospheric ozone is reasonably well monitored..." You are talking about surface ozone in this sentence so I would write "Surface ozone is reasonably..."

Changed.

L50-52: This list of satellite dataset is incomplete (missing are e.g. OMPS and TROPOMI for the most recent instruments) so I would write "These observations are *complemented with (combined) satellite observations from, e.g., GOME-2, ...”*

We changed this into:
(…) **satellite observations from instruments such as** (…)

L62-64: “Simultaneously international modelling initiatives...” I don’t understand this sentence, please, clarify.

This sentence is meant to address some of the main coordination and collaboration frameworks that have emphasis on various aspects which rely more heavily on modeling, both in air quality and climate change context. To clarify better we rephrased this sentence to :

Additional coordination with the emphasis on modelling activities related to tropospheric ozone have been established, for instance (…)

L77: “...individual measurements suffer...” Do you mean “...individual measurements which suffer...”?

No, here we refer to the impact of representativity of individual observations for drawing general conclusions, i.e. undersampling, or sampling bias. We clarify this better by writing:

This was shown useful as evaluations using individual measurements are subject to significant sampling biases

L81: “...particular constellations of pollution...” What do you mean by “constellations”?

We simply mean ‘pollution events’, as directly clarified in the consecutive sentence. The reviewer is correct that the wording is a bit awkward. We have rewritten this to:

… and to analyse particular pollution events such as those associated with heat waves…

L85: “However, all of these applications presume that the reanalysis is sufficiently accurate,...” What matter is that reanalysis is well characterized more than accurate.

Strictly speaking the reviewer is correct. When well characterized, users of respective reanalyses can take such information into account in their applications. On the other hand, if the characterization of biases is complex, because of changes in time and space, then the use of any such product is still hampered. Therefore, we argue that in practice a specification of the accuracy of the reanalysis may then be more desirable. We rewrite this into:

However, all of these applications presume that the reanalysis is sufficiently accurate, or, to the least, well characterized. Despite the range of observations assimilated into the respective systems, this is not necessarily ensured.

L118-119: CAMS-REAN and CAMS-iREAN acronyms are undefined.

Now defined slightly above:

… the ‘CAMS Interim Reanalysis’ (hereafter ‘CAMS-iRean’) (…) and recently the ‘CAMS Reanalysis’ (‘CAMS-Rean’)

L126: “NOx” => “NOx”
Changed

L129: “...changing constellation of ...” => “... the change in the observing system...”

Changed

Table 1: What are the output frequency of each product. Are the output snapshots or time averages?

The basic output frequency in the CAMS products is three-hourly for the 3D-fields evaluated here, as already specified at the end of Sec. 2.2. The TCR products adopt two-hourly output. This is already specified at the end of Sec. 2.4. We think this should do.

L156: => “The meteorological model version is CY40R2.”

Thank you. Changed to:

The meteorological model is IFS CY40R2.

L157: “In terms of ozone, observations from the following set of satellite instruments have been assimilated:...”

Changed, thank you.

L159: “Limb observations are instrumental to discriminate...” => “Profiles from limb instruments (MIPAS and MLS) are used to discriminate...” Could you explain how does limb profiles are used to discriminate the tropospheric and stratospheric contribution of the total column observations?

By assimilating both total and stratospheric columns, the tropospheric columns are indirectly constrained as the residue of both elements. We now change the manuscript on this aspect writing:

Profile observations from limb instruments (MIPAS and MLS) are used to constrain the stratospheric contribution of the total column. In combination with the assimilated total column retrievals this implies that also the tropospheric part is constrained (Inness et al., 2013).

L161-163: See my general comment above regarding MLS V3.4.

We clarify that this indeed refers to the version V3.4, see also above.

L211: Remove reference to Watanabe et al. since it is already provided 2 lines above.

Done, thank you

L341: “In the TCR systems,...” Move this info in Sect. 2.3.

Sentense has been moved to Sect 2.3.

Figure 2: What is the difference between the part in page 14 and 15? It seems to be the same.

This was indeed an duplication of plots, we apologize for this.

L376: “...both model and observations...” Which model? Do you mean the reanalyses? If yes, replace by “... the reanalyses and the ozonesondes...”

The reviewer is correct. Nevertheless the complete sentence is now removed as this statement is no longer correct when analyzing the partial columns from surface to 300hPa instead.

L377: same as above “modelled” or “analysed”?

We have updated this. Also elsewhere throughout the document we have revisited the use of *‘model’* and *‘modeled’*, and changed to ‘analysis’/’analyzed’ where appropriate.

L379-381: Is the poor correlation between reanalyses and observations due to the missing total column observations during the polar night? Since, as far as I know, none of the total column assimilated data are taken by emissions instruments thus failing to measure during the night?

The reviewer is correct that no total column (in CAMS), and also no TES profile retrievals (in TCR) are assimilated during polar nights. We discuss these aspects in more detail as part of Sec 4.3, see also the reviewer comments on this issue below (as well as in our response to his/her main comments).

L397-399: “During 2003 and 2004 both CAMS reanalyses...” Why? This is not related to GOME data since CAMS-REAN does not assimilate GOME.

The 2003-2004 discrepancy compared to other years, particularly at the 350 hPa level, was attributed to the use of early SCIAMACHY and NRT MIPAS O_3 retrievals, which are of poorer quality than the observations used lateron. The GOME issue was mostly related to the differences between the two CAMS reanalyses in 2003 at altitudes below 650 hPa. The manuscript was not fully clear on this. To clarify better, we rewrote this section:

During 2003 and 2004 both CAMS reanalyses show anomalously low springtime ozone, different to the rest of the time period, particularly at ~350 hPa. The different reanalysis performance statistics 2003 over the Arctic compared to later years is attributed to the use of early SCIAMACHY and NRT MIPAS O_3 retrievals, which are of poorer quality than the OMI MLS observations which have been used from August 2004 onwards, and reprocessed MIPAS data used from January 2005 onwards. CAMS-iRean also shows a large offset compared to observations and CAMS-Rean in 2003, particularly at altitudes below 650 hPa. This was attributed to the assimilation of GOME nadir profiles in CAMS-iRean, which has been omitted in CAMS-Rean (Inness et al., 2019).

L399: "...GOME observations..." => "...GOME nadir profiles...".

Changed, thank you.

L400-403: Why does CAMS assimilate MIPAS NRT and not the offline reprocessed products delivered by ESA?

The MIPAS NRT data were only assimilated for the period between January 2003 and February 2004, because no reprocessed CCI MIPAS data were available from the HARMOZ_MIPAS/fv0004 product for dates before 2005 from http://cci.esa.int/ozone. As already commented above, in future reanalyses this dataset should be harmonized to resolve this inconsistency, which is indeed an important issue. This is now also addressed specifically in the conclusion where we now write:

Discontinuities in the availability, coverage and *product version* of the assimilated measurements are also shown to affect the quality of the reanalysis, particularly in terms of temporal consistency, both in the CAMS and TCR-reanalyses.

L412-413: "Also both the observations and reanalyses indicate an upward trend of tropospheric ozone in the UTLS..." I don't see this from figure 4. Could you clarify?

This indeed cannot be seen from Figure 4, but is visible from the corresponding Figure S1 in the Supplementary material, presenting the O_3 monthly mean values over the given regions and altitude ranges. The NH polar region at 378 hPa shows a clear sign of an upward trend, both in observations and reanalyses. We now make explicit reference to this figure in the manuscript, which was missing indeed.

L431-433: "From 2011 onwards the correspondence with observations improves remarkably, despite the lack of TES measurements in TCR from June 2011 onwards."
Why?

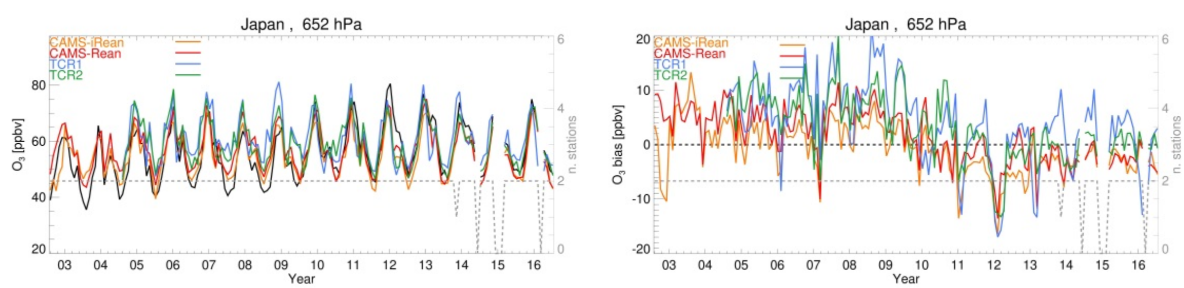


Figure R2: Absolute value (left) and mean bias (right) of O_3 at ~652 hPa against sonde observations over Japan.

The changes in bias characterization of the reanalysis is obvious from Figure R2, but the reason for this is not well understood. Not only the absolute values show an upward trend over the 2003-2016 time period (Figure R2, left), which seems absent in the reanalyses, but also there are changes in the observing system. We now write:

The changes in performance statistics for all reanalyses likely have multiple causes. This includes trends in the observed ozone (Verstraeten et al., 2015), associated to changes in Chinese precursor NO_x emissions (e.g. van der A et al., 2017). Also changes in the observing system are important to consider, particularly the reduction of assimilated TES measurements in TCR from 2010 onwards, and the row anomaly issues affecting assimilated OMI O_3 and NO_2, see also Sec. 2.5.

L434-444: I do not agree with most of what is written.
“In the tropics, ...” This is not true for CAMS-iREAN which generally underestimate the ozonesondes.
“... both CAMS reanalyses show a strong peak ...” In fact, TCRs also show a peak.
“...overestimation of up to 20 ppb.” None biases are going up to -20 ppb. I would rather say -15 ppb. Do not omit the sign of the bias in the comparison.
“This spike appears much weaker in TCR...” Does the reason not due to the fact that TCR also optimize surface emissions allowing the reduce the bias with observations?

But the authors does not discuss the fact that CAMS-iREAN seems to have the best agreement with ozonesondes during the whole period and they should comment on the reason for this.

We thank the reviewer for closely checking our analysis. We have updated the comment on the mean bias before 2012. Also the exceptional peak in 2015 was only visible at the ~850 hPa altitude, only for CAMS reanalyses, and to much lesser extent at ~650 hPa. We confirm that the sign of the bias (reanalysis-observation) is positive, and reaches 20 ppb. As the reviewer suggests, the discussion why TCR behaves differently than CAMS, with on average more acceptable O_3 values, is possibly not only due to the sampling issue, but can also be associated to better optimized NO_x emissions compared to those from GFAS, as used in CAMS.
The CAMS-iRean is not superior to CAMS-Rean at the 650 and 350 hPa altitude range; it is unfortunately not clear what is the reason for the better performance before 2012 at the 850 altitude range, although a likely explanation appears the change in MLS version used in CAMS-iRean from 1 January 2013 onwards.
In summary, following his/her comments, we change this section into:

In the tropics, all reanalyses except CAMS-iRean overestimate ozone at 850 hPa before 2012, with positive biases in the range 2.5-3 ppb. The different performance for CAMS-iRean from 2012 onwards is probably associated to the use of another version of the MLS retrieval product. Interestingly, both CAMS reanalyses show a strong peak in ozone at 850 hPa during the second half of 2015 (see corresponding Figure S1 in the Supplementary material), but with a zonally averaged overestimation of up to 20 ppb. This is associated to the strong El Niño conditions, and this particular spike was attributed to an over-estimate of ozone observed at the Kuala Lumpur station for October 2015. Here exactly the grid box affected by the extreme fire emissions in Indonesia for this period (Huijnen et al., 2016), as prescribed by the daily GFAS product, has been sampled. This peak appears much weaker in TCR. Possible explanations are lower optimized NO_x and CO emissions in TCR compared to those used in CAMS, resulting in

weaker ozone production, together with a coarser model resolution. At 650 hPa, the TCR reanalyses overestimate ozone almost throughout the reanalysis period (by 3.1–3.8 ppb on average), whereas the CAMS-Rean shows closer agreement with the observations (mean bias = 0.5 ppb, RMSE = 3.2 ppb). At ~350 hPa, the TCR-2 shows improved agreement compared with the earlier TCR-1, as confirmed by improved mean bias (from 4.3 to 0.6 ppb) and RMSE (from 6.6 to 5.7 ppb) although the temporal correlation remains relatively low.

L449: "332 hPa" => "382 hPa"?

Changed, thank you

L467-474: I see other reasons for the seasonal variations in the bias time series than those mentioned in this §. For CAMS products, their troposphere is not constrained by any data during the polar night since all of the assimilated nadir instruments are measuring UV sun-scattered light. For TCR, TES ozone data are only available at latitude lower than +/-72°. Could the author comment on that?

The reviewer is correct that there are no constraints on total O_3 column in the CAMS reanalyses during polar winter, neither tropospheric O_3 profiles from TES in the TCR reanalyses over the poles. Indeed the seasonal variations in the availability of satellite observations, in particular for the CAMS reanalyses, is bound to contribute to the seasonal cycle in their biases.
Likewise, if TES observations would have been available for this region then the bias in TCR-1 would probably have been much smaller. Nevertheless, as shown for the TCR-2 reanalysis, also a meaningful product with a mean bias (stddev) of within 2 (4.5) ppb at 650 hPa can be provided by optimizing the data-assimilation system, even if direct satellite observations are not available.
We revise the manuscript accordingly as follows:

The seasonal cycle in the biases can largely be attributed to the lack of O_3 total column observations during polar night, combined with a seasonal variation in model forecast biases. The TCR reanalyses largely underestimate ozone during austral summer and autumn in the lower troposphere. At 351 hPa, TCR-1 substantially overestimates ozone throughout the year because of large model biases and the lack of observational constraints. **This large positive bias was resolved in TCR-2 by improving the modelling framework.**

L473: "332 hPa" => "351 hPa"

Changed, thank you

Sect. 5.2: "Figure 6 presents the temporal variability..." Well, figure 6 is a scatter plot without any time axis (on the x-, y- or any colorbar) so I would change this sentence. Moreover, all the discussion in this §related to seasonal differences are not supported by Fig. 6. I understand that Fig. S3 could support this discussion but as being part of the supplement, it cannot be used for new discussion.

The reviewer is correct. We changed the formulation to better connect the discussion to the presented figures, and omit statements that largely rely on results presented in the Supplementary material. We have rewritten this section as follows:

Figure 6 presents scatter plots of monthly mean ozone from the reanalyses against those from the TOAR surface observations for various regions. The corresponding time series are given in Figure S3 in the Supplementary Material. As is clear from Figure S3, the main driver of the variation in magnitude of ozone concentrations in the reanalyses and observations in Figure 6 is the seasonal cycle. Over the Arctic, the general pattern in the seasonal variations is captured for all reanalyses (R between 0.58 and 0.72), although they all underestimate the increased ozone values during boreal spring.

Over Europe and the US, the CAMS reanalyses show the closest agreement with the observations (MB between -2.4 and 1.5 ppb, R>0.8). Furthermore, CAMS-REAN shows reduced negative biases for observed low ozone values compared with the CAMS-iREAN, which is in boreal winter and spring . The TCR reanalyses exhibit large positive biases over Europe and the US regions (MB between 6.7 and 17 ppb), with significantly lower biases in TCR-2. Over East Asia, all the reanalyses show positive biases in the range of 2.7 ppb (CAMS-REAN) to 10.5 ppb (TCR-1) and fail to reproduce the minimum concentrations in autumn. Still the temporal correlations are similar to most other regions (R between 0.79 and 0.83), associated with the stable seasonal cycle in both the reanalyses and observations. Over Southeast Asia, positive biases exist throughout the period, which are largest in TCR-1. For this region the TCR-reanalyses show lower temporal correlations (R between 0.39 and 0.49) compared to the CAMS reanalyses (R=0.68). Significant changes in the surface ozone biases are found in the TCR reanalyses over the SH mid latitudes, with reduced values after 2010.

The CAMS reanalyses capture well the temporal variability over the SH mid latitudes and Antarctic (R between 0.89 and 0.96), while CAMS-REAN shows a positive bias for observed high ozone values. This is associated to model biases austral winter (JJA), particularly during 2005-2013, Figure S3. The TCR reanalyses show a significant negative bias throughout the year except for observed low ozone values (during Austral summer) which results in lower temporal correlations (R~0.68).

L521: Here and at several other placed “R=0.89 – 0.96”? Do you mean “R between 0.89 and 0.96” or “Rï˘CO˝ [0.89,0.96]”? Or something else?

We refer to values between a minimum and maximum. We clarify this now by writing explicitly

R between 0.39 and 0.49 (etc)

L541: “We compute the interannual...” Do you mean the deseasonnalized anomaly for each region? See also the general comments.

As described above, we now make a more strict difference in our referencing to IAV, and to deseasonalized anomalies with respect to the mean value. Particularly, at the start of Sec. 5.3 we now write:

We assess the interannual variability (IAV) by computing the deseasonalized anomaly of surface ozone concentrations. For this, the 2005-2012 multi-annual monthly, regional mean surface ozone is subtracted from its corresponding instantaneous monthly, regional mean value, (…)

L652: Do you mean Fig. 12? So this is almost Fig. 3 without observations. Is it really the annual mean? It seems more to be a time series of monthly mean?

The reviewer is fully correct that this should have been reference to Fig. 12, and refers to monthly means rather than annual mean. This figure is analogue to Figure 3, but with the main difference that it much better reflects the average zonal mean, as it is not sampled for station locations. This figure has now been moved to the Supplementary Material, together with most of the contents of this section.

L669-670: The change in behaviour is clear above the SH polar latitude but less clear in SH midlatitudes.

The reviewer is correct, thank you. It should have written:
Particularly at the SH high-latitudes, but to lesser extend also at the SH mid-latitudes, there is a remarkable change in behaviour after 2013 in all reanalyses except TCR-1
But, following reviewer #3 we choose to remove this section from the main manuscript, in view of duplication and length. The figure is retained in the Supplementary Material.

Figure 13: Is it as Fig. 7 but for PC surface-300 hPa in south-east Asia and ENSO? "A 2-month smoothing". Do you mean a running mean or moving average? What is TSI?

Indeed a similar procedure has been followed to create Figure 13 as was done for Figure 7. For better clarity we now refer to 'deseasonalized anomalies'. The reference to 'TSI' was spurious, and has now been removed. Discussion of this figure has been moved to the end of the next section.

L742: "...annual mean..." For which year?

This actually refers to the multi-annual mean analogous to what is presented in Figure 14.

Figure 15 is very interesting but I would add the ozone sonde values in order to assess the quality of the reanalyses against the best estimation of the truth (i.e. the sondes).

This is a good suggestion. We now also compute the frequency distribution sampled for instantaneous sonde observations at three pressure levels. This indeed gives a quantitative impression of (differences in) reanalysis performances, as quantified by the total absolute difference between the frequency distributions of the reanalyses and observations. Nevertheless, an important drawback is that by sampling the analyses at the location and time of the observations the global representativity, which was central to this section is largely lost.

Therefore we choose to provide this evaluation as part of the supplementary material, figure S6. In the manuscript we now write:

A corresponding evaluation of the frequency distributions, but sampled at individual ozone sonde observations, is given in Figure S7 in the Supplementary material. Because of the different sampling approach the shape of the frequency distributions is different than was seen in Figure 15. Evaluation of the absolute differences *d* between analyzed and observed frequency distributions indicates that at 850 hPa the performance between the four reanalyses is very similar (*d* between 0.17 and 0.19), while at 650 hPa CAMS-Rean is superior (*d*=0.13). CAMS-iRean shows an under-estimate of the frequency of high ozone values (larger than ~55 ppbv) at 850 and 650 hPa, explaining the worst performance at 650 hPa (*d*=0.20). At 350 hPa the differences in performance are largest, with best correspondence to observations for CAMS-iRean (*d*=0.11), and worst for TCR-1 (*d*=0.43).

To aid the interpretation, Figure 15 is now presented in terms of bars.

L767: "The changing constellation..." I would rather say "The changes in the observing system..."

We change this, thank you for your suggestion.

L770: "This calls for a detailed evaluation of the capability of the current reanalyses of tropospheric ozone." Do you mean this is something to do in the future? Please, clarify.

Here we refer to our study. We change the sentence into:

This calls gives rise for a detailed evaluation of the capability of the current reanalyses of tropospheric ozone, as presented here.

L793-795: "In the TCR reanalysis, the chemical concentrations and precursor's emissions were simultaneously optimized through EnKF data assimilation, which was important in providing information on precursors' emissions variations (Miyazaki et al., 2014; 2017; 2019a; Kiang et al., 2018) and in improving the vertical profiles of ozone." Well, this is not shown in the paper so I would remove this comment from the conclusions.

We agree with the reviewer that this is not shown in this manuscript, and remove the sentence.

L800-803: "Meanwhile, the analysis ensemble spread ..." Well, again, the TCRs ensemble spread are not shown in the paper. Also, what do you mean with "4D-var could be used ..." Altogether, I don't understand the message in this sentence.

These sentences contain recommendations for further improvements, and are therefore not shown in the manuscript. To clarify better, we change the sentence to:

Furthermore, in future studies the analysis ensemble spread from EnKF can be regarded as uncertainty information about the analysis mean fields, indicating the need for

additional observational constraints. Likewise, in the 4-D Var system the contributions from individual retrieval products can be tested.

L413: The acronym UTLS must be defined.

We do this now at first appearance (sec 4.3)

L819: " ... a careful assessment of changes in the assimilation configuration..." Which what this paper should have done.

Here we do not agree with the reviewer. Our manuscript provides an a-posteriori evaluation of the reanalysis products, and as such provides various indications where changes in the tropospheric ozone reanalyses are linked to changes in the observing system. Our evaluations should be taken into account when designing an updated observing system and details regarding the data assimilation configuration in future reanalyses. To clarify better, we rewrite this section into:

We have shown that discontinuities in the availability, coverage and product version of the assimilated measurements affect the quality of any of the reanalyses, particularly in terms of temporal consistency. This is particularly important for assessing interannual variability. The influence of data discontinuities must be considered and where possible removed when studying interannual variability and trends using products from these reanalyses. To improve the temporal consistency in future reanalyses, a careful assessment of changes in the assimilation configuration, most prominently associated with ozone column and profile assimilation is needed, including a detailed assessment of biases between various retrieval products.

L822: "The assimilation of multi-species data influence..." This has not been addressed in the paper.

Analogous to our response above, our manuscript is not intended to assess in detail the impact of individual contributions of the data assimilation configurations on the quality of resulting reanalyses, such as multi-species assimilation, or issues regarding the CTM's. The reviewer is correct that this has not been analyzed in our manuscript, as this would require dedicated sensitivity experiments. Therefore we agree with the reviewer that we should be more accurate in our formulation. We now write:

The assimilation of multi-species data in both the CAMS and TCR configurations influences the representation of the entire chemical system, while the influence of persistent model errors in complex tropospheric chemistry continues to be a concern. Therefore, further improvements to long-term reanalyses of tropospheric ozone can be achieved by improving the observational constraints, together with a further optimization of model parameters, such as the chemical mechanism, emission, deposition, and mixing processes.

References

Davis, S. M., Hegglin, M. I., Fujiwara, M., Dragani, R., Harada, Y., Kobayashi, C., Long, C., Manney, G. L., Nash, E. R., Potter, G. L., Tegtmeier, S., Wang, T., Wargan, K., and Wright, J. S.: Assessment of upper tropospheric and stratospheric water vapor and ozone in reanalyses as part of S-RIP, Atmos. Chem. Phys., 17, 12743–12778, https://doi.org/10.5194/acp-17-12743-2017, 2017.

Munro, R., R. Siddans, W. J. Reburn, and B. J. Kerridge, Direct measurements of tropospheric ozone distributions from space, Nature, 392, 168–171, 1998.

Response to the third reviewer

We thank the referee for his/her short, but nevertheless useful, positive review, which contain various useful comments and suggestions. Here 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 paper inter-compared tropospheric ozone reanalysis products from CAMS, CAMS-Interim, TCR-1, and TCR-2. This study is of scientific importance and the research is well conducted. The presentation is generally clear with logic flow and convincing discussions. The paper provides an enhanced understanding of issues related to tropospehric ozone reanalysis products. I only have some minor issues for the authors to consider when revising their paper. Minor issues:
1. *In the abstract and conclusions, it is useful to summary where and when the reanalysis products perform strongest and weakest, in term of relative difference with ozonesonde data.*

In response, we now specify in the abstract a sentence on the evaluation against ozone sondes:

For instance, for the NH mid latitudes the tropospheric ozone columns (surface to 300 hPa) from the updated reanalyses show mean biases to within 0.8 DU (3% relative to the observed column) with respect to the ozonesonde observations.

Also in the conclusions we describe the main strengths of the reanalyses, and suggest potential application areas:

The well-characterized, small mean bias in tropospheric columns in these reanalyses suggest that they can be used to provide a climatology of present-day tropospheric ozone. This may serve as a reference for the present-day contribution of tropospheric ozone to the radiation budget, or may provide a climatology for a-priori ozone profiles as required for satellite retrieval products (e.g., Fu et al., 2018). The ability of the CAMS Reanalysis to capture the variability of (near-)surface ozone on multiple time scales, and for many regions over the globe, indicates it is fit for use as boundary conditions for hindcasts of regional air quality models.

2. *Figures and tables need more annotation.*

In response, we have extended the descriptions of (new) Figures 1, 4, 6, 7, 15.

Fig. 1, What are the boxed areas?

We now specify the regions used in the analyses in the legend of Figure 1, together with specification of the other regions.

Table 5-7, the area definitions can be provided in Fig. 1.

This is a good suggestion, thank you. We now provide this information, see above.

Fig. 4. Relative difference is more meaningful.

We prefer to stick to the absolute values here, to remain close to the physical quantity. Nevertheless, in our revisions we now report relative differences much more frequently, e.g. by adding bar-plots presenting the relative biases and standard deviations in the Supplementary material, and referring to this in our analyses, as well as in the abstract and conclusions.

Fig. 11, what is the time zone for this figure? How are the model errors removed?

For the diurnal cycle we use UTC, we now include this in the x-axis label in the Figure. The model bias was removed by subtracting the seasonal mean analysis bias with respect to the corresponding observations. We now write this explicitly.

3. *The definition of the tropopause needs some discussion and references.*

As was also commented by the other reviewers we have updated our analysis of tropospheric columns. This now refers to subcolumns from the surface to 300hPa. In this way we circumvent any potential ambiguity regarding the definition of the tropopause, and make the reanalysis products better comparable.

4. *The paper appears lengthy. Please shorten the paper and move less significant contents to Supplements.*

The reviewer is correct that the manuscript benefits from a more stringent priority in presenting material, thank you for your comment. In response, we decided to move most of Sec. 7 into the supplementary material. Only an assessment of the correlation with the ENSO is retained, as well as the concluding sentences which describe the consistency in time series between the various renalyses.

