

Response to reviewer #1 for: An Improved, Automated Whole-Air Sampler and Gas Chromatography Mass Spectrometry Analysis System for Volatile Organic Compounds in the Atmosphere

Brian M. Lerner^{1,2}, Jessica B. Gilman², Kenneth C. Aikin^{1,2}, Elliot L. Atlas³, Paul D. Goldan^{2, †}, Martin Graus⁴, R. Hendershot⁵, Gabriel A. Isaacman-VanWertz^{6,7}, Abigail Koss^{1,2}, William C. Kuster^{1,2, †}, Richard A. Lueb⁵, Richard J. McLaughlin^{1,2}, Jeff Peischl^{1,2}, Donna Sueper⁸, Thomas B. Ryerson², Travis W. Tokarek⁹, Carsten Warneke^{1,2}, Bin Yuan^{1,2}, Joost A. de Gouw²

¹Cooperative Institute for Research in Environmental Sciences, Boulder, CO, USA

²NOAA Earth System Research Laboratory, Chemical Sciences Division, Boulder, CO, USA

³University of Miami, Rosenstiel School of Marine and Atmospheric Science, Miami, FL, USA

⁴Universität Innsbruck, Institut für Atmosphären- und Kryosphärenwissenschaften, Innsbruck, Austria

⁵National Center for Atmospheric Research, Division of Atmospheric Chemistry, Boulder, CO, USA

⁶University of California at Berkeley, Department of Environmental Science, Policy, and Management, Berkeley, CA, USA

⁷Now at Massachusetts Institute of Technology, Departments of Civil and Environmental Engineering and Chemical Engineering, Cambridge, MA, USA

⁸Aerodyne Research, Inc., Billerica, MA

⁹University of Calgary, Department of Chemistry, Calgary, AB, Canada

[†]now retired

Reviewer #1

The paper “An Improved, Automated Whole-Air Sampler . . .” by Brian M. Lerner et al. is a generally well written and structured paper presenting a new, cryogen-free GCMS system capable of analyzing a large number of VOC compounds in fairly short analysis time and at low detection limits. It has demonstrated successful operation during the 2 flight campaigns SENEX (2013) and SONGNEX (2015) with impressive number of samples analyzed, and comparison results to other VOC online-techniques are presented.

The sampling and analytical systems are well described. Characterization experiments are presented and the DL and precision of the GC-MS system is determined as well as recovery efficiencies and humidification and storage effects of the canisters. The comparisons to PTRMS, H3O+-CIMS, and optical ethane detector are appreciated yielding an impressive data set for characterization of the various techniques. The quality of the VOC measurements by the new instrument is hard to assess for the reader but should generally be in the 15% + x range for high concentrations, and worse for OVOC, heavier aromatics, and at lower concentrations. This is not the state of the art with respect to high quality VOC measurements (see recent intercomparisons and GAW Data Quality Objectives (Plass-Duelmer et al., 2006; Rappenglueck et al., 2006, Hoerger et al., 2015), however, the here presented technique has the focus on short analysis time, many different VOC, and suitability for continental NH conditions. It

provides very experienced and excellent concepts and corresponding compromises with respect to data quality must be accepted.

The paper is highly recommended for publication, however, there are some general and specific comments to be considered for revision.

General comments

1. The authors do not, but should present an uncertainty budget covering all uncertainty contributions. This should separately describe the precision contributions and calibration issues of the GC-MS system, and then should set-up an uncertainty budget for the entire measurements covering inlet, canister issues, calibration and GC-MS for the various compounds. This needs to consider blank and bias issues (absolute contributions) as well as precision (reproducibility, relative contributions) related issues. Table 2 currently only lists partial uncertainty contributions and the unexperienced reader might take those numbers as the overall uncertainty. Quantitative estimates are possible based on the presented material, and in the end they characterize the capability of the new system. Furthermore, these are needed to quantitatively interpret the comparison results presented in section 4.

We agree with reviewer #1 that explicit uncertainties are required for the GC-MS separate from the entire canister sampling and analysis system. We have separated uncertainties in Table 2 into two tables (2 and 3). There is considerable discussion in specific comments below detailing changes we have made to our presentation of uncertainties.

2. This paper lacks some more thorough discussion versus existing “knowledge”. Comparisons with other canister conditioning procedures (e.g. Irvine group, GAW) and canister issues (e.g. GAW and AMOHA intercomparisons), airborne intake systems (e.g. FAAM), calibration procedures (e.g. GAW), etc. Though generally reliable, the choice of methods and materials often appears somewhat arbitrary and could be better justified by/discussed versus corresponding references.

We recognize reviewer #1’s point of a limited discussion of existing methods in canister procedures and have added several references in the discussion of canister preparation (see specific comment below).

3. Canister issues with OVOCs (alcohols, ketones. . .) and light alkenes are well known and should be made clearer in the paper. The assumption that OVOC issues are due to the humidification step is part of the story but also, adsorption/desorption artefacts on stainless steel surfaces and in water traps are known to occur. Thus, compounds showing artifacts should be more rigorously identified.

We have re-worked our description of total uncertainty from our canister analysis system, which demonstrates considerable artifact response from oxygenated VOCs. This is now made explicit in the text.

