Reviewer responses for 'Dilution impacts on smoke aging: Evidence in BBOP data'

We thank the reviewers for their helpful comments. To aid the review process, we are placing reviewer comments in black text, our responses in blue text, any changes to the text in red, and, in some instances, reproduce text from the previously submitted manuscript (*italic magenta*). We have numbered the reviewer comments to assist the conversation.

Due to the length of the reviews and responses we provide here the page numbers of the start of each review:
Review 1 and responses: page 3
Review 2 and responses: page 24

First we would like to note that we found a minor error in our code that calculates the locations of the lowest 10% of out-of-plume CO that we use to determine our background region. This error led to us not including all of the locations (indexes) of this background region for each flight. Fortunately, when we fixed the error, none of our conclusions changed and all values shifted only slightly. We have updated all figures, tables, and text that depends on background corrections and note that the changes in our moderate and strong correlation coefficients (see Fig. 2 for instance) do not exceed 8%.

We note the recent publication of Lee et al. (2020) that focuses on aerosol optical properties in a southwestern US wildfire that has also looked at differences between edge and core. We have added the following text in Sect. 3.1 (new text underlined for emphasis)

(Garofalo et al., 2019) segregated smoke data from the WE-CAN field campaign by distance from the center of a given plume and showed that the edges of one of the fires studied have less f_{60} and more f_{44} (not background-corrected) than the core of the plume; Lee et al. (2020) saw similar patterns in a southwestern United States wildfire.

And

We do not have UV measurements that allow us to calculate photolysis rates but the in-plume SPN1 shortwave measurements in the visible show a dimming in the fresh cores that has a similar pattern to f_{44} and the inverse of f_{60} (Fig. S26; the rapid oscillations in this figure could be indicative of sporadic cloud cover above the plumes). (Lee et al. 2020) similarly saw indications of enhanced photochemical bleaching at the edges of a southwestern United States wildfire when examining aerosol optical properties.

Lee, J. E., Dubey, M. K., Aiken, A. C., Chylek, P., & Carrico, C. M. (2020). Optical and chemical analysis of absorption enhancement by mixed carbonaceous aerosols in the 2019

Woodbury, AZ fire plume. Journal of Geophysical Research: Atmospheres, 125, e2020JD032399. https://doi.org/10.1029/2020JD032399

We have noticed that we did not include any discussion of the fit equations that we developed (Eqs. 4-5 in the revised manuscript), despite spending significant time on them in the text. We have included the following statements in the conclusions:

"We have developed fit equations that can weakly to moderately predict Δf_{60} , Δf_{44}, $\Delta O/\Delta C$, and mean aerosol diameter given a known initial (at the time of first measurement) total organic aerosol mass loading and physical age."

"We also suggest further refinement of our fit equations, as further variables (such as photolysis rates) and better quantification of interfire variability (such as variable emission rates) are anticipated to improve these fits."

Finally, we note that we have made updates to many SI figures. In order to hopefully keep this document more navigable, we only rarely have included an updated SI figure here and instead point the reviewers to our marked-up SI document to assess these changes. We have also made many small edits to the main text to improve sentence structure, readability, and grammar (as noted a few times specifically by reviewer 2).

Review 1

Overall, I find this an interesting paper that addresses an important topic and builds nicely on previous work by the authors. However, I have a number of concerns regarding the inherent assumptions made or implied throughout and how thoroughly they are justified, and regarding the consistency of the interpretations provided. I find there are also a number of areas where more detail is required. I think that this work might be publishable after substantial revision. My specific comments and questions follow below.

R1.1) L54: It is not clear to me how plume thickness controls gas-particle partitioning or particle coagulation rates. Both depend on concentrations, not thickness. I suggest the authors clarify whether they really mean "thickness" here and on L58.

We agree that "thickness" is vague and that "concentration" is more clear. We have changed "thickness" to "aerosol concentration" in both instances as we are really referring to the aerosol concentration.

R1.2) L65: Do oxidant concentrations not also depend on the composition of the plume?

Yes, this was an oversight on our part. We have updated the text to read:

"In turn, oxidant concentrations depend on shortwave fluxes (Tang et al., 1998; Tie, 2003; Yang et al., 2009) and the composition of the plume (Yokelson et al. 2009; Akagi et al. 2012; Hobbs et al. 2003; Alvarado et al. 2015)."

R1.3) L67: The authors cite Formenti et al. (2003) as support of dilution occurring. However, they might note that the particular conclusion in Formenti et al. (2003) really derives from the observation of a single, high concentration point for the "fresh" samples that controls the linear regression. If that point is excluded, the slopes of the fresh and aged EC vs. OC curves are nearly identical.

This point is a subtlety that we did not capture with our original statement. Upon re-review of Formenti et al. (2003), we see that the authors state "...as our data for the elemental versus organic carbon ratio suggest that organic carbon might have evaporated while in the atmosphere." (Sect 3.4) However, the authors do not directly explicitly connect evaporation with dilution in their manuscript, and we have chosen to remove this citation. We replace it with (Garofalo et al. 2019; Grieshop et al. 2009).

R1.4) L79: Much of this paragraph seems redundant with material already presented. I suggest it be streamlined. The only new information is the slightly greater information regarding coagulation.

We respectfully disagree and believe this paragraph stands alone--it connects prior discussion to climate-relevant aerosol properties, which have not been discussed yet.

R1.5) L94: I suggest that the authors here define what they mean by "initial." This is a critical feature of this study. Only later is it clear that "initial" means "the closest we got to the fire for a given flight."

This is another oversight on our part--we have updated the text to read:

"A range of initial (at the time of the first plume pass in the aircraft) plume OA mass concentrations were captured within these flights and sufficiently fast (1 second) measurements of aerosols and key vapors were taken."

R1.6) L112: The authors should note the size range of the SP-AMS measurements, and the size range of the SP2 measurements (L126).

We have added the following text (with added underlines as guides) for the SP-AMS:

"A Soot Photometer Aerosol Mass Spectrometer (SP-AMS) provided organic and inorganic (sulfate, chlorine, nitrate, ammonium) PM_1 aerosol masses (Canagaratna et al. 2007), select fractional components (the fraction of the AMS OA spectra at a given mass-to-charge ratio) (Onasch et al., 2012), and elemental analysis (O/C and H/C) (Aiken et al., 2008; Canagaratna et al., 2015). The SP-AMS had the highest sensitivity between 70-500 nm, dropping to 50% transmission efficiency by 1000 nm (Liu et al. 2007). "

And for the SP2:

" A Single-Particle Soot Photometer (SP2; Droplet Measurement Technologies) was used to measure refractory black carbon (rBC) between 80-500 nm (Schwarz et al. 2010) …"

R1.7) L125: The authors might also note that the atomic ratios are strongly affected by mixing of different air masses and the co-oxidation of different VOC precursors, which start at different points on a van Krevelen diagram. Different VOCs in the plumes will age on a variety of timescales, giving rise to an evolving O:C and H:C regardless of "aging" of the sort implied here. Mixing and co-oxidation affect the H:C, especially, making inferences of the "types of reactions occurring" challenging. This is discussed in (Chen et al., 2015). See later comment on the same subject.

We agree that we did not expand upon this discussion as much as we could have, and we thank the reviewer for the helpful reference. We have expanded this discussion as follows:

"O/C tends to increase with oxidative aging (Decarlo et al., 2008) whereas H/C ranges from increasing to decreasing with oxidative aging, depending on the types of reactions occurring (Heald et al., 2010). Changes in O/C and H/C are also influenced by mixing of different air masses and co-oxidation of different VOC precursors (Chen et al. 2015). Thus, tracking H/C with aging may provide clues upon the types of reactions that may be occuring; however, variable oxidation timescales can make inferences of this type difficult (Chen et al. 2015)."

In our analysis, we background-correct C, O, and H (creating ΔC, ΔO, and ΔH) and present the ratios as ΔO:ΔC and ΔH:ΔC. The mixing of background OA into the plume should have no direct impact on ΔO:ΔC and ΔH:ΔC (although there may be indirect impacts through changing chemistry).

R1.8) L130: The authors note that the supporting info provides "more details on the instruments used." I find this misleading. The information provided in the SI is extremely limited, hardly greater than that provided in this paragraph. I suggest the authors provide in the SI some discussion at least of instrumental uncertainties.

We agree that our SI is sparse on details of the BBOP instrumentation. Our coauthor Lawrence Kleinman's current ACPD paper also on BBOP aerosol properties has a significant amount of detail on the SP-AMS, the SP2, the FIMS, and trace gas instruments. We will refer the reader to this text for those details. As well, we flushed out our discussion in the SI:

The Fast Integrated Mobility Spectrometer (FIMS) characterizes particle sizes based on electrical mobility as in scanning mobility particle sizer (SMPS). Because FIMS measures particles of different sizes simultaneously instead of sequentially as in traditional SMPS, it provides aerosol size distribution with a much higher time resolution at 1 Hz (Wang et al., 2017). The relative humidity of the aerosol sample was reduced to below ~25% using a Nafion dryer before being introduced into the FIMS. Therefore, the measured size distributions represented that of the dry aerosol particles. The particle number concentration integrated from FIMS size distribution typically agrees with the CPC 3010 (Condensation Particle Counter) measurement (Kleinman et al. 2020) within ~ 15% when size distribution suggests that particles smaller than 10 nm contribute negligibly to the total number concentration. Thus, we estimate the uncertainty in the FIMS number concentration to be ~15%. The uncertainty in measured particle size is about 3% (Wang et al. 2017).

The Soot Particle Aerosol Mass Spectrometer (SP-AMS) is thoroughly detailed in Kleinman et al. (2020). Although it was not directly characterized for uncertainties during the BBOP campaign, we estimate uncertainties as follows. The AMS uncertainty is estimated

following the methods in (Bahreini et al. 2009) (first equation of their supplemental information), leading to 37% uncertainty for organics. The laser vaporizer adds additional uncertainty up to 20%. Thus summing the uncertainties in quadrature leads to a 42% uncertainty in organics. The Soot Photometer (SP2) had an uncertainty of 20%.

CO measurement uncertainties are detailed in Kleinmen et al. (2020): the Off-Axis Integrated Cavity Output Spectroscopy was found to have an accuracy of 1-2%, and the precision at ambient backgrounds of 90 ppb was 0.5 ppbv RMS (using a 1 second averaging).

An SPN1 radiometer (Badosa et al. 2014; Long et al. 2010) provided total shortwave irradiance, with a shaded mask applied following (Badosa et al. 2014). The data was corrected for tilt up to 10 degrees of tilt, following (Long et al. 2010). For tilt greater than 10 degrees these values are set to "bad". Instrument uncertainties are detailed in (Badosa et al. 2014).

Badosa, Jordi, John Wood, Philippe Blanc, Charles N. Long, Laurent Vuilleumier, Dominique Demengel, and Martial Haeffelin. 2014. "Solar Irradiances Measured Using SPN1 Radiometers: Uncertainties and Clues for Development." *Atmospheric Measurement Techniques* 7: 4267–83.

Bahreini, R., Ervens, B., Middlebrook, a. M., Warneke, C., de Gouw, J. a., DeCarlo, P.F., Jimenez, J.L., Brock, C. a., Neuman, J. a., Ryerson, T.B., Stark, H., Atlas, E., Brioude, J., Fried, A., Holloway, J.S., Peischl, J., Richter, D., Walega, J., Weibring, P., Wollny, a. G., and Fehsenfeld, F.C. (2009). Organic aerosol formation in urban and industrial plumes near Houston and Dallas, Texas. J. Geophys. Res., 114:D00F16.

Kleinman, L. I., Sedlacek III, A. J., Adachi, K., Buseck, P. R., Collier, S., Dubey, M. K., Hodshire, A. L., Lewis, E., Onasch, T. B., Pierce, J. R., Shilling, J., Springston, S. R., Wang, J., Zhang, Q., Zhou, S., and Yokelson, R. J.: Rapid Evolution of Aerosol Particles and their Optical Properties Downwind of Wildfires in the Western U.S., Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2020-239, in review, 2020.

Wang, J., Pikridas, M., Spielman, S. R., and Pinterich, T.: A fast integrated mobility spectrometer for rapid measurement of sub-micrometer aerosol size distribution, Part I: Design and model evaluation, J. Aerosol Sci., 108, 44-55, 10.1016/j.jaerosci.2017.02.012, 2017.

R1.9) L138: I suggest it be clarified how f60 and f44 are background corrected. Presumably this is not a straight difference, as the denominators ([OA]) differ. Is it, for example f60_corrected = (f60_plume*[OA]_plume – f60_bgd*[OA]_bgd)/[OA]_plume? If the authors used a straight difference, this must be justified as it does not seem appropriate to me. Similarly, more details on how the other intensive properties (O:C, H:C) are corrected are needed.

