Review of manuscript amt-2016-264 “Determination of zenith hydrostatic delays and the development of new global long-term GNSS-derived precipitable water vapor” by Xiaoming Wang and co-authors.

General comments

The conversion of GNSS ZTD to PWV requires the use of surface pressure data to estimate the hydrostatic delay component. Errors in the surface pressure add uncertainty in the PWV results and may lead to erroneous conclusions on climate variations. This manuscript investigates the accuracy of surface pressure data from two global datasets based on the ERA-Interim reanalysis: the GPT2w, a coarse spatial/temporal resolution (5° mean horizontal, with annual and semi-annual cycles) version of the reanalysis commonly used by geodesists for GNSS data processing, and the legacy reanalysis at high spatial/temporal resolution (80-km horizontal, 60 levels, 6-hourly). This topic is very important to the GNSS/climate community and the proposed study is pertinent to the AMT journal aims and scope. However, the approach followed by the authors needs substantial improvement and the results need be analysed more thoroughly to be really useful to the scientific community. Given the importance of the topic and work already done by the authors, my suggestion is a major revision. I give below the main issues which should be solved and improvements that should be brought to the organisation of results before publication.

Major comments

1. On the quality of the reference pressure observations

The accuracy of the reanalysis data is evaluated with respect to surface pressure observations available from the IGS and distributed by SOPAC. Nothing is said about the accuracy of the IGS data in the manuscript. Have these data been quality controlled? How can their accuracy be established to be suitable for serving as a reference at the level of climate requirements? It has been shown in past studies (Wang et al., 2007; Heise et al., 2009) that the IGS meteorological data are generally of poor quality. I urge the authors either to use another, validated, surface pressure dataset (e.g. ISPD, see the work of Lagler et al., 2013), or to thoroughly screen the IGS data and select a subset of high quality stations. The fact that many biases detected in the GPT2w data appear also in the ERA-Interim data at much higher spatial resolution (Fig. 3a-c) suggest that these biases might in fact be in the IGS pressure data.

2. On the interpolation methods for ERA-Interim data

Two methods are introduced in section 3 for interpolating the ERA-Interim data from the model grid to the GNSS site. The first one is based on the nearest grid point and the second one is using the 4 surrounding grid points. The motivation for comparing these two methods should be better explained and their results should be discussed and interpreted in a more comprehensive way. The impact of representativeness errors should also be discussed when comparing model data and observations. However, in its present form, I suspect a major issue in the results due an inconsistency between the vertical interpolations used in both methods. Whereas the first method is based on the standard formula – eq (1) - assuming and constant lapse rate (linear temperature variation with height in the troposphere), the second one follows Schüler, 2001, and uses an empirical formula – eq (4) or (5) – which is inconsistent with eq (1) and poorly validated for usage at global scale. Moreover,

the weighted interpolation from 2 model levels – eq (3), (6) and (7) is a commonly used approach for horizontal interpolation but is not a priori valid for vertical interpolation because it would not conserve mass (vertical pressure variations should satisfy hydrostatic equilibrium). Tracking back the origin and validity of these equations in Schüler, 2001, their usage for climate purposes appears highly questionable. I urge the author either to demonstrate in an appendix the validity of these equations at global scale or bring the vertical interpolation in line with the first method.

3. Objectives of the work and interpretation of the results

Though it is a priori obvious that GPT2w will give worse results than Era-Interim due to the difference in spatial and temporal resolutions, quantifying the spatial distribution of errors and decomposing them into different time scales (mean, seasonal, diurnal) is useful in an assessment study. In this respect, the Introduction should better state the overall aim of this study and introduce the requirements in terms of accuracy on the studied data for climate applications. Once the target accuracy is specified it is easier to conclude on the observed results. The reference to the E-GVAP Product Reference Document given P8 should thus be provided in the Introduction. Note however, that the E-GVAP requirements may not be adequate for global climate as they are only expressed in a single value in kg m-2 unit. Therefore the requirements should be complemented with GCOS recommendations and expressed either in % or consider different values in different climate zones.

Tables presenting results in latitude bands and plots of results as a function of latitude might be useful to give a synthetic and more legible view than the hard to read plots (Fig. 3 and similar) and lengthy and repetitive descriptions in the text (similar results for pressure, ZHD, and PWV).

The spatial distribution and temporal variations of pressure/ZHD (Fig. 5-8) are well known climatic features (e.g. Trenberth, 1981; Dai and Wang, 1999). The text and comments should be revised accordingly.

Trenberth, K. E. (1981), Seasonal variations in global sea level pressure and the total mass of the atmosphere, J. Geophys. Res., 86(C6), 5238–5246, doi:10.1029/JC086iC06p05238.

Dai, A., & Wang, J. (1999). Diurnal and semidiurnal tides in global surface pressure fields. Journal of the atmospheric sciences, 56(22), 3874-3891.

The ZTD data introduced in section 4.4 are not used in fact because the error in PWV due to surface pressure does not depend on ZTD but only on ZHD and the conversion factor PI. So the ZTD could be completely avoided in this study unless the relative PWV errors are computed, in which case the results would depend on ZHD and ZTD (and no longer on PI). I suggest that the authors present also the relative PWV errors which might also highlight shortcomings in the Polar Regions.

The authors conclude that ERA-Interim pressure data can be used globally for climate studies while GPT2w may be suitable only in the tropics. These conclusions are simply based on the E-GVAP thresholds and the results obtained from the comparison of 6-hourly data. However, it is obvious that for climate applications, it might often be sufficient to consider monthly means. Hence the random errors would be reduced accordingly and a larger number of sites might be considered. This study should thus provide also results for monthly mean data.

At the end of section 4.4.1, it is written that ERA-Interim data yield RMS errors < 0.5mm at 75 or 78% of the sites. What happens at the remaining 22 or 25%? Should these stations be blacklisted?

The discussion and conclusion must also take into account the presence of systematic errors.

4. On the presentation of results

In section 4 of the manuscript, the results for surface pressure, ZHD, and PWV, are presented successively. In each case, the biases and RMS errors characterizing the surface pressure difference between the tested model and the reference observations are presented. As attested by eq (11) and (13), an error in surface pressure produces a proportional error in ZHD and PWV which can be quantified almost exactly by the rule of thumb: 2.3 mm/hPa and 1 kg m-2 / 6.5 mm, respectively. As a consequence, the spatial distributions of biases and RMS errors presented in Fig. 4 and 10 are quasi similar to those shown in Fig. 3 and don't add information. This is also the case for Fig. 5 -8 (pressure and ZHD). I suggest that the authors combine the results in one figure when possible and add data axis (or colorbars) with multiple scales for pressure, ZHD, and PWV. This would avoid unnecessary duplication of figures and leave room for additional information.

Minor comments

- The IS unit for pressure is hPa (not mbar)
- The preferred unit for PWV is kg m-2 as mm may be mixed up with the ZHD unit.
- It is written P3L87 that the ERA-Interim data are available on 60 model levels, but later the equations referring to computed quantities refer to pressure levels (section 3.2 and 4). Please clarify.
- Section 4.3: it is not said which of the two ERA-Interim datasets is used.
- Section 4.3: it is not said how the annual and semi-annual amplitudes are computed.
- The information provided in Fig. 2 might be simply added in the captions of Fig. 1.
- Fig. 3 and alike are too small to be useful. Consider using full page width.
- P9L245: add a reference to the E-GVAP Product Reference Document rather than citing the website (http://egvap.dmi.dk/)
- Reference to Yao et al., GJI, 2014, is not complete and might be replaced with a more recent paper by Yao et al., Science China, 2015.