Author response to RC1 of Karu, E., Li, M., Ernle, L., Brenninkmeijer, C. A. M., Lelieveld, J., and Williams, J.: Atomic emission detector with gas chromatographic separation and cryogenic preconcentration (CryoTrap-GC-AED) for atmospheric trace gas measurements.

Here present an in-depth analysis of a potentially useful new instrument for atmospheric research, namely the CryoTrap-GC-AED system. The instrument was purchased to replace an ageing FID system, the aim being to simplify calibrations with the equimolar response ideally offered by this technique and to use the element specific signals with low associated noise to search for new species of atmospheric interest. In this paper we report that in fact the species measurable by this instrument do exhibit compound specific responses due to the sampling preparation steps, and that no exotic heteroatom containing species were discovered in the samples analyzed. Although apparently disappointing it is nonetheless important to report these findings. The initial manuscript submitted included the instrument description and several examples of how selected data was used to derive new atmospheric information (for example on the upper atmosphere budget of OCS). At the request of the editor this was split into this instrumental paper (here) and a separate data analysis paper. We agree with the reviewer that since no new species (containing of iodine, bromine, nitrogen, silicon, mercury, selenium and arsenic) were revealed in any of the sampled environments to date that a GC-MS targeting the compounds in the calibration gas would have served the same purpose. Although it should be noted the GC-MS could exhibit a much greater range of sensitivities to individual compounds. For transparency we state this point now clearly in the discussion as "L249: For the work presented here, a pre-concentrating GC-MS could have served as the same or better alternative for monitoring targeted species present in a calibration gas. Nevertheless, the AED system has advantages for searching exotic heteroatom molecules in the atmosphere, for the non-carbon elements which are in the detector wavelength range.". We take the view that it is important to document the instrument performance "warts and all".

Compound specific response – all compounds measured in the atmosphere were calibrated against a multi-component gas mixture, thus all the losses were accounted for.

Added to the manuscript at L190.

Identify unknowns by determining composition of a mixed halocarbon compound – Unfortunately, chlorine and fluorine do not land in the quantifiable 161 – 211 nm wavelength range (AED III HR has a restricted wavelength range, but all the wavelengths can be recorded simultaneously. AED II has a mechanically turning grating for measuring different wavelength groups in the range of 171 to 837 nm, thus nevertheless the simultaneous measurement of the whole range is not possible).

Added to the manuscript at L44.

Throughout my PhD, I was always on a search for new compounds in the group of iodine, bromine, nitrogen, silicon, mercury, selenium and arsenic but these elements were mostly under the quantifiable detection limit with the method developed, and in the regions measured.

Added to the manuscript at L251.

Might have been done more easily with a GC/FID or GC/MS – We agree that for the presented case these instruments would have been a better option, but as we were also searching for new heteroatom containing species in the samples, where AED detector has an advantage.

Added to the manuscript at L249.

AED capability to measure oxygen – not suitable for trace atmospheric measurements as the requirement is to use 10% methane 90% nitrogen reagent gas mixture. As this reagent gas produces a very high background signal it was not feasible to use it for trace gas measurements which was the aim of this project. This AED capability is suitable for oxygen measurements in oil and gas industry.

Added to the manuscript at L245.

L33 Even though "great care" was taken to minimize compound specific effects, such effects are later reported. Here the authors claim that a single RF could be applied to an unknown compound, but they don't indicate how they choose this RF, or if they use some average.

Compound specific RFs could be applied for a similar compound groups (pseudo-unknown compounds) within the uncertainty range. Even better approach would be to average similar compound group RFs to quantify the pseudo-unknown or a compound missing in the calibration standard.

Added to the manuscript at L35.

L41 The authors show a range of 161 – 211 nm for the range of the JAS AEDIII HR instrument. It seems from the literature of JAS that higher wavelengths are also accessible. Earlier AED models did have a much wider range. Can the authors clarify if this limited range is all that is available? Improved responses for some elements might be found at higher wavelengths. If the instrument is limited in wavelength range, perhaps the authors could comment on the advantages or disadvantages of a limited range versus an extended range instrument.

With the AEDIII HR wavelength range of only 161 – 211 nm can be measured simultaneously (light is focused in a horizontal plane onto two back-thinned CCDs). Due to a physical hardware gap between the two CCDs, there is a 7 nm gap in the wavelength spectra between 183 – 190 nm range). Therefore, e.g. chlorine and fluorine cannot be measured with the AEDIII HR version. For example, AEDII (previously developed by HP -> Agilent -> JAS GmbH) was able to measure in the range of 171 to 837 nm with a turning grating as mentioned above, therefore not all wavelengths simultaneously. Unfortunately, chlorine (sensitive at 479 nm) and fluorine (sensitive at 690 nm) are not in the measurement range of AEDIII HR. The advantage of the limited wavelength range is the simultaneous recording of the whole spectra with 0.01 nm step resolution.

