We appreciate the careful consideration of our manuscript by this reviewer. We have considered the points raised and revised our manuscript accordingly. Our detailed responses and all changes that have been made are presented below.

General comments

This paper provides a case study of size distributions and, to a lesser extent, coating thickness on refractory black carbon particles over oil sands activities in Canada. I think this work is publishable, after the following concerns are addressed.

Specific comments

(1) L10. This paper does not address mixing state, so much as it addresses the properties of rBC particles. I suggest that the authors revise to make more specific for their particular study.

Our response: We agree with the reviewer that this manuscript did not address mixing state of particles and instead, only the mixing state of rBC containing particles was investigated (based on coating thickness). It was confirmed that coating thickness retrieved from SP2 has been commonly used as a measure of mixing state for black carbon aerosols (e.g., Schwarz et al., 2008; Kondo et al., 2011; Laborde et al., 2013). It was also confirmed that the term "mixing state" was always used for rBC containing particles or black carbon aerosols throughout the manuscript. Therefore, we think the use of "mixing state" in this manuscript will not cause misunderstandings.

(2) L21. While this is technically correct, the authors are presenting this as if it is new information. This has been known since some of the first SP2 measurements.

Our response: We agree with the reviewer that this manuscript was not the first one challenging the NMD settings for fossil fuel BC in aerosol-climate models. But we think this manuscript indeed provides additional information for this argument, based on a distinct type of BC source (i.e., oil sands operations) that has rarely been investigated. Nonetheless, we removed the related descriptions, which appear to imply that this point is a new finding of the present study, from the Conclusions section.

(3) L23. The meaning of "consistent" is not clear. Does this mean the same shape? Same MMD? Same width?

Our response: "Consistent" means the same MMD, NMD and the corresponding distribution widths. This point was clarified in the revised manuscript.

(4) L26. The meaning of "doubled" is left ambiguous here. Did the coating thickness increase from 1 to 2 nm? Or from 50 to 100 nm? These are both doublings, but with very different implications. More detail is welcome.

Our response: The coating thickness increased from ~ 20 to 40 nm within three hours when the oil sands plume was transported over a distance of 90 km form the source area. Coating thickness values were added as suggested.

(5) L27. How can the authors be sure that the apparent increase is not due to entrainment of background air?

Our response: As shown in Figure 12a and discussed in Section 3.4, coating thicknesses were in general smaller for rBC particles in the background air compared to the in-plume rBC cores. Therefore, the increase of coating thickness for in-plume rBC cores observed during aging could not be attributed to the entrainment of background air.

(6) Abstract. I suggest the abstract is revised to make it abundantly clear that the BC is derived from activities associated with mining of the OS, and not from the OS directly. This differs from some of the other emissions that are observed in this region.

Our response: The change was made as suggested: "we focused on BC emissions from the oil sands (OS) surface mining activities in northern Alberta".

(7) L82. I'm not sure this is necessarily the case. I don't think this, for example. The authors here set up an argument simply to shoot it down. I suggest they focus on what they observed, and leave discussion for later in the manuscript.

Our response: This sentence was removed as suggested.

(8) Fig. 2. It is unclear to me why there are so few points on these graphs. The SP2 measures at much higher size resolution than is shown here.

Our response: In this study, individual rBC cores detected during a specific period were first grouped into different size bins and then fitted by a lognormal curve. The same size bin was used throughout the present study, e.g., for the facility-integrated results shown in Figure 3 (i.e., Figure 2 in the original manuscript) and for the time-resolved results shown in Figures 4, 5, 7, and 8. This point was clarified in the revised manuscript.

(9) L115. The meaning of "virtual screen" is unclear to me. This could be clarified. Consequently, I have a difficult time following the discussion starting line 270 or so. The authors could be clearer about what they are doing.

Our response: A virtual screen corresponds to a specific downwind distance from the oil sands source area and consists of level flight tracks perpendicular to the wind direction at multiple altitudes. This sentence was added to the caption of Figure 1 which shows an example for the transformation flights involving multiple virtual screens. Moreover, a composite Google Earth image (as shown below) was provided for an easier understanding of transformation flight and virtual screens.

