

We wish to express our appreciation to the referees for their thorough and thoughtful analyses of our manuscript. Based on the suggestions and comments of the referees many improvements were made in the revised version of the manuscript, and we are confident that the referees and editor will find that the new version addresses the concerns that were raised. Below we have responded to each of the comments made by the referees.

Response to Anonymous Referee #1:

Comment: Page 14132 line 20-27: A description of what will be discussed in a future paper does not seem appropriate here. I think at most one sentence advertising the paper is enough.

The description of the future paper (Part 2) has been reduced to a single sentence.

Comment: Page 14138: For the inter-comparison the R coefficients should be given as well, because they tell a lot about the quality of the measurements. A good agreement of the slopes shows only that the calibrations of the different instruments agree. Especially looking at the TOGA-WAS comparison of the oxygenated VOCs a large scatter in the data is obvious, which is not the case for the alkanes. This means that the sampling time artifacts cannot be the cause for this large scatter. Oxygenated VOCs are generally hard to measure and often don't seem to be very stable in canister samples. A discussion of the measurement quality of each compound class using the three different instruments should be added and a recommendation could be given, which instrument is the best for which compound class.

Because the regressions on the comparison plots are orthogonal distance regressions, there is no r -coefficient for the regression. However, standard linear-least squares regressions have been determined, and the r^2 value for the linear regression has been included on the plots in Tables 1 and 2 following the slope, intercept and uncertainties of the ODR to give further information regarding data comparisons.

The scatter in the TOGA v. WAS oxygenated NMOCs comparisons is discussed in the paper, and text has been added to the paper describing the uncertainties in the oxygenated NMOCs reported from the canister measurements as further explanation to the spread in the comparison plots. However, apart from pointing out the uncertainties in the WAS-measured oxygenated NMOCs, we prefer not to recommend a specific technique as the best to use for "each compound class" because this is not the focus of the paper; a recommendation such as this may be appropriate for an intercomparison paper but we feel it is not here. Further, the combination of the three different instrumental techniques reporting observations of numerous overlapping species provides internal validation to the measurements.

Comment: Page 14138: If the measurements of the different instruments agree, the fastest instrument should always be used for the analysis to get the best correlation determining the enhancement ratios.

A higher frequency of data acquisition does not necessarily imply better or more reliable data. A comparison of the results shown in Table 2 shows this quite clearly. High-frequency measurements are only reliably superior to or more useful than lower-frequency measurements when the period of interest is less than the sampling period of the slower instrument.

Comment: Page 14139 line 25: The term plume group is somewhat confusing. If I understand this sentence correctly a plume group is the same biomass burning plume that has been intercepted multiple times, which I would not call plume group, but simple a plume.

The terminology has been changed for the 40 identified plumes in Table 1 from “plume groups” to “plumes”.

Comment: Page 14140: A short explanation of what the MCE means should be added here. It is given only late in the paper and should be moved here.

The paragraph has been changed to include a description of the meaning behind modified combustion efficiency (MCE).

Comment: Page 14143 line 10: A short discussion about the possible urban sources of HCN for the Californian fires needs to be added here even though the enhancement ratios are lower compared to the Siberian and Canadian fires.

This has been added.

Comment: Page 14143: A discussion about the importance of acetonitrile as a biomass burning marker should be added. Especially since the enhancement ratios do not vary significantly between the source regions, acetonitrile is a good tracer.

Text has been added that reiterates that acetonitrile is a good biomass burning tracer based on the fact that the enhancement ratios don't vary significantly between the source regions.

Comment: Page 14144: The ratio of acetonitrile to HCN can be calculated very easily from the Tables and I don't think that has to be discussed explicitly here. It was explained that the enhancement ratio of acetonitrile with CO is relatively constant and HCN is rather variable, therefore it is pretty obvious that the acetonitrile/HCN ratio is variable in the same way as HCN/CO. So I suggest deleting lines 1-14.

The authors agree, and the paragraph has been deleted.

Comment: Page 14144 line 18: typo ARCTAS

The typo has been fixed.

Comment: Page 14145: Relatively small acetone production in all the observed plumes, even in the summer ones, suggests relatively low OH in the plumes. Is this consistent with the low acetaldehyde enhancements in the Asian plumes?

Yes.

Comment: Page 14148 line 7: Apel et al 2011 is not in the reference list.

It has been added.

Comment: Page 14148 and chapter 3.5: The chapter about the alkanes should be reorganized. The plume group 28 seems to be not only biomass burning in origin, and should be removed from this analysis. This plume group should certainly not be discussed at the beginning of this chapter. I would suggest doing the calculation of the alkane enhancement ratios without this plume and then discuss the results, after that briefly mention plume group 28. Chapter 3.5 should then be moved into the alkane section. Going back to the discussion of measurement results after the modeling part is not good for the flow of the paper. If correlations of the light alkane enhancement ratios versus propane are discussed, other compound classes should be treated in the same way. Correlations of the aromatics likely show the same behavior as the alkanes, whereas the oxygenates might not.

Sections 3.5 and 3.4 have been exchanged so the discussion regarding the correlations of NEMR immediately follows the discussions of individual NEMRs. Plume 28 has been eliminated from the discussion and the plots, simplifying the presentation of the NEMR correlations. Also, the comparison between the alkane NEMRs has been changed to use the NEMRs determined using the WAS data for all the C₂-C₅ alkanes. This has made the comparisons far more robust because the observations being compared are simultaneous, rather than “overlapping”.

The new section 3.4, now called “Ratios of NMOC enhancements” has been expanded to include a section comparing the NEMRs of aromatic hydrocarbons, both to each other and to propane. Regarding aromatics, we have also included plots of the comparisons of the NEMRs for benzene and propane, toluene and benzene, and C₈-aromatics and toluene (now Fig. 9), which all indicate less correlation than comparisons of NEMRs of C₂-C₅ alkane. This is not surprising based on lower emission rates and typically faster chemical loss rates in the atmosphere for these compounds. As expected, the relationships that do exist between aromatics are primarily dependent on the age of the plume, and are therefore primarily dependent on the chemistry occurring within the plume, and not the emissions. Similarly, comparisons for the oxygenated NMOC indicate even less correlation when all data are included, but by looking at only younger plumes (estimated to be < 2 days old), the relationships between the NEMRs to CO for various oxygenated NMOC are now explored, as is the impact of plume MCE.

