Atmospheric
Measurement
Techniques
Discussions

Open Access

EGU

AMTD

Interactive
comment

Interactive comment on "Comparison of ground-based and satellite measurements of water vapour vertical profiles over Ellesmere Island, Nunavut" *by* Dan Weaver et al.

Anonymous Referee #2

Received and published: 1 November 2018

Review of "Comparison of ground-based and satellite measurements of water vapour vertical profiles over Ellesmere Island, Nunavut" by D. Weaver et al.

Overview

The goal of this manuscript is to quantify and present the biases between satellite retrievals of water vapor (WV), ground-based WV measurements by FTIR, and balloon borne RH measurements by Vaisala RS92 radiosondes. The study is limited to a single location in the Canadian Arctic where the FTIR and RS92 measurements are performed.

General (Major) Comments

Printer-friendly version

Discussion paper

Figures 6 and 9 appear to be identical. It is impossible that they can look exactly the same given what they are meant to show and the obvious differences between the 125HR and RS92 profiles in Figure 5. Also, values stated in the text for specific satellite-RS92 differences don't match up with what's shown in Figure 9. See specific examples below for pages 15 and 16. Finally, Figure 9 shows difference profiles for MIPAS and SCIAMACHY vs RS92 while the text in Section 3.2.4 explicitly says that no MIPAS or SCIAMACHY measurements were coincident with radiosondes. As Figures 6 and 9 are the most important Figures in this paper, it became impossible to continue my review past page 15. My hope is that the authors not only include the correct Figure 9 in the next version, but also take to heart the remainder of my comments and those of the other reviewer(s) that will improve the paper.

I think there are also problems with some of the mean bias values in Tables 2 and 3. For example, for the MIPAS IMK retrievals (v5 and v7) at 12 km in Table 2. The mean difference from the 125HR is given as -0.3 ppmv and -1.4%. If the biases that produce these values are normally distributed, they imply that the mean MIPAS retrieval at 12 km is between 18 and 25 ppmv (-0.25/0.014 and -0.35/0.014). This is way too wet for stratospheric air, and is 3 to 4 times the mean MIPAS IMK retrieval at 12 km (approx. 6 ppmv) shown in Figures 6a and 9a. Another example of this problem is found in Table 3. Mean bias values for AIRS vs RS92 at 12 km are -2.0 ppmv and +5.2%. How can the mean absolute bias (ppmv) and mean relative bias (%) be of opposite signs if the biases are normally distributed? Either there are errors in the mean values presented in these Tables or the distributions of the differences that produce the mean biases are very skewed. If the former, please double check the Table values and make corrections. If the latter, quantifying the biases using Gaussian statistics (i.e., mean and standard error of the mean) is not warranted.

Correlation coefficients and correlation plots are of limited quantitative value in a paper focused on measurement biases between pairs of instruments. Two sets of measurements can be well correlated even though there are huge biases between them! Cor-

relation plots can show biases, but only qualitatively, so consider if the three Figures with correlation plots reveal any quantitative information not already revealed by the profile differences and/or time series of differences. If the correlation plots are deemed unnecessary (my opinion), some (if not all) of the Supplemental Figures could become part of the main manuscript. Please see my specific comments below for Page 15 Line 1 (P15 L1).

The Introduction describes the importance of water vapor in the UTLS and how accurate measurements of WV in the global UTLS are needed. The focus of the paper therefore seems drawn towards WV measurement biases in the UTLS. But this focus becomes lost when you start to compare WV measurements at altitudes as low as 1 km. Why do you apply the same spatial and temporal coincidence criteria to the stratospheric and lower tropospheric data even though the spatiotemporal variability of WV in these regions is very, very different? My advice is to focus this paper on the crucial UTLS region and leave out or downplay the lower tropospheric comparisons.

General Comments (through page 15)

P2 L20 what exactly does "modest vertical resolution" mean? Please be more quantitative here. The vertical resolution of FTIR measurements is very important information for this paper that compares satellite retrievals to the FTIR measurements.

P2 L20-22 Radiosonde humidity sensor measurements also require substantial corrections for solar radiation effects, calibration biases and slow response times in the cold UTLS. It surprises me that frost point hygrometers and lidars are not mentioned here even though the current global coverage of frost point hygrometer sounding sites is starting to surpass the coverage of FTIRs.

P2 L28 "assessing the accuracy and quality" - what does quality mean here if not accuracy?

P3 L1 I believe UT WV measurements will also be compared, not just those in the

stratosphere and lower mesosphere.

P3 L26 move lat/lon to L20 (description of Eureka location)

P4 L23 why is the humidity sensor "no longer able to report a meaningful value"? Is it the cold ambient temperature? Is it the low number density of WV? The solar heating effects on the sensor? Please be more specific.