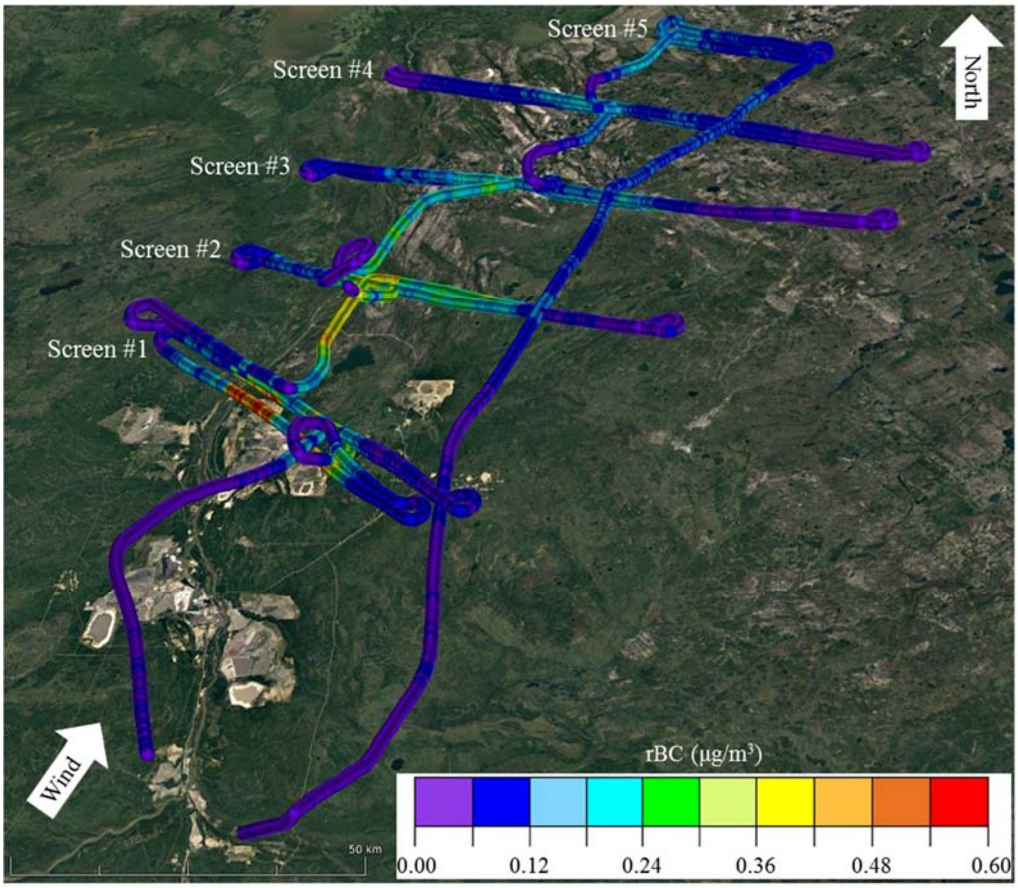

Figure R3_1. Composite Google Earth image showing flight track (colored by rBC mass concentration) for the transformation flight F_9/4.

(10) L123. A better citation regarding calibration would probably be one of the AMT papers by Laborde. Although I suppose they don't use regal black.

Our response: We agree with the reviewer that Laborde et al. (2012) was highly relevant to SP2 calibration. But we prefer to cite Cappa et al. (2012) here because regal black was used in it to calibrate the SP2.

(11) L129. The meaning of the "conversion" is unclear to me. Do the authors mean that they measure over the per-particle mass range X-Y fg/p and this corresponds to 70-260 nm? Could be clearer.

Our response: This sentence was revised to : "……the SP2's detection range for single particle rBC core mass (~ 0.3−16 fg) corresponded to an rBC size detection range of ~ 70−260 nm in terms of D_{MEV}.", which should be clearer.

(12) L134. Are the authors saying that a single, campaign average for F_{rBC} was found and applied? Or was it flight specific? Or something else?

Our response: Flight specific F_{rBC} were used to scale rBC concentrations in this study. This point was clarified in the revised manuscript.