Comment: Page 13149 line 15: Is the difference of the propane enhancement ratios between spring and summer consistent with the acetone production?

It is consistent, and this detail has been included in the text in the propane section of the alkane NEMR discussion section (3.3.4).

Comment: Page 14151 line 23: I don't understand what you mean with “more polluted conditions”. Are you referring to the mixing ratios in the biomass burning plumes or the atmospheric background conditions?

The text has been changed to clarify that the higher NO_x and O₃ in the CAL BB starting conditions are consistent with the more anthropogenically-influenced background conditions in California compared to Saskatchewan, and the terminology “more polluted” has been removed.

Comment: Page 14155 line 25: Why should averaging fire emissions over large areas lead to significant under- or overpredictions in emission inventories. This should only cause large scatter, when comparing individual plumes to model results.

This specific text has been removed, and the Conclusions have been made more specific with regards to our findings.

Comment: Table 3: Please add the new Akagi et al 2011 paper to the comparison.

Response: While we have added the Akagi et al., 2011 paper as a reference elsewhere in our paper, Table 3 and 4 lists our measured NEMR against several previously-reported values of field-observed NEMRs, and this is not what the Akagi reference includes. The Akagi reference is primarily a (very useful) fire terminology primer, as well as a very thorough summary of emission factors for a host of NMOC.

Comment: Table 4: Warneke et al 2009 is missing from the reference list. A very useful addition to this paper would be to discuss the relationship of some VOCs with the MCE. I think this would be more interesting than the results presented in Figure 10.

The Warneke et al., 2009 paper has been added to the reference list. Also, the results in Fig. 10 (now Fig. 8) have been simplified by removing plume 28 from the analysis. However, there is still considerable interest in the contents of Fig. 8, and the comparison of NEMRs in similar-structure NMOC families. Further, as mentioned earlier, the relationship between the NEMRs and MCEs is now explored in the new Fig. 10, the comparisons of the NEMRs to CO for oxygenated NMOC.

Response to Referee #2, Robert Yokelson:

Comment: ARCTAS characterized the trace gas composition of the biomass burning impacted boreal atmosphere in NH summer 2008. The total number of pyrogenic gases measured during ARCTAS greatly exceeded that in previous field campaigns in the boreal region with the instrumental improvements since NASA's last boreal forest fire mission (ABLE) being nothing short of spectacular. The measurements covered an extensive geographic range and are of high accuracy. The data should definitely be published. This paper is one of at least four papers that summarized the trace gas data from multiple instruments on the DC-8. All were multi-investigator, but none were comprehensive in including all the data. In Simpson et al the fresh plumes without signs of aging were carefully selected and used to compute initial emission ratios (ER) and emission factors (EF) as model input. They unfortunately did not include the TOGA gases though. In contrast, this paper presents the average normalized excess mixing ratios (NEMR) observed for each non-methane organic compound (NMOC) to CO (for fresh and aged plumes) over the course of the campaign, which depends on the average plume age and species lifetime. Because of the extensive geographic range sampled, these average NEMR give some idea of how polluted the boreal atmosphere was due to fresh or aged smoke. This paper however only includes a few of the WAS gases and not the HCHO. It's unclear to me if all the PTR-MS data was published or how much overlap the above papers have with papers by Singh et al and Vay et al. Since outsiders are too busy to read all the papers, my review will reflect that knowledge gap and it would be great if future missions could strive to produce comprehensive papers organized by themes.

The authors have added analysis of observed HCHO in this work to make it a more (albeit not absolutely) complete analysis of the NMOC in observed in BB during the ARCTAS field study.

Comment: I have some general comments on a few concerns I have with the paper, followed by a summary of the many things I like. I should state at the outset, that most of my concerns are largely an artifact of the flight planning, which was likely beyond the authors control. I discuss these concerns at

some length, not to lecture the present authors, but hopefully to impress upon the entire ARCTAS team that these issues may need to be acknowledged throughout all the ARCTAS publications.

It has been clear since the beginning of biomass burning research that it is highly variable – not unlike many natural phenomena. The highest variability may be seen during a single fire as a function of time because the flaming to smoldering ratio evolves causing high variability in individual samples or “intercepted plumes” as demonstrated at least all the way back to Lobert et al 1991 and Yokelson et al 1996 in lab studies. Real fires also normally flare-up and die-down as they burn in response to many different environmental variables, which changes their flaming/smoldering ratio. In fact, as the figures in Yokelson et al., (2008) show, the full range of variability cannot even be measured from a single platform in the field, with most of the range only being accessible to ground-based sampling.

If multiple samples of a single fire are acquired and one estimates fire average values, then one can examine fire-fire variation. Most papers provide the average EF for NMOC and the 1-sigma uncertainty in the mean, often in a vegetation classification scheme. The variability is often prominently displayed as a function of MCE or the CO/CO₂ ratio, which are both proxies for the flaming to smoldering ratio. This can be seen in Figure 3 of Yokelson et al., (2009), but has been ubiquitous in the fire literature since the early 1990s. Reporting the mean and uncertainty for a group of fire-average values is not done to be “variability deniers,” but simply to provide the data in compact, convenient form. It’s widely assumed that, in large-scale applications, the modelers will probably want to apply an average EF value to all detected fires in an ecosystem rather than having to generate a distribution of EF about the mean and applying that distribution of EF to a set of hotspots, or whatever. Based on standard statistical theory, it’s understood that 95% of the observations would likely fall within 2 standard deviations of the mean. Since the average literature value for one standard deviation is typically ~40% for an NMOC, the range would usually be at least +80% or a factor of 9. However despite this high variability there are reproducible ecosystem effects on average values. E.G. taking CH₃OH as a generic smoldering NMOC, the study-average CH₃OH EF for savanna fires is lower than the study average CH₃OH EF for tropical deforestation fires. Since vegetation maps are ubiquitous, this is an accessible approach to increase the spatial representativeness of the model. But in some applications, such as modeling individual plumes, the limitations of using average values may be severe and should be acknowledged.