Met opmaak: Breedte: 21 cm, Hoogte: 29,7 cm

An inter-comparison of tropospheric ozone reanalysis products from CAMS, CAMS-Interim, TCR-1 and TCR-2

Vincent Huijnen[1], Kazuyuki Miyazaki[2], Johannes Flemming[3], Antje Inness[3], Takashi Sekiya[4], Martin G. Schultz[5]

[1] Royal Netherlands Meteorological Institute, De Bilt, the Netherlands
[2] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA
[3] ECMWF, Shinfield Park, Reading, RG2 9AX, UK
[4] Research Institute for Global Change (RIGC), Japan Agency for Marine–Earth Science and Technology (JAMSTEC), Yokohama 2360001, Japan
[5] Jülich Supercomputing Centre, Forschungszentrum Jülich, Jülich, Germany

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

Abstract. Global tropospheric ozone reanalyses constructed using different state-of-the-art satellite data assimilation systems, prepared as part of the Copernicus Atmosphere Monitoring Service (CAMS-iRean and CAMS-Rean) as well as two fully independent Tropospheric Chemistry Reanalyses (TCR-1 and TCR-2), have been inter-compared and evaluated for the past decade. The updated reanalyses (CAMS-Rean and TCR-2) generally show substantially improved agreements with independent ground and ozonesonde observations over their predecessor versions (CAMS-iRean and TCR-1) for the diurnal, synoptical, seasonal, and decadal variability. For instance, for the NH mid latitudes the tropospheric ozone columns (surface to 300 hPa) from the updated reanalyses show mean biases to within 0.8 DU (3% relative to the observed column) with respect to the ozonesonde observations. The improved performance can likely be attributed to a mixture of various upgrades, such as revisions in the chemical data assimilation, including the assimilated measurements, and the forecast model performance. The updated chemical reanalyses agree well with each other for most cases, which highlights the usefulness of the current chemical reanalyses in a variety of studies. Meanwhile, significant temporal changes in the reanalysis quality in all the systems can be attributed to discontinuities in the observing systems. To improve the temporal consistency, a careful assessment of changes in the assimilation configuration, such as a detailed assessment of biases between various retrieval products, is needed. ~~Even though the assimilation of multi-species data influences the representation of the trace gases in all the systems and also the precursors' emissions in the TCR reanalyses, the influence of persistent model errors remains a concern, especially for the lower troposphere.~~ Our comparison suggests that improving the observational constraints, including the continued development of satellite observing systems, together with the optimization of model parameterisations, such as deposition and chemical reactions, will lead to increasingly consistent long-term reanalyses in the future.

1 Introduction

Both human activity and natural processes influence ~~T~~the global distribution of present-day tropospheric ozone, together with its interannual variability and trends.~~, plays an important role when describing the impact of human activity and natural processes on air quality and climate.~~ Amongst other factors, increments in surface ozone concentrations contribute to changes in air quality (e.g., Im et al., 2018), human health (Liang et al., 2018), and agriculture (van Dingenen et al., 2009). Owing to its radiative effects, tropospheric ozone is an important driver in climate ~~-~~change (Checa-Garcia et al., 2018~~), ~~). ~~even~~ Also it may affect ~~if no improvement in~~ long-range weather forecasts, even if in evaluations no improvement ~~-~~has been detected so far (Cheung et al., 2014). Considering its lifetime of a few weeks, tropospheric ozone can be controlled by both local and remote pollution sources through atmospheric chemical processes and long-range transport (Monks et al., 2015; Huang et al., 2017; Jonson et al., 2018), as well as stratospheric influx (e.g. Hsu and Prather, 2009; Knowland et al., 2017).

In addition to anthropogenic sources, natural processes such as El Niño–Southern Oscillation (ENSO) conditions affect tropospheric ozone production and loss terms, through changes in upwelling, convection, solar irradiance, humidity, and biomass burning emissions (e.g., Ziemke and Chandra, 2003, Inness et al., 2015). Other processes that potentially influence tropospheric ozone, ~~but are~~ which is generally considered of minor importance, are the quasi-biennial oscillation (Neu et al., 2014) and the North Atlantic Oscillation (Thouret et al., 2006).

Various types of datasets have been compiled to allow the analysis of the current state of tropospheric ozone and its changes over time. ~~Tropospheric~~ Surface ozone is reasonably well monitored through in-situ networks measuring surface concentrations

(Global Atmosphere Watch (GAW), Global Monitoring Division (GMD), European Monitoring and Evaluation Programme (EMEP), AirNow), as collected and homogenised by the Tropospheric Ozone Assessment Report (TOAR, Schultz et al., 2017). Above the ground ozone is monitored through ozone sondes, collected by World Ozone and Ultraviolet Radiation Data Centre (WOUDC; https://woudc.org/), and aircraft (In-Service Aircraft for a Global Observing System (IAGOS), Nédélec et al., 2015). These observations are complemented with (combined) satellite observations ~~of the~~ from instruments such as Global Ozone Monitoring Experiment–2 (GOME-2), Ozone Monitoring Instrument (OMI), Microwave Limb Sounder (MLS), Infrared Atmospheric Sounding Interferometer (IASI), Tropospheric Emission Spectrometer (TES). Here each retrieval product comes with its specific (vertical) sensitivity, which allows the derivation of tropospheric ozone columns as listed in Gaudel et al. (2018).

The multitude of observational datasets have led to observationally constrained assessments of the current state and trends in tropospheric ozone, for instance documented as part of TOAR (Schultz et al., 2017; Gaudel et al., 2018; Fleming et al., 2018; Tarasick et al., 2019). Recent studies have also shown decadal-scale changes in global tropospheric ozone using various observations, such as a shift in the seasonal cycle at northern hemisphere (NH) mid-latitudes and long-term trends over many regions (e.g., Parrish et al., 2014; Cooper et al., 2014; Gaudel et al., 2018; Fleming et al., 2018). Based on a combination of multiple ozone retrieval products, Ziemke et al. (2019) have inferred positive trends in tropospheric ozone trends, particularly in the 2005-2016 time period.

~~Simultaneously international~~ Additional coordination with the emphasis on modelling activities related to tropospheric ozone ~~modelling initiatives~~ have been established, for instance to analyse the contribution of ozone on air quality (AQMEII: Air Quality Modelling Evaluation International Initiative), the impact of long-range transport on air quality (HTAP: Hemispheric Transport of Air Pollution), and the impact of composition changes on climate change (CCMI: Chemistry-Climate Model Initiative). (e.g., Young et al., 2013; Morgenstern et al., 2018; Liang et al., 2018)

Following the concept of meteorological reanalyses such as ERA5 (Hersbach et al., 2018), observationally constrained reanalyses of the atmospheric chemical composition have been developed to provide time series of tropospheric and stratospheric ozone. A reanalysis is a systematic approach to create long-term data assimilation products by combining a series of observational datasets with a model. Advanced data assimilation, such as four-dimensional variational data assimilation (4D-VAR) and ensemble Kalman filter (EnKF), allows the propagation of observational information in time and space, and from a limited number of observed species, to an analysis of a wide range of chemical components. This can be used in reanalyses to provide consistent global fields that are in agreement with individual observations (Lahoz and Schneider, 2014; Bocquet et al., 2015). A reanalysis hence provides an instantaneous global image of atmospheric composition, together with its change over time and therefore serves in principle to analyse the mean state of the atmosphere, together with its variability and trends.

Applications of chemical reanalyses include comprehensive spatiotemporal evaluation of independent models, such as those developed in the framework of the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP, ~~(~~Young et al., 2013) and CCMI (Morgenstern et al., 2018). This was shown useful as evaluations ~~comparisons~~ using individual measurements ~~suffer~~ are subject to ~~from~~ significant sampling biases (Miyazaki and Bowman, 2017). In their study the

ACCMIP ensemble ozone simulations were evaluated using a chemical reanalysis, complementing the use of individual measurements for such purpose. The chemical reanalyses can also be used as an input to meteorological reanalyses, e.g. for radiation calculations (Dragani et al., 2018), and they can provide boundary conditions to regional-scale models and to analyse particular ~~constellations of~~ pollution events such as~~(e.g.,~~ those associated with heat waves or large-scale forest fires; Ordóñez et al., (2010), Huijnen et al., (2012, 2016)~~)~~. Finally, they can be used as a reference to identify to what extent particular periods and regions deviate from climatology, as provided by the reanalysis, as for instance also discussed in the series of the State of the Climate (Flemming and Inness, 2018).

However, all of these applications presume that the reanalysis is sufficiently accurate, or, to the least, well described.~~,~~ ~~which, despite~~ Despite ~~many years of research and~~ the range of observations assimilated into the respective systems, this is not necessarily ensured. Issues are multiple, and depend on the availability of observations, and on the modelling and data-assimilation framework with respect to the species and location under consideration. For tropospheric ozone reanalyses, state-of-the-art global analysis systems have been used to assimilate satellite-based observations, where satellite measurements have limited information on vertical profiles. In particular, the small measurement sensitivities to the lower troposphere makes it difficult to correct near-surface ozone. Advanced satellite retrievals provide improved vertical resolution to the troposphere (Cuesta et al., 2018; Fu et al., 2018), but the temporal coverage and vertical resolution of these retrievals is still limited, and their application in data assimilation remains a challenge (Miyazaki et al., 2019a). This also implies that constraints on other parts of the system (other trace gases, aerosol, their emissions, as well as meteorology, driving the tracer transport and its removal) will strongly affect the quality of the reanalysis. Simultaneous optimization of concentrations and precursors' emissions seems thus important in improving the analysis of lower tropospheric ozone (Miyazaki et al., 2012b). Furthermore, providing consistent time series over a decadal time-scales is challenging. The observational data from satellite instruments available for assimilation evolve over time with new instruments becoming available while others cease to exist, and different satellite retrieval products typically showing biases with respect to ground-based observations as well as with respect to each other.

In the framework of the Copernicus Atmosphere Monitoring Service (CAMS, https://atmosphere.copernicus.eu), ECMWF's Integrated Forecasting System (IFS) has been extended to include modules for atmospheric chemistry, aerosols and greenhouse gases. Using this system, three recent reanalyses have been released: the Monitoring Atmospheric Composition and Climate (MACC) reanalysis for the years 2003-2012 (Inness et al., 2015), the 'CAMS Interim Reanalysis' (hereafter 'CAMS-iRean') for the years 2003-2018 (Flemming et al., 2017) and recently the 'CAMS Reanalysis' ('CAMS-Rean') for the years 2003-present (Inness et al., 2019). Miyazaki et al. (2015) simultaneously estimated concentrations and emissions for an 8-year tropospheric chemistry reanalysis (TCR-1) for the years 2005–2012 obtained from an assimilation of multi-constituent satellite measurements using an ensemble Kalman filter (EnKF). TCR-1 has been used to provide comprehensive information on atmospheric composition variability and elucidate variations in precursor emissions, and to evaluate bottom-up emission inventories (Miyazaki et al., 2012, 2014, 2015, 2017; Ding et al., 2017; Jiang et al., 2019; Tang et al., 2019). A second version of the EnKF-based reanalysis (TCR-2) has been recently produced using an updated model and satellite retrievals for the years 2005–2018 (Kanaya et al., 2019; Miyazaki et al., 2019a; Thompson et al., 2019). For stratospheric constituents, several studies have been conducted to produce and compare stratospheric chemical reanalysis products (Davis et al., 2017; Errera et al., 2019).

Here we evaluate the ability of the two CAMS and two TCR atmospheric composition reanalysis data sets to constrain tropospheric ozone variability. We do not evaluate the MACC reanalysis here, because it has been extensively documented in the past (Inness et al. 2013; Flemming et al., 2017; Bennouna et al., 2019) and only covered the 2003-2012 time period. Furthermore it has been shown to suffer from significant spurious drifts in tropospheric ozone due to a bias-correction issue, which makes it less useful to assess its ~~multi-annual~~multiannual mean and inter-annual variability. In particular, Katragkou et al. (2015) discusses the ozone in the MACC reanalysis; while Inness et al., (2019) reports how CAMS ~~-REAN-~~Rean compares to CAMS ~~-iREAN-~~iRean and MACC reanalysis.

To assess the quality of these reanalysis products, with attention for the various potential ~~and in particular to evaluate their fitness for purpose for the various~~ types of application described above, this study evaluates tropospheric ozone for a range of independent in-situ observations: ozone sondes from ~~various~~ networks World Ozone and Ultraviolet Radiation Data Centre (WOUDC), NOAA Earth System Research Laboratory (ESRL), and Southern Hemisphere Additional Ozonesondes (SHADOZ), monthly mean gridded surface ozone as collected within TOAR, and individual surface ozone observations from the EMEP network.

In this study, we limit ourselves to tropospheric ozone in the reanalysis products, and only refer, where relevant, to interactions with other components in the reanalysis systems, such as nitrogen oxides (NO_x) and carbon monoxide (CO), and aerosols. Even though these four reanalysis products are not equally independent, each of their configurations show substantial differences which are bound to impact the performance of the reanalysis products. This intercomparison aims to reveal to what extend the reanalysis products agree, depending on region and time periods. Temporal consistency is an important aspect when assessing long-term time series and intercomparing individual years. At the same time this is a challenge because of the change in the observing system ~~ing constellation of satellite observations~~ used to constrain the reanalysis products over the course of a decade or more, all having different retrieval specifications (see also Gaudel et al., 2018).

Met opmaak: Subscript

Met opmaak: Niet Superscript/ Subscript

In the next sections we describe the various reanalysis products used in this paper (Sect. 2) and the observational data used for evaluation (Sect. 3). Evaluations against ozone sondes are presented in Sect. 4, and against TOAR gridded surface ozone and EMEP surface observations in Sect. 5, and Sect. 6, respectively. We continue describing the reanalysis products through assessment of their global spatial and temporal consistency ~~of their tropospheric column time series (Sect 7) and average concentrations~~ (Sect ~~8~~7)~~, to assess the spread~~. We end with discussions and conclusions in Sect ~~9~~8.

2. Chemical reanalysis products

The global atmospheric chemistry reanalysis products evaluated in this paper are listed in Table 1. The general configuration of the various data assimilation systems, together with details specific to tropospheric ozone analysis, are provided in the following subsections. For more detailed information on the specifications of the various reanalysis products the reader is referred to the references.

Table 1. Overview of recent reanalysis products

Name (reference)	Time period	Altitude range and horizontal resolution	Forecast model	Data assimilation scheme	Assimilated components
CAMS ~~-iREAN-~~ iRean (Flemming et al., 2017)	2003-2018	Up to 0.1hPa T159/L60	IFS(CB05) CY40R2	4D-VAR	CO, O_3, AOD
CAMS ~~-REAN-~~ Rean (Inness et al., 2019)	2003-present	Up to 0.1hPa T255/L60	IFS(CB05) CY42R1	4D-VAR	CO, O_3, NO_2, AOD

TCR-1 (Miyazaki et al., 2015; Miyazaki et al., 2017; Miyazaki and Bowman, 2017)	2005-2014	Up to 4.4 hPa T42/L32	MIROC-Chem Nudged to ERA-Interim	EnKF	CO, O_3, NO_2, HNO_3
TCR-2 (Miyazaki et al., 2019a, Kanaya et al., 2019)	2005-2018	Up to 4.4 hPa T106/L32	MIROC-Chem Nudged to ERA-Interim	EnKF	CO, O_3, NO_2, HNO_3, SO_2

2.1 The CAMS Interim reanalysis

In CAMS, the data assimilation capabilities in IFS for trace gases and aerosols relies on the four-dimensional variational (4D-VAR) technique, developed for the analysis of meteorological fields. The CAMS interim Reanalysis (CAMS~~-iREAN-~~iRean, Flemming et al., 2017) has been the intermediate reanalysis between the widely used MACC Reanalysis (Inness et al., 2015) and the recently produced CAMS reanalysis (Inness et al., 2019). The chemistry module as adopted in CAMS~~-iREAN~~ iRean is described and evaluated in Flemming et al. (2015). It relies on the modified CB05 tropospheric chemistry mechanism as originating from TM5 (Huijnen et al., 2010; Williams et al., 2013) which contains 52 species and 130 (gas-phase + photolytic) reactions; stratospheric ozone is modelled through the Cariolle parameterization (Cariolle and Teyssèdre, 2007). Anthropogenic emissions originate essentially from the MACCity inventory (Granier et al., 2011) with enhanced wintertime CO emissions over Europe and US (Stein et al., 2014). Monthly specific biogenic emissions originate from MEGAN-MACC (Sindelarova et al., 2014), but using monthly climatological values from 2011 onwards. Daily biomass burning emissions originate from GFASv1.2 (Kaiser et al., 2013). The meteorological model is ~~adopts~~ IFS CY40R2.

In terms of ozone, observations from the following set of satellite instruments have been assimilated: Solar Backscatter ULTa-Violet (SBUV~~-~~/2), OMI, MLS, GOME-2, SCanning Imaging Absorption spectroMeter for Atmospheric CHartographY (SCIAMACHY), GOME and Michelson Interferometer for Passive Atmospheric Sounding (MIPAS), see also Table 2. A variational bias correction (VarBC) scheme was applied to OMI, SCIAMACHY and GOME-2 retrievals of total ozone columns to ensure optimal consistency of all information used in the analysis. SBUV/2 and also profile retrievals from MLS and MIPAS were assimilated without correction. Note that no total columns are assimilated for solar elevations less than 6°, hence excluding polar winters.

Profile observations from limb instruments ~~(MIPAS and MLS)~~ in the range of 0.1-150 hPa for MIPAS and 0.1-147 hPa for MLS ~~Limb observations~~ are ~~instrumental~~ used to constrain the stratospheric contribution of the total column. In combination with the assimilated total column retrievals this implies that also the tropospheric part is constrained (Inness et al., 2013). ~~discriminate between the tropospheric and stratospheric contribution of the total column observations.~~ CAMS~~-iREAN-~~iRean uses observations from the MIPAS instrument for the period February 2005 - March 2012. MLS data on Aura have been used from August 2004 onwards, based on version 2 observations during August 2004-Dec 2012, and V3.4 from January 2013 onwards. V3.4 has a different specification of the vertical levels and observation errors compared to V2 (Schwartz et al., 2015 and https://mls.jpl.nasa.gov/data/v3_data_quality_document.pdf). Finally, note that in CAMS~~-iREAN-~~iRean no observations of NO_2 have been assimilated. CO has been constrained through assimilation of Measurement of Pollution in the Troposphere (MOPITT) total columns.

Table 2: Observations of ozone used in the CAMS ~~iREAN~~ iRean assimilation system

Instrument (satellite)	**Product**	**Data provider/version**	**Period**	**Reference**
SCIAMACHY (Envisat)	TC	ESA, / CCI (TC_SCIAMACHY/fv0300)(BIRA)	2003-01-01 to 2012-04-08	Lerot et al. (2009)
MIPAS (Envisat)	Prof	ESA , / NRT ESA, CCI(KIT) (HARMOZ_MIPAS)/fv0004	2005-01-27 to 2012-03-31	Von Clarmann et al. (2003, 2009)
MLS (Aura)	Prof	NASA /V2 NASA/V3.4	2004-08-03 to 2012-12-31 2013-01-01 to 2016-12-31	Schwartz et al. (2015)
OMI(Aura)	TC	KNMI /V3 KNMI /NRT	2004-08-03 to 2015-05-31 2015-06-01 to present	Liu et al. (2010)
GOME (ERS-2)	Prof	RAL	2003-01-01 to 2003-05-31	Munro et al. (1998)
GOME-2 (Metop-A)	TC	ESA, (CCI),BIRA /fv0100 ESA, (CCI),BIRA /fv0300 NRT	2007-01-23 to 2012-12-31 2013-01-01 to 2016-12-31 2017-01-01 to present	Hao et al. (2014)
GOME-2 (Metop-B)	TC	ESA, (CCI),BIRA /fv0300 NRT	2013-01-01 to 2016-12-31 2017-01-01 to present	Hao et al. (2014)
SBUV/2 (NOAA-14 – NOAA-19)	PC	NASA / v8.6 13L NRT 21L	2003-01-01 to 2012-12-31 2013-01-01 to present	Bhartia et al. (1996), McPeters et al. (2013)

Tabel met opmaak

Table 3: Observations of ozone used in the CAMS~~-REAN~~-Rean assimilation system

Instrument (satellite)	Product	Data provider/version	Period	Reference
SCIAMACHY (Envisat)	TC	ESA, CCI ~~(BIRA)~~ (TC_SCIAMACHY/fv0300)	2003-01-01 to 2012-04-08	Lerot et al. (2009)
MIPAS (Envisat)	Prof	ESA, NRT ESA, CCI~~(KIT)~~ (HARMOZ_MIPAS)/fv0004	2003-01-27 to 2004-03-26 and 2005-01-27 to 2012-03-31	Von Clarmann et al. (2003, 2009)
MLS (Aura)	Prof	NASA /V4	2004-08-03 to 2016-12-31	Schwartz et al. (2015)
OMI(Aura)	TC	KNMI /V3 KNMI /NRT	2004-08-03 to 2015-05-31 2015-06-01 to present	Liu et al. (2010)
GOME-2 (Metop-A)	TC	ESA, ~~(~~CCI~~),BIRA~~ /fv0100 ESA, ~~(~~CCI~~),BIRA~~ /fv0300 NRT	2007-01-23 to 2012-12-31 2013-01-01 to 2016-12-31 2017-01-01 to present	Hao et al. (2014)
GOME-2 (Metop-B)	TC	ESA, ~~(~~CCI~~),BIRA~~ /fv0300 NRT	2013-01-01 to 2016-12-31 2017-01-01 to present	Hao et al. (2014)
SBUV/2 (NOAA-14 – NOAA-19)	PC	NASA / v8.6 13L NRT 21L	2003-01-01 to 2013-07-07 2013-07-08 to present	Bhartia et al. (1996), McPeters et al. (2013)

Tabel met opmaak

2.2 The CAMS Reanalysis

The CAMS Reanalysis (CAMS~~-REAN~~-Rean; Inness et al., 2019) is the successor of the CAMS~~-iREAN~~-iRean. Compared to CAMS~~-iREAN~~-iRean, the horizontal resolution has increased to ~80 km (T255), while meteorology is now based on CY42R1. Emissions are largely similar to CAMS~~-iREAN~~-iRean, except that the monthly varying biogenic emissions have been used for the full time period. With respect to the CB05-based chemistry module, heterogeneous chemistry on clouds and aerosol has been switched on, as well as the modification of photolysis rates due to aerosol scattering and absorption (Huijnen et al., 2014).

As for assimilated ozone observations, data from a very similar set of instruments have been used as for CAMS~~-iREAN~~-iRean: SCIAMACHY, MIPAS, OMI, MLS, GOME-2, and SBUV/2, see Table 3. However, note that the CAMS-Interim Reanalysis additionally assimilated GOME profile observations during the first 5 months of 2003, which have not been assimilated in CAMS~~-REAN~~-Rean as it was found to lead to a degradation in the O_3 analysis. Different to CAMS~~-iREAN~~-iRean, CAMS~~-REAN~~-Rean also assimilated observations from the MIPAS instrument during 2003 and early 2004, although using a different version. Also frequently newer versions of the data have been adopted in CAMS~~-REAN~~-Rean compared to CAMS~~-iREAN~~-iRean, particularly for MLS observations the reprocessed version 4 has been applied throughout the full time period.

In CAMS~~-REAN~~-Rean also tropospheric NO_2 columns are assimilated, using observations from the SCIAMACHY (2003-2012), OMI (from October 2004 onwards) and GOME-2 (from April 2007 onwards) instruments. The same settings for the ~~A~~ variational bias correction were used in CAMS-Rean as in CAMS-iRean. ~~(VarBC) scheme was applied to OMI, SCIAMACHY and GOME-2 retrievals of total ozone columns to ensure optimal consistency of all information used in the analysis. SBUV/2 and also profile retrievals from MLS and MIPAS were assimilated without correction. Inness et al. (2019) provide an extended overview of the biases of various assimilated observations against the reanalysis. For ozone assimilation in particular, the following findings are most noteworthy for this study (see also Appendix C of Inness et al., 2019):~~

- Larger biases for SCIAMACHY observations in 2003 and early 2004, associated to issues with the early SCIAMACHY O_3 retrievals in this time period
- Larger departures for MIPAS data during 2003-2004 than after 2005, where CCI data was used
- Different behaviour of OMI data between 2009 and 2012, associated to a deterioration in the OMI row anomalies (Schenkeveld et al., 2017) which could not be filtered out in the CAMS assimilation procedure.
- An increasing bias correction for GOME-2A especially after January 2013, associated to a version change of the SBUV/2 data.

CAMS iREAN iRean and CAMS REAN Rean surface and tropospheric ozone are archived with a three-hourly output frequency.

2.3 Tropospheric Chemistry Reanalysis (TCR-1)

The TCR-1 data assimilation system is constructed using an EnKF approach. A revised version of the TCR-1 data is used in this study. A major update from the original TCR-1 system (Miyazaki et al., 2015) to the system used here and in (Miyazaki et al., 2017 (2017); Miyazaki and Bowman, 2017) is the replacement of the forecast model from CHASER (Sudo et al., 2002) to MIROC-Chem (Watanabe et al., 2011), which caused substantial changes in the a priori field and thus the data assimilation results of various species.

MIROC-Chem (Watanabe et al., 2011), considers detailed photochemistry in the troposphere and stratosphere by simulating tracer transport, wet and dry deposition, and emissions, and calculates the concentrations of 92 chemical species and 262 chemical reactions. The MIROC-Chem model used in TCR-1 has a T42 horizontal resolution (~2.8°) with 32 vertical levels from the surface to 4.4 hPa. It is coupled to the atmospheric general circulation model MIROC-AGCM version 4 (Watanabe et al., 2011). The simulated meteorological fields were nudged toward the 6-hourly ERA-Interim (Dee et al., 2011) to reproduce past meteorological fields.

The a priori anthropogenic NO_x and CO emissions were obtained from the Emission Database for Global Atmospheric Research (EDGAR) version 4.2 (EC-JRC, 2011). Emissions from biomass burning were based on the monthly Global Fire Emissions Database (GFED) version 3.1 (van der Werf et al., 2010). Emissions from soils were based on monthly mean Global Emissions Inventory Activity (GEIA; Graedel et al., 1993).

The data assimilation used is based upon on an EnKF approach (Hunt et al., 2007) that uses an ensemble forecast to estimate the background error covariance matrix and generates an analysis ensemble mean and covariance that satisfy the Kalman filter equations for linear models. The concentrations and emission fields of various species are simultaneously optimized using the EnKF data assimilation, see also Table 4.

For data assimilation of tropospheric NO_2 column retrievals, the version 2 Dutch OMI NO_2 (DOMINO) data product (Boersma et al., 2011) and version 2.3 TM4NO2A data products for SCIAMACHY and GOME-2 (Boersma et al., 2004) were used, obtained through the TEMIS website (http://www.temis.nl). The TES ozone data and observation operators used are version 5 level 2 nadir data obtained from the global survey mode (Bowman et al., 2006; Herman and Kulawik, 2013). TES ozone data was excluded poleward of 72 degree because of the small retrieval sensitivity, limiting data assimilation adjustments at high latitudes in the troposphere. Also Note note that the availability of TES measurements is strongly reduced after 2010, which led to a degradation of the reanalysis performance, as demonstrated by Miyazaki et al. (2015). The MLS data used are the version 4.2 ozone and HNO_3 level 2 products (Livesey et al., 2011 2018). Data for pressures of less than 215 hPa for ozone and 150 hPa for HNO_3 were used. The MOPITT CO data used are version 6 level 2 thermal-infrared retrieval (TIR) products (Deeter et al., 2013). A superobservation approach was employed to produce representative data with a horizontal resolution of the forecast model NO_2 and CO observations, following the approach of Miyazaki et al. (2012). No bias correction was applied to the assimilated measurements.