(13) L135. This paragraph needs to include discussion of the limitations of the scattering method for coating thickness. The SP2 can only determine coating thicknesses for rBC particles with core diameters above some minimum size, typically around 150-160 nm. This is a limitation of the mismatch between the incandescence and scattering lower limit. This is discussed later, but it belongs in the methods. There is no reason, in my opinion, to even present Fig. 10 (since it is wrong, as the authors note later). There is too great of potential for this to be misinterpreted by people who do not read the paper carefully. This figure must be removed from the paper if it is to be published. Or, perhaps, it could be moved to the methods and discussed here as a method limitation. But it does not belong in the results and discussion.

Our response: Discussions on the limitations of the scattering method for coating thickness determination (including the corresponding figure), which were presented in the Results and Discussion section in the original manuscript, were moved to the Method section as suggested.

(14) L174. This very long sentence would be better as a table. Or could at least be supported by a table.

Our response: A table was added as suggested, which also included rBC NMD following the suggestion given in Comment# 15.

(15) L185. For any study that reported a MMD and a width, the NMD can be calculated. So, even if not directly reported the information is, most likely, easily obtainable. I suggest that the authors go through the effort of extracting this information and including it as part of a table.

Our response: rBC NMD were calculated as suggested for the studies which reported both the rBC MMD and mass distribution width. The calculated results were summarized in Table 1, together with reported NMD values. The NMD range representative for urban emissions was updated accordingly.

(16) L190. This has been long known. The authors should rephrase to state "Our results support the well-known suggestion that rBC from fossil fuel is smaller than biomass burning" or something like that.

Our response: The change was made as suggested.

(17) L202. While I generally agree with the authors, they must note here that their measurements are limited by an ability to measure below ~ 70 nm. They do not know if there is some other mode at smaller sizes. Most likely there is not. But their data cannot prove this one way or another.

Our response: We agree with the reviewer that the SP2 used in this study could not detect rBC cores below ~ 70 nm and thus, we have to assume that there was only one size distribution mode for the rBC cores, which might be invalid. This limitation was discussed in the revised manuscript: "However, there is also a need to evaluate the unimodal assumption for black carbon size distribution (Liggio et al., 2012; Buffaloe et al., 2014), given the SP2's limited detection range of rBC core size.".

(18) Fig. 5. The authors should change panel (a) to have two different axes ranges. Right now the range for the width and F_{rBC} are way too large. This should be changed to a range of 0-1. And the MMD axis to the range 100-200.

Our response: The figure was revised as suggested. The axis ranges we finally chose were slightly different from the suggested values, to more clearly distinguish different data series in the figure.

(19) The authors should strongly consider revising their definition of a log-normal distribution to use the more common formulation from e.g. Seinfeld and Pandis that includes a 1/sqrt(sigma) in the prefactor. This makes the widths vary from 1 to greater than 1, and is much more commonly used by e.g. climate models (which seems to be a target of this study).

Our response: We agree with the reviewer that different fitting parameters are being used to describe a lognormal distribution. As pointed out by the reviewer, there are two fitting parameters that can be converted between each other, i.e., the distribution width used in this manuscript as well as in some other SP2-based studies (e.g., McMeeking et al., 2010) and the standard deviation used by Seinfeld and Pandis in their book entitled "Atmospheric Chemistry and Physics". For a lognormal rBC mass-size distribution, the mass distribution width

($Width_{mass}$) can be converted to the standard deviation of the distribution (σ_{mass}) by $\sigma_{mass} = \exp\left(Width_{mass} / \sqrt{2}\right)$. Similarly, $Width_{number}$ can be converted to the standard deviation of a lognormal rBC number size distribution (σ_{number}) by $\sigma_{number} = \exp\left(Width_{number} / \sqrt{2}\right)$. For a lognormal distribution, therefore, a distribution width of ~ 0.7 corresponds to a standard deviation of ~ 1.6. The "$Width_{mass}$ vs. σ_{mass}" and "$Width_{number}$ vs. σ_{number}" relationships were added to the manuscript.

(20) L235. If the authors happen to have CO measurements from the flights then they can explicitly test this dilution hypothesis.

Our response: We agree with the reviewer that CO data, if available, can help to evaluate the dilution hypothesis. This dilution hypothesis was removed to avoid any misunderstanding.

(21) General. The authors are providing more sig figs than appropriate (e.g. L263). This should be revised.

Our response: The change was made as suggested.

(22) Figures 9. This figure works alright, but would probably work better as set of box and whisker plots, if the authors so choose.

