Referee #1

This paper provides an overview of the POLARCAT Model Intercomparison Project (POLMIP) and presents the results concerning ozone and its precursors. The paper is accompanied by two additional manuscripts by Monks et al. and Arnold et al. that contain complementary analysis. The paper addresses the very important question of model comparison focusing on ozone and its precursors in the Arctic region. 9 global and 2 regional models with similar emissions are included in the comparison. Consistent differences exist between different models as well as between models and observations (aircraft observations from POLARCAT mission, ozonesondes, OMI NO2 columns, C2H6 and C3H8 surface measurements). As demonstrated nicely with tracers in the accompanying paper by Monks et al., the differences between models in CO in the Arctic are due mostly to differences in chemistry (OH fields are different) rather than differences in transport from source regions. The paper by Emmons et al. contains an analyses of some of the reasons for the different OH concentrations: different cloud coverage leading to different photolysis rates, differences in H2O mixing ratios. Other possible reasons, such as differences in ozone deposition velocities, are identified but not analysed. The comparisons with observations allow identifying some consistent model biases and lead to suggestions for possible improvements, e.g. emission inventories.

General Comments

This is a well-written article presenting an important inter-comparison of models in the context of a measurement campaign (POLARCAT). Important and interesting insights on the reasons of the inter-model differences are provided.

1) What is not entirely clear is that even though it is stated that the same emission inventories are used in all models, some differences are present. This is due to differences in the chemical mechanisms (some species are not explicitly modelled), but it is not clear if these are the only differences (e.g. the differences in GEOS-CHEM emissions).

All groups were asked to use the same emissions, but the implementation was not perfectly identical. Table 2 was included to document what was actually used in each model.

2) Some of the figures (especially related to the comparison with OMI NO2 columns) are somewhat hard to analyse.

We have increased the font size and increased image sizes for the OMI results and comparisons, along with other improvements in other figures.

3) It could be also helpful if the objectives of this paper with respect to the accompanying papers (especially Monks et al) could be somewhat more detailed in section 1.

We prefer to keep the focus of the introduction on the objectives of this paper. The Monks et al and Arnold et al papers are in the same special issue, so readily accessible to readers.

Some additional suggestions for improving the clarity of the text are provided in the specific comments below.

Specific Comments

Page 29335, line 13: change “slow mixing” to “slow vertical mixing”?
changed

Page 29335, line 18: add that the rapid advection follows isentropic surfaces, which can explain the layering
added

Page 29336, line 17: to be more clear, change “than transport does” to “ than differences due to different transport in the models”.
Changed to “than differences in transport in the models”

Page 29336, line 26: rephrase “following these”. It would be also nice to give the corresponding section numbers.
This paragraph has been re-written to more clearly identify the paper sections.

Page 29342, lines 28-29: Only one version of MOZART-4 results is shown starting with figure 7. Which of these versions is used?
The version with the lookup table is used. This is clarified in the MOZART-4 description.

Page 29345, line 20: I would suggest to change the title of section 4, for example: Overview of model characteristics and of the main model differences.
Good suggestion. Changed to ‘Overview of model characteristics and differences’.

Page 29346: line 21: remove “in” after “agree on”
corrected

Page 29347, line 19: should be Table 2?
Yes, corrected.

Page 29347, line 19: It is not clear why different emissions were used in GEOS-Chem. This is not explained in the model description section.
At the time of running these simulations, GEOS-Chem did not have an easy procedure for reading emissions from external netcdf files, so were unable to use the provided emissions files, but attempted to replicate them in their simulation.

Page 29348: line 18-19: the means were compared?
To determine the biases between model and observations the mean of each altitude bin was used. This has been clarified in the text: “Since the models, with roughly 0.5–1 km vertical layer spacing in the free troposphere cannot reproduce all of the observed structure, the ozonesonde data and model profiles were binned to 100 hPa layers. The mean of each bin between the surface and 300 hPa was used to calculate the bias between model and measurements for each profile.”

Page 29348: lines 23-25- Please indicate how the too strong transport from the stratosphere is consistent with the values lower than observations above 300 mb.

And also is the strong negative bias present for some models above 300mb real? Or is this due to the binning of levels across the tropopause (and the exact location of model levels in this zone of high vertical gradients)?
The figures have been re-plotted to not show values above 300 hPa, as the strong gradient across the tropopause makes the comparison difficult. Text modified, given in previous response.

Page 29349: section 5.2: why is the difference between GEOS-Chem and other models so large for ethane? According to table 2, the differences in total fluxes are small.
Probably due to the much lower OH in GEOS-Chem.