4. It is argued that the measurements of section 4 were not taken at the same time and differences may be due to the variability of VOC compounds. However, the online instruments operated at higher time resolutions than the sampling of canisters, and the scatter in those data could be used to identify the variability. The authors should discuss the deviations considering this information.

The authors agree that the intercomparison between the fast time-response on-line instruments and the WAS measurements shows significant scatter, perhaps larger than would be expected. This scatter is not limited to the data presented here, but has been demonstrated with other whole-air sampler system intercomparisons, both during aircraft campaigns, when time alignment would be expected to be less burdensome, and even during ground-based comparisons with *in situ* GC-MS systems (de Gouw et al., 2003; de Gouw et al., 2006; Plass-Dülmer et al., 2006; Apel et al., 2003; Pollmann et al., 2008; Hornbrook et al., 2011). The authors believe the scatter in the data presented here compares favorably with the previously published work, and therefore the discussion has scientific merit. We have added some discussion of this previous work in Section 4 to put our results into this context. See the response to Specific Comments below for further details.

Specific Comments

p. 2, l. 8: *Please, consider also citation of the recent GAW paper by Schultz et al., 2016.*

This citation has been added to the first sentence of this section.

p. 5: *Sampling is via a stainless steel hose, metal bellows pumps and SS manifold. Though the flows are high and corresponding fill times are short (3-15 s), there is sufficient contact time to non-electropolished and badly flushed (in hose) SS surfaces. Sampling artefacts, especially of OVOCs, are known to occur and the authors should verify how these were tested and under which level they are. (It is well recognized, that in the subsequent analysis Sulfinert-treated SS is used)*

The canister sampling manifold is constructed from electropolished stainless steel tubing, and this has been added to the text: “The compressor output is connected in series to six canister modules via welded manifolds of 4.6 mm ID electropolished stainless steel tubing with breakable connections . . .” The tests of the sample inlet surfaces for artifacts are inherent to the canister tests described in Section 3.4, as all of the canister fills require the use of the sampling path described. Sampling a humidified zero gas (ultra-zero air or ultra-high purity nitrogen) is not currently possible for us at the sample flow rates (> 30 slpm) typical for our system.

Only the surfaces of the connection hose and the bellows pump are not electropolished stainless steel, which are continuously flushed throughout the entire flight at a very high flow rate (>30 slpm). Additionally, this portion of the sampling system represents a small surface area and very short residence time compare to the canisters themselves. We have not been able to test every singular component of the iWAS system and have chosen to focus on characterizing the system as a whole with particular emphasis on the canister aging/humidity effects

p. 6, l. 24-27: *The impression is raised that only 6 seconds flushing time are used prior to trap loading to passivate surfaces, which is misleading. Thus, it is recommended to point out that an additional flushing time of 210 s is used to achieve equilibrated surfaces.*

An additional sentence has been added at the end of this paragraph to clarify the total flush time: “If the manifold pressure is greater than 30 psia, the system will continue to flush the manifold and GC sample inlet for 35 seconds before sample acquisition is initiated.”

p. 7, l. 23: Please provide Valco valve details (Rotor material, channel diameter) and operation temperature of valves and lines.

We have added the additional text to this paragraph: "All chromatography valves have stainless steel bodies with polyaryletherketone / PTFE rotors with 0.40mm diameter channel, without external purging. The valves and transfer lines (Restek, 1/16" OD Sulfinert-treated stainless steel) are housed within an oven that is held at a constant 80°C."

p. 11, l. 11: Please, provide title and Journal of reference Isaacmann-Van Weertz et al. (2016), if not published yet, please provide the main concepts.

See next response below.

p.11, 2nd §: Could the authors provide information on the quality of automated peak integration relative to hand-integration? Ideally, the results obtained by applying both methods on the same data should be compared with respect to scatter in correlation plots and estimated integration uncertainties. Part of the differences in section 4 could be due to the peak integration, this should be discussed.

We have added a section to the Supplemental Information that describes the integration software in more detail than the main text and presents an intercomparison between data produced via manually drawn baselines and automatic peak fitting.

p. 11, section 2.23: Could you please refer to other canister conditioning procedures, e.g. GAW Guidelines on Canister sampling (GAW Report 204), and discuss deviations from other procedures.

We have added references to GAW Report 204, as well as McClenny et al. (1991), Colman et al. (2001) and Plass-Dülmer et al. (2006) in section 2.23. We have significantly revised and expanded this section to clarify our cleaning procedure and to compare with descriptions published in the cited references (see next two responses below).

p. 11, l. 27-29: Which type and purity of water is used, is the VOC content analyzed, and how is the water brought into the canisters?

We use Sigma-Aldrich HPLC-grade water for both the canister cleaning and analysis systems, and this is now noted in the text in this section, as well as in section 2.2.1.

