Atmos. Chem. Phys. Discuss., 14, C11704–C11719, 2015
www.atmos-chem-phys-discuss.net/14/C11704/2015/

Atmospheric Chemistry and Physics Discussions

Interactive comment on "Multi-model study of chemical and physical controls on transport of anthropogenic and biomass burning pollution to the Arctic" *by* S. A. Monks et al.

S. A. Monks et al.

s.a.monks@leeds.ac.uk

Received and published: 31 January 2015

Interactive comment on "Multi-model study of chemical and physical controls on transport of anthropogenic and biomass burning pollution to the Arctic" by S. A. Monks et al. Anonymous Referee #2

This manuscript presents a comprehensive multi-model analysis of CO and O3 in the Northern Hemisphere with focus in the Arctic region, in the frame of the POLMIP project. Potential sources of model discrepancies and variability have been investigated in great details through analysis of simulated CO, O3 and OH, as well as CO-like

tracers throughout the Arctic troposphere using numerous atmospheric models (global and regional) which have been first evaluated against Arctic surface, aircraft and satellite measurements. This paper is an interesting and important work which covers many aspects (atmospheric transport and chemical processes) related to model deficiencies in tropospheric ozone and its precursors in the Arctic. The manuscript is well written and the topic is suitable for ACP.

I have one major comment related to the impact of using different model resolutions and meteorological data for evaluating model performance and inter-model differences, which should be addressed before final publication. The authors argue to evaluate the chemical and physical controls on transport of pollution to the Arctic through a multi-model analysis. However from literature reviews it is known that model resolutions may be influencing ozone and CO concentrations. This could, amongst others, explain the inter-model differences in the underrepresentation of ozone transported from mid-latitudes. Similarly, underrepresentation of ozone transported from stratosphere may also be influenced by vertical resolution between models, especially at the tropopause level. The different vertical and horizontal resolutions in models as well as the different meteorological datasets are mentioned in Table 1 but their impact in both model performances and inter-model comparison results is not discussed later. I understand that modelers are used to perform model runs with specific resolution and offline meteorology, and the complexity for running consistent models, but I do not think that only considering the same emission dataset is enough for arguing models as "directly comparable" and discussing performances and inter-comparison results without mentioning other probable causes of inter-model differences. I would suggest to better discuss your results in line with the differences in both resolutions and meteorological datasets. Even though the authors shortly mention the influence of the resolution by running WRF-Chem at two horizontal resolutions (100 km and 50 km), both of them are high-resolution simulations and they do not allow to evaluate the effect of coarse global model resolutions on multi-model vs. observations comparisons.

The authors would like to that the reviewer for taking the time to read the manuscript and suggest changes. The suggestions have been taken on board and have helped produce a better manuscript. The comments are addressed individually below.

Some similarities in concentrations of CO, O3 and OH were found between models, which were being run using the same meteorology. As mentioned here on pg 25306, L19-23:

'It is interesting to note that the models that use ECMWF reanalyses also have some of the lowest global mean OH concentrations (C-IFS, TM5 and TOMCAT), however, C-IFS and TM5 both use the same chemical mechanism, which has been shown produce lower OH concentrations compared to other mechanisms.'

However, as mentioned here, it is difficult to know whether these similarities are a result of the meteorology driving the models or other existing similarities without doing multiple models runs using different meteorology. This is a huge undertaking as models are often written using one type of meteorology and changing the forcing data would require large parts of the model to be rewritten to allow this. It would be an interesting point to pursue in more detail in future studies of atmospheric chemistry.

I have mentioned this by adding the following sentence to the above point: 'Further investigation into how sensitive model tropospheric OH and other trace gases, such as O3 and CO, are to model differences in meteorology would be worthwhile.'

In terms of resolution we do not see any substantial improvement in C-IFS compared to other models. This model is run at the highest horizontal (T159) out of the global models and one of the highest vertical resolutions (60 vertical levels). I also looked at the number of vertical levels in the models between 400 and 200 hPa to consider how important this would be for O3 transport for the stratosphere and found models had between 5 and 7 levels with 1 model having only 4 (LMDZ). From looking at the aircraft O3 in the UT compared to the models, the models with slightly more vertical levels show no improvement over the other models. The following has been added to

discuss resolution in more detail:

- As below (p. 25308, ll. 10-13): 'For the models shown in Figure 13, LMDZ and TOMCAT have a lower number of vertical levels below 700 hPa (9 and 8, respectively) compared to the other models (which have either 15 or 16 levels). However, they have very different vertical differences in the 25-day fixed lifetime tracer suggesting other processes apart from vertical resolution are important.'