Page 29349: section 5.3, first paragraph: Please specify if total NO2 columns are considered or rather mostly tropospheric columns.
Yes, just tropospheric columns are used. This has been added to the text.

Page 29349: line 25: please explain further the sentence starting with: The averaging kernels of the . . .
An explanation of the need for using averaging kernels has been added. I hope this is what the reviewer meant.

Page 29349: section 5.3, first paragraph: it would be very helpful to have some more information on the relative values of the averaging kernels at different levels. How much weight is on average given to the boundary layer as compared to the free troposphere and/or upper troposphere?
An explanation of the effect of the averaging kernel as applied in the DOMINO product is given in Eskes and Boersma (2003) who show that the range of sensitivities is large between 200 and 1000 hPa, with the lowest sensitivity towards the surface and largest sensitivity in the stratosphere. For instance, depending on the surface albedo the sensitivity may increase by roughly a factor 3 when moving from the boundary layer to the upper troposphere.
We have changed the text as follows:
"The transformation of the model profiles with the averaging kernels gives model levels in the free troposphere relatively greater weight in the column calculation. For instance, depending on the surface albedo the sensitivity to the upper free troposphere compared to the surface layer may increase by roughly a factor 3 (Eskes and Boersma, 2003). This means that errors in the shape of the NO_2 profile can contribute to biases in the total column."

Eskes, H. J. and Boersma, K. F.: Averaging kernels for DOAS total-column satellite retrievals, Atmos. Chem. Phys., 3, 1285-1291, doi:10.5194/acp-3-1285-2003, 2003.

Page 29350: line 6: replace "model bias" by "mean model bias"
changed

Page 29350: lines 17-20. Split the sentence by ending the first sentence after respectively. Please explain how it was decided whether 90% of NOx emissions originate from anthropogenic or biomass burning emissions. This is based on model

inventories used? Also should a threshold be used to filter the data in the plot? The ship emissions do not bring much information to these plots.
Sentence split. Added to the text that the 90% limit is based on the emissions inventory. Fig. 13 has been remade (see new figure below).

Page 29350, lines 23-25 – why not restrict the zone over NW Europe so that it does not include the North Sea? The strong emissions are only over land.
The European region has been moved to over land (see new Fig. 13 below).

Page 29350, lines 25-29: a conclusion seems to be missing. Is my understanding correct: model NO2 columns are underestimated over pollution hotspots, but the spread in model results does not allow to draw conclusions on the quality of the NOx emissions used. Also it would be helpful to understand what weight is given to the boundary layer NOx (see also comment above).
For most regions the model variance is low enough to draw conclusions – the anthropogenic emissions are too low and the fire emissions are too high. Sentence added:
"Since all models used the same NO emissions, the large variation between models (as seen in Figure 12) indicates differences in the rapid photochemical cycling between NO and NO_2 in the boundary layer."

Page 29351: lines 9-10: Split the sentence: Figure 15 shows . . . of the campaign. The flight tracks have . . .
Don't see that is needed.

Page 29351: line 25: add that data were binned in 1 km bins.
Added.

Page 29351: line 26: how was the measurement uncertainty calculated for the binned median values?
Sentence changed to: "The thick error bars represent the measurement uncertainty (determined by applying the fractional uncertainty reported in each measurement data file to the median binned value), while the thinner horizontal lines show the variation (25^{th} to 75^{th} percentile) in the observations over the flights."

Page 29352: line 6-7: the sentence starting with 'In the cases...' should probably be moved to the end of the previous paragraph. In Fig 19, only one measurement is considered. Also it might be helpful to add, that "more than one measurement" indicates measurements with different instrument/technique.
No, Fig. 19 shows the average of all measurement techniques for each compound. The observations from each instrument are shown separately in Figs. 16 and 18. Sentence re-written: "In the cases where a compound was measured by more than one instrument, the differences between the model and each observation were averaged over all the measurement techniques."

Page 29352, lines 10-11: I would not put the sentence 'consistent with' between parentheses: this is an important statement.

Parentheses removed.

Page 29352, lines 16 : after 'boundary layer' could add '(ARCTAS-B, fig. 18).
Added.

Page 29352, line 20 : missing 'some' before cases.
Added.

Page 29353: lines 16-18: the fine structure in OH is due mostly to what sub-grid processes? Cloud distribution and resulting photolysis? Fine scale plumes?
We believe clouds and photolysis are most frequently the cause of OH structure. Sentence re-written: "The distribution of OH is strongly affected by clouds and their impact on photolysis, which coarse-grid models cannot be expected to reproduce, however, these differences are likely averaged out in the binned vertical profiles."

