Author response to reviewer's comments on

"Intercomparison of atmospheric water vapour measurements in the Canadian high Arctic" by Weaver et al.

Reply to Anonymous Referee #3

We would like to thank the reviewer for their detailed and helpful comments. Replies are given below in blue italics.

The paper investigates the accuracy of water vapour measurements in the Canadian Arctic (Eureka) over the time span from 2006 to 2014. This is a highly important topic as the role of water vapour for Arctic Amplification is uncertain and the detection of water vapour trends and their spatio-temporal patterns is a major challenge. As such the paper provides an important contribution to the field. However, there are some major issues in terms of focus and technical aspects that require major revisions.

General Comments: The manuscript reports on the intercomparison of water vapour column and profile measurements observed by a suite of different instruments operated at two sites in the Eureka area: 1) The paper reports a "good agreement between all datasets with correlation coefficients > 0.9". I have two issues with this statement: i) How good is good enough? ii) Is the correlation coefficient the right quantity to judge good agreement? For i), the authors should look at the requirements for essential climate variables, the accuracies required to detect trends (see Serreze et al., 2012, JGR) and/or importance in terms of the radiative budget (see for example Cox et al., 2015 Nat. Commun.)? Are the data good enough to evaluate atmospheric reanalysis (see Perro et al., 2016, AMT)? The requirements feed back to the question which measure is the right one to judge agreement among data sets.

The correlation coefficient is used in this study alongside other measures used to show the agreement. We also examine the differences (absolute and percent) to assess agreement between datasets. In specific regards to the correlation coefficient, values greater than 0.9 are widely considered a threshold for the relationship as being "very high" (Hinkle et al., 2003). While not reported, the p-values for all comparisons are essentially zero ($p < 0.05$ is a common test for correlation significance). Moreover, comparisons using these quantities aligns this study with similar studies at other sites (e.g. Schneider et al., 2010, Palm et al., 2010, Buehler et al., 2012).

Often a combination of absolute and relative values is used value. This is especially important for low column water vapour (PWV) values there percentwise differences can become rather large. For instruments that have a certain detection limit, e.g. microwave radiometer, sun photometer, very large relative errors can occur. Another reason why in addition to the relative error and the correlation coefficient also the root mean square difference (RMS) should be used is the sensitivity of the correlation coefficient to the PWV spread. As the sample size and thus PWV spread varies in pairwise comparisons, the values achieved for different instrument comparisons are not comparable.

The paper has been revised to include the standard error of the mean when discussing the differences. In addition, the root mean square difference has been added to the tables summarizing the comparison results.

2) The focus of the paper is sometimes confusing. First, I thought the paper should introduce water vapour measurements in the Arctic in general but became puzzled in the introduction with a rather limited list of satellites omitting the most used microwave sounders that can also penetrate clouds. Next, emphasis was put on the MUSICA FTIR processing not even mentioning the measurement principle of solar absorption. As a side note I find the branding of MUSICA a bit overdone and would prefer a more generic naming convention as it is done for the other instruments. Then at the end of the introduction the new thermal infrared measurements are mentioned as a novelty in the outline. Therefore, the introduction section (p3l17 to P4l2) needs to clarify the overall goal.

The introduction has been edited to focus the discussion on the specific goals of the paper, which are to assess the available measurements of water vapour at the Eureka site.

The rationale for the MUSICA-specific references is because there are other possible FTIR retrieval approaches. For comparison results in tables and figures, the dataset is referred to as '125HR' in reference to the instrument. However, in describing the dataset, we consider it appropriate to maintain the MUSICA label because it is specific and ensures the reader connects the retrieval approach with the related literature.

The MUSICA dataset is described in Barthlott et al. 2017 and can be found in DOI.org/10.5281/zenodo.48902.