- Added to the discussion regarding Figure 2 Arctic surface CO comparisons: 'The C-IFS and MATCH models, which are run at the two highest horizontal resolutions (see Table 1), show no improvement over the other models in the simulation of CO.'

- Added to the aircraft O3 comparison discussion: 'The higher multi-model mean NMGE against the Falcon data is likely a result of the large negative bias seen in the upper troposphere, most notably in the TOMCAT, C-IFS and MATCH models. These models have a higher number of vertical levels compared to other models at this altitude suggesting vertical resolution is not the cause of the larger bias in these models.'

As you have already pointed out we did find large differences within plumes of high CO between the resolution WRF-Chem simulations suggesting there is a need to consider resolution as an important factor near emission regions. This has also been expanded upon as described in response to reviewer 1.

Specific comments:

p. 25290, l.30: I do not understand what do you mean with this sentence. Please explain or reformulate. Do you mean that in terms of available days when comparing with observations?

I can't find the sentence that the reviewer is referring to. L30 does not exist on p25290.

p.25290, ll.8-11: It is not clear to me what do you mean by "These tracers therefore had the same transport . . ." I guess you mean it by comparing, for each tracer, each TOMCAT run performed with monthly mean OH concentration fields taken from

each POLMIP model? Could you please clarify? For this sensitivity study, why did you choose the TOMCAT model which uses lower spatial and vertical resolutions than most of the POLMIP models?

For the first query, I have tried to make this clearer by rewriting the sentence: Changed from: 'For this study, a further set of these six tracers were calculated in the TOMCAT model, using the same emissions, with loss determined by monthly mean OH concentration fields taken from each POLMIP model, and kinetics of CO+OH loss (k = 1.44×10−13(1+[N2]/4.2×1019). These tracers therefore had the same transport but different loss rates, allowing the impact of OH differences on Arctic burdens of CO in the models to be examined through their differences in OH.' To: 'For this study, a further set of these six tracers were calculated in the TOMCAT model, using the same emissions, with loss determined by monthly mean OH concentration fields taken from each each POLMIP model, and kinetics of CO+OH loss (k ='.44×10−13(1+[N2]/4.2×1019). These tracers therefore had the same model transport, as calculated by TOMCAT, but different loss rates due to different OH fields, allowing the impact of model OH differences on Arctic burdens of CO to be examined through their differences in OH.'

For the second query, TOMCAT was chosen as this model had the capability to do these special experiments. Not all models have this set-up ready installed and tested. This method has been previously used and published in other papers (e.g. Monks et al., (2012); Patra et al., (2011)). As all the OH tracers run in TOMCAT had the same 'low' resolution these runs will still isolate the variability due to OH differences. I have added the following text:

'This technique has been used in previous TOMCAT model studies (Monks et al 2012; Patra et al., 2010).'

Added reference: 'Patra PK; Houweling S; Krol M; Bousquet P; Belikov D; Bergmann D; Bian H; Cameron-Smith P; Chipperfield MP; Corbin K; Fortems-Cheiney A; Fraser A; Gloor E; Hess P; Ito A; Kawa SR; Law RM; Loh Z; Maksyutov S; Meng L; Palmer PI;

Prinn RG; Rigby M; Saito R; Wilson C (2011) TransCom model simulations of CH 4 and related species: Linking transport, surface flux and chemical loss with CH 4 variability in the troposphere and lower stratosphere, Atmospheric Chemistry and Physics, 11, 12813-12837. doi: 10.5194/acp-11-12813-2011'

p. 25290, ll.16-18: Why using monthly means to evaluate your model simulations and not daily means since daily observations and hourly model output are available?

We only have hourly model output available for the period of the POLARCAT aircraft campaigns and therefore we have to use the monthly mean output to compare the models to the surface data over the whole year (see pg 25289 l28 to pg 25290 l1). As the surface data was being used to primarily compare the overall model and observed seasonality and CO/O3 concentrations rather than pollution events this was thought to be sufficient. (Please see below for a further related question)

p. 25290, ll.18-20: Could you please indicate the accuracy associated with the Arctic surface CO and O3 measurements?

