Response to interactive comment of Referee #2 on "Experimental total uncertainty of the derived GNSS-integrated water vapour using four co-located techniques in Finland" by E. Fionda et al.

Anonymous Referee #2
In black => referee observations
In red => our response

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).

We thank the reviewer for this thoughtful comment. We agree that the effect of different volumes sampled by the instruments are likely to be one major source of uncertainty and were not properly discussed in the work. For this reason, we have substantially reviewed the analysis and added a new subsection (3.1) with a discussion of the different volumes of atmosphere sampled by the three techniques. Table 2 and Fig. 3 were also added in which the different field of views are specified

With regard to the two options proposed by the reviewer, option (a) is difficult to implement because GPS networks are sparse, with distances of the order of a hundred kilometers. To try to address the second option (b) we added in Table 3 previous findings from studies where GPS, radiometers, and radiosondes were collocated. Unfortunately, a GPS receiver is not part of the AMF2 and therefore it was not co-located during this campaign.

We think the additional analysis and discussion make the paper stronger and we thank the reviewer for the suggestions.

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).

We agree with the reviewer. In the revised version we specifically refer to variability in the field of view and variability due to the fact that the instruments are not all strictly co-located.

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?

Retrieval coefficients are usually derived for convenience from a radiosonde ensemble in a representative climatic region. However, they could as well be derived from a set of simulations (for example temperature and humidity profiles from a regional model). As long as the broad climatic region is representative of the radiometer's location the retrieval coefficients should be well defined. Because we used more than 10 years of radiosondes data, the regional atmospheric variability should be well captured. In the figures below the vapor density and temperature profiles of the training dataset (Jyväskylä, ~8000 profiles, blue) are plotted together with the BAECC profiles (~800 profiles, red). The climatology of the training Jyväskylä dataset is well suited to represent the AMF2 location. Specifically, in the year 2014 the mean IWV in Jyväskylä was 13.1 kg/m2 vs. 13.6 kg/m2 in Hyytiälä.
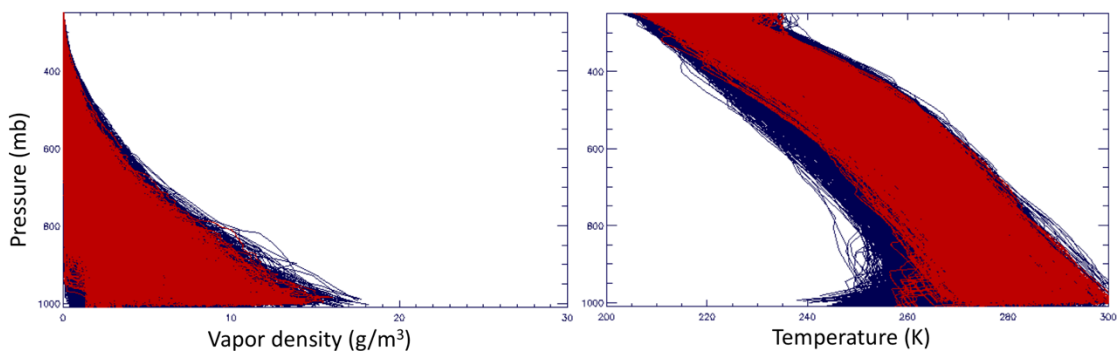
Can the linearity limitations observed with the 2-channel radiometer be explained/corrected? A temperature dependence is mentioned in the conclusions section though it is not discussed earlier. Please discuss this point in the text.

We added a discussion to this point in the text. It is our opinion that the slight temperature dependence observed in the 2CH data is due to the fact that the receiver of WVR1100 models are stabilized to 0.2 K (compared to the 30 mK of the 3CH). The antenna subsystem including the feedhorn is not thermally stabilized. Although a correction for residual temperature dependence of the parts that are not thermally stabilized (lens, feedhorn and waveguide) is routinely applied, there may be small residual dependencies.

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?

The RS92 radiosonde data used for the BAECC campaign were downloaded from the ARM Archive. It is our understanding that the data are processed with the standard Vaisala-provided algorithm and no additional corrections are applied. The vertical resolution is about 10 m and usually data are transmitted up to an altitude of 20 km. Standard quality control flag are applied and data quality

reports are issued to flag data that are considered incorrect. We added 2 references (Peppler at al, 2008, 2016) that explain in detail the quality control process at the Archive. We added these details in the text at the end of section 2.1.