3) The readability of the paper is poor due its structure that goes three times through the series of instruments (2. Instrumentation, 3. Comparison, 4. Discussion) and switches between the two sites and column & profile. I strongly recommend the authors to move the profile comparison into an appendix and concentrate on PWV. Anyhow only FTIR and radiosonde profiles are compared although also AERI has profiling capability. Profiling might be even an opportunity for a second paper that also investigates the interesting question is to what degree remote sensing can observe moisture inversions in the Arctic (see Serezze et al., 2012).

We agree that a comparison of profiles would be a useful study. The profile comparisons between the 125HR and radiosondes have been removed; we expect profile comparisons will be presented in a future publication.

4) The paper does not explicitly mention the problem of cloud sensitivity that prevents measurements from several instruments during long time periods. What are the implications then trying to derive trends? What is the percentage of cloudy cases? Sometimes only 64 coincidences where found. Will this give a significant result?

That clouds limit many of the measurement techniques in the study has been mentioned. See Section 2.1.1 (page 6 line 23) for the 125HR, Section 2.2 (page 11, line 12-13) for the sun photometers. An additional comment has been added to the discussion.

The implications of the clear-sky bias for trend analysis is of interest; however, this paper does not seek to analyze the trends of the available datasets. Water vapour trends at Eureka is an interesting topic for a future study.

5) Altitude difference: Measurements at two different sites 15 km and 610 altitude meters apart are used. On page 5 it is mentioned that this offers "an opportunity to investigate the effectiveness of different measurement techniques" but this is not addressed properly and no conclusions can be found in the manuscript. In fact, Fig. 17 shows the high correlation of 0.97 between sun photometers at the two sites indicating no strong differences. The authors need correct the altitude difference before doing any comparisons. This can easily be done by taking the mean of the absolute humidity (ground measurement) at both sites and calculating the hypothetical column to be added to the higher altitude site.

The correlation between the instruments is an indication of the relationship between the conditions at the two sites. However, the slope indicates a large difference exists, reflecting the larger total column at the lower altitude site, and the large scatter illustrates the substantial variability in water vapour even between two close measurements.

We disagree that correcting for the altitude difference is a simple matter of subtracting the partial column between the sites. This would ignore the significant meteorological differences between a near-sea level site in a valley and a windy site atop a ridge. Partial columns between the two sites have now been calculated, compared, and presented in the revised text using the radiosonde, GRUAN, and sun photometer measurements. The partial column between the two sites calculated using near-simultaneous sun photometer measurements reveal considerable variability with the partial column calculated using the radiosonde or GRUAN profiles. The partial column from the SPM compared to GRUAN, for example, has a mean agreement of 0.6 mm PWV (20.7%) using 203 coincident measurements. These differences are largest in the summer. Adding (or subtracting) the partial column between the sites measured by any of the three options available to facilitate comparisons between instruments would introduce substantial errors, particularly if some mean value were used instead of the closest available radiosonde or GRUAN profile (where they are available). If the nearest radiosonde or GRUAN measurements were used, in addition to the concerns raised by the observed disagreement with the sun photometer partial column, their sampling sparsity (e.g. 6 am and pm, local time) would bias the partial column subtracted from measurements throughout the day.

Instead of showing the direct comparisons, the following major changes have been made:

- *Direct comparisons between instruments at different sites have been removed.*
- *Comparisons between instruments at the Ridge Lab and the radiosondes (standard dataset and GRUAN) now use the radiosonde profile integrated only down to the Ridge Lab altitude (i.e. 610 m) to put both measurements on an equivalent altitude basis.*

- *Three datasets have been calculated for the partial column between the sites, using measurements from the radiosondes, GRUAN, and sun photometers (the difference between near-simultaneous SPM measurements at both sites). A new section discusses the partial column between the sites.*
- *The table and correlation plots summarizing comparison results have been split up into three: one for comparisons among Ridge Lab instruments, one for comparisons among 0PAL instruments, and one for comparisons between measurements of the partial column between 0PAL and the Ridge Lab.*

6) The microwave radiometer time series needs to be stopped in 2009 as the last calibration (inspection) was performed in 2008. A calibration needs to be performed roughly every 6 months – maybe up to a year. The comparisons with other instruments clearly show that the data are completely off in later years. No conclusions can be drawn from these data.

