Response to Anonymous reviewer #2

COMMENT:
This paper describes applying the regional CTM WRF-Chem run with 35 x 35 km grid to assess transport of wild fire emissions from Canada and anthropogenic emissions from north east north America to central/south Greenland. Attention is also focused on production of ozone during this transport and assessing impact on the ozone burden in high northern latitudes. The second phase of the IPY POLARCAT experiment in June/July is targeted for study because of the availability of airborne measurements from 4 different platforms. A unique aspect of the airborne data set is that several of the flights near Greenland sampled distinct plumes that had been characterized ~ 5 days earlier much closer to sources. Fresh wild fire emissions were sampled by the NASA DC-8 on 4 flights over north central Canada, and two profiles from MOZAIC over Philadelphia provide ozone and CO profiles in airmasses heavily impacted by anthropogenic sources.

As noted by anonymous referee 1, applying a regional CTM to these questions is probably the most noteworthy aspect of this study. The authors point out that several previous studies using global CTMs had not found BB plumes to be dominant sources of ozone in the Arctic troposphere, including several recent studies that also used POLARCAT observations to assess CTM performance in the Arctic and sub-Arctic during summer 2008. They suggest that global CTM simulations may underestimate ozone production in both urban and BB plumes partly due to the large grid cells used in such models, and test how well the regional CTM simulates the transport and evolution of several individual plumes that were sampled both fresh and aged. Having established some confidence in the skill of WRF-Chem to properly simulate plumes, the impact of all boreal fires in June-July is estimated to have increased ozone in the POLARCAT study region by about 5% from 6-9 km, compared to an 18% increase over the 2-6 km range attributed (by the model) to ozone produced in pollution plumes from north American anthropogenic sources.

The study is well designed and mainly well presented. I think it should be published in ACP after attention to a couple of relatively major suggestions (and a longer list of editorial comments.
RESPONSE:
The authors thank the reviewer for the helpful review and the careful reading of the manuscript. Each individual comment is addressed in detail below.

COMMENT:
Referee 1 suggests that the authors consider refining the statistical approach used to assess the impact of plumes (section 5.1) and I think I largely agree with these comments.
RESPONSE:
We agree with this comment and have addressed this in our response to reviewer #1.

COMMENT:
My own primary concern is that the authors should take a bit more care describing and assessing the comparison between observations and model estimates, especially in section 4 where they focus on establishing the skill of WRF-Chem to correctly transport and transform individual plumes.

My principal concern is that throughout section 4 the text is not consistent in making it clear when a statement is being made in reference to the observations, or in reference to the model estimates. For example, the first sentence of 4.1 states that "a first plume was measured. . .11:45-12:15" and "second . . .12:20-12:50" and refers to Fig 8 a,b. I assume that measured is referring to observations and suggest that there appear to be 2 separate CO plumes during the 6 km flight leg (~ 11:35-11:40 and 11:45-12:00) and a third plume with higher concentrations along most of the flight leg just below 4 km (12:10-12:30). In ozone observations I also see 2 enhancements along the 6 km leg (~11:30-11:35 and 11:45-12:00), then a third during the descent between level flight legs (~12:05-12:10). Note that only the second ozone peak coincides with CO peak, in fact CO decreases markedly in both the first and third intervals with enhanced ozone. (Authors point this out for the ozone peak/CO dip during the descent near the end of this paragraph, and discuss possible cause in the last paragraph of 4.1).
RESPONSE:
This section has been rewritten for clarity and has been revised to specific when the discussion refers to the model or the measurements.

COMMENT:
My point is that this opening sentence claims to be discussing the "measurements" but really seems to be describing features in the model output. Then, in the next sentence I am confused/concerned by the statement about "good agreement. . ..but the peak in CO occurs later in the model than in the measurements by 15 min". Confusion arises because I see 5 CO peaks in Fig 8a (3 in the measurements and 2 in the model, with temporal alignment only between the first modeled and second measured). So what is meant by "the peak"? I think they are pointing to third peak in the atmosphere, second one in the model, but it should not be up to me to work this out. We can leave it to personal judgment whether horizontal displacement on the order of 100 km between a plume sampled by the plane and one in the model world constitutes good agreement, and I do note that the authors discuss the challenge of accurately transporting small features several other places in the manuscript.
RESPONSE:
This section has been revised to clarify this discussion. It's important to note that considering the nature of atmospheric dynamics and chemistry – a shift of 15 minutes in the timing of the peak is very good agreement. We have revised the discussion of the peaks to more clearly distinguish between discussion of the model and measurements in this section.

COMMENT:
I have similar issues with the first paragraph of section 4.2. Authors state there are two obvious pollution plumes shown by CO, one at the end of the leg just above 7 km and then at the start of the 4 km leg. I easily see the first one, but the lower altitude enhancement is not so compelling. However, I do find discussion of the plumes sloppy, especially the statement "plume 1: peak CO of 120 ppbv". In the measured plume I see peak values > 140 ppb, and note that FireCOSens stays above 130 ppbv for 10-15 minutes. Regarding the second "plume", it is impressive how well the model does, and a good case is made for anthropogenic dominance, but 110 ppb is not an impressive plume and I have to note that the measured values rise back to nearly the same level at the end of the 4 km leg while the model estimates begin to fall off rapidly.
RESPONSE:
This is certainly true that the second peak in CO (with lower CO concentrations) is not as apparent in the measurements. We have re-written this section to highlight that model predicts the CO mixing ratios correctly, but only upon including anthropogenic emissions. Otherwise, the model predicts lower values than measured.