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

Although we are aware of the GRUAN program we are not familiar with the details of the GRUAN correction, therefore no ad-hoc corrections were attempted. It would be interesting to see, though.

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

This sentence was left over from a previous analysis were radiosondes from the station of Sodankylä were used to derive the coefficients. Sodankylä being at 180 m elevation, the radiosondes were extrapolated to the radiometer's elevation before deriving the coefficients. However, the Jyväskylä radiosonde station (which is the nearest available) is approximately at the same elevation as the GPS and radiometers stations at Orivesi and Hyytiala respectively. The elevations of the 3 sites were double checked using data from the NASA/ASTER Global Digital Elevation Map that has a quoted accuracy of 15 m and were found to agree with each other in within 10 meters. Therefore, no corrections were applied.

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.

The methodology applied in this work may actually reduce the uncertainty in the GPS IWV estimate when accurate temperature measurements are not available at the GPS location. Bevis' uncertainty estimate of about 2% was mostly attributed to uncertainty in the calculation of the weighted mean temperature of the water vapour which is dependent on surface temperature.

P. Basili et al (2001) proposed an alternative statistical approach, in which monthly average values of π were calculated from a historical database of radio-soundings. The estimated relative uncertainty for this approach was also in the 1.6 - 2 % range.

In absence of the surface temperature data at the GPS station and, because we are not sure of the uncertainty of surface temperature in NWP models the statistical approach is preferred in this specific case. Supplemental material is provided at the end of the responses to support this choice.

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.

Although we agree with the reviewer that there are a few points that display higher scatter, we have no objective way to remove those points without introducing some arbitrary criteria. Following the suggestion of the first reviewer Fig. 3 (now Fig. 4) was modified to show the time series of the differences between the water vapor measurements. The largest discrepancies are during June and July when we expect conditions of increased cloudiness and variability. Rather than removing the points we have added a brief discussion in section 3.2.

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).

To address the comment, the following changes were implemented: 1) Added section 3.1 where the different volumes sampled by the systems are discussed. Added Table 2 where the FOVs (fields of view), the estimated total uncertainty of each instrument from error propagation, and the variability of each instrument while looking at a night scene with little water vapor variability are reported.

Fig. 3 was added which shows pictorially the volumes sampled by the systems. Although it is difficult to separate the contribution to the differences due to the FOVs and to the distance between instruments we provided some support to our conclusions that the distance of 20 km is contributing to the increased variability by adding in Table 3 findings of previous studies where similar instruments were co-located. In the 3 studies reported (Basili, Liljgren, Mattioli) the standard deviation of the differences between the GPS and the radiosondes (or radiometers) was similar to the one between the sondes and the radiometers which supports the claim that the much larger variability between the GPS and the other instruments found in this study is due to the physical distance between the sensors.

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?

Proving randomness and independence is not easy, we base our assumption on the fact that the instruments are independent and measure water vapor based on very different physical principles, however they all are measuring a quantity that is highly correlated in space and time.

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.

We have revised the definition of representativeness error in the manuscript and modified the discussion. Because it is not possible to entirely separate the single contributions to the uncertainty, we have rephrased the conclusion to refer to a total uncertainty that includes a contribution from the FOV and a contribution due to the distance from the GPS.

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

Done and as a consequence the definition of RMSD (eq 7) was removed as this quantity was not used anymore. In the revised manuscript Figure 8 -> Figure 9

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?

We have not observed any drifts or dependency of the mean differences on time. For this reason, we think all the instruments performed nominally throughout the campaign. Following the comment from Reviewer #1 this sentence was deleted from the revised version.

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. -Done

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

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

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

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.

We agree with the reviewer comment. What we really meant is that they have no bias relative to each other as we can't prove that they are truly unbiased.

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

We thank the reviewer for this comment. Fig. 8 was not showing the correct data. We replotted the figure (now Fig. 9) that now shows the higher variability of the data with increasing water vapor. From Fig. 4 it can be seen that the higher variability occurs in June and July.

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?

Yes, we have rephrased. In this case it seems that because of the distance between the instruments the different volumes sampled by the systems resulted in larger variability of the differences.