P5 L2 why describe the Miloshevich et al. (2009) limits when Dirksen et al. (2014) improves the correction algorithms and expands the upper altitude limits of "meaningful" RH measurements by the RS92?

P6 L20 this would be a good place to mention the vertical resolution of the MUSICA FTIR WV profiles

P7 L20 "Correlations between ... were observed to be greater than ..." Why are correlations important in this inter-comparison? Two data sets can be extremely well correlated, even when there is a very large bias between them. Correlation is not a good measure of the agreement between two data sets.

P8 L2 what is the vertical resolution of ACE-MAESTRO WV retrievals in the UT and LS?

P9 L11 what is the vertical resolution of Aura MLS WV retrievals in the UT and LS?

P9 L25 what is the vertical resolution of Aura TES WV retrievals in the UT and LS?

P10 L22 Stiller et al. (2012) compared MIPAS with many types of WV instruments including frost point hygrometers, lidars, microwave radiometers and an FTIR, not just the CFH.

P10 L25 "suggest" and "might be" are very waffly terms. Are there 20-40% biases or not?

P11 L5 Weigel et al. (2016) also compared *SCIAMACHY* v3.01 (not MIPAS v3.01 as

written) to in situ instruments made from balloons (FPH) and aircraft (FISH), not just other satellite retrievals.

P12 L2 Closest in time or space? How did you determine the time stamp for FTIR spectra, which are often co-added for minutes or hours? Also, radiosondes reach 10 km about 30 minutes after they are launched, so how did you set the timestamps for the RS92 profiles?

P12 L9 if the results of comparisons using the closest satellite profile are similar to the results using all coincident profiles, why do you need to show the latter in Supplemental Tables?

P13 L15-18 "... effectively synthesizing a narrow weighting function, then is possible from any one channels. We use of the width ... to estimate a Gaussian smoother generally overestimates ..." These sentences are very poorly constructed. Please fix them.

P14 L4 Above, you stated that the FWHM approximates the vertical resolution of the measurement. So why then do the weighting functions for MLS have a FWHM or 1.0 km when the vertical resolution of MLS retrievals is more like 2-3 km?

P14 L22 Are the 8% and 6% mean differences significantly different from zero? In other words, what are the standard errors of these mean values? It they are not statistically different from zero I would hesitate to call them "biases" because you have no evidence that they are real biases, just mean differences that may equal zero.

P14 L28 I can't see any ACE-FTS differences between 6 and 9 km in Figure 6b that exceed 9 ppmv, so why do you say "was within 11 ppmv"? Also, why report differences for this altitude range when they change from negative to positive at 7 km then become much smaller (in ppmv) and consistent (in ppmv and %) at 8 km and above?

Figure 6 I suggest using fewer red and purple curves, as they are difficult to tell apart. Replace some of them with green, orange and gray. Also, I am guessing that you

discuss satellite-125HR mean differences at 6.4, 8.0 and 9.8 km because these are the altitudes of 125HR retrievals?

P15 L1 and Figure 8 I don't see the value of the correlation coefficients or the correlation plots. The focus of this paper is biases. Correlation coefficients can be near unity when biases between instruments are huge! The correlation plots reveal only qualitative information about biases. For example, the linear fits to ACE-MAESTRO vs 125HR show really awful correlations and essentially no quantitative information about biases. The AIRS panels show good correlations and (qualitatively) that AIRS is biased low at 6.4 and 8.0 km because most of the differences lie below the 1:1 line. What does this Figure (and Figures 11 & 12) show that the vertical profiles of mean differences and time series of differences don't show?

Figure 9 I cannot find a single difference between this Figure and Figure 6, even though they are meant to be showing differences from the RS92 sondes and 125HR, respectively. The two Figures appear to be identical, even when printed, stacked, and held up to backlighting. Are you sure Figures 6 and 9 are actually showing what they are intended to show? The only way they can be exactly the same is if the RS92 and 125HR mean differences are very close to 0 ppmv and 0%, which they are not (Figure 5). The mean differences presented in the text (P15 L7-8) and in Figure 9 do not agree. I suspect Figure 6 appears a second time as Figure 9 in this manuscript.

P15 L19 Your statement here "scatter around the zero line" contradicts what you just concluded, "a dry bias of approx. 10%". The dry bias in ACE-MAESTRO vs 125HR is apparent in Figure 7, so the "scatter" is not "around the zero line" as stated, otherwise there would be no bias.

P16 L10 "Differences as large as 13% are observed between 8 and 14 km." The suspicious Figure 9 shows no relative differences (AIRS-RS92) exceeding 5% between 8 and 14 km.

Unfortunately I cannot continue my review past this point because half of the state-

ments in this very important section are about the biases shown in Figure 9, which is
not really Figure 9.

Printer-friendly version

Discussion paper