Finally some work has looked at the variation in study average values for individual ecosystems. The compendium of Andrea and Merlet (2001), and the updated compendium of Akagi et al (2011a) both report high average study variability and even high variability study to study in some cases. This variability discussed above means that in studies that randomly sample different plumes of different ages, the photochemical changes would be expected to be masked to a large degree by the inherent variability. Even in studying a single plume, the photochemistry can be impossible to quantify if the source is changing. In particular, an aircraft is much faster than the wind, so sequential (long axis) sampling, if e.g. started at the source and then pursued downwind is likely to compare source smoke to downwind smoke emitted much earlier. This can mask the photochemical changes unless it can be shown that the source was burning in a fairly constant manner as in Yokelson et al 2009, or if the sampling is Lagrangian or pseudo Lagrangian as in Hobbs et al 2003 or Akagi et al 2011b. There is an additional problem that as plumes evolve the variability may increase. In any case, the NEMRs can be significantly distorted due to changes in the background air that the plume dilutes with. One can easily demonstrate in a spreadsheet that diluting a plume with constant background air will not effect the dX/dY values, the NEMRs in the plume. But, if X/Y changes in the entrained background air, then dX/dY no longer reflects the original source characteristics. This is seen in a few of the plumes in the present work that have non-physical MCE that are > 1. The uncertainty in sample ages is very significant and

requires some subtle analysis. In the best of circumstances, in a single plume the 1-sigma variability in the windspeed is usually 50% of the mean. When the plume source is physically located and samples in an isolated plume are aged based on distance downwind and windspeed, the uncertainty in age is quite high due to the uncertainty in the windspeed. However, because there are unlikely to be any "passing lanes" in a single smoke plume, at least the age order corresponds to the distance order and so the relative ages of samples are well-known. When sampling different plumes this advantage is lost. A plume thought to be 4 h old may only be 2 h old due to uncertainty in windspeed and a different plume thought to be 2 h old may really be 3 h old. Thus, the order of samples is no longer confidently known and the age uncertainty needs to be prominently acknowledged.

Another large source of age uncertainty that is hard to quantify likely exists in the ARCTAS data if samples were aged by running back trajectories back to active fire detections. Many fires are not detected as hotspots due to size or cloud cover (Yokelson et al., 2011, Burling et al 2011), with the latter being ubiquitous during ARCTAS. In addition, plume injection heights are largely unknown. So a sample may not be emitted at a hotspot that the back trajectory passes over, but instead at a different hotspot or at a fire with no hotspot.

For this reason, our estimated ages are given a significant uncertainty, and are meant only to distinguish fairly young fire plumes (0-1 days) from slightly aged (2-3 days) to fairly aged (4-8 days) plumes.

Comment: In summary, to do aging best, one needs lagrangian or pseudo-lagrangian flight plans or some evidence of a steady source and an isolated, individual plume that can be verified not to have it's photochemistry "reset" by mixing with other fresh plumes downwind.

If the goal is to instead compare average observed NEMRS, a best effort is needed to compare NEMRs for reactive species in similarly-aged samples. Because of variability in the NEMRs and uncertainty in age, it is unclear to me how well-constrained the measurement model comparison is. I'm not sure what the take home message of the modeling portion is as explained in more detail later. It surprises me that two key photochemical intermediates, HCHO and OH, were measured on the DC-8, but not discussed in this paper. These species are critical parameters to get right in models (Alvarado and Prinn, 2009, JGR). Thus, e.g. the model OH should be compared, even in a sentence, to the measured OH. The disconnect between the mission goals and what the flight plans could actually deliver seems to have led to some intellectual confusion in this and many ARCTAS papers. In this paper, the abstract may be an example of this. In outline form it seems to say that (1) variability is huge and (2) the variability has been underestimated, (3) our variability agrees well with previous studies, (4) the model agrees well with the hugely variable data. On first principles one would wonder if the latter means the line bisects the scatter or that the model can reproduce the scatter. Some other key species that are not mentioned were not measured in ARCTAS, e.g. organic acids, but their role should be acknowledged anyway. Further many NMOC still cannot be measured with current technology (Christian et al., 2003; Karl et al., 2007; Warneke et al., 2011). The uncertainty due to both these factors should be acknowledged.

The authors have more carefully delineated the issue of observed variability. Organic acids are now mentioned in the text, and the fact that there are many NMOC that are not measured by current technology has also been added.

Comment: The authors do a good job of comparing to a lot of studies, but a great deal of relevant previous work in the boreal ecosystem is unmentioned. E.G. Goode et al., 2000 (JGR), the entire NASA ABLE mission that studied boreal fires, Nance et al., 1993 (JGR), etc. The Goode et al paper and

references therein, may help the authors track down papers that provide important context and relevant info! Also modeling papers that are relevant are not cited as I point out below.

The literature survey is not intended to be a complete review of all previous measurements, as explained in the text.

Comment: It would be helpful if the authors could clarify even in a few sentences the rationale behind the TOGA species selection – could it be expanded to include more pyrogenic species?

The compounds that are reported from the TOGA system represent those that have been identified and included in calibration standards, and are by no means the only NMOC that can be observed by the system. For example, in 2010 the species list was expanded to include several monoterpenes. For future field studies, other species of interest are likely to be added, which could include additional pyrogenic species. For the ARCTAS data set, the possibility of extracting additional compound information partially relies on whether or not appropriate masses were scanned during the elution of that compound from the column. If so, then there would be a fair amount of work involved in adding it to the data set. The current list of compounds that have been included in the ARCTAS data archive are those that have been historically measured and reported, as well as the most-requested and most-useful across a number of applications. Text has been added to the manuscript in this regard.

Comment: Terminology issues: if EMR stands for “normalized excess mixing ratio” then what would you use to abbreviate excess mixing ratio? Since the NEMR abbreviation has been in the literature since Hobbs et al., 2003 it may be wise to stick with this admittedly clumsy term. I am trying to get people to use NMOC instead of VOC as a general term for organic gas-phase emissions from fires since about half of the organic gases are not VOC, but in fact “SVOC.”

The reported “normalized excess mixing ratios” are now being abbreviated NEMR, and the term VOC has been changed throughout the manuscript to NMOC. We agree that these terms should be used consistently in the literature, and are happy to comply.

Comment: I am not 100% sure the intercomparison was done right as explained in more detail below.

Comment: LRT [long-range transport – RSH] happens at all levels of atmosphere including BL too, just typically at different speed and direction if in BL instead of FT.

The text in the introduction has been changed to say that long-range transport in the UT and LS can “widely and rapidly” distribute emissions.

