Dear Editor,

Attached please find the revised version of our manuscript #amt-2018-279 where the marked-up manuscript version is also provided in order to show the changes made. All responses are given in italic and green font.

We appreciate the work of the referees, their questions and their suggestions for improvements. They have been adopted and/or further discussed as described in the following.

We first describe the overall changes in the revised manuscript and then we give responses to the individual comments from the referees.

1. We did additional work in order to investigate the reasons of the different trend values obtained from the GPS elevation 10° and 25° solutions. We estimated the IWV trends using the GPS data under three scenarios, i.e., (1) using the original GPS data without corrections for interventions, (2) corrections carried out for known interventions, i.e., due to known hardware changes, which are shown in Table 2, and (3) corrections also applied for unknown interventions which were detected, by a PMTred test, for the GPS 10° solution. A new Section (3.2.2) was added in order to have a discussion on all the unknown interventions detected by the PMTred test and the new Table 5 was provided to describe the corresponding mean IWV differences caused by the interventions. In addition, the root-mean-square (RMS) difference and correlations between the trends obtained from the GPS solutions, for the different scenarios, and the ones given by radiosonde and ERA-Interim data were summarized in Tables 7 and 8. Section 4.3 (IWV trend comparison) has therefore been significantly modified with all the new results.

2. Concerning that the results in terms of the effect on the trend comparisons were insignificant, and that only results for two GPS sites were presented in the original manuscript, all the parts that were related to (1) the tests using the alternative mapping functions and (2) the tests including second-order ionospheric corrections were removed.

3. The test of the use of elevation dependent weighting is now carried out on the entire dataset of 13 GPS sites.

4. As suggested by one of the referees, the old Figure 4 was removed from the revised manuscript. In addition, the old Figures 8 and 9 were removed and the statistics for the test using elevation dependent weighting (or not) are now presented in in the new Table 10.

Sincerely yours

T. Ning and G. Elgered

RESPONSES
Anonymous Referee #1

MAJOR COMMENTS
This observation immediately brings me to the largest drawback of the current study: the lack of scientific novelty. Although the manuscript is well written and the analysis is really done very carefully and detailed, taking into account all the related issues, the added value of this study with respect to the earlier Ning & Elgered (2012) results is really minor. Moreover, the authors also do not argue enough why there is a need for this follow-up study and what the major improvements/changes in the methodology of this study are, compared to the earlier one. This is rather disappointing. In the conclusions of both papers, the authors themselves give a hint of what a meaningful follow-up study would be: "It is important to carry out similar studies for other sites and especially from areas with different climates. Furthermore, the optimum cutoff angle (25) for the trend estimation may be different for GPS sites in different electromagnetic environments and sites at lower latitudes, where the distribution of observations as a function of elevation angle is different" (Ning et al., 2012). Especially taking into account that one of the authors has already done IWV time series analyses on a larger dataset (Ning et al., 2013, Ning et al., 2016b), it should be argued strongly in the manuscript why their suggestion has not been followed in this study.
Response: Thank you for the comments and suggestions. It is true that the old paper (Ning et al., 2012) had more novelty, on the other hand an important part in scientific research is also to assess and verify earlier findings. Our motivation with this paper is to go into detail rather than extending the area studied to include stations where we do not have the same knowledge about the stability of the stations. It therefore makes sense to us to use the same stations, but longer time series, plus another reference time series for the comparisons, i.e., the ERA-Interim from ECMWF. Although they are not fully independent from the RS data they are different. Other differences are that we now also investigate the impacts of introducing elevation dependent weighting and correcting for interventions in the GPS IWV time series.

In addition, our suggestion in Ning et al., (2012) to study other stations/regions was not just an announcement of our own possible future activities, but intended for the interested community. In the revised paper, we point out that a similar study was carried out by Baldysz et al. (2018) where they estimated the IWV for a time period of 20 years (1996-2015) using the data acquired from 20 GPS stations in the EUREF Permanent Network (EPN). They found a better agreement with nearby radiosonde-derived IWV trends for the 20° cutoff solution when compared to a 5° cutoff solution.