We calculated the background corrected f60 and f44 as follows (where $f = f_{60}$ or f_{44}):

$$\Delta f = \frac{(f_{in} * OA_{in}) - (f_{out} * OA_{out})}{\Delta OA} \quad \text{Eq. R1}$$

Similar, the $\Delta O/\Delta C$ and $\Delta H/\Delta C$ are calculated through (where X = O or H):

$$\frac{\Delta X}{\Delta C} = \frac{(X_{in\ plume} - X_{out\ of\ plume})}{(C_{in\ plume} - C_{out\ of\ plume})} \quad \text{Eq. R2}$$

We've added Eqs. R1-R2 as Eqs. 1 and 2 in the main text and have updated other equation numbers and references.

R1.10) L140: It would be helpful if in Figs. S2-S6 and S7-S11 the authors would number each plume so that the two can be related to each other. It would also help if the time-series were shown as an additional panel with the spatial plots, again so comparisons can be made. I think this is important because the authors discuss "plumes" here, but they do not discuss how it is, for example, that in a given transect there can be multiple maxima in CO. Does this imply there are two plumes? Or is this the same plume? What drives this behavior, and what might it indicate about the evolution of the plumes? What does it mean to define a "centerline" of the plume if there are clearly two distinct maxima on either side (see Fig. S3, for example).

We have included subplots for figures S2-S6 that show both the flight tracks colored by time in minutes as well as the leg numbers as designated in the BBOP database (as designated by the flight team). We've updated the x-axis of figures S7-S11 to be in minutes to allow for easier comparisons between the two. We agree that the "centerline" is an imperfect metric and is a limitation of this study. However, the centerlines have been determined using the most-concentrated portion of the aerosol number concentration, which did tend to be more clear (see e.g. Fig. S1). We added more text about the centerline, also following comments from R2.25:

"To estimate the physical age of the plume, we use the estimated fire center as well as the total FIMS number concentration to determine an approximate centerline of the plume as the smoke travels downwind (Figs. S1-S6). The centerline is subjectively placed to attempt to capture the most-concentrated portion of the total number concentration for each plume pass, as we focus on aerosol properties and their relations to dilution in this study. We use mean wind speed and this estimated centerline to get an estimated physical age for each transect. We did not propagate uncertainty in fire location, wind speed, or centerline through to the physical age, which is a limitation of this study."

We have also added the following text to the first paragraph of section 3 to discuss the potential of multiple plumes (underlines for the new material):

"We have divided each transect into four regions: between the 5-15 (edge), 15-50 (intermediate, outer), 50-90 (intermediate, inner), and 90-100 (core) percentile of ΔCO within each transect.

Fig. 1 shows the edge and core data, both averaged per transect, with Figs. S14-18 providing all four percentile bins for each flight. These percentile bins correspond with the thinnest to thickest portions of the plume, respectively, and if a fire has uniform emissions ratios across all regions and dilutes evenly downwind, these percentile bins would correspond to the edges, intermediate regions, and the core of the diluting plume. We use this terminology in this study but note that uneven emissions, mixing, and/or dilution lead to the percentile bins not corresponding physically to our defined regions in some cases. We note that some plumes show more than one maxima in CO concentrations within a given plume crossing, which implies that there may be more than one fire or fire front, and that these plumes from separate fires or fronts are not perfectly mixing. As well, in at least one of the fires (in flights '730a' and '730b'), the fuels vary between different sides of the fire, as discussed in Kleinman et al., (2020)."

R1.11) From Figs. S7-S11, it appears that the background [CO] varies from flight-to-flight. For example, in Fig. S7 the background is clearly lower than the 150 ppb threshold the authors have used, but in Fig. S9 it is barely sufficient. Why not define a flight-specific background [CO] based on the observations?

We agree that the background CO is variable from flight to flight. However, we performed a sensitivity analysis on the background CO cutoff (using a cutoff of 200 ppbv instead of 150 ppbv), shown in Fig. S20, and the results do not qualitatively change our conclusions. This is briefly discussed in lines 205-208, "*Figs. S13, S19-S21 show the same details as Fig. 2 but provide sensitivity tests to potential FIMS measurement artifacts (Fig. S13) and our assumed background CO and ΔCO percentile spacing (Figs. S19-S21). Although these figures show slight variability, the findings discussed below remain robust, and we constrain the rest of our discussion to the FIMS measurements, background and ΔCO percentile spacings used in Fig. 2.*"

R1.12) L156: The authors note that the instruments had various time lags, but it is not clear whether they were all adjusted to account for these varying time lags. This should be clarified. Also, it would be helpful if the authors clarified whether they really mean a "lag" but with a fast response time (i.e., two instruments both show sharp changes but are offset) or whether they are referring to some amount of smearing in which previous measurements affect the current measurement. From the FIMS discussion, it sounds as if they are actually talking about smearing (related to instrument response time) and not a lag.

The data was not time lag corrected, and we clarify this in the text now. Kleinman et al. (2020) provides further details on time lags--they did correct the data but note that "Time-shifts of 1-2 seconds are readily apparent as a degradation in correlation when comparing instruments. Maximizing correlations, however, does not accurately compensate for varying response time." From coauthor Kleinman's careful work and analyses, we believe that most of the instruments

display only a time lag, but that the FIMS displays both a time lag and some smearing. Given that analysis only using the first half of the FIMS data for each leg did not change our conclusions (see the methods section, specifically *"To test if these lags impact our results, we perform an additional analysis where we only consider the first half of each in-plume transect, when concentrations are generally rising with time (Figure S12-S13), and our main conclusions are unaffected."*) We have clarified in the text that the FIMS had additional smearing.

Kleinman, L. I., Sedlacek III, A. J., Adachi, K., Buseck, P. R., Collier, S., Dubey, M. K., Hodshire, A. L., Lewis, E., Onasch, T. B., Pierce, J. R., Shilling, J., Springston, S. R., Wang, J., Zhang, Q., Zhou, S., and Yokelson, R. J.: Rapid Evolution of Aerosol Particles and their Optical Properties Downwind of Wildfires in the Western U.S., Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2020-239, in review, 2020.

R1.13) L165: Further details regarding how the FIMS data were used to establish the centerline are needed. How were the number distributions used specifically? How were these determined for different transects to give a single straight line? Also, is wind speed as measured by the aircraft?

The centerline was subjectively determined to approximately capture the most-concentrated portion of the total number concentration for each plume pass, as we are focused on aerosol properties in this study (and their relation to concentration and dilution). We have added this text to the main document and do include this as a limitation of the study in the original text (new text underlined for clarity and along with some fixes to errors pointed out in R2):

"To estimate the physical age of the plume, we use the estimated fire center as well as the total FIMS number concentration to determine an approximate centerline of the plume as the smoke travels downwind (Figs. S1-S6). The centerline is subjectively placed to attempt to capture the most-concentrated portion of the total number concentration for each plume pass, as we focus on aerosol properties and their relations to dilution in this study. We use mean wind speed and this estimated centerline to get an estimated physical age for each transect. We did not propagate uncertainty in fire location, wind speed, or centerline through to the physical age, which is a limitation of this study."

R1.14) Fig. 1: The figure lacks error bars. Given the analysis, it would seem that precision-based propagated uncertainties would be appropriate, as the authors seem interested more in characterizing changes than they are absolute values. I suggest appropriate error bars are added.

We are quite hesitant to put forth a precision-based analysis. We are cautious to apply a known precision under ambient conditions to the sometimes extremely concentrated conditions of smoke plumes. For instance, our initial analyses included ozone measurements and UHSAS

particle size distribution measurements, but we had to remove both instruments due to unresolvable issues with interferences under plume conditions. The UHSAS became saturated--this saturation level may be changing as both a function of particle size and concentration (as was discovered from careful analysis of a UHSAS during strong pollution events during an indoor campaign and seen again during a controlled burn study; Erin Boedicker [Colorado State University; Farmer group], personal communication). Another issue is that propagating uncertainties assumes that precision is equivalent in all of the measurements. We are using multiple instruments so this assumption breaks down, as many instruments define and calculate precision differently. This makes a true apples-to-apples comparison (which is needed for propagation of errors) tricky or impossible. As discussed in response to other comments by this and the other reviewer, we have weakened the language of our results throughout due to these uncertainties.

R1.15) L182: While it seems that the 5-15 percentile values are primarily found at the physical edges of the plumes shown in the supplemental, as often as not the 90-100 percentile values exhibit bimodal behavior across a transect, often occurring relatively close to the physical edge. From what is shown, I do not believe it is justified to say that the 90-100 percentile "core" corresponds to the physical "core" of the plume as observed. I strongly suggest the authors to define a quantitative metric to relate the percentiles to the spatial distribution. Perhaps a normalized distance from the centerline.

We agree that the 5-15 and 90-100 percentiles do not perfectly line up to the physical edge and core, and state in the original manuscript (lines 177-182): "*These percentile bins correspond with the thinnest to thickest portions of the plume, respectively, and if a fire has uniform emissions ratios across all regions and dilutes evenly downwind, these percentile bins would correspond to the edges, intermediate regions, and the core of the diluting plume. We use this terminology in this study but note that uneven emissions, mixing, and/or dilution lead to the percentile bins not corresponding physically to our defined regions in some cases. However, the lowest two ΔCO bins tend more towards the physical edges of the plume and the highest two tend more towards the physical center of the plume (Figs. S2-S6).*"

(We note that we have added more material to the above quoted section, following comment R1.10). We argue that our 5-15, 15-50, 50-90, and 90-100 ΔCO percentile bins are our quantitative metric and that due to variable mixing between different smoke plumes as well as variable plume widths, defining a spatial relationship is not necessarily particularly informative. We add the following reminder to the manuscript in sect 3.1:

"We use the simple 'edge' and 'core' terminology throughout the following discussion but note that the 5-15 and 90-100 ΔCO percentile bins do not necessarily correspond to the physical (spatial) edges and cores of each plume. They instead correspond to the most CO-dense and least CO-dense portions of the plume.

R1.16) L191: I suggest the authors be more precise in their claims. The normalized number concentration in the "core" does not change with age, and at the edge the entirety of the change is observed from the first transect to the second. And there is perhaps an increase in diameter from the first transect to the next, but the diameter is constant (within variability) for all transects further downwind. Also, the deltaO/C does not increase with aging. The authors indicate that the delta_f44 changes with age, but it is not clear how this was determined. Was some sort of linear fit done? Is this just the difference between the first point and the last? Visually, the points look scattered about a flat line. Overall, for this discussion I think that the authors need to be more specific and precise and quantitative. As currently written, it is not always clear how the authors came to the conclusions that they did.

We agree that this flight shows weak trends for the majority of the metrics discussed, and that information on trends is only gained once all of the flights have been pooled together. Figure 1's primary purpose is to orient the reader to the different metrics and how they might look for a flight. We have changed this paragraph to read:

"Figure 1 shows that for this specific plume, ΔOA/ΔCO and ΔBC/ΔCO vary little with age for both the 5-15 and 90-100 percentile of ΔCO (p-values>0.5). A true Lagrangian flight with the aircraft sampling the same portion of the plume and no measurement artifacts (e.g. coincidence errors at high concentrations) would have a constant ΔBC/ΔCO for each transect. This flight and other flights studied here have slight variations in ΔBC/ΔCO (Fig. 1; Figs. S14-S18), which may be indicative of deviations from a Lagrangian flight path with temporal variations in emission and/or measurement uncertainties. The remaining variables plotted also show some noise and few clear trends, but it is apparent that the 5-15 and 90-100 percentiles do show a separation for many of the individual metrics. In order to determine the existence or lack trends for these metrics, we spend the remainder of this study examining each metric from all of the pseudo-Lagrangian flights together."

R1.17) L203: I find it exceptionally difficult to understand exactly what the authors have done with the Spearman rank-order correlation tests. The authors need to be much more specific. The authors have one value for (e.g.,) initial plume OA mass but then have multiple values for the deltaOA/deltaCO for each transect of a given plume. Then there are multiple plumes. How are the data merged to allow comparison across all plumes? Physical age makes more sense, as (for example) deltaOA/deltaCO can be regressed versus physical age for each plume. But, to me, how the other parameters are used (OA initial and deltaOAinitial) is unclear. Are all the initial OA values repeated for a given flight? Are the authors using only the initial values for the other parameters to compare with initial OA?

