Review 2 of Wiggins et al by Bob Yokelson

The authors monitored three stable trace gases (CO_2, CO, and CH_4) that were emitted by fires located upwind of a tower in Alaska. They derived emission ratios and emission factors for two of the gases (not sure why $EFCO_2$ was not reported?). The study sampled smoke, when present, 24/7 for a whole fire season so it has a big effective sample size compared to individual past studies. It was also sensitive to examples of much, if not most, of the lifecycle of the upwind fires with exceptions including e.g. intense combustion episodes that lead to free-troposphere injection and long-range transport. In theory, the most important use of this tower data is to test model predictions of smoke production and transport for the stable species measured. This is discussed a little and could be very valuable for future model evaluation in other papers. The work is new, very valuable, and should definitely be published with minor revisions as summarized next and also pointed out in the specific comments.

The study has some weaknesses, which need to be recognized in a more balanced discussion. In no particular order:

1. Towers can only monitor upwind fires limiting the range of sampling.

2. Any ground-based site may have some bias to smoldering or miss the type of emissions subject to long-range transport in the free troposphere. This is a difficult topic to achieve certainty on.

3. The uncertainty in the background at the tower is pretty large compared to the observed enhancements (in 2015) when far downwind and so the tower-based approach may only work in near-record fire years whose representativeness is unknown.

4. The initial emissions can only be measured for a few stable species but the vast majority of interesting fire products are reactive.

The current discussion is written as if the authors discovered potential sampling biases specific to geographic regions and platforms that have already been major concerns in mainstream thinking for decades. At the same time, they fail to emphasize the exciting finding, which is that past attempts to overcome the limitations of any one sampling platform appear to have worked pretty well according to the perspective provided by this novel, unique study. In other words, past compilations averaged together the results from multiple platforms in an attempt to overcome the limitations of using just airborne, ground-based, or lab data. The results in these compilations are virtually indistinguishable from the authors results for the two species they report, which is pretty remarkable. It inspires more confidence in the previous recommendations for countless other species reported in those compilations, which is good news from a fresh perspective. The authors miss the mark by instead dwelling on air/ground differences, which are worth pointing out, but were already well-known. I think the authors deserve credit for recognizing the unique opportunity they had to evaluate past recommendations, but mistakenly focus their discussion on the limitations of a subset of previous work. The value of validating previous recommendations is huge because past work was actually vastly more complete chemically and probed many other fire seasons and geographic areas. Imagine the millions of dollars it would cost to outfit a tower with instrumentation similar to that on the NASA DC-8 for just one summer and then maybe have a year like 2012 with no smoke or only downwind fires!

In addition, the study makes speculative, unsupported tangential claims about the particles from boreal fires despite the lack of any PM data. Despite validating previous recommendations, it is guessed that EFPM, and therefore health and climate effects, might be underestimated in models. However, the authors a) did not sample PM, b) may not have sampled the type of combustion that leads to long-range transport and wider impacts, c) did not consider secondary aerosol processes such as evaporation (see detailed comments), and d) fail to recognize that a PM network is in place that constrains the amount of PM in populated areas.

A brief warning, compared to other journals, ACP has pretty lax quality control and rarely sends papers back to the Referees for a second look. Thus the authors will be well advised to proofread future versions more carefully. There are typos that could be recycled or should have been caught by a spell-checker that I note along with other specific comments below by page and line number.

Specific comments format is page, line number: “comment”

1, 18: example typo, see page 6, line 35 EFCH4 is 5.3+/-1.8

1,22-24: How does smoke age impact sampling times? I.e. can’t you measure 24/7 from anywhere?

1, 24: high compared to what? not recommendations. How does “variable” inform a comparison? delete “high and variable”

1, 25: more prominent than what? Keep “prominent”, delete “more”, ”continuously” > “continuous”

1, 26: change “typical” to “a range of” since 2015 not a typical year according to authors.

1, 29: could add albedo and aerosol for completeness of overview here

1, 32 – 2, 2 – 2, 7: Exactly, but these “many” fires are forgotten about in the rest of the paper as it stands now.

2, 9: delete “future”

2, 11: delete “feedbacks”

2, 13: add “emissions of” before “specific” or it makes no sense.

2, 19: “have sometimes been” … Recommendations from Andreae weight all studies included equally, but the Akagi recommendations often consider amount of sampling, representativeness, quality of technique, etc. in recommendation as explained for each fire type in Sect 2. Users are encouraged to change the averaging formulas in the supplemental tables if justified for their application.

2, 22: “near and within” or “through” or “across”

2, 23: not just IR and WAS, other instruments include diode lasers, mass spec, and many others too, especially in ARCTAS.

