Referee #2

We thank the reviewer for his/her generally favorable view of the paper along with very detailed comments. After removing the duplicated sections of the review due to an apparent error in copy and pasting in the submission of the review, we have reproduced the review below, and included our responses (in bold).

GENERAL REMARKS

Analyzing the content and the context of the manuscript, I would say the study described in the manuscript has fulfilled 3 objectives: 1) Introduction of the models which are used in POLMIP; 2) Showing a comparison of the models with observations in the Arctic, and make conclusions about how good the models reproduce the observations; and 3) Conclusions about the emission data set.

I think the manuscript does a valuable job in comparing the results of model simulations with observations in the Arctic. It is important to estimate the reliability of CTMs in this remote but vulnerable region, which is impacted by pollution from different origins. Therefore, it is very nice that the study uses for this the extensive data set obtained during the International Polar year 2008. The manuscript is well written and agreeable to read.

However, the paper lacks analysis to attribute the differences between model results and observations to specific model components. It does it for some model components (complexity of the chemistry scheme to some degree, photolysis rates and cloud fields), but other possible contributing factors are only mentioned but have not been investigated (impact of dry/wet deposition, tracer transport scheme, convection and boundary layer parameterisations, vertical resolution in lowest 5 km, reduced grid near the poles, inclusion of stratospheric chemistry, ...). In the abstract is mentioned : "to quantify the differences in model chemistry and transport schemes." I read this as a focus of POLMIP (broader than this manuscript), and not only of this study. However, it gives the suggestion that it might all be investigated in this study. The abstract continues : "Differences in a number of model parameters are identified as contributing to differences in the modeled chemical species, including cloud fields and photolysis rates." But this makes arise the question : which others have been identified? Also, the analysis of cloud fields and photolysis rates feels a bit limited to an illustration. Further, in the conclusions is written : "However, numerous differences occurred among the model outputs due to the different chemical schemes and physical parameterizations such as convection, boundary layer mixing and ventilation, wet and dry deposition." However, at the end I have the impression that the contribution of these different parameterisations to the actual model differences is not quantified. Finally, in the first paragraph of the conclusions is already written : "Additional model diagnostics

are required to completely understand the differences among models." I think such suggestions should be made at the end of the conclusions.

I write this because, at the moment, as a reader one gets the impression that certain things will be investigated or analyzed, but in the end they are not. It would therefore be nice if the analysis could be improved and extended. What type of model components play a large role in the modeling of the chemical composition of the Arctic atmosphere? If this extension is not possible, then it should be explained why only the clouds and photolysis rates are explicitly illustrated. In that case also, make the text more clean and sober: focus only on the aspects you really investigate in the study. Just mention the ones you don't investigate (and possibly estimate their relevance based on other studies for the Arctic), but state from the beginning that they are not the focus of the study. Therefore the abstract, Section 4, and the Conclusions (Section 7) should be thoroughfully rewritten.
Attempts have been made to revise the abstract and conclusions so as to clarify the limited set of comparisons and evaluation performed in this paper. Discussion of additional diagnostics that would have been useful, etc., have been moved to the end of the Conclusions.

There are valuable conclusions about the emissions data set. However, these conclusions only have a value if this is a publicly available dataset, and which is intended to be used by others, or which at least can be investigated. Although the webpage http://bio.cgrer.uiowa.edu/arctas/emission.html was accessible, further links which would guide to the data did not work. The links on that web-page which did not work were : (i) "This inventory is available for download", and (ii) "Gridded emissions can be accessed at the University of Iowa ACESS web- site http://www.cgrer.uiowa.edu/EMISSION_DATA/index_16.htm, under the direction of Gregory Carmichael, Principal Investigator of this project."
We believe that this emission inventory is intended for use by any who are interested. It is freely available (without password control) from a website. It was developed for the ARCTAS and POLARCAT modeling groups, and anyone else interested. I don't believe the University of Iowa intends to prevent access to the inventory by having broken links and we have contacted them to ask that the links be repaired.
The direct link to the data is:
http://bio.cgrer.uiowa.edu/arctas/arctas/07222009/.

Further, I think the manuscript contains too many figures.
We believe the figures presented are all useful. The other reviewer suggested a number of improvements to the figures (which we have made), implying they saw value in all of the figures.

The structure of the paper could be improved. Now the Sections 4, 5, and 6 all contain results, so they might be brought together (as Sections 4 and 6 are short). Section 3 should have a more general title.
Sections 4, 5 and 6 were separated to highlight the rather different results

being presented. Section 3 is renamed: Model configurations and inputs. We believe this organization of the document helps the reader.

Below you can find : (i) a list of smaller general remarks, (ii) a list of detailed comments on the manuscript text, (iii) a list of detailed comments on the tables, and (iv) a list of detailed comments on the figures
Each of these comments is addressed below and in the revised manuscript.

SMALLER GENERAL REMARKS

1. Figures There are too many figures. In some figures, too small fonts are used. Some plots are too small. The layout of some of the figures should be improved.
We find all of the figures valuable (as addressed above) and have improved the layout and font sizes.

2. For MOZART, 2 things are unclear : (i) which photolysis rate calculation is actually used for the principal simulations? (ii) how are the clouds and the humidity calculated? Is CAM (nudged to GEOS-5) also used?
(i) The lookup table for photolysis is used for most simulations (this is clarified as explained below). (ii) MOZART-4 CTM contains a cloud parameterization, as described in the MOZART description papers. It is completely unrelated to the CAM simulations. CAM4-chem and CAM5-chem simulations for POLMIP are nudged with GEOS-5 meteorology.

3. Maybe mention a bit more explicitly which papers are companioning papers within POLMIP. I presume that Monks et al. [2014] and Arnold et al. [2014] are, but this is not explicitly said. Are there others?
These are the only 3 POLMIP papers to date and it is unlikely there will be others.

4. I would consequently use "wildfire" instead of "fire".
The FINN fire emissions include all open burning - both wildfires and prescribed agricultural burns that are detected by satellite, thus, we prefer to use the more general term.

5. Abbreviations Be consequent in how abbreviations are defined : first the long expression, followed by the abbreviation between brackets. It is probably not possible to always follow this rule, but try to be as consequent as possible. What strategy is followed for the first letter in the full expression : CTM (small letter in full expression), GCM (big letters in full expression), SLCP (small letters in full expression)?
We have tried to modify the text to follow this rule.

6. Homogenize the description of the different models (see DETAILED COMMENTS on TEXT)
Answered below.

7. LMDZ-INCA, LMDZ both are used in the text. Try to just choose one.
Corrected to LMDz-INCA.

DETAILED COMMENTS on TEXT

Please find below a list of detailed comments. Among these, there are quite some comments and suggestions about language use. Feel free not to follow these suggestions on language, but please give a good reason why you do not do so.
We have accepted the suggestions where they improve the precision of the grammar and readability, however, in some cases we kept our original language as we felt it conveyed our meaning more clearly.

p 29332 : Norrkoping -> Norrk\"oping (\" on the o)
Corrected

p 29332 : University Innsbruck -> University of Innsbruck
Corrected

p 29333, l 2-3 : "atmospheric chemistry observations" : this is a bit vague.
Change to "observations of atmospheric composition"

p 29333, l 5-11 : 5 times "differences" on only a few lines
rewritten

p 29333, l 15-23 : the order of these last three sentences is a bit strange. First, a sentence about usefulness of aircraft observations without conclusions. Second, a sentence about satellite observations with conclusions. Third, a sentence about aircraft observations (although) with conclusions. Maybe sentence one and three can be put together.
rewritten

new Abstract:
A model intercomparison activity was inspired by the large suite of observations of atmospheric composition made during the International Polar Year (2008) in the Arctic. Nine global and two regional chemical transport models participated in this intercomparison and have performed simulations for 2008 using a common emissions inventory to assess the differences in model chemistry and transport schemes. This paper summarizes the models and compares their simulations of ozone and its precursors, and presents an evaluation of the simulations using a variety of surface, balloon, aircraft and satellite observations. Each type of measurement has some limitations in spatial or temporal coverage or in composition, but together they assist in quantifying the limitations of the models in the Arctic and surrounding regions. Despite using the same emissions, large differences are seen among the models. The cloud fields and photolysis rates are shown to vary greatly among the models, indicating one source of the differences in the simulated chemical species. The largest differences among models, and between models and observations, are in NO_y partitioning (PAN vs. HNO_3) and in oxygenated volatile organic compounds (VOCs) such as acetaldehyde and

acetone. Comparisons to surface site measurements of ethane and propane indicate that the emissions of these species are significantly underestimated. Satellite observations of NO_2 from OMI have been used to evaluate the models over source regions, indicating anthropogenic emissions are underestimated in East Asia, but fire emissions are generally overestimated. The emission factors for wildfires in Canada are evaluated using the correlations of VOCs to CO in the model output in comparison to enhancement factors derived from aircraft observations, showing reasonable agreement for methanol and acetaldehyde, but underestimate ethanol, propane and acetone, while overestimating ethane emission factors.