2.4 Updated Tropospheric Chemistry Reanalysis (TCR-2)

An updated Chemistry Transport Model (CTM) and satellite retrievals are used in TCR-2 (Kanaya et al., 2019; Miyazaki et al., 2019a, 2019b; Thompson et al., 2019). A high-resolution version of the MIROC-Chem model with a horizontal resolution of T106 (1.1° x 1.1°) was used. Sekiya et al. (2018) demonstrated the improved model performance on tropospheric ozone and its precursors by increasing the model resolution from 2.8° x 2.8° to 1.1° x 1.1°. A priori anthropogenic emissions of NO_x and CO were obtained from the HTAP version 2 inventory for 2008 and 2010 (Janssens-Maenhout et al., 2015). Emissions from biomass burning are based on the monthly GFED version 4.2 inventory (Randerson et al., 2018) for NO_x and CO, while those from soils are based on the monthly GEIA inventory (Graedel et al., 1993) for NO_x. Emission data for other compounds are taken from the HTAP version 2 and GFED version 4 inventories.

Met opmaak: Subscript

Met opmaak: Subscript

The satellite products used in TCR-2 are more recent than those used in TCR-1, see Table 4. Tropospheric NO_2 column retrievals used are the QA4ECV version 1.1 L2 product for OMI (Boersma et al., 2017a) and GOME-2 (Boersma et al., 2017b). Version 6 of the TES ozone profile data was used. The MLS data used are the version 4.2 ozone and HNO_3 L2 products (Livesey et al., ~~2011~~2018). The MOPITT total column CO data used were the version 7L2 TIR/NIR product (Deeter et al., 2017). OMI SO_2 data of the planetary boundary layer vertical column L2 product were used as produced with the principal component analysis algorithm (Krotkov et al., 2016; Li et al., 2013). As in TCR-1, a super-observation approach to produce representative data with a horizontal resolution of the forecast model (1.1° × 1.1°) for NO_2 and CO observations was applied. As in TCR-1, no bias correction was applied to the assimilated measurements.

TCR-2 data was used to study the processes controlling air quality in East Asia during the KORUS-AQ aircraft campaign (Miyazaki et al., 2019a). Kanaya et al. (2019) demonstrated the TCR-2 ozone and CO performance using research vessel observations over open oceans. Thompson et al. (2019) used the TCR-2 data to help understanding of near surface NO_2 pollutions observed during the KORUS-OC campaign. Both for TCR-1 and TCR-2 the reanalysis data is archived on a two-hourly output frequency.

Table 4: Observations used for ozone assimilation in TCR-1, and in square brackets changes for TCR-2.

Instrument (satellite)	**Species**	**Product**	**Data provider/version**	**Period**	**Reference**
OMI (Aura)	NO_2 [+ SO_2]	TrC for NO_2 [PBL for SO_2]	DOMINO v2 [QA4ECV v1.1] for NO_2 [PCA v3 for SO_2]	2005-01-01 to present	Boersma et al. (2011) [Boersma et al. (2017a)] [Krotkov et al. (2016); Li et al. (2013)]
SCIAMACHY (Envisat)	NO_2	TrC	DOMINO v2 [QA4ECV]	2005-01-01 to 2012-03-29	Boersma et al. (2004) [(Boersma et al. (2017b)]
GOME-2 (Metop-A)	NO_2	TrC	DOMINO v2 [QA4ECV]	2007-01-01 to present	Boersma et al. (2004) [(Boersma et al. (2017b)]
TES (Aura)	O_3	Profile	v5 [v6]	2005-01-01 to 2011-06-04	Bowman et al. (2006); Herman and Kulawik (2013)
MLS (Aura)	O_3, HNO_3	Profile	~~V3.3~~ v4.2 [v4.2]	2005-01-01 to present	Livesey et al. (201~~1~~8)
MOPITT (Terra)	CO	Profile	v6 NIR [v7 TIR/NIR]	2005-01-01 to present	Deeter et al. (2013) [Deeter et al. (2017)]

Tabel met opmaak

2.5 Temporal consistency of the observing systems

Discontinuities in the observing systems, as specified in Tables 2-4, can cause significant temporal changes in the reanalyses quality. In 4D-Var systems this can be assessed through statistics based on analysis departures (observation minus analysis) and first guess departures (observations minus model first guess). Inness et al. (2019) provide an extended overview of the biases of various assimilated observations against the CAMS ~~r~~Reanalysis. For ozone assimilation in particular, the following findings are most noteworthy for this study (see also Appendix C of Inness et al., 2019):

- Larger biases for SCIAMACHY observations in 2003 and early 2004, associated to issues with the early SCIAMACHY O_3 retrievals in this time period
- Larger departures for MIPAS data during 2003-2004 than after 2005, where CCI data was used
- Different behaviour of OMI data between 2009 and 2012, associated to a deterioration in the OMI row anomalies (Schenkeveld et al., 2017) which ~~could~~unfortunately have not be filtered out in the CAMS assimilation procedure.
- An increasing bias correction for GOME-2A especially after January 2013, associated to a version change of the SBUV/2 data.

With respect to both TCR reanalyses which are based on the EnKF approach, important information regarding the reanalysis product is provided by the error covariance. The analysis ensemble spread is estimated as the standard deviation of the simulated concentrations across the ensemble and can be used as a measure of the uncertainty of the reanalysis product (Miyazaki et al., 2012). The uncertainty information on the analysis uncertainty is included in the TCR-1 and TCR-2 reanalysis products and this can be used to investigate the long-term stability of the data assimilation performance. In addition, the χ^2 test was used to evaluate the temporal changes in data assimilation balance (e.g. Ménard and Chang, 2000). Miyazaki et al (2015) demonstrated increased χ^2 for OMI NO_2 after 2010, associated with a decrease in the number of the assimilated measurements and changes in the super-observation error due to the OMI row anomalies. Furthermore, the decreased number of assimilated TES ozone retrievals after 2010 affected the long-term reanalysis characteristics. Before 2011 the analysis spread for ozone in the middle troposphere is about 1-3 ppb in the tropics and subtropics and 3-12 ppbv in the extratropics. The larger spread at lower latitudes could be attributed to the higher sensitivities in the TES ozone retrievals. From 2011 onwards the spread mostly becomes smaller than 3 ppb for the globe, which seems excessively small and is likely associated with the lack of effective observations for measuring the analysis uncertainties and with the stiff tropospheric chemical system. The obtained results indicate the requirements for additional observational information and/or stronger covariance inflation to the forecast error covariance for measuring the long-term analysis spread corresponding to actual analysis uncertainty.

3. Ozone observations used for evaluation

3.1 Ozone sondes

For evaluation of free tropospheric ozone data from the global network of ozone sondes, as collected by the WOUDC, is used, expanded with observations available from SHADOZ (Thompson et al., 2017; Witte et al., 2017) and ESRL. The observation error of the sondes is about 7–17% below 200 hPa and ±5% in the range between 200 and 10 hPa (Beekmann et al., 1994, Komhyr et al., 1995 and Steinbrecht et al., 1996). Typically, the sondes are launched once a week, but in certain periods, such as during ozone hole conditions, launches can be more frequent. Sonde launches are mostly carried out between 9:00 and 12:00 local time.

The ozone sonde network provides critical independent validation of the reanalysis products. Although the number of soundings varied for the different stations, the global distribution of the launch sites is expected to be sufficient to allow meaningful monthly to seasonal averages over larger areas. However, because of the sparseness of the ozone sonde network, we are aware that the evaluation based on ozone sonde observations can introduce large biases in regional and seasonal reanalysis performance (Miyazaki and Bowman, 2017).

The ~~model~~ reanalysis data ha~~s~~ve been collocated with observations through interpolation in time and space. Individual intercomparisons have been aggregated on a monthly and seasonal basis. The number of stations contributing to the monthly and regional means varies over the course of the reanalysis products, and is additionally reported as this is naturally an important consideration when assessing interannual variability of ozone biases. While we present time series from 2003 onwards in our figures, where CAMS starts to provide reanalysis products, for any of the statistics we only base this on the 2005-2016 time period (unless explicitly mentioned), to allow fair intercomparison between CAMS and TCR.

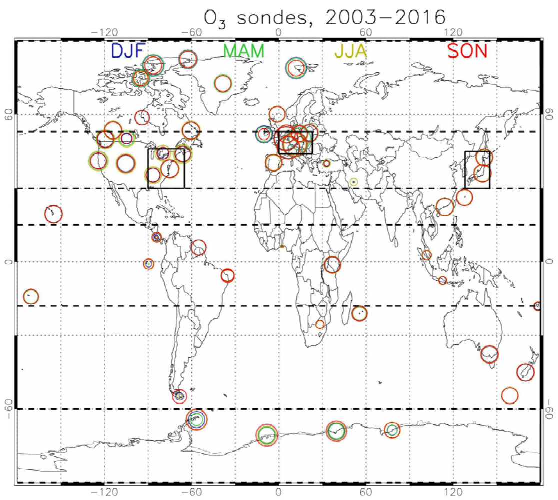


Figure 1: Location of ozone sondes contributing to the various regions~~, and indicated by the black boxes and the dashed lines~~. The size of circles indicates the relative number of observations contributing to the statistics for December–February (blue), March–May (green), June–August (yellow), and September–November (red). The dashed lines indicate the latitude bands representing the NH arctic region (53°N-90°N), NH mid latitudes (30°N -53°N), NH subtropics (15°N-30°N), Tropics (18°S-15°N), SH mid latitudes (60°S-18°S) and Antarctic (90°S-60°S). The black boxes indicate three additional regions in the NH-mid latitudes: Eastern US (90°W-65°W, 30°N-46°N), Western Europe (0°E-23°E), Japan (30°N-45°N, 128°E-145°E).

For spatial aggregation the choice is more difficult, depending on the characteristics of the species and availability of observations. Tilmes et al. (2012) defined an aggregation approach for ozonesonde locations based on the characteristics of the observed ozone profiles. We follow in part their aggregation approach, by adopting the European, Eastern US, Japan, and Antarctic regions. For several regions, the number of measurements could be insufficient to construct meaningful aggregates. Instead we define regions for the northern hemisphere (NH) subtropics ~~(15°N to 32°N)~~, the tropics ~~(18°S to 15°N)~~, southern hemisphere (SH) mid latitudes ~~(60°S to 18°S)~~ and Antarctic ~~(90°S to 60°S)~~, and combine the NH Polar regions to a single region ~~(60°N to 90°N)~~, see also Figure 1.

3.2 Surface ozone

We evaluate surface ozone against the TOAR database (Schultz et al., 2017), which provides a globally consistent, gridded, long-term dataset with ozone observation statistics on a monthly mean basis. The TOAR database has been produced with particular attention to quality control, and representativeness of the in-situ observations, in order to establish consistent, long-term time records of observations. TOAR provides a disaggregation of rural and urban stations. For our study we use the 2°×2° gridded monthly mean dataset representative for rural stations for the 1990-2014 time period. This allows easy intercomparison with monthly mean results from the various reanalysis products.

Note that in these comparisons we used rural observations only, because none of the reanalysis model resolutions is considered sufficient ~~any of the reanalysis model resolutions is considered too coarse~~ to resolve local concentration changes over highly polluted urban areas. Therefore the rural observations can be considered as more representative data for grid averaged concentrations. Nevertheless, neglecting urban observations could lead to biased evaluations particularly in cases where large fractions of the grid cells are associated to urban conditions, e.g. in megacities.

This TOAR dataset has a good global coverage, including stations over East Asia, and provides overall a constant, and good quality controlled data record up to 2014. Nevertheless, the number of records in this database decreases significantly for various regions on the globe after 2012. Therefore in our evaluation statistics we focus on the period before 2012, considering that the reduction in available observations afterwards hampers the intercomparison of ~~model~~ reanalysis performance between different years. Similar to the evaluation against ozone sonde observations, the statistics is computed for data from 2005 onwards.

3.3 EMEP observations

In order to assess the ability of the reanalysis products to represent spatial and temporal variability on a sub-seasonal and on regional scales, we additionally evaluate the reanalyses against ground-based hourly observations from the EMEP network (obtained from http://ebas.nilu.no/) for the year 2006. Although EMEP data are also included in the TOAR data product, this analysis allows for a complementary approach, in particular the assessment of pollution events during heat waves, but also evaluation of the diurnal cycles and spatial variability in the various products. The summer period of 2006 over Europe was characterized by a heat wave event (Struzewska and Kaminski, 2008). For this evaluation, we collocate the ~~model~~ reanalysis output spatially and temporally to the observations, using a reference 3-hourly time frequency. Considering the comparatively coarse horizontal resolution, which is not generally able to represent the local orography at the location of the individual observations, we match the model level with the same (average) pressure level at the location of the observations. Here we note that the CAMS reanalyses use a higher vertical resolution than TCR. This implies that for high-altitude stations also different (higher) model levels are sampled in the CAMS reanalyses compared with TCR. After this collocation procedure, we compute temporal correlation coefficients on a seasonal basis, using the temporally collocated 3-hourly ~~mode~~ reanalysis and observational data.

4. Evaluation against ozonesondes

4.1 Annually and regionally averaged profiles

Figure 2 provides an overview of the multi-annual mean ozone for the four reanalyses for the 2005–2016 time period. All reanalyses capture the observed vertical profiles of ozone from the lower troposphere to the lower stratosphere, with a regional mean bias of typically less than 8 ppb throughout the troposphere. Corresponding mean biases at 850, 650 hPa and 350 hPa are given in ~~Tables 5-7~~Figure 3, where the bias is defined as the reanalysis-observation, throughout this work. The normalized values, as scaled with the mean of the observations, are given in Figure S1 in the Supplementary Material. These ~~multi-annual~~multiannual, regional mean biases are below 3.7 ppb ~~(~7%)~~ at 850 hPa and 4 ppb ~~(~7%)~~ at 650 hPa, while normalized (absolute) biases are mostly ~~below~~ below 10%. ~~Generally~~For most regions, the CAMS reanalysis shows improvement against the CAMS interim reanalysis at 650 hPa and also 850 hPa, particularly for regions over the NH high- and mid-latitudes, as well as the SH-mid latitudes, but at the cost of a degradation (an emerging positive bias) towards the surface. TCR-2 shows a more mixed picture in this respect. Biases between TCR and CAMS are within a similar order of magnitude, but are not correlated in any way in sign or magnitude. For most of the major polluted areas in the lower troposphere, the biases are lower in the CAMS reanalysis than in the TCR reanalyses, probably due to its higher reanalysis model resolution and a better chemical forecast model performance. The annual mean ozone biases in TCR are relatively large in the tropics and SH high latitudes. After 2011, no TES tropospheric ozone measurements were assimilated, which could lead to enhanced ozone biases, as demonstrated by Miyazaki et al. (2015). Assimilation of MLS measurements does not noticeably influence the tropospheric ozone analysis in the tropics. In the NH subtropics and the tropics regions the reanalyses show some larger deviation against sonde observations at lower altitudes, which was traced to comparatively large biases at the Hong Kong and Kuala Lumpur stations. Note~~, in the NH subtropics and tropics~~ that in these regions~~,~~ the ozonesonde network is sparse, while the spatial and temporal variability of ozone is large, which ~~could~~ limits our understanding of the generalized reanalysis performance (Miyazaki and Bowman, 2017). At high latitudes, the large diversity in the reanalysis ozone could be associated with the lack of direct tropospheric ozone measurements in all of the systems. ~~In the TCR systems, TES ozone data was excluded poleward of 72 degree because of the small retrieval sensitivity, limiting data assimilation adjustments at high latitudes in the troposphere.~~

Overall, this evaluation shows that the biases from these reanalysis products are ~~much~~ smaller than those reported from recent CTM simulations. E.g. Young et al. (2013) present median biases across ACCMIP model versions at 700 (500) hPa up to 10 (15)%, depending on the region. This demonstrates that the reanalysis of tropospheric ozone fields is generally well constrained by assimilated measurements for the globe.

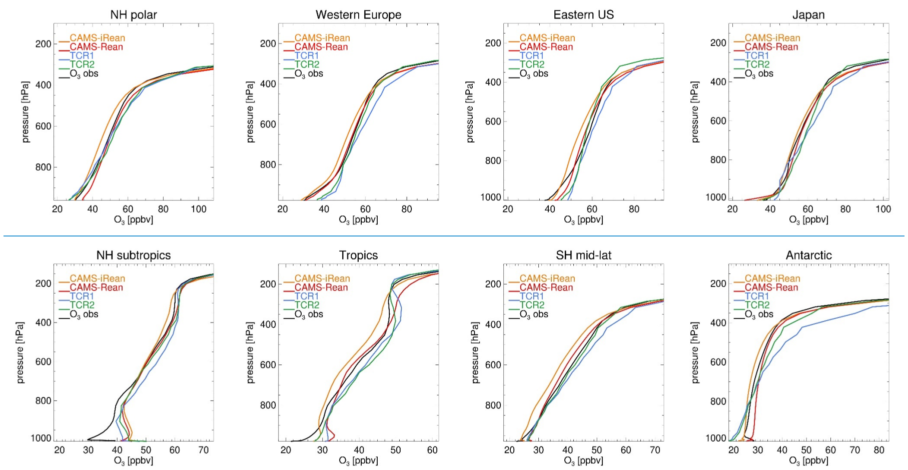
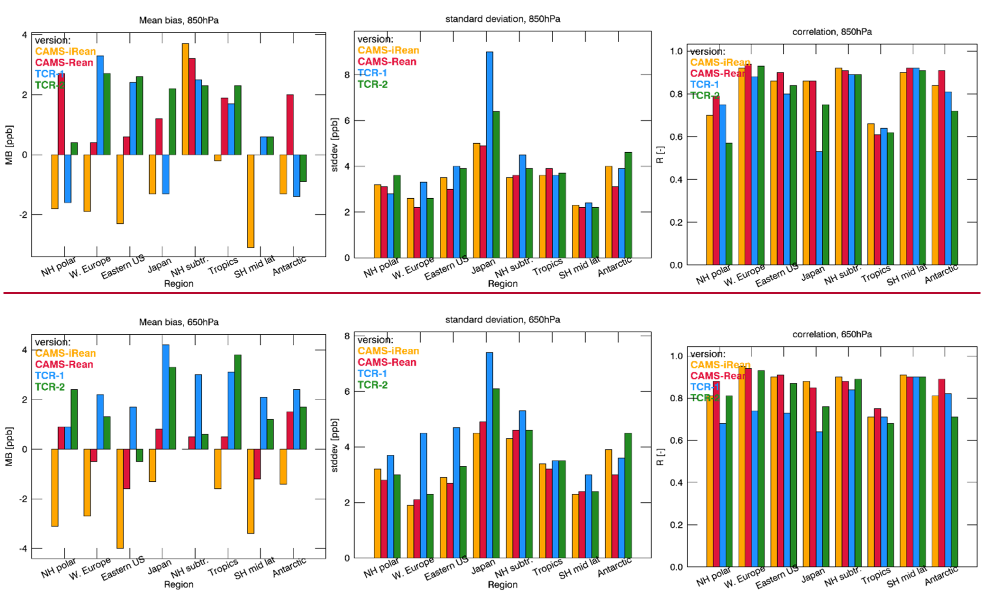


Figure 2: Evaluation of regional mean ~~multi-annual~~multiannual mean O_3 profiles against ozone sondes, averaged over the 2005–2016 time period.



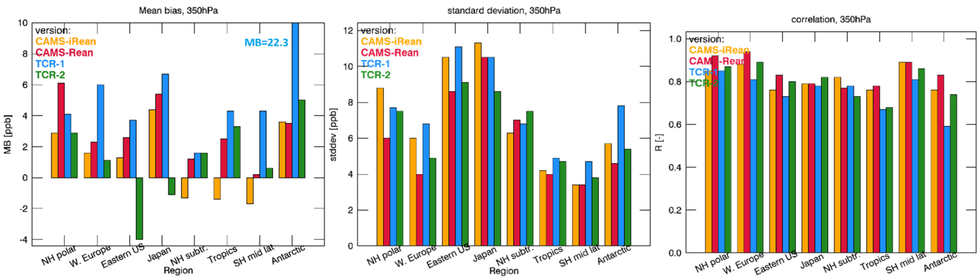


Figure 3: Evaluation of ozone mean bias (reanalysis-observation, left), standard deviation of the unbiased differences (middle) and temporal correlation R (right) for the four reanalysis products at 850 (top), 650 (middle) and 350 (bottom) hPa against sondes, computed for various regions, for the 2005-2016 time period.

Met opmaak: Standaard

~~**Table 5. Evaluation of ozone mean bias, RMSE (both in ppb) and temporal correlation R for the four reanalysis products at 850 hPa against sondes, computed for various regions, for the 2005-2016 time series.**~~

Met opmaak: Standaard

~~**Table 6. Same as Table 5 but for 650 hPa.**~~

~~**Table 7. Same as Table 5 but for 350 hPa.**~~

~~**Table 8. Same as Table 5 but for tropospheric columns in units DU.**~~

4.2 Time series of zonally averaged O_3 tropospheric columns

Collocated ~~tropospheric~~ partial columns from the surface up to 300 hPa, hereafter for brevity referred to as 'tropospheric columns', have been compared to ~~tropospheric~~ partial columns ~~(surface up to 300 hPa, hereafter for brevity referred to as 'tropospheric columns')~~ derived from the sonde observations. An intercomparison of the monthly and zonally mean tropospheric columns sampled at the observations is given in Figure ~~3~~4. The corresponding performance statistics is given in Figure 5. Here, the standard deviation (stddev) is computed based on the unbiased differences between the reanalyses and sonde observations, and provides a metric of the quality of the monthly mean variability in the reanalyses.

Normalized statistics are provided in Figure S2 in the Supplementary Material. Note that the figures also contain information on the number of sonde stations that are included in the evaluation for individual months. ~~Here the tropopause has been defined as the altitude where ozone exceeds 150 ppb for each of the individual products.~~

Outside the polar regions all reanalyses capture the magnitude of the zonal mean tropospheric column to within a MB of within 1.8 DU, and the stddev between 0.8 and 1.3 DU ~~root-mean-square (RMS) of 0.8–2.1 DU~~ depending on the reanalysis product~~, see also Table 8~~. For most regions and performance metrics, the updated reanalyses outperform their predecessor versions. For instance, for the NH mid latitudes the MB is -0.3 DU (1.2%, when normalized with sonde observations) for CAMS-Rean and 0.8 DU (3%) for TCR-2, which was earlier -1.2 DU (CAMS-iRean) and 1.8 DU (TCR-1).

Largest uncertainties are found for the polar regions, with MB within 2.6 DU and the ~~RMS~~ stddev ranging between 1.~~4~~5 (CAMS-~~REAN~~-Rean) to 2.1~~6~~ (TCR-1) DU, corresponding to up to ~~~15~~12% of the average O_3 tropospheric column. Over the SH mid latitudes the reanalyses show similar features as over the Antarctic, with normalized mean biases within -1DU (-5%, CAMS-~~iREAN~~-iRean) and 1.5 DU (+10%, TCR-1). The normalized standard deviations over the SH mid latitudes are within 7%, marking a considerably better ability to capture temporal variability than over the Antarctic. ~~marked by a relatively large under-estimate for CAMS-iRean by 1 DU. The stat~~

~~In contrast, in~~ In the tropics the MB ranges within -0.6 to 1.2 DU, and the ~~RMS~~ stddev is ~~up to~~about 1.~~1–1.5~~0DU, or ~5% of the average O_3 tropospheric column. ~~Except for the NH mid latitudes and Antarctic region the seasonal cycle in both model and observations is not very pronounced.~~ The temporal correlation between ~~modelled~~ analyzed and observed tropospheric columns is correspondingly highest (R>0.90) for the NH mid-latitudes, but still relatively low for the Antarctic region (R<0.80~~0.84~~) for all reanalyses. This relatively poor temporal correlation over the Antarctic, despite the strong seasonal cycle, does indicate difficulties of the reanalyses to reproduce a consistent seasonality over the full time series, as described in more detail in the following sections.

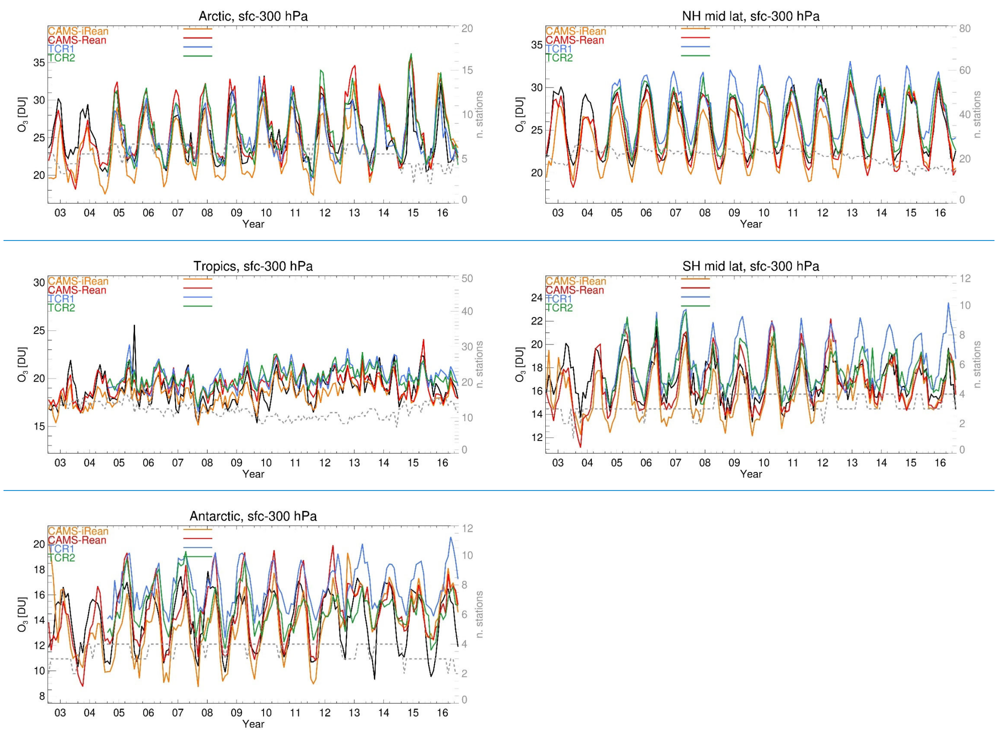


Figure 4~~3~~: Evaluation of zonally averaged monthly mean ~~tropospheric~~ partial columns (surface – 300 hPa) against sonde observations. Observations are in black. The gray dashed line refers to the number of stations that contribute to the statistics (right vertical axis).

