Interactive comment on "Intercomparison of atmospheric water vapour measurements in the Canadian high Arctic" *by* Dan Weaver et al.

Anonymous Referee #1

Received and published: 23 January 2017

The manuscripts deals with ground-based measurements of water vapour at a high latitude (80N) site. Some new data are presented and a broad comparison is made. The work is of special importance due to the sensitivity of the Arctic region to climate change and the crucial role of water vapour as a main feedback mechanism, in combination with the fact that there are few permanent research sites inside the Arctic region.

The main strengths of the manuscript are the high number of observation techniques considered, the introduction of new AERI retrievals and a clear demonstration that the MUSICA retrievals have deteriorated for dry sites. The manuscript also treats the difference at the site between standard radiosonde data and the GRUAN version (not clear if these are new results or not). These topics fit well with the scope of AMT. I have no basic doubts about the amount of work, or its quality, that has gone into the

manuscript, but the results lack generality and the presentation should be improved.

The only general results I can identify is the conclusion around the MUSICA retrievals. Other contributions are either (1) quite specific for the site or (2) are not described in sufficient detail. The basic results are of course obtained at the site (and have intrinsic value), but there is very little discussion on how the results agree with other comparisons. For example, are these results consistent with the ones presented by Buehler et al (2012) and Palm et al (2010) for other high latitude sites? That is, to solve point (1) the results should be properly compared to what has been found elsewhere.

Regarding point (2) I am mainly thinking about the new AERI retrievals. I simply don't think it is possible to recreate the implemented algorithm from the description provided. For example, the "changes" discussed around line 4 of page 10, are these applied on the total column, or for layers? I assume the later, but there is no information on what layers that are used. (If the total column is perturbed, the description makes a simple thing very complex.) I can not really understand the treatment of the self-broadening part (end of page 10) of the water vapour optical depth. I would also like to see a demonstration that the cloud optical thickness can be as high as 0.6 without having a significant impact on the water vapour retrieval. That is, Section 3.1.2 should be revised to ensure that the retrieval algorithm can be understood and duplicated by others. However, I don't know if this can be achieved without making the manuscript too long, and an additional article must be considered.

Some general comments around the presentation of the results:

Obviously incorrect/bad comparisons should not be included. This is not the case now and the reader is easily confused by a high share of irrelevant numbers. I am here thinking about comparing columns with different start altitudes (i.e. Ridge lab vs. 0PAL), and results coming from the microwave radiometer after fall of 2010 (when its calibration started to fail).

In many places the sign of found difference is unclear. Referring to Eq. 1, it is not clear

what measurement that has been taken as X and Y. Table 3 is one example. It says "comparison of 125HR", but does this mean that 125HR is X or Y? I get the impression that the authors tend to select the mentioned instrument as X, but I can not find any general statement on this point. I would also say that the standard approach is to put the identified reference as Y, and recommend to use this scheme. (Also note that all differences are positive in Table 3. Is that really correct?) Another example is that the text in Sec 3.3.1 does not not fit with a positive difference (i.e. should it not be -4.6%). Please check carefully the complete manuscript to avoid all uncertainty about the sign of differences.

Another problem can be illustrated by line 24 on page 15: "the mean percent difference is 11.4+-13.2%". As written like this, the value 13.2% must be taken as the uncertainty of estimated mean difference. That is, no significant difference has been found in this case! However, I guess that 13.2% refers to something else. Based on Table 3 I assume that this is the standard deviation of the sample variation (that is of interest if the measurement precision are investigated, but not for the mean difference). Further, are 1 or 2 standard variations given? Please revise the manuscript on these points.

The manuscript contains a high number of figures. Several figures can be removed without any important loss of information, such as Fig. 4, 7a, 11 and 15. The results displayed are either not important for the discussion, or are already covered by values in tables.

Use the term bias more carefully. For example, the only possible interpretation of the sentence in Sec 4.1.6 is that GRUAN has a dry bias, i.e. that it its mean is too low! If this is what is actually meant, this results should be stressed (and e.g. be part of the abstract). However, I assume it is an example of poor use of the term bias. Note that the correct meaning of bias is a deviation from the true mean, but it is frequently used incorrectly in comparisons like this. As it has become standard, I can't refuse the authors to use bias also in the later manner, but the text must then be very precise to avoid all risk for misinterpretation.

Paragraphs consisting of a single sentence should be avoided.

Some comments on a more detailed level:

The abstract should be more specific, i.e. more hard facts, and less vague comments such as "Good agreement is observed" and "A variety of biases and calibration issues are revealed". The fact that the MUSICA retrievals have deteriorated at the site shall be clearly expressed, even if this a negative result.

Page 7, line 22: What signifies "statistical errors"?

Page 7, line 25: N is not defined in the text.

The difference between standard and extended MUSICA should be expressed more clearly. The text indicates that it is only the SZA threshold differs, but this is not consistent that are differences between the two versions for SZA < 78.5.

Page 8, line 5: How is this sensitivity defined?

Page 8, line 8: What is XCO2?

Page 8, line 19-20: PCA in itself does not remove any noise! It is just a transformation the measurement. The text implies that a filtering is done after PCA, and it is then critical to describe how many components that are left. Another example on that the AERI algorithm is not described properly.

Page 13, line 4-5: "two satellite instruments" is left from an earlier version.

Page 14, line 13-17: This test, does it apply closest or all version of measurement matching?

Page 14, line 14: Add [] around mm, to be consistent with the next unit.

Sec 3.4: With satellite part removed, why having Sec 3.4.1 and 3.5, and not just 3.4? For these profile comparisons you must also discuss the vertical resolution and the measurement response of the MUSICA retrievals. I assume that the good values at 15

km just reflects a low variability around a priori.

Page 17, line 14: 2ppm/6% implies a mean VMR at 15 km of 30 ppm. Is that correct?

Page 18, line 4-5: Don't understand what this sentence tries to stress.

Page 18, line 23-24: How can the agreement be close, when there is a consistent wet bias? That is, statements such as good and close agreement are objective and should be avoided. Just present the numbers.

Page 19, line 9: "showed" refers to what? Earlier in the text or another article?

Page 23, line 4: The statement is correct, but it does not justify to operate multiple remote sensing instruments. My point here is that a microwave radiometer gives highest possible temporal coverage. If the microwave radiometer can not be used (due to precipitation), there is no other measurement technique to use instead. Other techniques can have lower random or systematic error, but that is another question. That is, the logic of the paragraph does not hold.

Finally, I must be bold and ask if really all authors have actively contributed to the manuscript? I ask this as the short summary just uses "I" and "my", indicating that just the first author has been active. My standpoint is that an active participation is required for co-authorship (i.e. persons should not be added just for politeness).

———————————————