p 29334, l 3-5 : Two aspects are mentioned ((i) heat transport from lower latitudes, (ii) local radiative forcing). The second one is really about climate change, while the first one is even true without climate change. Try to formulate more precisely.
Sentence changed to: "Arctic temperatures are affected by increased heat transport from lower latitudes and by local in-situ response to radiative forcing due to changes in greenhouse gases and aerosols."

p 29334, l 8 : aerosol -> aerosols (as on line 5)
Changed

p 29334, l 27-28 : twice "significant" in the same sentence
Changed

p 29335, l 4 : "it" refers to "Arctic" I presume. It would be clearer to explicitly write it.
Corrected

p 29335, l 12-16 : maybe add "only" before "in close proximity", and "mainly" before "retaining only". Otherwise the last 3 parts of the sentence do not fit together well.
Changed

p 29335, l 21 : "slow" before "mixing"?
Changed

p29335, l22: and more->or more
Changed

p 29335, l 25 : "Climate" -> "of Climate" in the definition of POLARCAT
Corrected

p 29336, l 5 : "focused" -> "focuses"
Corrected

p 29336, l 5 : "this comparison" : does this refer to the whole POLMIP, or just to this paper? Is POLMIP limited to gas phase chemistry evaluation?
The comparisons performed within POLMIP are primarily of gas phase compounds. Text clarified.

p 29336, l 9-10 : "evaluate ... with ... observations" : is this correct language use?
Kept original language.

p 29336, l 13-14 : "by methyl chloroform observations and emissions" : maybe "by methyl chloroform observations and its emission estimates"
Changed

p 29336, l 17 : "than transport does" -> "than differences in transport"

Changed to "than differences in transport in the models"

p 29336, l 23 : "efficiency in" -> "efficiency of"

Changed

p 29336, l 24 - p 29337, l 2 : I would suggest to add section numbers in this paragraph

p 29336, l 26 : "of all of the models to observations" -> "of model results with observations"

Changed

p 29336, l 26-28 : This sentence is a bit strange as an ozonesondes is more an instrument, while "NMHC" and "compounds" are species. I would suggest to homogenize the sentence.
Changed

p 29336, l 28 : "emissions" -> "the emissions"
not changed

p 29337, l 4 : "collaboration of experiments" : is this the correct description?
"Consortium" might be better.

p 29337, l 12-13 : "of each mission" -> "for each mission"
Changed

p 29337, l 13-14 : "transported to the Arctic" : shouldn't it be after "wildfire plumes" (if the measured wildfire plumes are not the ones going to the Arctic, than the sentence is correct I think)
Corrected

p 29337, l 17-18 : abbreviation precedes the full name, while different for ARCTAS
Changed – ARCPAC placed in parentheses.

p 29337, l 20 : is the word "sources" needed?
'sources' removed.

p 29337, l 21 : fire plumes -> wild fire plumes
Changed

p 29337, l 24 : "Spring" -> "spring", maybe cancel "in spring" as the dates make this clear (between 30 March and 11 April).
Removed 'in spring'.

p 29338, l 1 : shouldn't GRACE be explained : Greenland Aerosol and Chemistry Experiment?
Yes, thank you.

p 29338, l 2 : "Greenland" maybe not needed as already mentioned on page 29337, l 25. On the other hand, it is maybe good to repeat it.
Not changed

p 29338, l 5 : here again the explanation follows the abbreviation. Maybe it is unavoidable due to the YAK-part.
Not changed

p 29338, l 10 : the section "3 Models" seems to be more general than just about models. Another title might be more appropriate.
True. Section title changed to 'Model configurations and inputs'

p 29338, l 12 : twice "output" -> maybe change the second one into "monthly mean species distributions and diagnostics"
Good suggestion.

p 29338, l 14-15 : there is some tension between "All the models" and "with a few exceptions". Maybe change in "Most of the models"
Changed to: "A single emissions inventory was specified for use by all of the models"

p 29338, l 16 : "global" - except WRF?
Re-worded: "Each global model was run at its standard resolution …"

p 29338, l 16 : "meteorology" is vague; maybe "meteorological forcing". In addition to differences in "chemistry scheme", "meteorology", and "deposition schemes", the models probably also differ in vertical distribution of BB emission distribution, tracer transport schemes, detrainment/entrainment rates in convection, etc.
True. 'meteorology and deposition schemes' changed to 'meteorological forcing and other parameterizations'.

p 29338, l 17-18 : twice "output", and "a number" should be avoided. Maybe : "... included monthly mean distributions of mixing ratios and some other diagnostics"
Changed to 'monthly mean species distributions and diagnostics'

p 29338, l 18 : "evaluation" -> "an evaluation"

not changed

p 39338, l 20 : "a smaller number" -> "a limited number"
not changed

p 29339, l 2 : www.ceip.at -> http://www.ceip.at as on p29335, l 28
corrected

p 29339, l 3-5 : maybe the sentence can be rewritten to have "speciation" only once
First 'speciation' changed to: 'emissions for specific hydrocarbons'

p 29339, l 4 : is in this paper VOCs used to describe the same set as NMHC (defined on p 29336, l 27)?
No, NMHC generally refers to alkanes, alkenes, etc. (compounds with only carbon and hydrogen atoms) while VOCs includes oxygenated hydrocarbons (NMHC plus acetone, formaldehyde, acetaldehyde, etc., etc.). Thus, it seems useful to make the distinction.

p 29339, l 6-8 : "provided daily" : maybe also mention that they are given to/used in the models on daily basis too.
Changed to start sentence with 'Daily biomass burning …'

p 29339, l 7 : INventory to agree with FINN (maybe put N as a capital letter)
Must have been changed by the typesetter. Corrected.

p 29339, l 10 : "these" -> "the ARCTAS", because "these" is confusing as different emission datasets are mentioned just above
Changed.

p 29339, l 11 : "showed" -> "showed that"
not changed.

p 29339, l 17 : "usual" -> "standard"
not changed.

p 29339, l 19 : "intercomparison" -> "POLMIP intercomparison" to make clear that it is not just about this paper
Changed.

p 29339, l 19 : "dynamics" -> "tracer transport"
Changed.

p 29339, l 26 : "far removed" -> far away
I like my words.

p 29339, l 28 : "This offset in location produces differences in atmospheric composition" : I think I understand what is meant, but it should be expressed more clearly

Re-written: Since anthropogenic and biomass burning emissions have different relative amounts of CO, NOx and VOCs, the offset in location of the two source types leads to significant differences in atmospheric composition within these regions. These differences have particular relevance in the analyses of Monks et al. and Arnold et al. that use these tracers.

p 29340, l 4 : "March through August" -> looking at Fig. 1, I would rather say "March through July"
Yes.

p 29340, l 10-11 : meteorology -> "origin of meteorological data"
Changed to: 'origin of meteorological analyses'

p29340,l12: Table1->Table2
Removed that sentence (about lightning emissions, as it was mentioned above).

p 29340, l 27 : "integrated forecasting system" -> "Integrated Forecasting System"
Corrected

p 29340, l 27-28 : I would add "(ECMWF)" after "European Centre for Medium Range Weather Forecasting" (it is later used, e.g., on p 29342, l 20)
Added

p 29341, l 4-5 : is it necessary to mention "and applies the emission and dry deposition fluxes as part of the vertical diffusion scheme"?
No. removed.

p 29341, l 5-6 : is this the best way to describe this?
It is technically correct.

p 29341, l 2 : CTM is defined here while "chemical transport model" is already used earlier (e.g. p 29334, l 27). The definition should be given there.
Changed.

