Authors' response to Reviewer 1's comments

We thank the reviewer for the helpful and constructive comments. Our responses to the reviewer's comments are included below following each reviewer's comment. The reviewer's comments are shown in *italic*, and our responses are in blue.

This manuscript describes a detailed evaluation of Arctic tropospheric ozone in two regional chemical transport models using a variety of surface and vertical profile measurements for the year 2015. The study undertakes a thorough observational comparison, followed by a detailed investigation of the role of halogen chemistry in controlling model ozone, an analysis of wildfire emission contributions, and finally a regional Arctic tropospheric ozone budget analysis from one of the two models.

Among other aspects, the detailed comparison of high resolution ozone simulations with hourly data is an important advance on many previous studies that have evaluated coarse global-scale models with monthly mean observations. The presentation of comparison of model simulations that include detailed halogen chemistry against simulations with this removed is also very informative, as are the investigations of sensitivity to assumptions in the model Br mechanisms. In addition, evaluation of the structure of modelled and observed ozone vertical profiles using aircraft observations during springtime is of great benefit. These results are of high value to the Arctic atmospheric composition and modelling communities.

Generally, I do not have any major concerns or reservations with the paper. The analysis presented is very thorough, and the manuscript is well written. I recommend that the manuscript is suitable for publication in ACP once the following comments have been addressed.

General comments

The paper is very long, however I recognise that the analysis presented is very thorough. I have made one suggestion of where text could be shortened by avoiding separation of parts of the results that could be better linked (see below).

R: See response to the specific comments below.

In a couple of places more could be done to compare the performance of the models presented here with previous model assessments of ozone using the same datasets. I have highlighted a couple of examples below in my specific comments.

R: See responses to the specific comments below.

Specific comments

Line 61: Archibald et al., (2020) is not a primary reference for the impacts of ozone on health and ecosystems. Can the authors provide alternative references for these two aspects of ozone impact.?

R: We have updated the references to those pertinent to the impact of ozone on health (Fleming et al., 2018; US Environmental Protection Agency, 2013; World Health organization, 2013) and ecosystems (Ainsworth et al., 2012; Mills et al., 2011, 2018).

Line 65: "changes [..].. in the transport pattern from lower latitudes" This needs to be more explicit to provide context. i.e. "changes in the patterns of transport of ozone and precursors from lower latitudes".

R: We have incorporated the reviewer's suggestion.

Line 88: "The ability of current models to simulate Arctic tropospheric O3 has been evaluated in several studies (e.g., Monks et al., 2015b; Shindell et al., 2008; Whaley et al., 2023)" What is meant by "current" in this context (given that citations from 2015 are relevant here)? I agree with the need to cite some of these older studies, since it is not clear that the models have improved substantially in this time. Maybe omit "current" and rephrase as "... evaluated in several previous and recent studies..."

R: We have rephrased as suggested by the reviewer.

Line 190-193: Does this imply that within the European domain ECLIPSE emissions are not used (replaced by EMEP)? What is the motivation for this? Is it simply more information from higher resolution? How different are the emissions?

R: It is correct that over Europe the ECLIPSE emissions were replaced by the EMEP emissions in the DEHM simulations. There were several reasons for this replacement. Over Europe, the EMEP emission inventories have higher spatial (0.1° x 0.1° as compared to ECLIPSE at 0.5° x 0.5°) and temporal resolution (EMEP provides yearly emissions from 1990 to 2022 as compared to the 5 years intervals from 1990 to 2020 for ECLIPSE). ECLIPSE6b was developed in 2018-2019, while the EMEP emissions are continuously updated each year not only for the newly added year but also for previous years retrospectively (i.e., previous inventories are revised when new activity data for the different European countries or updated emissions factors are available). There are differences in emissions between the ECLIPSE and EMEP inventories over Europe both in spatial distribution (due to the difference in spatial resolution and allocation) and in total amount (most pronounced for NOx and NMVOC, where ECLIPSE emissions are significantly higher than EMEP emissions). The

DEHM model and other European models such as the EMEP model have intensively used and tested the EMEP emissions in many projects, where the model simulations for Europe shows good performance (Brandt et al., 2012; Simpson et al., 2012). We have revised the text to "The anthropogenic emissions from the ECLIPSE v6b dataset at 0.5° x 0.5° resolution (Klimont et al., 2017) are used for the portion of the model domain outside Europe, while for the areas over Europe the emissions from the European Monitoring and Evaluation Programme (EMEP) expert database with 0.1° x 0.1° resolution are used (see https://www.ceip.at/)".