The following additional text has been added to section 2.2.3 to further clarify the water addition process: "Water for both nitrogen humidification and water vapor addition is HPLC-grade (Sigma-Aldrich), in a bubbler heated to nominally 35°C. After the final evacuation step of the cleaning cycle, all sample canisters are opened for 2.5 minutes in order to be filled to ~15 hPa with water vapor. This is done by shutting off the nitrogen flow to the water bubbler so that only the headspace over the water reservoir is available to fill the evacuated canisters with water vapor. Water is delivered to the canisters via stainless steel tubing. The final fill pressure varied by less than 5% for each campaign. The cleaning

and humidification procedure is based upon a survey of canister preparation methods presented in the literature (McClenny et al., 1991; Colman et al., 2001; Plass-Dülmer et al., 2006; WMO, 2012), albeit at a slightly lower bake-out temperature than the range cited (70-80°C). After the SONGNEX field campaign, the cleaning oven was rebuilt to operate at 75°C.”

We did not test the water used in the cleaning oven for VOC content directly, but have now started to do so. When operating in the field, we use the same Sigma-Aldrich water to fill the bubbler on the analysis rack and to humidify a nitrogen stream for calibration experiments. If either of these samples show an unacceptable level of VOC contamination we empty, clean, and refill all water reservoirs with a new batch of water that has been boiled and bubbled with nitrogen before use, whenever possible.

p. 11, ll. 29. . .: *Please make clear what the argument is: when pressurizing ambient air to 3450 hPa, water vapor will condense at relative humidity above some 30% in the canisters anyways, so why change the humidification process? Which problems are expected when there is a condensed water phase in the canister versus a layer of adsorbed water molecules on the inner canister-surface? Were issues with water soluble VOC observed when a condensed water phase can be expected?*

The following text was added to this section: “This was switched to using water vapor for the SONGNEX campaign to ensure consistency of total water content in the canisters between the field campaigns. The presence of condensed water in the sample canisters is expected to have a deleterious impact upon soluble oxygenated VOCs (e.g. alcohols) (Ochiai et al., 2002). Further discussion of water effects is presented in Section 3.4.4 below, and the ambient water mixing ratios of the collected samples from each field campaign is described in Supplemental Information.”

A new section was added to Supplemental Information, describing the observed water mixing ratio for samples collected during SENEX and SONGNEX, with particular emphasis of the differences between boundary layer samples.

p. 12, section 3.1: *There are various possible contributions to drifting sensitivity of VOC analysis, comprising (1) MS sensitivity drifts, (2) trapping efficiency changes, (3) mass discrimination and nonlinearity in MS, and (4) active sites on the Al₂O₃ column to partly destroy compounds, etc. The authors only consider (1) and (2). The chosen calibration method is reasonable as it follows drifts run by run, though it is not usual. Usual is the use of calibration gases and interpolation of the sensitivity changes. No justification for the choice of normalization and interpolation is provided, based on the data. Thus, it is left open whether this is really the best choice. Furthermore, no thorough uncertainty estimate for the applied calibration factors is provided for the various VOC. Please, provide more information on the quality of thus performed calibrations, e.g. by plotting quality charts with time series of the various normalization factors based on each halocarbon analyzed on each column (e.g. three NF for column 1). Further, the calibrations of hydrocarbons could be displayed the same way by plotting their respective ratios to the initial raw peak area. Thus, similar temporal changes are expected given the cause of sensitivity change is (1) or (2). Deviation and scatter should then be discussed with respect to uncertainty of individual calibrations. This could be provided in the Supplemental material and only the result should go into the paper.*

While there are various possible contributions to drifting sensitivity of VOC analysis, the use of halocarbons in both ambient air and in the secondary standard are the most efficient way of describing

and accounting for the largest source of sensitivity drift on the described GC-MS system. The other 57 hydrocarbons in the secondary standard provide a measure of all other sources of sensitivity drift within the system as each hydrocarbon has a unique trapping efficiency, solubility, affinity to the Al₂O₃ column, and the hydrocarbons are measured on a range of masses from m/z 26 to 120. This allows for additional corrections to be applied on a compound-by-compound basis for a large number of reported species if necessary, but that has rarely been needed and was not necessary for the datasets described here.

The secondary standard is used as a single point determination of the sensitivity changes for individual compounds throughout the field campaign and are continuously compared to long term measurements of this same tank on the GC-MS. The secondary standard is also used to tie back to the much more detailed multi-point calibration experiments described in Section 3.2 which are generally conducted immediately prior to and after a field campaign, as they are too time intensive to conduct during the field campaign. Our older GC-MS system (Goldan et al., 2004) has utilized the same normalization procedure and the resulting sensitivities for the vast majority of the species has not changed by more than 10% (well within the stated measurement uncertainties of the older GC-MS) over the last 10 years. Other groups have also used this normalization technique (Karbiwnyk et al., 2003). Both instruments are run in a way as to increase the stability/reproducibility of the measurements over the long-term with less regard to maximizing the overall sensitivity over more brief periods of time (i.e., during a singular field campaign).

To demonstrate the utility of our normalization method, we have added a pair of time series plots to Supplemental Information, showing the change in raw counts measured over time for both a halocarbon and a hydrocarbon measured in our secondary standard.

The following text was inserted into Section 3.1:

“An example time series for the raw instrument response for a halocarbon and hydrocarbon species in the secondary standard measured on each channel is provided in Supplemental Information (Figure S3).”