p 29341, l 7 : "model convective precipitation" -> "convective precipitation"
Not changed as it is important to emphasize the lightning parameterization is not based on observations and is consistent with the model physics.

p 29341, l 7 : "the C-shaped profile" -> "a C-shaped profile"
Not changed. A specific profile is given in Pickering et al.

p 29341, l 10 : 3-D is defined later (page 29343, l 26)
It seems obvious that these are all 3D models, so removed in both places.

p 29341, l 19 : "includes an" -> "includes "
ok.

p 29341, l 24 : GMI : abbreviation first
not changed.

p 29341, l 25 : I would replace "chemical transport model" by "CTM"
ok.

p 29341, l 27-28 : is it necessary to mention "with all the emissions from the specified inventory". I would rather only mention the exceptions to this rule.
Removed.

p 29342, l 1 : "several" : are there more than these 2? If so add "e.g.," at the beginning, or ", ..." at the end.
Added e.g.

p 29342, l 5 : "but" -> "and"
not changed.

p 29342, l 9 : LMDz should be after the full expression
changed.

p 29342, l 9 : the definition of GCM, should not use capital letters to be in agreement with the definition of CTM : so general circulation model
changed.

p 29342, l 11 : ORCHIDEE should be after the full expression
kept the original.

p 29342, l 16 : "gas phase" versus "gas-phase" (both used in the text)
2 words replaced with hyphenated.

p 29342, l 18-19 : I would write "sulfate" instead of "sulfates"
changed.

p 29342, l 20 : "6h" too cryptic -> "6-hourly"
changed.

p 29342, l 23 : "global annual total" -> "global annual total emission"
changed to 'annual lightning emissions total'

p 29342, l 24 : abbreviation before long expression
not changed.

p 29342, l 25 : use the abbreviation CTM (because it is introduced earlier on p 29341, l2)
changed.

p 29342, l 26-28 : should it be mentioned that the specified emissions have been used?

Replaced 'For these simulations all of the emissions were from the specified inventory (i.e., online MEGAN was not used for biogenic emissions).' with 'While MOZART-4 includes the capability to calculate biogenic isoprene and terpenes using the MEGAN algorithms, the specified monthly mean emissions were used for POLMIP.'

p 29342, l 28-29 : which one is actually used for the analysis here? In Fig. 4, they are presented both: but what with the rest of the analysis?
Generally the LUT run was used. Added at this point: 'Unless otherwise stated, the results shown here are from the LUT simulation.'

p 29343, l 3-5 : this sentence is identical to the one for GMI, except for the list of oxygenated hydrocarbons. Maybe try a slightly different formulation.
Sentence re-written.

p 29343, l 5 : "is the same"; but the CAM-chem explication mentions stratospheric chemistry"? Are the aerosols identical?
Re-written to clarify that the tropospheric gas-phase chemistry is the same in MOZART-4 and CAM-chem (there are other differences).

p 29343, l 9 : abbreviation before definition
not changed.

p 29343, l 16 : "It includes ..." -> "In total, the TM5 chemical mechanism ..."
changed

p 29343, l 18-19 : is it worth mentioning this specificity, as other models might also do it (but just don't mention it in their description)?
True. Sentence removed.

p 29343, l 22 : "as" -> "than"
changed

p 29343, l 23 : "by a fourth-order polynomial function" -> "as a function of"
not changed

p 29343, l 20-25 : is the lightning parameterization of TM5 the same as the one of C-IFS? But apparently different aspects are stressed.
The lightning parameterization in C-IFS as applied for the POLMIP evaluations was in fact very similar to what was applied in TM5, except for a different scaling factor between convective precipitation and lightning flashes. Slight changes have been made to the C-IFS description: "Lightning emissions in C-IFS are based on the model convective precipitation (Meijer et al., 2001) and use the C-shaped profile suggested by Pickering et al. (1998), and follows the same implementation as TM5, except that the lightning emissions are scaled to give a global annual total of 4.9 Tg N yr^{-1}."

and the TM5 description:
"NO_X production from lightning is calculated using a linear relationship between lightning flashes and convective precipitation (Meijer et al., 2001), using a C-shaped profile suggested by Pickering et al. (1998). …"

p 29343, l 25 : has TOMCAT a name definition? 3-D should be defined earlier.
No definition for TOMCAT. 3-D removed (see above)

p 29344, l 1 : "Extended Tropospheric chemistry" : shouldn't chemistry start with a capital letter?
changed

p 29344, l 3 : aerosol -> aerosols
not changed

p 29344, l 5 : the ";" between the references -> "and"
changed

p 29344, l 9 : full expression for MATCH?
"Multiple-scale Atmospheric Transport and Chemistry Modeling System" added.

p 29344, l 9 : chemistry transport model -> CTM
changed

p 29344, l 16: "in (Andersson et al. 2007)" -> "in Andersson et al. (2007)"
changed

p 29344, l 16 : "evaluation" -> an "evaluation"
not changed

p 29344, l 17 : this should not be "we" as it does not refer to all authors
changed to '… data … were used …'

p 29344, l 20 : I would add "only" after "but"
changed

p 29344, l 20 : "reaching about 16 km" -> "reaching about 16 km high"
changed to 'reaching to about 16 km'

p 29344, l 21 : I would skip "In addition to the standard daily POLMIP emissions"
not changed

p 29344, l 23 : "annual global total" -> "annual global total emission"
not changed as it is clear the sentence is about emissions.

p 29344, l 23-25 : why specifically mentioning the DMS emissions?

Because they are different from the specified inventory.

p 29344, l 25 : Maybe something can be mentioned about the boundary conditions for tracer concentrations at the 20N boundary?
Added: "Monthly average results for 2007 and 2008 from global model runs using MOZART at ECMWF in the MACC (Monitoring Atmospheric Chemical Composition) project were used as both upper and 20N chemical boundary conditions."

p 29344, l 26 : has WRF a full name?
Full name added: The Weather Research and Forecasting model with Chemistry

p 29345, l 5 : fire -> wildfire
not changed – FINN includes all open burning, including prescribed agricultural fires

p29345,l5: 1->1x1
not changed. Seems clear to me.

p 29345, l 13 : GOCART (definition after abbreviation)
not changed

p 29345, l 17 : N. America -> North America
not changed

p 29345, l 24 : "meteorology fields -> "meteorological fields"
not changed

p 29345, l 25-27 : shouldn't it be LMDz instead of LMDZ?
changed

p 29346, l 3-4 : How can you calculate water vapour only based on surface water fluxes? Is there a cloud parameterization in MOZART? Is there CAM behind?
MOZART does have a cloud parameterization. It is based on the MATCH CTM developed by P.Rasch and M.Lawrence at NCAR many years ago (not SMHI-MATCH used in this paper).

p 29346, l 15 : "that are" -> "which are" ; "source of" -> "source for"
not changed

p 29346, l 21 : "agree on in the location" -> "agree on the location" or "agree in the location"
changed

29346, l 26 - p 29347, l 2 : by mentioning explicitly dry deposition, one can give the impression that dry deposition is determining for the tropospheric ozone differences. Especially the April differences in the NH will only by slightly impacted

by the deposition scheme, I presume. Isn't the influx from the stratosphere determining: is there a difference in performance among models which prescribe O3 at the top of the model, and those using explicit stratospheric chemistry?
The discussion of differences in ozone distributions has been expanded to mention strat-trop exchange and chemistry, as well as identifying deposition as important to the lower troposphere. There does not seem to be a consistent difference between models with simulated or specified stratosphere.

p 29346, l 28 : "surface layer" : or is it meant "boundary layer"? The surface layer is often just a fraction (1 tenth) of the boundary layer.
'surface layer' changed to 'boundary layer'

p 29347, l 4 : magnitude -> should be little bit more specific like "concentration/mixing ratio/value"
added 'concentration'.

p 29347, l 5-6 : I would think that the other models possibly also show a maximum in the tropical mid- to upper-troposphere. But when that maximum is lower than 2x10-6 it will not be visible in this type of plots. Maybe change : "maximum" -> "a maximum higher than 2x10e-6".
Added 'greater than 2x10-6'.

p 29347, l 7 : "have" -> "reach OH concentrations of"
changed

p 29347, l 11 : "a number of compounds" -> I would suggest to be more specific
changed to ' the time series of ozone and its precursors'

p 29347, l 14 : CO was already used on page 29334, so the definition should come earlier
definition added in Introduction

p 29347, l 17 : C2H6 already mentioned earlier, so should be defined earlier
it was mentioned only in the model descriptions; not changed

p 29347, l 17-19 : can differences in transport contribute to these differences?
Possibly, but the long lifetime of ethane makes that less important, I think.