Table 1: It would be useful to add to this table the temporal resolution of the data measured and/or used in the study. Could this perhaps be added into column 3? Similarly for sondes, what is the approximate vertical resolution of the data?

R: We have added the appropriate information on data temporal resolution in Table 1. For ozonesondes, the data obtained from HEGIFTOM has the vertical resolution varying from a few meters to a few tens of meters over the lowest 5 km of the atmosphere. We have added this information at the end of section 2.2.3.

Section 3.1 and Fig. 2 discussion. There is no mention of the low ozone simulated in both models over the northern Eurasian region during winter. This is also evident in Figure 3, which highlights the winter months as being the time of the minimum at the surface. Is this the impact of ozone titration by Eurasian NO emissions in winter?

R: We have added "Both models simulated low surface O3 concentrations over northern Eurasia and northern Europe during winter. The low ozone can be argued to be attributable to reduced photo-chemical production and enhanced titration by NO emissions from local sources within the darker and shallower boundary layer during winter, as well as dry deposition". This is supported by examining the O3 tendencies from one of the models (GEM-MACH) as discussed in section 4.3.

Figure 4 - Would it be possible to add a legend to the figure labelling the coloured lines used?

R: Will do.

Line 557: This text describes the statistical evaluations for the comparisons shown in Fig. 4. I am not sure this needs to be separated from the presentation of performance of the models in the previous paragraph. The text could be combined to reference the statistics as part of the discussion of model performance. This would also help qualify several subjective terms such as "compare well" (e.g. line 551).

R: Here we present the comparison of O3 time series first to examine how model simulations are capable of capturing the seasonal and synoptic variabilities in surface O3 observed at the Arctic sites in a broad sense. This is followed by statistical evaluation to quantify and characterise model performance, which leads to discussions on several processes affecting the model performance.

We tried to combine the statistical evaluation with the discussion on time series comparison as suggested by the reviewer, but it did not improve the clarity and readability. However, we have reorganised this section, moving the original discussion on the impact of the model parameterization of dry deposition on model performance at the two northern European boreal sites to combine with the discussion on statistical evaluation, as the low bias of modelled O3 is characterised there. We have also eliminated the discussion on uncertainty in model representation of biogenic VOC emissions and its possible impact on model performance, for the sake of shortening the length of the paper. This reorganisation was also in response to Reviewer 2's suggestion to break up the long opening paragraph of 3.2. Regarding the subjective term "compare well" used in describing the time series comparison, we have added more explicit descriptives such as, "capturing the observed seasonal and synoptic scale variation (also evident from the statistical evaluation shown in Table 2)".

Line 575: The authors make the statement that the comparisons shown demonstrate improved model performance compared with similar evaluations using global models. Would it be possible to be more quantitative, given that previous studies have used the same surface sites for evaluation and will have quoted e.g. mean bias values (notwithstanding the use of different time resolution data)?

R: The statement was made largely based on a qualitative comparison of between similar model-observation time series from the present study and those presented in previous studies mostly comprised of large-scale global models (e.g., Monks et al., 2015; Whaley et al., 2023; Young et al., 2018). Unfortunately, very little standardized statistics for these models have been presented in such a way that can be compared quantitatively with the results in this study. Only Whaley et al. (2023) provided annual (normalised) mean bias from the multi-model ensemble at selected Arctic ground sites, based on monthly mean values (see their Figure 6). Such a metric has issues with error cancellations amongst models and seasonal variations which can give a false impression of accuracy. Ideally, a suite of robust metrics which can more accurately identify the true seasonal and sub-seasonal variations and associated errors between models and observations is desired. We have revised the text to "Overall, the two regional models seem to demonstrate better skills in capturing the observed seasonal variations in the Arctic surface ozone, compared to the large-scale global atmospheric chemistry models reported in previous assessments (e.g., Law et al., 2023; Whaley et al., 2023; Young et al., 2018) where the models showed large spread in simulated surface O3 concentrations and inability to reproduce the observed seasonal cycles at some of the Arctic sites. Besides the implementation of the processes involved in springtime bromine chemistry in the Arctic, the better performance from the two independent regional models in this study can be attributed, at least in part, to better resolved atmospheric dynamics and boundary layer processes modelled at finer spatial and temporal scales".