Added to the manuscript at L44 and L108.

The newer development by the JAS company AED III Wide Range (WR) measures much wider range with movable optics. Thus, simultaneous recording of the whole spectral range is once again not possible (different elements must be recorded in different wavelength range groups). Therefore, also the resolution is lower and likely also the sensitivity (they do not state the sensitivity yet).

https://www.jas.de/en/products/jas/aed/aed-iii-wr-wide-range-spectrometer/

L47 The authors state that the performance of atomic emission lines for multiple compounds is discussed. However, no information is provided if the authors characterized different potential atomic emission lines for each atom. Why were these particular lines chosen? Or if different ratios of combustion gases were tested to evaluate the impact of plasma conditions on the results.

The most sensitive atomic emission lines in this setup and in the range of this detector were chosen. Iodine (163 nm), sulfur (181 nm), carbon (193 nm), bromine (163 nm) and nitrogen (174 nm). No secondary atomic emission lines for each element were used. The optimal ratio of hydrogen at 12 psi (0.83 bar) and oxygen (15 psi (1.03 bar) supply pressure was used throughout the measurements.

Added to the manuscript at L49 and L101.

L56 Are the sample inlet lines temperature controlled? Is an internal standard used? Did you determine the maximum sample size possible before the most volatile measured compound breaks through the trap?

All the sample inlet lines entering the CryoTrap instrument were at constant lab temperature and same length inlet lines were used for calibration standards and the sample lines, thus any losses here would be accounted for. No internal standard was used for these measurements. As the liquid nitrogen based trapping at -60°C to the main Tenax trap retains the analytes well, no breakthrough of the most volatile measured compounds was noticed with the maximum sample size of 1 800 mL tested.

Added to the manuscript at L120.

L64 Could you specify the sample and other flows used during sample collection/transfer?

Nominal sample trapping flow of 200 mL min^{-1} was used and recorded. Helium flush flow rate was set to 100 mL min^{-1}. Trap module 1 to trap module 2 transfer flow of 10 mL min^{-1} was used.

Added to the manuscript at L69, L71 and L72.

L96 Perhaps also specify purity requirements for the H_2 and O_2.

For the reagent gases extra ultra-high purity H_2 (EUHP purity > 99.99999% by a Parker Balston Hydrogen Generator, model H2-300, Parker Hannifin Corporation, USA) and ultra-high purity O_2 (UHP, purity 99.9999%, Westfalen, Germany) were used.

Added to the manuscript at L57.

L114 Could you specify what backflush and bakeout times were required based on the carryover experiments? What level of carryover is observed from ppb level to zero level samples?

With the 15 min bakeout step after the injection the modules 1 and 2 are heated to 150°C and 210°C respectively and back-flushed with He flow (75 ml min^{-1}), whereas the bulkhead heaters are kept at 150°C. After this bakeout no carryover was measured.

Added to the manuscript at L122.

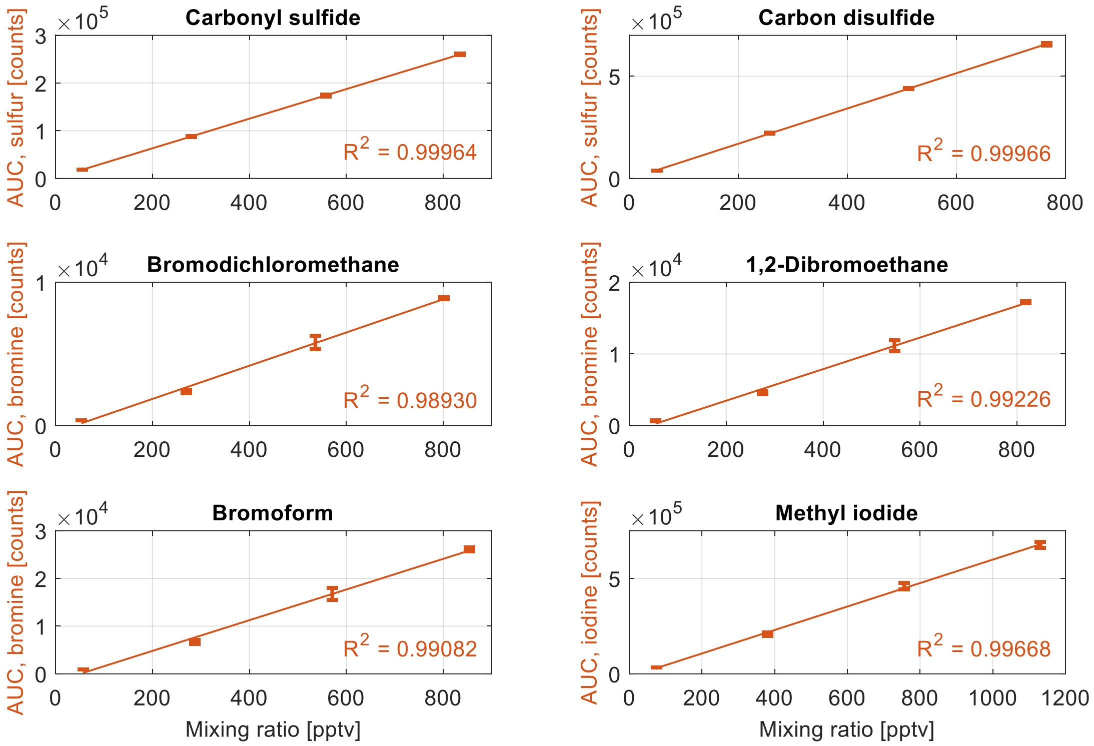
L116 Only two standards are described, but it says that three were used. Please clarify. Later a NOAA ambient air calibration standard was mentioned. Is this number 3? If so, how does this compare with dilutions of the Riemer or NPL standards?