p29347,l19: Table1->Table2?
corrected

p 29347, l 21-22 : inverse order of H2O2 and hydrogen peroxide
I don't understand why this is so important.

p 29348, l 2-3 : strange sentence, improve it.
Strange? I see nothing wrong with: The differences among models are further explained below with regard to comparisons to observations.

p 29348, l 15 : "The hourly model output" -> "Hourly model output"
not changed

p 29248, l 21 : can one learn something from showing additionally RMSE-profiles?
I think the direct bias (preserving the sign) is more informative in evaluating the model performance.

p 29248, l 24-25 : what type of upper boundary conditions have been used in SMATCH? Maybe also refer to Figs. 5 and 7 where the different behavior of SMATCH was noticeable.
The source of the upper boundary conditions was added to the description of SMHI-MATCH.

p 29348, l 25 - p 29349, l 1 : GEOS-chem : can this be linked to the different emissions? Is there a reason for this?
It is likely that the low ozone in GEOS-Chem is related to the high uptake of HO2 on aerosols.

p 29349, l 10 : provide -> provides
corrected

p 29349, l 12 : "over a range of latitudes" : "over a range of mid- to high latitude stations"
changed to 'over a range of northern mid- to high latitudes'

p 29349, l 15 : "altitude" : has for some of the stations a model level different from the lowest level been chosen?
Yes, to account for the surface elevation of the coarse model grids being lower than the sampling site.

p 29350, l 1-3 : While the former sentence indicates an advantage of using averaging kernels, this sentence mentions a possible disadvantage. Therefore I would replace "also" by something like "on the other hand", or start with "however".
This paragraph has been re-worded based on comments by Referee #1 and this sentence has been modified.

p 29350, l 4-5 : I think this introductory sentence should be improved.
I don't see it as deficient.

p 29350, l 10 : "fires" -> "wild fires"
not changed, as explained previously

p 29350, l 14-16 : can this be more specific?
The specifics are listed in the previous sentences.

p 29350, l 17 : "median", while the figure 13 caption says "mean"
The bias plots have been removed from Fig.13, so this is irrelevant now.

p 29350, l 19 : I think this "while"-construction is not so lucky; "regions" : I'd rather use the word "pixel" again, or "grid point"
'not so lucky'? 'while only regions …' changed to 'and only pixels'

p 29350, l 21-22 : "Figure 13a and c" -> "Figures 13a and c"
corrected

p 29350, l 23 : median
correct, not mean

p 29350, l 24-26 : "Northwest" , while "north east" on line 9
corrected (one word, lowercase)

p 29350, l 29 : "East-China", while "East China" on line 13 (but here it is more an adjective)
corrected

p 29350, l 29 : "indicating a large uncertainty introduced by the models" : this looks like models introduce uncertainty. I'd rather say that our knowledge is uncertain.
That is what we meant – the models all have the same NO emissions but end up with different NO2 column amounts. Added sentence to clarify: 'Since all models used the same NO emissions, the large variation between models (as seen in Figure 14) indicates differences in the chemistry and transport processes affecting NO and NO2.'

p 29351, l 5 : "for forests" -> "for forest fires"
not changed

p 29351, l 6 : "the hourly output" -> "hourly output"
not changed

p 29351, l 11 : I would put (A1, A2) immediately behind ARCTAS-A
seems clear enough as is

p 29351, l 19 : GRACE -> to "-GRACE"
not changed

p29351,l20: fire->wildfire
not changed

p 29351, l 22 : "with" -> "and"
not changed

p 29351, l 24-25 : the same -> in the same way
changed

p 29351, l 25 : "were" -> "was"
corrected

p 29351, l 26 : "measurement uncertainty" : is this the measurement uncertainty on one single observation? Or is it a reduced uncertainty as the observations shown are already the mean/median over a large number of individual observations?
It is the measurement uncertainty of a single measurement.

p 29351, l 22 - p29352, l 2 : In this short paragraph of text, three large figures are introduced. In the following sentence "To make a more quantitative ...", one temporarily gets the impression that these profiles as such will not be discussed anymore further. However, in the next 2 pages, one refers several times to the profiles. To avoid this initial misconception, I would start the sentence "To make a more ..." by something like "in addition".
Yes, that is better.

p 29352, l 12 : "indicating" -> "indicating that"
changed

p29352,l14: 100%->90%
corrected

p 29352, l 17 : (Fig. 14) -> (Fig 18.)
No, we meant the OMI NO2 comparison in Fig. 14.

p 29352, l 18 : hugely : I would use another word
vastly?

p29353,l2: "oftoofineascale"->"ofatoofinescale"?
of a scale too fine?

p 29353, l 7-8 : I would replace the first "with" by "and"
changed

p 29353, l 8 : ethanol is not shown in figure 19?
No, but it is in Figs 16 and 18.

p 29353, l 9 : a poor job of : can you express this differently?
Maybe…

p 29353, l 10 : much closer is true for CARB, but not for B
but B is on average better than A1 and A2

p 29353, l 16 : refer to Figs. 16 and 18, after "For ARCTAS"
added

p 29353, l 16-18 : however if the conditions are homogeneous, it should not have a large impact. And if the flight legs are long, part of the observations will be in/below/above clouds, and averages can than still make sense.
Yes, so that is why it is shown. Sentence has been re-written.

p 29353, l 18 : These average biases -> "The average biases"
changed

p 29353, l 20 : I would add "(see Fig. 18)" after "In summer"
added at end of sentence

p 29254, l 4 : fires -> wild fires
changed

p 29354, l 7-8 : is the "," before "back trajectories" correct?
Yes, this is a list of the several techniques used by Hornbrook et al.

p 29354, l 15 : "make" -> "makes"
corrected

p 29354, l 21 : express 252-258 E as 92-98 W, etc ...
Is that necessary?

p 29354, l 21 : between the surface and 850 hPa (to avoid confusion)
changed

p 29355, l 1 : 252-258E : express as ... W
not changed

p 29355, l 19-20 : maybe you can add "POLMIP"
ok

p 29355, l 23 : "driven to at least some degree by observed ..." -> "driven by, to some degree at least, observed ..."
added commas without rearranging text

p 29355, l 23-24 : meteorology -> meteorological data/fields
not changed

p 29355, l 25 : "occurred among the model outputs" -> "occurred in the model results"
not changed

p 29356, l 3 : I would not write "completely", as that is probably a too high expectation
changed to 'better'

p 29356, l 1-6 : this is a very weak conclusion : just illustrating differences, and requiring for "additional model diagnostics in the future". This contradicts with the suggestion in the abstract (line 5-6) "to quantify the differences in model chemistry and transport schemes". Do we learn something about transport schemes? Suggestions for new/other research should be stated at the end of the conclusions.
The abstract was perhaps overstated. '… quantify the differences' has

been changed to 'assess the differences'. The second sentence of the abstract is referring to POLMIP in general, and Monks et al concludes the transport schemes are similar among the models. The comments about additional diagnostics have been moved to the end of the Conclusions.

p 29356, l 7-9 : Why is this a reason?
'model errors' changed to 'emissions errors'. The full sentence is: 'While the extensive suite of aircraft observations in 2008 at high northern latitudes is extremely valuable for evaluating the models, they cannot uniquely identify the source of emissions errors, as the Arctic is influenced by many sources at lower latitudes.'

p 29356, l 14 : indicate -> indicate that
OK.

p 29356. l 19 : "dynamics" : this contradicts a bit the sentence on p 29355, l 23-25, where it was suggested that all model represent the "dynamics of the study year".
'dynamics' changed to 'boundary layer parameterizations'.

p 29356, l 29 : OVOCs should be defined. Or is VOC meant?
Changed to VOCs.

New Conclusions:

Eleven global or regional chemistry models participated in the POLARCAT Model Intercomparison Project (POLMIP), allowing for an assessment of our current understanding of the chemical and transport processes affecting the distributions of ozone and its precursors in the Arctic. To limit the differences among models, a standard emissions inventory was used. All of the models were driven, to at least some degree, by observed meteorology (GEOS-5, NCEP or ECMWF), and therefore represented the dynamics of the study year (2008).