It is ideal to perform regular calibrations of instruments; however, in practice this is not always possible at a remote location. In this specific case, the remoteness of the site contributes to the difficulty of regular maintenance. The MWR data are publicly available. We feel it is useful to document and discuss the tradeoffs between having measurements without ideal maintenance or not having measurements. That said, we agree that using the MWR to compare with other measurement techniques being validated after it has become clear that there is a calibration problem is misleading.

The instrument comparisons now use the MWR dataset prior to July, 2010. This cut-off point was selected because the first sign of MWR measurement accuracy loss occurs in the fall of 2010 (which appears to affect low-PWV measurements). Figure 20, which had demonstrated the loss of calibration, has been modified but retained. The aim is to justify the cut-off of the MWR dataset in summer 2010 instead of using the full timeseries, and also to show the calibration problem for the benefit of other data users.

7) I am not convinced that all 20 figures are really needed to convey the message of the paper (see specific comments). In fact the conclusion section is rather weak. What are the major results? The paper started showing some climatology from radiosondes – so which additional information is provided by the ground-based remote sensing measurements?

The following figures have been cut from the paper in an effort to make the manuscript more succinct:

- *Figure 4. Its intention had been to illustrate the variability of the water vapour column from year to year, which is substantial and not otherwise shown in the manuscript.*
- *Figure 7 (a) has been removed, as it duplicates the information in Fig. 7 (b).*
- *Figure 12, which had shown correlation plots of AERI vs. co-located instruments, has been removed. The new Fig. 11 and 12, which replace the original Fig. 13 (large set of all correlation plots) show all correlation plots from the Ridge Lab and 0PAL, respectively.*
- *Figures 14, 15, and 16 have been removed because the profile comparisons have been removed.*
- *Figure 17, showing the comparison of SPM at Ridge Lab and 0PAL have been removed, as comparisons between instruments located at different altitudes are no longer presented.*
- *Figure 18, showing the timeseries of differences between the current and previous MUSICA retrieval versions.*
- *In addition, Figure 20 (b) has been removed, as it offers similar information as Fig. 20 (a).*

The AERI data are recommend due to their sampling – the data set can be used to study sampling issues. What products do you recommend a potential user except "to be explored"? Are the data good enough to look at longwave downward feedback? How does the water vapor climatology compare to Barrow (cf. Perro et al., 2016)? 8)

Additional comments have been offered in the conclusions. The MUSICA retrieval of the 125HR spectra is also a valuable dataset, but users should be aware of the small observed wet bias. The Eureka radiosondes are used for a variety of investigations; the conclusions of this paper indicate users should be aware of a likely small dry bias, particularly during summer. Climatological comparisons to other Arctic sites has been noted for future work.
Several minor error like missing references occur.

The manuscript has been carefully checked to ensure references are appropriately listed.

I also hope that it is a typo on page 12, line 20 that says that relative humidity is converted to mixing ratio (dimensionless) and then integrated and converted to PWV. PWV is the vertical integral of absolute humidity (kg/mˆ3) so mixing ratio needs to multiplied with air density to give absolute humidity and then integrated.

Since the conversion between mixing ratio and number density profiles is common and standard in the atmospheric measurement community, the details were omitted. The explanation has been expanded as suggested.

Specific Comments:

Title: Mention Eureka as the regional focus is limited

The title now refers to the site to clarify the specific geographic context of the study.

P3l4 Reference Soden et al missing in reference list

The Soden reference has been added.

P3l14: say that WV trend from Lesins is derived by radiosonde. I am a bit reluctant of the quality of 1960's 1970's radiosonde. Would be good to also refer with more detail to the work by Serreze et al. (2012)

We have clarified that the humidity trend reported by Lesins et al. 2010 was derived from Eureka radiosoundings. i.e. the text now reads "the total column of water vapour at Eureka increased by 10 ± 3% between 1961 and 2007, according to radiosoundings (Lesins et al., 2010)".