This has been added into a new Table as suggested by reviewer 1 (Table 2).

p. 25291, ll. 10-12: What are the main differences between older versions of MOPITT and version 6? Is it possible to use the bias previously calculated by Deeter et al. (2010) with in situ estimates to version 6 of MOPITT, which could be used later in the analysis?

The most important updates in version 6 are, as already stated on page 25291, L1-5: 'Version 6 is the latest release and uses an updated a priori based on a climatology of 2000–2009 output from the CAM-Chem model (Deeter, 2013). It also has an increased sensitivity to lower tropospheric CO by exploiting measurements in the near-infrared and thermal infrared (Deeter et al., 2011).'

Since submission there has been a new paper evaluating V.6 of the MOPITT data. I have removed the following sentence to reflect this:

'Older versions of MOPITT have been validated against other observations and shown to measure CO to within a few ppbv of in situ estimates, with the largest bias occurring at 400 hPa (Deeter et al., 2010).'

And replaced it with:

' The latest version of the MOPITT retrieval has been shown to have a bias of between -5% to 9%, with the largest positive bias occurring at the surface and the largest negative bias occurring at 400 hPa (Deeter et al., 2014).'

Added reference: Deeter, M. N. and Martinez-Alonso, S. and Edwards, D. P. and Emmons, L. K. and Gille, J. C. and Worden, H. M. and Sweeney, C. and Pittman, J. V. and Daube, B. C. and Wofsy, S. C., The MOPITT Version 6 product: Algorithm enhancements and validation., Atmospheric Measurement Techniques, 7, 3623-3632, 10.5194/amt-7-3623-2014, 2014.

p.25292, ll. 5-10: Could you please indicate the uncertainties for CO and O3 measured by the POLARCAT aircrafts?

This has been added into a new Table (Table 2).

p. 25292, ll.19-22: It is not correct here to indicate that NMGE "gives the mean model bias over the vertical column", in fact it corresponds to specific altitude ranges sounded by instruments or to two specific pressure levels (300 and 700hPa) when comparing models with MOPITT. Please reformulate.

I have changed the sentence to clarify this point from:

'Model performance against the observations is summarised in Fig. 10 using the normalised mean gross error (NMGE), which gives the mean model bias (regardless of sign) over the vertical column or over the year as a percentage of the observed concentrations.'

To:

'Model performance against the observations is summarised in Fig. 10 using the normalised mean gross error (NMGE). This gives the mean model bias (regardless of sign) over the vertical profile (when comparing to aircraft data) or over the whole year at a certain altitude (when comparing against surface and MOPITT data) as a percentage of the observed concentrations.'

p. 25293, ll.3-5, Figure 2: Since daily observations for surface measurements are available, you could add in the figure error bars corresponding to the observed variability (3*σ) associated with monthly means. Similarly, you could add on the 3*σ model variability associated with monthly means for each run using the hourly instantaneous output files for model simulations. This should help to better evaluate the seasonal variations and observations-models differences in your Section 4.1.1. Models can capture the monthly means but misrepresent the amplitude of the associated variability. It would be also worth to indicate and discuss here that observations – multi-model mean differences are larger than inter-model differences.

As stated above, hourly output from the models is not available for the full year or for any full month so it is not possible to calculate the monthly mean standard deviations for the models to compare to the observations. This would be a useful endeavour for a future model study. The standard deviation of the daily O3 data have been analysed and I have added error bars to the surface plots in Fig. 2. I have also changed the figure caption accordingly. As the CO is collected by flasks daily data is not available. I looked at the event data, but this didn't match the monthly mean data provided by NOAA GMD, for this reason I do not show the event data as well as the means.

p. 25293, l.8, Figure 3: This figure is not fully described in the text. I do not understand what do the scatter points represent? I guess that they should correspond to each monthly means, but then, why only 4 scatter points by model run are represented for the CO Taylor diagrams against only 2 scatter points by model run for O3 and not 12 points for both CO and O3? Please explain the figure and describe better the results.