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.

If the reviewer has no objection we rather remove the top panel and keep the bottom panel of Fig 6 that shows the bias compared to all systems. In the revised manuscript, Figure 6 -> Figure 7

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

-Removed

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

-Done. In the revised manuscript Figure 3 -> Figure 4

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

-Done

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

Supplemental information on the conversion of ZWD to IWV using a statistical conversion factor π

It is possible to retrieve the IWV from ZWD calculated by GNSS observation using the well-known following relationship:

$$IWV = \pi * ZWD \tag{1}$$

The conversion coefficient π is a "constant" of proportionality which depends on several physical parameters, among which the weighted mean temperature of the water vapor (Tm):

$$Tm = \frac{\int (e/T)dz}{\int (e/T^2)dz} \tag{2}$$

$$\pi = 10^6/\rho_l R_v \left[\left(k_3/T_m \right) + (k_2 - mk_1) \right] \tag{3}$$

where ρ_l is the density of liquid water, R_v is the specific gas constant for water vapor, m is the ratio of molar masses of water vapor and dry air, k_1, k_2, k_3 are constants (77.6036, 64.79, 3.77x10^5), T is the air temperature, e is the partial pressure of water vapor, and dz has units of length in the zenith.

Computation of the conversion coefficient
Eq. 1 assumes that the Wet Delay is entirely due to water vapour and liquid water traveled by the GNSS signal to reach the ground receiver [1]. Bevis et al 1994 [2] showed that in most practical conditions the uncertainty of π is mainly due to Tm. When Tm is predicted from surface temperature on the basis of a linear regression, Bevis estimated a relative error in π of the order of 2%.

P. Basili et al (2001) [3] proposed an alternative statistical approach, in which monthly averaged π values were calculated from a "historical" database of radio-soundings. This approach was developed to overcome the lack of surface temperature measurements at many GPS sites. The approach was tested using a GPS from Cagliari, Italy. The first two columns of Table S1 report the computed values of π with associated uncertainties computed for each month. It can be noticed that the relative error (Std. Dev., %) of the monthly averages ranges between 1.14 % (October) and 1.61% (January), and using a constant value of π = 0.152 (yearly value) produces a relative error of 2.15 %.
In the same work π was computed following the Bevis approach by computing Tm with a linear regression on surface temperature. The relative error of π in this case was equal to 1.33%, which is in the same range (1.14 % - 1.61 %) of the errors obtained by the statistical approach (see Table S1).

Due to the lack of surface temperature data at the GPS station, the statistical approach proposed in [3] was adopted in this work. The performance of the chosen approach was analyzed looking at two long-term RAOB Finnish databases in Sodankylä (2000-2014) and Jyväskylä (2003-2014). Supplemental Table S1 reports the computed statistical monthly conversion coefficient (π) values with associated uncertainties at the two stations. The results are similar to those for Cagliari with average monthly relative error ranging between 1.24 % and 1.76 % in Sodankylä (2.72 % when using all year values) and between 1.15% and 1.83 % in Jyväskylä (2.77 % when using all year values). Supplemental Figure S1 shows the seasonal dependence of the conversion coefficient at the 3 sites.

Based on these results a statistical approach can provide good estimates π when climatologic information is available and surface temperature measurements are not available at the GNSS station.

Table S1 – Statistical conversion coefficient (π) values for Cagliari [3], Sodankylä, and Jyväskylä