Sorry for the typo. Two standards were used for characterization experiments. For the IAGOS-CARIBIC sample measurement procedure, the National Oceanic and Atmospheric Administration (NOAA) ambient air calibration standard was used for calibrating carbonyl sulfide to the NOAA-scale.

L131 Not really sure what linear up to 4 orders of magnitude means? Could you please provide a table of response factors determined for a range of different classes of compounds over the 4 orders of magnitude that were tested. Can you also do a similar table for the calibration levels used for the characterization experiments (e.g. 50 – 1000 pptv).

With this sentence it is meant that linear calibration can be used at least from 1 pptv to 10 ppbv. It was a conservative measurement range, probably the linear range is even larger.

An example calibration curve for 6 species below:



L137 I am curious how the instrument drifted under normal use. Is there a significant drift in system response over time?

The most significant drift comes from discharge tube wear, but this was accounted for with frequent calibration standard measurement. Between different discharge tubes and the discharge fitting quality the background signal varied significantly.

L144 The LOD determination can be done in a variety of ways. In a system that is free of adsorption or other artifacts, the equation 1 might be a reasonable extrapolation to determine detection limit. However, the system described does have some compound specific characteristics, so it would be helpful to verify the reported LODs with actual measurements near that LOD (e.g., within a factor of 3).

For the LOD calculations the lowest practically achievable calibration points were used. As the compounds of interest in the unknown samples were almost always in the real calibration range, thus should not have a large effect.

L152 It is not clear how the average RF that is calculated was actually used for quantitation.

The average RF was used for unknown analyte determination derived from a linear calibration graph (area under curve [counts] vs. MR [pptv]).

L154 Were other carbon emission lines tested? If so, what were the different characteristics? Section 3.1 This section describes results that begs further investigation and discussion. The range of reported response for carbon in different compounds is very troubling. What is reported is not the equimolar response that is claimed for the system, and it suggests big problems somewhere. A factor of >2 between benzyl chloride and butyl nitrate seems to me to be a big problem that needs to be evaluated further. To me, the differences suggest issues of stability or standard drift, though system artifacts can't be ruled out. The observed differences could be checked with independent standards, or at least compare calculations of mixing ratios based on both carbon and heteroatoms in the molecule. Further, the large variability within each response factor also seems to be a problem (36% for butyl nitrate, 43% for benzyl chloride). It is not clear to me how these variations are useful for quantitative analysis.

Overall, the response factor uncertainty (to 1 sigma) is quite variable, from reasonable few percent to much higher values depending on compounds. The authors also note that the standard deviations from the Apel-Riemer standard is about 2 x that of the NPL standard. The authors speculate that this may result from changing discharge tubes, though one could also speculate that the dilution system used for the Apel-Riemer tests had some stability problems. If the problem is from discharge tubes, wouldn't this suggest a significant problem for routine application?

Only the carbon emission line at 193 nm as the most sensitive atomic emission line in the range was tested. The big RF differences could arise from different loss and trapping efficiencies in the whole analytical pathway as well as plasma atomization efficiency. Furthermore, the extra step of diluting the nominal 50 ppbv Apel-Riemer-2015 standard is an extra point for higher uncertainty compared to the direct trapping of the nominally 4 ppbv NPL-2017 calibration standard. Also, the stability differences within the calibration standards bottles could also play a role.

Added to the manuscript at L189 and L198.

Paragraph beginning L178. This paragraph describes the crux of the problem with this paper and with the proposed method. The nominal advantage of the equimolar response of an AED is not found in this system, presumably because of sample preparation/separation issues. I don't think it lies with the AED, but this could be checked by additional testing. My suggestion would be to locate the source of the problems and determine what needs to be fixed to improve the accuracy and precisions of the whole instrument. As presented, the analytical system described has some basic flaws. It can, as the authors point out, be used if one calibrates each compound. But this is essentially no different from use of an FID, which is substantially less complicated. (I would be interested to know if the cryo-enrichment unit was tested using an FID detector, and how this might compare.) The average RF is precise to about 10% (only 1 SD), and this offers no advantage over other common analytical systems.