Specific:

Comment: 1. The authors could provide a very useful list of EF for the TOGA species that were not included in Simpson et al. The formula is easy to implement in a spreadsheet. There is no need to redo the whole carbon mass balance as involving CO₂, CO, CH₄, and the TOGA NMOC will get close enough to total carbon to get a great approx to the EF. However, the authors should be careful to select only fresh plumes (with MCE < 1) for EF for the reactive species.

The text now includes an acetaldehyde EF in the acetaldehyde discussion in Sect. 3.3.3, and also a recommendation for the ethanol EF later in the same section.

Comment: 2. HCN and CH₃CN ER to CO – this section is already good – although the authors might explore the impact of eliminating plumes with MCE >1 and discuss how their results compare to those of Simpson et al., (2011).

The results from Simpson et al. (2011) have been added to Tables 3 and 4, and are now discussed in Sect. 3.3.

Comment: 3. It's useful to see TOGA confirmation of the biomass burning source for infrequently observed emissions also reported in Simpson et al such as ethanol and the discussion could be slightly expanded to compare to that paper.

Again, Simpson et al. (2011) has been included for comparison.

Comment: 4. The intercomparison between TOGA, PTR-MS, and WAS may need to be tweaked (vide infra), but is of great value.

See comments later regarding the instrument comparisons.

Comment: 5. The authors do not extrapolate Boreal results to global scale! This shows excellent judgment since boreal fires and tropical fires differ in many important ways.

Thanks, and agreed.

Comment: 6. The comparison of observed NEMRs is useful, but it should be clear what is being compared and when needed, the authors should tighten up or clarify the comparison of NEMRs for reactive species to literature values to ensure that they are comparing fresh to fresh or "old to old" i.e. similar ages.

There is a significant amount of work shown between the Tables of data (with estimated plume ages included for each plume), and the analysis including Figs. 7 - 9 that includes ages, including a discussion regarding the fact that plume ages are not solely responsible for the differences in the observed NEMRs, particularly for the longer-lived alkanes.

Comment: 7. The authors work to compare the ratios of NMOC to each other instead of CO - with age is not a panacea for eliminating variability, but definitely has high potential for useful simplifications since similar compounds are often produced by similar combustion processes and "track" with each other (Yokelson et al 1996). As time allows, the authors may wish to explore this further.

This section (now Sect. 3.4) has been expanded to include comparisons of aromatics, comparisons of oxygenated NMOCs, and impacts of MCE.

Comment: 8. The variability is not news and as community we strive to see the underlying patterns despite variability. Thus, perhaps it's worth a few minutes to try binning or other standard mathematical approaches to discerning underlying trends in variable phenomena?

The work in Sect. 3.4 decomposes the observed variability by comparing emissions of NMOCs from similar-structure families, and by looking at impacts of enhancements to CO with regards to the fire stage/MCE, and there are certainly trends that were found.

Comment: 9. Is there any potential that the body of ARCTAS data, properly de-resolved, could be used to validate global model output or satellite retrievals (especially the many new products from ACE, IASI, TES, but also MODIS AOD, etc)?

Possibly, but this is currently beyond the scope of this work.

Comment: Specific comments in order of appearance “P” is followed by the “last two” page numbers and “L” indicates the line on that page. P29, L17-21: Usually the abstract highlights new findings. My comments on the abstract are previewed in my general comments above. Here I point out specifics. The long established inherently variable nature of fires is presented in the abstract, first as if it is new finding and then it is stated that this was already in the literature.

The text in the abstract and elsewhere to this effect has been changed to highlight the new findings, and eliminate reiterating established details.

Comment: P129, L21-3: “However, this variability is often diluted in the literature when individual observations are averaged to generate an overall regional EMR from a particular study.”

1. Most studies report the stdev of the mean, which is a fairly standard way to report variability. Are they saying it is important to instead report the full range instead of letting people calculate the range from +/- 2-3-sigma. 2. The averaging is done because modelers generally ask for a best estimate of the average. Are they saying models should instead use random number generators and generate a large range of EF to apply randomly to inferred fire locations – and should those random numbers then average to the central estimate and would it make it any difference in the model result? interesting q. 3. Some regions have very diverse ecology or land-use. Thus the trend has been to develop average EF for ecosystems or land-cover types rather than “regions.” In summary, what specific concrete point are the authors trying to make here?

In our findings from ARCTAS, we see a lot of variability in the observed NEMRs, but as shown in Figs. 8-10, there are sometimes patterns between the lines that exist not as random noise and unresolved variability, but rather as trends in the emissions to CO between like-structure NMOC families, or trends with MCE. The fairly common practice of taking the average and standard deviations of all observations gives the impression of a randomly variable mixture, and the natural but fine-grain variability from which we can tease out trends becomes lost. We are even guilty of it in this paper, by taking literature data and averaging two, three, or even several data points to provide a single number (although we have made an effort to retain sometimes two or three data points where applicable) from a specific paper for comparison’s sake, but what we have found is that there is sometimes a non-randomness to the variability that should not be overlooked. It is for this reason that plumes from a given flight were often analyzed separately, based on changing ratios to CO.

Comment: P29, L24: “generally consistent within a given region” : : : what is meant by “generally consistent?” I think previous studies demonstrated that there are some differences in the average EF values between gross ecosystem types. I don’t think the authors of most previous works intended any more than that.

This text has been eliminated.

Comment: P29, L25: What is meant by “earlier assumptions”? Do they mean earlier measurements are in error or are they referring to some other “assumptions?” this needs to be specified.

This text has also been eliminated.

Comment: P30, L3-5: The oxidation products of NMHCs and OVOCs are typically OVOCs so the sum of all OVOC/CO likely can't decrease over 2.5 days although individual precursors often do. Unless what you are saying is essentially that most of the NMOC (and their oxidation products) are very short lived. Can the authors tell us what the model says happened to the oxidation products and what they were? Unless the modeling analysis included organic acids and the unidentified species as input and products the statement might need qualification by referring to "measured" or individual species.

The text has been altered to read "low molecular weight carbonyls...".

Comment: L4-7: During 2008 there were a lot of fires in northern California in high elevation coniferous forests that are similar to boreal forests in many aspects. The authors should check on the likely fuels for their samples (including for Asia) – if the samples being compared had similar age, the similarities observed could either reinforce or downplay the idea of characteristic averages for ecosystems

This has been addressed.