We see that our original text here is confusing and misleading. We have attempted to clarify it. We are using a single value for $\Delta OA_{initial}$ for each transect within a Lagrangian set of transects which is obtained from the first transect of the set. If a flight has two Lagrangian sets of

transects, there will be a different value of $\Delta OA_{initial}$ used for the two sets of transects, each again obtained from the first transect of each set. The original text may have been interpreted that we used $OA_{initial}$ but we did not--we have clarified that. We use the changing values of $\Delta OA/\Delta CO$, Δf_{60}, Δf_{44} $\Delta H/\Delta C$, and $\Delta O/\Delta C$ as they age downwind to compare with initial OA. We have updated this text (also following suggestions made in R1.20):

"Also included in Fig. 2 are the Spearman rank-order correlation tests (hereafter Spearman tests), which are tests for monotonicity. The Spearman tests show correlation coefficients for each flight set (Table S1) with the initial ΔOA of a flight set ($\Delta OA_{initial}$) against $\Delta OA/\Delta CO$, Δf_{60}, Δf_{44} $\Delta H/\Delta C$, and $\Delta O/\Delta C$ as each variable ages downwind. We also include Spearman tests for the calculated physical age of the smoke for each flight set against these same variables. The R values are labeled $R_{\Delta OA,initial}$ and R_{age}, respectively, in Fig. 2. For the correlations with $\Delta OA_{initial}$, all transects in a given Lagrangian set of transects have the same $\Delta OA_{initial}$ value; for flights with two Lagrangian set of transects, each set has its own $\Delta OA_{initial}$ value. Correlating to $\Delta OA_{initial}$ provides an estimate of how the plume aerosol concentrations at the time of the initial transect impact plume aging (aging both before and after this initial transect)."

R1.18) L213: What does it mean for something to "evaporate off through heterogeneous aging?" Things can evaporate, or they can be heterogeneously oxidized. These are distinct processes.

We agree that the language here is misleading, and have updated the text to read:

" Δf_{60} generally decreases with plume age (R_{age} = -0.26; a weak correlation), consistent with the hypotheses that Δf_{60} may be evaporating because of dilution, undergoing heterogeneous oxidation to new forms that do not appear at m/z 60, and/or having a decreasing fractional contribution due to condensation of other compounds."

R1.19) L210: The authors note that the changes in deltaOA/deltaCO with aging are small. A recent review by the authors (Hodshire et al., 2019) indicates a variety of reasons for such behavior. Another recent paper (Lim et al., 2019) introduces another potential reason for this behavior, specifically potential biases in the measurement of OA as the particle composition evolves. Have the authors considered this?

We agree that variable collection efficiency and related measurement artifacts could in theory bias OA measurements. We realized that we did not include in the original manuscript the characterized collection efficiencies (CE) of the SP-AMS, found to have two different efficiencies for when the laser was on (CE=0.76) or off (CE=0.5) and we include those details in the text now. We did not characterize any changes in efficiency with aging. This is an on-going topic of debate within the AMS community (and is addressed within the SI of the abovementioned paper from our group, Hodshire et al. 2019), and we briefly address it as a limitation of this study. We have included these details in the methods section:

"It [the SP-AMS] was characterized to have a collection efficiency of 0.5 when the laser was off and 0.76 when the laser was on during the BBOP campaign, and these corrections have been applied to the data. We do not attempt to characterize whether the collection efficiency of the SP-AMS changes as the aerosol ages. This may be a limitation of this study, as collection efficiency has been recently observed to decrease with aging within a laboratory study of biomass burning (Lim et al. 2019). However, no consistent evidence of changing collection efficiencies in field studies exist yet."

R1.20) With reporting the Spearman's correlation coefficient I suggest the authors use consistent language that links to typical interpretation of the level of significance (that a relationship is monotonic). For example, a value of -0.25 (as determined for f60) might be considered "weak" while a value of 0.54 (for f44) is "moderate." Also, the authors might note when introducing the Spearman's test that it is a test for monotonicity.

Thank you for these suggestions. We now note in the text that the Spearman tests are a test for monotonicity when we first mention it in the text, and have added the following definitions that we use throughout the text each time we discuss an R value (and we also have updated our language for R^2 to reflect these categories as well as emphasizing that R^2 is explaining a given fraction of the variance):

"We define the following categories of correlation for the absolute value of R: 0.0-0.19 is 'very weak', 0.2-0.39 is 'weak', 0.4-0.59 is 'moderate', 0.6-0.79 is 'strong', and 0.8-1.0 is 'very strong' (Evans, 1996)."

Evans, J. D. (1996). Straightforward statistics for the behavioral sciences. Thomson Brooks/Cole Publishing Co.

R1.21) There appears to be a good deal of flight-to-flight variability in behavior, from Fig. 2. This raises a question of how much of the inferred behavior (from the Spearman's test) derives from fairly strong changes in one flight. The authors might consider testing the sensitivity to their analysis by determining Spearman's coefficients when systematically leaving out individual flights or transects one at a time. This would give a broader sense of the robustness of the results, given the notable scatter.

We have performed the Spearman's test for R_{age} and $R_{\Delta OA,\ initial}$ for all metrics of Figure 2 leaving one flight out at a time. The results are summarized in Table S2. We add the following text when we first introduce the R values:

"As individual flights show scatter in the metrics of Fig. 2 (Figs. 1, Figs. S14-S18), we also include $R_{\Delta OA,initial}$ and R_{age} for each metric of Fig. 2 systematically sequentially removing one

flight from the statistical analysis. These results are summarized in Table S2. In general, removing single flights does not change our conclusions, particularly when correlations are moderate or stronger."

We provide the range of these results within the text as each metric is discussed.

R1.22) L217: Nitpicky, but compounds do not "contain f44." Certain compounds fragment in such a way that they show up at m/z 44 in the AMS. But overall this sentence is a run on with a second half that does not logically follow from the first. The sentence starts by talking about a balance between condensation and evaporation but shifts abruptly to note something about heterogeneous oxidation or particle-phase reactions. I suggest the authors clarify the point they are aiming to make here.

This is a reasonable point and we have updated the text to here to read (including updates as suggested by reviewer 2's comment R2.43):

"Δf_{60} generally decreases with plume age (R_{age} = -0.26), consistent with the hypotheses that Δf_{60} may be evaporating because of dilution, undergoing heterogeneous oxidation, and/or having a decreasing fractional contribution due to condensation of other compounds.. In contrast, Δf_{44} generally increases with age (R_{age} = +0.5) for all plumes with available data. It appears for the plumes in this study that although there is little change in $\Delta OA/\Delta CO$, loss of compounds that contain f_{60} fragments (as captured by the SP-AMS) is roughly balanced by condensation of more-oxidized compounds, including those that contain compounds with f_{44} fragments, such as carboxylic acids. This observation suggests the possibility of heterogeneous or particle-phase oxidation that would alter the balance of Δf_{60} and Δf_{44}."

R1.23) L219: The authors note that deltaOA/deltaCO does not change much. This would be consistent with the little mass loss that the authors note from heterogeneous oxidation here, correct? Are the authors aiming to make a point more specifically about the efficiency with which heterogeneous oxidation might degrade the f60 signal and not about mass loss? I find it unclear.

We are trying to note that heterogeneous chemistry is relatively slow (for near-field aging) and shouldn't significantly contribute to evaporative or compositional changes. We have added text to emphasize that point more clearly:

"However, estimates of heterogeneous mass losses indicate that after three hours of aging for a range of OH concentrations and reactive uptake coefficients, over 90% of aerosol mass is

anticipated to remain, indicating that heterogeneous loss has limited effect on aerosol composition or mass (Fig. S23; see SI text S2 for more details on the calculation). Hence, the evaporation of f_{60} being balanced by gas-phase production of f_{44} may be the more likely pathway."

R1.24) Laboratory observations (Cubison et al., 2011;Hennigan et al., 2011;Hodshire et al., 2019;McClure et al., 2020) have demonstrated that the f60 and f44 of freshly emitted particles vary over large ranges dependent on the fuel type and specific burn condition. Is it not possible that the differences in deltaf60 and deltaf44 between flights result from intrinsic differences in the emitted particle properties? The authors seem to discount this without explicit justification when they state that their interpretation assumes that "emitted deltaf60 and deltaf44 do not correlate with deltaOAinitial." Might there not be an initial correlation, as this might indicate some difference in the burn conditions or the particular fuel mix? I can certainly believe that "evaporation and/or chemistry likely occurred before the time of" the first measurements, however it is not clear to me that the observations as presented here demonstrate this conclusively. Also, given that different sources produce particles that have different initial f60 and f44, would they be expected to exhibit the same deltaf60 and deltaf44 even if initial OA and dilution were identical? Is there evidence that this is expected?

The reviewer makes reasonable points here and we agree that these are alternative hypotheses that should be explicitly discussed in the manuscript. Reviewer 2 made similar comments in R2.47. Unfortunately, lacking direct measurements of the emissions, we cannot explore this hypothesis in any detail. And we do find it compelling that less-dense plumes do show higher f44/lower f60 than more-dense plumes, which supports our hypothesis of aging prior to the transect. We have added the following text to Sect. 3.1:

"We note that each fire may emit particles with variable initial f_{44} and f_{60} values, as has been observed in laboratory studies (Hennigan et al. 2011; Cubison et al. 2011; McClure et al. 2020), which adds to scatter within the data. It is possible that variability in f44 and f60 may also contribute to the observed correlations with ΔOA_{intial}; however, this would require that higher f_{44} emissions are correlated with lower emissions rates and/or faster dilution rates (and visa versa for f_{60}). Lacking direct emissions measurements, this hypothesis cannot be further explored in this work."

To the reviewer's last query ("Also, given that different sources produce particles that have different initial f60 and f44, would they be expected to exhibit the same deltaf60 and deltaf44 even if initial OA and dilution were identical? Is there evidence that this is expected?"), we would not expect the same Δf_{44} and Δf_{60} under those circumstances and thus variability from emissions likely contributes to the noise of our fit parameters. We do include a brief discussion

on this in the text in Sect. 3.1 within the discussion of our fit parameters (with new minor edits addressing comments from reviewer 2):

"The scatter is likely due to variability in emissions due to source fuel or combustion conditions, instrument noise and responses under the large concentration ranges encountered in these smoke plumes, inhomogeneous mixing within the plume, variability in background concentrations not captured by our background correction method, inaccurate characterizations of physical age due to variable wind speed, and/or deviations from a true Lagrangian flight path."

R1.25) L243: I disagree with the authors interpretation of the van Krevelen diagram here. The authors interpret this in a process based way related to chemistry. However, this does not account for the fact that this is, likely, ultimately a mixing experiment wherein primary OA is being increasingly mixed with secondary OA. This cannot be interpreted in terms of functional group addition. Additionally, it is not clear that a plot of deltaO/deltaC vs deltaH/deltaC should behave in the same way as a plot of O/C vs H/C. The authors must demonstrate the equivalency of these.

We think that the reviewer has interpreted our work to mean that we have calculated delta(H/C) and delta(O/C) (we did not calculate this), rather than delta(H)/delta(C) and delta(O)/delta(C) (which is the calculation we did do). We hope that our response to the earlier reviewer comment R1.9 clarifies this matter. We remind the reader in the text of this here, (underlines for new additions):

A Van Krevelen diagram of ΔH/ΔC versus ΔO/ΔC (Fig. S27) indicates that oxygenation reactions or a combination of oxygenation and hydration reactions are likely dominant (Heald et al., 2010) (recalling that ΔH/ΔC and ΔO/ΔC are calculated by background-correcting the individual elements before ratioing; Eq. 1)

It is true that any non-linear changes in chemistry and composition will mean that our delta(H)/delta(C) and delta(O)/delta(C) method will not perfectly isolate the elemental factors from smoke, and we add this disclaimer in the methods:

"We note that any non-linear changes in chemistry and composition between the plume and background will not perfectly isolate the elemental factors in smoke."

R1.26) deltaO/deltaC ratios: I am somewhat surprised that these values are positive. O:C ratios of fresh biomass burning tend to be around 0.3-0.4 whereas O:C of background OA are typically large. (The same is true for f44.) The authors should comment on the very fact that their deltaO/deltaC values are positive.

We think that the reviewer has interpreted our work to mean that we have calculated delta(H/C) and delta(O/C) (we did not calculate this), rather than delta(H)/delta(C) and delta(O)/delta(C)

(which is the calculation we did do). delta(H)/delta(C) and delta(O)/delta(C) represent estimates of the H:C and O:C of the smoke OA, which cannot physically be negative; and it would be highly unlikely that delta(H), delta(C), and delta(O) are negative as this would require the background concentration of these elements to be higher than the plume concentrations We hope that our response to the earlier reviewer comment R1.9 clarifies this matter and explains the positive values.

R1.27) Eqn. 2: First, what is the justification for this functional form? Is there some other form that would better explain the data? Second, in terms of utility, is it really most useful to predict the delta values, as these will depend explicitly on the background, which may vary between locations? Do the authors expect these relationships will prove robust and applicable to other regions? Would these be appropriate at night as well as during the day? The authors have not been able to distinguish between dilution-driven changes and oxidation-driven changes, so there may be distinct day/night differences? When would they expect them applicable? How could these parameters assist specifically in biomass burning models? Presumably such models would aim to be processed based, differentiating between oxidation and dilution.

Reviewer 2 had similar questions in comment R2.56). We do not agree with the comment about delta values here. The delta values mean that the background has been subtracted off in an attempt to isolate the smoke contributions. Hence, in the absence of non-linear interactions between the smoke and background species, the delta values do not depend on the background. The non-delta values (the smoke+background values) much more explicitly depend on the background.