2, 24: I'm not checking this number of fires, but note past work coves a variety of places and years, which is good.

2, 22-30: This is a nice overview of limitations of aircraft sampling, but equal attention is needed on limitations of fixed surface sites as noted in general comment.

2, 31: I would change "surface tower" to "fixed surface site" to make it more general and include the work by Collier, Gilman, Selimovic et al cited just below. Selimovic et al., 2019a is now just "2019" and "2019b" is now "2020."

Selimovic, V., Yokelson, R. J., McMeeking, G. R., and Coefield, S.: Aerosol mass and optical properties, smoke influence on O_3, and high NO_3 production rates in a western US city impacted by wildfires, J. Geophys. Res., 125, e2020JD032791, 2020.

2, 34: delete "]."

2, 37: add "of" before "smoldering"

2, 40: "fromfrom"

3, 3: fyi, smoldering converts solid biomass to gases, flaming oxidizes some of those gases. Yokelson et al., 1996, 1997

3, 6-7: Actually no way to have an open fire with low oxygen so delete "in a high oxygen environment."

3, 3 – 3, 13 and 3, 14 – 3, - 21: good overviews

3, 24: 5,858,000 30 s samples would be almost 3 million minutes, >48,000 hours, or >2000 days all within a ~90 day period! 58,000 samples is only 20 days….?

3, 26-28: "Analysis of these data indicate that smoldering processes may have a higher contribution to total wildfire emissions from North American boreal forests than previous estimates derived from aircraft measurements." Out of place as a result in the intro and also comes across as a random change of subject.

4,4 move sentence till after next on or rephrase as "… data stream we used …"

4,1-17: Take a few sentences to explain the data collection and analysis better and refer to tables. Clarify the following:

1) If you shifted to make continuous data, the time base would get further and further off or have jumps making it harder to compare to model?

2) The instrument sampled for 30 s then did something else for "<15s" then repeated until 50 minutes was up?

3) If 30 or more of the 30 s samples within one 50 min interval each had CO > 0.5 ppm the series was denoted as an emission factor event as shown in tables?

4) elevated CO for less than 30 of the 30s samples was ignored?

5) no emission factor events were allowed to span two different 50 min intervals?

6) How does the sample size criteria impact continuity?

More important than justifying any choice as the best choice is to explain once clearly what was done in section 2.1, how the instrument sampled and how data was reduced and tie that explanation to the Tables – making sure tables are called out in right order.

4, 21: Correlation among these species occurs for all combustion, including traffic in Fairbanks, but hopefully low anthropogenic influence at tower.

4, 23-24: So assumed a flat background for CO and CH4 for the whole summer regardless of wind direction, etc. rather than fitting a baseline from before to after each peak? Aren't ecosystem CH4 fluxes potentially variable?

4, 24-26: So the model reproduced 2012 when few fires occurred and then was run with 2015 input to get a 2015 calculated background?

4, 34-36: Even if the calculated background level changes slowly it could be the wrong level. Fractional uncertainty in the fire excess CO2 is roughly the uncertainty in the background (~3 ppm from Fig. 2a) divided by the size of the enhancement (~15 ppm from Fig 2b) for about 20% uncertainty on average? Or, if you just want one ER for the whole season you could just integrate the excess over the whole summer or do regression on the whole summer and get uncertainty from the uncertainty in the slope. Computing integrals for the whole summer might be a step closer toward a flux-based EF? Could be interesting to see how the result of that approach differs?

4, 37: did you get a slope for each 30 data-point+ "interval" and are "intervals" individual peaks or could they be partial or multiple peaks? Are intervals typically associated mainly with one fire?

5, 1-9: I did not check formulas, but got same EF results for CO and CH4 from reported ER so probably no typos? Also, why not report EFCO2?

5, 17: "the sampled combustion processes"

5, 24-26: Varying plume injection height within the boundary layer may not impact result at tower a lot if PBL well-mixed, but it excludes injection into the free troposphere during intense combustions and arguably would reduce the importance of long range transport, which is highlighted in the intro and conclusions.

5, 28: "isis" hacked your paper:)

5, 33: It's not dark yet at 6 pm in summer in AK? But with this definition, 10% at night seems low, is there GOES FRP to back that up?

5, 35: “83% of detected fire activity”

6, 2: “roughly consistent” i.e. almost a factor of two different

6, 3-13: Nice modeling application here. Were the individual fire contributions too mixed-fire events at the tower computed on a whole season daily, hourly, or interval basis? Some large fires may not have grown much on the day they impacted the tower?

6, 14: units are not immediately understandable, maybe explain in a bit more detail?