The Serreze et al. 2012 work has been noted and cited.

P3l23, Say clearly that Eureka is no GRUAN station

The text in Section 2.5.1 now explicitly states that Eureka is not a GRUAN station and that this dataset was produced for this study. i.e. "Eureka is not a GRUAN site; however, a subset of the Eureka radiosonde data has recently been processed using the GRUAN technique for this study".

P4l15: Reference Adams et al., 2012 is missing

Adams et al. 2012 has been added to the references.

P5l20: This paragraph is disconnected. Why is there a need to mention water isotopes if nothing is done with them within the paper.

The measurement of isotolopologues is a major difference between the 125HR and other instruments at PEARL/Eureka. While it is not the data product under scrutiny in this study, it is worthwhile for the reason to be aware of this additional capability when considering whether water vapour measurements using the FTIR adds value to the existing suite of instruments.

P6l5: Nothing has been said about the emission FTIR (AERI) before

The emission FTIR (AERI) was mentioned in the first section of the introduction, see p.4 line 5-6.

P6l18: The single sentence as no meaning and should be omitted.

We assume the reviewer intended P6l19, as P6l18 is a section title. This section division has been removed.

P7l3: Paper by Rodgers (2000) not mentioned

This reference has been added to the text and the reference list.

P8l16: What is the standard software? Manufacture, reference?

AERI instruments and the software to run them are proprietary technology developed by ABB Inc. This has been noted in Section 2.2.

P9l14: It is not true that 10 km is the border when radiosondes can't measure humidity accurately any more. Probably you mean the stratosphere but the tropopause height varies quite a bit so this level should depend on the actual sounding.

This has been re-worded. The AERI retrieval procedure chose 10 km as a consistent best approximation for the limit of where the radiosondes are able to report a relative humidity value. The available GRUAN data (between Sept. 2008 and Mar. 2013) reports that the mean altitude where the water vapour mixing ratio reaches 5 ppmv or less is 10.4 km, so this approximation is reasonable for the purpose of a retrieval a priori.

It may be of interest to note that the GRUAN-processed radiosonde data reveals that 99.4% (σ = 0.6%) of the water vapour column above Eureka is below 10 km. Also, the mean tropopause height above Eureka is 8.7 km (σ = 1.2 km) according to the GRUAN-processed radiosonde data, and 8.6 km (σ = 1.1 km) according to NCEP data (2006 – 2016).

P9l14: Because radiosondings are also in cloudy sky conditions you're a priori will have a moist bias. P9l22: After developing atmospheric profiles sounds odd: better stick to constructing P10l2- how large are the uncertainties in percent? P10l20-29: I think I understand what you want to say but it is really difficult to follow. Somehow the issue of vertical (layers) and temporal interpolation become mixed. I suggest to revise this carefully and also use paragraphs to separate different thoughts, for example before the "forward model". Here a sentence to introduce its purpose would be helpful. Add spectral before resolution when you talk about matching.

The AERI retrieval description has been revised to add clarity. Additional details are now available in a new appendix.

P11l14 and all instruments: you give a bias error but what about random uncertainty?

Where available, we have added details available about the uncertainties in the text. This is not available from all datasets, e.g. all the AERONET sun photometer error values were "N/A" at the PEARL and 0PAL sites.

P11l20: "in most conditions" -> during non-precipitating clouds. You mention short-term variability but you use 30 min averages – see Steinke et al. 2015 ACP.

The quoted text has been revised as suggested for greater specificity.

For the purposes of this work, we used 30 minute averages (1) to provide a time interval similar to other instruments, (2) to reduce error, and (3) to create a manageable number of data points.

The comparisons have been updated to use 5 minute MWR averages. This aligns the measurement time more closely with other instruments. The results are similar.

P11l27: There are also radiosondes which measure dewpoint temperature.