Section 3.1: The Whaley et al., (2023) study presented evaluation of a set of global models against ozone sonde data (Figure 8 in their paper). It would be informative to make some sort of reference / comparison to this in putting the results presented by the authors into context.

R: We believe that the reviewer is referring to section 3.3 (Ozone vertical profiles comparison with ozonesondes). Yes, a reference to their study would be informative on how the regional models in this study compare with the large-scale models in that study. To do this, we have included a new figure in supplementary material (SF. 5; in the revised version) showing similar plots of relative difference (or the normalised mean bias, NMB) between modelled and observed ozone profiles which can be directly compared to their study. Again, the simulated O3 from the two regional models included in this study are shown in much better agreement with the ozonesonde observations than the large-scale models in Whaley et al. (2023), with significantly smaller relative difference (or NMB), generally well within +/-25% over the lowest 5 km of the atmosphere, compared to +/-50% shown in Whaley et al. (2023). We have included the following statement in section 3.3 of the revised manuscript:

"Whaley et al. (2023) compared model simulated vertical profiles (using monthly mean model output) from 12 different large-scale models to the ozonesonde measurements from the same group of sites as we examined here (see their Figure 8 and Figure S1). We have plotted the profiles of seasonal relative difference between model simulations and observations (or NMB) in SF.5, which can be compared with the results shown in Whaley et al. (2023). Again, the two regional models here show better skills in simulating the observed O3 vertical profiles over the lowest 5 km of the atmosphere examined here (having considerably smaller biases, generally well within +/-25%, compared to the large spread of relative difference, +/-50%, in the same altitude range amongst the large-scale global models)".

Figure 10 - A minor point, but maybe it is worth spelling out "interquartile range" (IQR) in the legend or caption.

R: Done.

Figure 11 - It might help in comparison of the different sensitivity simulations to provide some quantitative metrics for the comparisons with observations (i.e. mean bias / r2 values).

R: We have added a couple of statistical metrics (normalised mean bias and Pearson correlation coefficient) for reference (in revised Figure 11). However, Figure 11 is mainly to illustrate how the model's ability in simulating the observed ODEs is influenced by parameter settings that controls the snowpack bromine production (Φ_1) and bromine production through heterogeneous reactions on atmospheric aerosols (aerosol surface area). A more pertinent approach would be to design metrics to quantify how the frequency and intensity of the observed ODEs being replicated by model simulations. This will be pursued in our future investigations.

Line 955 - The Arnold et al., (2015) evaluation of fire-impacted O3/CO enhancement ratios are also based on monthly mean large-scale Arctic enhancements, so these could be more directly compared with results presented here (i.e. they are also not plume specific enhancements).

R: Thanks for pointing this out. It is true that Arnold et al. (2015) evaluation of the O3-to-CO enhancement ratios due to boreal wildfires were based on modelled monthly mean. However, their ΔO3/ΔCO was evaluated from a linear fit to the scatterplot of O3-to-CO over model grids identified as dominated by wildfire influences (determined by means of fixed-lifetime CO tracers introduced in their model simulations to track transport of tracers emitted from wildfire TR_{fire} and from anthropogenic sources $TR_{anthrop}$; fire-impacted grids are those where $TR_{fire}/(TR_{fire}+TR_{anthrop}) > 0.67$). In our case, the enhancement ratios ΔO3/ΔCO were directly calculated from the pair of model runs with and without the wildfire emissions, similar to the approach of Pfister et al. (2006) and Thomas et al. (2013). Pfister et al. (2006) also examined different approaches in evaluating the enhancement ratios, namely, an approach using a linear fit to the O3-CO scatterplot (similar to Arnold et al., 2015) vs. an enhancement approach based on excess mixing ratios of O3 and CO (similar to this study). They showed that when the variability in the background concentration levels was well characterised, the enhancement approach would be more robust and accurate in evaluating the fire-influenced ΔO3/ΔCO ratios than the approach based on scatterplot. The latter is subjected to the uncertainty in isolating the enhancement due to wildfires from those due to non-wildfire (anthropogenic) emissions, as a result of different NOx/CO emissions ratios between boreal wildfires and anthropogenic sources. We have replaced the previous sentence, "Note, however, here the excess ratios are evaluated based on monthly mean over a broad area while the previously reported values were mostly evaluated within plumes and for a short time period (e.g., duration of field campaign)" with the following,

