Dear reviewer #2,

thank you for reviewing our article on trace gas measurements in the UTLS using the airborne chemical ionization mass spectrometer AIMS. We appreciate the corrections and new suggestions to the manuscript. Particularly the background measurements during flight and the inflight calibration were added and discussed in more detail in the script. We have added new parameters to Table 1 and compare our measurements to other measurements made in stratospheric air.

Below we address the comments point by point. Our answers are written in italic/cursive letters.

Kind regards,
Tina Jurkat

Comments and answers to Referee #2

The paper describes the setup and initial results of an airborne CIMS instrument measuring
SO2, HNO3, HCl and HONO. The instrument inlet, ionization region, calibration and background conditions are described and implemented for initial field measurements.
The work utilizes a previously known ionization method which it describes in detail in the introduction but does not continue to discuss much comparison of their results to previous work. I have only minor problems with the manuscript and recommend it for publication once the questions/comments below are addressed or amended.

P13568 L22. Replace "exemplarily" with "for example"
P13569 L13. Replace "great" with "large"
We made the suggested changes.

P13570 L23. The author cites a number of nitric acid measurements, although other possible references which are more recent and more relevant to the work here for airborne measurements), for example Le Breton et al (2014). It is worth comparing other works measurements of HNO3, HONO (Zhang et al 2009 as an example), HCl and SO2 etc in the results section particularly with respect to instrument statistics such as sensitivity, LOD and e-folding time where possible.
We included a comparison to previous work with other instruments in the calibration section. Some of the measurements mentioned by the reviewer are mainly performed in humid tropospheric air, which makes the comparison difficult. AIMS is optimized for measurements in the UTLS region.
"Detection limits, inferred from ground measurements are between 10 and 15 pptv for a running average of 20 s and water vapour mixing ratios below 100 ppmv. Compared to Marcy et al., (2005), the detection limit for HCl is a factor of two higher, which may result from different ionization techniques. Response times of former measurements using shorter inlet lines are marginally shorter than for AIMS (Neuman et al., 2001;Zondlo et al., 2003). However calibration standards are generally added directly to the mass spectrometer instead of the inlet line which enhances the speed for the calibration and background signal but does not capture possible interferences and losses. Inflight calibrations however significantly reduce the uncertainty when compared to uncalibrated measurements (Fiedler et al., 2011). The sensitivity of CIMS techniques differ by an order of magnitude (Neuman et al., 2000; Veres et al.,

2008; Roberts et al., 2010; Le Breton et al., 2014). AIMS sensitivities are in the upper range which is mainly a result of high reagent ion count rates and mixing between ion source gas flow and sample flow. Thus despite the high sensitivities of AIMS, background noise levels still dominate the detection limit. HONO measurements with detection limits of 3 pptv have been reported (Heland et al., 2001) for daytime HONO measurements in unpolluted regions (Zhang et al., 2009). In comparison, the AIMS detection limit for HONO measurements is higher however with a better time resolution optimized for measurements of high concentrations with fast flying jet aircraft."

P13572 L7. Change ".Especially" to ", especially"
We have made the suggested changes.

P13572 L20. Were there any tests done to see how much residual calibrant gas was in the inlet line after the calibration was stopped? If the calibration gas is to flow through the inlet line, surely the sticky nature of the gasses in question would hold a memory for a substantial time period when performing calibrations in flight. This links to a question further in the paper; it would be useful to see in flight calibration data rather than show and talk about implementing laboratory calibrations as they are not representable for atmospheric flying conditions. Further information and graphics are needed for the reader to appreciate the validity of the inflight calibration system rather than referring to laboratory calibrations. An e-folding time for the in-flight calibrations should also be calculated.
We agree that adding a demonstration of the inflight calibration would be valuable for the script. Here we use the t_{90} times which is the time until 90% of the stable signal has been reached and call it response time.