Correct. Unfortunately, until now the CryoTrap system was not tested with an FID instrument. I agree, this would give significantly more insight to both, the CryoTrap preconcentration and thus indirectly also to the AED detector. Some comparison measurements were made between the CryoTrap-GC-AED and the GC-FID instrument where same sample air was analyzed. The results agreed within 10%. Nevertheless once again, as all the reported compounds were analyzed against calibration gas with the same parameters as unknown samples, thus all the losses were accounted for.

Added to the manuscript at L190.

Section 3.2 Why is chlorine not included in this section? Was this not measured due to wavelength limitations? If available, please add Cl atom responses to this Table 2, and discuss in this section. It would have been great to also have Cl responses for the Apel Riemer standard to help diagnose the system. Also, I was trying to compare relative C and halogen, N, or S responses, but couldn't match all compounds between Table 1 and Table 2. For example, bromodichloromethane and acrylonitrile are in Table 2, but not listed in Table 1. Why?

Yes, unfortunately chlorine is not in the range of this AED detector. The reason some compounds are missing from the carbon emission line determined response (Table 1) is mostly due to the co-eluting of the compounds on the C193 nm chromatogram, but clearly separated e.g. on the bromine chromatogram. In the case of the missing compounds, they will be added to Table 1.

L204. This sentence further explains the fundamental problem with the proposed method. "The different element count scales were non-related to each other." It seems to me that this negates the advantages of AED over other methods. Note for example CS_2 and OCS. Ratio of C response between compounds (OCS/CS_2): C ratio = 637/696 = 0.915; S ratio=342/476 = 0.718. How these differences are *handled in practice needs clarification. It is not specified in the later examples, for example, if C or S emission lines are both used for the CS2 and OCS measurements, or one or the other. The authors might also wish to check on influence of CO2 on the response factors. I found a 1994 reference (Swan and Ivey, DOI: 10.1002/jhrc.1240171203) that discuss use of AED for ambient S compound analysis. Interestingly, their S ratio response between OCS and CS2 was 2932/1745 = 1.68, something quite different from that reported here (note: should add this reference to Introduction).*

All element emission lines were used independently and only one element's wavelength (the most sensitive for every specific compound) was used for calibration and quantification of the analytes of interest.

Added to the manuscript at L220.

Thank you for this reference, this is added to the manuscript.

Added to the manuscript at L27.

Section 3.3 The authors compare the advantages/disadvantages of the AED system to other common analytical systems. Given the problems outlined, the main advantage seemed to be the detection of compounds with low FID response, particularly COS. This is not particularly compelling, as other (simpler) options are available. The calculated detection limits are useful for ambient measurement of the NMHC, OCS, and perhaps several oxygenated species. LODs of halogenated species is too low for ambient measurement in most environments. The authors further argue that the multielement capabilities of the system are useful for screening gas samples. I would agree, but the selective transmission of compounds through their system would make quantitative estimations problematic, especially for exotic compounds. It would have been much more interesting to see how this system might be used to measure volatile arsenic, selenium or silicon compounds, which are not routinely done by other methods. Furthermore, the issue of the deterioration of the plasma tube and potential effects on atomic emission would also seem to be a big disadvantage for routine analysis. Given that the instrument has been in use for several years, the authors may be able to comment on that in more detail.

Agreed. During the measurement campaigns, we were always on the search for more exotic arsenic, selenium and silicon compounds but unluckily they were not detectable in the regions we measured and within the sensitivity of these elements for the system. CryoTrap-GC-MS is a better alternative for tasks monitoring targeted species. For example, the AED measurements could reveal exotic species from volcanic emission environments. This is now stated in the discussion section.

Added to the manuscript at L250.

Section 4 Case Studies. The case studies presented show that the AED system can produce data that appears reasonable. It would be important to help validate the analytical system if it was compared to currently validated techniques, or if the instrument was used in some sort of multi-lab comparison study. The forest study was mostly for terpenes and sulfur species. Were any new S species (or other heteroatom) species found in this environment? For the CARIBIC flights, one might get a better sense of sample to sample reproducibility if the data were presented in a graphical time series that also included flight altitude and perhaps ozone as a secondary tracer. Or, if the samples were measured in more than one lab, a comparison of the data for OCS would be of interest.

Until now, multi-lab comparison study has not been conducted but it would help to further characterize the instrument. In the boreal forest study, no new sulfur or other heteroatom species were detected with the method used. For the IAGOS-CARIBIC OCS, currently the CryoTrap-GC-AED is the only instrument measuring this species.