A new figure (S3) was added to Supplemental Material, with the following figure caption:

“Time series showing change in measured raw counts for multiple analysis of a secondary standard over 3.5 days. Panel a) shows changes in response for two species from channel #1, CF₂Cl₂ and propane; panel b) shows changes in response for two species from channel #2, CCl₄ and benzene.”

p. 13, section sensitivity: *Issues have been identified with the linearity of the GCMS, potentially the linearity of the water trap, and with commercial standards yielding differences of 3-18% which is partly substantially larger than the stated uncertainty of the standards. I guess the stated uncertainty is at 95% confidence, whereas the differences are as observed. Is the message here that the stated uncertainties are wrong or that the instrument yields up to 18% unexplained variation? The water trap equilibration should affect water soluble and insoluble compounds very differently, and the authors should provide corresponding information. The calculation method of the total uncertainty of the calibration method of 12% is not transparent. Furthermore, it is not clear at what confidence level this uncertainty is stated. It is amazing that the authors consider all compounds' calibrations equal uncertain though they should show very different behavior. It is recommended to provide calibration uncertainties for the individual compounds.*

We agree with reviewer #1 that the general discussion of uncertainty in the original manuscript is unclear. We have attempted to improve this by describing the individual components that make up the overall uncertainty for each reported species, as shown in the revised Table 2. More directly, the discussion of the large variance in PAMS standards is misleading, as we are inappropriately averaging observed uncertainties across species when these should be treated independently, as the reviewer recommends. Indeed, the large uncertainty observed in ethane in the PAMS standards led to an artificially high statement of uncertainty for other species. We have added a brief discussion of the variance of ethane observed in the PAMS standards, separate from other species in those standards.

“Secondary gas standards are exchanged and analyzed with the NOAA Global Monitoring Division (GMD) Halocarbon and other Trace Gases (HATS) group on an informal basis every one to two years to establish the veracity of the stated gas standard concentrations. This process led to the discovery of the misstated ethane mixing ratio in our current primary PAMS standard (14% higher than stated). Accounting for additional measurement errors of flows of the dynamic dilution system, 1% for analyte and 2% for dilution, we define the calibration accuracy as the uncertainties of concentration and flow added in quadrature. These values are listed in Table 2. We have left the larger uncertainty in ethane accuracy in our current description of the GC-MS performance, as we are continuing to evaluate ethane standards.”

We have added to the discussion of the non-linearity in Section 3.2 and provided an example of the method of calculating non-linear sensitivity. We have also added a quantitative description of non-linearity to Table 2.

The following text has been inserted into Section 3.2:

“When plotted with linear-log scaling, the behavior is sigmoidal, in that sensitivity is constant at low mixing ratios, then transitions to a higher sensitivity at high mixing ratios. The sensitivity can be well-described with an exponential function:

$$Sens = Sens_0 + A/e^{\left(\frac{nct}{\tau}\right)} \quad (3)$$

where *Sens* is the sensitivity at a given normalized count signal (*nct*), *Sens*₀ is the sensitivity at low mixing ratios and τ is the normalized count signal midway between low and high sensitivity. An example of solving for the non-linear sensitivity (for m,p-xylenes) is presented in Figure 5, where a seven point calibration curve is shown spanning 0.025 – 1.0 ppbv mixing ratio. The linear fit statistics for the data points indicate a very good fit (Fig. 5a), with a small uncertainty of the slope (<2%) and $r^2 = 0.999$. However, replotting the data on a log-log scale (Fig. 5b) shows that the fit does a poor job describing the data collected at the lowest mixing ratios. Solving for individual sensitivities at each calibration mixing ratio, simply by dividing normalized counts by mixing ratio and plotting versus normalized counts on a logarithmic shows the sigmoidal behavior described above (Fig 5c). This data is well-described with equation (3), and comparing calculated mixing ratios found with the linear and non-linear sensitivities (Fig 5d) shows that the non-linear sensitivity provides an excellent match across the entire dynamic range of the calibration. We also provide a measure for the non-linearity, using the ratio of $A:Sens_0$, for all species reported in Table 2; for the earliest eluting compounds no non-linearity is observed.”

There appears to be a trend in non-linearity with later eluting compounds within each class, although it is not clear what is driving this trend. Reviewer #1 suggests that water soluble species should be more susceptible to this effect, which may be true for alcohols based on the values in Table 2. Other oxygenated species show no more pronounced non-linearity than hydrocarbons.

Further work on alternative water traps (as mentioned in the text) has identified significant losses to PEEK surfaces in the water trap, as a redesigned water trap with PFA surfaces has performed with reduced non-linearity. Further assessment of the water trap is underway and beyond the scope of the work here, and will be presented in a future manuscript.

We agree with reviewer #1 that a more explicit description of the uncertainty of the calibration method, independent of the accuracy of the calibration standard, is required. We have added a measure of calibration method uncertainty for all species reported in Table 2, based upon the 1σ confidence interval of the calibration curve (e.g. see Figure 5c). This is defined as a percentage uncertainty as measured at the secondary standard mixing ratio (275 pptv).