6, 18 & 20: Useful to define emission factor “event” or “period” earlier when describing how data stream analyzed?

6, 20-25: This is a cool analysis and useful that likely represents a lot of work! Not a criticism, but the finding that 27% of smoke impacting the tower was emitted at night, but the model assumes 10% of total AK smoke was emitted at night kind of shows the difficulty in proving representative sampling. Or what else does it mean? One general philosophy for dealing with this quandary has traditionally been to sample in multiple ways and synthesize the results; and simultaneously take the differences between approaches as a rough estimate of overall uncertainty. This is sort of what happens when using a literature average/stdev, while I acknowledge weighted averages can be better than straight averages in some cases.

6, 29: Are these 55 events the same as the EF events or periods? Are they all < ~50 minutes long? If the CO rose for two 50 minute periods and then fell for two 50 minute periods, is that one peak an emission factor event or is it 4 events? The data reduction can easily be spelled out clearly at the outset for folks that did not do the calculations and might wonder. Has the table of events been called out yet?

6, 29-30: The definition of an event earlier was lasting ~900 or more s? Here all the events lasted 50 minutes? So each hourly measurement interval with high enough CO was an event? I think it might be easy to take a few sentences above to just spell out how data was analyzed. Then I look at Table 2, are these the events and is N the number of 30 s increments? Maybe explain that earlier and include if each of these events is separated by a clean period?

6, 31: it would be interesting to see range in CH_4/CO also in this sentence.

7, 8: “within” should be “with”? Table 3 called out by mistake? Also on line 16?

7, 20: diddid

7, 21, 22, 23: Variability < 5% probably not significant. Were events actually different fires? What is meant by flux-weighted estimates? Accounting for fuel consumption rate in a weighted average EF or windspeed at tower? The highest flux periods at the fire may produce high injection altitudes.

7, 25-27: Figure 7 shows some big peaks at tower, but not in model (doy ~188) or modeled peaks not seen at tower. The text says the model confirms elevated CO was primarily from fires. So I guess “primarily” signals > 50% and signals rough agreement? The authors stand by the unmodeled peaks being due to fires? How was it possible to get the fires contributing to the signal at the tower when the model did not capture a peak?

7, 28: “likelycaused”

7, 34: average distance weighted by fractional contribution?

7, 36: What is meant by “integrate emissions from multiple fires through the full planetary boundary layer”?

7, 39: > 8% in Table 3

8, 1: delete “significantly”

8, 3: 4646%

8, 8: Andreae-associated recommendations averaged the values from studies using different platforms partly in recognition of bias being possible for any one platform. Akagi et al pioneered splitting extratropical forests into boreal and temperate. They (Sect 2.3.2) actually used a pretty complex scheme averaging smoldering fuels from lab studies by fuel type rather than by study to get a ground-based average, which was then averaged with airborne results for an overall average roughly consistent with about 70% of overall fuel consumption by smoldering. They mentioned evidence that smoldering might be even more important. They devised formulas to estimate compounds measured only in lab or air and invited users to modify any of the formulas in their Table S2 if they preferred. Remarkably, their default recommendations are almost indistinguishable from this work. Regarding “important” differences on P8, L15, keep in mind that modelers determine the level of detail that works for them and it often involves model domain, scope of study, availability, reliability, and complexity of operational input, but also completeness, i.e. they need ERs/EFs for more than 2 species!

8, 13 re Table 1: Good idea to parse out data by location and platform and nice overview of data collected. Note Yokelson et al 1997 is missing (used in Akagi Table S2). Boreal peat was burned in Stockwell et al., 2015. Double check if Siberian fires were wild or prescribed, I think at least some were prescribed. Split Siberian fires out by air or ground? Siberian average row has possibly wrong total? Remove line numbers from number of fires column, “McMeeking has two capital “M”s, etc…

Stockwell, C. E., Veres, P. R., Williams, J., and Yokelson, R. J.: Characterization of biomass burning emissions from cooking fires, peat, crop residue, and other fuels with high-resolution proton-transfer-reaction time-of-flight mass spectrometry, Atmos. Chem. Phys., 15, 845-865, doi:10.5194/acp-15-845-2015, 2015.

Yokelson, R.J., D.E. Ward, R.A. Susott, J. Reardon, and D.W.T. Griffith, Emissions from smoldering combustion of biomass measured by open-path Fourier transform infrared spectroscopy, J. Geophys. Res., 102, 18865-18877, 1997.

8, 15: “measurement technique” should be “sampling strategy” to be consistent and precise?