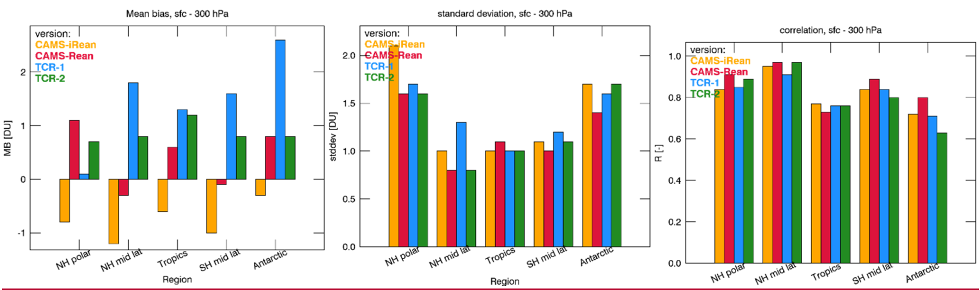


Figure 5. Analogous to Figure 3 but for partial columns from surface to 300hPa, in Dobson Units (DU).

4.3 Time series of regionally averaged O_3 biases at multiple altitudes

Figure ~~4~~6 shows time series of monthly mean ozone biases against ozone sondes at three pressure levels (~850, 650, and 350 hPa), aggregated for the predefined regions. ~~These~~ The panels in this figure~~s~~ give an indication of the stability of the reanalyses against sonde observations during the 2003–2016 time period. The corresponding timeseries with monthly mean concentration values, showing the seasonal cycle, is given in Figure S~~1~~3 in the Supplementary material. As in the previous section, persistent changes in the number of stations may contribute to changes in biases over the course of the fourteen year time interval. The mean bias, ~~RMSE~~ stddev and temporal correlation for each of these time series ~~have been~~ are given in ~~Tables~~ Figure 3~~5-7~~. Based on these evaluations we note the following: Over the **NH polar** region, CAMS~~-REAN~~-Rean shows a small positive bias in the lower troposphere (2.7 ppb at 850 hPa for the 2005-2016 ~~multi-annual~~multiannual mean), particularly during the springtime (5.0 ppb when averaged over March~~A~~-May). During 2003 and 2004 both CAMS reanalyses show anomalously low springtime ozone, different to the rest of the time period, particularly at ~350 hPa. ~~CAMS-iREAN shows a large offset compared to observations and CAMS-REAN in 2003. This is attributed to the assimilation of GOME observations in CAMS-iREAN, which has been omitted in CAMS-REAN (Inness et al., 2019).~~ The different ~~error~~ reanalysis performance statistics ~~for 2003~~ over the Arctic compared to later years is ~~furthermore~~ attributed to the use of early SCIAMACHY and NRT MIPAS O_3 retrievals, which are of poorer quality than the OMI MLS observations which have been used from August 2004 onwards, and reprocessed MIPAS data used from January 2005 onwards. Combined with total column retrievals, assimilation of such stratospheric profiles has been shown to also affect the tropospheric contribution (Inness et al., 2013). CAMS~~-iREAN~~-iRean shows a large offset compared to observations and CAMS~~-REAN~~-Rean in 2003, particularly at altitudes below 650 hPa. This was~~is~~ attributed to the assimilation of GOME nadir profiles ~~observations~~ in CAMS~~-iREAN~~-iRean, which has been omitted in CAMS~~-REAN~~-Rean (Inness et al., 2019).

Furthermore, before 2014 CAMS~~-iREAN~~-iRean shows lower values than CAMS~~-REAN~~-Rean, while for 2014 to 2016 the two CAMS reanalyses are much more alike. This offset before 2014 results in a slight negative bias against observations at ~850 hPa over the Arctic, and a significant negative bias at ~650 hPa, and is attributed to the use of a different version of the MLS retrieval product from V2 to V3.4, Flemming et al. (2017). The TCR reanalyses underestimate the lower tropospheric ozone after 2011, which could be associated with the lack of TES measurements during the recent years. At higher altitudes (650 and 350 hPa) differences between the reanalyses are relatively smaller. On average at 650 hPa CAMS~~-iREAN~~-iRean shows a slight underestimation (-3.1 ppb), while CAMS~~-REAN~~-Rean and TCR-1 bias is below 1 ppb, and slightly larger for TCR-2 (2.4 ppb). At 350 hPa all reanalysis products perform overall similar. At this altitude a considerable inter-annual variability is visible in the observations, see Figure S3 in the Supplementary Material, which appears to be well captured by the reanalysis products, with temporal correlations in the order R=0.85 (for TCR-1) to R=0.92 (CAMS~~-REAN~~-Rean). Also both the observations and reanalyses indicate an upward trend of tropospheric ozone in the Upper Troposphere - Lower Stratosphere (UTLS), as also confirmed by Williams et al., (2019).

Over **Western Europe** the CAMS reanalyses show good correspondence to the observations at 850 hPa from 2004 onwards, with mean biases of -1.9 (CAMS-iREAN-iRean) and 0.4 ppb (CAMS-REAN-Rean). The TCR reanalyses overestimate ozone at lower altitudes, particularly in TCR-1 before 2010, which shows positive biases at 850 hPa of up to ~15 ppb, with an average over the full time period of 3.3 ppb. Such overestimates suggest a strong influence of the forecast model performance for the boundary layer (e.g., mixing and chemistry), while the optimization of the emission precursors was not sufficient to improve the lower tropospheric ozone analysis. At ~650 and ~350 hPa, the chemical reanalyses reproduced well the observed seasonal and interannual variations. As an exception, TCR-1 overestimates ozone for some cases, especially in winter. In contrast, the CAMS reanalyses show average (absolute) biases less than 3.3 ppb at all pressure levels.

Over the **Eastern US**, all the reanalysis products show similar RMSE stddev values at ~850 hPa (3.0–4.7 0 ppb), which is associated with positive model analysis biases, mostly during summer by 0.3–6.8 ppb. Such model biases have also been reported in dedicated other model studies (e.g., Travis et al., 2016), which could be associated with model errors, for instance, excessive vertical mixing and net ozone production in the boundary layer. The annual mean bias for the reanalyses ranges between -2.3 and 2.6 ppb. A decrease in the observed ozone concentrations at ~850 hPa after 2014, associated to a change in the number of contributing stations in this evaluation, leads to a general and consistent over-estimate in all of the reanalyses. A similar agreement with the observations was found in the middle troposphere compared to the lower troposphere, with RMSE stddev ranging between 2.93.0 and 4.74.9 ppb, while at ~350 hPa the RMSE stddev ranges between 8.6 and 11.1 9 and 11.7 ppb.

Over **Japan**, all reanalyses on average overestimate ozone at 850 hPa and 650 hPa before 2011, with relatively large positive biases in TCR-1 and TCR-2 at 650 hPa (7.9 and 6.9 ppb, respectively, when averaged for the 2005-2010 time period). From 2011 onwards the correspondence with observations improves remarkably, despite the lack of TES measurements in TCR from June 2011 onwards. The changes in performance statistics for all reanalyses likely have multiple causes. This includes trends in the observed ozone (Verstraeten et al., 2015), associated to changes in Chinese precursor NO_x emissions (e.g. van der A et al., 2017). Also changes in the observing system are important to consider, particularly the reduction of assimilated TES measurements in TCR from 2010 onwards, and the row anomaly issues affecting assimilated OMI O_3 and NO_2, see also Sec. 2.5.

In the **tropics**, all reanalyses except CAMS-iRean all of the reanalysis products overestimate ozone at 850 hPa before 2012, with large positive biases in CAMS-REAN and TCR-1 at 850 hPathe range 2.5-3 ppb. The different performance for CAMS-iRean from 2012 onwards is probably associated to the use of another version of the MLS retrieval product. Interestingly, both CAMS reanalyses show a strong peak in ozone at 850 hPa (and to lesser extent at 650 hPa) during the second half of 2015 (see corresponding Figure S3 in the Supplementary material), but with a zonally averaged overestimation of up to 20 ppb. This is associated to the strong El Niño conditions, and this particular spike was attributed to an over-estimate of ozone observed at the Kuala Lumpur station for October 2015. Here exactly the grid box affected by the extreme fire emissions in Indonesia for this period (Huijnen et al., 2016), as prescribed by the daily GFAS product, has been sampled. This peak appears much weaker in TCR, . probably Possible explanations are lower optimized NO_x and CO emissions in TCR compared to those used in CAMS, resulting in owing to the lack of direct ozone measurements together with underestimated weaker ozone production, together with a and coarser reanalysis model resolution. At 650 hPa, the TCR reanalyses overestimate ozone almost throughout the reanalysis period (by 3.1–3.8 ppb on average), whereas the CAMS-Rean shows closer agreement with the observations (mean bias = 0.5 ppb, RMSE stddev = 3.2 ppb). At ~350 hPa, the TCR-2 shows improved agreement compared with the earlier TCR-1, as confirmed by improved mean bias (from 4.3 to 0.6 ppb) and RMSE although similar stddev (from 6.64.9 to 5.74.7 ppb). although Also the temporal correlation remains relatively low.

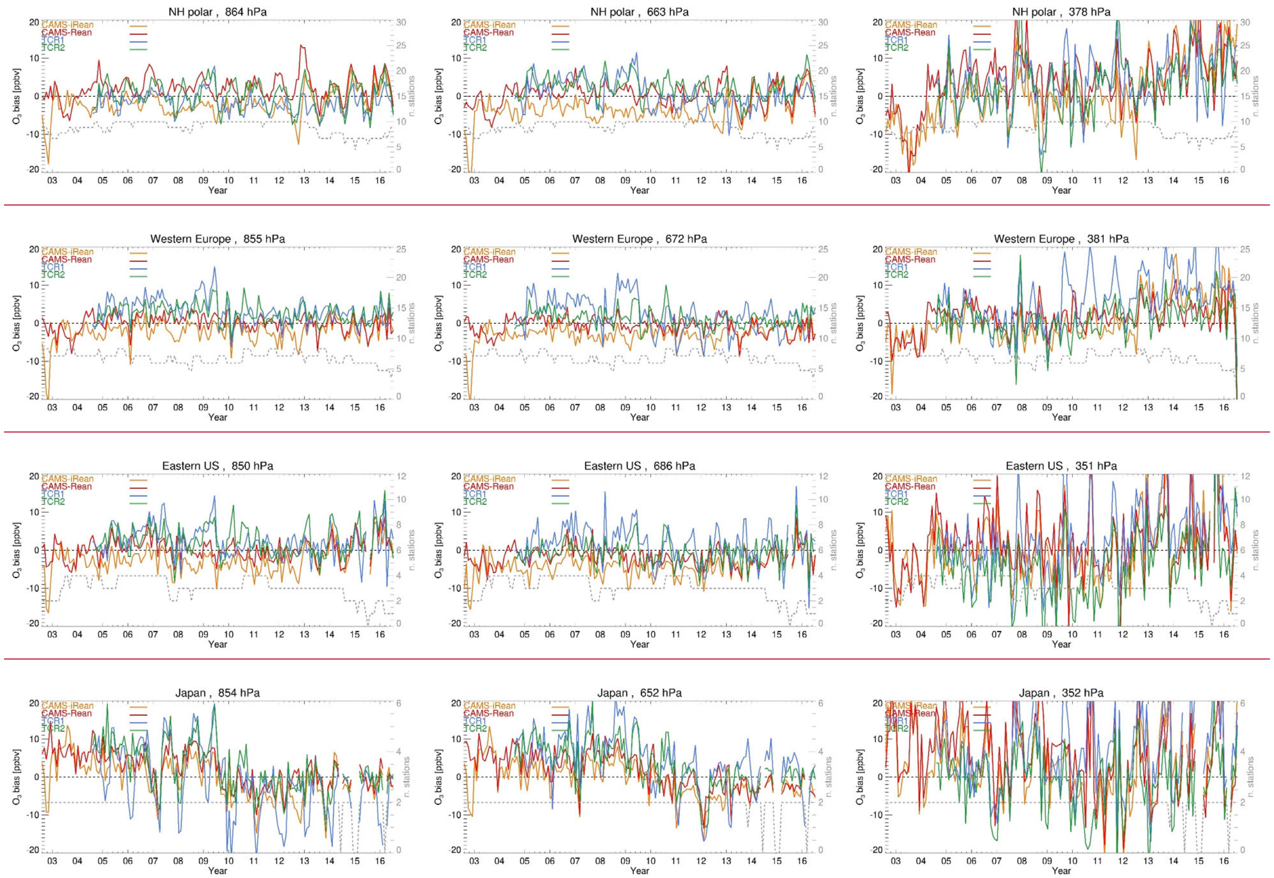
Met opmaak: Subscript

Met opmaak: Subscript

Met opmaak: Subscript

Over the **SH mid-latitudes** an overall ~~remarkably~~ good correspondence is ~~generally~~ obtained for all reanalyses, but particularly CAMS~~-REAN~~-Rean and TCR-2, throughout the troposphere. This is marked by the lowest magnitudes for ~~RMSE~~ stddev and highest for the temporal correlations, for any of the three altitude ranges compared to the statistics in other regions. Nevertheless, CAMS-iRean still underestimates ozone before 2012 in the lower and middle troposphere, whereas TCR-1 overestimates it particularly at ~~332~~ 382 hPa after 2010. Furthermore, CAMS-iRean and CAMS-Rean suffer from relatively large negative biases before 2005, particularly at 382 hPa. This is attributed to similar causes as have been discussed for the Arctic region.

~~In contrast, l~~A large diversity among the system~~s~~ performance is seen over the **Antarctic**. As in the Arctic region, free tropospheric O_3 in the CAMS reanalyses is comparatively poorly constrained during 2003, as consequence of the use of the NRT data product from MIPAS and early SCIAMACHY data in the assimilation. Also in the period between the end of March and the beginning of August 2004 no profile data were available for assimilation, leading to a temporary degradation in the reanalysis performance.



Met opmaak: Breedte: 29,7 cm, Hoogte: 21 cm

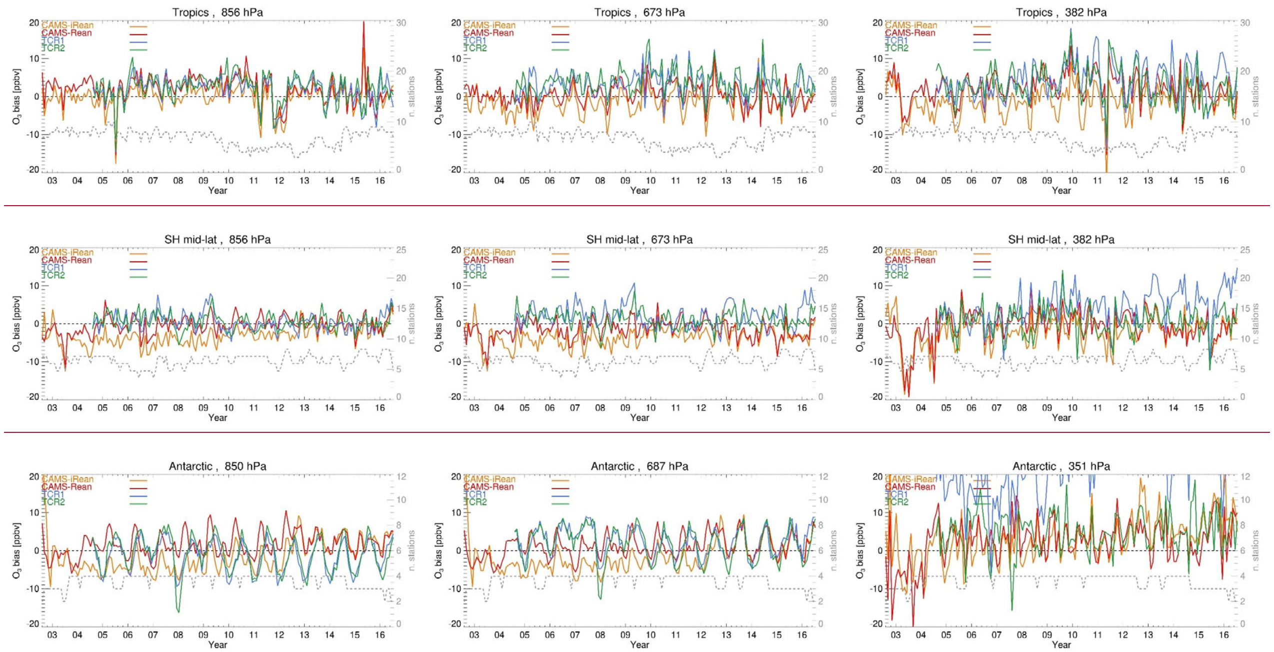


Figure 64. Time series of regionally and monthly aggregated ozone biases at different altitudes (850, 650 and 350 hPa), sampled at ozone sonde locations., against ozone sonde observations (black). As in Figure 4, the gray dashed line refers to the number of stations that contribute to the statistics (right vertical axis).

Before 2013, CAMS ~~iREAN~~ iRean underestimates the low ozone values in the lower and middle troposphere during austral spring, while CAMS ~~REAN~~ Rean overestimates it during austral winter. Afterwards, both systems show very similar results, also in overall better agreement with the observations, even though an overestimate during austral spring remains. Reasons for the change in behaviour in CAMS ~~iREAN~~ iRean is ~~a~~ the change MLS version from ~~v2~~ V2 to ~~v3~~V3.4 after 2012. Furthermore both CAMS ~~iREAN~~ iRean and CAMS ~~REAN~~ Rean are affected by a change from 6L SBUV to 21L NRT data in January and July 2013 respectively, which appears to contribute significantly to the changes in the bias. The seasonal cycle in the biases can largely be attributed to the lack of O_3 total column observations during polar night, combined with a seasonal variation in model forecast biases. The TCR reanalyses largely underestimate ozone during austral summer and autumn in the lower troposphere. At 351~~32~~ hPa, TCR-1 substantially overestimates ozone throughout the year (22 ppb on average) because of large model biases and the lack of observational constraints~~, which~~. This large positive bias was resolved in TCR-2 by improving the modelling framework.

In conclusion, evaluation of the tropospheric ozone reanalyses against ozone sondes has revealed the following:

- The updated reanalyses show on average improved performance compared to the predecessor versions, but with some notable exceptions, such as an increased positive bias over the Antarctic in CAMS-Rean versus CAMS-iRean. Over the Antarctic the TCR-2 strongly improved upon TCR-1, despite the lack of direct observational constraints.
- For individual regions or conditions CAMS Reanalysis and TCR-2 show different performance, but averaged for all regions of similar quality. Best performance, in terms of mean bias, standard deviation and correlation, for the updated reanalyses is obtained for the Western Europe, Eastern US and SH mid latitude regions (both normalized mean bias and standard deviation below 8% at 850 and 650 hPa). Relatively worst performance is found for the Antarctic region, with normalized standard deviation up to 18%. This is likely associated to the fewer observational constraints in the polar regions compared to the other regions.
- In terms of temporal consistency, the CAMS Reanalyses show degraded performance over the polar regions during 2003 and 2004, due to lower quality MIPAS and SCIAMACHY data usage. CAMS-iRean also shows a change in performance statistics in the polar regions from 2014 onwards, associated to a changes in the MLS retrieval product versions. Furthermore, both CAMS-Rean and CAMS-iRean are affected by the change in the SBUV/2 product versions in 2013.

With the reduced data-availability from TES from 2010 onwards the TCR tropospheric ozone products show changes in their performances. Remarkably, TCR-1 and TCR-2 show overall slight improvements from 2010 onwards. This is marked by reduced positive biases in the lower troposphere over NH-mid-latitude regions and may be attributed to biases in the TES retrieval product, combined with changes in the OMI product, see also Sec. 2.5. Additional Observing System Experiments (OSEs) are needed to identify the relative roles of individual assimilated measurements on the changes in reanalysis bias.

Met opmaak: Breedte: 21 cm, Hoogte: 29,7 cm

5. Validation against TOAR surface observations

We evaluated the reanalyses against monthly mean, gridded surface observations filtered for measurements performed at rural sites, as compiled in the TOAR project (Schultz et al., 2017). These evaluations reveal the ability of the reanalysis products to reproduce near-surface background ozone concentrations in terms of mean value and variability, both temporally, on seasonal to annual time scale, and spatially, for various regions over the globe.

5.1 ~~Multi-annual~~Multiannual mean

Figure ~~5~~7 shows a map with ~~multi-annual~~multiannual mean ozone observations from the TOAR database, for the 2005–2012 time period, as well as the corresponding normalized biases in surface ozone for the reanalysis products. Detailed maps for North America, Europe and East Asia are given in Figure S4~~2~~ in the Supplementary Material, while the corresponding regional mean biases are given in Table 9.

The TCR-reanalyses show significant positive biases for many regions, with ~~multi-annual~~multiannual mean biases of 11.0 ppb and 6.8 ppb over the Eastern and Western US, and 6.7 ppb over Europe in TCR-2. These biases can mainly be attributed to model errors. Mean biases in the CAMS-reanalyses are generally smaller (1.5 ppb and -0.2 ppb for Eastern and Western US, respectively, -1.8 ppb for Europe), but still show substantial spatial variations, as quantified by the root-mean-square of the ~~multi-annual~~multiannual mean differences across the various regions, which is 8.9 ppb and 6.1 ppb for Eastern and Western US, and 5.6 ppb over Europe for the CAMS Reanalysis (18, 11 and 11 ppb for TCR-2 for these regions). The mean bias is negative over the Arctic, Europe and the Western US and positive over East Asia and Southeast Asia in both versions of the CAMS reanalyses. The positive regional mean biases over the major polluted regions are reduced by 35 to 55% in TCR-2 as compared with TCR-1. Likewise, the negative biases over the Arctic, Europe, the Western US, and SH mid and high latitudes are reduced by more than 25% in CAMS~~-REAN-~~Rean as compared with CAMS~~-iREAN-~~iRean, illustrating overall improvements for the newer reanalyses.

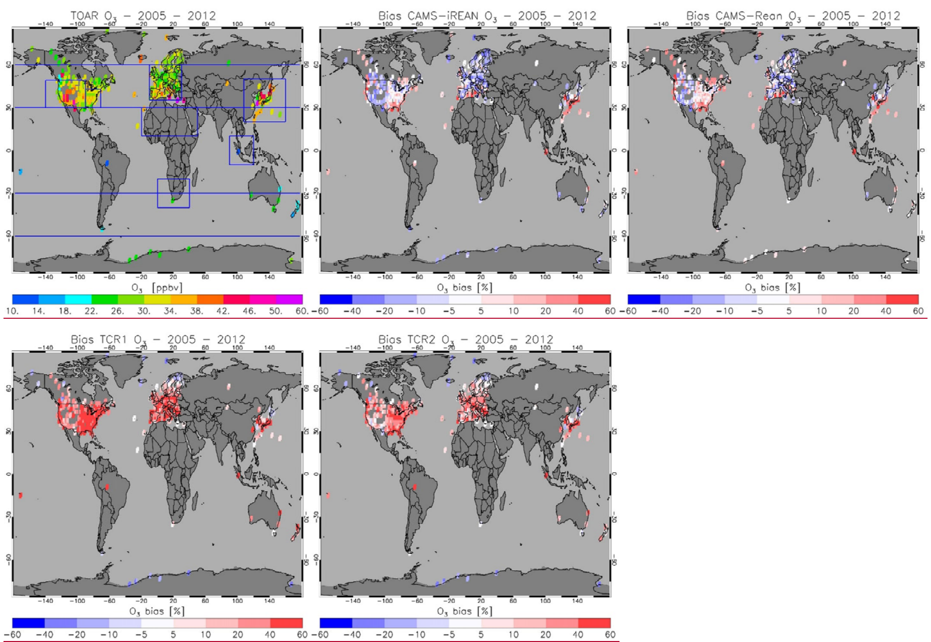


Figure ~~7~~5: ~~Multi-annual~~Multiannual (2005-2012) mean surface ozone from TOAR (upper left), along with corresponding ~~relative~~ normalized mean bias with respect to the observations for the reanalyses.

Table 9. Mean bias (ppb) for the reanalyses against TOAR monthly mean, regional mean observations for the 2005-2012 time period, as sampled for observations in the specified regions indicated in Figure ~~4~~7.

	Arctic	Europe	Eastern US	Western US	Southeast Asia	East Asia	SH mid-latitudes	Antarctica
CAMS~~-iREAN-~~iRean	-4.5	-2.4	0.1	-1.9	5.6	4.5	-2.2	-3.5
CAMS~~-~~	-1.5	-1.8	1.5	-0.2	6.7	2.7	-0.1	1.1

~~REAN~~- Rean								
TCR-1	-1.8	11.2	17.2	12.9	15.8	10.5	3.6	-5.7
TCR-2	-2.3	6.7	11.0	6.8	7.4	7.6	1.5	-5.4

5.2 Variability in regionally averaged surface ozone

Figure 8~~6~~ presents scatter plots of monthly mean ~~the temporal variability from the~~ ozone (2005-2012) from the reanalyses against those from the TOAR surface observations ~~at the surface~~for various regions. The corresponding time series, ~~are~~ given in Figure S5~~S3~~ in the Supplementary Material~~.~~, indicate that the main driver of the variation in magnitude of ozone concentrations in the reanalyses and observations is associated to the seasonal cycle. Over the Arctic, the general pattern in the ~~seasonal~~ monthly variations is captured for all reanalyses (R between~~=~~ 0.58 and -0.72), although they all underestimate the increased ozone values during boreal spring. ~~A remarkable positive bias is seen in springtime 2013 in CAMS-REAN, as was also visible in the bias plots against ozone sondes in the lower free troposphere, Figure 4.~~

Over Europe and the US, the CAMS reanalyses show the closest agreement with the observations (MB between ~~=~~ -2.4 ~~–~~and 1.5 ppb, R>0.8~~),~~). ~~furthermore~~ CAMS-Rean shows improved ~~showing reduced~~ ~~model~~ negative biases for observed low ozone values compared with the CAMS-iRean, which is in boreal winter and spring. ~~in CAMS-REAN compared with the CAMS-iREAN.~~ The TCR reanalyses exhibit large positive biases ~~in boreal summer~~ over Europe and the US regions (MB between 6.7 ~~–~~and 17 ppb), with significant~~ly lower biases~~ improvements in TCR-2. Over East Asia, all the reanalyses show positive biases in the range of 2.7 ppb (CAMS-~~REAN~~-Rean) to 10.5 ppb (TCR-1) and fail to reproduce the minimum concentrations in autumn. Still the temporal correlations are similar to most other regions (R between ~~=~~0.79 ~~–~~and 0.83), associated with the stable seasonal cycle in both the reanalyses and observations. Over Southeast Asia, positive biases exist throughout the period, which are largest in TCR-1. For this region the TCR-reanalyses show lower temporal correlations (R~~=~~ between 0.39 ~~–~~and 0.49) compared to the CAMS reanalyses (R=0.68).

Significant changes in the surface ozone biases are found in the TCR reanalyses over the SH mid latitudes, with reduced ~~values~~ positive biases after 2010.