P. 14, section 3.3: *The parameters described in this section appear to be the repeatability and detection limit for very specific conditions. This should not be mixed up with the precision as a measure of the random uncertainty of the analytical system (see comment to Fig. 5 below). Please provide the mixing ratios of the secondary standard measurements. The detection limit corresponds to conditions of standard dilution measurements of the analytical system and should be clearly separated from the detection limit in canister measurements, which is further enhanced by blank issues of the canisters for various compounds. What is missing here is a precision and detection limit for the entire measurements of canisters (see also general comment).*

The authors recognize that the distinction between analytic system limit of detection and canister system limit of detection is important, and we have revised the manuscript to distinguish between the two. Table 2 has now been divided into two tables (2 and 3). Table 2 now describes uncertainties relevant to the GC/MS system alone, and includes non-linearity (as described above), precision, calibration uncertainty and total uncertainty stated in relative + absolute terms. Table 3 now describes the uncertainties associated with the canister system, namely blank levels, replicates, retrieval efficiency and total uncertainty (relative + absolute). We are grateful to the reviewer for this comment, as the GC/MS can and has been used to collect data without the canister system, and the separation of uncertainty descriptions is very helpful.

The following text has been added to section 3.1 to describe the secondary standard in greater detail: “The mixing ratio of the secondary standard is 275 pptv for all hydrocarbons with 7% uncertainty for each compound. The secondary standard also contains the halocarbons used for normalization at the following mixing ratios with 6% uncertainty: $C_2F_2Cl_2 = 140$ pptv, $CFCl_3 = 26.2$ pptv, CFC-113 = 12.0, $CCl_4 = 87.5$ pptv.”

We have inserted an additional section 3.4.4 (Total uncertainty), describing the total canister uncertainty for the SONGNEX campaign:

“Using the information from the canister tests and the total analytical uncertainty reported in section 3.3, we can describe the total uncertainty for measurements reported from the iWAS/ACCBAR system for SONGNEX, reported as relative plus absolute uncertainty [% + pptv]. For each compound, total relative uncertainty for the canisters is the quadrature summed analytical relative uncertainty and percentage differences of the replicate and Rtv Eff. Absolute uncertainty is the quadrature sum of analytical DL and the standard deviation of the canister blanks. Relative uncertainties for the canister measurements reported here typically range between 10-20% for hydrocarbons and alkyl nitrates. During SONGNEX, most oxygenated species had significantly larger uncertainties due to uncertain retrieval efficiencies, with only methyl ethyl ketone and methacrolein having acceptable uncertainty levels. Further canister testing evaluations are required to determine if improved canister preparation and handling will allow for reduced uncertainties of oxygenated species.”

p. 15, l. 9-11 and Fig. 5: *Please provide the range of deviation in the form of +/- (x ppt + y%(mixing ratio)) for 95% of the data, because below 10 ppt the deviation is determined by an absolute deviation rather than a relative deviation. Fig.5 and the suggested evaluation of the scatter should be applied to all compounds analyzed on both channels. The deviations do very well characterize the precision and detection limit of the analytical system for real ambient air without sampling and canister issues. A rough inspection of Fig. 5 by eye makes clear, that detection limit and precision stated in Table 2 (1 ppt DL and 3% precision) appear not realistic. The authors are encouraged to provide a realistic estimate of the random uncertainty contributions and prove the consistency with the data.*

We agree with reviewer #1 here, and have made changes to our statements of uncertainty (see previous answer). We have added the following text to Section 3.3:

“Total analytic uncertainty for each reported species is the sum in quadrature of the uncertainties of the calibration standard and sensitivity function, and the precision. Total uncertainty must also account for absolute uncertainty at low mixing ratios from DL, so that our total uncertainty is stated as relative uncertainty (%) + absolute uncertainty (pptv). Total uncertainty for all species is listed in Table 2. Other than ethane (17%), the range of relative analytical uncertainties for species reported ranges between 6-11%, and is typically less than 10%.”

We have also added error bars to Figure 5 (now Figure 6) to better present the uncertainties for the measurement of n-hexane from each channel. We agree with reviewer #1 that the original statement of uncertainty was incorrect, and the new analysis does well to describe the data presented in Figure 6. We have updated the caption to Figure 6 to:

“Intercomparison of n-hexane measurements made by the two different channels of iWAS-ACCBAR system during SENEX 2013. Data are shown with error bars based upon total analytical uncertainty presented in Table 2.”

We certainly agree with reviewer #1 that an analysis like that presented in Figure 5 would be very informative for other species besides n-hexane. Unfortunately, that is the only species that we currently quantify on both channels of the GC-MS. We will strongly consider including additional species in future work, although this would be limited to a subset of C5-C6 saturated hydrocarbons.