Therefore, we propose to add Figure 5 as a comparison to the ground calibrations shown in Figure 4. In Figure 5, the time series of the ratio of product to reagent ion [P-]/[E-] for HCl, HNO_3 and SO_2 is shown for two campaigns TACTS/ESMVal in 2012 and POLSTRACC in 2015/2016. During this 12 min calibration sequence obtained during TACTS/ESMVal (Figure 5a), 5.8 ppbv of HNO_3 and 2.6 ppbv HCl was added to the total sample flow. The two gases were successively added to investigate the effects of the presence of other trace gases on the ion ratio. SO_2 was continuously added. The short increase of SO_2 after HNO_3 addition is due to unstable flows. Figure 5b shows a calibration sequence obtained during POLSTRACC where we added 3.5 ppbv HNO_3 and 2.6 ppbv HCl simultaneously to the total sample flow. Both calibrations were followed by a background measurement. Due to an improvement of the calibration set up (shorter lines, additional valves, continuous operation of the permeation ovens) during the second campaign, a clearer picture of the inlet line response times (t_{90}) evolves. Both campaigns had the same inlet flow and lengths. While inlet response times t_{90} of 15 to 20 s for background measurements are the same, the calibration response time was improved for the POLSTRACC campaign.

The two empty squares in Fig 4b represent the inflight calibration points which are in general agreement with the laboratory calibrations.
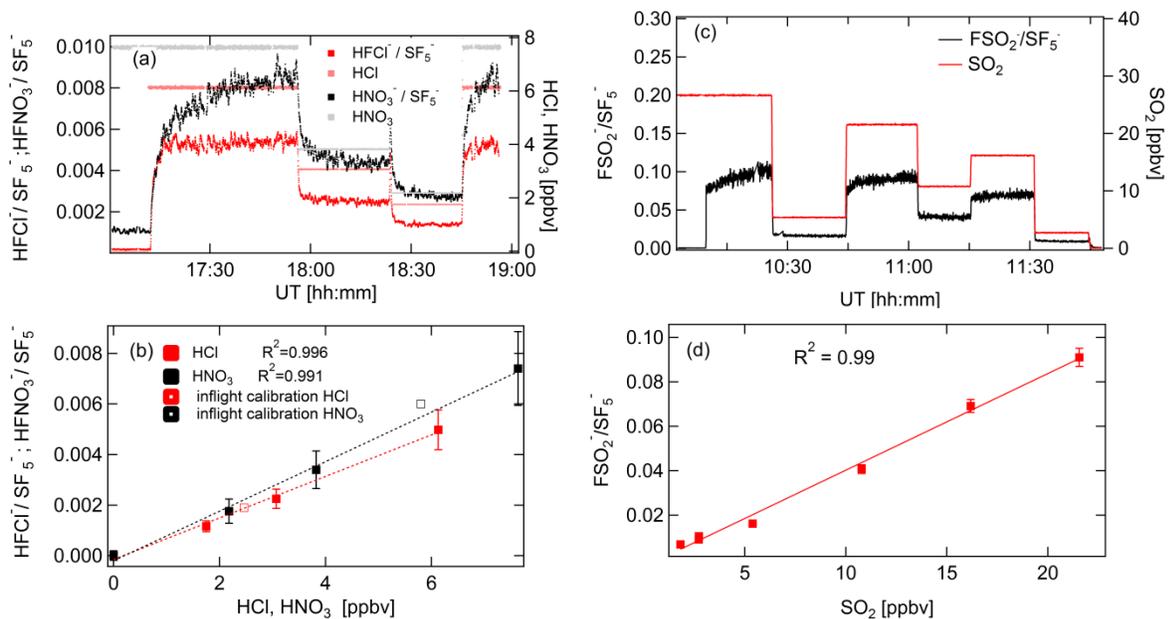
Figure 4 Laboratory calibrations of HNO_3, HCl and SO_2. (a) Time series of product to reagent ion ratios for HCl (red) and HNO_3 (black) are shown. While HCl reaches a plateau within a few minutes, HNO_3 needs more time for passivation. Once the sampling line is passivated, the plateau for a given concentration is reached faster for both trace gases (here after 18:45 h). Additionally the concentrations of the added calibration gas are shown in light red and grey. For comparison, inflight calibrations of HNO_3 and HCl are included, represented by the empty square. (b) Calibration curves for HCl (red) and HNO_3 (black). The calibration is linear up to high concentrations of 6 ppbv which can occasionally be found in the stratosphere. The error bars denote the precision of a single measurement. Figure 4 (c) and (d) is analogue to (a) and (b). In contrast to HCl and HNO_3, SO_2 is calibrated for a wider range of concentrations to account for high mixing ratios in volcanic or aircraft plumes.