The text has been revised to include this fact.

P12l5: Before you said that the radiosonde can't measure humidity above 10 km but now you say that 99.4 is below 15 km. Better give value for 10 km.

The earlier comment about using a humidity value to 5 ppmv above 10 km instead of the radiosondes was in reference to a decision that affected the AERI retrieval procedure. Both statements have been reworded.

P12l10: Better give uncertainty due to rounding in absolute humidity to relate it to humidity.

The conversion between relative humidity and the quantity discussed in the paper (either mixing ratio or precipitable water) is temperature dependent. The text has been revised to comment on the uncertainty reported by the cited paper, which the reader can refer to for detailed analysis.

P12l19: One sentence does not make a paragraph!

That is a matter of stylistic opinion, not a matter of grammar. For example, a detailed accounting of influential English writers in the 19th century found that most used single-sentence paragraphs (Herbert, 1894). Further, the authoritative The Oxford Essential Guide to Writing explicitly endorses single-sentence paragraphs as useful in a variety of occasions, such as cases where it "may serve as an emphatic paragraph" (see chapter 12, "Basic structure", in Kane, 2000).

P12l19: Important: One sentence does not make a paragraph!

In addition to the above references, Government of Canada Translation Bureau notes that this is a myth, and that "occasionally, one sentence may be all a paragraph needs." (Peck, 1997)

This particular piece of text has been revised to add more details about the conversion between relative humidity and precipitable water vapour to address another reviewer comment.

P12l22: No single subsection 2.4.1 without 2.4.2. Please say how many radiosondes are processed according to GRUAN. Why not more?

Table 1 notes the details of the available datasets, including the number of radiosondes processed by GRUAN. The reason for there being fewer GRUAN measurements than radiosoundings is discussed in Section 2.4.1 (i.e. page 12, lines 28-29). The GRUAN processing requires the raw data files; however, a complete record of these raw files was not available for this study. This is a regrettable limitation the authors have no control over.

P13l1: What does close mean? If you say that 64 km are close when I argue that the difference between both sites (15 km) is negligible.

The horizontal distance between the two sites is relatively negligible. However, the vertical distance and the potential meteorological differences make the two sites distinct.

P13l3: section 2.5. Very short: Maybe add some information on WV conditions. What is the impact on sampling due to instrument limitations.

A further comment about water vapour conditions at Eureka has been added (to what is now Section 2.6), noting the sampling issue and the range of PWV value observed by the datasets in this study.

P14l10: In order to decide what coincidence criteria to use one should not look at quality of the intercomparison but at the time scales of water vapour variability at the site (e.g. autocorrelation), e.g. from high resolution, continuous time series or reanalysis. Twelve hours for sure is too long. . .cf Steinke et al., 2015, www.atmos-chemphys.net/15/2675/2015)

The coincidence criterion is also limited by available measurements. We feel the coincidence-finding approach used in this study is a reasonable balance between finding enough measurements to compare with and minimizing the time difference where possible.

If 10 minutes had been used, for example, there would be few, if any, coincidences between the radiosondes and the solar-viewing instruments for most of the year. Most comparisons involved time differences significantly shorter

than the maximum allowable 2 hour coincidence criterion because only the closest pair out of all possible coincidences meeting the criterion were used.

Indeed, the comparison between the 125HR and the Ridge Lab sun photometer shows nearly identical results when the coincidence criterion is reduced to 30 minutes compared to 120 minutes (2 hours). The mean time difference between 125HR and SPM measurements in the case of the 120 minute coincidence criterion (presented in the study) is 5.4 minutes (standard deviation of 11.7 minutes). The comparison between the 125HR and the E-AERI at the Ridge Lab similarly has a mean time difference of 5.7 minutes (standard deviation of 12.6 minutes).

Only when comparisons involve the radiosonde or GRUAN data are the time differences, on average, greater than 10 minutes. Even among comparisons with the radiosondes and GRUAN data, the majority of coincident measurements are within 30 minutes. The only exception is the 125HR vs. radiosonde comparison, where only 43% of coincident pairs are within 30 minutes (the mean time difference is 55.0 minutes).