We do agree that it's as yet unclear whether these fits are appropriate for other regions of the world as well as day/night differences. We tried a large number of mathematical fits and these equations (Eqs. 2-3 in the original text; Eqs. 4-5 in the updated text) performed the best. They do not have a direct physical meaning. The parameters would need significantly more testing to be applicable for models, and we have added the following text to address these comments:

"Eqs. 4-5 performed the best out of the mathematical fits that we tested. These equations do not have a direct physical interpretation but may be used as a starting point for modeling studies as well as for constructing a more physically based fit. There may be another variable not available to us in the BBOP measurements that can improve these mathematical fits, such as photolysis rates. We do not know whether these fits may well-represent fires in other regions around the world, given variability in fuels and burn conditions. We also do not know how these fits will perform under nighttime conditions, as our fits were made during daytime conditions with different chemistry than would happen at night. We encourage these fits to be tested out with further data sets and modeling. These equations are a first step towards parameterizations appropriate for regional and global modeling and need extensive testing to separate influences of oxidation versus dilution-driven evaporation."

R1.28) When the authors report the Pearson's coefficients, are these constrained to go through the origin? The authors show only the 1-1 lines, but visually it seems that any linear fit to the calculated vs. observed relationship will have a non-zero intercept unless constrained. In this context, having a good r^2 value is simply an indication of a linear relationship but it is not an indication of the goodness of the calculated vs. observed. Instead, the authors would need to provide some metric such as normalized mean bias. As presented, I am not convinced that the r^2 values are particularly meaningful.

We do not constrain the Pearson's coefficients to go through the origin. We have now calculated the normalized mean bias (NMB) and normalized mean error (NME), as the normalized mean bias is likely to be small given that we're minimizing the linear fit. We include the NMB and NME values in our Figures 3 and S28-29. We have updated figures and figure captions accordingly. We add the following sections of text to Sect. 3.1 and 3.2:

1. (Section 3.1) "We do not constrain our fits to go through the origin. To provide further metrics of goodness-of-fit, we also include the normalized mean bias (NMB) and normalized mean error (NME) in percent for each metric of Fig. 3. The NMB values are very close to zero (which is anticipated as linear fits seek to minimize the sum of squared residuals). The NME is more variable, at 18.8% for Δf_{60}, 14.9% for Δf_{44}, and 10.4% for $\Delta O/\Delta C$."

2. (Section 3.1) "Other functional fits were explored, with

$$ln(\Delta X) = a\ ln(\Delta OA_{initial}) + b\ ln(Physical\ age)\ + c \qquad \text{Eq. 5}$$

(Fig. S28 and Table S4 for the fit coefficients) and $\Delta N_{initial}$ in the place of $\Delta OA_{initial}$ in Eq. 42 (Fig. S29 and Table S5 for the fit coefficients) providing similar correlation values and NMB and NME values for Δf_{60}, and Δf_{44}, and $\Delta O/\Delta C$."

3. (Section 3.2) We also perform the functional fit analysis following Sect. 3.1 (Eq. 4; where X is $\underline{D_p}$ in this case). The fit can also weakly predict greater than 30 percent of the variance in $\underline{D_p}$ (R_p^2 and R_s^{2s} of 0.36 and 0.31 and NME of 5.6%; Fig. 3d) but does not well-predict $\Delta N_{40\text{-}300\ nm}/\Delta CO$ (not shown). We show the functional fit for $\underline{D_p}$ for the alternative fit equation (Eq. 5) in Fig. S28 and Table S4. We also show the functional fits for $\underline{D_p}$ for Eq. 4 with $\Delta N_{initial}$ in place of $\Delta OA_{initial}$ in Fig. 29 and Table S5.

R1.29) L263: It is not clear to me what the authors are getting at when they state that aged deltaf60 and deltaf44 show scatter, limiting the predictive skill of measurements available from BBOP. They had just discussed how there are "moderate goodness of fits." It seems now that

they are contradicting themselves. Or perhaps they are just providing more context for what "moderate" means.

We have updated our language when discussing the correlation metrics to be consistent throughout, following comment R1.20). Reviewer 2 had similar comments in R2.57), and we provide our response to that comment here:

We were referring to the aging values of Δf60 and Δf44, we were not careful in our language here though. "Limiting the predictive skill" was perhaps not the best phrase to use--we are trying to argue that the scatter in the measurement data is likely contributing to the limited predictive power of our current mathematical fits. We note that the p-values for these fits for Δf60 and Δf44 (as well as the other variables in Fig. 3, mean Dp ΔO/ΔC) and are both less than 0.01 and we argue that our fits provide valuable information on how physical age and a metric for plume size (here, initial OA at the time of the first measurement) impact Δf60 and Δf44. We now note in the text that the p-values are <0.01 for all fits and we have updated this section to read:

"The aging values of Δf_{60} , Δf_{44} , and ΔO/ΔC show scatter (Figs. S14-18), which likely contributes to the limited predictive power of our mathematical fits. The scatter is likely due to variability in emissions due to source fuel or combustion conditions, instrument noise and responses under the large concentration ranges encountered in these smoke plumes, inhomogeneous mixing within the plume, variability in background concentrations not captured by our background correction method, inaccurate characterizations of physical age due to variable wind speed, and/or deviations from a true Lagrangian flight path. Eqs. 4-5 performed the best out of the mathematical fits that we tested. These equations do not have a direct physical interpretation but may be used as a starting point for modeling studies as well as for constructing a more physically based fit. There may be another variable not available to us in the BBOP measurements that can improve these mathematical fits, such as photolysis rates. We do not know whether these fits may well-represent fires in other regions around the world, given variability in fuels and burn conditions. We also do not know how these fits will perform under nighttime conditions, as our fits were made during daytime conditions with different chemistry than would happen at night. We encourage these fits to be tested out with further data sets and modeling. These equations are a first step towards parameterizations appropriate for regional and global modeling and need extensive testing to separate influences of oxidation versus dilution-driven evaporation.."

R1.30) L273: While the authors state here that highest initial deltaOA generally has the lowest normalized number concentrations, this seems to contradict their near zero Spearman's coefficient reported in Fig. 2. In fact, the authors state this two lines later. This needs to be revised. Either there is a correlation or there is not.

This is a good point--we have omitted this statement as it is not consistent with the observations. Instead we state:

"Although we would anticipate that plume regions with higher initial ΔOA would have lower normalized number concentrations due to coagulation, a few dense cores have normalized number concentrations comparable or higher than the thinner edges, leading to no correlation with $\Delta OA_{initial}$."

R1.31) L276: Is variability in number emissions really "noise?" It seems like an inherent feature.

We have changed "noise" to "unexplained variability" in the text.

R1.32) L278: Does the particle size really increase for "all" plumes, or does it statistically increase when considered across all plumes? There seem to be some lines in the graph that are basically flat when considered individually; thus, I am not certain that the "all" applies.

This is a good point. We have deleted the 'all' reference and have modified the text to read:

"The mean particle size between 40-262 nm, $\underline{D_p}$ (Eq. 31), is shown to statistically increase with aging when considered across the BBOP dataset…"

R1.33) L280: As mentioned above, have the authors considered other potential artifacts in their deltaOA/deltaCO that might lead to this parameter remaining flat while the apparent particle size increases? I suggest this be discussed in the context of the authors' conclusion that coagulation drives the size change.

We agree that this caveat is appropriate to discuss here. We have added the following parenthetical remark:

"(We acknowledge that ΔOA/ΔCO may be impacted by measurement artifacts as discussed in Sect. 2. For instance, if the collection efficiency of the AMS is actually decreasing with age, then ΔOA/ΔCO would be increasing and the increases in mean diameter will be due to SOA condensation as well as coagulation.)"

R1.34) L283: The authors have been assuming that it is acceptable to use as an "initial" OA and particle concentration the value measured in the closest transect for each flight. Given this assumption, it is unclear why the authors now indicate it is essentially inappropriate to estimate an initial particle diameter from the closest transect to use for comparison with the model of Sakamoto et al. (2016). If the assumption is poor for one variable how is it justified that it is okay for two other variables?

The reviewer has a good point that our logic seems inconsistent here. We add the following text when we first introduce the concept of $\Delta OA_{initial}$ in Sect. 3.1:

We note that $\Delta OA_{initial}$ does not actually represent the true initial emitted OA from each fire, but instead serves as a proxy for the general fire size, intensity, and emission rate (as presumably larger, more intensely burning fires will have larger mass fluxes than smaller, less intensely burning fires). Thus, $\Delta OA_{initial}$ and other "initial" metrics referred to in this study are not to be taken as emission values, and direct comparison to studies with direct emissions values is not appropriate, as dilution and chemistry may occur before the initial flight transect, which we discuss further below.

We also modify the specifically mentioned section:

"Sakamoto et al. (2016) provide fit equations for modeled D_p as a function of age, but they include a known initial D_p at the time of emission in their parameterization (rather than 15 minutes or greater, as available to us in this study)" (underline added to point out new text)

R1.35) Equation 2: What units must the time have?

Good call--for the fit coefficients, time is in hours. We have now included this when introducing the fit equation.

R1.36) L290: Nucleation is generally more favorable when existing particle surface area is smaller, as the condensation sink is reduced. Might this also be an explanation for the greater incidence of nucleation near plume edges?

Yes absolutely--we have added this possibility to the discussion.

"As well, nucleation is more favorable when the total condensation sink is lower (e.g. reduced particle surface area) (Dal Maso et al. 2002), which may occur for outer portions of plumes with little aerosol loading."

R1.37) L294: The authors note that the nucleation mode "appears to be coagulating or evaporating away as the plumes travel downwind." It would be useful if they show this explicitly in some way. Which figures should the reader look at specifically and which intersects? I find this overall too vague and suggest that it needs to be made more explicit.

We have examined this statement and Figs. S7-S11 (number size distribution plots) and upon further consideration do not think that it's strictly apparent what is happening to the smallest particles downwind--quite often the nucleation mode appears to be persistent even at final

transects. We have removed the statement and have moved the first half of this sentence to earlier in the paragraph (underlines to emphasis text that has been moved into this sentence):

Nucleation-mode particles appear to be approximately one order of magnitude less concentrated than the larger particles, and primarily occur in the outer portion of plumes, although one day did show nucleation-mode particles within the core of the plume (Fig. S11).

R1.38) L303: Again, does "thicker" here mean "more concentrated"? Thickness, which I would interpret to mean some spatial thickness, is not discussed in this paper as best I can tell. Regardless, the authors cannot conclude that deltaN/deltaCO is lower for "thicker" plumes since their Spearman's coefficient is essentially zero.

The reviewer is correct--following similar comments above, we delete this statement.

R1.39) L308: Again, how can the authors rule out differences in the initial conditions that are independent of physical or chemical aging? This seems to be an underlying assumption throughout this entire study, but I do not find that the authors have really justified this assumption. Given how central it is to everything, I strongly suggest that an explicit discussion must be included wherein the authors review the evidence for and against their assumption.

We have added more text and qualifiers to section 3 addressing this issue, following comments R1.24 and R2.47. We add the following text to this discussion:

"We were unable to quantify the impact on potential interfire variability in the emission values of the metrics studied here (such as variable f_{60} and f_{44}). We anticipate that being able to capture this additional source of variability may lead to stronger fits and correlation."

And

"We also suggest further refinement of our fit equations, as further variables (such as photolysis rates) and better quantification of interfire variability (such as variable emission rates) are anticipated to improve these fits."

Minor:

R1.40) L47: It might be more accurate to say that the smoke plumes dilute through entrainment of background air rather than that they dilute and entrain background air.

Thank you--this is similar to comment R2.14) and we have clarified this sentence, addressing both reviewer comments:

"Dilution through entrainment of regional background air can cause vapors and particles emitted from fires to rapidly evolve as smoke travels downwind"

Review 2

Summary:

This manuscript uses airborne data of wildfire smoke plumes, measured as pseudo-lagrangian transects of the plumes during the 2013 BBOP field campaign. Physical ages of the plumes ranged from approximately 15 minutes to 2-4 hours.

The authors analyze the oxidation state (through f44, f60, O/C, and H/C) as well as mean particle diameter and the OA/CO emission ratio of aerosol in terms of physical plume age and the aerosol's proximity to the plume core. They demonstrate enhanced chemical aging/oxidation at the edges of plumes that they argue is related to enhanced photolysis in more dilute BBOA-containing air.

Only a couple studies have discussed the effects of chemical aging in terms of plume thickness and edge-to-core position. This is a very informative and fascinating approach and is a great use of archived data BBOP data to build upon previous modeling research. The paper is well cited and the figures are generally aesthetically pleasing. Please don't be dismayed by the criticism to follow as I tend to focus on the things that need to be fixed. There are a lot of good observations and analysis in this paper which I don't, but maybe should, highlight.