P. 16, l. 22: *It is hard to understand why contamination of the canisters due to the humidification process should have been different for SONGNEX and SENEX. Both needed water which might have been contaminated. Please, provide proof why SENEX is not affected by blanks.*

Unfortunately, water is the easiest source of potential contamination in our system as it is virtually impossible to get VOC-free water in the relatively large quantities and in easily-acquired means that are necessary for field measurements and yet the water is critical to the canister system as detailed in the literature. The humidification system between the 2 field campaigns had different water sources (i.e., different batches and ages of the HPLC water) and a few modifications were introduced to the system in order to provide concentrated water vapor for the SONGNEX campaign. We don't have canister blank data from SENEX, so we have offered our best explanation. All of the OVOC data that we reported for SENEX showed good agreement with the PTR-MS data suggesting that contamination from the canister system was not an issue during this campaign. During the course of these studies and the system development, we have recognized the importance of ensuring that the canisters are as clean as we expect them to be and have added several new tests to our standard operating procedures that will be detailed in forthcoming manuscripts.

P. 17, I. 17-22: *Earlier studies have seen some problems with light alkenes in canisters, also Table 2 indicates 33% standard deviation in ethene retrieval, propene values are not reported. I'd recommend more caution in the statement with regard to light alkenes (also see Plass-Duelmer et al., JGR, 2006). The intercepts in Fig. 7 should be given with units. The oxygenated VOC show substantial deviations. Though it is understandable that the authors claim issues with the humidification system, they do not provide evidence that other reasons for these issues can be excluded. Thus, at the current state, the canister technique appears inappropriate for alcohols and many of the ketones. This should be clearly stated.*

We agree with the author that the canister system is currently inadequate for the quantification of alcohols and other oxygenated species. We will make this point more clear by adding the following text: Section 3.4.2: "Oxygenated species were not well-behaved with aldehydes and ketones showing large scatter, excluding methyl ethyl ketone and methacrolein."

Section 3.4.4: "During SONGNEX, most oxygenated species had significantly larger uncertainties due to uncertain retrieval efficiencies, with only methyl ethyl ketone and methacrolein having acceptable uncertainty levels. Further canister testing evaluations are required to determine if improved canister preparation and handling will allow for reduced uncertainties of oxygenated species."

P. 17, I. 26: *The statement "entire system" should be more careful as it does not comprise the step in-situ to canister.*

We agree with reviewer #1 that "entire system" is too broad a term. We have changed the text here from "entire system" to "canister analysis system."

Section 4 (see also general comment 1): *Though there is a generally good impression from the comparisons with fairly good correlations, the authors do not provide quantitative estimates of the consistency of the scatter plots with the uncertainties of the analyses. This would be of high importance in order to prove the understanding of the analytical system and the uncertainty budget. Figure 9 displays considerable scatter, especially at low concentrations, and as mentioned above, there should be a description of the uncertainty and scatter as combination of an absolute and a relative quantity. This is important for judging the VOC system capability especially at lower concentrations. Furthermore, it is hard to understand comparisons based on different scales, as demonstrated for ethane (p. 20, I. 8).*

Though it is surprising that nowadays ethane calibration still appears to be an issue, this opens the question of how well defined are the scales for all VOC compounds and accordingly the results of the presented measurements. The scientific community is aware of the importance of good calibration methods since various intercomparisons and there have been standardizations in quality assurance like in Global Atmosphere Watch. The authors are encouraged to comment on their calibration with respect to the standards set by, e.g. GAW.

The authors appreciate the discussion of an improved description of uncertainty in the canister measurements, and have provided a more appropriate and descriptive report of canister uncertainties in Table 3, as discussed above. The scatter plots shown in Figure 9 (now Figure 10) show good agreement within the stated relative uncertainties of the measurement. While we certainly aspire to the analytical standards of GAW (Schultz M.G., 2015), we recognize that our canister system does not currently meet those standards across all species reported. However, we believe that we do well to characterize our current uncertainties, and plan to improve the sampling system towards the goal of further reducing our uncertainties. In an effort to provide context for the reported

p. 20, 2nd §: *Different ratios may be due to different scales or systematic errors in the analytical instruments. As similar compounds are compared in Fig. 11, no differences in the analytical system are expected and this kind of study is appropriate to characterize the scale ratios and the scatter in the data, but less the capability of the analytical system.*

We included the inter-comparison of the ground-based, is-situ NOAA GC-MS system and the airborne iWAS/ACCBAR system for the 2 different air sheds in order to highlight the observed consistency between the 2 different instruments and sampling techniques operating on 2 different platforms over a period of 5 years. Each instrument was independently calibrated and used independent normalization factors, yet the ratios for the 4 different VOCs are in very good agreement.

Table 2: *As stated above, the table should contain absolute and relative information on the repeatability, specify ideal analytical system conditions (DL), list information on compounds from both columns, and discuss some of the unusual data in the text (ethane and propane show highest scatter of alkanes in Rtv Eff though they are expected to be most stable, several alkanes are significantly changing over time based on Replicate and standard deviations (e.g. propane 0.93 with s.d. of 0.01), ethene indicates large scatter in Rtv eff and what about other light alkenes, aldehydes should be strictly characterized as canister artifacts, how is a negative value for acetone understandable (Rtv Eff). . .). Another Table should sum the overall precision (random uncertainty contributions with specified confidence) and calibration uncertainty (systematic, 95 % confidence) for canister sampling and analysis by the presented system for all VOC compounds.*

We are grateful to reviewer #1 for this suggestion and have split Table 2 into Tables 2 and 3. See discussion above for further details on these changes.