Our response: We prefer to keep this figure as is. First, time-resolved lognormal fitting was not performed for all the flights. Second, the fitting parameters shown in this figure, e.g., for F_9/4, correspond to the various lognormal curves shown in Figures 7 (c) and 8 (c); and these fitting curves represent screen-averaged rather than time-resolved results.

(23) L299. Again, this is not generally believed. I don't believe this. I just take this to mean that background measurements are more impacted by biomass burning emissions than are urban measurements for many of the measurements that have been made. In the current study, the authors are simply sampling a particular part of the atmosphere where this is not the case. It seems that in their environment that the background rBC is dominated by, most likely, emissions from the OS activities. Thus, the size distribution of the background and plume look similar. The authors' discussion here focuses much too much on "processing," in my opinion, when the bigger issue is "emissions." I think this paragraph needs revision. Overall, I think that the authors are over complicating something that seems to me quite simple.

Our response: We agree with the reviewer that the statement "It is commonly believed that…… " is exaggerated. We also agree with the reviewer that the in- vs. out-of-plume comparison of rBC size distributions was less straightforward to support our argument that "not all aging processes will change rBC size distribution”. This paragraph was revised to: “……the rBC MMD was found to be 20 nm higher for aged urban plumes from Nagoya, Japan compared to fresh emissions from the same urban area (Moteki et al., 2007). Therefore, it has been argued that rBC size distribution tends to shift toward larger sizes during aging (e.g., McMeeking et al., 2010). Results from the present study, especially the comparison of rBC size distributions among successive flight screens (Figure 10), indicate that this is not necessarily the case……”.

(24) L338. I have substantial concerns that the 130 nm particles are still too small for robust sizing that will be (mostly) bias free. Scattering by an e.g. 120 nm rBC particle with a 5 nm coating (130 nm total) will be very different than that from an 80 nm rBC particle with a 25 nm coating, for example. This can lead to biases in interpretation. The larger one goes, the less this is an issue. The most robust studies limit analysis of coating thickness to >150 nm, unless it is explicitly demonstrated that a smaller threshold is.

Our response:

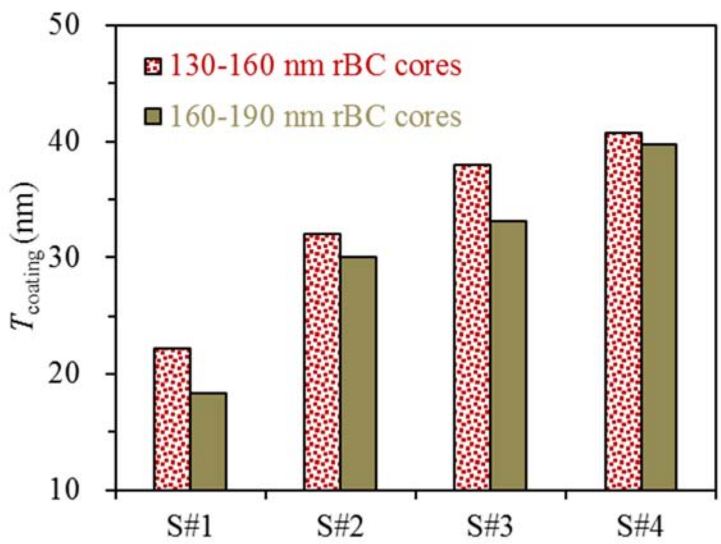

Figure R3_2. Comparison of coating thicknesses ($T_{coating}$) for in-plume rBC cores in two different D_{MEV} ranges (i.e., 130−160 nm and 160−190 nm) during the transformation flight F_9/4. Results from successive flight screens are shown separately. Coating thicknesses determined for 160−190 nm rBC cores are smaller than $T_{coating}$ of 130−160 nm rBC cores, which can be attributed to the limitation that the detection range of $T_{coating}$ is rBC D_{MEV} dependent. Evolutions of $T_{coating}$ exhibit the same pattern for rBC cores in the two different D_{MEV} ranges. However, both the count of rBC cores that can be assigned a coating thickness and the fraction of rBC cores that can be assigned a coating thickness are higher for the D_{MEV} range of 130−160 nm. Therefore, 130−160 nm rBC cores are used in the main manuscript for discussions on $T_{coating}$.