I believe that many of the conclusions are likely true, however the way the data was analyzed does not always support this and I have made quite a few comments regarding this. In my opinion, a focus should be made on comparisons within transect sets regarding how things evolve with physical age and generalizations of plume cores vs plume edges instead of on bulk regressions (Spearman's correlations) which are not particularly convincing (either low R-values or R-values reflective of outlier data). Additionally, there seems to be a lot of contradicting statements made in interpreting the results. This is potentially a very good and interesting paper relevant to the subject areas of ACP and eventually should be published, but obviously will require significant edits.

General Comments:

R2.1) Figures are aesthetically pleasing but could use some minor changes.

We have followed both reviewers' specific suggestions in ensuing comments to the best of our abilities and scientific agreement.

R2.2) Format of citations need to be fixed.

We agree that a number of our in-text citations came through poorly. We apologize and have fixed these to the best of our abilities.

R2.3) There are a lot of typos and issues with word choice which will need to be fixed before final publication.

We have responded to specific comments in both reviews and have done a thorough check of our document before resubmission.

R2.4) I am curious, how wide were the plumes and how long did it take to fly through them? It seems like you explored whether instrument lags affected your results, but during a transect did the physical age of the leading the plume edge vary significantly from the edge when you left the plume?

The plumes are approximately 5-50 km wide (using the Haversine method; this can be observed in Figs. S2-S6). We now note this in the methods section:

"The plumes spanned from approximately 5-50 km wide (Figs. S2-6)."

If the flights were perfectly Lagrangian, the physical age would be the same from leading plume edge to trailing plume edge. The plane was travelling at 100 m s^{-1} on average, and thus took ~50-500 s (0.8 to 8.3 minutes) to cross, and general uncertainty in physical age is larger than this. We note this at the end of the section:

"We use the mean wind speed and this estimated centerline to calculate an estimated physical age for each transect, and this physical age is assumed to be constant across the transect, as plume crossings took between 50-500 seconds."

R2.5) I think you can better clarify how you estimate physical age. In the supplementary files, the "core" trajectory is a straight line, presumably because you use a single wind speed and direction, but the core of the transect frequently does not lie on that line. Could this be improved with Hysplit/WRF models? Would that help the core of the transect fall along the dashed line?

There were a few other comments on our physical age estimate (see R1.10, R1.13, and R2.25). We have modified the text to read:

"To estimate the physical age of the plume, we use the estimated fire center as well as the total FIMS number concentration to determine an approximate centerline of the plume as the smoke travels downwind (Figs. S1-S6). The centerline is subjectively placed to attempt to capture the

most-concentrated portion of the total number concentration for each plume pass, as we focus on aerosol properties and their relations to dilution in this study."

We agree that our approximate center-line is not perfect. However, the resolution and uncertainty of models like Hysplit or WRF are great enough that we do not have confidence that they would perform any better as the model/reanalysis meteorology may have errors.

R2.6) Data are broken down into physical age and further into fringe-vs-core (such as shown in Figure 1). These data-points represent a range of data subsample in time and space and therefore should include error bars representing the variance in data represented by each data point as well as the measurement uncertainty.

Reviewer 1 had similar comments in R1.14) for figure 1. We provide our response to them here and argue that these comments are appropriate for figure 2 as well.

We are quite hesitant to put forth a precision-based analysis. We are cautious to apply a known precision under ambient conditions to the sometimes extremely concentrated conditions of smoke plumes. For instance, our initial analyses included ozone measurements and UHSAS particle size distribution measurements, but we had to remove both instruments due to unresolvable issues with interferences under plume conditions. The UHSAS became saturated--this saturation level may be changing as both a function of particle size and concentration (as was discovered from careful analysis of a UHSAS during strong pollution events during an indoor campaign and seen again during a controlled burn study; Erin Boedicker [Colorado State University; Farmer group], personal communication). Another issue is that propagating uncertainties assumes that precision is equivalent in all of the measurements. We are using multiple instruments so this assumption breaks down, as many instruments define and calculate precision differently. This makes a true apples-to-apples comparison (which is needed for propagation of errors) tricky or impossible. As discussed in response to other comments by this and the other reviewer, we have weakened the language of our results throughout due to these uncertainties.

R2.7) Df60 and Df44 are known to vary in primary emissions, even in laboratory experiments where nascent soot can be analyzed (i.e. not after 10+ minutes of aging). However, a key assumption in many of the conclusions seems to be that all primary BBOA has the same initial Df60 and Df44. This is a problem when the authors try to support their conclusions.

Reviewer 1 had similar concerns in comment R1.24. We did not intend to make that assumption, but it is possible that a reading of our manuscript gives the impression that we implicitly are making that assumption. We do not expect the same Δf_{44} and Δf_{60} for each fire, and thus

variability from emissions likely contributes to the unexplained variability of our fit parameters. We do include two more brief discussions on this in the text in Sect. 3.1 within the discussion of our fit parameters:

"We note that each fire may emit particles with variable initial f_{44} and f_{60} values, as has been observed in laboratory studies (Hennigan et al. 2011; Cubison et al. 2011; McClure et al. 2020), which adds to scatter within the data. It is possible that variability in f44 and f60 may also contribute to the observed correlations with ΔOA_{intial}; however, this would require that higher f_{44} emissions are correlated with lower emissions rates and/or faster dilution rates (and visa versa for f_{60}). Lacking direct emissions measurements, this hypothesis cannot be further explored in this work."(this comment and R1.24)

"The scatter is likely due to variability in emissions due to source fuel or combustion conditions, instrument noise and responses under the large concentration ranges encountered in these smoke plumes, inhomogeneous mixing within the plume, variability in background concentrations not captured by our background correction method, inaccurate characterizations of physical age due to variable wind speed, and/or deviations from a true Lagrangian flight path." (this comment and R1.24)

R2.8) The use of Spearman's rank-correlation is fine as you may not expect linearly increasing/decreasing values with physical (or even chemical) age. But it needs to be clearly stated that this is a test of monotonically increasing/decreasing values, which does not give the same predictive interpretations as a Pearson's correlation. Interpretation of Spearman's correlation coefficients and the strength of these coefficients, in many cases, do not support the interpretations presented in this work. Part of this is because the authors chose to combine all data from all flights together for the regressions. This means that data representing older physical age of a plume with high initial concentrations is mixed together with data representing young physical age but low concentrations. The result is that there is not a strong relationship between these parameters (e.g. DN/DCO) and physical age (or DOAinitial). If these transects were normalized in some other way, maybe these statements may be more supportive of the Conclusions.

We now note in the text that the Spearman tests are a test for monotonicity when we first mention it in the text. We agree that mixing data in the fashion described may limit our statistical analysis. However, the fit equations and results of Figure 3 do get at the combined effects of age/concentration. Given that those fits show initial promise and that the results of Figure 1 do show some moderate trends, we argue that there is value in our methods. The reviewer asks more specific questions regarding normalization in comment R2.35, and we refer the reviewer to our response there for further details.

R2.9) The supplementary text provides very little additional information. There seems to be some confusion regarding methodology which could be explained in more detail here. I would suggest a cartoon of a flight path showing how you chose your background for a Transect.

We agree that the original SI was too sparse. We have expanded the SI section to include more information about the campaign instrumentation, following reviewer 1's comment R1.8, and we refer reviewer 2 to comment R1.8. We have included the locations of each flight's background (lowest 10% of CO) in Figures S2-S6.

R2.10) Were all supplementary sections/figures referenced in the text? I lost count.

We verified before submission that all SI figures and text were referenced in the text; we have re-verified before our current re-submission.

Specific Comments:

R2.11) L30: Be more specific about what you mean by "smoke concentrations… aging markers,number, diameter."

We have updated the text to read:

"Here, we use observational data from the BBOP field campaign and show that initial smoke organic aerosol mass concentrations can help predict changes in smoke aerosol aging markers, number concentration, and mean diameter between 40-262 nm."

R2.12) L34-35: You state that it is not quantifiable how diluted a plume is when first measured; does this contradict the next statement that (hence) the initially measured (number?) concentration is a proxy for dilution?

We agree that this text is confusing and have clarified it:

"However, the extent to which dilution has occurred prior to the first observation is not a directly measurable quantity. Hence, initial observed plume concentrations can serve as a rough indicator of the extent of dilution prior to the first measurement, which impacts photochemistry and aerosol evaporation."

R2.13) L37: Do you mean "increases in oxidative tracers" or that the oxidation-state of OA at the edges was higher?

The latter--we've clarified the text (and split the original long sentence into two):

"We further find that on the edges, the oxidation state of organic aerosol has increased and has undergone more decreases in a marker for primary biomass burning organic aerosol. "

R2.14) L44-47: "...rapidly evolve as smoke travels downwind, diluting and entraining background air." I think you mean that dilution and entrainment can rapidly cause aerosol & vapor evolution, but that is not how it reads.

Thank you--this is similar to comment R1.40) and we have clarified this sentence, addressing both reviewer comments:

"Dilution through entrainment of regional background air can cause vapors and particles emitted from fires to rapidly evolve as smoke travels downwind"

R2.15) L49: I think you mean "dilution at time of measurement".

Thank you--we have added this.

R2.16) L54: Does this refer to radiative fluxes?

Yes--we have updated this phrase to "shortwave radiative fluxes"

R2.17) L 55-57: Please fix the brackets around citations.

Fixed.

R2.18) L93: Should read "aging and oxidation of OA mass and aerosol number concentration and mean Diameter."

We agree that this sentence is hard to parse; we've updated it:

"Here, we present smoke plume observations from the Biomass Burning Observation Project (BBOP) campaign of aerosol properties from five research flights sampling wildfires downwind in seven pseudo-Lagrangian sets of transects to investigate the evolution of OA mass and oxidation state, aerosol number, and aerosol mean diameter."

R2.19) L112: 20-262 nm size range is not ideal, but I guess it is what you have.

We agree that we would have preferred a larger size range.

R2.20) L134-135: Also background correct m/z=44 and m/z=60?

Reviewer 1 was also unclear on our background-corrections and calculations for f_{60} and f_{44} (comment R1.9). We repeat our response here:

We calculated the background corrected f60 and f44 as follows (where $f = f_{60}$ or f_{44}):

$$\Delta f = \frac{(f_{in} * OA_{in}) - (f_{out} * OA_{out})}{\Delta OA} \qquad \text{Eq. R1}$$

Similar, the $\Delta O/\Delta C$ and $\Delta H/\Delta C$ are calculated through (where X = O or H):

$$\frac{\Delta X}{\Delta C} = \frac{(X_{in\ plume} - X_{out\ of\ plume})}{(C_{in\ plume} - C_{out\ of\ plume})} \qquad \text{Eq. R2}$$

We've added Eqs. R1-R2 as Eqs. 1 and 2 in the main text and have updated other equation numbers and references.

R2.21) L 136: Conceptually, where does the lowest 10% of CO occur? Just outside of the plume as the plane circles back through? Is the background fairly constant for a flight leg? Do you adjust background each time the plane turns around and goes back to transect the plume again?

Figures S7-S11 (white solid line in each figure) indicate that the CO outside of the plume is fairly constant. We do not adjust the background each time but instead use the lowest 10% for the entire flight path once the plane has reached the fire until the plane leaves the fire/smoke complex. The location of the lowest 10% varies from flight to flight and from leg to leg, but often occurs on the flight portion furthest from the smoke plume of each leg. As was noted in the text, we did sensitivity analyses of our results to our assumptions about background and in-plume CO values and our conclusions were not changed.

R2.22) L137: Is elemental O, H, and C calculated from O/C, H/C & OA or is H/C and O/C calculated from the elemental O, H, C concentrations? Aiken et al (2007) estimate it in the later (Eqn 1).

We calculate elemental O, H, and C using O/C, H/C, and OA , assuming that all of the OA mass was from O, C, and H. We have added the following: "Elemental O, H, and C are calculated

using the O/C and H/C and OA data from the SP-AMS (assuming all of the OA mass is from O, C, and H),..." (underline ours)

R2.23) L 139: Typo ("…, we but do not…")

Fixed

R2.24) L164-165: Sentence grammar

Updated to:

"The true fire location and center at the time of sampling is likely different than the MODIS estimates, depending on the speed of the fire front."

R2.25) L165-167: Why use the FIMS # distribution to determine plume center? Why not [CO], [mrBC], total number concentration, etc? In the supplemental figures, it says the center-flow is determined by number concentration (not distribution).

We have made an error here--we do use the total FIMS number concentration to determine our plume center and have updated the text to reflect that. We use aerosol number as this study is focused on aerosol properties as a function of dilution amount. We have updated the text here (also following points made in R1.13):

"To estimate the physical age of the plume, we use the estimated fire center as well as the total FIMS number concentration to determine an approximate centerline of the plume as the smoke travels downwind (Figs. S1-S6). The centerline is subjectively placed to attempt to capture the most-concentrated portion of the total number concentration for each plume pass, as we focus on aerosol properties and their relations to dilution in this study. We use mean wind speed and this estimated centerline to get an estimated physical age for each transect. We did not propagate uncertainty in fire location, wind speed, or centerline through to the physical age, which is a limitation of this study."

