Review of manuscript amt-2018-161 “Experimental total uncertainty of the derived GNSS-integrated water vapour using four co-located techniques in Finland” by Ermanno Fionda, Maria Cadeddu, Vinia Mattioli, and Rosa Pacione.

General comments

This study compares IWV estimates measured with 4 different instruments in Finland over a period of 9 months. Three of the instruments (2 microwave radiometers and a radiosonde system) are collocated and the fourth one (GPS) is distant by 20 km. The paper presents a statistical analysis of IWV differences and discusses the results in terms of inter-system biases and random errors including “representativeness errors”. The study reports a dry bias in the radiosonde data during day, a bias source in the 2-channel radiometer data, and a tendency in the GPS IWV differences to increase when IWV increases. The latter effect is explained as “representativeness errors” due to the 20-km distance between the GPS receiver and the other systems. Emphasis is made on this result but two other factors are expected to contribute as well to the observed results: differences in the sensitivity and volume of sensed atmosphere between GPS and the other techniques and instrumental errors specific to each technique. Since these factors are not quantified, it seems difficult to assess the third one. A rigorous way of evaluating the impact of the distance between remote techniques requires using identical instruments at both ends. The discussion and conclusions of this study are impaired by this deficiency. To overcome this major weakness, the authors could consider the following options:

(a) The differences in IWV due to a 20-km distance can be evaluated from two identical instruments (e.g. a pair of GPS receivers) separated by approximately this distance in the same climatic region;

(b) The differences in IWV due to the different sensitivity and volume of sensed atmosphere between GPS and other techniques can be evaluated from co-located measurements (e.g. from past campaigns).

Specific comments

Give a clear definition of the “representativeness errors” in the introduction (Page 2 Line 27). Strictly speaking, it should include the differences in the sensitivity and volume of sensed atmosphere between GPS and other techniques and the instrumental errors specific to each technique (see e.g. Buehler, AMT, 2012).

The MWR retrieval algorithm used radiosonde data from a station located 90-km apart. Could there be a “representativeness error” impacting the results? Did you compare the operational radiosonde observations to those from the BAECC campaign used in this study?

Can the linearity limitations observed with the 2-channel radiometer be explained/corrected? A temperature dependence is mentioned in the conclusions section though it is not discussed earlier. Please discuss this point in the text.

Give more details about the BAECC radiosonde data. What is the vertical resolution of the profiles? How are they quality-controlled? Is any data processing/filtering or bias correction applied?

Since the radiosonde measurements were made with a RS92 sonde type, did you consider applying the GRUAN bias corrections?

How are the IWV data adjusted for the differences in altitude? (P5 L21)

Why is the GPS conversion factor (pi) used as a monthly mean value and not from a linear regression with 2-m temperature following the method of Bevis et al. 1992? The uncertainty in the GPS IWV estimates could probably be reduced applying the Bevis method or time-varying Tm estimates from a NWP model.

It seems to me that your GPS IWV data contain outliers. Figure 4 and Figure 7 show large deviations in the IWV differences which involve the GPS data. The outlying values should be removed from the analysis.

The central description and discussion of results in section 3 should be supported by a statistical analysis including specifically the contributions from atmospheric variability (temporal/spatial) and the different sources of differences (representativeness, etc.) between measurements (see e.g. Ning et al., AMT, Vol. 9, p79-92, 2016).

According to your interpretation of the results, you assume (implicitly) that the differences are random and independent (e.g. between GPS and MWR3C). Can you bring evidence of this?

It is not clear how you estimate the representativeness errors contribution (P8 L12-14; P9 L17-19) Given the above assumptions, the standard deviations should be added/subtracted quadratically.

Figure 8 left panel: it might be more interesting to plot the SD rather than the RMSD to support the above discussion.

In the conclusion section it is stated that the comparison of the two radiometers gives confidence in the "stability" of these measurements but this point is not discussed earlier in the paper. Can you elaborate a bit more on this?

Technical comments

P1 L25 and 26: you refer incorrectly to IWV "variability" above or below 20 kg/m2 in place of simply "IWV". Please correct.

P6 The quantity defined by equation (5) should be referred to as the "mean difference"

P7 L23-24: the comment about the log-normal PDF is not used later, this sentence can be removed.

P7 L25: the comment on the agreement between measurements is elusive. Please remove it.

P8 L9: the GPS and MWR3C are "unbiased" maybe you want to write "have a mean different close to zero". To prove that an instrument is unbiased requires to compare its observations to a calibrated reference standard. However, you may argue here that the instruments provide independent observations which may have specific biases, however it is unlikely that their biases would perfectly cancel out, so it is much more likely here that they are both unbiased.

P8 L15: "higher variability in summer": this is not obvious from Figure 9 right panel.

P8 L16-19: you suggest to include a "representativeness" error of 1-2 kg/m2 to account for the spatial variability of GPS measurements. How would you do that? Shouldn't this error depend on the distance to the GPS measurements?

P8 L20-23: the comment of the bottom panel of Fig. 6 doesn't add any further information about the radiosonde bias at 11 UTC. Remove this comment as well as the bottom panel of Fig. 6 which shows merely a combination of the lines plotted in the upper panel.

P9 L21-31: these two paragraphs are not useful in the conclusion section.

Figure 3: the series are superposed and cannot be distinguished. Maybe show only one series to illustrate the IWV variations over the study period.

Figure 4: it is common practice to draw the common reference data set (here GPS) on the x-axis. Please change the 3 plots.

It may be useful to add a scatter plots of RAOB vs. MWR3C and MWR2C vs. MWR3C.