Comment: L20, NMOC probably a better term as many of the emissions are SVOC?

Amended.

Comment: L25, Lefer et al 1994 and Goode et al 2000 (JGR) reported on the NMOC emissions and their subsequent evolution in boreal forest fire plumes. Reid et al 1998, Hobbs et al 2003, and Yokelson et al 2003 (JGR) reported on the NMOC emissions and their evolution in tropical plumes. Akagi et al 2011a (ACP) cite additional useful references in this respect.

There are many papers with one or two (or even several) compounds that we could have included for comparison, but for the sake of space, we chose to limit the literature comparison to approximately 12 references with multiple NMOC enhancement ratios to compare against the ARCTAS observations.

Comment: P31, L3: The new fire emissions inventory draws on a review by Akagi et al., 2011a, which explains the NMOC differences in detail – it's subtle. Not sure this sentence got it exactly right.

The VOC v. NMOC concern has been address throughout the paper, so this should eliminate problematic differences in terminology.

Comment: P31, L9: NOAA had a P3 in ARCPAC. Was the NASA P3 in ARCTAS then?

Yes. There were two P-3s in Fairbanks in April, one of NOAA's P-3s, and the NASA P-3. The NASA P-3 was also involved in ARCTAS-CARB and ARCTAS-B.

Comment: P32, L8-9: The authors point out that plumes are extremely variable and then state that by randomly intercepting fresh and aged plumes they studied plume aging in an "ideal manner." I think my general comments above show why the ARCTAS data is not well-suited for studying plume evolution. It may also be helpful to see the discussion of plume evolution in section 3.5 of Akagi et al., 2011a.

We never intended to suggest that the BB plumes intercepted during ARCTAS were sampled in an ideal manner, but having not suggested otherwise, we agree that it could be construed that it was being implied. Text has been added that explicitly states that the BB plume sampling was not ideal due to the number of competing mission objectives.

Comment: L19-26: An unusually extensive section on a paper in preparation, which could be pared down to customary length.

The discussion of part 2 of this work has been decreased to a single sentence.

Comment: P33, L19-26: TOGA undoubtedly made high quality measurements of the species indicated. It is interesting to me that the species list is a small fraction of the species that would likely have appeared in the raw chromatograms or that have been purportedly measured by GC-MS in other studies. Is it possible to give a sentence explaining how this specific suite of compounds was chosen? At this point it seems a bit eclectic or random and not targeted at the main emissions of biomass burning such as CO₂, CO, CH₄, C₂H₄, C₃H₆, etc? (I realize that ARCTAS had multiple objectives and that the reported species are very interesting and important!)

There are a number of factors that have been used to determine which compounds the TOGA can be used to quantify. Efforts are currently being made for the newest version of TOGA (the HIAPER-TOGA HAIS instrument designed for the NSF/NCAR G-V) to be able to quantify HCHO, C₂H₆, and C₃H₆, but these are not trivial for a system that is being cycled once every two minutes. For low molecular weight compounds, the starting temperature is crucial, and a sub-ambient room temperature is ideal. Light alkenes and heavier alkanes may also be quantifiable using the current methodology. Ideally, a Time-of-Flight (TOF) mass spectrometer will allow us to expand our list of compounds significantly, but currently the systems are all outfitted with quadrupole mass spectrometers.

Comment: P34, L3-7: The PTR-MS did not monitor a lot of significant BB species that it could have monitored as can be seen by referring to Karl et al 2007 or Yokelson et al., 2008. I am not sure why and not criticizing the selection, but perhaps something could be inserted to say "full mass scans were not implemented to increase time resolution and S:N for the following species selected because : : ." The reduced suite of species may need to be acknowledged when drawing conclusions from the data.

This assessment is correct. While TOGA and WAS provided low-time resolution VOC data, the PTR-MS was taken onboard the NASA DC-8 to get VOC data with a high time resolution in the 10-to-20 second range. This is not sufficient to perform full mass scans with good S/N. Thus a sub-set of signals was selected for which the PTR-MS method has been well-validated in the past.

Comment: P34, L16-P35, L6: could refer to Simpson et al 2011, ARCTAS BB paper in here.

The reference has been added.

Comment: P36, L3: Might want to mention here that Mason et al (2001 and 2006) used the NCAR Master Mechanism at least twice to investigate BB plumes previously. Daniel Jacob modeled boreal fire emissions in ALE and Alvarado did in ARCTAS. Stu McKeen and Mike Trainer also modeled boreal forest fire plumes in JGR and Science, respectively. Yokelson et al., (1999) showed simple demo of including HCHO in a BB plume model.

A reference to the two Mason papers has been added, as they also made use of the Master Mechanism box model to investigate oxygenated NMOCs in biomass burning plumes.

Comment: P36-P38, section 3.1, Figs 1 and 2: As the authors themselves seem to point out, it's potentially misleading to compare absolute amounts between instruments that sample at different times in a structured environment. The text seems to imply that the largest differences are due to timing issues, so does that obscure detection of the actual instrumental differences present? There is a simple, standard method to compare instruments in a heterogeneous environment that works well though. For continuous instruments, one compares the integrated excess amounts across the whole plume (as seen by each instrument) to compensate for response time differences as explained most recently by Karl et al., 2007 or Yokelson et al., 2009. For intermittent instruments there are two cases. If the plume transit time is much longer than the gap between samples and the concentrations are varying slowly one again computes the integrated excess amounts and compares those. If the plume width and sampling intervals are comparable one typically compares dX/dCO from each instrument where the CO is carefully matched temporally to the measurement interval for X. For instance the excess TOGA acetone divided by the excess TOGA-merge CO should be compared to the excess WAS acetone divided by the excess WAS-merge CO (or better yet, the WAS CO and TOGA CO if they are available). This can be tedious, but it yields a meaningful comparison. An idea: something like $dAcetone/dMeOH$ (TOGA) versus $dAcetone/dMeOH$ (WAS) vs $dAcetone/dMeOH$ (PTR) etc might help and be quicker? It might be useful to try this for a few of the outliers and see if it affects the results?

The comparison shown in the paper is not intended to be a rigorous intercomparison of the techniques. Such an undertaking would in itself be a full paper. While we understand that this is not the most foolproof of methodologies, the brief comparison we present is meant to highlight where the measurement techniques agree well, and where issues may exist. However, we do agree that for the most accurate comparisons between the NEMRs from each instrument (Table 2), the NEMRs for the WAS measurements should be calculated against CO averaged to the same timescale of their measurements. Thus, all WAS-data NEMRs have been recalculated using the UCI-merge, and all discussions, tables, and figures referencing these numbers have been amended accordingly.