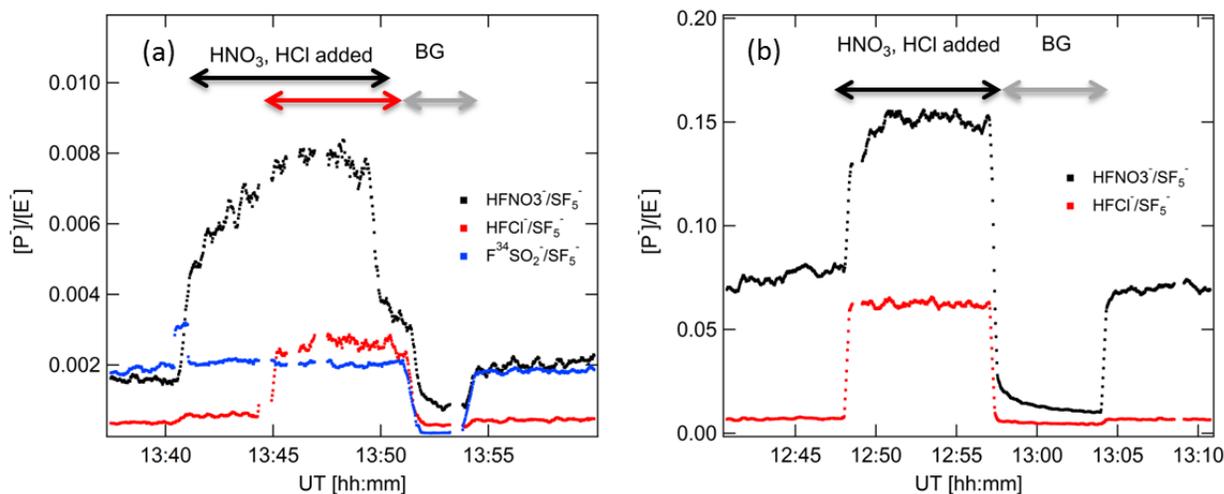Figure 5 (a) Inflight calibrations of HCl, HNO_3 and SO_2 during TACTS/ESMVal. The time series of the ratio of product to reagent ion $[P^-]/[E^-]$ for a 12 minute calibration sequence is shown. 5.8 ppbv HNO_3 was added to the total sample flow for approximately 10 min and stabilizes after 4 min. After 3 min, 2.6 ppbv HCl was added for 8 min and showed a stable signal after 20 s. 2.9 ppbv SO_2 was added continuously. A background measurement (BG) was performed shortly after the calibration sequence with t_{90} of 15 s for

SO$_2$ and HCl, and 20 s for HNO$_3$, respectively. (b) Inflight calibration of HCl and HNO$_3$ during POLSTRACC. 2.6 ppbv HCl and 3.5 ppbv HNO$_3$ were added simultaneously for 8 min. t$_{90}$ for the calibration signal is 15 s for HCl and 20 s for HNO$_3$. Background measurements were performed shortly after the calibration with t$_{90}$ of 15 to 20 s. Improvements of the calibration system between the two campaigns significantly enhance the speed of the calibration measurements. While inlet response times t$_{90}$ for background measurements are the same, the calibration response time was improved.

P13573 L18. What is the highest altitude this instrument can be operated at? The 33hPa inlet pressure maintenance will give an absolute theoretical limit, but is there a limit set by the smallest flow you receive due to the dilution flow?
We haven't tested the instrument for lower pressures than 110 hPa since the instrument doesn't face lower pressures when operated on HALO. The minimum pressure where we can operate AIMS is 33 hPa plus the pressure drop in the inlet line which is not quantified but probably some ten hPa The pressure drop further depends on the bypass flow, a higher bypass flow would result in a larger pressure gradient along the inlet line.