The CAMS reanalyses capture well the temporal variability ~~seasonal cycle~~ over the SH mid latitudes and Antarctic (R~~=~~ between 0.89 ~~–~~and 0.96), while CAMS-~~REAN~~-Rean shows a positive bias for observed high ozone values. This is associated to ~~during~~ model biases austral winter (JJA), particularly during 2005-2013, Figure S5. The TCR reanalyses show a significant negative bias throughout the year except for observed low ozone values (during Austral summer) which results in lower temporal correlations (R~0.68).

The free tropospheric intercomparison at different altitudes, as presented in Figure 3~~reported in Tables 5 to 7~~, already indicated generally larger biases ~~with decreasing altitude near the surface~~at 850 hPa compared to 650 hPa. This can be understood as near-surface ozone concentrations are less well constrained by the satellite data products used in the assimilation, and they depend strongly on local conditions such as precursor emissions, deposition, vertical mixing, and chemistry, which are difficult to parameterise at the model grid scale (Sekiya et al., 2018).

An important example of a driver for local variability is the emissions from forest fires which in the CAMS reanalyses are provided through daily-varying GFAS emissions. This has been shown to capture to a good degree the carbon monoxide and aerosol from fire plumes, although larger uncertainties exist in the NO_x emissions, e.g. Bennouna et al. (2019).

In summary, CAMS-Rean shows the best ability to capture the regional mean surface ozone and its variability, while particularly TCR-2 (and to lesser extent also TCR-1) shows positive biases and reduced correlations. Particularly good performance is seen over the western US (R=0.95, MB=-0.2), while over east, and particularly southeast, Asia the performance is poorest.

Met opmaak: Subscript

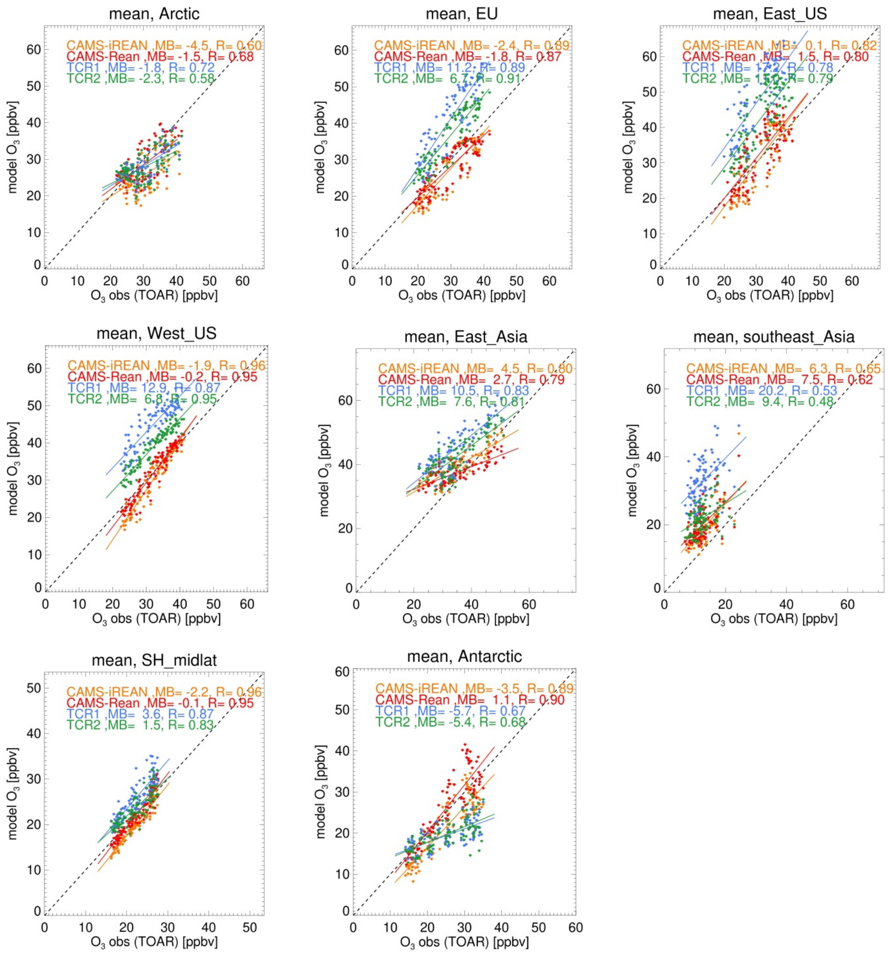


Figure ~~6~~8: scatter plot of surface ozone against TOAR observations for the 2005-2012 regionally averaged, monthly mean time series. Also the mean bias and temporal correlation is given.

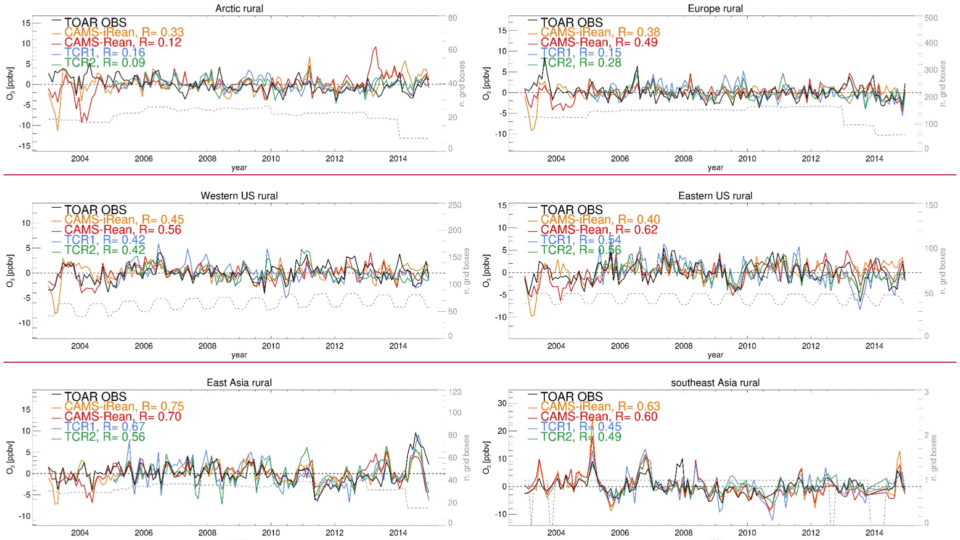
5.3 Interannual variability of regionally averaged surface ozone

We ~~compute~~ assess the interannual variability (IAV) by computing the deseasonalized anomaly of surface ozone concentrations. For this, ~~subtracting~~ the 2005-2012 ~~multi-annual~~multiannual monthly, regional mean surface ozone is subtracted from its corresponding instantaneous monthly, regional mean value, both for the reanalyses and for the TOAR observations, see Figure ~~7~~9. By doing so, we remove the ~~model~~ analysis bias, as well as the seasonal cycle. No clear long-term trends are visible in the regional mean surface ozone concentrations. Nevertheless, the observations reveal distinct deviations from the 8-year mean value, which point at temporary anomalies in meteorological conditions and/or emissions. Note that large fluctuations in the time series can also occur due to changes in the observation network. Therefore, when evaluating the temporal correlations between observed and ~~modelled~~ analyzed anomalies we exclude individual months with low data coverage, defined as months where the number of grid boxes with observations is less than half of its average number for the complete time series.

Overall, the reanalysis anomalies are in reasonable general agreement with those seen in the observations, with better skill for regions at low latitudes compared to those at high latitudes. Also for 2003–2004 the CAMS reanalyses mostly show larger deviations than justified from the observations, particularly the first months for CAMS~~-iREAN~~-iRean. This is attributed to the inconsistencies in the assimilated satellite retrieval products as already described. Also the observed positive anomaly associated to the 2003 heatwave period over Europe is therefore not equally seen from the CAMS reanalyses, but with an offset (see also Bennouna et al., 2019). For later years, the magnitude of the ~~IAV~~ anomalies correspond~~s~~ better to the observations. Over the Arctic the temporal correlation is generally low (R<0.33). For Europe CAMS~~-REAN~~-Rean shows a largest correlation (R=0.49). For the Eastern US region all reanalyses follow an extended dip during 2009, as seen from the observations, and also a second dip during 2013, particularly captured by TCR-2, also resulting in relatively good temporal correlations (R~~=~~ between 0.4 ~~—~~and 0.64). Also, in the Western US the temporal correlations are acceptable (R between ~~=~~0.42 and~~–~~ 0.56). Over East Asia the correlations are relatively high (R~~=~~ between 0.56 ~~—~~and 0.75), and likewise for the station in Indonesia (southeast Asia) with R=0.45 – 0.63. Here all reanalyses capture the increases in surface ozone in early 2005 and late 2006, and the decrease in 2010.

Over the SH mid latitudes and Antarctic the ozone reanalyses show overall a relatively poor temporal correlation (R<0.37), particularly for TCR (R<0.23). For these regions the TCR reanalyses show larger anomalies during 2007-2009 as compared with observations, whereas the CAMS reanalyses show larger anomalies from 2012 onwards. Figure 9~~5~~ suggests that this is particularly caused by the change in system behaviour after 2012, as already described in Sec 5.2 evaluating the tropospheric ozone over the Antarctic. As was the case there, for surface ozone the CAMS reanalyses in fact show a better match to the observations from 2013 onwards.

In conclusion, the reanalyses considered here show some skill to capture IAV in monthly mean ozone surface concentrations, in particular for the tropical, sub-tropical and NH mid-latitude regions. In these regions the signal of the observed ozone variability is also larger than for the comparatively stable Arctic conditions. Here the performance is hampered due to changes in the overall ~~model~~ bias of the analyses over time.



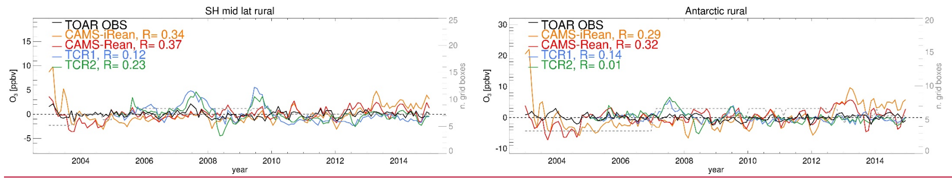


Figure ~~7~~9: Time series of regional, monthly mean ozone anomalies against those derived from the TOAR observations. The dashed line indicates the number of TOAR 2°×2° grid boxes that contribute to the statistics. Also the temporal correlation as computed for the 2005-2014 time series is given.

6. Evaluation of surface ozone in 2006

To assess the ability of reanalyses to cope with local situations, and specific meteorological conditions, we analysed their performance over Europe in 2006, with a focus on the ability to capture the diurnal and synoptic variability during the heat wave event that affected large parts of Europe during July 2006 (Struzewska and Kaminski, 2008). Here we use the ground-based observations from the EMEP network. For this evaluation we note that these large-scale models do not represent local orography. Therefore we select the appropriate model level depending on its pressure level, which is representative for mean pressure at the observation site (Flemming et al., 2009). Figure ~~8~~10 presents the evaluation at two EMEP stations in Great Britain, during July 2006, illustrating the general performance of the reanalyses for this situation. The Lullington Heath station (50.8° N, 0.2° E, 120 m.a.s.l.) is located in a Nature Reserve area, near the coast south of London. Great Dun Fell observatory (54.7° N, 2.4° W, 847 m.a.s.l) is located on a mountain summit, approximately 15 km north of Manchester. Both stations show enhanced levels of ozone in the first part of July, as well as during 16-20 July. In contrast to Great Dun Fell, Lullington Heath shows a pronounced diurnal cycle. For this evaluation, the reanalyses are sampled at different model levels (see figure caption). Note that the TCR reanalyses have fewer model levels towards the surface than the CAMS reanalyses. All reanalyses capture both the diurnal and synoptic variability with a significant improvement in TCR-2 compare to TCR-1, while the CAMS reanalyses are more alike. Particularly for Lullington Heath, the CAMS reanalyses and TCR-2 show remarkably small biases (MB < 3.6 ppb). Also at Great Dun Fell the synoptic variability is generally well captured, particularly for the CAMS reanalyses and TCR-2.

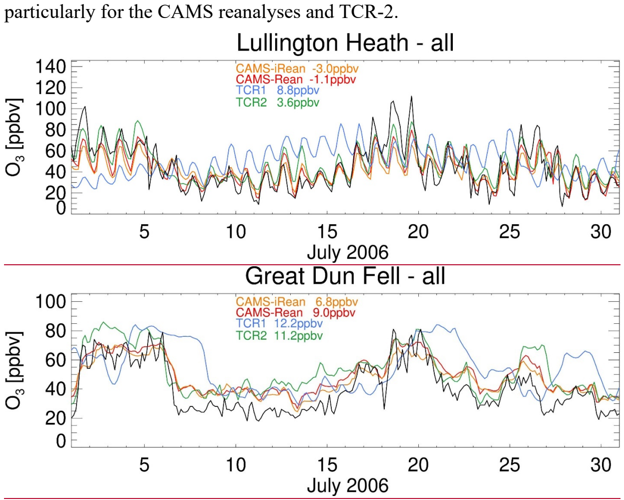


Figure ~~8~~10: Time series of reanalyses against ozone observations at two EMEP stations in Great Britain during July 2006: Lullington Heath (~~left~~top, 120m a.s.l., model level 3 (CAMS) and 1 (TCR)) and Great Dun Fell (~~right~~bottom, 847m a.s.l., model level 8 (CAMS) and 3 (TCR)). Also given are mean biases.

A more quantitative assessment of the ~~models'~~ ability of the reanalyses to capture the ozone variability is presented in Figures ~~9~~11 and 1~~0~~2, which show a graphical presentation of the temporal correlation coefficient at EMEP stations for December-January-February (DJF) and June-August (JJA) 2006, computed interpolating the ~~model~~ reanalyses and observational results onto a common 3-hourly time frequency.

In the DJF period, regionally averaged correlation coefficients range from 0.45 (TCR-1) to 0.58 (CAMS-~~iREAN~~-iRean). Comparatively high correlations were found over western Europe (particularly over the southern part of Britain), with R>0.8 for the CAMS reanalyses, and R>0.6 for TCR. The lower correlations over the regions in the TCR reanalyses could be associated with its coarser model resolution.

For the summer period (JJA, Figure 12~~0~~), temporal correlations are overall higher than in the winter period, most markedly by better correlation statistics over south-western, eastern and northern Europe. This is due to the more pronounced diurnal cycle during summer and results in generally consistent correlation over any of the stations across the European domain. The average values range between R=0.61 (TCR-1) and 0.68 (CAMS-~~iREAN~~-iRean). Only stations sampling ozone around the Mediterranean are consistently poorly captured, with R<0.5.

Temporal correlations for the March-April-May (MAM) and September-October-November (SON) seasons are in-between those for DJF and JJA~~,~~. ~~the~~ The CAMS-~~REAN~~-Rean correlations are on average lower by ~0.02 than those of CAMS-~~iREAN~~-iRean, while TCR-2 has systematically improved ~~temporal correlation~~ by 0.02–0.05 over TCR-1.

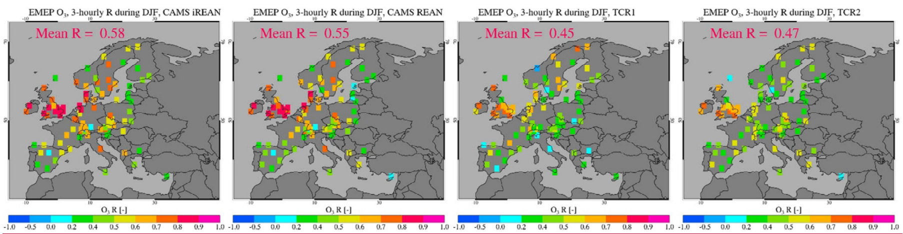


Figure ~~9~~11: Correlation coefficients computed for 3-hourly DJF 2006 at EMEP stations for the four reanalyses. The mean value, based on correlations computed for all individual stations, is given.

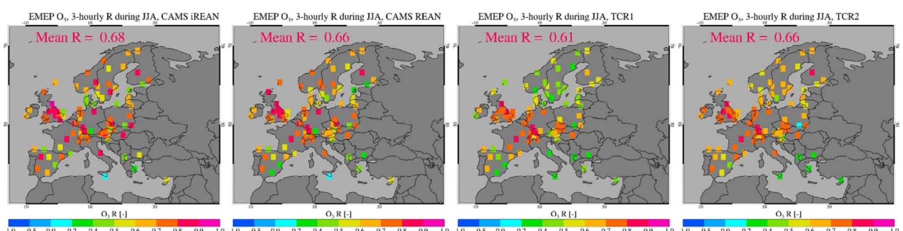


Figure 1~~0~~2: As figure ~~9~~11, but for JJA.

A closer look at the diurnal cycle for different seasons and regions over Europe is given in Figure 13~~1~~. In this figure the ~~model~~ seasonal mean reanalysis biases have been subtracted in order to assess ~~the model~~ their ability to capture the diurnal cycle only. All reanalyses generally capture the diurnal variability, and its variation across latitude region and season. For instance, all reanalyses show little diurnal variability for Northern European stations during DJF, although the CAMS-based reanalyses (and particularly CAMS-~~REAN~~-Rean) show enhanced night-time O_3, which is not in TCR nor in the observations. Except for isoprene, no diurnal cycle in O_3 precursor emissions has been adopted in the CAMS reanalyses, which contributes to biases in the diurnal cycle. Note, however, that CAMS-~~REAN~~-Rean shows a comparatively large mean bias for these conditions, of -8 ppb (CAMS-~~iREAN~~-iRean bias is -6 ppb).

The diurnal cycle is generally larger for CAMS ~~iREAN~~ iRean than CAMS ~~REAN~~ Rean, overall showing better correspondence to the observations. Particularly over middle and southern Europe during DJF the CAMS reanalyses show a larger diurnal cycle than those obtained with TCR, also better matching to the observations. For MAM differences between the reanalyses are rather small, while during JJA the TCR-2 and CAMS ~~iREAN~~ iRean show largest diurnal cycle across Europe, best matching again to the observations.

In summary, all reanalyses capture the synoptic to diurnal variability, as illustrated by the assessment of the heatwave event in July 2006. Still there are considerable differences in performance, depending on the reanalysis, region and season. While CAMS-iRean and CAMS-Rean perform mostly similar, for TCR-2 a considerable improvement was found compared to TCR-1. Overall better temporal correlations are obtained for the summer period compared to winter, and also for Western Europe compared to the Mediterranean region. Further improvements can be obtained by a better description of surface processes, including emissions and deposition, together with higher spatial resolution modelling.

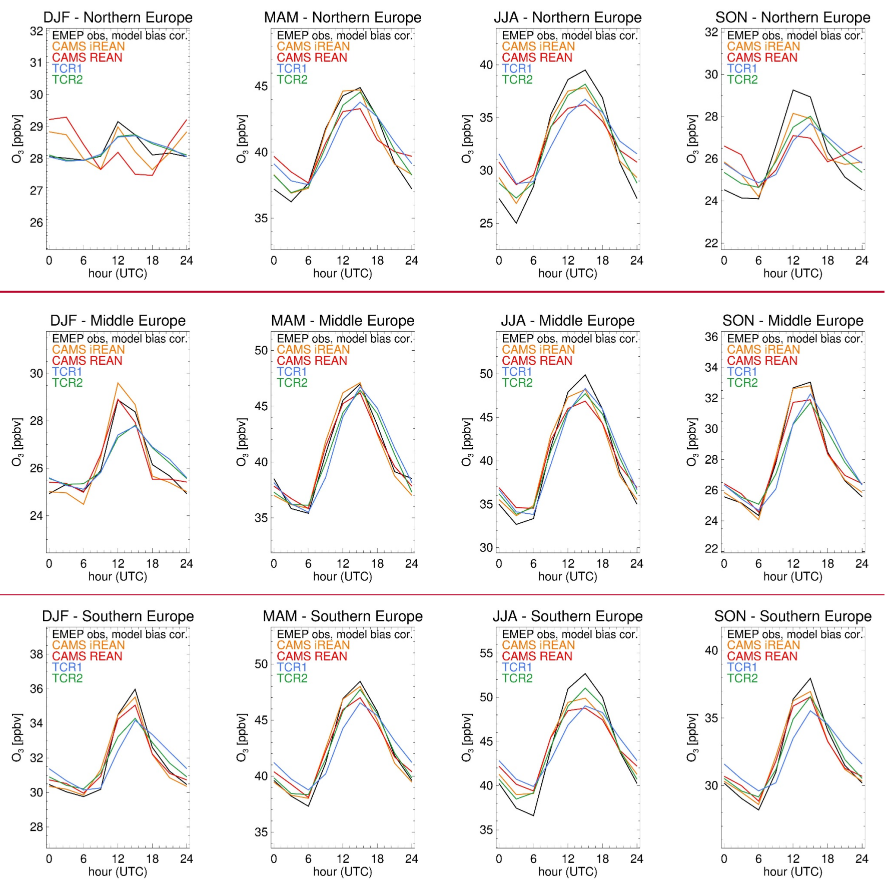


Figure 13~~4~~: Plots of seasonal mean diurnal cycle against EMEP observations for 2006. Middle-Europe is here defined as the region between 35°N and 45°N, with Northern (Southern) Europe at higher (lower) latitudes. Note that the ~~model~~ reanalysis bias has been removed. Model level selection is through matching of the ~~model~~ reanalysis pressure with pressure level of the station sites.

7. Time series of tropospheric ozone columns

Given the detailed validation results of tropospheric ozone profiles (Section 4), this section aims to demonstrate the potential value of the reanalysis products for studies of temporal changes in tropospheric ozone columns associated with changes in chemical and meteorological conditions for different regions of the world. Compared to global tropospheric column products derived directly from observations (e.g. Ziemke et al. 2019), reanalysis products have the potential to better include variations in near-surface ozone, provided that precursor emissions, deposition and chemical conversion are well constrained in the reanalysis.

Met opmaak: Kop 1

Figure 11 shows time series of the annual mean partial ozone columns from the surface up to 300 hPa, for five zonal bands. From this, CAMS-iREAN shows an offset until mid 2013, followed by closer correspondence to the other reanalyses at every zonal band, in particular to the CAMS Reanalysis. The anomalously low columns in CAMS-iREAN before 2013 is due to a switch in the use of MLS data from V2 to NRT V3.4 (Flemming et al., 2017) together with the switch in the version of SBUV in 2013. While these switches implied the introduction of a positive offset in the CAMS-iREAN O_3 *total* columns with respect to CAMS-REAN and observations (Inness et al., 2019), the increased tropospheric columns in fact show overall a better correspondence to CAMS-REAN from this date onwards. The better consistency between CAMS-iREAN and CAMS-REAN could also be seen from the evaluation against sondes, Figures 3 and 4. Note however that from the sonde evaluations there is no overall indication that the CAMS reanalyses perform worse for the period from 2013 onwards, it can rather be characterized as a change in its error statistics.

CAMS-REAN and TCR-2 agree well over the NH extra-tropical regions, but show significant discrepancies over the tropics, with TCR-2 being 0.7 DU (2005) up to 1.8 DU (2016) larger than CAMS-REAN. Considering that tropospheric columns are already overall higher in CAMS-REAN than those derived from in-situ observations (Table 8), this suggests an overall slight over-estimate in TCR-2, particularly in the later period. Whereas CAMS-REAN is close to TCR-2 until 2009, it is closely correlated to the lower tropospheric columns in CAMS-iREAN from 2013 onwards. Nevertheless, the magnitude of the seasonal cycle is consistent for all reanalyses.

Over the SH extra-tropics, CAMS-REAN shows a good consistency with TCR-2, even though for the period before 2013 the amplitude in the seasonal cycle is a little larger. Both at mid-latitudes and high-latitudes there is a remarkable change in behaviour after 2013 in all reanalyses except TCR-1. From 2013 onwards the seasonal cycle is much weaker at mid-latitudes while essentially absent over the Antarctic. This change is largest for both CAMS reanalyses, but also visible in TCR-2, particularly over the Antarctic. Also the evaluations of tropospheric columns against sonde observations show changes in error statistics from 2013 onwards, see also Sec. 4. This shows once again the significant impact of changes in the observing system used to constrain tropospheric ozone, which may have difficulties to cope with the comparatively low magnitudes of tropospheric ozone columns over the Antarctic (~15 DU) compared to the Arctic (~ 26 DU).

Figure 12: Intercomparison of regionally averaged monthly mean partial columns up to 300 hPa.

Figure 13: Anomalies in O_3 partial columns (surface to 300 hPa) in four reanalyses, as compared to the MEI index for two regions: Southeast Asia (90° E - 120° E; 10° S - 20° N) and ENSO3.4 over the Eastern Pacific (120W-170W; 5S-5N). A 2-month smoothing has been applied to the reanalysis data, (as for the MEI index and TSI). Temporal correlations are given in the legends for comparison to MEI. Correlations are calculated on monthly data for the 2005-2016 time period.

Figure 13 shows the anomalies in monthly mean ozone tropospheric columns (surface to 300 hPa) over two regions of the tropics. These anomalies are computed by subtracting the reanalysis-specific mean seasonal cycle based on the 2005-2016 time series. When comparing the anomalies with the Multivariate ENSO Index (MEI), (Wolter and Timlin, 1998), we find a significant correlation with R ranging between 0.6 (CAMS-REAN) and 0.65 (TCR-2) for the Southeast Asia region. A strong anti-correlation for the eastern Pacific region is found with R between -0.70 (CAMS-REAN) and -0.78 (TCR-2). The CAMS iREAN shows a lower correlation for this region, possibly associated with the jump in offset around the beginning of 2013, whose magnitude is significant in comparison to the signal. The high correlation over the Southeast Asia region is associated with enhanced fire emissions, and associated ozone production, during El Nino conditions over Indonesia (Inness et al., 2015), together with suppressed convection, while the anti-correlation over the Eastern Pacific is related to enhanced convection (Ziemke and Chandra, 2003).