The rather small dataset of only 13 sites, with rather similar IWV field properties I assume (at least belonging to a similar geographical region), weakens the significance of the analysis and possible conclusions: if correlations are based only on 13 points (Figure 7) or even on 8 points (Figures 8 & 9), is it, from a statistical point of view, meaningful to ascribe differences in the correlation coefficients between trend estimates to different elevation cut-off angles used, differences in datasets (RS vs ERAInterim), elevation-angle-dependent weighting? Are the differences in the correlation coefficients statistically significant to draw the conclusion that "In fact when compared to the trends obtained from the radiosonde and the ERA-Interim (only in this study) data a higher correlation coefficient and a lower RMS differences are seen for the 25 solution. Therefore the high cutoff angle is desired to be used in order to estimate the long-term trend in the IWV" (page 12, lines 379-380)?
Response: As described in overall changes, we now estimated IWV trends using the GPS data under three different scenarios in order to investigate trend differences in detail. Based on the new results, we modified the conclusion in the revised paper as: "using different elevation cutoff angles is a valuable diagnostic tool that can be used for validation purposes and detection of possible site problems, such as multipath impacts. When we use the GPS data to monitor the long-term change in the IWV, e.g., as linear trends, it is recommended to apply at least two different elevation cutoff angles in the data processing. Ideally the IWV trends obtained from the two significantly different elevation cutoff angle solutions should be the same if there are no significant long-term changes in the multipath impacts, or any other elevation dependent phenomena that affects the observations."

Also based on the results in no case, or for no site, we found a trend that shows an obvious error when compared to the radiosonde and ERA-Interim data when the higher cutoff angle is applied suggesting that if there is a large difference in estimated trends from two different cutoff angles, the use of the result from the higher cutoff angle is more representative since systematic effects are less probable. This is also explicitly stated in the conclusion of the revised paper.

A last major concern is that the manuscript needs more interpretation for some fundamental statements or issues. I will give some more examples in my minor comments here below, but I already want to highlight two of them here: first, in Section 4.1 ("IWV intercomparison"), the discussion immediately starts with describing the standard deviations of the differences between IWV GPS and IWV RS/ERA-Interim, but no description or explanation is given of the IWV biases between GPS and RS, GPS and ERA-Interim, GPS elev 10 and GPS elev 25 (and RS and ERA-Interim, see below). Secondly, in Section 4.3 ("IWV trend comparison") , describing Fig. 6: give an explanation about the difference of the trend estimations between the 10° and 25° solutions for (i) both white noise + time correlated behavior separately and (ii) between the white noise + time correlated behavior (if there are any, difficult to see on lower panel of Fig. 6 for majority of the sites).
Response: In the revised paper, text was added in order to discuss the results of IWV inter-comparison in terms of both bias and standard deviations. In addition, the results and corresponding discussions from the comparison between two GPS elevation solutions, and RS and ERA-Interim were also added. In terms of trend uncertainties a few more sentences were added in order to address the issue brought up by the referee.

MINOR COMMENTS

* Given the fact that the sample of sites is rather geographically constrained, I would add "at Swedish and Finnish sites" (or "at high-latitude sites", or "in a sample of Scandinavian sites") to the title.
Response: The title of the revised paper is changed to "Trends in the atmospheric water vapour estimated from GPS data acquired in Sweden and Finland and analyzed using different elevation cutoff angles".

* Page 3, lines 65-67: here, you should really point out what the added value of this study is compared to Ning and Elgered (2012) and why there is a need for this follow-up study.
Response: See the response to the major comments.

* Page 5, lines 141-144: why do you only select 2 stations for the test using the alternative mapping functions and for the test including second-order ionospheric corrections? Will this be significant enough to draw conclusions? What will be the sample bias for the conclusions?
Response: In the revised paper the results for the test using the alternative mapping functions and for the test including second-order ionospheric corrections are removed.