Page 29354: line 17: the exact location and strength of model sources could be also added to the list of reasons why the modelled and observed plumes do not overlap in space and time.
Added.

Page 29354: lines 19-22: were all pixels in the selected zones affected by fires? If not, why not take only pixels inside the plumes from fires? Will taking pixels not in the plumes affect the calculated slopes?
No, not all grid points on each day are affected by the fires. We neglected to say the calculated slopes were determined for points with CO > 150ppb, and thus the non-fire points are ignored. The text has been updated to add this point.

Page 29355: line 9: "fire emissions" or should it be "fire emission factors"? Not having compared CO in these airmasses (observations vs plumes), it seems that one cannot extend the conclusions to emissions but should talk only about emission factors.
Quite true. 'emissions' has been changed to 'emission factors'.

Page 29355: line 17: as in the previous comment. Should it be emission factors rather than emissions?
Yes, corrected.

Page 29356: line 26: replace "ozonesondes" by "ozone"
Changed.

Figure 8: The station Narragansett seems not to be used in later figures (9 and 10)
Narragansett had much less frequent sampling (4 sondes in April, 3 in June-July), so was not used. An explanation has been added to the text

Figure 9-10: Comments on these figures: showing both: the individual measured profiles and mean biases on the same figures does not really facilitate the analysis of these figures. Can they be split to show the mean profiles (observations + models) and biases separately? This would also allow zooming in on the biases. It would also be nice to have in these figures some information on the standard deviations. It is understood that it would be hard to include this information for these plots for all

models at the same time. But it could be presented at least for the observed data instead of showing the individual ozonesonde profiles. Showing individual observed profiles probably does not bring much useful information compared to mean + standard deviation. Showing mean + standard deviation would however simplify the figures.
The ozonesonde plots have been remade, replacing the individual sonde profiles with the mean and standard deviation for each site. (see below)

Figure 13: indicate in the legend that left column for April and right for June-July.
Added to caption.

Figure 14: indicate in the legend that the box plots show model results! They include data from all models?
The caption stated "Summary of model mean…", but has been re-written. Observations are also shown.

Figure 15: It seems that Grace and POLARCAT France measurements are not used in this paper. Is there a reason for this? This could be mentioned in the legend. Please add also AP next to ARCPAC.
The GRACE and France experiments had limited gas-phase measurements available so added little to the conclusions already presented. So in the interest of clarity and space they were not included. AP is defined in the Figure 19 caption.

Figures 16-18: not easy to distinguish between thick and thin error bars? Use error bars with vertical lines? Also replace ARCTAS-A by ARCTAS A1? WRF can be removed from the legend in figure 16?
Figures 16-18 have been updated.

Figure 17: nomenclature: "P3" not used in figure 15, for clarity might be helpful to remove it.
OK.

Figure 19: why OH was not included?
Limited space.

Table 2: Change title from Emissions to Global emissions. Indicate in the legend that the regional models were not listed, as the global values cannot be provided.
Changed.

