

Response to reviewer #2 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 #2

This a very worthwhile paper describing a new GC-MS system for efficient analysis of multiple VOC species from whole air samples. This long-standing methodology remains heavily relied on in atmospheric chemistry research and continual improvements in technology are good to see. One paper can't cover every aspect of the measurement science and there are still major gaps in knowledge on how to best calibrate, storage, dry and data process VOC samples. In large part of the paper uses best practices when it is not suggesting new or improved ways of working. The core of the improved capability here appears to be the Stirling cooler trapping system and a fast two-column separation – the improving ease of field operation and increasing sample throughput. With some additional information and clarifications that paper should be published.

The paper places most of the emphasis for external validation of the method against co-measurements made using PTR-MS. It is very understandable, since this instrument was available on the same campaigns. But when demonstrating a new GC system's measurement performance against a characterized reference then PTR doesn't really seem the best choice. Rather comparing to PTR seems to just add in a whole load of new uncertainties from that technique as well. Only a modest number of VOCs can be measured by both systems, and even then quite often it is a summation of compounds. GC-MS (and sometimes FID) remains the gold standard method, even if laborious and slow, so how does this new more efficient instrument compare against the previous WAS and GC method? Something in the paper on that would give a much better measure of 'progress'. Is this new system just faster, or is it faster, more sensitive and more precise? The authors have a huge track

record of measurements going back many years, so it would be genuinely helpful to know how this system compares against notionally similar GC principles used previously. Is there a trade off between speed and precision, when compared to off the shelf instruments?

We have not operated our system simultaneously with other WAS systems, so a direct comparison is not possible. Our goals for the new system were primarily for a faster system with reduced operator oversight and a minimal decrease in instrument performance, either sensitivity or precision. Additionally, we have sought to optimize the chromatography and detector behavior to improve automated peak fitting with TERN, rather than to improve sensitivity and/or precision. For example, wider overlapping peaks can often be trivially fit by TERN, so have become somewhat disincentivised from optimizing peak separation to baseline.

While we could provide normalized sensitivities for our previous GC-MS system to allow comparison, the sample volumes are not equivalent (350 vs 240 standard cm³ for the old and new systems, respectively). For example, a better column oven design with the new system allows us to heat the columns faster and more consistently than the old design allowing for better retention time reproducibility and better throughput of later eluting compounds (i.e., expanded our analytical capabilities in less chromatography time). Shorter transfer lines that are all heated should also improve chromatographic separation and overall sensitivity for the lower-vapor pressure species, but we have yet to do a direct comparison of the sensitivities of each instrument.

With the restatement of total uncertainties, separated into analytical system and canister measurement, we have been able to describe more realistic uncertainties for the GC-MS system, which are typically less than 10% for most species reported here. We believe this level of performance is acceptable for a field-deployed GC-MS system, although we certainly will continue to seek to improve the systems with further work.

It was surprising that for an instrument where the new thermal desorption developments are so central to that there is no demonstration of breakthrough volumes and trapping efficiency. The trapping volume is small and it relies primarily on cold brute force to achieve retention. Breakthrough volumes for say acetylene would be interesting. On P10 it was still a little unclear how much air was actually sampled on to each trap. 240 mL each or shared between two traps? There is some reference on P13 to changing trapping efficiencies suggesting that retention is not 100% for all species.

Sample volumes are 240 standard cm³, flowed at 60 sccm. High mixing ratio calibration experiments (up to nominal 180 ppbv mixing ratio) with C2-C4 hydrocarbons (ethane, ethene, propane, propene, ethyne, n-butane) showed no decrease in sensitivity with increasing mixing ratio throughout the calibration range, indicating no breakthrough. This calibration range was selected to cover the largest mixing ratios encountered during the SONGNEX 2015 campaign, with the largest single reported VOC mixing ratio of 124 ppb for propane. We cannot speak to mixing ratios greater than 180 ppbv or to increased sample volumes, as we currently have data for neither. We can perform these experiments in the future as warranted.

The following text has been added to Section 3.2 in the discussion of non-linearity:

“This behavior is the opposite of what is typically observed when analyte breakthrough occurs at the sample trap, and we have tested the instrument up to 180 ppbv with a mixture of light hydrocarbons most susceptible to breakthrough (ethane, ethene, propane, propene, ethyne, n-butane) with no observed decrease in sensitivity; this mixing ratio is larger than any we have observed in ambient air.”