Figure 10a: *Benzene appears to go down to zero for the CIMS instrument whereas for the iWAS not. Benzene at 0 ppt seems unrealistic for continental air over US and points to a bias problem of the CIMS. This also underlines the importance of considering the uncertainties of the compared methods.*

The authors agree that intercomparison of WAS airborne measurements with the *in situ* CIMS measurement introduces additional uncertainties that are not well-characterized. We suggest that is not limited to the discussion here, but has been previously published by our group and other authors (de Gouw et al., 2003; de Gouw et al., 2006; Hornbrook et al., 2011). We present the Figures 10 and 11 here to demonstrate not only the performance of these complementary techniques but also to allow the reader to evaluate the improving capability of the measurements. To that end, we have added the following text to provide better context to the instrument intercomparisons:

“The intercomparisons with *in situ* instruments presented here show significant scatter, especially when compared to recent WAS validation work (Apel et al., 2003; Hoerger et al., 2015). It should be recognized that the scatter in the data shown here is not unique, but is typical in other presentations of comparisons between *in situ* and WAS measurements aboard aircraft (de Gouw et al., 2006; Hornbrook et al., 2011), as well as ground-based comparisons with fast time-response and GC-MS systems (de Gouw et al., 2003; Plass-Dülmer et al., 2006; Pollmann et al., 2008).”

Figure 11: the symbol colors are hardly discernable (grey and light blue) and should be changed.

We agree with the reviewer that this figure is difficult to interpret with the original color scheme. We have changed the grey color for this figure (now Figure 12) to make it more legible.

References (from reviewer #1)

Hoerger, C. C., Claude, A., Plass-Duelmer, C., Reimann, S., Eckart, E., Steinbrecher, R., Aalto, J., Arduini, J., Bonnaire, N., Cape, J. N., Colomb, A., Connolly, R., Diskova, J., Dumitrescu, P., Ehlers, C., Gros, V., Hakola, H., Hill, M., Hopkins, J. R., Jäger, J., Junek, R., Kajos, M. K., Klemp, D., Leuchner, M., Lewis, A. C., Locoge, N., Maione, M., Martin, D., Michl, K., Nemitz, E., O’Doherty, S., Pérez Ballesta, P., Ruuskanen, T. M., Sauvage, S., Schmidbauer, N., Spain, T. G., Straube, E., Vana, M., Vollmer, M. K., Wegener, R., and Wenger, A.: ACTRIS non-methane hydrocarbon intercomparison experiment in Europe to support WMO GAW and EMEP observation networks, *Atmos. Meas. Tech.*, 8, 2715-2736, doi:10.5194/amt-8-2715-2015, 2015.

Schultz MG, Akimoto H, Bottenheim J, Buchmann B, Galbally IE, Gilge S, Helmig D, Koide H, Lewis AC, Novelli PC, Plass-Dülmer C, Ryerson TB, Steinbacher M, Steinbrecher R, Tarasova O, Tørseth K, Thouret V, Sabatier P, Zellweger C : The Global Atmosphere Watch reactive gases measurement network. *Elementa: Science of the Anthropocene* 3: 000067. doi: 10.12952/journal.elementa.000067

C. Plass-Dülmer, N. Schmidbauer, J. Slemr, F. Slemr, and H.D’Souza, European hydrocarbon intercomparison experiment AMOHA part 4: Canister sampling of ambient air, *J. Geophys. Res.*, 111, D04306, doi:10.1029/2005JD006351, 2006.

B. Rappenglück, E. Apel, M. Bauerfeind, J. Bottenheim, P. Brickell, P. Cavolka, J. Cech, L. Gatti, H. Hakola, J. Honzak, R. Junek, D. Martin, C. Noone, C. Plass-Dülmer, D. Travers, and D. Wang, The first VOC intercomparison exercise within the Global Atmosphere Watch (GAW), *Atmos. Environ.* 40, 7508-7527, 2006.

GAWReport 204, WMO, Geneva, <http://www.wmo.int/pages/prog/arep/gaw/documents/FINA>