COMMENT:
The discussion of Fig 10 d-f (fourth paragraph of 4.2) really got my blood pressure up. This entire paragraph ignores the fact that the cross sections are model products. Comparison to Fig 8 makes it clear that not all of the features in Fig 10 discussed in the text are real!
RESPONSE:
The original text contained a typo and in addition was very confusing, we thank the reviewer for pointing this out. Figure 10f shows measured lidar profiles, not a result from the model. Therefore, this section has been updated to indicate the Figure 10d-e are profiles extracted from the model and to highlight that 10f is the corresponding ozone lidar measurement. The discussion of the lidar measurements has been updated to be more clear when the measurements are being discussed.

COMMENT:
Note, I am not saying that the model has to be perfect before it can be useful, I am just urging the authors to be more precise when making comparisons between the observations (also not perfect) and the model estimates.
RESPONSE:
The entire section on comparing model to measurements has been revised for clarity, to provide a more precise comparison between the model and measurements.

MINOR COMMENTS:

COMMENT:
29707, 13 contribute almost equally as–→contribute nearly as much as
RESPONSE:
This has been updated

COMMENT:
29708-709 Surprised there is no mention of TOPSE in this survey of previous work. Tropospheric ozone was the key motivator for this mission that extended over nearly 5 months.
RESPONSE:
We have added a statement and reference to TOPSE:
"Earlier studies focused on Arctic ozone in the late winter-spring and ozone production during the spring equinox (e.g. Atlas et al., 2003)..."

Atlas, E. L., Ridley, B. A., and Cantrell, C.: Tropospheric Ozone Production about the Spring Equinox (TOPSE) Experiment: Introduction, J. Geophys. Res., 108(D4), 8353, doi:10.1029/2002JD003172, 2003.

COMMENT:
29709, 26 during the June-July 2008 —→ during June-July or during the June-July 2008 study period
RESPONSE:
This has been updated

COMMENT:
29710, 10-12 Confusing as written. There are essentially no sources at Summit, so what is meant by "downwind of source regions at Summit"?
RESPONSE:
This has been updated to say - "Previously, global models were also shown to underestimate ozone and CO in the remote Arctic during summer, for example at Summit in central Greenland, a receptor region for mid-latitudes pollution "

COMMENT:
29712, 10-12 Please clarify exactly what was done in the noFire and noAnthro runs. This sentence suggests that sources inside the WRF domain were shut off, but what about inside MOZART? If BB and anthro sources stayed on, seems that updating BC for WRF from MOZART would still transfer the impacts into the smaller domain.
RESPONSE:
This has been updated to include the sentence: "The model runs used the same MOZART boundary conditions, therefore the sensitivity runs highlight changes due to emissions within the model domain (Canadian fires and northeastern US anthropogenic emissions)."

COMMENT:
29712, 26 Do not need "summary" and "summarized" in same sentence. 29713, 25 ozone lidar profiles–→ ozone profiles measured by lidar
RESPONSE:
This has been updated to "shown in Table 1"

COMMENT:
29714, 3-4 what is meant by "2 ppbv, 2%" Is this 2 ppb plus 2%, or 2 ppb or 2%, whichever is larger
RESPONSE:
The precision of the measurement is 2 ppb - therefore we have updated this by removing "2 %" in the text.

COMMENT:
29714, 8 choose between "upward looking" and "zenith-viewing" but not both
RESPONSE:
This has been updated as suggested

COMMENT:
29714, 14-16 not clear to me why being above clouds or thick aerosol would impact lidar retrievals above the plane. It is obvious that being in, or below, an aerosol plume or clouds could be a problem for the lidar depending on the optical depth.
RESPONSE:
This sentence refers to when the measurement is above the clouds, not the airplane. To clarify this we have changed the sentence to read:
"Only measurements between the aircraft and clouds or thick aerosol layers are included here, because ozone retrievals in/above such layers is not possible."

COMMENT:
29714, 25 by adding CO—→with added CO
RESPONSE:
This has been updated as suggested

COMMENT:
29717, 15 while that aircraft targeted–→ while many of the DC-8 sorties targeted
RESPONSE:
This has been updated

COMMENT:
29717, 28 Not sure I would start this sentence "In contrast," Previous sentence just pointed out that the model has too little PAN, probably because formation from NOx is too slow. So, one would expect there to be too much NOx in the model. Quantitatively, the excess NO is much greater than the missing PAN, so there is probably also a problem with emissions or vertical mixing, but qualitatively the 2 problems are not inconsistent with each other.
RESPONSE:
The reviewer is correct, these are not contradictory statements therefore this has been changed to "In addition".