R2.26) L170: Fix heading

Fixed

R2.27) L189: Measurement uncertainty should be plotted in Figures (sum of variance in data represented by each data point + uncertainty in each instrumental recording)

This comment is similar to comments R2.8 and R1.14, and we refer the reviewer to our responses there.

R2.28) L189-190: Changes in f60 and f44 should be provided as fractional (as displayed on axis of Figure 1, etc). Relative changes (%) are confusing.

Reviewer 1 pointed out in comment R1.16 that much of the discussion in this paragraph (lines 185-194 of the original document) was not well-posed. We have deleted this discussion and replaced it with:

"Figure 1 shows that ΔOA/ΔCO and ΔBC/ΔCO vary little with age for both the 5-15 and 90-100 percentile of ΔCO (p-values>0.5). A true Lagrangian flight with the aircraft sampling the same portion of the plume and no measurement artifacts (e.g. coincidence errors at high concentrations) would have a constant ΔBC/ΔCO for each transect. This flight and other flights studied here have slight variations in ΔBC/ΔCO (Fig. 1; Figs. S14-S18), which may be indicative of deviations from a Lagrangian flight path with temporal variations in emission and/or measurement uncertainties. The remaining variables plotted also show some noise and few clear trends, but it is apparent that the 5-15 and 90-100 percentiles do show a separation for many of the individual metrics. In order to determine the existence or lack trends for these metrics, we spend the remainder of this study examining each metric from all of the pseudo-Lagrangian flights together."

R2.29) L192: Replace "number concentration" with either "normalized number concentration" or "DN40-262 nm /DCO".

Thank you, we have changed this to read as "normalized number concentration"

R2.30) L192: I only see a decrease in DN40-262 nm /DCO between ~0.6 and 1.0 hours physical age. Saying that it decreases with age implies a consistent trend. For Dp, this trend is hard to tell if it is statistically significant.

Reviewer 1 had similar issues with this paragraph in comment R1.16 (see also comment R2.28 above) and we have modified the discussion entirely:

"Figure 1 shows that ΔOA/ΔCO and ΔBC/ΔCO vary little with age for both the 5-15 and 90-100 percentile of ΔCO (p-values>0.5). A true Lagrangian flight with the aircraft sampling the same portion of the plume and no measurement artifacts (e.g. coincidence errors at high concentrations) would have a constant ΔBC/ΔCO for each transect. This flight and other flights studied here have slight variations in ΔBC/ΔCO (Fig. 1; Figs. S14-S18), which may be indicative

of deviations from a Lagrangian flight path with temporal variations in emission and/or measurement uncertainties. The remaining variables plotted also show some noise and few clear trends, but it is apparent that the 5-15 and 90-100 percentiles do show a separation for many of the individual metrics. In order to determine the existence or lack trends for these metrics, we spend the remainder of this study examining each metric from all of the pseudo-Lagrangian flights together."

R2.31) L197: What do you mean by "available …"?

By available, we mean when instruments were taking measurements--we have gaps in the measurement data. We have added the following parenthetical statement:

"(Some transects do not have data available for specific instruments.)"

R2.32) L197-199: Really long sentence. I have had to read it 6-7 times to parse out what is shown.

We have updated this to:

"Fig. 2a-e show available $\Delta OA/\Delta CO$, Δf_{60}, Δf_{44} $\Delta H/\Delta C$, and $\Delta O/\Delta C$ edge and core data versus physical age for each transect for each flight of this study. We color each line by the mean ΔOA within a ΔCO percentile bin from the transect closest to the fire, $\Delta OA_{initial}$."

R2.33) L200-201: Physical age is the distance between the transect-center to the fire-center divided by the average windspeed? So does 0 physical age imply infinite or 0 windspeed?

It would imply that the measurement is directly over the fire center (fire center - transect center = 0), we've clarified this in the text:

"We note that although some of the physical ages appear to be at ~0 hours (e.g. over the fire)..."

R2.34) L203: The "…correlation coefficients (R) with initial plume OA mass,…" is not shown. Do you mean to say that this is represented by DOAinitial?

Reviewer 1 had similar concerns in comments R1.17 and R.20. We copy our discussion here:

We see that our original text here is confusing and misleading. We have attempted to clarify it. We are using a single value for $\Delta OA_{initial}$ for each transect within a Lagrangian set of transects which is obtained from the first transect of the set. If a flight has two Lagrangian sets of transects, there will be a different value of $\Delta OA_{initial}$ used for the two sets of transects, each again

obtained from the first transect of each set. The original text may have been interpreted that we used $OA_{initial}$ but we did not--we have clarified that. We use the changing values of $\Delta OA/\Delta CO$, Δf_{60}, Δf_{44} $\Delta H/\Delta C$, and $\Delta O/\Delta C$ as they age downwind to compare with initial OA. We have updated this text (also following suggestions made in R1.20):

"Also included in Fig. 2 are the Spearman rank-order correlation tests (hereafter Spearman tests), which are tests for monotonicity. The Spearman tests show correlation coefficients for each flight set (Table S1) with the initial ΔOA of a flight set ($\Delta OA_{initial}$) against $\Delta OA/\Delta CO$, Δf_{60}, Δf_{44} $\Delta H/\Delta C$, and $\Delta O/\Delta C$ as each variable ages downwind. We also include Spearman tests for the calculated physical age of the smoke for each flight set against these same variables. The R values are labeled $R_{\Delta OA,initial}$ and R_{age}, respectively, in Fig. 2. For the correlations with $\Delta OA_{initial}$, all transects in a given Lagrangian set of transects have the same $\Delta OA_{initial}$ value; for flights with two Lagrangian set of transects, each set has its own $\Delta OA_{initial}$ value. Correlating to $\Delta OA_{initial}$ provides an estimate of how the plume aerosol concentrations at the time of the initial transect impact plume aging (aging both before and after this initial transect)."

R2.35) L202-204: Is the Spearman coefficient for concatenation of all data points from all transects? If so, I am not sure it would make sense to do this way. Spearman's test tests for monotonically increasing/decreasing values. Given that each transect set starts at a different initial value you wouldn't expect the grouped transect sets to display a strong R-value. If you want to use Spearman's test in this way, for Rage you could normalize each normalized value to the initial normalized value to get a % change and plot that in Figure 2 and relevant supplementary figures.

We do agree that variability in emissions will lead to a different initial value of $\Delta OA_{initial}$. However, changes to the smoke aerosol (coagulation, dilution, evaporation, chemistry, etc.) should be occurring before the time of the first measurement, and using $\Delta OA_{initial}$ helps show that. If the changes in the factors in Figure 2 between the time of emission and the first transect are affected by the plume density, this would lead to an increase in the Spearman $R_{\Delta OA,\ initial}$. Of course, we are still impacted by variability in emissions within our current methods, and we have added further disclaimers throughout the text following reviewer comments. As the reviewer mentions, this scatter at the time of the first transect does reduce the Spearman R_{age}, but because plume-density-dependent aging prior to the first transect is one of the potentially interesting findings of this study, we feel that it is important to not normalize our changes. We have added the following text to Sect. 3.1:

"We note that scatter in $\Delta OA_{initial}$ leads to weaker R_{age} values than would be obtained if we normalized changes with aging to the first (normalized) value. However, as plume-density-dependent aging prior to the first transect is one of the potentially interesting findings of this study, we feel that it is important to not normalize our changes further."

R2.36) L206: Spell out “Figs.” And lower case.

Fixed

R2.37) L207-208: Type in list “…FIMS measurements AND BACKGROUND and DCO percentile Spacings…”

We have updated this section. We also have changed “background CO” to “in-plume CO threshold value”, as the latter is accurate and background CO is misleading.

“Figs. S13, S19-S21 show the same details as Fig. 2 but provide sensitivity tests to potential FIMS measurement artifacts (Fig. S13) and our assumed in-plume CO threshold value (set to 150 ppbv for Figs. 1-3; Sect. 2) and ΔCO percentile spacing (Figs. S19-S21). Although these figures show slight variability, the findings discussed below remain robust and we constrain the rest of our discussion to the assumptions made for the FIMS measurements, in-plume CO threshold value, and ΔCO percentiles used in Fig. 2.”

R2.38) L209: Previous line said you would only discuss FIMS, background and DCO.

We see that this sentence is confusing, we intend that our assumptions used in Fig. 2 about the FIMS measurements, CO, and delta(CO) percentiles will be used throughout the rest of the study. We have clarified the text:

“Although these figures show slight variability, the findings discussed below remain robust and we constrain the rest of our discussion to the original assumptions made for the FIMS measurements, in-plume CO threshold value, and ΔCO percentiles used in Fig. 2.”

R2.39) L209-210: RDOA,initial just says 0 in figure.

Thank you for catching this, the R value is 0 here and we have updated the text:

“In general, both the cores and edges show little change in ΔOA/ΔCO with physical aging, with $R_{\Delta OA,initial}$ and R_{age} at 0 .02 and 0.03… “

R2.40) L209-210: This figure shows orders of magnitude changes in DOA/DCO with age. I think you mean there is not a clear positive or negative trend (as stated in the first clause of the next sentence), not that there is no change.
We have updated the text from “show little change” to “do not show any positive or negative trend”.

R2.41) L212: Here and elsewhere, spell out "vs." Check grammar.

We have fixed the vs. errors and have done a thorough grammar check. We have made many small changes to improve readability and grammar.

R2.42) L213: For positive R values, consider putting a "+" sign in front of the value.

This does improve clarity and we have updated the positive R values to have a + sign throughout the manuscript.

R2.43) L214-218: Consider breaking this into multiple, shorter sentences. Check for redundancy with L212-214, i.e. a negative R value means there is a decreasing trend.

We have updated this section (including suggestions following reviewer 1's comment R1.22). We removed the sentence in L212-214, as it is redundant, and incorporated the R values into the updated text:

"Δf_{60} generally decreases with plume age (R_{age} = -0.26), consistent with the hypotheses that Δf_{60} may be evaporating because of dilution, undergoing heterogeneous oxidation, and/or having a decreasing fractional contribution due to condensation of other compounds.. In contrast, Δf_{44} generally increases with age (R_{age} = +0.5) for all plumes with available data. It appears for the plumes in this study that although there is little change in $\Delta OA/\Delta CO$, loss of compounds that contain f_{60} fragments (as captured by the SP-AMS) is roughly balanced by condensation of more-oxidized compounds, including those that contain compounds with f_{44} fragments, such as carboxylic acids. This observation suggests the possibility of heterogeneous or particle-phase oxidation that would alter the balance of Δf_{60} and Δf_{44}."

R2.44) L214-218: Is it only evaporation or condensation (phase changes) happening or does O attack volatile and semivolatile species (levoglucosan) changing its molecular composition to more oxidized/refractory species without a phase change?

Reviewer 1 made similar comments in R1.23. We answer this comment and the next comment (R2.45) as well as R1.23:

We are trying to note that heterogeneous chemistry is relatively slow (for near-field aging) and shouldn't significantly contribute to compositional changes. We have added text to emphasize that point more clearly:

"However, estimates of heterogeneous mass losses indicate that after three hours of aging for a range of OH concentrations and reactive uptake coefficients, over 90% of aerosol mass is anticipated to remain, indicating that heterogeneous loss has limited effect on aerosol composition or mass (Fig. S23; see SI text S2 for more details on the calculation). Hence, the evaporation of f_{60} being balanced by gas-phase production of f_{44} may be the more likely pathway"

R2.45) L218-220: If you didn't expect a change in normalized-OA anyway based on your model, why do you suggest a balance between evaporation particle mass loss and condensation mass gain?

The evaporative loss may be driven by dilution and the condensation may be driven by production of lower-volatility species from oxidation of either evaporated POA or more-volatile SOA precursors.

R2.46) L221: Those are not very strong R values to base your interpretations on, but I wouldn't expect them to be for the reasons discussed above. This statement is not particularly true for f60.

We have unified our language when discussing R and R^2 values throughout the text, following reviewer comment R1.20 as well as this comment.

R2.47) L224: But you just said that Df60 and Df44 correlate with DOAinitial. Differences in your initial Df60 or Df44 don't necessary need a mechanistic explanation. We see variance these parameters in fresh emission in laboratory experiments and would expect to also see variance in primary emissions of wildfires. This is not good support for your next conclusion (that aircraft observations are missing evaporation and/or condensation).