While the extensive suite of aircraft observations in 2008 at high northern latitudes is extremely valuable for evaluating the models, they cannot uniquely identify the source of emissions errors, as the Arctic is influenced by many sources at lower latitudes. However, several conclusions can be drawn about the emissions inventory used in this study. Based on the comparisons to aircraft observations and the NOAA surface network data, emissions of CO, ethane and propane are clearly too low. The comparisons to satellite retrievals of OMI NO_2 show a few regions of consistent model errors that indicate that anthropogenic NO_X emissions are underestimated in East Asia, while fire emissions are overestimated in Siberia. Large differences are seen among the model NO_2 tropospheric columns over Europe and China, thus limiting the conclusions that can be drawn regarding the accuracy of the emissions inventory. The large range in modeled NO_2 (where NO_X emissions were the same) also indicates that model chemistry and boundary layer parameterizations can significantly impact NO_X chemistry. More accurate emissions inventories might greatly improve many of the model deficiencies identified in this study. Emissions inventories modified based on inverse modeling results, as well as results of this study, will be used in future work as one step in improving model simulations of Arctic atmospheric

composition.

The simultaneous evaluation of the models with observations of reactive nitrogen species and VOCs has illustrated that large differences exist in the model chemical mechanisms, especially in their representation of VOCs and their oxidation. Most of the models showed a negative bias in comparison to ozone observations from son- des and aircraft, with a slightly larger difference in April than in summer. The models frequently underestimated ozone in the free troposphere by 10–20 ppb **in the comparison with ozonesondes. In addition, 10–30% negative model biases were seen in comparison to the mid-troposphere aircraft ozone measurements. Comparisons for ozone precursors such as** NO_X**, PAN, and VOCs show much greater biases and differences among models. It appears numerous factors are the causes of these model differences. The differences among model photolysis rates and cloud distributions indicate some of the possible causes for differences in modeled OH, which leads to differences in numerous species and ozone production and loss rates.**

Some differences among the simulated results are likely due to different physical parameterizations such as convection, boundary layer mixing and ventilation, wet and dry deposition. Additional model diagnostics are required to better understand the differences among models. For example, comparison of the wet deposition rates and fluxes of a number of compounds could be informative in understanding the budgets of NO_y**,** HO_X **and VOCs.**

Evaluation of chemical transport models with numerous simultaneous observations, such as those of the POLARCAT aircraft experiments, can assist in a critical assessment of ozone simulations and identify model components in need of improvement. Model representation of the oxidation of VOCs and the NO_y **budget can have a significant impact on ozone distributions. Future chemical model comparisons should consider evaluation of VOCs and reactive nitrogen species as an important component of the evaluation of ozone simulations.**

DETAILED COMMENTS on the TABLES :

Table 1 : The abbreviation "bb" should be defined in the caption or in the text. BIGALK, BIGENE, ... should be defined (or referenced).
Changed.

Table 2 : Try to use the same naming for acetone and methanol as in Table 1.
Changed.

Table 3 : For WRF-Chem, the number of levels is not indicated. In the chemistry description for WRF-Chem is written "MOZART" : is this "MOZART-4"?
Number of levels added. 'MOZART' is a standard chemical option in WRF-Chem. It is based on MOZART-4.

DETAILED COMMENTS on the FIGURES :

Figure 2 : The colours for CAM4-chem and CAM5-chem are very similar (as in other figures). Can this be changed? Units for pressure should be mentioned [hPa]. Is the unit mmol/mol for water vapour correct (as often water vapour mixing ratio is

expressed in kg/kg or g/kg)? If possible, it would be nice if some extra values where given on the y-axis (e.g. 200, 500, 700 hPa).
The colors for the 2 CAM-chem versions have been changed slightly.
The water vapor units are correct. Units added to y-axis labels.

Figures 3-4 : These plots are too small.
Will try to enlarge them.

Figures 3-6 : Units for Pressure should be mentioned (on the y-axis or in the caption), and if possible extra values should be indicated on the y-axis.
Units added.

Figure 7 : It would increase the readability of the figure when the names of the species are mentioned on the top of every individual plot (instead of on the y-axis). Only having on the y-axis "mixing ratio [ppbv]" would be ok. The common title "50-70N ZA 700 hPa" is not nicely integrated in the figure. The text in the caption is possibly enough to make this clear.
Will try to make these prettier.

Figure 9 and 10 : It would be nice to have the station latitude (and possibly the longitude) in the top of every individual plot. If possible, it would be nice to have some more pressure values on the y-axis. It would be nice to have the number of sondes indicated in the plots. It is mentioned in the text that there were daily launches during April, but it seems that for some stations there were much less than 30 profiles available.
Some more values added to y-axis, and number of sondes added. The map shows the location of the sites, so have not added latitudes to this plot.
The range of dates for the sondes (April 1-19 and June 25-July 12) was amended in the text to be more specific.

Figure 11 : Try to use for the longitude the "... W" notation if the longitude is between 180 E and 359 E.
Not changed.

Figure 12 : I don't know if the general title for this figure is needed - I would rather use the figure caption for this. Possibly add "upper left panel" after "OMI tropospheric column NO2". The figure is in general too small to read the values and units on the colourbar.
The plots of the individual model biases have been replaced with the multi-model median bias, so there are now just 2 panels (see below).

Figure 13 : The text in this figure is too small. It would be nice to indicate the boxes also in the bias figures (although it is true that the reason for their definitions comes from the OMI plot). The actual link between the individual boxes in Fig. 13 and individual descriptive names used in Fig. 14 and in the text (NW-Europe, NE-US., E-China, Japan, S. Korea, W. Asia, E. Asia, Canada, East. Sib.) are never explicitly made.

Maybe one should, or (i) indicate the names in the figures next to the boxes, or (ii) make a table giving the coordinates of the boundaries/corners of the areas.
Removed the panels of the bias, so only 4 panels are shown – anthro- and biomass burning-filtered pixels for each season (see below).

Figure 14 : A too small font is used in this figure. The word "model" is difficult to read in the upper right blue box. Somewhere should be mentioned that one looks at column values. I would replace "whiskers to" -> "whiskers show" or "and whiskers".
Figure has been remade (see below).

Figure 15 : Within the figure I would also write "ARCTAS-CARB instead" of "ARCTAS- C" to limit ambiguity.
Changed.

Figure 16 : Species names should be shown at the top of the figures. There are also x-axis labels winch overlap. One should try to improve the writing of "j o3 o1d" into "j(o3->o1d)" and "j no2" into "j(no2)".
Figure remade (see below).

Figure 17 : The mentioning of "ARCPAC P3 April 11-21" is not very elegant with respect to the rest of the figure (same comment as for Fig. 7). It would also be more practical for the reader if the names of the species should be mentioned at the top of the individual plots instead of below the x-axis.
Figure remade (see below).

Figure 18 : There are overlapping or too close values on the x-axis for some of the plots. Name of the species should be mentioned on the top of each individual plot, instead of on the x-axis. The x-axis should contain "mixing ration [ppbv]" or just "[ppbv]".
Figure remade (see below).

Figure 19 : The way "Campaign" (horizontal) and "Model Bias (%)" (vertical) are added in the figure should be improved. Eventually change the caption to contain the information, e.g. : "Mean bias" -> "Mean bias (%)" and "A1 : ARCTAS ..." -> "The campaigns are ..." . I would also write "ARCTAS-A1" and "ARCTAS-A2" instead of "ARCTAS-A" twice.
Caption changed as suggested.

Figure 20 : In the caption, I think Enhancement Ratio can be written with small letters. Does it make sense to also plot the uncertainty on the estimate from the models? Or does that give no interesting information?
The 1-sigma uncertainty in the derivation of the slopes from the model results are smaller than the symbols in Fig. 20. It is difficult to make further estimates of the model uncertainties.