Comment: P36: Is there a reference for exactly how the merges are done? Is the data carefully lined up with response times in mind or just "overlapped" Is it better to compare dX/dY for the PTR-MS, TOGA, or WAS in some of the ways I propose just above?

There is a reference that describes how the NASA merges are done for INTEX-B/MILAGRO: Kleb et al., 2011. It has been added. There are criteria for comparing measurements, and for a case such as the TOGA and WAS data, "for paired measurements with different integration time intervals, the shorter integration time measurements were merged into the longer time interval when measurements at the shorter time interval overlapped at least 50% of the longer time interval." Accordingly, the text of this paper has been adjusted to avoid implying that a few second overlap could create a comparison issue.

Comment: P36, L17-18: Are these comparing all data or just data in BB plumes?

The comparisons include all measured data, and the only exception to this, isoprene + furan, is mentioned specifically. However, the first sentence of the paragraph has been changed to make this clearer.

Comment: P36, L26: If this is the final best result to partition furan and isoprene in boreal BB plumes the authors could compare to the isoprene plus furan and the split found to that in Karl et al 2007 and Christian et al 2003 (JGR).

We have added calculated ratios for isoprene, and for furan and “other alkenes” based on comparisons of the BB data for the PTR-MS isoprene + furan and have compared them to the signal split reported by Karl et al., 2007.

Comment: P38: Again, because of these timing issues it's not that meaningful to compare absolute amounts, as explained above.

See comments above regarding the comparisons between instruments.

Comment: P38, L22-24: So in the end all the instruments are assigned equal accuracy – sounds fair enough for now.

That is the authors' intent.

Comment: P39, L19: It's a good idea to eliminate mixed plumes from an analysis of fire emissions and the authors likely did a good job of this, but I'm not sure NO_x is an anthropogenic tracer? BB produces a large fraction of global NO_x.

Agreed, however the NO_x and toluene concentrations that were used as a red flag for anthropogenic impact far exceeded normal expected BB levels and ratios, which is why we used the term “strongly-enhanced.” When in doubt, the data were omitted. For the most part, this involved eliminating observations of the mixture of smoke and urban emissions that filled the LA basin during June 22-26, and was thus very difficult to resolve the BB.

Comment: P39, L23-24: A lot of chemical evolution can happen in 2 hours see Reid et al 1998; Goode et al., 2000; Yokelson et al 2003, 9; Hobbs et al 2003; Alvarado et al 2010; Akagi et al 2011b. Also, note as a general aging issue that 2 hours flying down the long axis of a plume can represent almost 1000 km of plume length. At a windspeed of 10 km/hour this would represent the fire output over a time span of 100 hours or ~ 4 days.

The authors are aware of this issue, and wish to confirm that individually-studied BB plumes were identified based on similar chemical ratios, with respect to location. One of the longer plumes that was close to two hours in length involved several short spirals and circles over a small region very close to a single fire. By definition, a plume that involved significant chemical changes would have been separated into at least two separate ‘plumes’ to quantify the differences.

Comment: P39, L25: As explained above, there will be considerable uncertainty in the plume age due to missing hotspots or incorrect guesses at the injection altitude, except in the case of the McKay Lake fire where the source was actually located and sampled. Note, that fire exhibited high source variability (Alvarado et al., 2010).

We have added text to Sect. 3.2 explaining that the estimated ages are merely approximations that are meant to differentiate between relatively fresh emissions, those that are 2-3 days in age, and those that

are aged up to a week or more. We suggest an uncertainty in the estimated plume ages of approximately +/- 50%.

Comment: P39, L26-27: the word "similar" is vague. Past studies tended to look out the window and see what was burning and classify the fires by vegetation type. In this study what are the new classification parameters and the category limits? Is there a connection between region and land cover type? Also, what is the logic for the grouping? 1) What are they trying to achieve by groups? 2) How does the grouping strategy accomplish the goal? Slightly more detail on the grouping strategy is given in the conclusions, but this is good place to clarify what is meant by "similar" or "chemical composition."

The smoke interceptions were grouped together based primarily on source region, and then, as explained above, by having similar ratios of NMOC to CO. Plumes in which the ratios of acetonitrile/CO or HCN/CO or even acetone/CO changed significantly were separated into two or more plumes to investigate separately. We have modified the text to make this clearer.

Comment: P40, L3: I don't think this study measures "fire-averages" I think the term here would be "group average" or "regional average"?

The text has been changed to read "plume emission average", and the authors have also changed the terminology from "plume group" to "plume".

Comment: P40, L9: Table 1 has numerous MCE > 1, which is problematic. As explained above, a simple calculation shows that you can dilute a plume with background air and dX/dY is preserved as long as the background air is constant in X and Y. Once X/Y in the background air changes, dX/dY no longer reflects the original characteristics of the plume. For diluted aged plumes, the effect can be huge. Since MCE cannot be > 1 by definition at the source of a fire, and the authors define plumes as having positive dCO , an MCE > 1 means the dCO_2 is a negative number smaller in absolute value than dCO . Since CO_2 is the main emission of fires the negative dCO_2 can only arise as an artifact due to changes in local background CO_2 during transport and thus CO_2/CO in the background likely changed during transport. That means X/CO likely changed as well where X is any random analyte. The original fire ratios have probably been corrupted in the NEMR from these plumes, but they are still valuable measurements of how BB affects the atmosphere. Thus, the results are worth presenting, but it might be best to eliminate them from any derivation of the attributes of fires.

Unfortunately this is very true. For some of the plumes, the plots of CO vs. CO_2 are wonderfully anti-correlated, and the reason for this is not clear. For others, the plots simply do not correlate at all. For the aged plumes, we agree that there is possibly a mixing of CO from other air masses, which could definitely disrupt the NEMRs determined from the plumes, but the CO and CO_2 measurements are other possibilities for uncertainty. For all plumes with MCEs > 1, with the exception of plume 28, the MCEs were removed from Table 1. For plume 28, although it is anti-correlated, it is still equivalent to unity within the uncertainty of the regression, so its MCE was included in the table. However, because of the overall suspicious nature of plume 28, it has been left out of the remaining calculations and discussions.