* Page 5, lines 145-148: again, why did you not consider the entire dataset of 13 sites for processing the data with an elevation dependent weighting function?
Response: We now have included all 13 GPS sites for the test of including an elevation dependent weighting function. The updated results are presented in the revised paper.

* Page 6-7: Section 2.3: here, you should add that radiosonde data are assimilated in ERA-Interim and that these are not completely independent datasets, when used as references (as is done in this study). This was a component that I was missing in the analysis: how do the ERA-Interim and radiosonde IWV measurements/trends compare? Is there a need for having two reference datasets in this study?
Response: For ERA-Interim, data acquired from many observing systems (i.e., numerous satellite instruments, radiosondes, buoys, and other components) were assimilated. Radiosonde data are a fraction of data assimilated into ERA-Interim. In the revised paper, two sentences are added in order to address this issue: "For ERA-Interim, data acquired from many observing systems (i.e., numerous satellite instruments, radiosondes, buoys, and other components) were assimilated. Even though radiosonde data are a fraction of data assimilated into ERA-Interim, there are still many other observing systems contributing to the entire dataset. Therefore, we use ERA-Interim as a second reference dataset for the IWV comparisons with GPS data."

We think that the two datasets shall be used because it is not obvious that one is superior to the other. In general, one could assume that the ERA-Interim shall be more accurate but if accurate radiosonde data are acquired within a certain distance to the GPS site it may be that they are more representative.

* Page 6, line 186: which version of ERA-Interim are you using? I thought the horizontal resolution of ERA-Interim is about 80 km? Please specify!
Response: It was a mistake. The horizontal resolution of ERA-Interim is 80 km and the corresponding correction has been made in the text.

* Page 6, line 188: please give some details about the horizontal interpolation applied by EMOSLIB: weighted? How many pixels are considered?
Response: We added a link: https://confluence.ecmwf.int/display/EMOS/Grid+point+to+Grid+point+Interpolation in the revised paper as a reference for details about the horizontal interpolation.

* Page 6, lines 188-189: please give some numbers of the difference between the model height and GPS antenna height for your sample of sites. This might enlighten why you use an altitude correction between GPS and ERA-Interim, but not between GPS and RS.
Response: The motivation for the altitude correction between GPS and ERA-Interim was because we used the pressure and mean temperature given by ERA-Interim for converting ZWD to IWV for the GPS data. So the altitude correction between GPS and ERA-Interim was motivated by the need to do height corrections for the temperature and pressure. The difference in altitude between GPS and RS is relatively small (see Table 1). The resulting IWV offset will be systematic which has no impacts on the result when comparing GPS trends to the ones given by RS. This is now explicitly stated in the text.

* Page 7, lines 199-206: if you want to mention other studies using IWV time series obtained from ERA-Interim, please use also the most recent ones and certainly those in this Special Issue: Parracho et al., ACP 2018, Pacione et al., AMT 2017, and to a lesser extent Berckmans et al., ACP 2018, Van Malderen et al., ACP 2018 (not peer reviewed yet).
Response: Parracho et al., ACP 2018 and Pacione et al., AMT 2017 are now cited in the updated manuscript.

* Page 7, lines 218: speaking about the launch times of radiosondes: did you found different results when treating the 0 h (nighttime) and 12h (daytime) measurements separately in your analysis, also when compared with GPS and ERA-Interim?
Response: The tests have been done and no significant differences in terms of trend correlations have been found. A few sentences are added in the revised paper to discuss the results.

* Page 8, lines 231-232: did you try the other way around: have you applied the PMTred test on the differences time series between GPS and ERA-Interim and did you find offsets which could not be linked to interventions?
Response: After corrections for the know interventions, the time series of the differences between GPS 10° solution and ERA-Interim data was applied the PMTred test. Three more interventions were detected for three stations which are not related any known changes. Discussions on the corresponding results are added in the revised paper.