New Figures

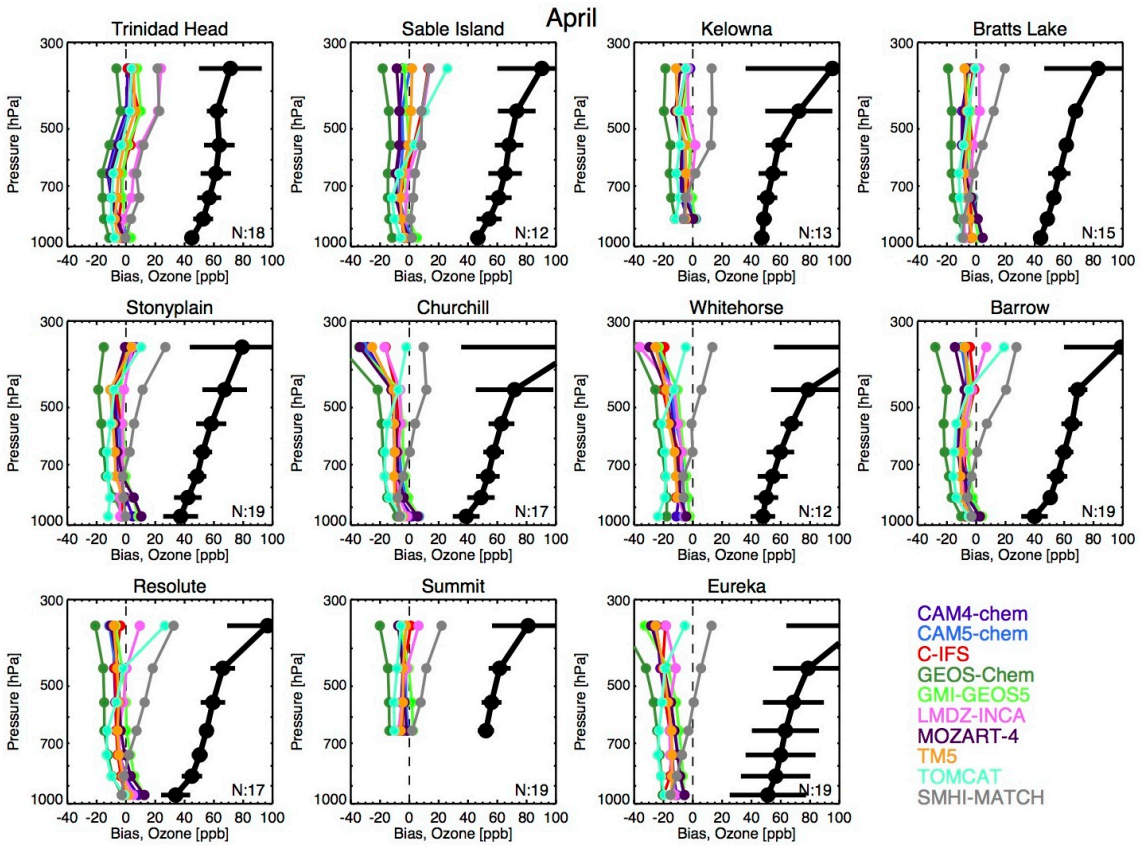

Figure 9. Comparison of models to ozonesondes for April, showing mean and standard deviation of the observations (black line) and the mean bias (colored lines) for each model at each site (Tarasick et al., 2010; Thompson et al., 2011). Results shown for only surface to 300 hPa for clarity. The number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

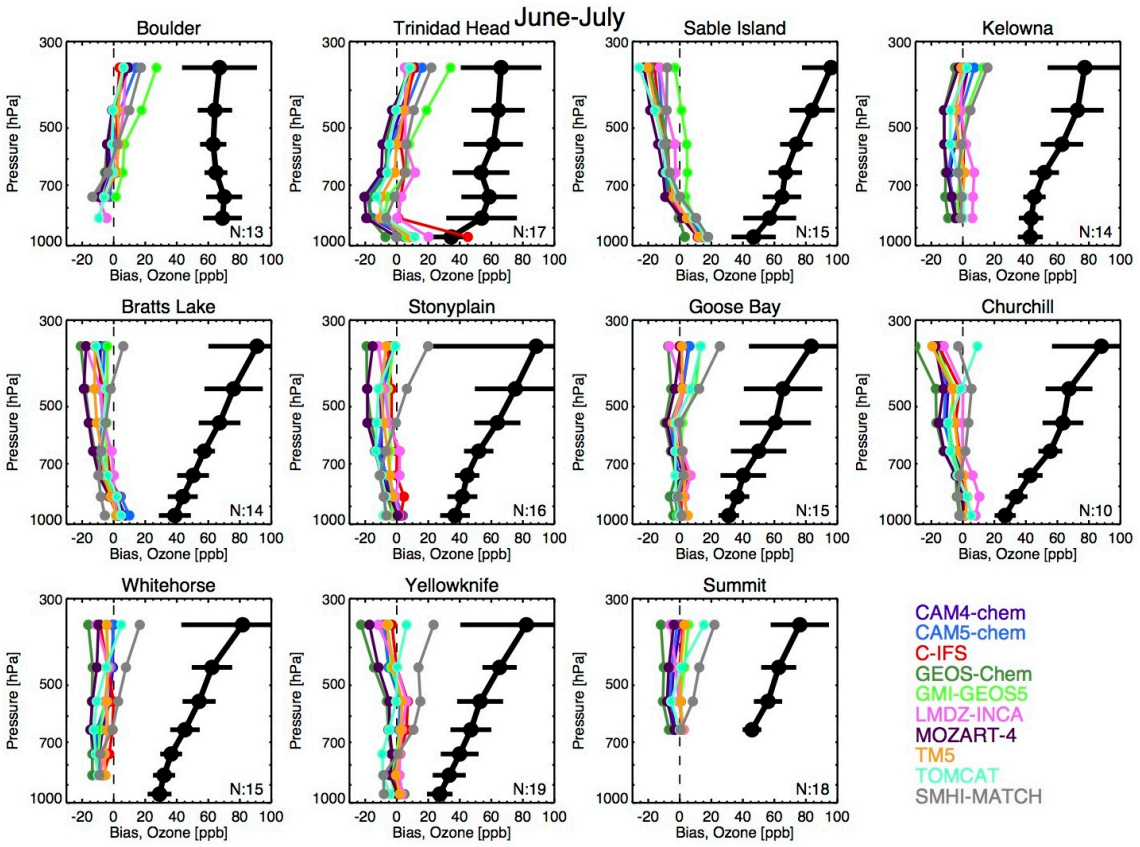

Figure 10. As Fig. 9, but for June–July.

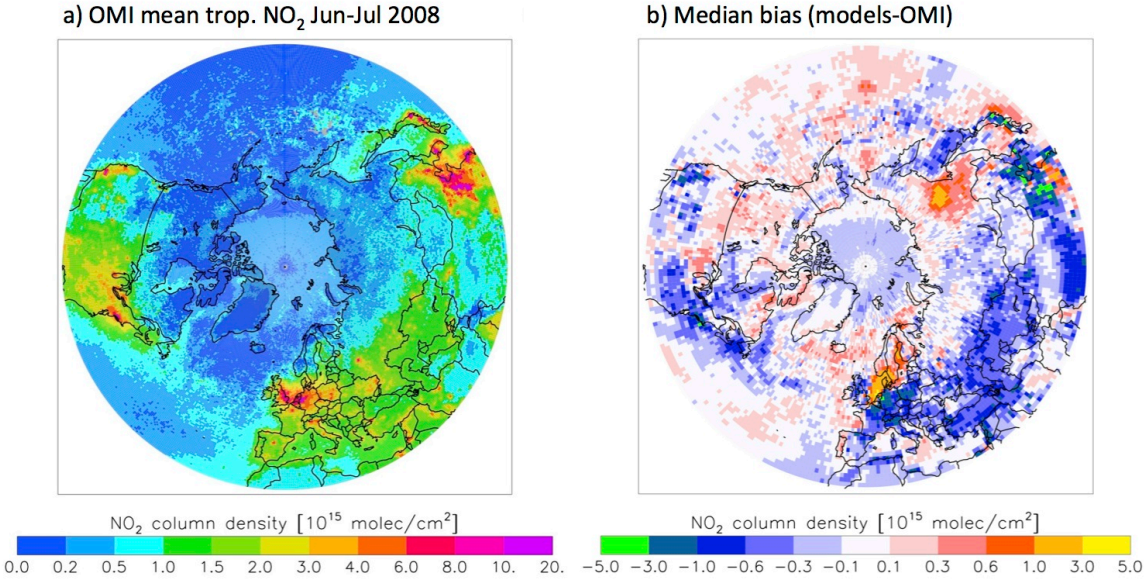

Figure 12. **a)** OMI tropospheric column NO_2 and **b)** median of the model biases, both for 18 June–15 July.