The points correspond to overall performance of each model (according to colour) at each station (a=Alert, b=Barrow, p=Pallas, s=Summit, z=Zeppelin). The colours represent the different models. The caption has been expanded to make this clearer. The description of the Taylor diagram was removed from the main body of text due to co-author comments saying it distracted from the results. This is described in the figure caption.

p. 25293, ll. 15-17: I do not understand this sentence. It would mean that even if large monthly deviations from the annual mean are real and that models capture them, you will get a high RMSD instead of a value close to 0. This would be nonsense. Do you rather mean that the statistic error is weighted by monthly deviations from the observed monthly means? Could you please clarify and reformulate?

As this description is causing confusion, I have changed the caption of Fig. 3 to say: 'mean model bias as a function of the variations around the annual mean'.

I have also added the equation to the caption to show how it is calculated (RMSD= 1/n* $\Sigma((m - \hat{A}\acute{r}m) - (o - \hat{A}\acute{r}o))$).

I have also changed the text on line 15-17 to:

'As this error statistic is also a function of the monthly deviations from the annual mean (see Figure 3 caption), the models with a higher RMSD are the models that do not capture the amplitude of the seasonal cycle well.'

p. 25293, ll. 17-18: I think that figure 4 is not really useful here since you do not discuss the values of the min., max., 25th and 75th percentile except the median biases. Similarly to figure 6 you could add on figure 2 the monthly mean percent bias. Figure 4 should appear later in Section 5.4.1. or reorganize these sections.

I think this is a valid point. However, due to the length of the paper I wouldn't like to increase the size of plots by adding more panels. I also can't reorganise the sections as I would then split up the sections that discuss the variability in the CO/O3 and the

CO-like tracers. I think it's important to discuss the variability in the tracers at the same time as the variability in the simulated model CO and O3 even if it means the reader will need to refer back to an earlier plot.

p. 25294, ll. 26-29: The better agreement of GEOS-Chem with observations at Barrow in March could be an artifact. In contrast with observations, no decrease in ozone in March is simulated by GEOS-Chem. The monthly mean values are underestimated over the year making the increasing simulated concentrations better matching with the observations in March. Such an underestimation is as well observed at Summit. I suggest to rewrite this sentence and to clarify the role of the halogen chemistry in the GEOS-Chem simulations.

This is related to a comment by reviewer 1. I have changed the paragraph to address this :

'At Barrow, observations show a decrease in O3 in March due to well-known halogen-induced ozone depletion events (ODE) at this location (Barrie et al., 1988; Helmig et al., 2007). In contrast, most models simulate increasing concentrations between February and April, most likely due to the lack of halogen chemistry in the models, resulting in low correlations at Barrow. The GEOS-Chem model, which does include halogen chemistry, is still not able to capture the seasonal transition between February and April even though in March the absolute concentrations agree well. This model also simulates O3 that is lower than the other models and observations between late-spring/early-summer, resulting in a larger negative bias against observations compared to other models. It is not clear what is causing this bias in GEOS-Chem, however, it occurs at the same time when simulated CO is higher than simulated by the other models suggesting it is related to the model chemistry, not transport.'

p. 25295, ll. 16-19: Which model has been excluded from the MOPITT – multi-models comparison? Why?

WRF-Chem is not included in any other analysis apart from the POLARCAT aircraft

comparisons (Fig 7 & 8). This is due to the high resolution of the model dictating that it can only be run over selected regions. It was mentioned on page 25289, L11-13:

'The WRF-Chem model was run for the period of the POLMIP summer campaign at two different horizontal resolutions (100 and 50 km) to show the impact of resolution on Arctic simulations of chemistry and transport.'

This sentence has been expanded to try and make this point clearer:

'The WRF-Chem model was run at two different horizontal resolutions (100 and 50 km) to show the impact of resolution on Arctic simulations of chemistry and transport. Due to the computational expense of this model, these simulations were performed over a region and time period covering only the summer POLARCAT flights (using the MOZART concentration fields as boundary conditions outside of this region and period).'

I have also added the following to the Fig. 5 caption: 'The multi-model mean CO is calculated from all 10 global models and the MATCH hemispheric model.'

p. 25296, ll. 26-28: What does the shadow grey area on Figure 6 represent?