References for response to reviewer #1

- Apel, E. C., Calvert, J. G., Gilpin, T. M., Fehsenfeld, F., and Lonneman, W. A.: Nonmethane Hydrocarbon Intercomparison Experiment (NOMHICE): Task 4, ambient air, *Journal of Geophysical Research: Atmospheres*, 108, n/a-n/a, doi: 10.1029/2002JD002936, 2003.
- Colman, J. J., Swanson, A. L., Meinardi, S., Sive, B. C., Blake, D. R., and Rowland, F. S.: Description of the analysis of a wide range of volatile organic compounds in whole air samples collected during PEM-Tropics A and B, *Anal. Chem.*, 73, 3723-3731, doi: 10.1021/ac010027g, 2001.
- de Gouw, J. A., Goldan, P. D., Warneke, C., Kuster, W. C., Roberts, J. M., Marchewka, M., Bertman, S. B., Pszenny, A. A. P., and Keene, W. C.: Validation of proton transfer reaction-mass spectrometry (PTR-MS) measurements of gas-phase organic compounds in the atmosphere during the New England Air Quality Study (NEAQS) in 2002, *Journal of Geophysical Research-Atmospheres*, 108, doi: 10.1029/2003jd003863, 2003.
- de Gouw, J. A., Warneke, C., Stohl, A., Wollny, A. G., Brock, C. A., Cooper, O. R., Holloway, J. S., Trainer, M., Fehsenfeld, F. C., Atlas, E. L., Donnelly, S. G., Stroud, V., and Lueb, A.: Volatile organic compounds composition of merged and aged forest fire plumes from Alaska and western Canada, *Journal of Geophysical Research: Atmospheres*, 111, n/a-n/a, doi: 10.1029/2005JD006175, 2006.
- Goldan, P. D., Kuster, W. C., Williams, E., Murphy, P. C., Fehsenfeld, F. C., and Meagher, J.: Nonmethane hydrocarbon and oxy hydrocarbon measurements during the 2002 New England Air Quality Study, *Journal of Geophysical Research-Atmospheres*, 109, doi: 10.1029/2003jd004455, 2004.
- Hoerger, C. C., Claude, A., Plass-Duelmer, C., Reimann, S., Eckart, E., Steinbrecher, R., Aalto, J., Arduini, J., Bonnaire, N., Cape, J. N., Colomb, A., Connolly, R., Diskova, J., Dumitrean, P., Ehlers, C., Gros, V., Hakola, H., Hill, M., Hopkins, J. R., Jäger, J., Junek, R., Kajos, M. K., Klemp, D., Leuchner, M., Lewis, A. C., Locoge, N., Maione, M., Martin, D., Michl, K., Nemitz, E., O'Doherty, S., Pérez Ballesta, P., Ruuskanen, T. M., Sauvage, S., Schmidbauer, N., Spain, T. G., Straube, E., Vana, M., Vollmer, M. K., Wegener, R., and Wenger, A.: ACTRIS non-methane hydrocarbon intercomparison experiment in Europe to support WMO GAW and EMEP observation networks, *Atmos. Meas. Tech.*, 8, 2715-2736, doi: 10.5194/amt-8-2715-2015, 2015.
- Hornbrook, R. S., Blake, D. R., Diskin, G. S., Fried, A., Fuelberg, H. E., Meinardi, S., Mikoviny, T., Richter, D., Sachse, G. W., Vay, S. A., Walega, J., Weibring, P., Weinheimer, A. J., Wiedinmyer, C., Wisthaler, A., Hills, A., Riemer, D. D., and Apel, E. C.: Observations of nonmethane organic compounds during ARCTAS − Part 1: Biomass burning emissions and plume enhancements, *Atmos. Chem. Phys.*, 11, 11103-11130, doi: 10.5194/acp-11-11103-2011, 2011.
- Karbiwnyk, C. M., Mills, C. S., Helmig, D., and Birks, J. W.: Use of Chlorofluorocarbons as Internal Standards for the Measurement of Atmospheric Non-Methane Volatile Organic Compounds Sampled onto Solid Adsorbent Cartridges, *Environ. Sci. Technol.*, 37, 1002-1007, doi: 10.1021/es025910q, 2003.
- McClenny, W. A., Pleil, J. D., Evans, G. F., Oliver, K. D., Holdren, M. W., and Winberry, W. T.: Canister-based methods for monitoring toxic VOCs in ambient air, *Journal of the Air & Waste Management Association*, 41, 1308-1318, 1991.
- Ochiai, N., Tsuji, A., Nakamura, N., Daishima, S., and Cardin, D. B.: Stabilities of 58 volatile organic compounds in fused-silica-lined and SUMMA polished canisters under various humidified conditions, *Journal of Environmental Monitoring*, 4, 879-889, doi: 10.1039/b209210g, 2002.

Plass-Dülmer, C., Schmidbauer, N., Slemr, J., Slemr, F., and D'Souza, H.: European hydrocarbon intercomparison experiment AMOHA part 4: Canister sampling of ambient air, *Journal of Geophysical Research: Atmospheres* (1984–2012), 111, doi: 10.1029/2005JD006351, 2006.

Pollmann, J., Helmig, D., Hueber, J., Plass-Dülmer, C., and Tans, P.: Sampling, storage, and analysis of C2–C7 non-methane hydrocarbons from the US National Oceanic and Atmospheric Administration Cooperative Air Sampling Network glass flasks, *Journal of Chromatography A*, 1188, 75-87, doi: <http://dx.doi.org/10.1016/j.chroma.2008.02.059>, 2008.

Schultz M.G., A. H., Bottenheim J., Buchmann B., Galbally I.E., Gilge S., Helmig D., Koide H., Lewis A.C., Novelli P.C., Plass-Dülmer C., Ryerson T.B., Steinbacher M., Steinbrecher R., Tarsova O., Tørseth K., Thouret V., Zellweger C.: The Global Atmosphere Watch reactive gases measurement network, *Elem Sci Anth*, 3:000067, doi: 10.12952/journal.elementa.000067, 2015.

WMO: GAW Report 204: Standard Operating Procedures (SOPs) for Air Sampling in Stainless Steel Canisters for Non-Methane Hydrocarbons Analysis, in: GAW, 204, World Meteorological Organization, Geneva, Switzerland, 24, 2012.