A comment has been added to the method section to make the implications of the matching approach more clear.

P14l25: warm atmospheric temperatures do not "hold much greater quantities".. With same relative humidity the atmosphere contains more water vapour. The pages 15-17 are rather lengthy descriptions listing number from figures and tables.

The sentence has been revised to state: "PWV values near the equator, where a warm atmosphere can hold greater quantities of water vapour, can be as large as 50 mm.

P18l9: What is excellent correlation?

Correlation coefficients above 0.9 are commonly interpreted to indicate a 'very high' or excellent correlation strength. (see, e.g. Hinkle et al. 2003).

P18l18: "demonstrates significant influence of altitude " : you need to correct for that! The correlation is 0.97 for the sun photometer – so not too bad.

The manuscript has been revised to exclude comparisons between instruments at different altitudes to satisfy comments of another referee. However, radiosonde and GRUAN partial columns starting at the altitude of the Ridge Lab have been calculated for comparison to Ridge Lab measurements.

P19l26: You argue that the bias is not to bad for a global data set but this depends what you want to do with it. . .

The decision to keep the consistent retrieval approach for all global sites is outside the scope of this study. For details on the goals of the MUSICA approach to the retrieval, see Schneider et al. 2016.

P19l29: Seasonal pattern? I guess it is more a dependence on PWV itself and temperature.

Yes, the seasonal pattern is largely influenced by the seasonal temperature changes.

P20l17: The radiosondes drifts away and the MWR looks at zenith!

The radiosonde and MWR comparisons are discussed in the subsequent paragraph.

P21l9: The statement does not make sense. Without an altitude correction you can't compare measurements at two different altitude sites.

Comparisons between instruments at different sites have been removed.

P26l5: what does cited 2011 mean?

That particular resource evolves with new data. The resource was accessed in 2011. The reference format has been updated to make clear it is an online resource.

P28l24: latex commands

The citation text accurately reproduces the title of the article. It is not a latex command.

Table1: Make clear that GRUAN is a subset of the radiosndes. Why not more?

The description of the GRUAN dataset in Section 2.4.1 states in the opening sentence (page 12 line 22) that "A subset of the radiosonde data has been analyzed using the technique developed by GRUAN". Section 2.4.1 (page 12 lines 260-further explains that there are gaps in the raw measurement file records (which is beyond the authors' control). The GRUAN processing requires the raw files, and this is the reason for the gaps and limits on the GRUAN processed data.

Table2: What type of accuracy, e.g. absolute, relative, bias? Use same order as in the ordering of the sections. Why here MUSICA and not 125HR?

Most instances where MUSICA is used to describe the dataset have been changed to 125HR to ensure consistency, except where it is necessary to refer specifically to the specific FTIR retrieval technique used in this study. In this case, MUSICA was used because the accuracy of the retrieval is discussed in a series of published articles. The entry has been revised to "125HR (MUSICA standard)" and "125HR (MUSICA extended)".

Table3: Why did you use the smoothed RS for calculating columns? Wha is it not mentioned in the text but in the table caption?

The smoothed radiosonde profiles were used to compare to the 125HR because this puts them on the same vertical resolution. This is mentioned in Section 3.1.

Explanation for the entroes in the table is missing. I can only guess what is what? Otherwise I like the table – why not do the same for the other site?

A complete description of the entries in the table has been added. In addition, the format of the tables summarizing the comparison results have been revised for greater clarity and readability.

Table4 – does not make any sense without correction for altitude difference. Here you describe the different values but not in table 3.

Comparisons between instruments at different sites have been removed, except for radiosondes as noted earlier.

Fig2: horizontal lines could be integrated in Fig.1 and then the figure could be omitted.