I have added the following to the caption: 'The grey shaded area shows 1σ of the MOPITT concentrations over the regions at the relevant pressure level.

p. 25297, ll.23-26: If comparison between the Arctic and the source regions is difficult during winter, for the rest of the year, MOPITT data over the Arctic are characterized by around one full level of information allowing comparisons between total columns measured in the Arctic with source regions. Because of the lack of MOPITT vertical sensitivity in the Arctic, showing the monthly mean MOPITT CO at 700 and 300hPa in figure 6 does not make sense. You should better evaluate total column model performance over main source regions relative to the Arctic.

I did calculate the total columns for Figure 6, but these were not shown as they do not show any additional information. I have added these to supplementary information.

I have edited the manuscript to say: 'The mean DOFS over the Arctic region (see top panel of figure 6) is not much more than 1, indicating the MOPITT retrievals have little vertical information in the Arctic, which is due to the lack of thermal contrast between the ground and the atmosphere. This makes evaluation of model performance in the Arctic more difficult.'

p. 25299, ll. 16-21: It would be worth here to comment and to point probable reasons explaining why all models with lower and higher OH underestimate OH in the lower and the middle troposphere.

I have been careful to not go into too much detail comparing the model and observed OH due to the high uncertainty of the measurements ($\pm$40%). This is why I say the models 'lie within the 25th and 75th percentiles' in the middle troposphere. I have added a few things to this section (ARCTAS-A spring measurements) to try and expand upon this subject.

I have expanded the following sentence to draw attention to the high uncertainty in the measurements: 'The ability of the models to capture this vertical structure is highly varied, with correlations ranging from 0.63 to 0.98. However, OH measurements have a particularly high uncertainty (see Table 2) and the short lifetime of OH makes comparisons with coarse global models difficult.'

Added sentence: 'Emmons et al., (2014) showed that the models with higher OH also had higher photolysis rates of O3 to O1D and that a subset of models (CAM4-Chem, CAM5-Chem, MOZART-4) exhibited a relationship between higher photolysis rates and lower cloud cover fraction. These are the models that show higher OH concentrations in the upper troposphere against the ARCTAS data.'

Also expanded (in bold): 'Model water vapour concentrations show good agreement with the observations apart from in the upper troposphere, where there is evidence that the models overestimate observed concentrations. This may be contributing to a possible overestimation of OH in this region in some models.'

p. 25300, ll. 22-25: It is not possible to distinguish the two WRF-Chem simulations, same color is used. I am now using different colours.

p. 25305, ll. 9-10, Figure 11: Why didn't you show the zonal mean OH from MATCH?

I have added MATCH OH to figure 11.

p. 25306, ll. 11-15, Figure 12: Why is LMDZ-INCA not shown in the correlation plots? For panels c) and d), I would suggest to plot annual mass-weighted tropospheric mean OH concentration in the northern hemisphere excluding the Arctic instead of global mean to investigate drivers in controlling inter-model OH differences over the Arctic and over the source regions.

For LMDZ-INCA, the modelling group didn't provide the model level interfaces so I was unable to calculate the grid box mass, which I need to weight the OH concentrations. This is why I excluded LMDZ from figure 12.

Different latitude ranges were originally looked at (e.g. 0-60N), however, they weren't included in the paper as they didn't offer any additional information. The global ones were included as they can be compared to previous studies. I have added the region 0-90N to supplementary information table.

p. 25307, ll. 20-25, Figure 13: Why didn't you plot seasonal mean differences for GEOS-Chem and TOMCAT?

Figure 13 shows the 25-day fixed lifetime tracers in order to evaluate transport efficiency in the models (this excludes any biases in model chemistry). GEOS-Chem is not shown as these tracers were not provided by this modelling group (see Table 1).

The following has been added to the caption:

'(GEOS-Chem is excluded as the 25-day fixed lifetime tracers experiments were not performed.)'

TOMCAT is included (in red), however, there was an error in the label. This has now

been resolved.

p. 25308, ll. 10-13: Apart from vertical convection, what about the role of vertical resolution in the POLMIP models for explaining a part of the discrepancies in the magnitude of the tracer?

'The models all agree on the overall distribution and seasonal changes, however, there are some clear discrepancies in the magnitude of the tracer concentrations between models. It is likely that these vertical transport differences will play a role in the simulated Arctic concentration differences seen throughout Sect. 4.'