* Page 8, lines 243-244: “We found that trends are not affected by which reference period that is chosen” ! please quantify, give a number here.
Response: We modified the sentence to “We found that trends are not affected by which reference period that is chosen (the maximum difference is smaller than 0.01 kg/m^2/decade) whereas the overall mean differences compared to the other techniques will change."

* Page 9, line 282: what is the correlation between the ZTD and IWV trends?
Response: The correlation between the ZTD and IWV trends is 0.94 which is now clearly stated in the revised paper. This value is however not important for the study itself, it just informs us that a large part of the trends in the ZTD is caused by IWV changes and less by long term trends in the ZHD (mean sea level pressure).

* Page 10, lines 311-314: higher trend values are obtained here, i.e. for the longer time period, than in the Ning and Elgered (2012) study. This might be a little out of scope of this paper, but does this fact means that there is an enhanced moistening in your sample of sites? And if so, what might be the cause for this?’
Response: The reason was actually given in the next sentence. In the revised paper, we clarify the issue by modifying the sentence to “the lower consistency between the trends (large SD values and small mean trends) shown in Ning and Elgered (2012) is because of the shorter time period of the data (14 years) meaning that the trend shows more sensitivity to deviations from the model especially in the beginning and in the end of the selected time series.”

* Page 10, lines 323-325: “A higher correlation coefficient (0.9) and a lower RMS difference (0.09 kg/(m2 decade)) are seen for the elevation 25_ solution than the ones (0.53 and 0.18 kg/(m2 decade)) for the 10_ solution”. What is causing those differences, to your opinion? And are these differences significant enough to conclude that the 25 solution should be used for IWV trend estimation?
Response: As described for the overall changes, additional works have been carried out in order to investigate the trend differences in detail. In addition, we modified the conclusion in the revised paper and stressed that the use of different elevation cutoff angles is a valuable diagnostic tool that for validation purposes and detection of possible site problems, such as multipath impacts.

* Page 11, lines 345-347: what is the reason for the slightly worse trend correlations and RMS differences when using GPS data with elevation dependent weighting? And are the differences significant?
*Response: The new results, after including all 13 sites in the test of elevation dependent weighting, show no significant differences seen for the GPS 10° solution while the agreement of the trend become significantly worse when the weighting was applied for GPS 25° solution. The RMS difference increase, by more than 40 %, from 0.14 to 0.22 and 0.20 kg/(m2*decade), respectively when compared to the radiosonde data and the ERA-interim data. The corresponding correlation coefficients decrease, by over 35 %, from 0.77 to 0.56, and 0.80 to 0.59, respectively. Our interpretation is that the elevation dependent weighting will only be useful when the observations have larger impacts from multipath effects using low elevation angles. For the observations*

acquired from higher elevation angles, 25° in our case, the multipath impact is expected small. As a result, the elevation weighting will weaken the contribution of data points just above 25° and the geometry will be weakened even further.

*Regarding the significance of the differences, due to the fact that the uncertainty of the correlation coefficient is limited by the small dynamical range of the trends (and that we only have 13 data points), we instead calculated the uncertainty, or standard error (SE), of the RMS difference in order to give the confidence level of the change in RMS differences. This has now been explicitly stated in the text of the revised paper by the sentence "the largest improvement seen in RMS differences is from 0.20 kg/(m^2*decade) to 0.14 kg/(m^2*decade) which is significant with an approximate 67 % confidence level."*