P13575 L5. Have any in-house comparisons of this ion source to a Po210 source been performed? It would be nice to see some evidence of the reduction in noise and in comparison to the radioactive source and also proof that radical backgrounds can be reduced by applying a lower potential.
Unfortunately no direct comparison could be achieved between the two ion source types. We propose to add another plot to show the dependence of the background signal on the ion source voltage applied to the needle (Supplement Material Figure 1). With increasing ion source voltage the background signal as well as the standard deviation which is directly linked to the detection limit increases by a factor of more than 3 between 400 and 3500V.
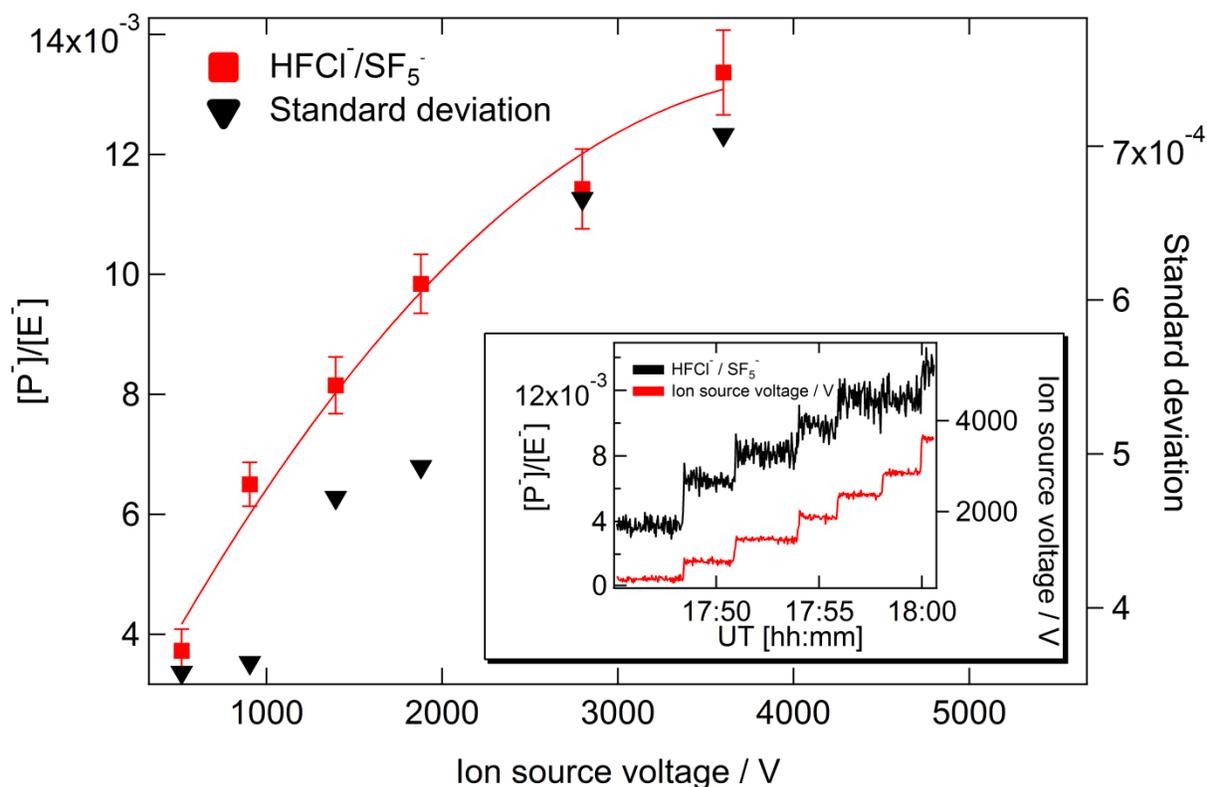
Figure 1 The ratio of product to reagent ions [P⁻]/[E⁻] (here HFCl⁻/SF₅⁻) during background measurements is shown for different ion source voltages. The inset shows the timeline of the signal, while the larger plot shows the direct correlation of the average values of [P-]/[E-] and the standard deviation of the background signal for different ion source voltages. With increasing voltage both background level and standard deviation of the background increase. Generally voltages between 400 and 1400V are used for the AIMS ion source.

P13576 L5. Replace "in" with "of"
We have made the suggested changes.

P13578 L9. The author infers a low resolution of the mass spectra. Can we have anidea of what "low resolution is"? Can the author calculate an average peak width ofFWHM. It is hard to put the infringement of other masses into perspective if the reader does not know if the resolution is what to be expected of a quadrupole.
The full width half maximum for mass 55 (HFCl-) is 1 amu, which results in a mass resolving power of R=m/Δm=55. We have included this information in the text.

P13580 L14. The addition of the calibration directly to the ambient sampling line can cause large errors, as the author states, if the ambient concentration changes. Line 7 states the HNO3 calibration flow concentration is 2.9 ppb. Figure 7 shows a time series of data with an average concentration of 3 ppb. Correctly, the author states the calibrant concentration should be similar to that of the ambient, although the variability in figure 7 would give a very large error if a calibration was performed during this flight period. The author needs to state when calibrations were performed, how times were chosen

to limit the ambient variability and also why it would not be a better option to have a calibration line or overflow through the background line so the ambient concentration was not a factor in the calculation.