The very low temperatures necessitate both a water removal step and also one for CO₂. Whilst the general methods used are reported in other literature it would be interesting to have a comment on whether any hysteresis is introduced by either. Any carry over for example on the coated beads of silica? Is there any residue if switching from a high ppb-level sample to a ppt sample. This is one of the big challenges for aircraft analysis systems since their samples can rapidly span a very wide dynamic range compared to surface-only sampling.

We regularly test for this behavior with system blanks, similar to the canister blanks described in Section 3.4.2, using humidified zero air or nitrogen between canister modules. We have found minimal evidence of carry-over from one sample to the next. We also test this during calibration experiments as well, where we calibrate at a high mixing ratio and then run without the calibration mixture (i.e., humidified N₂). The cleaning process of the sample and water traps and sampling manifold were designed to reduce any carry-over effects from sample to sample regardless of the mixing ratio. Finally, the in-situ comparison of iWAS vs. PTR-MS show no evidence of this occurring for the species that we have investigated.

Whilst the focus of the paper is naturally on the GC-MS and component parts it would be valuable to have some additional information on tests conducted on the sampling system itself. A critical and often ignored part of a WAS is the air compression system pump itself – without that there is no sample to analyse. Whilst I couldn't find any details on-line for on the 28823-11 Senior Aerospace pump, most variants of this kind do not have completely sealed bellows systems, instead having pin-hole pressure relief holes in the bellows to prevent motor stall at low inlet pressures. Is this the case here, or was this relief welded shut? A consequence of this pump design can be that a proportion of the air delivered to the canisters is from the surrounds of the pump, not the inlet tubing. What tests historically have been performed to show no contamination from this pump or losses – I appreciate this might already be published elsewhere in earlier work. It is an exceptionally difficult thing to do, but any insight that can be gained on testing of this kind would be valuable to other readers.

The following additional description of the compressor has been added to the text: “The 28823-11 compressor is a modified version of the “off-the-shelf” 28823-7 model available from Senior Aerospace. The modifications are a fully sealed stainless steel bellows (no pressure relief pin-holes) and the replacement of the pressure relief valve on the pump faceplate with a 1/8” NPT-tapped hole that is subsequently plugged with a stainless steel fitting. The modified stainless steel bellows is leak-tested with He to 1×10^{-6} cc sec⁻¹ by the manufacturer prior to assembly, and all wetted surfaces are cleaned with methanol/ethanol.”

We have not performed tests with the compressor beyond what is presented here. We have been unable to provide an adequate flow rate of humidified zero air or UHP nitrogen to the compressor under normal operating conditions (>35 slpm). We are currently running blank test with the system using greatly throttled flow rates of less than 5 slpm, but these tests result in fill times in excess of 1 minute per canister. We hope to publish the results of these tests in a future manuscript.

Reviewer 1 has already highlighted the need for an uncertainty budget and some further statistical details and I would also endorse these requests. Not only is this essential to understand how and where the data can subsequently be used, but it would provide a very helpful guide to where future instrumentation development is likely to yield the most beneficial gains.

See discussion of a new presentation of error analysis separating the analysis system from the canister system as described above in response to reviewer #1. Briefly, we have separated Table 2 into two tables: Table 2

describes uncertainties associated with the GC-MS system alone, while Table 3 describes additional uncertainties from the canisters and total uncertainty associated with WAS sampling and analysis.

P2. Line 17 – *it is other detection methods for GC that are less sensitive, not that alternative chromatographic methods are used for VOC analysis.*

Line 17 has been changed to “chromatograph detection methods.”

P3 line 5. *Lift capacity means needs defining here, not later on P8.*

Lift capacity was defined later in the original paragraph. This paragraph has been restructured so that lift capacity is clearly defined after its first usage.

P6 line 16. *What method, chemicals or technique is used to generate ultra zero air?*

A description of the commercial “ultra zero” air grade and vendor has been added to the text. We also describe the type of water used for humidification.

P7 line 22. *What sort of valves and materials, were these externally purged?*

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.”