Reviewer 1 had similar concerns in comments R1.24 and 1.39. Our response to both R1.24 and R2.47 is:

Both reviewers makes reasonable points here and we agree that these are alternative hypotheses that should be explicitly discussed in the manuscript. Reviewer 2 made similar comments in R2.47. Unfortunately, lacking direct measurements of the emissions, we cannot explore this hypothesis in any detail. And we do find it compelling that less-dense plumes do show higher f44/lower f60 than more-dense plumes, which supports our hypothesis of aging prior to the transect. We have added the following text to Sect. 3.1:

"We note that each fire may emit particles with variable initial f_{44} and f_{60} values, as has been observed in laboratory studies (Hennigan et al. 2011; Cubison et al. 2011; McClure et al. 2020), which adds to scatter within the data. It is possible that variability in f44 and f60 may also contribute to the observed correlations with ΔOA_{intial}; however, this would require that higher f_{44}

emissions are correlated with lower emissions rates and/or faster dilution rates (and visa versa for f_{60}). Lacking direct emissions measurements, this hypothesis cannot be further explored in this work."

To Reviewer 1's last query ("Also, given that different sources produce particles that have different initial f60 and f44, would they be expected to exhibit the same deltaf60 and deltaf44 even if initial OA and dilution were identical? Is there evidence that this is expected?"), we would not expect the same Δf_{44} and Δf_{60} under those circumstances and thus variability from emissions likely contributes to the noise of our fit parameters. We do include a brief discussion on this in the text in Sect. 3.1 within the discussion of our fit parameters (with new minor edits addressing comments from reviewer 2):

"The scatter is likely due to variability in emissions due to source fuel or combustion conditions, instrument noise and responses under the large concentration ranges encountered in these smoke plumes, inhomogeneous mixing within the plume, variability in background concentrations not captured by our background correction method, inaccurate characterizations of physical age due to variable wind speed, and/or deviations from a true Lagrangian flight path."

We also address this issue in the conclusions. We have added more text and qualifiers to section 3 addressing this issue, following comments R1.24 and R2.47. We add the following text to this discussion:

"We were unable to quantify the impact on potential interfire variability in the emission values of the metrics studied here (such as variable f_{60} and f_{44}). We anticipate that being able to capture this additional source of variability may lead to stronger fits and correlation."

And

"We also suggest further refinement of our fit equations, as further variables (such as photolysis rates) and better quantification of interfire variability (such as variable emission rates) are anticipated to improve these fits."

R2.48) L227: Is this logic circular? That differences in DOAinitial is due to different emission fluxes?

Differences in $\Delta OA_{initial}$ (which is the ΔOA of the first flight transect, not the ΔOA directly emitted from the fire) can stem from a variety of reasons beyond emission fluxes. We include some further reasons in our original text, copied here: "*The differences in $\Delta OA_{initial}$ between plumes may be due to different emissions fluxes (e.g., due to different fuels or combustion phases), or plume widths, where larger/thicker plumes dilute more slowly than smaller/thinner*

plumes; these larger plumes have been predicted to have less evaporation and may undergo relatively less photooxidation (Bian et al., 2017; Hodshire et al., 2019a, 2019b)."

R2.49) L228: should not be a comma after the bracket.

Fixed

R2.50) L231 & 234: Reference format needs to be changed.

Fixed

R2.51) L234: Grammar. Reference to figure in Garofalo should be something like "(Fig. 6 in Garofalo et al, 2019)"

Fixed

R2.52) L235-236: Isn't that why you normalize?

The lack of trends from physical edge to core is most likely due to inhomogeneous mixing (which will not be improved by subtracting background concentrations), which is our next sentence, repeated here for reference:

This could be as CO concentrations (and thus presumably other species) do not evenly increase from the edge to the core for many of the plume transects studied (Figs. S2-S6).

We have added "... the remaining plumes do not show a clear trend from the physical edges to cores" (underline ours) to this statement to emphasize that we are discussing the physical transect, rather than the divisions made by ΔCO percentile bins.

R2.53) L237-239: You imply that patterns of f60 and f44 compared to shortwave irradiance is related by photolysis rates. I don't necessarily agree with this interpretation. If the plume is thicker it means that a higher fraction of aerosol mass is from the fire and because fire-emitted aerosol has higher f60 and lower f44 than background a simple mechanism of mixing explains your observations.

Our f60 and f44 values are background corrected (please see section 2 and newly added equation 2; comment R1.9), which should correct for mixing. We are also not trying to draw any firm conclusions here, but are pointing out observational similarities (underline added for quick reference): *We do not have UV measurements that allow us to calculate photolysis rates but the in-plume shortwave measurements in the visible show a dimming in the fresh cores that has a*

similar pattern to f_{44} and the inverse of f_{60} (Fig. S26; the rapid oscillations in this figure could be indicative of sporadic cloud cover above the plumes).

R2.54) L242-243: DO/DC and f44 are both proxies for OOA and would be expected to have the same trends. DH/DC and f60, while not conceptually the same, both reflect primary BBOA and would also be expected to show the same trends It is a little redundant to analyze both sets.

We have reviewed a significant amount of biomass burning (BB) literature and have noted that many studies examine f44/f60 or O:C/H:C or both. Furthermore, while f44/f60 are popular within AMS BB measurement studies, models currently can only predict O:C/H:C. We chose to include both for completeness and ease of comparisons to other datasets in future studies. We agree that it's unsurprising to see similarities between the DO/DC and Df44 and DH/DC and Df60 results, given their relations, particularly for DO/DC and Df44. We have added the following text within this paragraph:

"Given that Δf_{44} and $\Delta O/\Delta C$ are both metrics for OA aging (Sect. 2), it is unsurprising that we see similar trends between them."

R2.55) L242-243: See issues raised earlier regarding interpreting Spearman's test results for these data sets.

We refer the reviewer to our responses on comments R2.8 and R.35.

R2.56) L249-264: You should provide explanation for why you used these equations to try and fit f44and f60. Is there a conceptual justification for them? Do they have meaning outside of a mathematical fit?

We tried a large number of mathematical fits and these equations (Eqs. 2-3 in the original text; Eqs. 4-5 in the updated text) performed the best. They do not have a direct physical meaning, and we have added the following to the end of this discussion:

"Eqs. 4-5 performed the best out of the mathematical fits that we tested. They do not have a direct physical interpretation but may be used as a starting point for modeling studies as well as for constructing a more physically-based fit."

R2.57) L263-268: What do you mean by "Aged Df60 and Df44"? Does "limiting the predictive skill"mean that your fits are not particularly informative?

We were referring to the aging values of $\Delta f60$ and $\Delta f44$, we were not careful in our language here though. "Limiting the predictive skill" was perhaps not the best phrase to use--we are trying

to argue that the scatter in the measurement data is likely contributing to the limited predictive power of our current mathematical fits. We note that the p-values for these fits for Δf60 and Δf44 (as well as the other variables in Fig. 3, mean Dp ΔO/ΔC) and are both less than 0.01 and we argue that our fits provide valuable information on how physical age and a metric for plume size (here, initial OA at the time of the first measurement) impact Δf60 and Δf44. We now note in the text that the p-values are <0.01 for all fits and we have updated this section to read:

"The aging values of Δf_{60}, Δf_{44}, and ΔO/ΔC show scatter (Figs. S14-18), which likely contributes to the limited predictive power of our mathematical fits. The scatter is likely due to variability in emissions due to source fuel or combustion conditions, instrument noise and responses under the large concentration ranges encountered in these smoke plumes, inhomogeneous mixing within the plume, variability in background concentrations not captured by our background correction method, inaccurate characterizations of physical age due to variable wind speed, and/or deviations from a true Lagrangian flight path. Eqs. 4-5 performed the best out of the mathematical fits that we tested. These equations do not have a direct physical interpretation but may be used as a starting point for modeling studies as well as for constructing a more physically based fit. There may be another variable not available to us in the BBOP measurements that can improve these mathematical fits, such as photolysis rates. We do not know whether these fits may well-represent fires in other regions around the world, given variability in fuels and burn conditions. We also do not know how these fits will perform under nighttime conditions, as our fits were made during daytime conditions with different chemistry than would happen at night. We encourage these fits to be tested out with further data sets and modeling. These equations are a first step towards parameterizations appropriate for regional and global modeling and need extensive testing to separate influences of oxidation versus dilution-driven evaporation."

R2.58) L264-265: typos/grammar

Fixed

R2.59) L271-272: The decrease in normalized number concentration with physical age mostly appears to be caused by 2-3 outlier measurements (the initial points for leg 730b edge, the initial value of another edge, and the tailing value of leg 726a 1). This does not seem like a statistically robust claim and I think the R value verifies it. Lines 275-277 seem to agree with my Assessment.

We agree with this assessment--reviewer 1 has asked us to be more precise in our language for reviewer comments (please see R1.20) and we have noted that this is a weak correlation within these sentences. We also note that reviewer 1 asked for a test in which we leave one flight out,

sequentially, to see how each R value changes (comment R1.21). We have done this and include language in the text as well as Table S2, summarizing the results.

R2.60) L273-274: "generally have lower normalized … by the time of the first measurement". This implies that there was a measurement made before the first measurement. Please explain.

We are merely trying to comment on our observations from the data here.We do not think that our text is implying that there's a measurement before the first measurement--perhaps this is made more clear by changing the phrase "by the time" to "**at** the time", and we have changed our text thusly.

R2.61) L273-274: "plume edges and cores with the highest DOA generally have lower normalized number concentrations…" This is not true based on figure 2f. The two lowest DOAinitial values (white dashed lines) have two of the highest DN/DCO values.

We respectfully point out that our quoted text here is discussing "highest ΔOA and low ΔN/ΔCO" whereas the reviewer is pointing out "lowest ΔOA and highest ΔN/ΔCO"--the two arguments are consistent with each other.

R2.62) L279: Evaporation (mass loss/time) is, partially, a function of available surface area. Since small particles have a higher surface area-to-volume, it is plausible that evaporation will decrease the number of small particles more than large particles and therefore increase the mean particle size. You state this possibility of preferential loss of small particles on lines 293-295.

This is a reasonable point--if evaporation is gas-phase mass-transfer limited, evaporation will decrease the size of smaller particles more than larger particles. However, this case would only lead to an increase in the mean diameter if a significant number of small particles shrunk to below 40 nm, removing them from the calculation of the mean Dp. And if evaporation is in quasi-equilibrium, evaporation is independent of surface area. However--the organic mass of the plume does not change significantly, so we do not have evidence to support this hypothesis for the increase in mean Dp. We have added the following text to this discussion:

"OA evaporation will decrease D_p if the particles are in quasi-equilibrium (where evaporation is independent of surface area) (Hodshire et al. 2019b). However, if evaporation is kinetically limited, smaller particles will preferentially evaporate more rapidly than larger particles, which may lead to an increase in D_p if the smallest particles evaporate to below 40 nm(Hodshire et al. 2019b). The plumes do not show significant changes in ΔOA/ΔCO (Fig. 2a), indicating that coagulation is likely responsible for the majority of increases in D_p."

R2.63) L282-283: should be R_P^2 instead of R^2_P.

We've fixed this formatting here and elsewhere in the text.

R2.64) L282-283: you were previously using R and not R2 (L272, Fig 2, etc). In my opinion, this is fine and depends on how you use them, but I have been reviewed differently. Did you intend to calculate R and R2? Please check to make sure that you they are used and calculated correctly. I only state this because there are a number of typos in the manuscript and want to make sure that this is not one.

We did indeed intend to calculate R^2 here. Calculating R previously was useful to indicate the sign of the correlation whereas here with R^2 we intend to show what fraction of the variability is captured, since all fits are positively correlated. We have added the following text:

We show R^2 here to indicate the fraction of variability captured by these fits, whereas calculating R for the trends in Fig. 2 indicate the direction of the correlation.

R2.65) L287: Do you mean "legs" instead of "days"?

We have updated this text to "Lagrangian set of transects" to match the language of our other text.

R2.66) L294: Replace "~" with "approximately"

We have updated this instance of '~' and all others in the text for consistency.

R2.67) L301-302: As mentioned above, I do not agree that the data supports the statement regarding correlation. I think there is a lot of good analysis in this paper and I don't think you need to make this statement.

We update this text to be more subjective and consistent with our terminology added in response to R1.20:

"We find that although $\Delta OA/\Delta CO$ does not correlate with $\Delta OA_{initial}$ or physical age, Δf_{60} (a marker for evaporation) is moderately correlated with $\Delta OA_{initial}$ (Spearman rank-order correlation tests correlation coefficient, $R_{\Delta OA,initial}$, of +0.43) and weakly correlated with physical age (Spearman rank-order correlation tests correlation coefficient, R_{age}, of -0.26). Δf_{44} and $\Delta O/\Delta C$ (markers for photochemical aging) increases with physical aging (moderate correlations of R_{age} of +0.5 and +0.56, respectively) and are inversely related to $\Delta OA_{initial}$ (moderate correlations of $R_{\Delta OA,initial}$ of -0.55 and -0.45, respectively)."

R2.68) L302-304: I also do not agree that the data supports the statement regarding DN/DCO.

We have removed the latter half of this sentence, which is consistent with edits made previously in the manuscript.