Calibrations under highly variable atmospheric conditions can be tricky. The calibration is preferably done in regions where small gradients of the background concentrations are expected. If these assumptions change during the calibration, the calibration has to be repeated. During measurements in Antarctic processed air, the background can be an issue, therefore the calibration was done three times during this flight and compared. We have included this remark in the script. The chemilumiscence detector gives additional information on the NOy variability which is congruent with the HNO$_3$ background in the stratosphere. Calibration with synthetic air is an option but may not capture the atmospheric conditions and inlet surface history (humidity, temperature, other possible reactants like NH$_3$, sea salt etc.).

P13581 L1. An error of 12-16% is given for the concentrations in the calibration gas, is an overall % error for the measurement available?

The overall uncertainty is a combination of the accuracy and the precision and it's between 20 and 35 % depending on the time resolution.

P13586 L17. I disagree with the statement that noise calculations are generally determined in the laboratory. The paper is publishing aircraft measurement data, a significantly more difficult task than laboratory data and there is obviously a large amount of work to achieve peak performance of an instrument on the aircraft. I therefore believe that aircraft calibrations and backgrounds should be implemented into the data. The system has been set up to perform in flight calibrations; therefore the data is available to calculate in flight noise levels. The detection limits should also be recalculated using in flight calibrations.

We agree and add two new columns to Table 1: detection limit in flight and response time of AIMS (time until 90% of a stable signal (calibration or background) is reached) to the measurement of the individual trace gases. In addition to the inflight calibration plot, we suggest to add a new figure on inflight background measurements. To clarify how the background was retrieved we propose to add Figure 6 to the script and Figure 2 to the supplement material. The background of HNO$_3$ and HCl was best described with the following fit function

$$y = y0 + Ae^{-(x-x0)/t}$$

with characteristic time scales of t=28-50 min. Thus the background was generally at a constant level after one hour depending on the humidity of the local environment where HALO was stationed. We propose to demonstrate the effect of the background decrease with an additional plot (Figure 2) showing the behavior of the background measurements for the flight on 11 September 2012. Additionally we show a 5-min background measurement sequence (Figure 6) and added two columns in Table 1. Detection limit in flight and response time were derived from these background measurements.
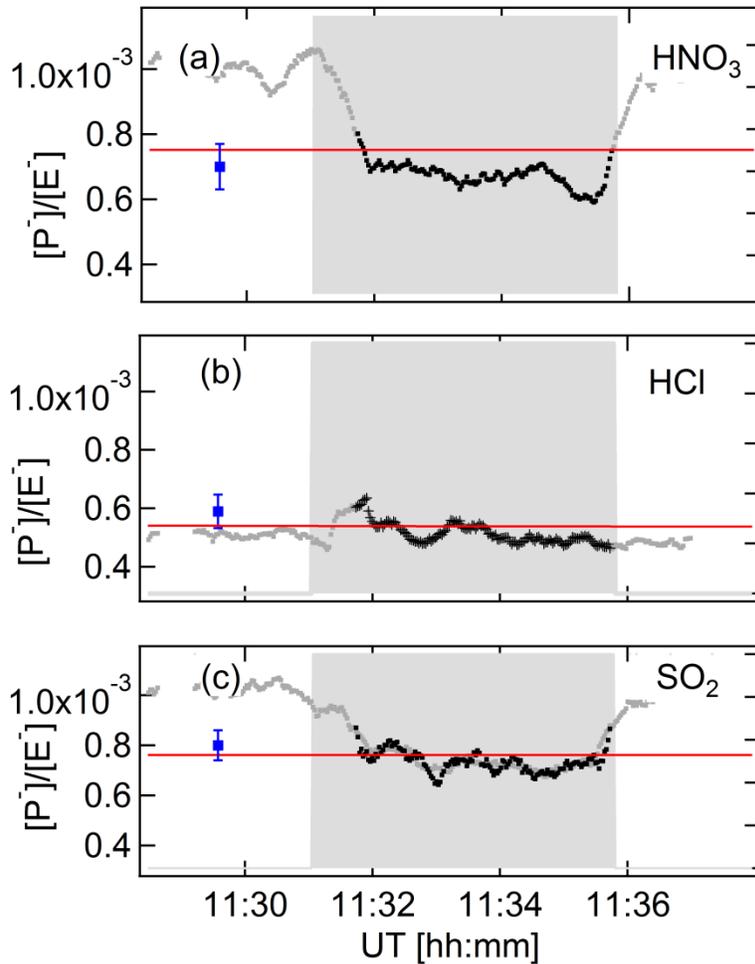
Figure 6 Ion ratios during background measurements for (a) HNO₃,(b) HCl and (c) SO₂ in tropospheric air during the flight on 11 September 2012 at 11:31 UT. Grey shaded areas show sequences where synthetic air was introduced. The red curves represent the values applied to correct for the instrumental background during atmospheric measurements. Except for the first hour of the flight, the background can be described by a constant value of 0.75e-3 for HNO₃, 0.54e-3 for HCl and 0.8e-3 for SO₂. For comparison, the ion ratio averages inferred during laboratory measurements are given (blue squares). Generally, flight and ground based measurements agree within the given variability of the laboratory measurements. Background measurements in stratospheric air are demonstrated in Figure 5.