The size range mentioned here (i.e., 130−160 nm) is for the rBC core, rather than for the whole rBC containing particle. After accounting for coating, the size range of the whole rBC containing particle was at least ~ 170 nm, recalling that rBC coating thicknesses varied between ~ 20−40 nm for the different flight screens during F_9/4. We also investigated the coating thicknesses determined for larger rBC cores (i.e., the rBC cores in the D_{MEV} range of 160−190 nm; D_{MEV} indicates the mass equivalent diameter of the rBC core). As shown in the figure above, evolutions of coating thickness exhibited the same pattern for rBC cores in the D_{MEV} ranges of 130−160 and 160−190 nm. We prefer to present coating thickness results for rBC cores in the D_{MEV} range of 130−160 nm, mainly because the fraction of rBC cores that can be assigned a coating thickness ($F_{assigned}$) was the highest for this D_{MEV} range (Figure 11 in the revised manuscript).

In addition, we also examined the counts of rBC cores that can be assigned a coating thickness for the D_{MEV} ranges of 130−160 and 160−190 nm, i.e., N_{145} and N_{175} (145 and 175 correspond to the centers of these two D_{MEV} bins). N_{175} were found to be much smaller than N_{145}. For example, the N_{145} to N_{175} ratios were ~ 3−5 for successive flight screens of F_9/4; and moreover, N_{175} were as low as ~ 100 for the out-of-plume conditions during F_9/4. We think that ~ 100 rBC containing particles are too few for a quantitative analysis of coating thickness. This is other reason why we prefer to use 130−160 nm rBC cores for $T_{coating}$ discussion.

The discussions above were reflected in the revised manuscript.

(25) L352. The authors cannot simultaneously argue for more "aged" air outside the plume (and supported by NO_x/NO_y) and non-OS local emissions. Or, if they are going to do so, they need to make a more concrete and persuasive argument here in my opinion.

Our response: Only the argument on non-OS local emissions was kept in the revised manuscript.

(26) L356. Why would the authors not compare in/out of plume T^* values based on their NO_x/NO_y, rather than the "screen"? A lot of this discussion would benefit from a more direct link to photochemical age. Right now, the concept of photochemical age is seemingly left behind to earlier discussions, but it also belongs here in my opinion.

Our response: The dependence of rBC coating thickness on photochemical age will be presented elsewhere, together with quantitative discussions on the evolution of non-rBC containing particles (as mentioned in our response to the second specific comment raised by the second Referee).

(27) L376. In this discussion, the authors should note more explicitly that they are working from the small-size side of things, compared to these other studies. Generally, one might expect

LEO to be more robust for larger particles, with larger signal. Of course, there is a limitation because as particles get too large the scattering detector is saturated and a full Gaussian cannot be fit. Overall, the point that there is uncertainty on the order of 10% in the optical diameter from the LEO fit is worth reporting, but the authors should note the issue that they are using small particles.

Our response: The uncertainties reported here were derived from all the detected non-rBC containing particles, including both relatively large and small ones. Nonetheless, we agree with the reviewer that the LEO fit should be more robust for larger particles. We also agree with the reviewer that "small particles" were used in this manuscript for the discussions on coating thickness, which can potentially lead to biases in interpretation. As mentioned in our response to Comment# 24, evolutions of coating thickness exhibited the same pattern for relatively "small" and "large" rBC cores, indicating that our interpretation of coating thickness and the related conclusions should be valid. The discussions above were reflected in the revised manuscript.

(28) Fig. 11b/L363. While true, the authors neglect that the OA/rBC is smaller for the in-plume conditions overall. While not all OA will be coated on rBC, that the OA/rBC is so much smaller in plume might lead one to think that the coating amount on the rBC in plume should be smaller compared to out of plume, opposite what is reported. This should be discussed, in my opinion.

Our response: The following sentences were added as suggested: "It should also be noted that the out-of-plume OA are dominated by pre-existing secondary organic aerosols formed from biogenic precursors (Liggio et al., 2016), which do not contribute to the formation of coating materials on rBC cores. This explains why the out-of-plume conditions have higher OA/rBC ratios but in general lower *T** compared to in plumes.".