New Figures

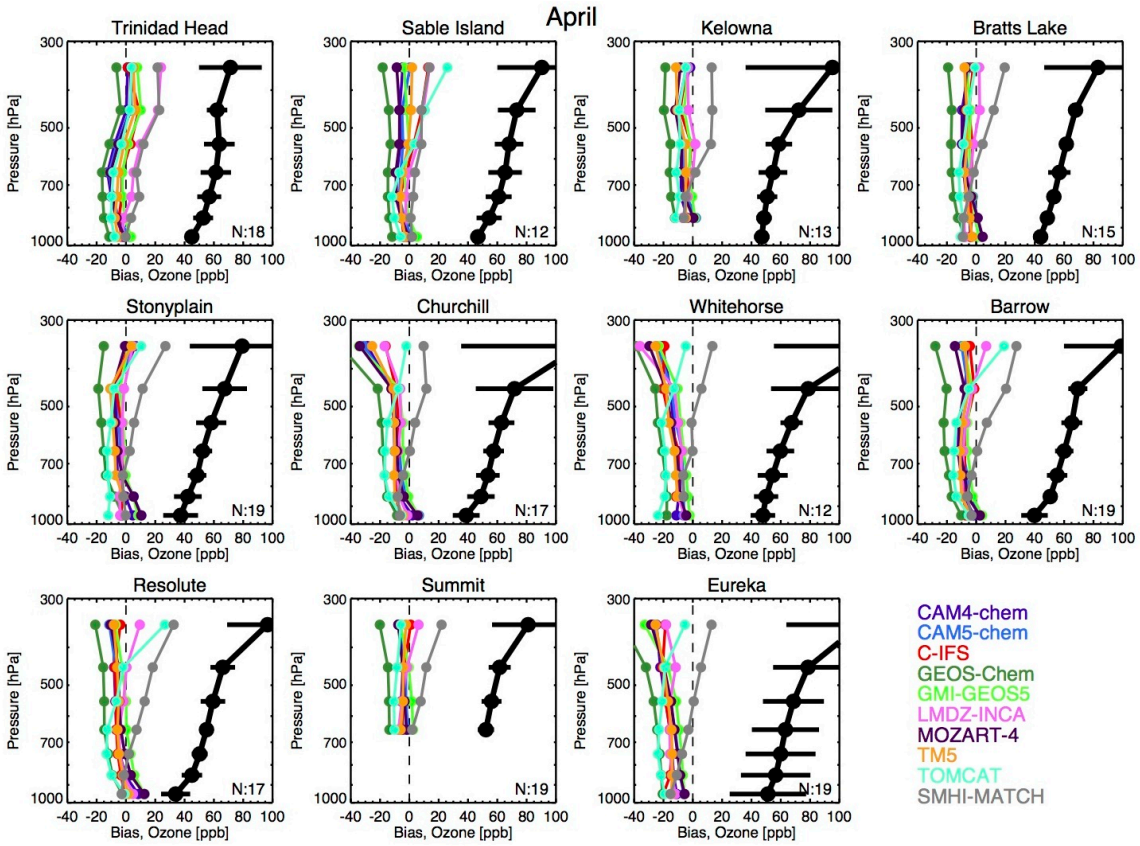

Figure 9. Comparison of models to ozonesondes for April, showing mean and standard deviation of the observations (black line) and the mean bias (colored lines) for each model at each site (Tarasick et al., 2010; Thompson et al., 2011). Results shown for only surface to 300 hPa for clarity. The number of sondes for each site is indicated in the lower right corner of each panel.

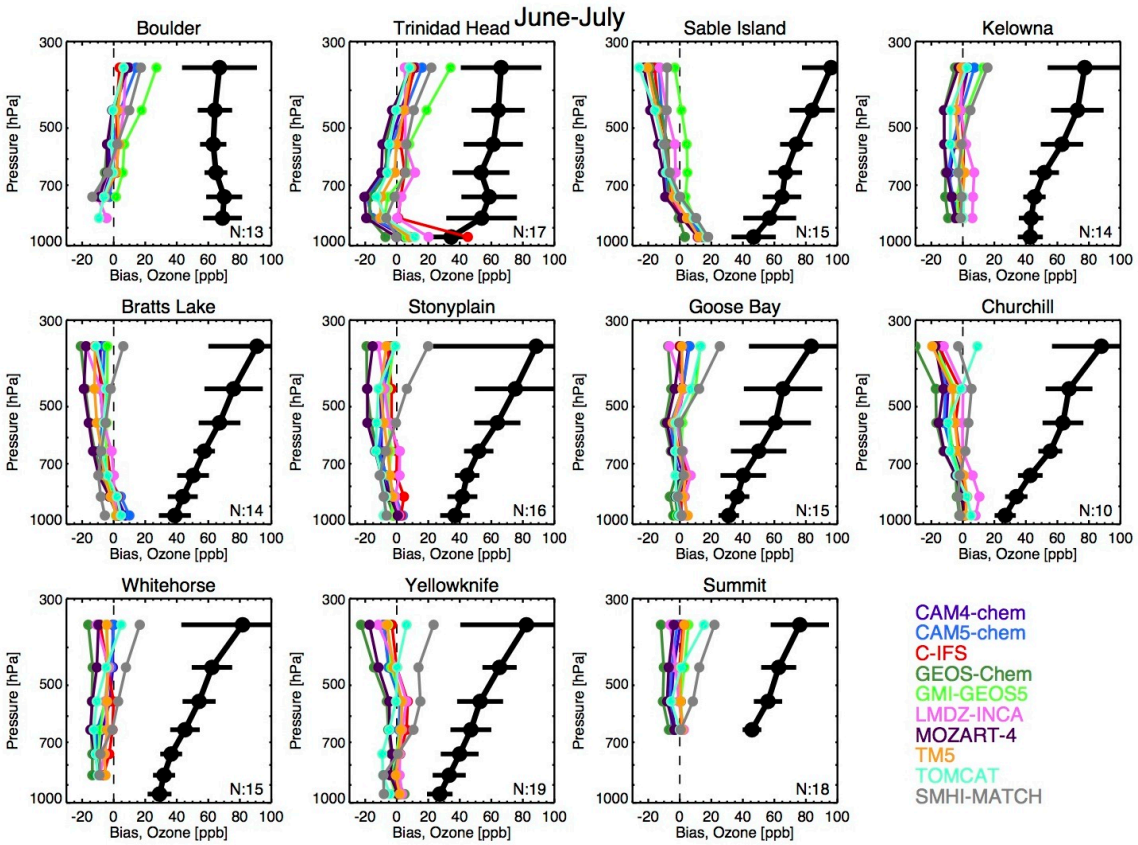

Figure 10. As Fig. 9, but for June–July.