P11 – *can something more quantitative be said about the precision of the integration from the new automated software compared to previous methods? The text refers to improved precision and accuracy, but doesn’t say anything else.*

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

P13. Line 13. *How are these reference gas dilutions achieved and do they include any reference materials that create traceability to other international standards? Can more information be given on how the 1 ppm PAMS 57 mix is diluted down to ambient levels, or is it used as is?*

We have expanded the description of our dilution system for calibrations. The following text has been inserted into Section 3.2 in two places:

“Dilutions are made using a dynamic dilution system, consisting of two high flow mass flow controllers to provide UHP N2 flow both through and bypassing a glass bubbler filled with HPLC water [Sigma Aldrich], and a low flow MFC or crimped stainless steel capillary tube to provide standard gas to the system. The low flow MFC

is used only for hydrocarbon standards (e.g. PAMS 57 component mix) as oxygenated gas standards through MFCs has resulted in the observation of unstable peak areas that are resolved by using crimped capillaries at controlled temperature. The high flow MFCs are controlled to provide total dilution flow between 0.5 – 10 slpm while allowing humidity to be varied between 0-80% RH. For system calibrations, the relative humidity is typically held between 35-40% for all data points. The analyte flow is controlled at multiple setpoints between 0.3 - 3 sccm with the low flow MFC, while crimped capillaries allow constant flow within the same range. Flows are measured using DryCal flow meters for dilution flows and a soap bubble flow meter for analyte flows. All tubing used in the dynamic dilution system is 6.2mm OD PFA, and a 10m loop of PFA tubing is placed between the dilution system and instrument to ensure adequate mixing before sampling.”

“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.”

P13 *Section 3.2 would be better titled calibration, since that is what the section is about and no sensitivities at actually presented. There is a discussion of a non-linear set of compound responses, and an undefined reference to linearity when mixing ratios are low compared high. Could this be sharpened up with some quantitative numbers for various indicative compounds?*

We have re-titled this section “Calibration” as suggested. We have also added an additional discussion of the non-linearity in the text, along with an example calculation of non-linear sensitivity. We have inserted a new figure (5) as well as added a measure of the non-linearity to Table 2.

The following text has been inserted in 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*), *Sens₀* 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.”

P14 *The total uncertainty in the calibration method is given as 12%, which is quite high relative to the DQOs set out for example by GAW. Any comments on this? I found it noteworthy, if a little depressing, that commercial standards for C2-C5 NMHC gave difference up to 18%. Can any manufacturers be named and shamed.*

We have restated the total analytical uncertainty, which we believe better describe the true uncertainties for the analysis system. See the discussion above in our response to reviewer #1.

The authors would prefer to not disclose the manufacturer of the standard. We will note that we have purchased multiple calibration standards previously from the vendor that have been demonstrated to be within their stated accuracy ranges. This simply highlight the need that calibrations avoid reliance upon a single tank or a single manufacturer whenever possible. In this case, diversity and cross calibrations were a critical aspect to discovering the discrepancy between standards for ethane.

P14 – *for limits of detection it would be very helpful to the reader, and for future comparison against other researchers instruments, if an example chromatograms with trace level peaks could be included as a figure – perhaps in the Supplementary materials.*

We are happy to add an additional figure (Figure S4) to the Supplemental Information section. We inserted the following sentence in Section 3.3: “An example chromatogram is shown in the Supplemental Information section (Figure S4) for a calibration measurement at 26 pptv mixing ratio of the 57-component PAMS mixture described in Section 3.2.”

We have included the following figure description in the Supplemental Information:

“Example TIC chromatogram of 57-component PAMS standard diluted to 26 pptv in humidified UHP N2, the lowest of seven calibration levels tested prior to the SONGNEX field campaign. Replicates of measurements of the lowest calibration point are used as a basis for calculating limits of detection. Note that the chromatogram scaling is 1/10 of the chromatogram shown in Figure 4.”

P15. *The detection and precisions are given in section 3.3 and Table 2, but it is difficult to square these values with for example the inter-column comparison data in Figure 5. The quoted precision uncertainty of 7% clearly isn't appropriate to apply to the lowest abundance samples and an absolute value for each compound needs quoting. Figure 7 also suggests the quoted precisions for the method look a little generous. Giving explicit uncertainties at the lowest concentrations is really important, since this type of VOC measurement often data ends up being (inappropriately) used in tracer-tracer ratios.*

Total uncertainties has been restated as a percentage + limit of detection value, as suggested by reviewer #1. As noted above, we have revamped our statements of uncertainty as suggested by both reviewers. We are appreciative of these suggestions, as well agree that it improves the description of the system.