COMMENT:
29718, 9-27 This paragraph and Fig 3 e-h shows that WRF has less (some cases much less) of all the NMHC than the sampled atmosphere near the BB sources. Is this difference significant for ozone production in the model? Seems you point out this potential problem here, but never come back to it.
RESPONSE:
This is a very important point, the model is too low in all VOCs, but still reproduces ozone mixing ratios after 5 days of transport and mixing. We can conclude either: (1) that ozone chemistry is buffered and the VOC concentrations sustain realistic ozone production chemistry, or (2) that this is a lower limit of ozone production in plumes due to the low VOCs. We have added a sentence later in the discussion to address this point:
"It is important to note that while NMHCs in fresh fire plumes (discussed earlier for the DC8 measurements) are under-predicted by the model the resulting ozone profiles in the Arctic region agree very well with the measurements. Even so, we note that ozone production in fire/anthropogenic plumes, as represented in WRF-Chem, may be a lower limit for ozone production in the atmosphere due to the under-prediction these species in fire plumes."

COMMENT:
29719, 25 determine if this—→determine how much this (how could having the source wrong not impact the evolution of the plume to some extent?)
RESPONSE:
This has been updated as suggested as suggested to read “determine how much this this discrepancy in the boundary layer impacts aged plumes in the model.”

COMMENT:
29721, 12 and Fig 6f Appears that the model has 2-4 times more NO than the atmosphere from 6-8 km. This would not be "good" agreement if you were assessing measurements by two different instruments.
RESPONSE:
The shape of the NO profile from the model and measurements are in good agreement, however the reviewer is right that the model has more NO than measured when comparing with the average profile between 6-9 kilometers. For the specific Falcon-20 flights used for this study (4 July, 7 July, 8 July) the majority of measurements were taken during long flight legs at ~5 km (4 July), 7.7 km (7 July), 8.7 km (7 July), and between 7-7.8 km (8 July). Therefore, the standard deviations for these altitudes are more reliable, due to the larger number of measurements. At these altitudes, the model is within one standard deviation of the measurements (see profiles between 4.5-5 km, 7.5-8 km, and 8.5-9 km. The points that disagree have far fewer measurements; therefore the standard deviation of the measurements is less reliable for evaluation of the model. To address this, the following text has been added to this section:
“In the region between 6-9 km, the model has higher NO mixing ratios than measured. However, for the altitudes with long flight legs (See SI, Fig. 3) comprising the majority of measurements (5 km, 7.5-8 km, and 8.5-9 km) the model is within one standard deviation of the measurements.”

COMMENT:
29721, 14 and Fig 6e Likewise a matter of perspective whether a factor of 2 (or roughly 100 ppt) offset between modeled and measured PAN is a "small positive bias" or a reason to be concerned.
RESPONSE:
This has been updated to read:
"The modeled PAN (Fig. 6e) agrees with measurements in the upper-troposphere, but has a positive bias in the mid- and lower-troposphere."

COMMENT:
29722, 20 On the 5 July—→On 5 July
RESPONSE:
This has been updated as suggested

COMMENT:
29723, 23-24 "air mass present. . .. . .that do not"–→that does not
RESPONSE:
This has been updated as suggested

COMMENT:
29724, 13 not sure what is meant by "vertical stretching of the tropopause"
RESPONSE:
This has been rewritten to clarify that this is a tropopause folding event.

COMMENT:
29725, 10-11 "The ozone peak. . .at 13:30. . .in the measurements (yellow box, Fig.8d)" I see just a tiny bump in observed ozone inside the yellow box, well before 13:30. There is a broad enhancement outside the yellow box (over 60 ppb from about 13:40-13:50). Turning to the model, there is a similar enhancement entirely inside the yellow box, but this is all well before 13:30.
RESPONSE:
We agree this was very confusing – for the measurements the ozone enhancements from the stratospheric intrusion are in the gray box, as noted by the reviewer. This has been updated along with the other text in Section 4.

COMMENT:
29730, lines 20 and 24 sloppy nomenclature is a little confusing here. The delta CO and delta ozone values are defined as the difference between base run and other runs with certain sources turned off. So, it is not possible to calculate average slope or ratio of deltas just for the base run.
RESPONSE:
This has been updated to note that the enhancements are calculated using difference between runs (with-without emissions), following taking differences to calculate $\Delta O3$ and ΔCO the distribution functions and the slopes are calculated.

COMMENT:
29734, 3-6 Have made pretty strong case that the FireCOSens (2 x CO emissions in BB) seems better than the base run, so why not use the delta values from those runs. Pretty sure Ref 1 made similar suggestion.
RESPONSE:
The results from this run have been added to Figure 12 of the paper, with the slopes of the best-fit lines. We have included both the base and FireCOSens run to highlight how the distributions and slopes change with addition CO emissions from fires.

COMMENT:
29734, 11-12 Think about redrafting this sentence. It is not at all clear what "their" refers to.
RESPONSE:
This has been updated to read:
"The model results are used to quantify the amount of ozone produced from anthropogenic and fire emissions north of 55 °N."