"The large variability in estimated wildfire impacted ΔO3/ΔCO enhancement ratios from various studies can arise from the different approaches used in evaluating the enhancement ratios. By comparing between a scatter technique (based on a linear fit to the O3-CO concentration scatterplot) and an enhancement technique (based on the evaluation of O3 and CO excess mixing ratios due to wildfire emissions), Pfister et al., (2006) showed that the ΔO3/ΔCO ratios evaluated using the scatter technique were affected by the selection of biomass-burning-impacted air masses and the degree of mixing in the considered air masses. Much higher enhancement ratios were found in anthropogenic-combustion-impacted air masses than in the boreal-wildfire-impacted air masses, due to the difference in NOx/CO emissions ratios between these source types. Pfister et al. (2006) also showed that when the variability in the background concentration levels was well characterised, the enhancement technique would be more robust and accurate in evaluating the fire-influenced ΔO3/ΔCO enhancement ratios."

Page 47: Discussion of PAN/CO enhancement ratios. In the Arnold et al., (2015) study, a difference in PAN/CO enhancement values was identified between models forced using different reanalyses products (models forced using GEOS-5 data displayed lower enhancements compared with models forced by ERA-Interim data). It would be interesting to know how the models presented compare and if they are consistent with the Arnold et al., (2015) values according to the meteorological dataset used (for DEHM using ERA-5 for example).

R: The DEHM meteorology is generated by WRF driven by ERA5 reanalysis. However, as the sensitivity simulation on wildfire emissions was not performed by the DEHM group, we were not able to carry out the same wildfire impact analysis on the DEHM simulations. The ΔPAN/ΔCO

enhancement ratios due to boreal wildfires from this study (based on GEM-MACH simulations), ~3 – 4 pptv/ppbv at the lowest model level, increasing with height to 6 – 7 pptv/ppbv near 700 hPa, are comparable to the PAN/CO enhancement ratios reported in Arnold et al. (2015) from the group of models driven by the ECMWF meteorological reanalysis. We have added this reference in the revised manuscript.

Editorial / typographical corrections

Line 79: "variations in the Arctic tropospheric O3" Omit "the".

R: revised as suggested.

Line 458: Better as "...varying degrees of complexity."

R: revised as suggested.

References

Arnold, S. R., et al., Biomass burning influence on high-latitude tropospheric ozone and reactive nitrogen in summer 2008: a multi-model analysis based on POLMIP simulations, Atmos Chem Phys, 15, 6047–6068, https://doi.org/10.5194/acp-15-6047-2015, 2015.

Whaley, C. H., et al., (2023), Arctic tropospheric ozone: assessment of current knowledge and model performance, Atmos. Chem. Phys., 23, 637-661, https://doi.org/10.5194/acp-23-637-2023.

Ainsworth, E. A., Yendrek, C. R., Sitch, S., Collins, W. J., and Emberson, L. D.: The Effects of Tropospheric Ozone on Net Primary Productivity and Implications for Climate Change*, Annu. Rev. Plant Biol., 63, 637–661, https://doi.org/10.1146/annurev-arplant-042110-103829, 2012.

Brandt, J., Silver, J. D., Frohn, L. M., Geels, C., Gross, A., Hansen, A. B., Hansen, K. M., Hedegaard, G. B., Skjøth, C. A., Villadsen, H., Zare, A., and Christensen, J. H.: An integrated model study for Europe and North America using the Danish Eulerian Hemispheric Model with focus on intercontinental transport of air pollution, AQMEII Int. Initiat. Eval. Reg.-Scale Air Qual. Models - Phase 1, 53, 156–176, https://doi.org/10.1016/j.atmosenv.2012.01.011, 2012.

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. Anthr., 6, 12, https://doi.org/10.1525/elementa.273, 2018.

Mills, G., HAYES, F., SIMPSON, D., EMBERSON, L., NORRIS, D., HARMENS, H., and BÜKER, P.: Evidence of widespread effects of ozone on crops and (semi-)natural vegetation in Europe (1990–2006) in relation to AOT40- and flux-based risk maps, Glob. Change Biol., 17, 592–613, https://doi.org/10.1111/j.1365-2486.2010.02217.x, 2011.

Mills, G., Pleijel, H., Malley, C. S., Sinha, B., Cooper, O. R., Schultz, M. G., Neufeld, H. S., Simpson, D., Sharps, K., Feng, Z., Gerosa, G., Harmens, H., Kobayashi, K., Saxena, P., Paoletti, E., Sinha, V., and Xu, X.: Tropospheric Ozone Assessment Report: Present-day tropospheric ozone distribution and trends relevant to vegetation, Elem. Sci. Anthr., 6, 47, https://doi.org/10.1525/elementa.302, 2018.