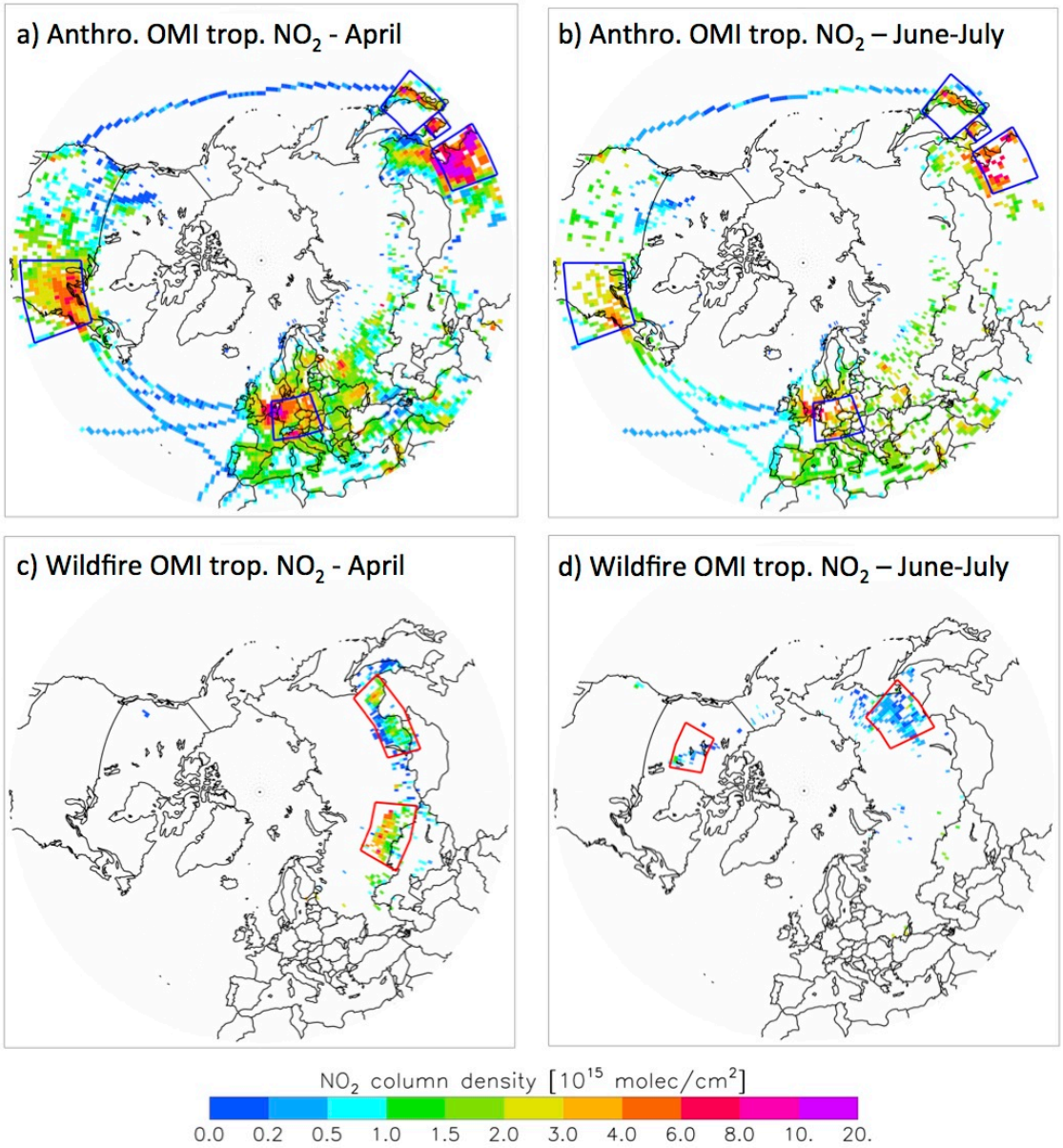

Figure 13. OMI NO_2 filtered for dominant anthropogenic emissions **(a, b)** and fire emissions **(c, d)**. Boxes indicate the regions for which biases have been calculated in Fig. 14. Panels **a, c** show April, panels **b, d** show June-July.

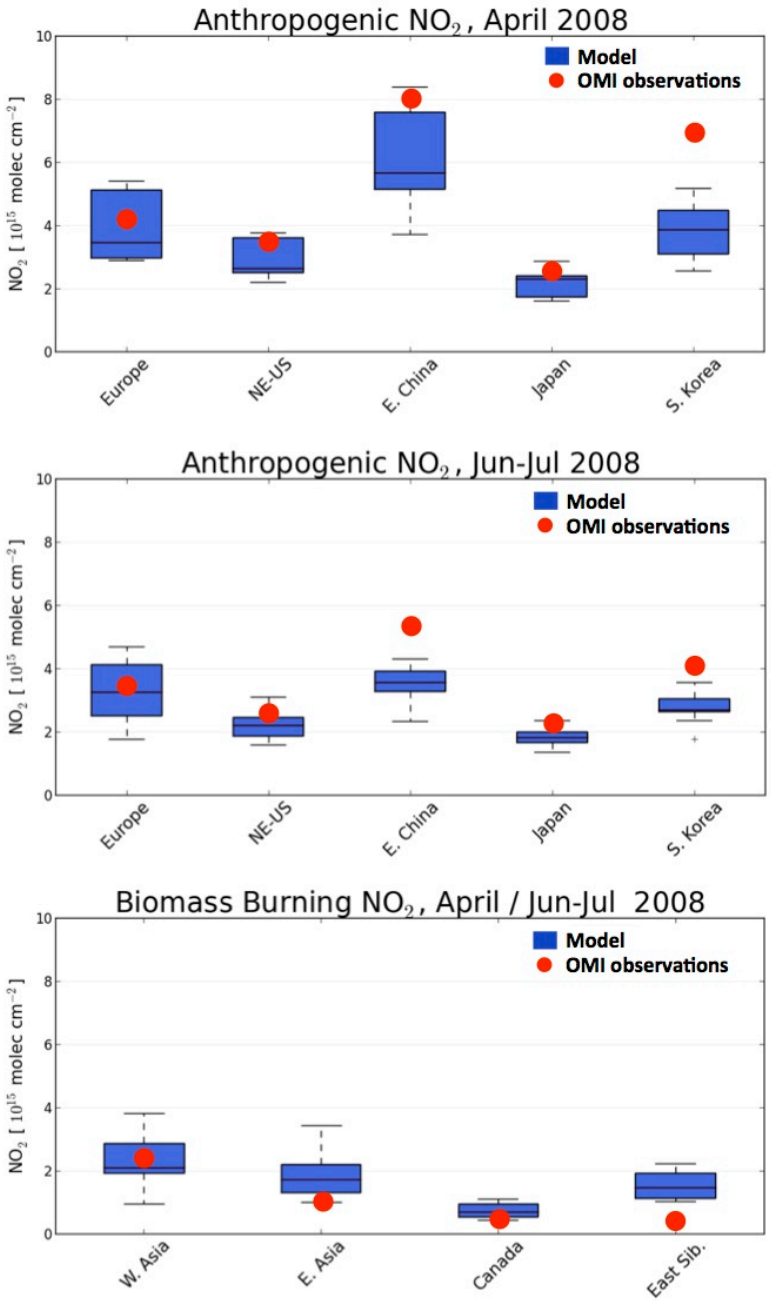

Figure 14. Summary of the regional means from each model and the OMI NO_2 tropospheric columns for each region indicated in Figure 13. (a) Anthropogenic emissions in April and (b) June–July, (c) biomass burning in both seasons. Red circles are mean OMI NO_2 observations for the region; box plots show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