As for the NEMRs from the plumes with MCEs that are inexplicably > 1, we are reluctant to eliminate them from further fire attributions based on the fact that the measurements that are used for the MCE calculation are not infallible. As well, for some of the plumes in question, the proximities to the sampled BB were fairly small, and the resulting estimated ages are fairly low, so we believe that dilution is not the likely culprit. Finally, by eliminating data with correlations having an $r^2 < 0.5$, which also eliminated

any negative slopes, and reporting uncertainties based on the linear fits, we have confidence in our reported NEMRs.

Comment: P40, L17: "long-lived" better term than "useful"?

It is better, and has been changed.

Comment: P40, L19: This was confusing because I thought the authors were using EMR to stand for "excess mixing ratio" but I guess its "normalized excess mixing ratio." They are not the same thing. P40, L21 and L24: maybe add an "N" to "EMR" throughout as this is confusing otherwise and excess mixing ratios don't compare to ER. P40, L27: Again at this point in a first reading I get confused as it seems you are now changing the definition of EMR? Or add the word "normalized" and an N to acronym. One needs to read carefully to get this straight. I recommend reading Akagi et al., 2011a (ACP) as a quick course in fire terminology (and plume aging).

The misused acronym has been amended throughout.

Comment: P40, L26: This is one of several places where I wonder why the OH measurements were not used.

The OH measurements tell us what the OH is at the measurement location, but not what the OH has been for the last hour, day, or week.

Comment: P41, L2: 0.2 days as in 4.8 hours is too long for a NEMR to be an ER for some species (see the OH on the ARCTAS archive and think about the rate constants for e.g. isoprene. See the decay plots in Hobbs et al., 2003 (JGR).

Agreed – the sentence in question has been omitted.

Comment: P41, L15-19: Initially one wonders what the selection criteria and purpose are for Figure 6 since it is presented without discussion, but then it does come up again later. At some point it might be instructive to compare e.g. the Toluene data to the Toluene lifetime at the measured OH (2 days?) to illustrate how the variability masks the photochemistry? It could be made slightly clearer how the colors line up with plume groups 1-40.

The purpose for the plot in Fig. 6 (now Fig. 7) is presented at the beginning of the paragraph, and discussion regarding the grouping of the plumes listed in Table 2 into the colors shown in Fig. 7 has been added.

Comment: P45, L2: "there is"

This sentence has been modified for other reasons, and the missing word is no longer an issue.

Comment: P45, L27: Is production the right word here? It seems odd that the values which are lowest by far would be associated with production.

Yes. Although acetaldehyde is relatively short-lived in the atmosphere, there is secondary formation within plumes as they age, decreasing the observed relative loss rate.

Comment: P46, L2: there's the expected factor of 10, which is also not too badly approximated by the authors mean ± 2 quoted standard deviations. E.G. four standard deviations is "48" whereas the observed range is "49." P46, L7: Yokelson et al., (2009) report CH₃OH/CO (Table 2) ranging from 12.5 - 48.1 mmol/mol, where those are fire-average values. The study-average (not including one point from C-130) is 27.3 \pm 10.2 (n=17). The high variability is shown to be correlated with MCE in Figure 3 in agreement with general comments above.

A comparison of the observed methanol NEMRs to the observations from Yokelson et al. (2009) has been made in the Sect. 3.4 discussion regarding the oxygenated NMOC and MCEs.

Comment: P46, L9-25: Discussion of ethanol is interesting and useful and can be expanded to compare to Simpson et al.

The discussion has been expanded to include the values from Simpson et al. (2011).

Comment: P47, L3: Different OH may be the main factor, but may want to recognize a possible contribution from potentially different spring vs summer fuels and/or fire behavior in Asia.

This possibility is now included.

Comment: P48, L7: see Burling et al., (2011) (ACPD) on isoprene. I would eliminate "our group" and just cite Apel et al 2011, but add it to the references as well.

Done.

Comment: P48, L12-14: Here is a good opportunity for a comparative discussion regarding the findings of Simpson et al who concluded there was a characteristic ethane/CO ratio for boreal fires? It would be useful to clarify the different assumptions that lead to different conclusions?

Done.

Comment: P48, L16 and 21: On Plume 28. A fresh BB plume, or a combustion plume of any kind with MCE>1 is not physically realistic. Plume 28 is also an outlier for MEK. Plumes such as these should probably be weeded from the analysis. This could help make the paper shorter and less complex. Unless this outlier is telling us something important? Sometimes the outlier is the best point of all!

Plume 28 has been disregarded from the majority of the analysis because of its outlier status, but we have kept it in the overall plume list (Tables 1 and 2) because it still had many of the criteria for BB.

Comment: P50, L17: Another place where bringing in the OH could be useful.

Again, the observed OH is not equivalent to the average OH for the entire lifetime of the plume.

Comment: P51, L2: Good place to acknowledge SVOC by switching to NMOC, especially since the SVOC are more likely to become particulate organic matter.

Done.

Comment: P51, L6: It's not just believed, but long known that a single OVOC can be both primary and secondary, e.g. HCHO from combustion and from isoprene, CH4, etc, oxidation.

The text has been changed to eliminate the uncertainty of the statement.

Comment: P51, L9&11: An important example of OVOC (organic acids) evolution in plumes is shown/reviewed in Yokelson et al., (2009).

The example and reference have been added.

Comment: P51, L20: Why 2.5 days?

The model simulations were initiated in the middle of the day, and run to the end of the third day.

P51, L22-3: What are the corresponding plume group #s that were chosen as CALBB and CANBB and why were they chosen to represent time zero?

The CAL BB starting conditions were taken from a 10-minute period included in plume 18 on 26-Jun, from 15:53-16:02 UTC. The CAN BB starting conditions were from plume 25, which is also a 10-minute period. The two starting conditions were selected as they were both low-altitude observations that were high in BB tracers (CO, acetonitrile, HCN). Additionally, the CAL BB data were also relatively high in NO_x, and VOCs, O₃, PAN, etc. They were chosen for time "zero" not necessarily because they included minute-old BBs emissions, but because as a 10-minute average, they included what were identified as some of the youngest plumes that were observed during ARCTAS.

Comment: P51, L25: Compare model to measured for OH and HCHO?

