

Editor comments on acp-2018-1170, revised manuscript (version 4), Interpreting the time variability of world-wide GPS and GOME/SCIAMACHY integrated water vapour retrievals, using reanalyses as auxiliary tools, by Roeland Van Malderen, Eric Pottiaux, Gintautas Stankunavicius, Steffen Beirle, Thomas Wagner, Hugues Brenot, and Carine Bruyninx

The word “tools” in the title of the manuscript is not very nice. I suggest replacing “...using reanalyses as auxiliary tools” by something like “by comparison with reanalyses”.

P4L20-24: This sentence can be removed as it makes reference to other data sets which do not fit the scope of the study.

P4L28-29: “the 5-minute resolution IGS repro 1 ZTD estimates are downsized to a 6-h time resolution” downsized is not a proper term and does not explain how you converted the 5-min data to 6-hourly data. Please explained if either you selected the 5-min values that fall exactly on the 00, 06, 12, or 18 UTC times, or the nearest in time within some window, or if you averaged several 5-min values over some interval centred on the UTC times. Note that using the 5-min values instead of time-averaged values results in noisier time series with more gaps. Though this may not be a major issue when monthly means are computed afterwards, it should be mentioned should this be the way you handled the time sampling.

P5L7-12: why do you use surface data from ERA-Interim? They need to be extrapolated from the model surface to the height of the GPS stations whereas pressure level data can be interpolated with higher accuracy to the height of the stations. Extrapolation of the model surface pressure to the altitudes of the GPS stations is known to introduce spurious seasonal signals and is not recommended (see my presentation at the GNSS4SWEC workshop in Reykjavik, 2016).

Bock O., A reference IWV dataset combining IGS repro1 and ERA-Interim reanalysis for the assessment of homogenization algorithms, 3rd COST ES1206 Workshop, 8-11.03.2016, Reykjavik, Iceland.

The motivation for including the NCEP/NCAR reanalysis 1 is not explained. This reanalysis has many known deficiencies in the representation of the water cycle and has been supplemented rapidly by NCEP/DOE reanalysis 2 and later by more modern global reanalyses (e.g. CFSR, MERRA, MERRA-2, JRA-55). Please justify the use of NCEP/NCAR reanalysis 1 or consider removing the results as they don't add much to the interpretation of time variability of the GPS IWV data.

P6L13-14: “The use of the NCEP/NCAR reanalysis is restricted to sensitivity analysis purposes because of its coarser spatial resolution.” This statement should be revised since you actually report results using this reanalysis, namely at the end of Section 3 and in Section 4. But again, given the resolution limitations and known issues with this reanalysis I think these results are not relevant.

Here are the full references to our past studies comparing evaluating the NCEP/NCAR and NCEP/DOE reanalyses mentioned by one of the reviewers:

Bock, O., M.-N. Bouin, A. Walpersdorf, J.P. Lafore, S. Janicot, F. Guichard, A. Agusti-Panareda (2007), Comparison of ground-based GPS precipitable water vapour to independent observations and Numerical Weather Prediction model reanalyses over Africa. Q. J. R. Meteorol. Soc., 133, 2011-2027, DOI: 10.1002/qj.185

Bock, O., and M. Nuret (2009) Verification of NWP model analyses and radiosonde humidity data with GPS precipitable water vapor estimates during AMMA. Weather Forecast., 24: 1085-1101 DOI:10.1175/2009WAF2222239.1

Section 3: Two of the referees questioned about the time sampling and missing data issues and I could not find clear answers to these questions. When you compare the monthly mean IWV data from the three datasets and report statistical results (bias, standard deviation of differences, etc.) it is important to know how you computed the monthly means and how you handled the missing data. For example, did you put a limit on the minimum number of 6-hourly values entering into the monthly means? Did you compute the biases as the mean of differences or difference of means? (i.e. are the time series time-matched beforehand or not?).

Why did you choose to illustrate only the correlations (Fig. 2) in section 3? Please justify.

P8L15: you report a R^2 value of 0.975 here but in the Appendix the value in the first row is 0.962. Shouldn't it be the same?

P8L17-22: "surface pressure has a larger impact than other variables". Based on the material in the Appendix, provide a quantitative estimate of uncertainty for the IWV data that are used in this study.

P8L20-21: ERAI is not a "local" data but represents averages over the size of grid cells ($0.75^\circ \times 0.75^\circ$)

Maybe a reference to Bock and Parracho, ACPD, 2019 (this Special Issue) can be useful here regarding the representativeness issues.

P8L26: "Looking at the biases" not easy since you don't show them. Please reformulated. Maybe provide results in a Table to support this discussion.

P9L3: "The standard deviations are smallest between GPS and ERA-interim": did you compare the standard deviations (variability)? Or do you (improperly) refer to the standard deviation of differences here?

P9L4: "Here, the impact between the different observations times (= satellite overpass times) at the sites for GOMESCIA compared to GPS and ERA-interim should be highlighted." Please explain the problem and evaluate the impact. A reference to Alraddawi et al., AMT, 2018 (this Special Issue) may be useful here.

P9L12: you should choose whether you think the results are "very similar" or "slightly worse". And I don't understand how the IWV comparison with NCEP can say something on the conversion using ERAI data? It may be that GPS IWV converted from ERAI data is more consistent with ERAI IWV data because the same reanalysis is used.

Section 4: it is not clear if you used monthly data or data with a higher time sampling.

P10L25: I think the NCEP results are not relevant because of the above-mentioned limitations. Moreover, since the results are only mentioned and not actually shown, I suggest to remove this sentence.

P17L27: I am not sure it is a good idea to use the NCEP results in the correlation analysis because of the above-mentioned limitations. Why don't you use the ERAI data instead?

Appendix: The presentation of Table A1 is not clear. I think it would be more clear to write out the compared data sources, e.g., for T_m and P_s in section [b] horizontally: T_m =ERA, P_s =ERA, and vertically: T_m =ERA, P_s =SYNOP; etc. Move the comparisons of T_m and P_s values from sections [b, c, d] to dedicated sections. For these comparisons the units are not mm. Section [a] is also a special case which should be specified as IWV=ERA; and IWV=GPS with T_m =ERA, P_s =ERA.

The text of the Appendix should be slightly revised. Strictly speaking, the sensitivity is quantified by the partial derivatives of IWW wrt T_m , T_s , and P_s , i.e. it is actually addressed only in the first paragraph (note that this paragraph should also be clarified and completed, especially quantify the impact quoted in the last sentence; maybe a reference to the PhD report of A. Parracho would be relevant here for further reading on this topic). The rest of the Appendix quantifies uncertainties (and not sensitivity) to the use of various datasets. This is useful but remains specific to the tested datasets. It might be interesting to add also the comparison between NCEP and ERA parameters (mentioned in the discussion but not reported in the Table).

Parracho, A. C., (2017) Study of trends and variability of atmospheric integrated water vapour with climate models and observations from global GNSS network, PhD report, Université Pierre et Marie Curie, Paris, France, <http://www.theses.fr/2017PA066524>