~~Figures with anomalies in the monthly mean tropospheric ozone columns, together with their standard deviations are provided in Figure S4 in the Supplementary material. Table 10 presents an evaluation of the correspondence in this IAV between the four reanalyses. This is quantified as the correlation in the tropospheric ozone column anomalies for the four reanalyses. For southeast Asia CAMS-REAN is highly correlated to CAMS-iREAN (R=0.93), and likewise TCR-1 with TCR-2 (R=0.90). Lower, but still clear correlations are obtained particularly between the CAMS reanalyses and TCR-2 (R>0.82). Likewise for the ENSO_3.4 region CAMS-REAN is well correlated to CAMS-iREAN and TCR-2, but poorer correlation is found between CAMS-iREAN and the TCR-reanalyses (R<0.57). Over the entire tropical region, correlations between the various reanalyses are relatively low (e.g. R between CAMS-REAN and TCR-2 is 0.29), caused by the small signal, suggesting little robust information. Still, correlation between TCR-1 and TCR-2 is remarkably larger (R=0.73) than between CAMS-REAN and CAMS-iREAN (R=0.17). Generally smaller standard deviations in the monthly anomalies are found in the updated reanalyses.~~

~~We focus here on correlations between the ozone anomalies from the updated reanalyses (CAMS-REAN and TCR-2). Over the Arctic, and also the Eastern US, these are R=0.60 and R=0.63, respectively, giving some confidence in the robustness of this IAV signal. Over Eastern Asia and Europe, these correlations decrease to 0.52 and 0.42. Over the Antarctic little correlation is remaining (R=0.33), implying that indeed any IAV from the reanalyses should be considered with care. Different reanalyses do not provide a consistent signal. Occasionally (e.g. over Antarctic) better correlations between reanalyses of the same family is found, but for instance over Europe and the Arctic the correlation between CAMS-REAN and TCR-2 is still better.~~

~~Table 10: correlation coefficient R of the interannual variability in tropospheric O_3 columns between the four reanalyses, as computed for the 2005-2016 monthly mean time series in tropospheric O_3 columns from the surface to 300 hPa for different regions.~~

7~~8~~. Global spatial and temporal consistency between reanalyses

Figure 14 shows the ~~multi-annual~~multiannual mean together with an evaluation of its multi-~~model~~ system standard deviation, at different altitude levels. The standard deviation is computed from the ~~multi-annual~~multiannual means of the four reanalyses, and provides a quantification of general agreement between reanalyses. The standard deviation at 850 and 650 hPa is relatively large over South America, Central Africa and Northern Australia, with values exceeding 6 ppb in the lower and middle troposphere. Normalized to local mean O_3 from the CAMS Reanalysis, the standard deviation values at 850 hPa reach 20% over Australia and up to 50% over South America and Central Africa. At 650 hPa these maximum ratios decrease to approx. 10% (Australia) and 20% (South America and Central Africa). These results suggest that the representation of biomass burning emissions and its impacts on ozone production are largely different among the systems. Also large uncertainties in biogenic emissions likely contribute. In TCR, the optimization of NO_x emissions can have strong impacts on the lower and middle tropospheric ozone, in contrast to the CAMS configuration which applies prescribed anthropogenic and biogenic emissions, combined with the daily varying biomass burning emissions. In addition, different representations of convective transport over the continents can lead to diversity in the vertical profile of ozone among the systems.

At 350 hPa, the multi-~~model~~ system standard deviation is large over Central Africa, South America and over the Arctic and Antarctic, which could reflect different representations of deep convection along with biomass burning emissions at low latitudes, and polar vortex, stratospheric ozone intrusions and chemistry treatment at high latitudes among the systems.

The absolute differences between the two most recent reanalyses, TCR-2 and CAMS-~~REAN~~-Rean, are also shown. Apart from the regions mentioned above, differences are significant around Alaska and Siberia, regions with tropospheric ozone influenced by biomass burning events and where observational constraints at such high latitudes are more limited. Such larger discrepancies once again highlight the importance of the forecast model performance in the reanalysis system as discussed in Miyazaki et al. (2019b), especially when direct observational constraints on tropospheric ozone are insufficient.

An evaluation of the consistency across the four reanalyses to describe the seasonal cycle of tropospheric ozone columns, and its interannual variability, is given in the Supplementary Material, Figure S6. From this, the difference in zonal mean partial columns (surface – 300hPa) in the tropics is quantified: TCR-2 is on average higher than CAMS-Rean by 0.7 DU (2005) up to 1.8 DU (2016), corresponding to approx. 3 to 8% of the annual mean column in this region.

Met opmaak: Subscript

Frequency distributions of the ~~multi-annual~~multiannual mean ozone concentrations in the four reanalyses at three altitude levels are given in Figure 15 and summarize the general differences discussed above. In the lower and mid-troposphere the CAMS reanalyses show a larger frequency of O_3 values below 30 ppb (850 hPa) and 45 ppb (650 hPa) compared to particularly TCR-1, but also TCR-2. This is associated to lower ozone in the CAMS reanalyses over the tropical regions. At 350 hPa the CAMS reanalyses and TCR-2 agree ~~to a large extent~~ reasonably in their frequency distribution, with CAMS-iRean showing the largest frequency of relatively low (35-55 ppb) O_3 values ~~. Only~~ and instead TCR-1 ~~shows overall~~ a larger ~~occurrence~~ frequency of ~~O_3~~ values in the range 70-100 ppb compared to the other reanalyses ~~at the expense of primarily lower O_3 values~~. This is associated to a positive ~~model~~ reanalysis bias in this altitude range (see also Table 7). A corresponding evaluation of the frequency distributions, but sampled at individual ozone sonde observations during the 2005-2016 period is given in Figure S7 in the Supplementary material. Because of the different sampling approach the shape of the frequency distributions is different than was seen in Figure 15. Evaluation of the sum in absolute differences *d* between analyzed and observed frequency distributions indicates that at 850 hPa the performance between the four reanalyses is very similar (*d* between 0.17 and 0.19), while at 650 hPa CAMS-Rean is superior (*d*=0.13). CAMS-iRean shows an under-estimate of the frequency of high ozone values (larger than ~55 ppbv) at 850 and 650 hPa, explaining the worst performance at 650 hPa (*d*=0.20). At 350 hPa the differences in performance between reanalyses are largest, with best correspondence to observations for CAMS-iRean (*d*=0.11), and worst for TCR-1 (*d*=0.43).

Deseasonalized anomalies in monthly mean ozone tropospheric columns (surface to 300 hPa) have been computed over various regions by subtracting the reanalysis-specific mean seasonal cycle based on the 2005-2016 time series. Figure 16 presents the reanalysis anomalies together with the Multivariate ENSO Index (MEI), (Wolter and Timlin, 1998) for two regions. Tropical tropospheric ozone variations during El Niño conditions are in part associated with enhanced fire emissions, and corresponding ozone production, over Indonesia, together with suppressed convection (Inness et al., 2015), while the anti-correlation over the eastern Pacific is related to enhanced convection (Ziemke and Chandra, 2003). We find a significant correlation with R ranging between 0.6 (CAMS-Rean) and 0.65 (TCR-2) for the Southeast Asia region. A strong anti-correlation for the eastern Pacific region is found with R between -0.70 (CAMS-Rean) and -0.78 (TCR-2). The CAMS iREAN shows a lower correlation for this region, possibly associated with the jump in offset around the beginning of 2013, whose magnitude is significant in comparison to the signal.

An assessment of the consistency between all reanalyses to describe the deseasonalized anomalies in various regions is given in the Supplementary Material, Figure S8 and Table S1, in terms of the correlations in their anomalies. Specifically, the correlations between CAMS-Rean and TCR-2 over the Arctic, and the Eastern US, are R=0.60 and R=0.63, respectively, giving some confidence in the robustness of this IAV signal in these reanalyses. For various other regions correlations are R=0.52 (Eastern Asia), R=0.42 (Europe) and R=0.33 (Antarctica). Also, when averaged over the full tropical zonal band the correlation decreases to R=0.33, i.e. much smaller than correlations between CAMS-Rean and TCR-2 for the sub-regions Southeast Asia (R=0.82) and ENSO_3.4 (R=0.78). This implies that many of the IAV signals in the reanalyses should be considered with care.

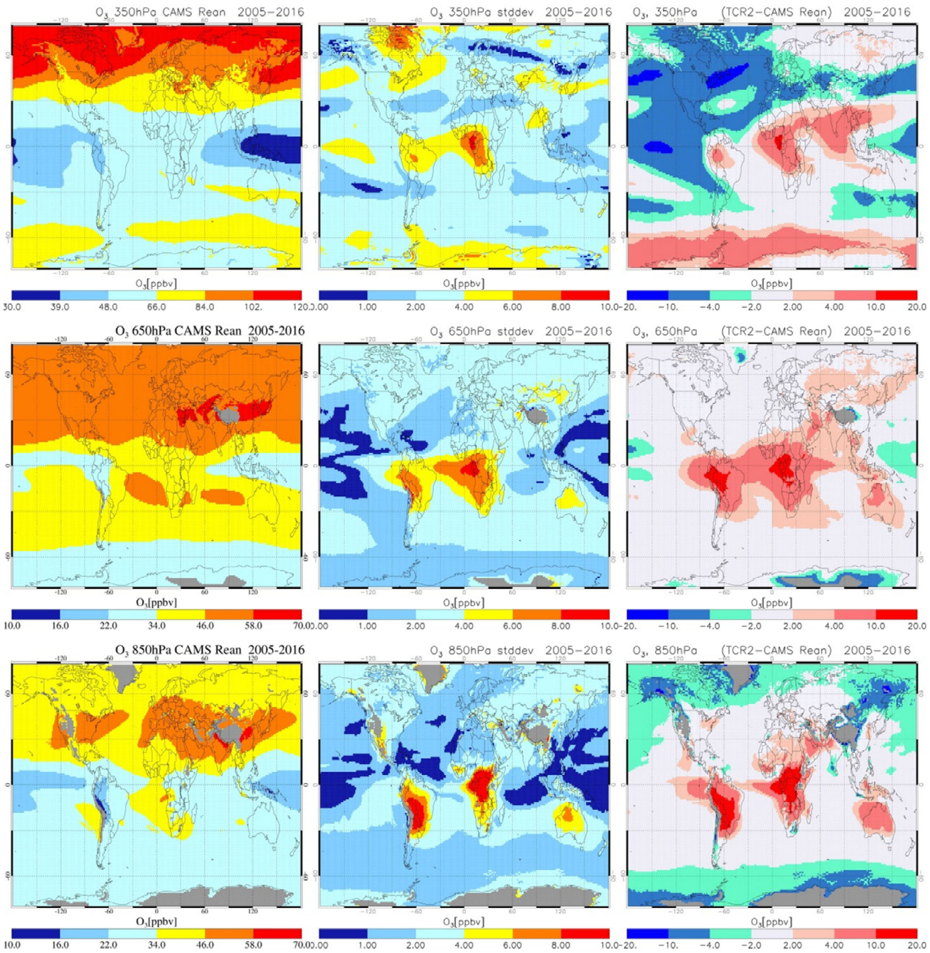


Figure 14: Left: ~~multi-annual~~multiannual mean O_3 at 350, 650 and 850 hPa for the CAMS Reanalysis at 650 hPa over 2005-2016. Middle: standard deviation in the multiannual means for the four reanalyses. Right: absolute difference between TCR-2 and CAMS Reanalysis, all in units ppb.

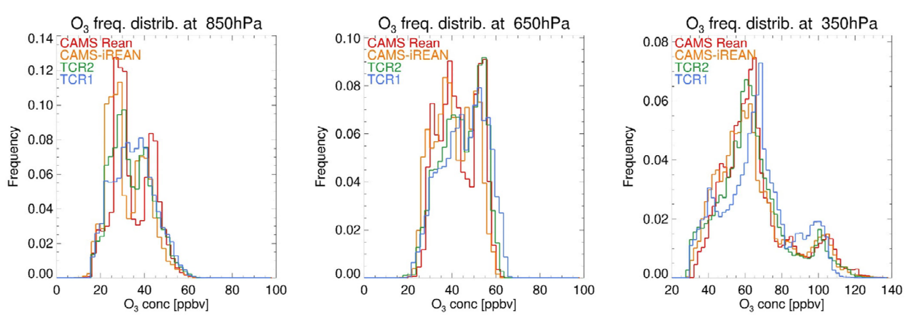


Figure 15: Area-normalized frequency distributions of ~~multi-annual~~multiannual (2005-2016) mean O_3 mixing ratios at ~~350~~ 850 (left), 650 (middle) and ~~8~~350 hPa (right) for the four reanalyses.

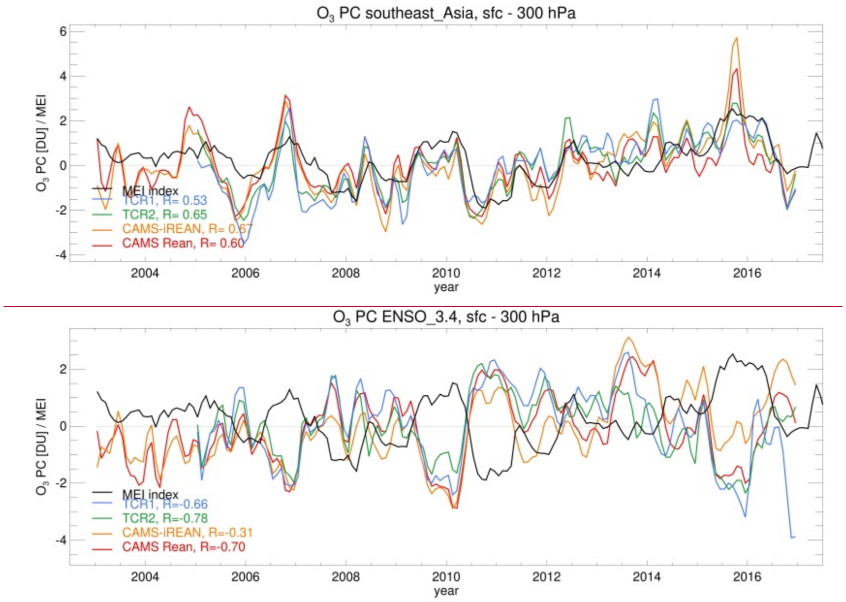


Figure 16: Deseasonalized anomalies in O_3 partial columns (surface to 300 hPa) in four reanalyses, as compared to the MEI index for two regions: Southeast Asia (90° E - 120° E; 10° S - 20° N) and ENSO3.4 over the Eastern Pacific (120W-170W; 5S-5N). A 2-month smoothing has been applied to the reanalysis data, (as for the MEI index). Temporal correlations with respect to the MEI index are given in the legends. Correlations are calculated on monthly data for the 2005-2016 time period.

~~8~~9. Conclusions and discussion

Four tropospheric ozone reanalyses have been compared in this paper, namely CAMS-iRean, CAMS-Rean, TCR-1, and TCR-2. A range of independent observations was used to validate the quality of the chemical reanalyses at various spatial and temporal scales. These reanalyses aim to capture individual large-scale events, such as heat waves or wildfires, and at the same time aim to provide a globally consistent climatology of present-day composition. This implies stringent requirements on their temporal consistency. The ~~changing constellation~~ changes in the observing system~~of satellite observations~~, combined with often their ~~often~~ limited sensitivity to tropospheric profiles and in particular the boundary layer, imply a significant dependency on the global chemistry model, its transport scheme, and its emissions, and makes the generation of any long-term chemical reanalysis challenging. This ~~calls~~ gives rise for a detailed evaluation of the capability of the current reanalyses of tropospheric ozone, as presented here.

Consistent with Inness et al., (2019), our evaluation also shows substantial improvement of CAMS-Rean over CAMS-iRean in the free troposphere, as quantified by lower ~~RMS~~ mean biases, ~~errors~~ standard deviations and higher correlations to ozone sonde observations, and better temporal consistency in ~~multi-annual~~multiannual time series of tropospheric ozone columns. For instance, averaged over the NH mid latitude region the mean bias in tropospheric ozone columns (surface to 300 hPa) is -0.3 DU (corresponding to approx. 1% of observed tropospheric column) for CAMS-Rean, which was 0.8 DU (3%) in CAMS-iRean.

At the surface the CAMS~~-REAN~~-Rean ~~is also~~ has generally ~~better than~~ improved with respect to CAMS-iRean~~CAMS-iREAN~~, assessed through evaluations of monthly mean surface concentrations against TOAR observations~~, .~~ ~~although~~

Nevertheless, similar performance of both CAMS reanalyses was seen for hourly to sub-seasonal variability assessed with EMEP observations over Europe for the year 2006~~, and in a few regions CAMS-iREAN showed a better diurnal cycle representation~~. The improved performance in the free troposphere can be attributed to a mixture of various upgrades, including revisions in the chemical data assimilation configuration, the chemistry mechanism, meteorological driver, model resolution, biogenic emissions.

Significant ~~temporal~~ changes in the quality of the ozone reanalys~~i~~es ~~in CAMS-iREAN across 2013~~ for different years have been attributed to changes over time in the observing system. Both CAMS reanalyses suffered from the use of relatively poor SCIAMACHY and MIPAS data products before 2005, which improved afterwards.~~,~~ ~~particularly~~ Also across 2013 in CAMS-iRean was affected by a switch of MLS version 2 to version 3.4. ~~In the CAMS system the MLS ozone profile measurements play a crucial role in constraining the partial column of ozone in the stratosphere. But as ozone total column observations are assimilated too, any changes in the MLS observations also affect the tropospheric ozone column in the CAMS reanalyses.~~ In both CAMS reanalyses a change to the vertical resolution of the assimilated SBUV/2 data during 2013 had a negative impact on the consistency of multiannual tropospheric ozone time series, particularly in polar regions. Inness et al. (2019) had noticed such a change in performance, but had not yet identified the responsible observational dataset.

Compared with TCR-1, TCR-2 shows better agreements with independent observations throughout the troposphere, including at the surface. Similar to the CAMS reanalyses, for the NH mid latitudes the mean bias in tropospheric columns against ozone sondes improved from 1.8 DU (7%) in TCR-1 to 0.8 DU (3%) in TCR-2. The improvements can be attributed to the use of more recent satellite retrievals and to an improved model performance, mainly associated with the increased model resolution. In spite of the good agreement with ozonesonde measurements in the free troposphere, the surface ozone reanalysis exhibits large positive biases over Europe and the United States. Also, the lack of the TES measurements led to a ~~degradation of~~ change in the reanalysis performance after 2010 for many regions in the lower and middle troposphere. ~~, while none of total column measurements of ozone was assimilated in the TCR systems. In the TCR reanalysis, the chemical concentrations and precursor's emissions were simultaneously optimized through EnKF data assimilation, which was important in providing information on precursors' emissions variations (Miyazaki et al., 2014; 2017; 2019a; Kiang et al., 2018) and in improving the vertical profiles of ozone.~~ Changes in the NO_2 observing system, including the OMI row anomaly after December 2009 and the limited temporal coverage of SCIAMACHY and GOME-2, are also considered to affect long-term consistency. The data assimilation diagnostics indicate the need for additional observational constraints, possibly combined with stronger inflation of the forecast error covariance, to improve the long-term reanalysis performance and to measure the actual analysis uncertainty.

Whereas free tropospheric ozone reanalyses agree well with independent observations, towards the surface larger biases have been found for many parts over the globe.~~this depends more on the model performance and emissions, and larger biases have been found in surface ozone analysis for many parts over the globe~~. A large spread at high latitudes could also be associated with the limited constraints from (tropospheric) ozone measurements. In these conditions the reanalyses depend more on the model performance and their emissions. Recently developed retrievals with high sensitivity to the lower troposphere (e.g. Deeter et al., 2013; Fu et al., 2018; Cuesta et al., 2018) would be helpful in improving the analysis of the lower troposphere. ~~Meanwhile,~~ Furthermore, in future studies the analysis ensemble spread from EnKF can be regarded as ~~the~~ uncertainty information about the analysis mean fields, indicating the need for additional observational constraints.~~, whereas~~ Likewise, in the 4-D Var system ~~could be used to test~~ the contributions from individual retrieval products can be tested.

We have demonstrated that the recent chemical reanalyses of CAMS~~-REAN~~-Rean and TCR-2 agree well with each other and with the independent observations in the majority of cases. This highlights the usefulness of the current chemical reanalyses in a variety of studies. For instance, the well-characterized, small mean bias in tropospheric columns in these reanalyses suggest that they can be used to provide a climatology of present-day tropospheric ozone. This may serve as a reference for the present-day contribution of tropospheric ozone to the radiation budget, or may provide a climatology for a-priori ozone profiles as required for satellite retrieval products (e.g., Fu et al., 2018). The ability of the CAMS Reanalysis to capture the variability of (near-)surface ozone on multiple time scales, and for many regions over the globe, indicates it is fit for use as boundary conditions for hindcasts of regional air quality models.

Meanwhile, our intercomparisons suggest that the model ~~performance~~ configuration can still ~~lead~~ explain differences ~~to discrepancies~~ in the ozone reanalys~~e~~is ~~quality among the systems~~. For instance, differences in the representation of convective transport over the continents and those in the precursor's emissions, as well as differences in the chemical scheme, lead to substantial differences in the vertical profile of ozone and ozone production, such as over Central Africa and South America. Here the standard deviation in annual mean ozone at 850 hPa reaches up to 50% of the multi-reanalysis mean~~as discussed in Miyazaki et al. (2019b)~~. The relatively coarse horizontal resolution ~~of~~ in any of the global reanalysis configurations ~~global models~~ could also cause significant ~~model~~ errors at urban sites. ~~A coarse vertical resolution additionally has larger impacts on the quality of tropospheric ozone around the UTLS. Thus, although the reanalysis dataset provides comprehensive information about interannual variability in tropospheric ozone,~~ Therefore both the data assimilation settings and the model performance are critical in improving the tropospheric ozone analysis and obtaining consistent data assimilation analysis, especially for the lower troposphere.

We have shown that ~~D~~discontinuities in the availability, ~~and~~ coverage and product version of the assimilated measurements ~~are also shown to~~ affect the quality of any of the ~~reanalysis~~reanalyses, particularly in terms of temporal consistency~~, both in the CAMS and TCR-reanalyses~~. This is particularly important for assessing interannual variability.~~, and the usability of such reanalysis products for model evaluation.~~ The influence of data discontinuities must be considered and where possible removed when studying interannual variability and trends using products from these reanalyses. To improve the temporal consistency in future reanalyses, a careful assessment of changes in the assimilation configuration, most prominently associated with ozone column and profile assimilation is needed, including a detailed assessment of biases between various retrieval products.

The assimilation of multi-species data in both the CAMS and TCR configurations influences the representation of the entire chemical system, while the influence of persistent model errors in complex tropospheric chemistry continues to be a concern. ~~Also changes and biases in assimilation of precursor trace gases, such as NO_2, could influence temporal consistency in reanalyses of tropospheric ozone. Validation of various trace gases from the chemical reanalysis products can be used to better identify potential sources of error in the reanalysis ozone fields. Furthermore,~~ Therefore, further improvements to long-term reanalyses of tropospheric ozone can be achieved by ~~increasing~~ improving the observational constraints, together with ~~the~~ a further optimization of model parameters, such as the chemical mechanism, emission, deposition, and mixing processes.~~, could lead to more consistent data assimilation fields, hence further improving long-term reanalyses.~~

Data availability

The CAMS reanalyses data are freely available from https://atmosphere.copernicus.eu/ (last access: 18 October 2019).
The TCR-1 reanalysis is available from https://ebcrpa.jamstec.go.jp/~miyazaki/tcr/ , the TCR-2 reanalysis is available from https://tes.jpl.nasa.gov/chemical-reanalysis/ ~~https://ebcrpa.jamstec.go.jp/tcr2/index.html~~ .

Met opmaak: Engels (Verenigd Koninkrijk)

Author contributions

VH and KM designed the study and wrote large parts of the manuscript. VH performed the evaluations and analyses. JF and AI provided the CAMS-Reanalysis data, KM and TS provided the TCR-Reanalysis data. MGS provided the TOAR data, and contributed to its interpretation. All co-authors contributed to the writing and the analyses.

Acknowledgements

We acknowledge the use of data products from the NASA AURA, EOS Terra, and Aqua satellite missions. Part of the research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. We also acknowledge the free use of tropospheric NO_2 column data from the SCIAMACHY, GOME-2, and OMI sensors from http://www.qa4ecv.eu and. This work was supported through JSPS KAKENHI grant number 18H01285 and the Environment Research and Technology Development Fund (2-1803) of the Ministry of the Environment, Japan, and by the Post-K computer project Priority Issue 4 - Advancement of meteorological and global environmental predictions utilizing observational Big Data. The Earth Simulator was used for simulations as "Strategic Project with Special Support" of Japan Agency Marine-Earth Science and Technology. Part of this work was performed at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. The Copernicus Atmosphere Monitoring Service is operated by the European Centre for Medium- Range Weather Forecasts on behalf of the European Commission as part of the Copernicus Programme (http://copernicus.eu, last access: 15 March 2019). We thank all research and agency teams who provided data to WOUDCC, SHADOZ, TOAR, and EMEP.

References

Bennouna, Y., M. Schulz, Y. Christophe, H.J. Eskes, S. Basart, A. Benedictow, A.-M. Blechschmidt, S. Chabrillat, H. Clark, E. Cuevas, H. Flentje, K.M. Hansen, U. Im, J. Kapsomenakis, B. Langerock, K. Petersen, A. Richter, N. Sudarchikova, V. Thouret, A. Wagner, Y. Wang, C. Zerefos, Validation report of the CAMS global Reanalysis of aerosols and reactive gases, years 2003-2017, Copernicus Atmosphere Monitoring Service (CAMS) report, CAMS84_2018SC1_D5.1.1-2017_v1.pdf, 2019, doi:10.24380/xjhk-zt69.

Bhartia, P. K., McPeters, R. D., Mateer, C. L., Flynn, L. E., and Wellemeyer, C., Algorithm for the estimation of vertical ozone profiles from the backscattered ultraviolet technique, J. Geophys. Res., 101(D13), 18,793– 18,806, 1996.

Bocquet, M., Elbern, H., Eskes, H., Hirtl, M., Žabkar, R., Carmichael, G. R., Flemming, J., Inness, A., Pagowski, M., Pérez Camaño, J. L., Saide, P. E., San Jose, R., Sofiev, M., Vira, J., Baklanov, A., Carnevale, C., Grell, G., and Seigneur, C.: Data assimilation in atmospheric chemistry models: current status and future prospects for coupled chemistry meteorology models, Atmos. Chem. Phys., 15, 5325-5358, https://doi.org/10.5194/acp-15-5325-2015, 2015.

Boersma, K. F., Eskes, H. J., Dirksen, R. J., van der A, R. J., Veefkind, J. P., Stammes, P., Huijnen, V., Kleipool, Q. L., Sneep, M., Claas, J., Leitão, J., Richter, A., Zhou, Y., and Brunner, D.: An improved tropospheric NO2 column retrieval algorithm for the Ozone Monitoring Instrument, Atmos. Meas. Tech., 4, 1905-1928, https://doi.org/10.5194/amt-4-1905-2011, 2011.

Boersma, K. F., Eskes, H., Richter, A., De Smedt, I., Lorente, A., Beirle, S., Van Geffen, J., Peters, E., Van Roozendael, M. & Wagner, T., QA4ECV NO_2 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, 2017a.

Boersma, K. F., Eskes, H., Richter, A., De Smedt, I., Lorente, A., Beirle, S., Van Geffen, J., Peters, E., Van Roozendael, M. & Wagner, T., QA4ECV NO_2 tropospheric and stratospheric vertical column data from GOME-2A (Version 1.1) [Data set]. Royal Netherlands Meteorological Institute (KNMI). http://doi.org/10.21944/qa4ecv-no2-gome2a-v1.1, 2017b.

Boersma, K. F., J. van Geffen, H. Eskes, R. van der A, I. De Smedt, M. Van Roozendael, H. Yu, A. Richter, E. Peters, S. Beirle, T. Wagner, A. Lorente, T. Scanlon, S. Compernolle, J.-C. Lambert, Product Specification Document for the QA4ECV NO2 ECV precursor product, Royal Netherlands Meteorological Institute (KNMI), available at http://temis.nl/qa4ecv/no2col/QA4ECV_NO2_PSD_v1.1.compressed.pdf, 2017c

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, https://doi.org/10.5194/acp-7-2183-2007, 2007.