R2.69) L304: You don't need to keep specifying that diameter size range of 40-262.

We removed this mention of the size range.

R2.70) L306-308: I don't like saying this, I don't agree that your data support this statement. The only way that differences in Df44initial, Df60initial and DO/Cinitialsupport this statement is if all primary OA from all wildfires have the same value which has been shown to not be true.

We respectfully disagree here--variability in the emitted oxidation markers from fire to fire is most likely random, and yet we see correlations despite the random variability. The only way this comment would be true is if the emitted oxidant markers are correlated with OA emission rates, fire size, and/or dilution rates prior to the first transect--there is currently no evidence for this. We choose to keep this statement as is. We note that in Sect 3.1 we have the following statement (and have added additional text to further emphasize these points, underlined here to clearly show what's been added):

Differences in Δf_{60} and Δf_{44} for the nearest-to-source measurements indicate that evaporation and/or chemistry likely occurred before the time of these first measurements (assuming that emitted Δf_{60} and Δf_{44} do not correlate with $\Delta OA_{initial}$; there is currently no evidence for this alternative hypothesis).

R2.71) Figure 1: Change "BC" to "rBC" in the legend and axis. Also in Figures S14-S18

We have changed all mentions of 'rBC' in the text to 'BC' to be consistent with our figure notation and note in the text when Fig. 1 is introduced that BC is for the refractory BC from the SP2.

R2.72) Figure 1: Change DN/DCO to DN40-262 nm /DCO to be consistent with text.

We have noticed our inconsistency of $\Delta N/\Delta CO$ vs. $\Delta N_{40\text{-}262\ nm}/\Delta CO$ throughout our figures. We had originally divided our analysis into $\Delta N_{40\text{-}262\ nm}/\Delta CO$ vs $\Delta N_{<40nm}/\Delta CO$ but did not include the $\Delta N_{<40nm}/\Delta CO$ analysis in the final paper. We apologize for these inconsistencies and have

changed all instances in the text and figures to simply ΔN/ΔCO. We have done the same for D_p vs $D_{p,40-262\,nm}$ (updating all mentions of the latter to the former).

R2.73) Figure 2: Caption should be "function of physical age"

Good catch, thank you. Fixed

R2.74) Figure 2: This figure is pretty confusing. If I look at Figure S2, I see that for leg 726a there were 2 sets of transects with each comprising of 4 transects. So, theoretically, the same air mass was sampled 4 times corresponding to 4 different physical ages. So a line in figure contains ~4 data points which correspond with either the edge or core of a transect in the transect set? Am I reading this correct?

For the flights that have 2 Lagrangian sets of transects or days with 2 separate flights ('726a', '730a', and '730b'), Figure 2 will contain one line for each Lagrangian set of transects downwind. The physical age is assumed to be constant across a given flight transect (see comment R2.4 for further discussion on this), as mentioned in the manuscript with minor edits for clarity,

"We use the mean wind speed and this estimated centerline to calculate an estimated physical age for each transect, and this physical age is assumed to be constant across the transect, as plume crossing took between 50-500 seconds".

We include the following text to clarify the reviewer's other comments on Figure 2 here:

"Flights with two sets of pseudo-Lagrangian transects ('726a' and '730b') have two separate lines in Fig. 2, one for each set."

R2.75) How does the white dashed line in 2a go backwards in physical age?

The white line in 2a is for flight '809a'. Figure S5 (S5 of the original submission) shows that 2 legs essentially overlap. We have added subpanels to Figs. S2-S6 that indicates the time-of-flight for each flight. However, the leg slightly further from the fire occurred first in the flight so it has a calculated age slightly older than the next leg, as the calculation depends in part on distance from the fire. This is a limitation of our method. We have added the following text to the first paragraph of Sect. 3.1:

"As well, two legs for flight '809a' nearly overlap (Fig. S5), with the leg that is further from the fire occurring first in the flight path, leading to an apparent slight decrease in physical age for the sequential leg (see e.g. the white dashed line in Fig. 2a)."

R2.76) Figure 2: Change to RDOA,initial instead of double subscript to be consistent with that used in text.

Thank you for catching this--we have updated these labels.

R2.77) Figure S1: I don't see a black star or dashed line.

We have added these to the figure, thank you. We include the new version of Fig. S1 after comment R2.80.

R2.78) Figure S1: Leg number not indicated. ("The numbers are the leg number")

We have removed this reference . Figures S2-S6 now include the leg numbers, and this is reflected in these figure captions.

R2.79) Figure S1: I would suggest that you use a different symbol and symbol color for the MODIS thermal anomalies so that it contrasts with the color code of the # concentration.

We have changed our color palette for the number concentration to 'plasma', which hopefully provides enough contrast.

R2.80) Figure S1: Please change the colorcode to a color-blind friendly one.

We have changed our color palette for the number concentration to 'plasma'. We include the updated figure and caption below, as reference.

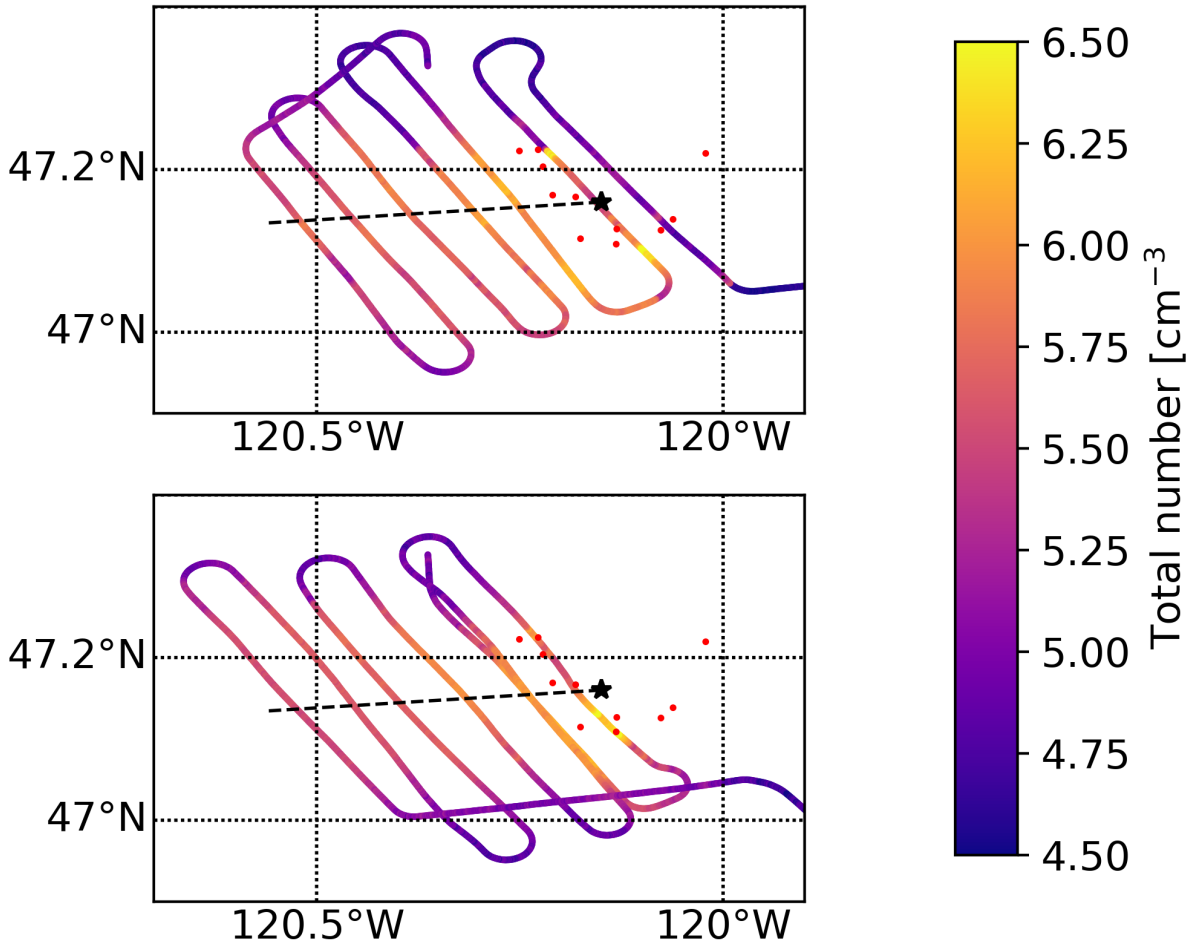


Figure S1. The flight path for flight '730b', colored by the FIMS total number concentration. The red dots are MODIS fire/thermal anomalies. The black star indicates the approximate center of the fire and the black dashed line indicates the approximate centerline of the plume, estimated by the number concentration.

R2.81) Figure S5: Is the black star the fire center for 8/9/2013 or 8/8/2013? The caption does not say what symbol is used for 8/8/2013, only that "The black star indicates the approximate center of the fire…"

We do not show the fire location on 8/8/2013 or 8/10/2013; we instead are estimating the fire center on 8/9/2013 (black star) using MODIS images from 8/8/2013 and 8/10/2013. We have added in a green star to this figure to indicate the approximate fire center on 8/8/2013.

R2.82) Figure S24-S25: The y-axis scale changes between graphs, with a wide range for data that do not look like they have much variation (leg 730a) and a smaller range for others (730b). Is this why there is not consistent patterns in 730a and 730b?

We have tightened the y axes on the subpanels that had too much whitespace. We thank the reviewer for pointing this out.

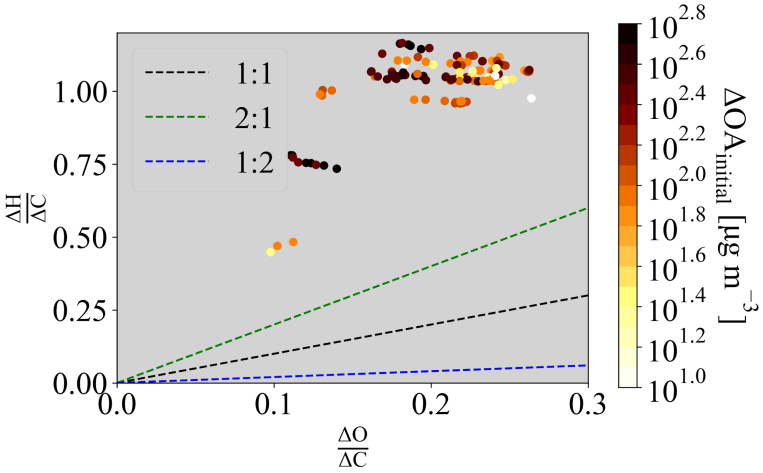
R2.83) Figure S26: is shortwave irradiance a measure of photo-chemical rate, the amount of scattering/absorbing aerosol above you, or a combination of both?

In this study, we're using the total shortwave irradiance as measured by an SPN1 (Long et al., 2010). The shortwave irradiance is a function of solar angle and scattering/absorption prior to the measurement. While it is not a measure of the UV wavelengths that drive photochemistry, we are using it as a rough proxy for these wavelengths so that we can look at how photolysis rates may vary across the flight path. We note this in our original text: "*We do not have UV measurements that allow us to calculate photolysis rates but the in-plume SPN1 shortwave measurements in the visible show a dimming in the fresh cores that has a similar pattern to f_{44} and the inverse of f_{60} (Fig. S26; the rapid oscillations in this figure could be indicative of sporadic cloud cover above the plumes).*"

Long, C. N., A. Bucholtz, H. Jonsson, B. Schmid, A. Vogelmann, and J. Wood (2010): A Method of Correcting for Tilt from Horizontal in Downwelling SW Measurements on Moving Platforms, TOASJ, 4, pp.78-87, doi: 10.2174/1874282301004010078

R2.84) Figure S27: Please complete the drawing of the Van Krevelen diagram with the 1:1, 2:1, and 0.5:1 lines

Literal 1:1, 2:1, and 0.5:1 lines are rather uninformative, as can be seen in the below figure.



We think that the reviewer may have intended constant lines of oxidation, as shown in Figure 1 of Heald et al. (2010) (their red and blue lines). Heald et al. (2010) chose a starting point of H/C=2 at O/C=0 (which is the case for long alkanes), which upon visual inspection is not an appropriate starting point for our data. We do not know exactly what the appropriate H/C and O/C starting point for primary biomass burning OA is, given variability in the emissions during BBOP and literature values. We do not add these lines of oxidation for this reason. We note that reviewer 1 had confusion with this figure, and we refer reviewer 2 to R1.7 and R1.25 for further details.

Heald, C. L., Kroll, J. H., Jimenez, J. L., Docherty, K. S., Decarlo, P. F., Aiken, A. C., Chen, Q., Martin, S. T., Farmer, D. K. and Artaxo, P.: A simplified description of the evolution of organic aerosol composition in the atmosphere, Geophys. Res. Lett., 37(8), doi:10.1029/2010GL042737, 2010.