References

Buffaloe, G. M., Lack, D. A., Williams, E. J., Coffman, D., Hayden, K. L., Lerner, B. M., Li, S. M., Nuaaman, I., Massoli, P., Onasch, T. B., Quinn, P. K., and Cappa, C. D.: Black carbon emissions from in-use ships: a California regional assessment, Atmos. Chem. Phys., 14, 1881–1896, 2014.

Cappa, C. D., Onasch, T. B., Massoli, P., Worsnop, D. R., Bates, T. S., Cross, E. S., Davidovits, P., Hakala, J., Hayden, K. L., Jobson, B. T., Kolesar, K. R., Lack, D. A., Lerner, B. M., Li, S. M., Mellon, D., Nuaaman, I., Olfert, J. S., Petäjä, T., Quinn, P. K., Song, C., Subramanian, R., Williams, E. J., and Zaveri, R. A.: Radiative absorption enhancements due to the mixing state of atmospheric black carbon, Science, 337, 1078–1081, 2012.

Kondo, Y., Matsui, H., Moteki, N., Sahu, L., Takegawa, N., Kajino, M., Zhao, Y., Cubison, M. J., Jimenez, J. L., Vay, S., Diskin, G. S., Anderson, B., Wisthaler, A., Mikoviny, T., Fuelberg, H. E., Blake, D. R., Huey, G., Weinheimer, A. J., Knapp, D. J., and Brune, W. H.: Emissions of black carbon, organic, and inorganic aerosols from biomass burning in North America and Asia in 2008, J. Geophys. Res., 116, D08204, doi:10.1029/2010JD015152, 2011.

Laborde, M., Crippa, M., Tritscher, T., Jurányi, Z., Decarlo, P. F., Temime-Roussel, B., Marchand, N., Eckhardt, S., Stohl, A., Baltensperger, U., Prévôt, A. S. H., Weingartner, E., and Gysel, M.: Black carbon physical properties and mixing state in the European megacity Paris, Atmos. Chem. Phys., 13, 5831–5856, 2013.

Laborde, M., Mertes, P., Zieger, P., Dommen, J., Baltensperger, U., and Gysel, M.: Sensitivity of the Single Particle Soot Photometer to different black carbon types, Atmos. Meas. Tech., 5, 1031–1043, 2012.

Liggio, J., Gordon, M., Smallwood, G., Li, S. M., Stroud, C., Staebler, R., Lu, G., Lee, P., Taylor, B., and Brook, J. R.: Are emissions of black carbon from gasoline vehicles underestimated? Insights from near and on-road measurements, Environ. Sci. Technol., 46, 4819–4828, 2012.

Liggio, J., Li, S. M., Hayden, K., Taha, Y. M., Stroud, C., Darlington, A., Drollette, B. D., Gordon, M., Lee, P., Liu, P., Leithead, A., Moussa, S. G., Wang, D., O'Brien, J., Mittermeier, R. L., Brook, J. R., Lu, G., Staebler, R. M., Han, Y., Tokarek, T. W., Osthoff, H. D., Makar, P. A., Zhang, J., Plata, D. L., and Gentner, D. R.: Oil sands operations as a large source of secondary organic aerosols, Nature, 534, 91–94, 2016.

McMeeking, G. R., Hamburger, T., Liu, D., Flynn, M., Morgan, W. T., Northway, M., Highwood, E. J., Krejci, R., Allan, J. D., Minikin, A., and Coe, H.: Black carbon measurements in the boundary layer over western and northern Europe, Atmos. Chem. Phys., 10, 9393–9414, 2010.

Moteki, N., Kondo, Y., Miyazaki, Y., Takegawa, N., Komazaki, Y., Kurata, G., Shirai, T., Blake, D. R., Miyakawa, T., and Koike, M.: Evolution of mixing state of black carbon particles: aircraft measurements over the western Pacific in March 2004, Geophys. Res. Lett., 34, L11803, doi:10.1029/2006GL028943, 2007.

Schwarz, J. P., Spackman, J. R., Fahey, D. W., Gao, R. S., Lohmann, U., Stier, P., Watts, L. A., Thomson, D. S., Lack, D. A., Pfister, L., Mahoney, M. J., Baumgardner, D., Wilson, J. C., and Reeves, J. M.: Coatings and their enhancement of black carbon light absorption in the tropical atmosphere, J. Geophys. Res., 113, D03203, doi:10.1029/2007JD009042, 2008.