Anonymous Referee #2

General comments

This manuscript reports on the impact of GPS data elevation cutoff angles and other processing options on IWV trend estimates in Scandinavia. This work is very similar in the concepts, ideas, and methods to a previous publication by the authors (Ning and Elgered, IEEE, 2012). The processing options are evaluated based on the correlation coefficients (and RMS differences) between the GPS trends and radiosonde and ERA-Interim trends at 13 GPS sites. Compared to their earlier paper, differences are with the length of the GPS series (20 yrs compared to 14), the use of ERA-Interim as a second validation dataset, and the test of other processing options (mapping functions, correction of higher order ionospheric effects, and elevation-depending weighting). The longer time series and the use of a second validation dataset yield more statistical confidence into the new results. However, the conclusions remain unchanged and the authors still recommend using a 25° cutoff angle rather than 10° (though only these two cutoff angles are tested in this study) and note that the other processing options that were tested are insignificant. Little new knowledge is brought actually compared to the earlier paper. One or both of following directions should be considered to increase the relevance of this study: 1) investigate the reasons of the different trend values found for the different cutoff angles by inspecting carefully the differences in the estimated IWV time series (are the differences due to drifts in the time series? If yes what could be the reasons? Are they due to multiple offsets due e.g. to documented or undocumented equipment changes?); 2) extend the study to other regions/climates where the sensitivity to cutoff angle and/or the other processing options tested here might be different. One can note that this was suggested by the authors themselves in this manuscript and in their previous publications.

Response: Thank you for the comments and suggestions. It is true that the concept of the current study is similar to the old paper (Ning et al., 2012), on the other hand an important part in scientific research is also to assess and verify earlier findings. Our motivation is to go into detail rather than extending the area studied to include stations where we do not have the same knowledge about the stability of the station. It therefore makes sense to use the same stations, but a longer time series, plus another time series for the comparisons, i.e., ERA-Interim from ECMWF. Other differences are that we now also investigate the impacts of introducing elevation dependent weighting and correcting for interventions in the GPS IWV time series.

As the response to your first suggestion, we carried out additional work in order to investigate the differences in the estimated IWV time series. For example, we examined the possible impacts by the receiver changes. In addition, as described in the overall changes, after corrections for all known interventions in the GPS data we also carried out a PMTred test, for the GPS 10° solution, in order to detect the unknow interventions. The result indicates that the largest improvement was obtained if we applied the corrections on both GPS solutions for the interventions caused by known hardware changes (i.e., antenna, radome and microwave absorbing material). A slightly better agreement with the trends from other data sets, and between the two GPS solutions is seen after a further corrections on the GPS 10° data for the interventions detected by a PMTred test.

Detailed comments

Given the small number of GPS sites used in this study, the statistical significance of the computed correlations and RMS differences is rather small (though not quantified). It seems thus hazardous to draw general conclusions on the choice of the cutoff angle. More insightful analysis is indeed required to convince the readers to use a 25° cutoff angle for trend estimates, especially since the general tendency in the GPS community is to use lower cutoff angles (typically between 3° and 10°) and the IWV comparison (GPS vs. radiosondes and ERA-Interim) shows that the biases and standard deviations increase when the cutoff angle is increased. Moreover, it should be recognised that trend estimates are sensitive to small changes in the mean bias and extremes at the beginning and end of the time series, and thus conclusions based on trend estimates can be tricky. A case by case analysis may help understanding the reasons why trend estimates change between 10 and 25° cutoff angles at some of the 13 sites investigated (e.g. OVE0, OLKI) and may strengthen the conclusions.
Response: As described for the overall changes, after corrections for all known interventions, we applied PMTred test on the GPS 10° solution. When the further corrections were taken on the GPS 10° solution for all unknown interventions detected by PMTred test, the trend agreement between the two GPS solutions is improved. Based on the results we found in no case and for no site the trends drift significantly from the radiosonde and ERA-Interim data when the higher cutoff angle is applied. This is also explicitly stated in the conclusion of the revised paper.

Why are the data not reprocessed for all the cutoff angles, e.g. between 5° and 30° or more, as in Ning and Elgered, 2012?
Response: In the previous study, i.e., Ning and Elgered, (2012) we searched for an optimum cutoff angle and therefore we did many different solutions. As stated earlier our ambition with this study was to assess the previous results by using a longer time series and adding ESA-Interim data, so for practical reasons we choose only two different cutoff angles. Additionally, in the conclusion we do not give a suggestion for an optimum cutoff angle. Instead, based on results, we claim that using different elevation cutoff angles is a valuable diagnostic tool that can be used for validation purposes and detection of possible site problems, such as multipath impacts.