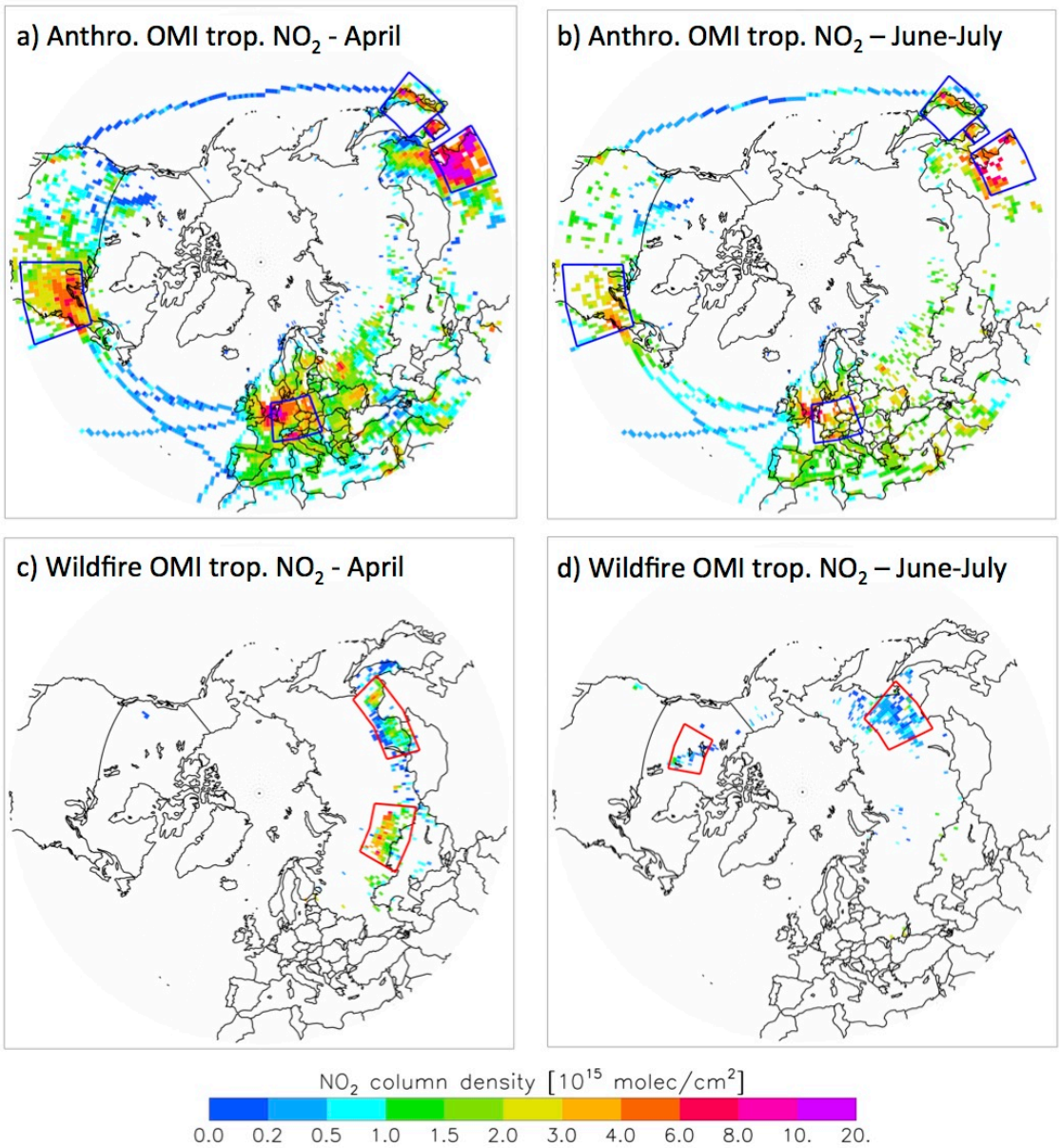

Figure 13. OMI NO_2 filtered for dominant anthropogenic emissions **(a, b)** and fire emissions **(c, d)**. Boxes indicate the regions for which biases have been calculated in Fig. 14. Panels **a, c** show April, panels **b, d** show June-July.