Checa-Garcia, R., Hegglin, M. I., Kinnison, D., Plummer, D. A., & Shine, K. P. (2018). Historical tropospheric and stratospheric ozone radiative forcing using the CMIP6 database. Geophysical Research Letters, 45, pp 3264-3273. https://doi.org/10.1002/2017GL076770

Cheung, J. C. H., Haigh, J. D., and Jackson, D. R. (2014), Impact of EOS MLS ozone data on medium-extended range ensemble weather forecasts, J. Geophys. Res. Atmos., 119, 9253– 9266, doi:10.1002/2014JD021823.

Cooper, OR, Parrish, DD, Ziemke, J, Balashov, NV, Cupeiro, M, Galbally, IE, Gilge, S, Horowitz, L, Jensen, NR, Lamarque, J-F, Naik, V, Oltmans, SJ, Schwab, J, Shindell, DT, Thompson, AM, Thouret, V, Wang, Y and Zbinden, RM. 2014. Global distribution and trends of tropospheric ozone: An observation-based review. Elementa: Science of the Anthropocene 2. DOI: 10.12952/journal.elementa.000029

Cuesta, J., Kanaya, Y., Takigawa, M., Dufour, G., Eremenko, M., Foret, G., Miyazaki, K., and Beekmann, M.: Transboundary ozone pollution across East Asia: daily evolution and photochemical production analysed by IASI + GOME2 multispectral satellite observations and models, Atmos. Chem. Phys., 18, 9499-9525, https://doi.org/10.5194/acp-18-9499-2018, 2018.

Davis, S. M., Hegglin, M. I., Fujiwara, M., Dragani, R., Harada, Y., Kobayashi, C., Long, C., Manney, G. L., Nash, E. R., Potter, G. L., Tegtmeier, S., Wang, T., Wargan, K., and Wright, J. S.: Assessment of upper tropospheric and stratospheric water vapor and ozone in reanalyses as part of S-RIP, Atmos. Chem. Phys., 17, 12743-12778, https://doi.org/10.5194/acp-17-12743-2017, 2017.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N. and Vitart, F., The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q.J.R. Meteorol. Soc.*, 137: 553–597. doi:10.1002/qj.828, 2011.

Deeter, M. N., Martínez-Alonso, S., Edwards, D. P., Emmons, L. K., Gille, J. C., Worden, H. M., Pittman, J. V., Daube, B. C., and Wofsy, S. C.: Validation of MOPITT Version 5 thermal-infrared, near-infrared, and multispectral carbon monoxide profile retrievals for 2000–2011, *J. Geophys. Res. Atmos.*, 118, 6710– 6725, doi:10.1002/jgrd.50272, 2013

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.

Ding, J., Miyazaki, K., van der A, R. J., Mijling, B., Kurokawa, J.-I., Cho, S., Janssens-Maenhout, G., Zhang, Q., Liu, F., and Levelt, P. F.: Intercomparison of NOx emission inventories over East Asia, Atmos. Chem. Phys., 17, 10125-10141, https://doi.org/10.5194/acp-17-10125-2017, 2017.

Dragani, R., Benedetti, A., Flemming, J., Balsamo, G., Diamantakis, M., Geer, A.J., Hogan, R., Stockdale, T., Ades, M., Agusti-Panareda, A., Barré, J., Bechtold, P., Bozzo, A., Hersbach, H., Hólm, E., Kipling, Z., Inness, A., Letertre-Danczak, J., Massart, S., Matricardi, M., McNally, T., Parrington, M., Sandu, I., Soci, C., Vitart, F.: Atmospheric Composition priority developments for Numerical Weather Prediction. ECMWF Technical memorandum 833, doi:10.21957/5e0whui2y , 2018.

Errera, Q., Chabrillat, S., Christophe, Y., Debosscher, J., Hubert, D., Lahoz, W., Santee, M. L., Shiotani, M., Skachko, S., von Clarmann, T., and Walker, K.: Technical note: Reanalysis of Aura MLS Chemical Observations, Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2019-530, in review, 2019.

Fleming, Z.L., Doherty, R.M., von Schneidemesser, E., Malley, C.S., Cooper, O.R., Pinto, J.P., Colette, A., Xu, X., Simpson, D., Schultz, M.G., Lefohn, A.S., Hamad, S., Moolla, R., Solberg, S. and Feng, Z.: Tropospheric Ozone Assessment Report: Present-day ozone distribution and trends relevant to human health. *Elem Sci Anth*, 6(1), p.12. DOI: http://doi.org/10.1525/elementa.273, 2018.

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 ECMWF technical memorandum 590, doi: 10.21957/z1hxhdhgo, 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, https://doi.org/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.

Flemming, J. and Inness, A.: Carbon Monoxide [in "State of the Climate in 2017"], B. Am. Meteorol. Soc., 99, S59–S61, 2018.

Fröhlich, C. and Lean, J., Solar radiative output and its variability: evidence and mechanisms. The Astron. Astrophys. Rev. 12: 273. https://doi.org/10.1007/s00159-004-0024-1, 2004.

Fu, D., Kulawik, S. S., Miyazaki, K., Bowman, K. W., Worden, J. R., Eldering, A., Livesey, N. J., Teixeira, J., Irion, F. W., Herman, R. L., Osterman, G. B., Liu, X., Levelt, P. F., Thompson, A. M., and Luo, M.: Retrievals of tropospheric ozone profiles from the synergism of AIRS and OMI: methodology and validation, Atmos. Meas. Tech., 11, 5587-5605, https://doi.org/10.5194/amt-11-5587-2018, 2018.

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

Graedel, T. E., Bates, T. S., Bouwman, A. F., Cunnold, D., Dignon, J., Fung, I., Jacob, D. J., Lamb, B. K., Logan, J. A., Marland, G., Middleton, P., Pacyna, J. M., Placet, M., & Veldt, C., A compilation of inventories of emissions to the atmosphere, Global Biogeochem. Cy., 7, 1–26, 1993.

Hao, N., Koukouli, M. E., Inness, A., Valks, P., Loyola, D. G., Zimmer, W., Balis, D. S., Zyrichidou, I., Van Roozendael, M., Lerot, C., and Spurr, R. J. D.: GOME-2 total ozone columns from MetOp-A/MetOp-B and assimilation in the MACC system, Atmos. Meas. Tech., 7, 2937-2951, https://doi.org/10.5194/amt-7-2937-2014, 2014.

Hersbach, H., de Rosnay, P., Bell, B., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Alonso-Balmaseda, M., Balsamo, G., Bechtold, P., Berrisford, P., Bidlot, J.-R., de Boisséson, E., Bonavita, M., Browne, P., Buizza, R., Dahlgren, P., Dee, D., Dragani, R., Diamantakis, M., Flemming, J., Forbes, R., Geer, A. J., Haiden, T., Hólm, E., Haimberger, L., Hogan, R., Horányi, A., Janiskova, M., Laloyaux, P., Lopez, P., Munoz-Sabater, J., Peubey, C., Radu, R., Richardson, D., Thépaut, J.-N., Vitart, F., Yang, X., Zsótér, E., and Zuo, H.: Operational global reanalysis: progress, future directions and synergies with NWP, ERA Report Series, Number 27, available at: https://www.ecmwf.int/node/18765 (last access: 15 March 2019), 2018.

Hsu, J. and Prather, M. J.: Stratospheric variability and tropospheric ozone, J. Geophys. Res.-Atmos., 114, D06102, doi:10.1029/2008jd010942, 2009.

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.

Hunt, B. R., Kostelich, E. J., & Szunyogh, I., Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter, *Physica D*, 230, 112–126, 2007.

Huijnen, V., Flemming, J., Kaiser, J. W., Inness, A., Leitão, J., Heil, A., Eskes, H. J., Schultz, M. G., Benedetti, A., Hadji-Lazaro, J., Dufour, G., and Eremenko, M.: Hindcast experiments of tropospheric composition during the summer 2010 fires over western Russia, Atmos. Chem. Phys., 12, 4341-4364, https://doi.org/10.5194/acp-12-4341-2012, 2012.

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, 2016.

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., Baier, F., Benedetti, A., Bouarar, I., Chabrillat, S., Clark, H., Clerbaux, C., Coheur, P., Engelen, R. J., Errera, Q., Flemming, J., George, M., Granier, C., Hadji-Lazaro, J., Huijnen, V., Hurtmans, D., Jones, L., Kaiser, J. W., Kapsomenakis, J., Lefever, K., Leitão, J., Razinger, M., Richter, A., Schultz, M. G., Simmons, A. J., Suttie, M., Stein, O., Thépaut, J.-N., Thouret, V., Vrekoussis, M., Zerefos, C., and the MACC team: The MACC reanalysis: an 8 yr data set of atmospheric composition, Atmos. Chem. Phys., 13, 4073-4109, https://doi.org/10.5194/acp-13-4073-2013, 2013.

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, M., Agustí-Panareda, A., Barré, J., Benedictow, A., Blechschmidt, A.-M., Dominguez, J. J., Engelen, R., Eskes, H., Flemming, J., Huijnen, V., Jones, L., Kipling, Z., Massart, S., Parrington, M., Peuch, V.-H., Razinger, M., Remy, S., Schulz, M., and Suttie, M.: The CAMS reanalysis of atmospheric composition, Atmos. Chem. Phys., 19, 3515-3556, https://doi.org/10.5194/acp-19-3515-2019, 2019.

Jiang, Z., B. C. McDonald, H. Worden, J. R. Worden, K. Miyazaki, Z. Qu, D. K. Henze, D. B. A. Jones, A. F. Arellano, E. V. Fischer, L. Zhu, and K. F. Boersma, Unexpected slowdown of US pollutant emission reduction in the last decade, Proceedings of the National Academy of Sciences, 201801191; DOI: 10.1073/pnas.1801191115, 2018.

Jonson, J. E., Schulz, M., Emmons, L., Flemming, J., Henze, D., Sudo, K., Tronstad Lund, M., Lin, M., Benedictow, A., Koffi, B., Dentener, F., Keating, T., Kivi, R., and Davila, Y.: The effects of intercontinental emission sources on European air pollution levels, Atmos. Chem. Phys., 18, 13655-13672, https://doi.org/10.5194/acp-18-13655-2018, 2018.

Kanaya, Y., Miyazaki, K., Taketani, F., Miyakawa, T., Takashima, H., Komazaki, Y., Pan, X., Kato, S., Sudo, K., Inoue, J., Sato, K., and Oshima, K.: Ozone and carbon monoxide observations over open oceans on R/V Mirai from 67° S to 75° N during 2012 to 2017: Testing global chemical reanalysis in terms of Arctic processes, low ozone levels at low latitudes, and pollution transport, Atmos. Chem. Phys., https://doi.org/10.5194/acp-2018-1354, in press, 2019.

Katragkou, E., Zanis, P., Tsikerdekis, A., Kapsomenakis, J., Melas, D., Eskes, H., Flemming, J., Huijnen, V., Inness, A., Schultz, M. G., Stein, O., and Zerefos, C. S.: Evaluation of near-surface ozone over Europe from the MACC reanalysis, Geosci. Model Dev., 8, 2299-2314, https://doi.org/10.5194/gmd-8-2299-2015, 2015.

Knowland, K., Ott, L., Duncan, B., and Wargan, K.: Stratospheric Intrusion-Influenced Ozone Air Quality Exceedances Investigated in the NASA MERRA-2 Reanalysis, Geophys. Res. Lett., 44, 10691–10701, https://doi.org/10.1002/2017GL074532, 2017.

Lahoz, W. A. and Schneider, P.: Data assimilation: making sense of Earth Observation, *Front. Environ. Sci.*, 2, 16, doi:10.3389/fenvs.2014.00016, 2014.

Lerot, C., M. Van Roozendael, J. van Geffen, C. Fayt, R. Spurr, G. Lichtenberg, and A. von Bargen (2009), Six years of total ozone column measurements from SCIAMACHY nadir observations, Atmos. Meas. Tech., doi:10.5194/amt-2-87-2009.

Liang, C.-K., West, J. J., Silva, R. A., Bian, H., Chin, M., Davila, Y., Dentener, F. J., Emmons, L., Flemming, J., Folberth, G., Henze, D., Im, U., Jonson, J. E., Keating, T. J., Kucsera, T., Lenzen, A., Lin, M., Lund, M. T., Pan, X., Park, R. J., Pierce, R. B., Sekiya, T., Sudo, K., and Takemura, T.: HTAP2 multi-model estimates of premature human mortality due to intercontinental transport of air pollution and emission sectors, Atmos. Chem. Phys., 18, 10497-10520, https://doi.org/10.5194/acp-18-10497-2018, 2018.

Liu, X., Bhartia, P. K., Chance, K., Froidevaux, L., Spurr, R. J. D., and Kurosu, T. P.: Validation of Ozone Monitoring Instrument (OMI) ozone profiles and stratospheric ozone columns with Microwave Limb Sounder (MLS) measurements, Atmos. Chem. Phys., 10, 2539-2549, https://doi.org/10.5194/acp-10-2539-2010, 2010.

Livesey, N. J., Read, W.G., Wagner, P. A., Froidevaux, L. , Lambert, A., Manney, G. L., Millán Valle, L. F., Pumphrey, H. C., Santee, M. L., Schwartz, M. J., Wang, S., Fuller, R. A., Jarnot, R. F., Knosp, B. W., Martinez, E., Lay, R. R.: Aura Microwave Limb Sounder (MLS) Version 4.2x Level2 data quality and description document, Tech. Rep. JPL D33509 Rev. D, Jet Propul. Lab., Pasadena, CA, 2018.

~~Livesey, N. J., Read, W. G., Froidevaux, L., Lambert, A., Manney, G. L., Pumphrey, H. C., Santee, M. L., Schwartz, M. J., Wang, S., Cofield, R. E., Cuddy, D. T., Fuller, R. A., Jarnot, R. F., Jiang, J. H., Knosp, B. W., Stek, P. C., Wagner, P. A., and Wu, D. L.: Aura Microwave Limb Sounder (MLS), Version 3.3 Level 2 data quality and description document, Tech. Rep. JPL D-33509, Jet Propul. Lab., Pasadena, CA, 2011.~~

McPeters, R. D., P. K. Bhartia, D. Haffner, G. J. Labow, and L. Flynn: The version 8.6 SBUV ozone data record: An overview, J. Geophys. Res. Atmos., 118, 8032–8039, doi:10.1002/jgrd.50597, 2013.

Ménard, R. and Chang, L.: Assimilation of Stratospheric Chemical Tracer Observations Using a Kalman Filter. Part II: χ^2-Validated Results and Analysis of Variance and Correlation Dynamics. Mon. Wea. Rev.,1282672–2686, https://doi.org/10.1175/1520-0493(2000)128<2672:AOSCTO>2.0.CO;2

Miyazaki, K., Eskes, H. J., & Sudo, K., Global NOx emission estimates derived from an assimilation of OMI tropospheric NO_2 columns, *Atmospheric Chemistry and Physics*, 12, 2263-2288, doi:10.5194/acp-12-2263-2012, 2012a.

Miyazaki, K., Eskes, H. J., Sudo, K., Takigawa, M., van Weele, M., & Boersma, K. F., Simultaneous assimilation of satellite NO_2, O_3, CO, and HNO_3 data for the analysis of tropospheric chemical composition and emissions, *Atmospheric Chemistry and Physics*, 12, 9545-9579, doi:10.5194/acp-12-9545-2012, 2012b.

Miyazaki, K., H. Eskes, K. Sudo, & C. Zhang, Global lightning NOx production estimated by an assimilation of multiple satellite datasets, *Atmospheric Chemistry and Physics*, 14, 3277-3305, 2014.

Miyazaki, K., Eskes, H. J., & Sudo, K., A tropospheric chemistry reanalysis for the years 2005–2012 based on an assimilation of OMI, MLS, TES, and MOPITT satellite data, *Atmos. Chem. Phys.*, 15, 8315-8348, doi:10.5194/acp-15-8315-2015, 2015.

Miyazaki, K. & Bowman, K., Evaluation of ACCMIP ozone simulations and ozonesonde sampling biases using a satellite-based multi-constituent chemical reanalysis, *Atmos. Chem. Phys.*, 17, 8285-8312, https://doi.org/10.5194/acp-17-8285-2017, 2017a.

Miyazaki, K., Eskes, H., Sudo, K., Boersma, K. F., Bowman, K., & Kanaya, Y. (2017), Decadal changes in global surface NOx emissions from multi-constituent satellite data assimilation, *Atmos. Chem. Phys.*, 17, 807-837, doi:10.5194/acp-17-807-2017b.

Miyazaki. K., T. Sekiya, D. Fu, K. W. Bowman, S. S. Kulawik, K. Sudo, T. Walker, Y. Kanaya, M. Takigawa, K. Ogochi, H. Eskes, K. F. Boersma, A. M. Thompson, B. Gaubert, J. Barre, L. K. Emmons: Balance of emission and dynamical controls on ozone during KORUS-AQ from multi-constituent satellite data assimilation, *Journal of Geophysical Research: Atmospheres*, 124, 387–413. https://doi.org/10.1029/2018JD028912, 2019a.

Miyazaki, K., Bowman, K. W., Yumimoto, K., Walker, T., and Sudo, K.: Evaluation of a multi-model, multi-constituent assimilation framework for tropospheric chemical reanalysis, Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2019-645, in review, 2019.

Monks, P. S., Archibald, A. T., Colette, A., Cooper, O., Coyle, M., Derwent, R., Fowler, D., Granier, C., Law, K. S., Mills, G. E., Stevenson, D. S., Tarasova, O., Thouret, V., von Schneidemesser, E., Sommariva, R., Wild, O., and Williams, M. L.: Tropospheric ozone and its precursors from the urban to the global scale from air quality to short-lived climate forcer, Atmos. Chem. Phys., 15, 8889–8973, https://doi.org/10.5194/acp-15-8889-2015, 2015.

Morgenstern, O., Stone, K. A., Schofield, R., Akiyoshi, H., Yamashita, Y., Kinnison, D. E., Garcia, R. R., Sudo, K., Plummer, D. A., Scinocca, J., Oman, L. D., Manyin, M. E., Zeng, G., Rozanov, E., Stenke, A., Revell, L. E., Pitari, G., Mancini, E., Di Genova, G., Visioni, D., Dhomse, S. S., and Chipperfield, M. P.: Ozone sensitivity to varying greenhouse gases and ozone-depleting substances in CCMI-1 simulations, Atmos. Chem. Phys., 18, 1091-1114, https://doi.org/10.5194/acp-18-1091-2018, 2018.

Munro, R., R. Siddans, W. J. Reburn, and B. J. Kerridge, Direct measurements of tropospheric ozone distributions from space, Nature, 392, 168–171, 1998.

Nédélec, P, Blot, R, Boulanger, D, Athier, G, Cousin, J, Gautron, B, Petzold, A, Volz-Thomas, A and Thouret, V. 2015. Instrumentation on commercial aircraft for monitoring the atmospheric composition on a global scale: the IAGOS system, technical overview of ozone and carbon monoxide measurements. Tellus B 67. DOI: 10.3402/tellusb.v67.27791

Neu, J. L., Flury, T., Manney, G. L., Santee, M. L., Livesey, N. J., Worden, J.: Tropospheric ozone variations governed by changes in stratospheric circulation. Nature Geoscience 7, pp 340–344 (2014). Doi: 10.1038/ngeo2138

Ordóñez, C., Elguindi, N., Stein, O., Huijnen, V., Flemming, J., Inness, A., Flentje, H., Katragkou, E., Moinat, P., Peuch, V.-H., Segers, A., Thouret, V., Athier, G., van Weele, M., Zerefos, C. S., Cammas, J.-P., and Schultz, M. G.: Global model simulations of air pollution during the 2003 European heat wave, Atmos. Chem. Phys., 10, 789-815, https://doi.org/10.5194/acp-10-789-2010, 2010.

Parrish, DD, Lamarque, J-F, Naik, V, Horowitz, L, Shindell, DT, et al.. Long-term changes in lower tropospheric baseline ozone concentrations: Comparing chemistry-climate models and observations at northern mid-latitudes. J. Geophys. Res . 119(9): 5719–5736. DOI: 10.1002/2013JD021435, 2014.

Schultz M. G., Schröder, S., Lyapina, O., Cooper, O., Galbally, I., Petropavlovskikh, I., et al.: Tropospheric Ozone Assessment Report: Database and Metrics Data of Global Surface Ozone Observations, Elem. Sci. Anth., 5, 58, https://doi.org/10.1525/elementa.244, 2017.

Schultz, M. G., Schröder, S., Lyapina, O., Cooper, O. R., Galbally, I., Petropavlovskikh, I., Schneidemesser, E. V., Tanimoto, H., Elshorbany, Y., Naja, M., Seguel, R. J., Dauert, U., Eckhardt, P., Feigenspan, S., Fiebig, M., Hjellbrekke, A.-G., Hong, Y.-D., Kjeld, P. C., Koide, H., Lear, G., Tarasick, D., Ueno, M., Wallasch, M., Baumgardner, D., Chuang, M.-T., Gillett, R., Lee, M., Molloy, S., Moolla, R., Wang, T., Sharps, K., Adame, J. A., Ancellet, G., Apadula, F., Artaxo, P., Barlasina, M. E., Bogucka, M., Bonasoni, P., Chang, L., Colomb, A., Cuevas-Agulló, E., Cupeiro, M., Degorska, A., Ding, A., Fröhlich, M., Frolova, M., Gadhavi, H., Gheusi, F., Gilge, S., Gonzalez, M. Y., Gros, V., Hamad, S. H., Helmig, D., Henriques, D., Hermansen, O., Holla, R., Hueber, J., Im, U., Jaffe, D. A., Komala, N., Kubistin, D., Lam, K.-S., Laurila, T., Lee, H., Levy, I., Mazzoleni, C., Mazzoleni, L. R., McClure-Begley, A., Mohamad, M., Murovec, M., Navarro-Comas, M., Nicodim, F., Parrish, D., Read, K. A., Reid, N., Ries, L., Saxena, P., Schwab, J. J., Scorgie, Y., Senik, I., Simmonds, P., Sinha, V., Skorokhod, A. I., Spain, G., Spangl, W., Spoor, R., Springston, S. R., Steer, K., Steinbacher, M., Suharguniyawan, E., Torre, P., Trickl, T., Weili, L., Weller, R., Xiaobin, X., Xue, L., and Zhiqiang, M.: Tropospheric Ozone Assessment Report, links to Global surface ozone datasets, PANGAEA, https://doi.org/10.1594/PANGAEA.876108, 2017.

Schwartz, M., Froidevaux, L., Livesey, N. and Read, W. (2015), MLS/Aura Level 2 Ozone (O3) Mixing Ratio V004, Greenbelt, 25 MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed: [20180718], 10.5067/Aura/MLS/DATA2017, 2015.

Sekiya, T., Miyazaki, K., Ogochi, K., Sudo, K., and Takigawa, M.: Global high-resolution simulations of tropospheric nitrogen dioxide using CHASER V4.0, Geosci. Model Dev., 11, 959-988, https://doi.org/10.5194/gmd-11-959-2018, 2018.

Sudo, K., Takahashi, M., and Akimoto, H.: CHASER: a global chemical model of the troposphere 2. Model results and evaluation, J. Geophys. Res., 107, 4586, doi:10.1029/2001JD001114, 2002.

Struzewska, J. and Kaminski, J. W.: Formation and transport of photooxidants over Europe during the July 2006 heat wave – observations and GEM-AQ model simulations, Atmos. Chem. Phys., 8, 721-736, https://doi.org/10.5194/acp-8-721-2008, 2008.

Tang, W., Arellano, A. F., Gaubert, B., Miyazaki, K., and Worden, H. M.: Satellite data reveal a common combustion emission pathway for major cities in China, Atmos. Chem. Phys., 19, 4269-4288, https://doi.org/10.5194/acp-19-4269-2019, 2019.

Thompson, A.M, R. M. Stauffer, T. P. Boyle, D. E. Kollonige, K. Miyazaki, M. Tzortziou, J. R. Herman, C. N. Jordan, B. T. Lamb, B. N. Duncan, Comparison of Near-surface NO_2 Pollution with Pandora Total Column NO_2 during the Korea-United States Ocean Color (KORUS OC) Campaign, , J. Geophys. Res., in review, 2019.

Thouret, V., Cammas, J.-P., Sauvage, B., Athier, G., Zbinden, R., Nédélec, P., Simon, P., and Karcher, F.: Tropopause referenced ozone climatology and inter-annual variability (1994–2003) from the MOZAIC programme, Atmos. Chem. Phys., 6, 1033-1051, https://doi.org/10.5194/acp-6-1033-2006, 2006.

Tilmes, S., Lamarque, J.-F., Emmons, L. K., Conley, A., Schultz, M. G., Saunois, M., Thouret, V., Thompson, A. M., Oltmans, S. J., Johnson, B., and Tarasick, D.: Technical Note: Ozonesonde climatology between 1995 and 2011: description, evaluation and applications, Atmos. Chem. Phys., 12, 7475–7497, https://doi.org/10.5194/acp-12-7475-2012, 2012.

Travis, K. R., Jacob, D. J., Fisher, J. A., Kim, P. S., Marais, E. A., Zhu, L., Yu, K., Miller, C. C., Yantosca, R. M., Sulprizio, M. P., Thompson, A. M., Wennberg, P. O., Crounse, J. D., St. Clair, J. M., Cohen, R. C., Laughner, J. L., Dibb, J. E., Hall, S. R., Ullmann, K., Wolfe, G. M., Pollack, I. B., Peischl, J., Neuman, J. A., and Zhou, X.: Why do models overestimate surface ozone in the Southeast United States?, Atmos. Chem. Phys., 16, 13561-13577, https://doi.org/10.5194/acp-16-13561-2016, 2016.

van der A, R. J., Mijling, B., Ding, J., Koukouli, M. E., Liu, F., Li, Q., Mao, H., and Theys, N.: Cleaning up the air: effectiveness of air quality policy for SO2 and NOx emissions in China, Atmos. Chem. Phys., 17, 1775–1789, https://doi.org/10.5194/acp-17-1775-2017, 2017.

Van Dingenen, R, Dentener, FJ, Raes, F, Krol, MC, Emberson, L & Cofala, J: 'The global impact of ozone on agricultural crop yields under current and future air quality legislation', Atmospheric Environment, vol. 43, no. 3, pp. 604-618. https://doi.org/10.1016/j.atmosenv.2008.10.033, 2009.

Verstraeten, W. W., Neu, J. L., Williams, J. E., Bowman, K. W., Worden, J. R., and Boersma, K. F.: Rapid increases in tropospheric ozone production and export from China, Nat. Geosci., 8, 690–695, doi:10.1038/ngeo2493, 2015.

von Clarmann, T., et al., Retrieval of temperature and tangent altitude pointing from limb emission spectra recorded from space by the Michelson Interferometer for Passive Atmospheric Sounding (MIPAS), J. Geophys. Res., 108(D23), 4736, doi:10.1029/2003JD003602, 2003.

von Clarmann, T., Höpfner, M., Kellmann, S., Linden, A., Chauhan, S., Funke, B., Grabowski, U., Glatthor, N., Kiefer, M., Schieferdecker, T., Stiller, G. P., and Versick, S.: Retrieval of temperature, H2O, O3, HNO3, CH4, N2O, ClONO2 and ClO from MIPAS reduced resolution nominal mode limb emission measurements, Atmos. Meas. Tech., 2, 159–175, doi:10.5194/amt-2-159- 2009, 2009.