Table 1. For a new instrument technical paper I believe these values should be compared to existing publications to allow the reader to put the statistics into perspective of how the work here has advanced the measurement capabilities of these compounds. I also feel the statistics should be calculated using in flight data.

We agree that comparing our measurements to other measurements is needed and add a passage in section 4.1 and 4.2 on comparison with other measurements. As new parameters in Table 1 we prefer to add response time (time until 90% of the calibration signal is reached) and detection limit in flight and discuss it in the manuscript.

Figure 3. This graph would be much clearer to the reader if it had 3 separate y axis soeach MS could be looked at separately and compared in the vertical component. It is currently difficult to see much of the information provided.

We adress this issue by separating the tropospheric spectrum and the stratospheric spectrum and present each in comparison to the background spectrum. We hope this will help to make the over-all picture clearer.
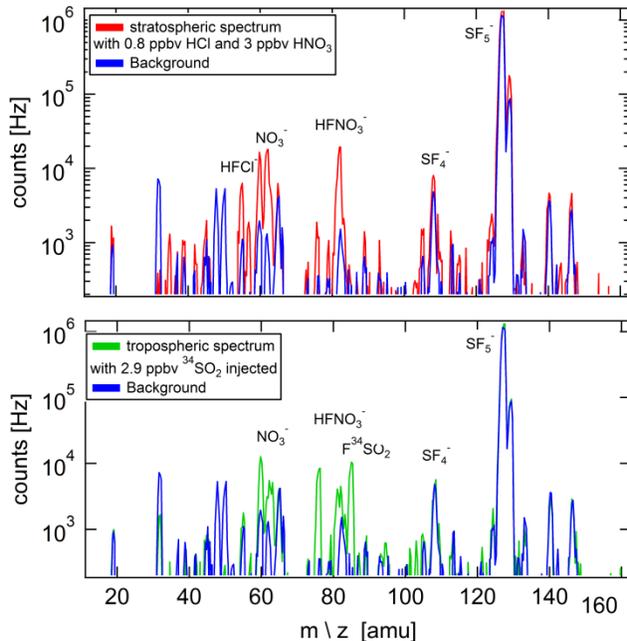
Figure 3 A spectrum of AIMS-TG obtained in stratospheric (upper panel) and tropospheric (lower panel) air south of Cape Verde during TACTS/ESMVal is shown. The SF_5^- reagent ion at 127 and 129 amu is most prominent. Nitric acid and hydrochloric acid with their respective ions ($HFNO_3^-$ (m/z 82) and $HFCl^-$ (m/z 55 and 57)) are enhanced in the stratosphere (upper panel). The isotopically labeled SO_2 is detected as the $F^{34}SO_2^-$ ion (m/z = 85 amu) (lower panel). At m/z 146 small amounts of SF_6^- are also present. The stratospheric and tropospheric spectra are compared to a background spectrum taken while synthetic air was introduced in the inlet line.

Figure 4. It would be better to see in flight calibrations and backgrounds here so the reader can evaluate the performance of the instrument in flight. What are the e-folding times of each species.

We agree and suggest to add Figure 5 and Figure 6 with two inflight calibrations and a background measurement plot. The response time of the inlet with respect to the three trace gases has been listed in Table 1 and is discussed in the manuscript.

Figure 5. Could H2O concentrations be shown on this plot. It is mentioned that it is important to the ionization efficiency so would be good to see that there are no strong correlations of t series with H2O. Figure 7. The R2 of this fit could be useful to show here.

We would like to keep the plot simple and refrain from adding another variable to the already crowded plot. Generally H_2O was below 45 ppmv throughout the whole flight except for the ascent and descend where data were removed below 8 km. Thus no effect of water vapor is expected during this flight. We added a statement on the H_2O mixing ratios.

P13593 L12. Replace "stabely" with "stably"

We have made the suggested changes.