The initial conditions were constrained by the observations as explained in the text, including OH and HCHO. Comparisons of the modeled and observed NEMRs of formaldehyde are included in Fig. 12. However, measurements were not made of OH over three days under these conditions, so a comparison is not possible.

Comment: P52, L15-17: Yokelson et al., 1999 (JGR) also demonstrated the effects of OVOC on plume chemistry and explained why OVOC are enhanced in BB plumes (the fuel is almost 50% oxygen as opposed to zero percent oxygen for FF). Yokelson et al., cite a Singh et al., (1995) Nature paper that gave a nice explanation of some mechanistic reasons why chemistry in plumes with OVOC has to be different.

Yokelson et al., 1999 was added as a reference.

Comment: P52, L18-20: Since the OVOC precursors are more abundant than the NHMC precursors it's a foregone conclusion that, except perhaps for HCHO (which has multiple precursors), that the individual OVOC will decrease as the plume ages. But how about the sum of OVOC? I'd expect the oxidation of OVOC to make other OVOC, which may not even have been measured in the plume? And it should be briefly acknowledged that this whole exercise ignores organic acids and other unmeasured species, that are important both as precursors and major secondary products and seen in BB plumes since Goode et al 2000 thru Akagi et al 2011b submitted. They were not measured in ARCTAS, but HCHO was, which is also

important (Simpson et al. 2011) and not explicitly mentioned here. Yokelson et al., (1999) showed big effects from including HCHO initial emissions in a very simple BB plume photochemical model.

The text has been altered to account for the compounds included and not included in the model. And formaldehyde, which was previously constrained in the model, is now discussed in the analysis.
Comment: P52, L21: Again HCHO is a product of isoprene, methane, methanol, etc oxidation. It may go up with aging as seen in a BB plume by Akagi et al., (2011b).

HCHO is now included in the discussion and Fig. 12, and does not increase with age in either model simulation because the loss is greater than the production.

Comment: P53, L1-2: If Figure 6 is retained (which I think it should be) here would a great place to indicate the expected decay of these species based on the measured OH for comparison to the scattered observations.

The measured OH does not tell us what the average OH in the plume has been.

Comment: P53, L11: My first guess would be that in real plumes the variability in NEMR would increase with aging due to different OH, light intensity, background mixing etc. Of course, the reactive species can all go very quickly to very small values that look similar on a plot with a large y-scale?

See response for the following comment below.

Comment: P53, L13-15: It seems surprising that the outcome is largely independent of the input and it may be a coincidence in that increasing NO_x increases O₃ to a point and then further NO_x increases "poison" the O₃ chemistry as discussed in Mason et al., 2001. Are the author's model results consistent with the message of the well known, traditional NO_x/NMOC O₃ isopleths?

Agreed. Upon greater analysis, there are some key differences in the rate of NMOC losses, and in the production of ozone within the two plumes. The analysis of the model output could be expanded significantly, but because the emphasis for this paper is on the NMOC evolution, to maintain this focus the discussion has been limited to the NMOCs.

Comment: P53, L13-20: It would be helpful to summarize what the message is here. It's probably not that it doesn't matter what emissions are used in atmospheric models, but it's hard to be sure as written.

The text in question has been amended to summarize and simplify the message.

Comment: Section 3.5: This is very promising. The ratios of NMOC to each other as a function of plume age may take out some of the variability from the analysis and could be explored further as a way to extract important new findings or enhanced comparisons to the model. E.G. it is known that the alkanes are mostly made by glowing, many of the OVOC are mostly made by pyrolysis, etc (Yokelson et al 1996, 97). Thus the NEMRs between NMOC that are similar to each other in structure should be less scattered in many cases than the NEMRs of NMOC to CO, but nonetheless, the NMOC within a "similar origin group" can have different OH rate constants and so the evolution of their ratios may be a useful window into the aging!

The discussion regarding the ratios of alkane NEMRs in Sect. 3.5, now Sect. 3.4, has been expanded significantly. As well, a similar analysis of ratios of NEMRs to CO for aromatic NMHC and oxygenated NMOC has been included, including the relationship of oxygenated NMOC emissions to the observed MCE.

Comment: P55, L14-16: This is an important result of the paper, the measurements of rarely measured species.

Agreed, and for this reason, the result has been added to the abstract.

Comment: P55, L19-20: Actually literature ranges in NEMR are well over an order of magnitude. For instance, Yokelson et al., (1999) cite a range of 600 for acetic acid EFs. I agree with the authors that there surely are some papers in the literature that underestimate variability, but they should be identified specifically rather than referring to the "literature" as a whole.

The authors agree that the primary message in the conclusions is not necessarily clear, and it has been overhauled to convey a more direct message regarding the findings of this work.

Comment: P55, L21-25: I think I disagree with the message here, but the overall message of the conclusions is vague. Here the authors seem to claim that fires are more variable than recognized in the literature (inaccurately I think), but in lines 13-14, the "good agreement" with the literature is emphasized. The authors should decide if they made quantitative progress in determining the seasonal, spatial, temporal, chemical, etc, resolution needed to adequately represent BB emissions, and if so give the results in specific, quantitative language. E.g. see Korontzi et al., (2003) or van Leeuwen et al., (2011).

As per the previous comment response, this text has been changed.

Comment: P55, L23: A paper with this "perception" should perhaps be quoted. Also a region and a fuel type are not similar concepts.

This text has been changed.

Comment: P55, L25-6: "confirmed" more accurate than "shown." Regarding the next sentence: using the mean can't over or under estimate the mean, unless the mean is wrong. But of course, in any particular plume the "mean" may not be observed. So this is like the weather vs climate.

This text has also been changed.

Comment: P56, L3-8: "The Californian BB plume was significantly more polluted than the Canadian BB plume, with 10x greater NOx, twice the O3 concentrations, and 4greater VOC and CO concentrations." Are the concentrations referred to here the background, excess, or absolute in plume? Unless the differences are in the background it's hard to believe they make no difference in the outcome. The implications of that would be enormous.

The differences are in the background. This has been clarified in the text.

Comment: Fig 7: maybe expand the caption or refer to discussion in text?

The caption was expanded.

Comment: Fig 8: Change "VOC" to "Measured NMOC" to acknowledge that not everything was (or can be) measured. (Akagi et al. 2011a; Warneke et al., 2011; Christian et al., 2003; Karl et al., 2007).

Done.