Why is only antenna, radome, and microwave absorber changes considered as GPS interventions? Did the authors check that receiver changes do generate breakpoints? Moreover, in several places in the manuscript, the breakpoints in the GPS series not explained by antenna, radome, and microwave absorber changes are attributed to environment changes resulting in changes in multipath. Firstly, this attribution may be wrong because receiver changes are ignored. Secondly, the attribution to multipath is pure speculation as no additional observation/data/information is provided to support this hypothesis. After the receiver changes are checked, I recommend to call the remaining breakpoints 'unknown' or 'undocumented' unless a true multipath diagnostic is provided.
Response: Johansson et al., (2002) performed a number of solutions with two different GPS receivers connected to the same antenna and determined differences at the 1 mm level or less in coordinates (~0.05 kg/m^2 in IWV). Beside the references, in the revised paper we included a plot (Figure 3) to show the time series GPS-ERA-

Interim differences for three sites with the vertical lines indicating the date of the receiver changes. The figure demonstrates that the impact of receiver changes on the estimated IWV is not noticeable.

Regarding the choice of period for correction of the GPS interventions, the one with the smallest bias compared to ERA-Interim might not be the best choice since ERA-Interim itself may contain biases. Why didn't the authors use the more recent period following their previous work (Ning and Elgered, 2012)?
Response: In the previous work only the Onsala IWV data were corrected for an intervention. In that case it was clear from independent VLBI data that the later period had more accurate data. The reason for the intervention was to replace a conical radome which was known to produce an IWV bias. In addition to that the selection of the reference time period for the correction has insignificant impact on the trend estimated. In the revised paper, this is clearly addressed by the sentence" We found that trends are not affected by which reference period that is chosen (the maximum difference is smaller than 0.01 kg/m^2/decade) but the overall mean differences compared to other techniques will."

The mapping function test and second-order ionospheric corrections performed on only two sites are not significant and don't add anything to the study as the impact of these parameters is known from past studies to be small in the study area. If to be mentioned, they may simply be included in the discussion section along with the elevation weighting results (Fig. 9 unnecessary).
Response: In the revised paper the results from the test using the alternative mapping functions and for the test including second-order ionospheric corrections are removed.

Use statistical tests to assess the significance of trend estimates and differences.
The authors recommend to compare the trends computed from two different cutoff angle elevation solutions. What should be done when they yield significantly different values?
*Response: In the revised paper, we calculated the uncertainty, or standard error (SE), of the trend RMS difference in order to give the confidence level of the change in RMS differences. This has now been explicitly stated in the text of the revised paper by the sentence "the largest improvement seen in RMS differences is from 0.20 kg/(m^2*decade) to 0.14 kg/(m^2*decade) which is significant with an approximate 67 % confidence level."*

Data from ERA-Interim and radiosondes should be intercompared and checked for inhomogeneities as well. Why didn't the authors use a homogenized radiosonde dataset? (e.g. Dai et al., 2011)
Dai, A., J. Wang, P. W. Thorne, D. E. Parker, L. Haimberger, and X. L. Wang (2011), A new approach to homogenize daily radiosonde humidity data, J. Clim., 24, 965–991.
Response: Via personal communication with June Wang we have found out that the work was not continued and therefore we cannot have the homogenized radiosonde data covering the investigated time period.

Table 5: are the ZHD trends significant? They are not discussed in the text. Figure 4: is this figure really useful?

Response: In the revised paper, a few sentences are added to discuss ZHD trends: "The significance of the ZHD trends does not really matter since they are calculated accurately using measurements of the ground pressure. Therefore, they are not important when subtracted from the ZTD and hence have almost no impact on the accuracy of the ZWD and the IWV trends."

We agree that Figure 4 is not necessary, and it has been removed from the revised paper.

Figure 8, 9: unnecessary figures, but the statistics could be included in a Table.

Response: Figures 8 and 9 have been removed and the statistics are presented in Table 10.