| Months	Cagliari (I)		Sodankyla (FIN)		Jyvaskyla (FIN)	
	π	Std. Dev. (%)	π	Std. Dev. (%)	π	Std. Dev. (%)
1	0.149	**1.61**	0.147	**1.62**	0.148	**1.74**
2	0.148	**1.48**	0.147	**1.69**	0.148	**1.83**
3	0.149	**1.39**	0.147	**1.60**	0.149	**1.80**
4	0.150	**1.23**	0.150	**1.52**	0.151	**1.44**
5	0.152	**1.20**	0.152	**1.76**	0.154	**1.68**
6	0.154	**1.29**	0.154	**1.53**	0.156	**1.44**
7	0.156	**1.17**	0.157	**1.24**	0.158	**1.15**
8	0.156	**1.24**	0.156	**1.27**	0.157	**1.26**
9	0.154	**1.22**	0.154	**1.38**	0.155	**1.26**
10	0.153	**1.14**	0.151	**1.62**	0.153	**1.54**
11	0.151	**1.51**	0.149	**1.54**	0.151	**1.62**
12	0.149	**1.42**	0.148	**1.66**	0.149	**1.67**
All years	**0.152**	**2.15**	**0.151**	**2.72**	**0.152**	**2.77**
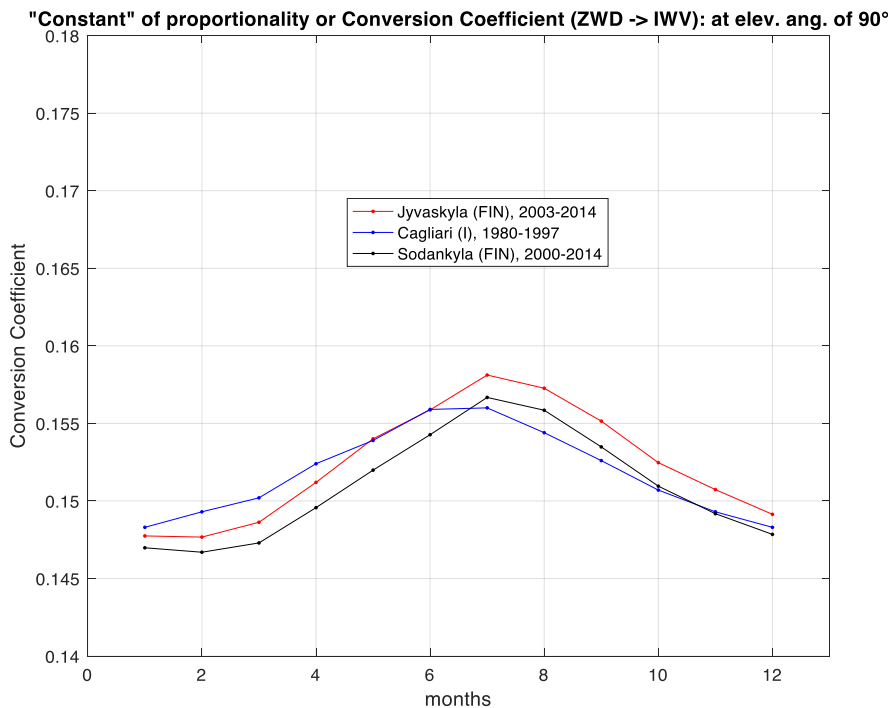
Fig. S1 – Monthly mean of the statistical conversion coefficient (π), calculated at black: Sodankylä, red: Jyväskylä, and blue: Cagliari.

References
[1] J. Duan, M. Bevis, P. Fang, Y. Bock, S. Chiswell, S. Businger, C. Rocken, F. Solheim, T. Van Hove, R. Ware, S. Mclusky, T.A. Herring, and R.W. King, "GPS meteorology: Direct estimation of the absolute value of precipitable water", J. Appl. Meteorol. vol. 35, pp. 830-838, 1996.
[2] M. Bevis, S. Businger, S. Chiswell, T. A. Herring, R. A. Anthes, C. Rocken, and R. H. Ware, "GPS meteorology: Mapping zenith wet delays onto precipitable water," J. Appl. Meteorol., vol. 33, pp. 379–386, 1994.

[3]   P. Basili, S. Bonafoni, R. Ferrara; P. Ciotti, E. Fionda and R. Ambrosini, "Atmospheric water vapor retrieval by means of both a GPS network and a Microwave Radiometer during an Experimental campaign in Cagliari, Italy, in 1999", IEEE Geoscience Remote Sensing, vol. 39, N.11, pp. 2436-2443, Nov 2011.

[4]   M. Bevis, S. Businger, T. A. Herring, C. Rocken, R. A. Anthes, and R. H. Ware, "GPS meteorology: Remote sensing of atmospheric water vapor using the Global Positioning System," J. Geophys. Res., vol. 97, pp. 787–801, 1992.

[5]   O. T. Davies and P. A. Watson, "Comparison of integrated precipitable water vapor obtained by GPS and radiosondes," Electron. Lett., vol. 34, pp. 645–646, 1998.