Watanabe, S., Hajima, T., Sudo, K., Nagashima, T., Takemura, T., Okajima, H., Nozawa, T., Kawase, H., Abe, M., Yokohata, T., Ise, T., Sato, H., Kato, E., Takata, K., Emori, S., & Kawamiya, M., MIROC-ESM 2010: model description and basic results of CMIP5-20c3m experiments, *Geosci. Model Dev*., 4, 845–872, https://doi.org/10.5194/gmd-4-845-2011, 2011.

Williams, R. S., Hegglin, M. I., Kerridge, B. J., Jöckel, P., Latter, B. G., and Plummer, D. A.: Characterising the seasonal and geographical variability in tropospheric ozone, stratospheric influence and recent changes, Atmos. Chem. Phys., 19, 3589-3620, https://doi.org/10.5194/acp-19-3589-2019, 2019.

Wolter, K., and M. S. Timlin: Measuring the strength of ENSO events - how does 1997/98 rank? Weather, 53, 315-324, 1998.

Young, P. J., Archibald, A. T., Bowman, K. W., Lamarque, J.-F., Naik, V., Stevenson, D. S., Tilmes, S., Voulgarakis, A., Wild, O., Bergmann, D., Cameron-Smith, P., Cionni, I., Collins, W. J., Dalsøren, S. B., Doherty, R. M., Eyring, V., Faluvegi, G., Horowitz, L. W., Josse, B., Lee, Y. H., MacKenzie, I. A., Nagashima, T., Plummer, D. A., Righi, M., Rumbold, S. T., Skeie, R. B., Shindell, D. T., Strode, S. A., Sudo, K., Szopa, S., and Zeng, G.: Pre-industrial to end 21st century projections of tropospheric ozone from the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP), Atmos. Chem. Phys., 13, 2063-2090, https://doi.org/10.5194/acp-13-2063-2013, 2013.

Ziemke, J. R., and S. Chandra (2003), La Nina and El Nino—induced variabilities of ozone in the tropical lower atmosphere during 1970–2001, Geophys. Res. Lett., 30, 1142, doi: 10.1029/2002GL016387, 3.

Ziemke, J. R., Oman, L. D., Strode, S. A., Douglass, A. R., Olsen, M. A., McPeters, R. D., Bhartia, P. K., Froidevaux, L., Labow, G. J., Witte, J. C., Thompson, A. M., Haffner, D. P., Kramarova, N. A., Frith, S. M., Huang, L.-K., Jaross, G. R., Seftor, C. J., Deland, M. T., and Taylor, S. L.: Trends in global tropospheric ozone inferred from a composite record of TOMS/OMI/MLS/OMPS satellite measurements and the MERRA-2 GMI simulation , Atmos. Chem. Phys., 19, 3257-3269, https://doi.org/10.5194/acp-19-3257-2019, 2019.

Supplementary material to: An inter-comparison of tropospheric ozone reanalysis products from CAMS, CAMS-Interim, TCR-1 and TCR-2

Vincent Huijnen[1], Kazuyuki Miyazaki[2], Johannes Flemming[3], Antje Inness[3], Takashi Sekiya[4], Martin G. Schultz[5]

[1] Royal Netherlands Meteorological Institute, De Bilt, the Netherlands
[2] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA
[3] ECMWF, Shinfield Park, Reading, RG2 9AX, UK
[4] Research Institute for Global Change (RIGC), Japan Agency for Marine–Earth Science and Technology (JAMSTEC), Yokohama 2360001, Japan
[5] Jülich Supercomputing Centre, Forschungszentrum Jülich, Jülich, Germany

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

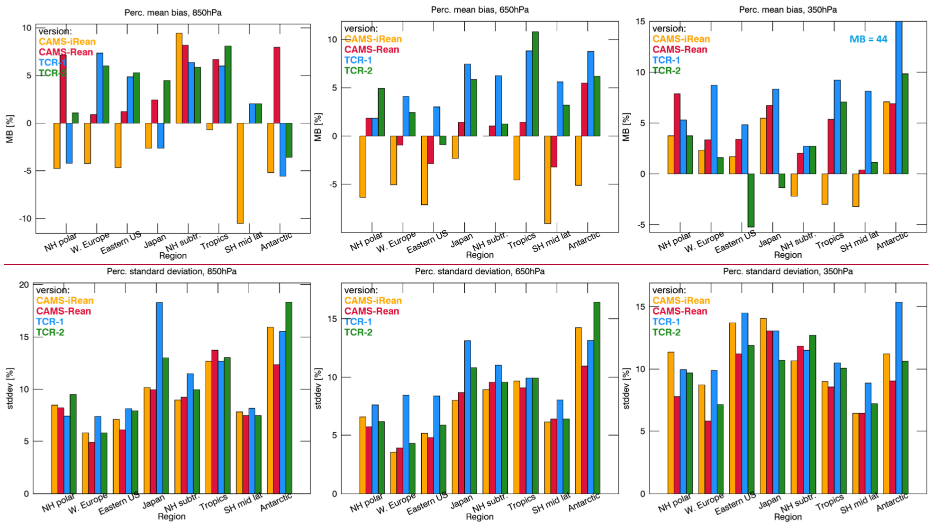


Figure S1. Evaluation of ozone normalized mean bias (reanalysis-observation, top) and standard deviation of difference (bottom) for the four reanalysis products at 850, 650 and 350 hPa against sondes, computed for various regions, for the 2005-2016 time period.

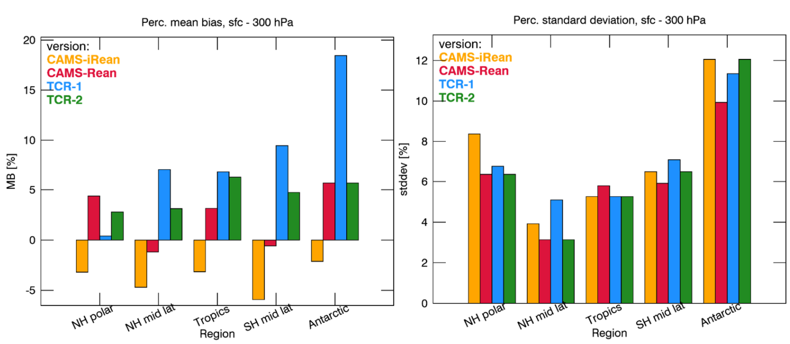
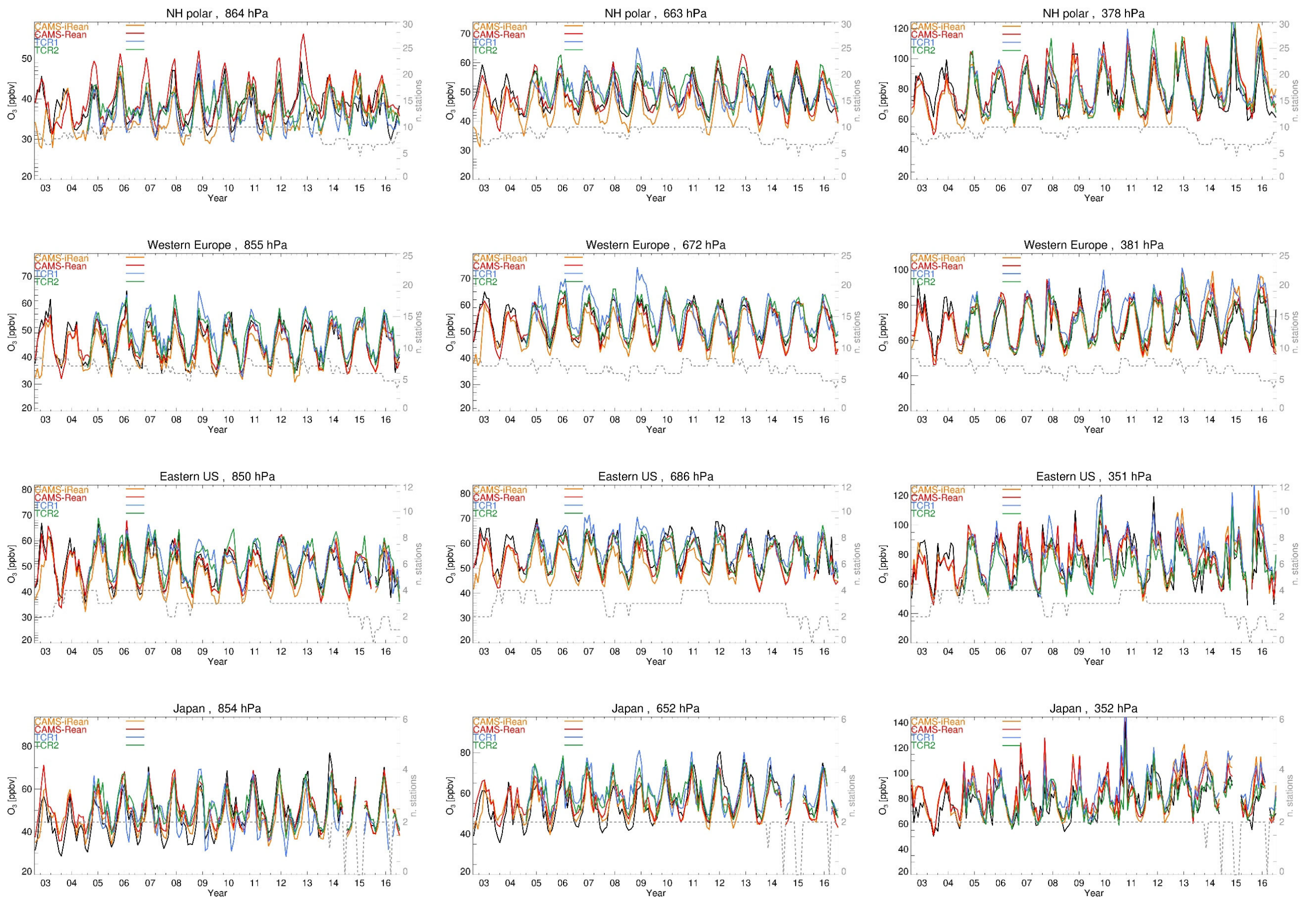


Figure S2. Evaluation of ozone partial columns (surface – 300 hPa) for the four reanalysis products for the 2005-2016 time period against sonde observations within five latitude bands. Left: normalized mean bias (reanalysis-observation), right: normalized standard deviation of differences.


NH polar , 864 hPa
NH polar , 663 hPa
NH polar , 378 hPa
Western Europe , 855 hPa
Western Europe , 672 hPa
Western Europe , 381 hPa
Eastern US , 850 hPa
Eastern US , 686 hPa
Eastern US , 351 hPa
Japan , 854 hPa
Japan , 652 hPa
Japan , 352 hPa
CAMS-iRean
CAMS-Rean
TCR1
TCR2
O_3 [ppbv]
n. stations
Year

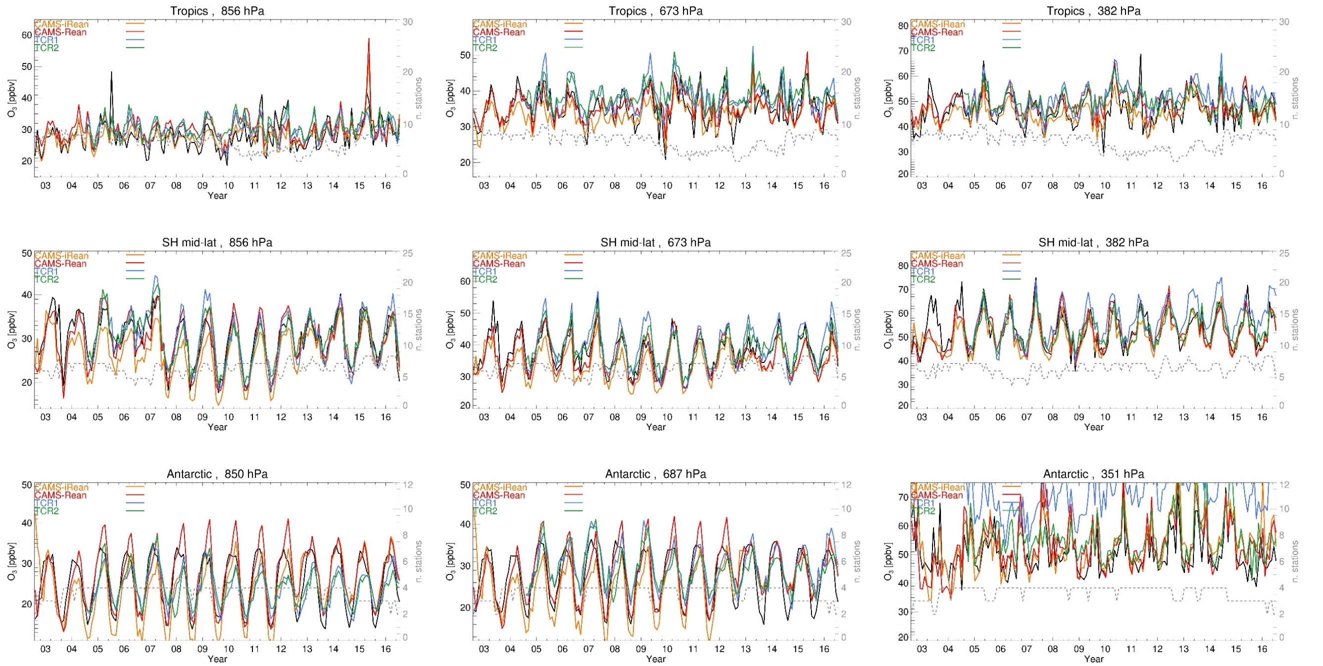
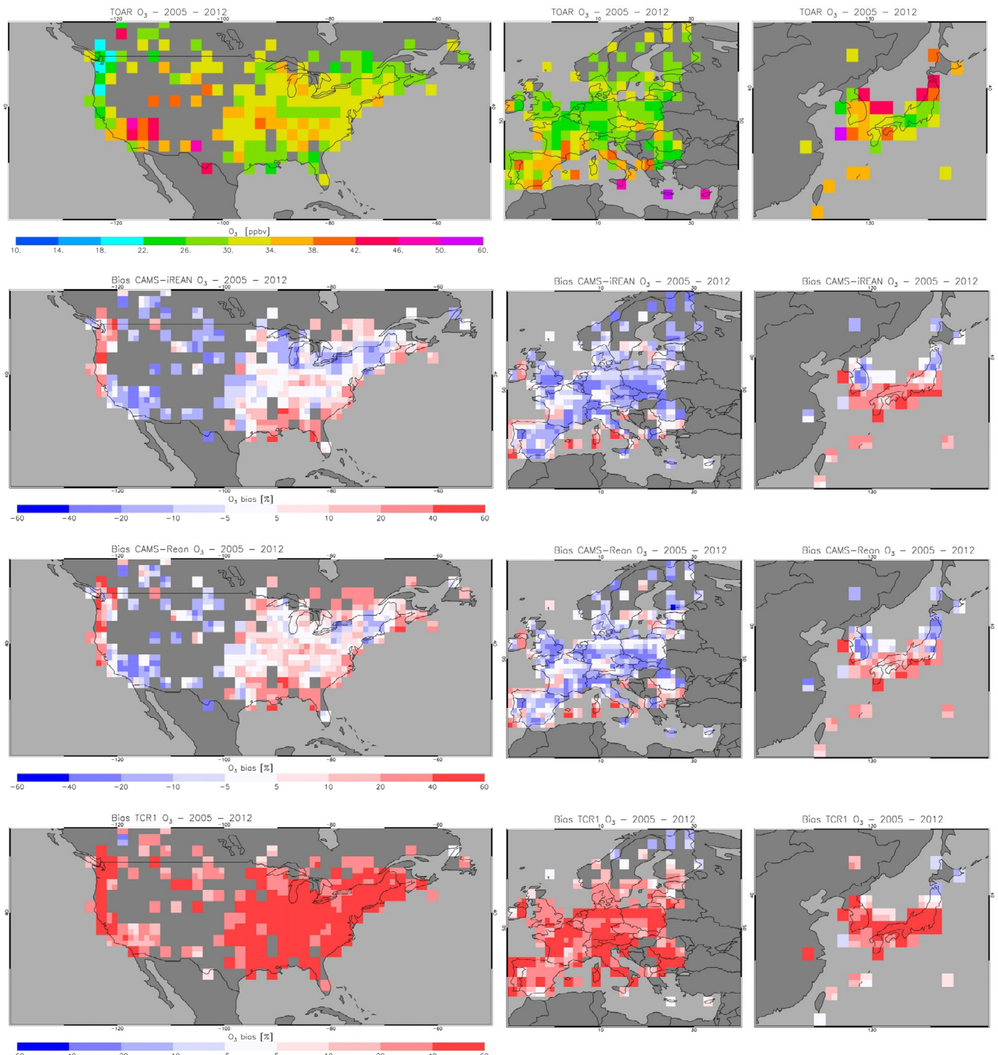


Figure S~~1~~3. Time series of regionally and monthly aggregated ozone concentrations at different altitudes (850, 650 and 350 hPa), sampled at ozone sonde locations, against ozone sonde observations (black). **The gray dashed line refers to the number of stations that contribute to the statistics (right vertical axis).**



TOAR O3 - 2005 - 2012
O3 [ppbv]
10. 14. 18. 22. 26. 30. 34. 38. 42. 46. 50. 60.
Bias CAMS-iREAN O3 - 2005 - 2012
Bias CAMS-Rean O3 - 2005 - 2012
Bias TCR1 O3 - 2005 - 2012
O3 bias [%]
-60 -40 -20 -10 -5 5 10 20 40 60

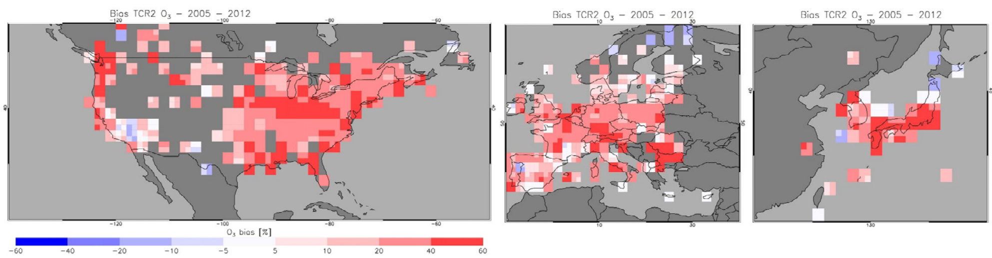


Figure S42: Multi-annual (2005-2012) mean surface ozone from TOAR for three regions, (top figures), along with corresponding relative mean bias (reanalysis-observation) for the reanalyses CAMS-iRean, CAMS-Rean, TCR1 and TCR2, respectively.

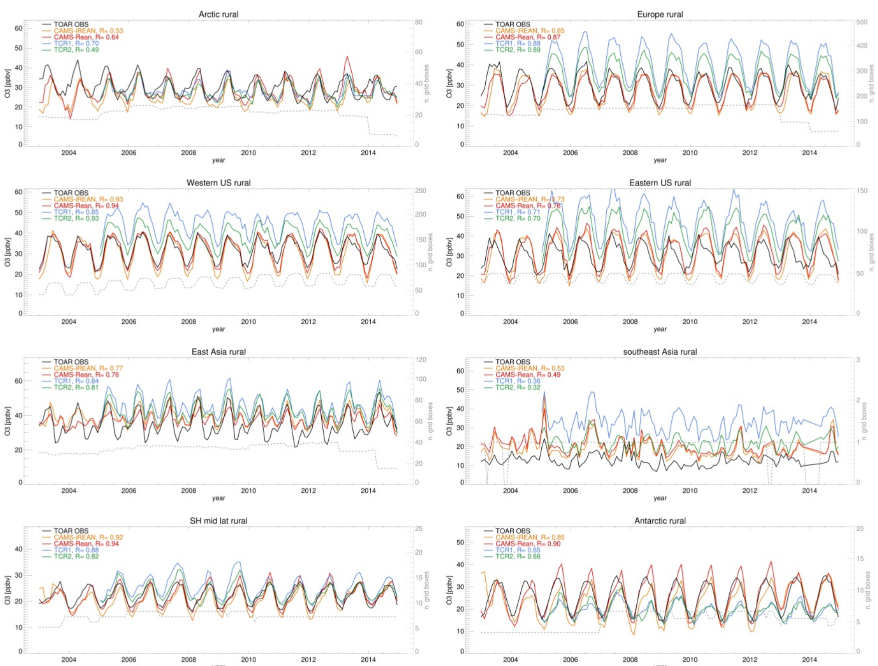


Figure S35: Time series of regional, monthly mean surface ozone against TOAR observations. The dashed line indicates the number of TOAR 2°×2° grid boxes contributing to the statistics (see also right axis). Also the temporal correlation for the 2005-2014 time series is given in the figure legends.

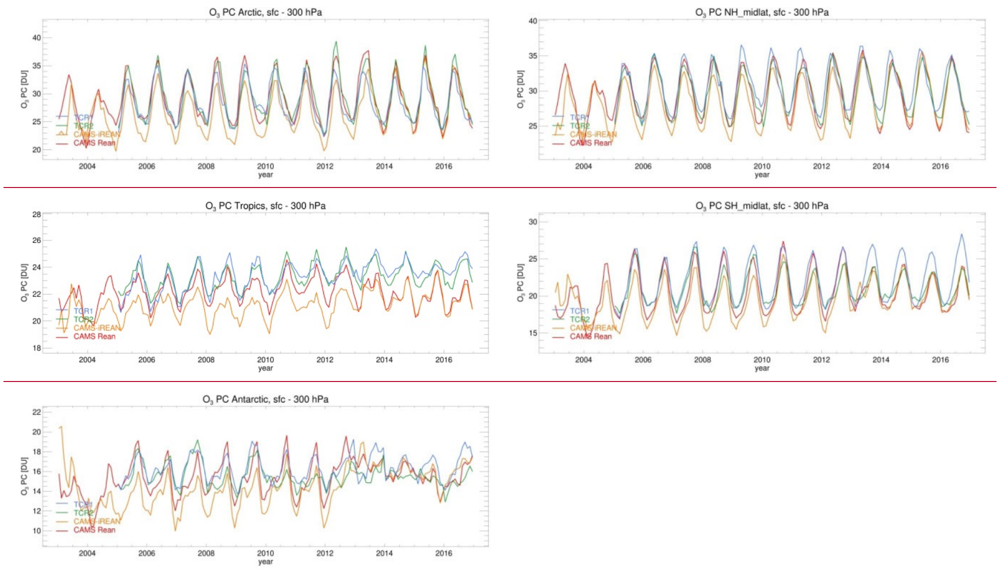


Figure S6: Intercomparison of regionally averaged monthly mean partial columns up to 300 hPa.

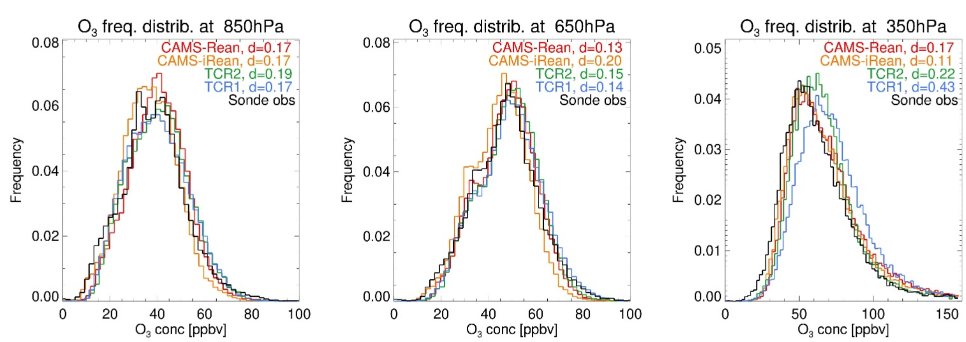


Figure S7: Normalized frequency distributions of O_3 mixing ratios sampled at individual ozone sonde observations between 2005 and 2016 at 850 (left), 650 (middle) and 350 hPa (right) for the four reanalyses, together with the corresponding frequency distribution for the observations. The sum of the absolute differences between the frequency distribution of the reanalyses and observations is also given as *d*.

Met opmaak: Standaard

Figure S84: Anomalies in monthly mean O_3 partial columns (surface to 300 hPa) in four reanalyses, averaged for six regions: Arctic (>60°N), Eastern US (90°W – 70°W; 30°N - 43°N) Europe (10°W-30°E; 35°N-60°N), East Asia (108°E-160°E, 20°N-50°N), Tropics (30°S-30°N) and Antarctic (>60°S). Standard deviations for monthly mean anomalies are given, computed for the 2005-2016 time period.

Table S1: correlation coefficient R of the interannual variability in tropospheric O_3 columns between the four reanalyses, as computed for the 2005-2016 monthly mean time series in tropospheric O_3 columns from the surface to 300 hPa for different regions, see Figure S8.

Southeast Asia	CAMS-Rean	CAMS-iRean	TCR-2	TCR-1
CAMS-Rean	1.00	0.93	0.82	0.73
CAMS-iRean		1.00	0.83	0.76
TCR-2			1.00	0.90
TCR-1				1.00
ENSO_3.4	CAMS-Rean	CAMS-iRean	TCR-2	TCR-1
CAMS-Rean	1.00	0.83	0.78	0.73
CAMS-iRean		1.00	0.57	0.48
TCR-2			1.00	0.77
TCR-1				1.00
Arctic	CAMS-Rean	CAMS-iRean	TCR-2	TCR-1
CAMS-Rean	1.00	0.50	0.60	0.35
CAMS-iRean		1.00	0.48	0.26
TCR-2			1.00	0.46
TCR-1				1.00
Europe	CAMS-Rean	CAMS-iRean	TCR-2	TCR-1
CAMS-Rean	1.00	0.33	0.42	0.07
CAMS-iRean		1.00	0.47	0.24
TCR-2			1.00	0.40
TCR-1				1.00
Eastern US	CAMS-Rean	CAMS-iRean	TCR-2	TCR-1
CAMS-Rean	1.00	0.41	0.63	0.45
CAMS-iRean		1.00	0.54	0.46
TCR-2			1.00	0.64
TCR-1				1.00

Eastern Asia	CAMS-Rean	CAMS-iRean	TCR-2	TCR-1
CAMS-Rean	1.00	0.52	0.52	0.29
CAMS-iRean		1.00	0.69	0.55
TCR-2			1.00	0.59
TCR-1				1.00
Tropics	CAMS-Rean	CAMS-iRean	TCR-2	TCR-1
CAMS-Rean	1.00	0.17	0.29	0.01
CAMS-iRean		1.00	0.55	0.45
TCR-2			1.00	0.73
TCR-1				1.00
Antarctic	CAMS-Rean	CAMS-iRean	TCR-2	TCR-1
CAMS-Rean	1.00	0.65	0.33	0.39
CAMS-iRean		1.00	0.16	0.46
TCR-2			1.00	0.56
TCR-1				1.00