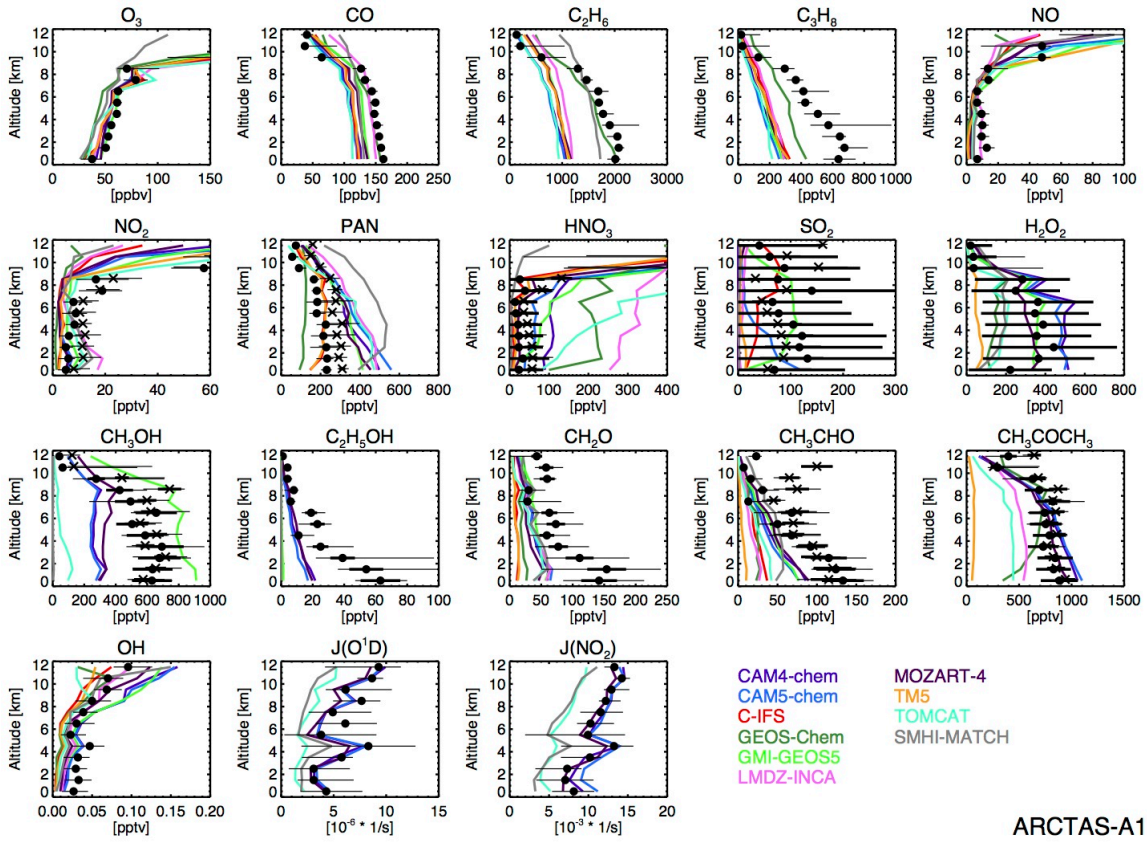

Figure 16.

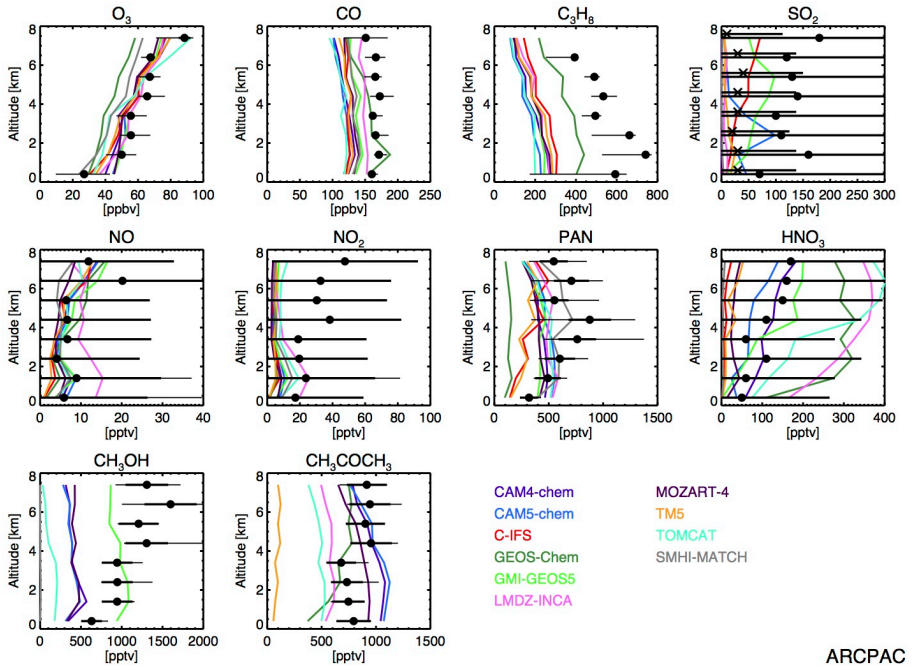

Figure 17.

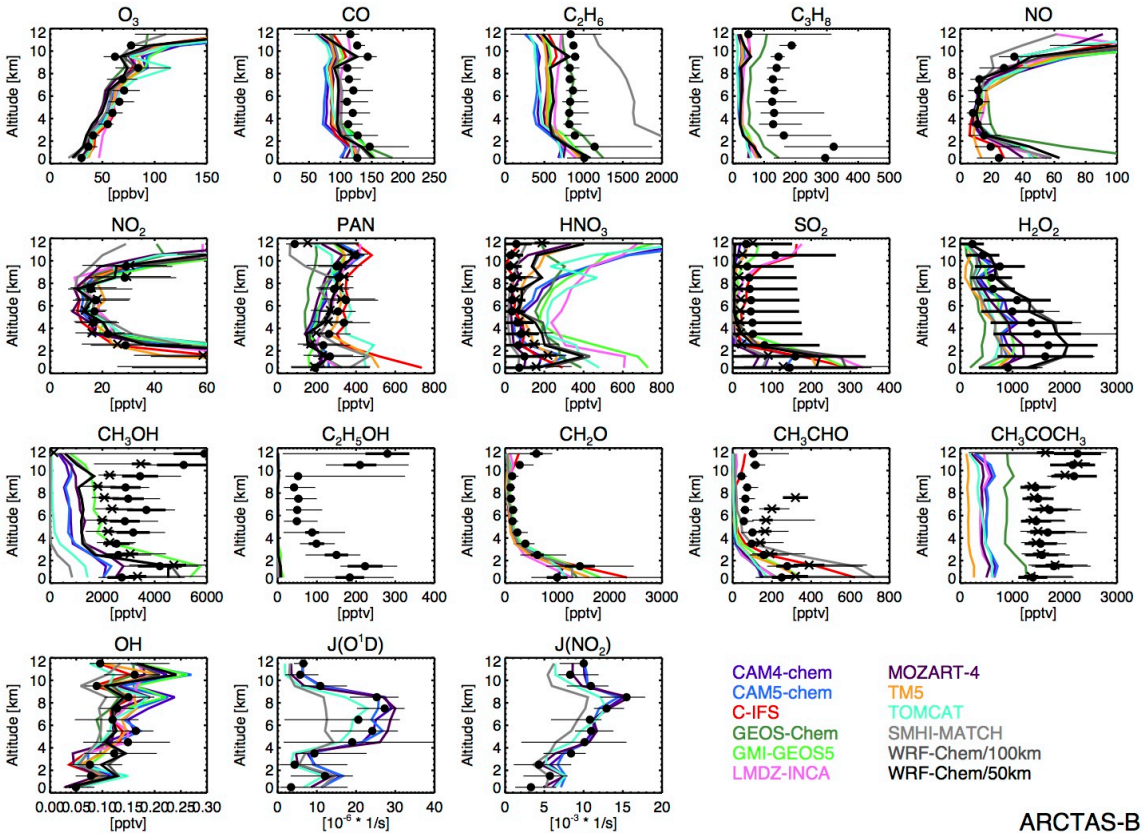
O3
CO
C2H6
C3H8
NO
NO2
PAN
HNO3
SO2
H2O2
CH3OH
C2H5OH
CH2O
CH3CHO
CH3COCH3
OH
J(O1D)
J(NO2)
Altitude [km]
[ppbv]
[pptv]
[10-6 * 1/s]
[10-3 * 1/s]
CAM4-chem
CAM5-chem
C-IFS
GEOS-Chem
GMI-GEOS5
LMDZ-INCA
MOZART-4
TM5
TOMCAT
SMHI-MATCH
WRF-Chem/100km
WRF-Chem/50km
ARCTAS-B

Figure 18.