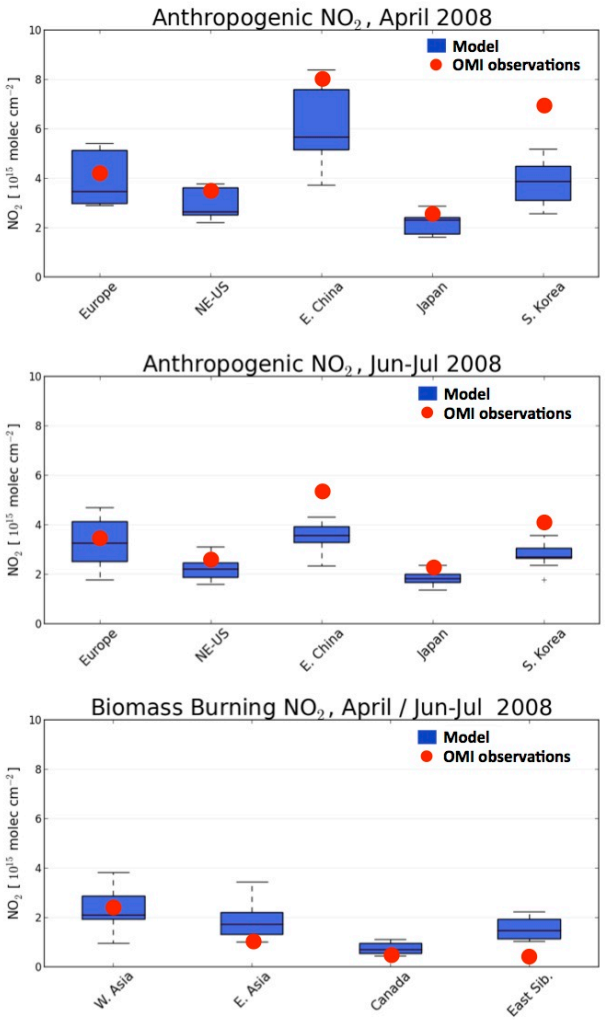

Figure 14. Summary of the regional means from each model and the OMI NO_2 tropospheric columns for each region indicated in Figure 13. (a) Anthropogenic emissions in April and (b) June–July, (c) biomass burning in both seasons. Red circles are mean OMI NO_2 observations for the region; box plots show median, 25th and 75th quartiles, whiskers to 5th and 95th percentiles of the model means.

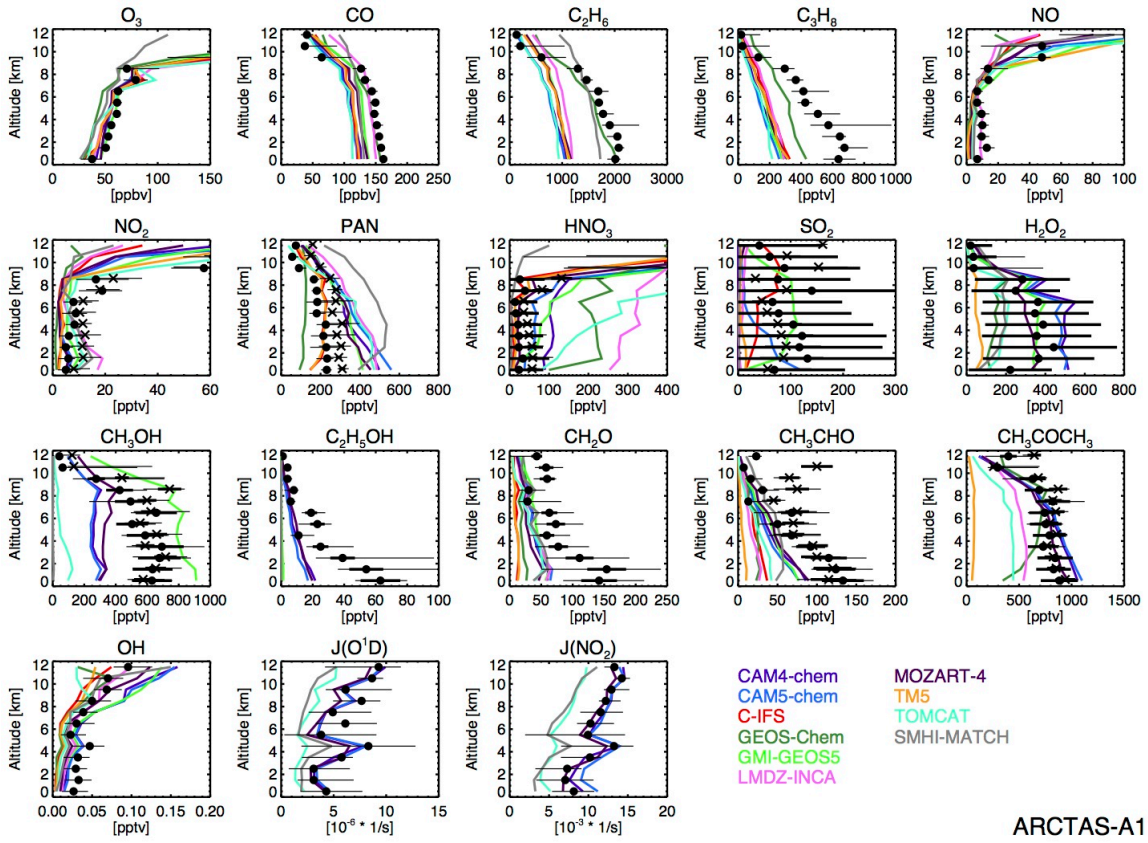
O3
CO
C2H6
C3H8
NO
NO2
PAN
HNO3
SO2
H2O2
CH3OH
C2H5OH
CH2O
CH3CHO
CH3COCH3
OH
J(O1D)
J(NO2)
Altitude [km]
[ppbv]
[pptv]
[10-6 * 1/s]
[10-3 * 1/s]
CAM4-chem
CAM5-chem
C-IFS
GEOS-Chem
GMI-GEOS5
LMDZ-INCA
MOZART-4
TM5
TOMCAT
SMHI-MATCH
ARCTAS-A1