8, 17-36: The overview of air versus ground sampling of sources is pretty good, a little disorganized but all the most important points emerge clearly! A few points to add could be: Aircraft can replicate tower-based sampling with downwind vertical profiles, but not on a continuous basis like a tower. Also, any aircraft bias toward flaming combustion may actually be partly okay if it weights the EF towards times of higher fuel consumption, relevant to author's desire for flux-based EFs? Flaming always entrains some smoldering, and the entrainment footprint is larger with more intense flaming. Best not to oversimplify a complicated situation.

8, 29: "weak or non-existent" convection columns (aka "updraft cores"). Mostly true for fresh RSC emissions, so "usually" is a good qualifier since some RSC may get to aircraft altitude by non-fire uplift or be sampled in rare missed approaches.

8, 35: "yet rarely" is okay – a fresh RSC sample would require "a really good drill on the front of the plane" to quote a DC-8 pilot.

9, 4: "combustion of"

9, 4-5: Organic soils were focus of lab study of Yokelson et al., 1997 and included in Stockwell et al., 2015 during FLAME-4.

9, 7: "should" > "could" or "might" (see above on models)

9, 8: At least five lab studies burned boreal fuels, the CO/CO2 ratios for FLAME-4 for black spruce and boreal peat are in supplement of Stockwell et al., 2015. Listing what fuels were included in averages in Table 1 would be helpful.

9,12 "McMeeking"

9, 18-29: The claim of different ERs is not strongly supported. The quoted (Table 1), purely surface-based sampling of Siberian fires had *lower* CO/CO2 than the authors NA work, and, even more remarkably, only about half the CO/CO2 ratio as the studies that included some airborne sampling of Siberian fires. So maybe better to say, the ecosystems differ and the emissions might as well, but not enough data to know yet.

Also work on the Siberia/NA differences goes back to at least 1993 when the Bor Island Experiment was started. Differences in Siberian and North American boreal fires were noted in publications 20-24 years ago with hundreds of references cited and a more recent review on that:

Goldammer, J.G., and V.V. Furyaev. 1996. Fire in ecosystems of boreal Eurasia. Ecological impacts and links to the global system. In: Fire in ecosystems of boreal Eurasia (J.G. Goldammer and V.V. Furyaev, eds.), 1-20. Kluwer Academic Publ., Dordrecht, 528 pp. https://link.springer.com/chapter/10.1007%2F978-94-015-8737-2_1

E.S.Kasischke and B.J.Stocks, eds. 2000. Fire, climate change, and carbon cycling in the boreal forest. Ecological Studies 138, Springer-Verlag, Berlin-Heidelberg-New York, 461 p.

Goldammer, J.G. (ed.) 2013. Prescribed Burning in Russia and Neighbouring Temperate-Boreal Eurasia. A publication of the Global Fire Monitoring Center (GFMC). Kessel Publishing House, 326 p. (ISBN 978-3-941300-71-2). http://www.forestrybooks.com/

9, 23: “hotter” okay, but there is no single temperature that defines any landscape fire, more aggressive flaming is probably what is meant.

9, 30: “Stronger” than what? Not a complete thought. Here the work goes off on a random tangent rehashing a long-recognized issue. Concerns about air/ground bias are discussed in Andreae and Merlet, 2001, which supports this with the following citation:

Andreae, M. O., E. Atlas, H. Cachier, W. R. Cofer, III, G. W. Harris, G. Helas, R. Koppmann, J.-P. Lacaux, and D. E. Ward, Trace gas and aerosol emissions from savanna fires, in Biomass Burning and Global Change, edited by J. S. Levine, pp. 278 – 295, MIT Press, Cambridge, Mass., 1996.

Previous recommendations by Akagi and Andreae appear to have compensated adequately for this issue according to this studies results to the extent that we are ever likely to know. The authors could claim that they have investigated the extent of platform-based bias in additional detail and present a useful contribution in that way, but the issue of the existence of differences is not a new finding.

Perhaps an appropriate header is: “A detailed examination of tower versus airborne sampling”. Either include or don’t include the enigmatic data from Siberia and make a new, useful point if you can, perhaps: a) mean difference is “X”, or b) surprisingly no conclusion.

9, 38 – 10, 1: The authors have good evidence that tower-based platforms see more smoldering that the aircraft studies to date (in NA) and that is useful, but you don’t know for sure if the tower might under-estimate flaming or why the Siberian data is enigmatic. And “previous reports” should be changed to “the average of previous airborne studies” since “previous reports” could imply all studies.