This is an interesting suggestion; however, we have decided to keep the figures as they are. Combining the information about vertical distribution with the seasonal and overall mean profiles would create a busy figure that is difficult to read. In addition, we feel it is worthwhile to note how the vertical distribution of water vapour changes seasonally, as shown in Fig.2 (b).

Fig3: add mean and std of each data set

Adding this information to the side of each subplot was tried; however, it cluttered the figure. This was particularly the case because some subplots include multiple datasets (e.g. the three AERI datasets). The information has been added to Table 1, which summarizes each data set. However, these values have to be interpreted carefully, as they are significantly affected by the instruments' different sampling (e.g. the clear sky bias of the FTIR and sun photometer). This point has also been added to Section 2.6.

Fig4. Shows typical PWV variability – what are you aiming at? If you want to perform a climatological analysis you should use all the different instruments. You could for example also show the mean values for interannual variability. How does the frequency of occurrence look like. For what is currently shown in the paper the figure is not necessary.

Figure 4 has been removed. This comment offers interesting suggestion for consideration in a future publication examining water vapour trends at Eureka.

Fig. 5 but especially Fig. 6 are not necessary

One of the goals of the article is to assess the accuracy of a new water vapour dataset, the MUSICA retrieval using PEARL 125HR spectra. The averaging kernel and sensitivity shown in Fig. 5 is critical to understanding the retrieval's ability to capture information about the state of the atmosphere. Indeed, its inclusion was explicitly supported by another referee's comments for this reason.

Fig. 6 is necessary to illustrate the implications of relaxing the standard MUSICA SZA filter, primarily that this increases the uncertainty. The use of a non-standard MUSICA dataset requires explanation and justification. This may inform other high-latitude sites' use of the MUSICA retrieval results.

Fig. 10: I do not see the value of that graph – the table already contains all the data and the temporal distribution is already shown in Fig. 9.

Differences between datasets cannot be reduced simply to a mean value, a standard deviation, and the RMS difference. Some of the difference timeseries show seasonality; others do not. Some show spikes in the differences, others do not. These are relevant characteristics to consider when assessing the agreement of measurements that are not captured by the table values.

Fig.14: What happens in the lowest kilometre? This is not described in the text. Maybe I don't understand the most right figure as sigma is not explained but a standard deviation of more than 100% - actually above 2000% in the plot is not tolerable. . .or di you mean ppmv? ..also some other figures not necessary ..

The units should indeed to ppmv. This was a regrettable typo.
This figure has been removed to satisfy recommendations of another reviewer.

References

Barthlott, S., Schneider, M., Hase, F., Blumenstock, T., Kiel, M., Dubravica, D., García, O. E., Sepúlveda, E., Mengistu Tsidu, G., Takele Kenea, S., Grutter, M., Plaza-Medina, E. F., Stremme, W., Strong, K., Weaver, D., Palm, M., Warneke, T., Notholt, J., Mahieu, E., Servais, C., Jones, N., Griffith, D. W. T., Smale, D., and Robinson, J.: Tropospheric water vapour isotopologue data ($H_2{}^{16}O$, $H_2{}^{18}O$, and $HD^{16}O$) as obtained from NDACC/FTIR solar absorption spectra, Earth Syst. Sci. Data, 9, 15-29, doi:10.5194/essd-9-15-2017, 2017.

Hinkle DE, Wiersma, W., Jurs, S.G. (2003). *Applied Statistics for the Behavioral Sciences*. 5th ed. Boston: Houghton Mifflin.

Lewis, E. H. (1984). *The History of the English Paragraph*. Chicago: The University of Chicago Press.

Kane, T. S. (2000). *The Oxford Essential Guide to Writing*. New York: Oxford University Press.

Peck, F. (1997). *Style Myths*. Public Works and Government Services Canada, Translation Bureau. Online. URL: http://www.btb.termiumplus.gc.ca/tpv2guides/guides/favart/index-eng.html?lang=eng&lettr=indx_autr8jiZHa-hDjSg&page=9llxSK21d3Cw.html