Figure 16.

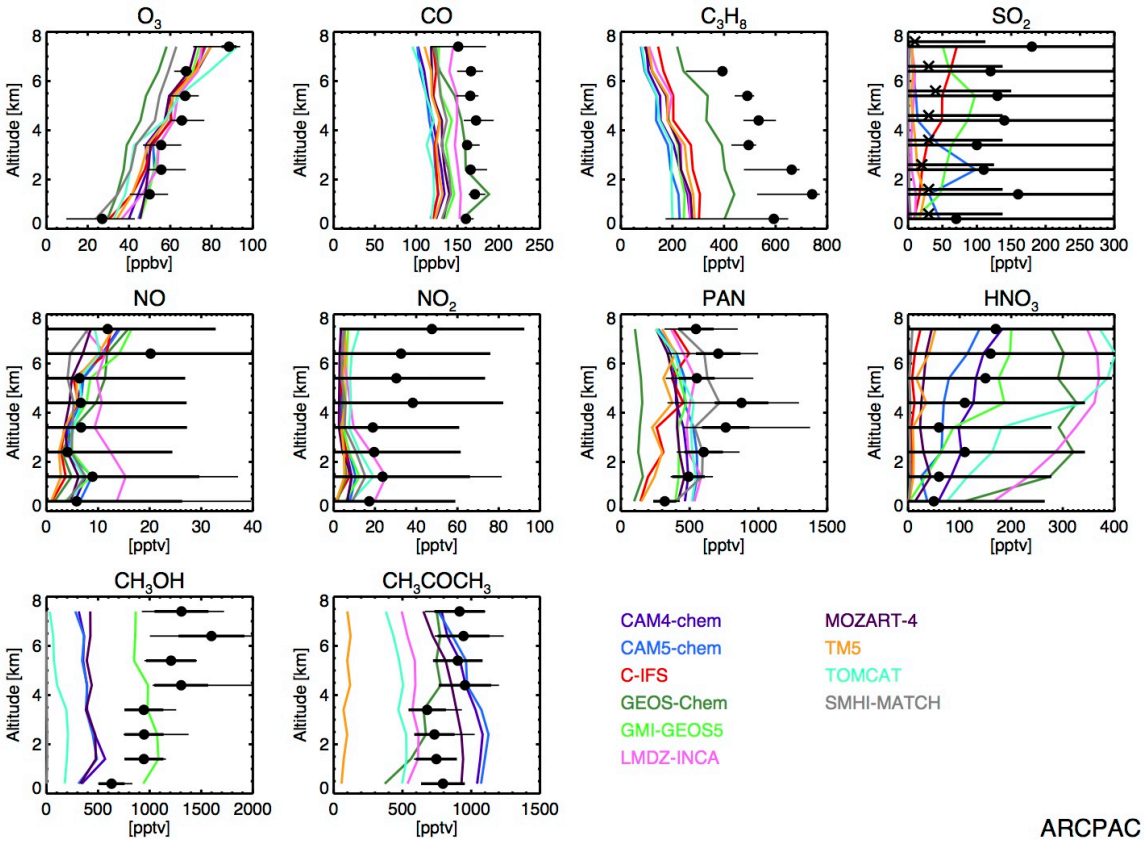
O3
CO
C3H8
SO2
NO
NO2
PAN
HNO3
CH3OH
CH3COCHO
Altitude [km]
[ppbv]
[pptv]
CAM4-chem
CAM5-chem
C-IFS
GEOS-Chem
GMI-GEOS5
LMDZ-INCA
MOZART-4
TM5
TOMCAT
SMHI-MATCH
ARCPAC