10,1-2: at a minimum change to “some previous” , “some flaming” , “some residual”

10, 1 – 8: Showing that the tower and aircraft got different overall average CO/CO_2 ratios is straightforward and useful. But both platforms could have some error so proving that 100% of the error in representativeness is with the aircraft is not really doable. Every fire that impacted the tower also, undoubtedly produced some emissions that did not impact the tower due to wind shifts, altitude, or whatever, it’s just basic common sense. The most exciting thing about this work is not even stressed. That is, by measuring downwind of many fires burning at all stages of their life cycle around-the clock, the authors have created a high-quality data set for evaluating fire emissions models performance at a regional level (as in Selimovic et al., 2019; 2020). I.e. AKFED did pretty well integrating the effects of many fires around the clock and predicting the tower “point CO” specifically. What would need to be changed in AKFED/STILT to improve performance could be a great follow-on study along with how do larger-scale models such as GFED, GFAS, FINN, etc., perform against the tower observations! Regardless of the “real, unknown total fire emissions,” the signal at the tower is well-measured now and very useful to test models!

10, 9: This next paragraph repeats some of the material in the previous paragraph. If the text survives editing, change “crown” to “surface fuels” since the NW Territory crown fire experiment found that often the fires propagate in surface fuels followed by torching

10, 10: not a sentence, delete “that” to fix?

10, 11: substantial contributions of RSC were stressed by Bertschi et al 2003 in their Table 3.

10, 12: not only FRP-based, but any thermal signal. Smoldering involves temperature high enough to saturate the 3.9 micron channel if widespread enough, but obstruction is more of an issue for deep smoldering.

10,16: “previously thought” by who? If you mean the studies mentioned directly above, no comparison is possible unless you also have the relative consumption of above- and below-ground fuels. If you mean a previous compilation, Akagi et al estimated 70% of all boreal fuel was consumed by smoldering based on an MCE similar to this work.

10, 18: “most” is not “all” and MISR only looks at 1030 AM long before both the most intense combustion and diurnal cycle fuel consumption peak. I would change “most” to “many” or “some”

10, 19: “length scale”?

10, 20 – 21: good point there is time for vertical mixing, if the atmosphere is not too stratified.

10, 22-25: This sentence is not accurate as Collier et al sampled smoke up to 48 hours old and Selimovic et al 2019; 2020 sampled smoke from fires in the range 20-800 km upwind.

10, 24: if text survives “band” > “and”

10, 26-35: This discussion is oversimplified since dry weather can make larger fuels that tend to smolder more likely to burn. See section 2.4 in Akagi et al and the papers referenced therein by Hoffa et al., 1999, Shea et al., 1996, Kauffman et al., 1998; 2003, etc. Hot, dry weather can increase smoldering. Note also that most airborne studies occurred in years that were arguably more “typical”.

10, 40: I don’t think any new ideas “emerged”, but the authors work can help continue to evaluate some long-recognized issues and maybe help reduce uncertainty.

11, 1: it always makes sense to “report” what you measured and studies of Siberian fires report their location. “Using” regionally-specific EFs might make sense, but is a separate decision for the modelers that is hard because few measurements have occurred in Russia where research access is super-problematic. A colleague had their canisters confiscated by the Russian military and “filled for them at undisclosed locations.”

11, 5 – 11: This is not that big a deal, the complex averaging scheme of Akagi gave almost the same answer as this study or the simple averaging scheme of Andreae. Adding this studies extensive results in a weighted or simple average to the “evolving literature average” is BAU and will have little impact on the average; though it is important to be clear about how things are synthesized.

11, 12 - 18: It is not clear what is meant by flux-weighted EF? Aircraft measurements may in fact be weighted towards times with high fuel consumption rates. Fires can produce multiple plumes. A flux of emissions in models results from the fuel consumption rate assumptions.

11, 18-23: This is just stating obvious that if we could measure everything, we'd know more. I would very strongly recommend deleting sections 4.3 (and 4.4) and instead have a section to highlight the exciting model evaluation now possible with what you *already* measured.

11, 24: "larger" than what?

4.4 is all speculation about PM, which was not even measured and the section doesn't consider SOA or PM evaporation where the latter was significant in Selimovic et al., 2019, 2020, and references there-in. Also, health impacts are based on measured PM and this study does not suggest the regional PM networks are inaccurate.

11, 27: higher than some studies doesn't equal higher than "previously thought".

11, 28 – 29: Long range transported smoke was not sampled in this study and that type of smoke may actually be better sampled from aircraft.

11, 25 – 36: delete, all speculation, not a topic or result of this paper.

12, 3: after "Our results" I would delete the rest and fill in the valuable insights that you actually learned about the AKFED model. I.e. it underestimated nighttime combustion impacts at the tower, it captured X of the Y peaks, seasonal average CO was within Z%, etc… Highlight potential for additional, future model evaluation and improvement.