Monks, S. A., Arnold, S. R., Emmons, L. K., Law, K. S., Turquety, S., Duncan, B. N., Flemming, J., Huijnen, V., Tilmes, S., Langner, J., Mao, J., Long, Y., Thomas, J. L., Steenrod, S. D., Raut, J. C., Wilson, C., Chipperfield, M. P., Diskin, G. S., Weinheimer, A., Schlager, H., and Ancellet, G.: Multi-model study of chemical and physical controls on transport of anthropogenic and biomass burning pollution to the Arctic, Atmos Chem Phys, 15, 3575–3603, https://doi.org/10.5194/acp-15-3575-2015, 2015.

Pfister, G. G., Emmons, L. K., Hess, P. G., Honrath, R., Lamarque, J.-F., Val Martin, M., Owen, R. C., Avery, M. A., Browell, E. V., Holloway, J. S., Nedelec, P., Purvis, R., Ryerson, T. B., Sachse, G. W., and Schlager, H.: Ozone production from the 2004 North American boreal fires, J. Geophys. Res. Atmospheres, 111, https://doi.org/10.1029/2006JD007695, 2006.

Simpson, D., Benedictow, A., Berge, H., Bergström, R., Emberson, L. D., Fagerli, H., Flechard, C. R., Hayman, G. D., Gauss, M., Jonson, J. E., Jenkin, M. E., Nyíri, A., Richter, C., Semeena, V. S., Tsyro, S., Tuovinen, J.-P., Valdebenito, Á., and Wind, P.: The EMEP MSC-W chemical transport model – technical description, Atmos Chem Phys, 12, 7825–7865, https://doi.org/10.5194/acp-12-7825-2012, 2012.

Thomas, J. L., Raut, J.-C., Law, K. S., Marelle, L., Ancellet, G., Ravetta, F., Fast, J. D., Pfister, G., Emmons, L. K., Diskin, G. S., Weinheimer, A., Roiger, A., and Schlager, H.: Pollution transport from North America to Greenland during summer 2008, Atmos Chem Phys, 13, 3825–3848, https://doi.org/10.5194/acp-13-3825-2013, 2013.

US Environmental Protection Agency: Integrated Science Assessment for Ozone and Related Photochemical Oxidants, Office of Research and Development, Research Triangle Park, NC, 2013.

Whaley, C. H., Law, K. S., Hjorth, J. L., Skov, H., Arnold, S. R., Langner, J., Pernov, J. B., Bergeron, G., Bourgeois, I., Christensen, J. H., Chien, R.-Y., Deushi, M., Dong, X., Effertz, P., Faluvegi, G., Flanner, M., Fu, J. S., Gauss, M., Huey, G., Im, U., Kivi, R., Marelle, L., Onishi, T., Oshima, N., Petropavlovskikh, I., Peischl, J., Plummer, D. A., Pozzoli, L., Raut, J.-C., Ryerson, T., Skeie, R., Solberg, S., Thomas, M. A., Thompson, C., Tsigaridis, K., Tsyro, S., Turnock, S. T., von Salzen, K., and Tarasick, D. W.: Arctic tropospheric ozone: assessment of current knowledge and model performance, Atmos Chem Phys, 23, 637–661, https://doi.org/10.5194/acp-23-637-2023, 2023.

World Health organization: Review of evidence on health aspects of air pollution – REVIHAAP final technical report, World Health Organization, 2013.

Young, P. J., Naik, V., Fiore, A. M., Gaudel, A., Guo, J., Lin, M. Y., Neu, J. L., Parrish, D. D., Rieder, H. E., Schnell, J. L., Tilmes, S., Wild, O., Zhang, L., Ziemke, J., Brandt, J., Delcloo, A., Doherty, R. M., Geels, C., Hegglin, M. I., Hu, L., Im, U., Kumar, R., Luhar, A., Murray, L., Plummer, D., Rodriguez, J., Saiz-Lopez, A., Schultz, M. G., Woodhouse, M. T., and Zeng, G.: Tropospheric Ozone Assessment Report: Assessment of global-scale model performance for global and regional ozone distributions, variability, and trends, Elem. Sci. Anthr., 6, 10, https://doi.org/10.1525/elementa.265, 2018.