Figure 17.

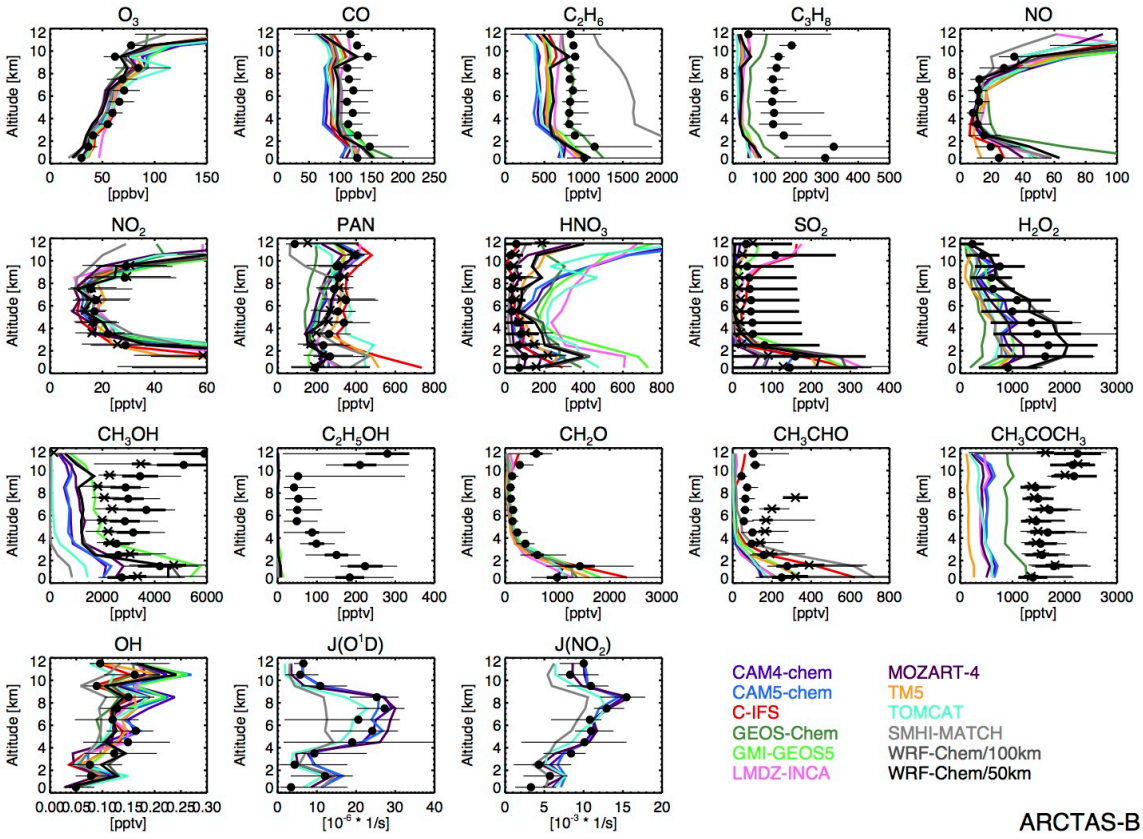
O3
CO
C2H6
C3H8
NO
NO2
PAN
HNO3
SO2
H2O2
CH3OH
C2H5OH
CH2O
CH3CHO
CH3COCH3
OH
J(O1D)
J(NO2)
Altitude [km]
[ppbv]
[pptv]
[10-6 * 1/s]
[10-3 * 1/s]
CAM4-chem
CAM5-chem
C-IFS
GEOS-Chem
GMI-GEOS5
LMDZ-INCA
MOZART-4
TM5
TOMCAT
SMHI-MATCH
WRF-Chem/100km
WRF-Chem/50km
ARCTAS-B

Figure 18.