I have added the following after this:

'These differences could be caused by a variety of model differences that can influence the transport of tracers, such as large-scale advection schemes, parameterised convection and the vertical and horizontal resolution of each model. For the models shown in Figure 13, LMDZ and TOMCAT have a lower number of vertical levels below 700 hPa (9 and 8, respectively) compared to the other models (which have either 15 or 16 levels). However, they have very different vertical differences in the 25-day fixed lifetime tracer suggesting other processes apart from resolution are important.'

p. 25308, ll. 20-25, Figures 14 and 15: I do not see contributions from MATCH and GEOS-Chem. Which models do include the six regional tracers and which do not? Figure 13 shows that MATCH includes them all and TOMCAT does not and Figure 14 and 15 shows the contrary, this is confusing. Please correct or clarify. In Figure 15, since all tracers are included within the TOMCAT model, even if GEOS-Chem does not include the regional tracers, you could use its OH field within TOMCAT. Did you exclude GEOS-Chem from the multi-model analysis because of different anthropogenic emission dataset? Please clarify?

See Table 1 for a list of the models that have 25-day lifetime tracers: 'Y in the tracers column indicates that idealised tracers from model have been used for the transport

and chemistry analysis in Sect. 5'

TOMCAT does include the tracers and is included in Fig. 13 but there was a mistake in the label of fig 13. This has now been corrected.

GEOS-Chem didn't provide the tracers for the 25-day lifetime experiment. As I use these fixed lifetime tracers and the OH tracers run within the TOMCAT model for a statistical comparison between transport and OH differences I needed to keep the two samples the same. For this reason I excluded GEOS-Chem from the OH tracer experiments run within TOMCAT. As the GEOS-Chem group also didn't use the same emissions then I don't think it makes sense to include this model for calculating the statistics.

As for MATCH, they did provide the 25-day fixed lifetime tracers so I could show them in figure 13, however, as we couldn't run the OH tracer experiments using only a partial OH field (due to MATCH being a regional model) this model was also excluded from the analysis performed in this Section 5. This was reflected with a N in the table 1. I have changed the caption in Table 1 to reflect this:

'(N.B. MATCH provided the 25-day tracers to be included in Figure 13, however, they were excluded from the analysis in Figure 14-17 as the regional model OH field could not be used in the global TOMCAT model to calculate the OH tracers needed to perform the statistical analysis.)

Table 1: What is the number of vertical levels used in WRF-Chem? Does the model use 27 vertical levels from the surface to 50 hPa similarly to Thomas et al. (2013)? Please insert that in the Table?

The WRF-Chem model was extended to use 65 vertical levels up to 50 hPa for the POLMIP simulations. This has now been added to the Table 1.

Technical corrections: p. 25283, l.26: "at the Barrow." ! "at Barrow." Changed P.25295, l.3: "with Shindell et al. (2008)" ! "with Shindell et al. (2008)." Changed p. 25296,

l.14: "Shindell et al. (2006)," ! "Shindell et al. (2006)" Changed p. 25306, l. 11: "have the the highest" ! "have the highest" Changed p. 25336, Figure 3: Specify that the diagrams represent Arctic surface comparisons "at Barrow and Zeppelin". It seems that the RMSD contours are plotted in blue dotted contours, not black. See earlier comment: I have changed the caption of Fig. 3 to state which stations are used and 'black' has been changed to 'blue' p. 25339, Figure 6: The grey area is not mentioned in the caption/text. Added to caption p. 25340, Figures 7 and 8: Please, use different colors for WRF-Chem_100 and _50 km resolution simulations. This has been changed p. 25345, Figure 12: I do not see the OLS regression black lines in panels a), b) and d). Figure caption/text is not clear and too long: no differences between description of panel a) and panel c), and most of the explanation should be indicated in Section 5.1 instead of including it in the caption. The black line was underneath the red line, I have changed the red line to be dashed so you can see both regression lines. I have shortened the caption for figure 12 and removed the following lines: - 'Removing MOZART-4 from analysis shown in (b) results in a r2 value of 0.91, which is also significant above the 90%CL.' - 'however, results are similar when using a climatological tropopause as described by Lawrence et al. (2001).' - 'The full list of regression coefficients (b) and coefficients of determination (r2) are shown in Table S1.' -> 'See Table S1 for more details.'

Interactive comment on Atmos. Chem. Phys. Discuss., 14, 25281, 2014.