Reviewer #1: https://doi.org/10.5194/acp-2021-797-RC1

This manuscript investigated the multi-temporal-scale variabilities and trends of IWV and assessed six commonly-used atmospheric reanalyses (CFSR, ERA5, ERA-Interim, JRA55, MERR2, and NCEP2) over Europe using IWV time series from 108 GPS stations for more than two decades. I have the following comments:

Reply: Thanks a lot for your comments.

Main comments:

1. The authors have taken into account of vertical IWV adjustment. However, the height system of GPS is different from that of the reanalyses. I'm not sure if the authors have considered the unification and differences of the different height systems?

Reply: You are right. Ellipsoidal height is usually used in GPS data processing, whereas geopotential height is used in atmospheric reanalyses. To calculate meteorological variables from reanalyses at the location of a GPS station, geopotential height of the station should be used. In this study, the height system conversion was carried out as follows:

We first converted the reference of the GPS station's altitude from ellipsoid to Mean Sea Level (MSL) by using geoid model Earth Gravitational Model 2008 (EGM2008; Pavlis et al., 2012):

$$H_{or} = H_{el} - N, \quad (1)$$

where the H_{or} and N are the orthometric and geoid heights in metre, respectively.

We then adjusted the altitude of the GPS station (with a latitude of φ) by considering gravity variations (Dirksen et al., 2014; Wang et al., 2016; World Meteorological Organization, 2018):

$$H_{gp} = \frac{\gamma_s(\varphi)}{9.80665} \cdot \frac{R(\varphi) \cdot H_{or}}{R(\varphi) + H_{or}}, \quad (2)$$

$$\gamma_s(\varphi) = 9.780325 \frac{1 + 1.93185 \times 10^{-3} \cdot \sin(\varphi)^2}{(1 - 6.69435 \times 10^{-3} \cdot \sin(\varphi)^2)^{0.5}}, \quad (3)$$

$$R(\varphi) = \frac{6.378137 \times 10^6}{1.006803 - 6.706 \times 10^{-3} \cdot \sin(\varphi)^2}. \quad (4)$$

Reference

Dirksen, R. J., Sommer, M., Immler, F. J., Hurst, D. F., Kivi, R., and Vömel, H.: Reference quality upper-air measurements: GRUAN data processing for the Vaisala RS92 radiosonde, Atmos. Meas. Tech., 7, 4463–4490, https://doi.org/10.5194/amt-7-4463-2014, 2014.

Pavlis, N. K., Holmes, S. A., Kenyon, S. C., and Factor, J. K.: The development and evaluation of the Earth Gravitational Model 2008 (EGM2008), Journal of Geophysical Research: Solid Earth, 117, https://doi.org/10.1029/2011JB008916, 2012.

Wang, X., Zhang, K., Wu, S., Fan, S., and Cheng, Y.: Water vapor-weighted mean temperature and its impact on the determination of precipitable water vapor and its linear trend, Journal of Geophysical Research: Atmospheres, 121, 833–852, https://doi.org/10.1002/2015JD024181, 2016.

World Meteorological Organization: Guide to Instruments and Methods of Observation, Measurement of Meteorological Variables. (WMO-No. 8), 1, 2018.

2. In the manuscript, the authors used the difference time series between ERA5 IWV and GPS IWV to visually detect the breaks in GPS IWV, so the potential significant differences may be eliminated since the homogenization, also this may be the reason why the ERA5 outperforms than other reanalyses. Are these breaks based on ERA5 IWV still significant, are there any other reanalyses used for the homogenization process?

Reply: We agree with you that using ERA5 alone for the homogenization of GPS IWV is unfair to the evaluation of the other five reanalyses. Therefore, in the revised manuscript, we used all the six atmospheric reanalyses for the detection. The procedure is as follows. For each GPS station, we first tested the changepoints by referring to its log file and identify them by using the detection tool developed by Wang (2008) based on the comparisons of the monthly mean IWV values of GPS and each reanalysis. A changepoint is accepted if it is reported by the tool in at least three GPS–reanalysis comparisons. Its amplitude is calculated as the average of those reported by the comparisons.

Regarding the undocumented changepoints, we determined them carefully. For each GPS station, we first identified the undocumented changepoints automatically reported by the detection tool based on the IWV monthly mean comparisons of GPS and each reanalysis. If similar changepoints within six months are reported by at least three GPS-reanalysis comparisons, they are considered as the changepoints from the GPS IWV series. Then, they are combined into one at the median month, and its amplitude is calculated as the average of those reported by the comparisons. By using all reanalyses in the changepoint detection tool, we hope to minimize the effect of these changepoints on the results here, although it cannot completely rule out that identical changepoints appear in different reanalyses by ingesting the same observational datasets through data assimilation. This limitation has been mentioned.

Reference

Wang, X. L.: Penalized Maximal F Test for Detecting Undocumented Mean Shift without Trend Change, Journal of Atmospheric and Oceanic Technology, 25, 368–384, https://doi.org/10.1175/2007JTECHA982.1, 2008.

3. The spatial resolution contributes to most of the representativeness differences, such as the ERAI provides the products with higher spatial resolution (i.e. 0.25°) than the product used in this paper (0.75°). The conclusion that ERA5 has the best performance on the representativeness differences is questionable. This needs more clarification or convincing statements.

Reply: It is possible to download ERAI data at 0.25° spatial resolution, though its native resolution is 0.75°. However, the 0.25° ERAI is obtained from a bilinear interpolation of its native spatial resolution of 0.75°. Therefore, using 0.25° instead of 0.75° for the ERAI data will not really bring any benefit. An explanation from ECMWF is quoted as follows in italic type:

https://confluence.ecmwf.int/display/CKB/Does+downloading+data+at+higher+resolution+improve+the+output

Does downloading data at higher resolution improve the output?

When you download CAMS data, C3S data and other data from ECMWF, you can obtain the output data on its archived grid or on a Cartesian lat/long grid at a custom resolution.

You can specify a higher output resolution than the archived resolution, but the resulting data will not contain any more information than the original, it has merely been interpolated[1] to a higher resolution. This makes the output look smoother, but does not increase the accuracy or the precision of the data. However, if you choose to interpolate to a coarser resolution than the archived resolution you should be aware that the data can be aliased, unless care was taken to avoid this.

*For **ERA-Interim** atmospheric data the point interval on the native Gaussian grid is about 0.75 degrees. You can specify a custom grid on the data server web interface, or using the ECMWF WebAPI or using the MARS client (if you have access to it). On the web interface the default grid for ERA-Interim is lat/long, with a default resolution of 0.75x0.75 degrees (about 80km), approximating the irregular grid spacing on the native Gaussian grid.*

*For **ERA5** HRES atmospheric data the point interval on the native Gaussian grid is about 0.28 degrees. You can download ERA5 data using Python and specify a custom grid and resolution in your script. You should set the horizontal resolution to slightly lower than 0.28 degrees (about 30km), for example to 0.25 degrees, approximating the irregular grid spacing on the native Gaussian grid.*

[1] When data is interpolated, all continuous fields (e.g. precipitation, temperature) are interpolated by bilinear interpolation, and discrete fields (e.g. vegetation, precipitation type, soil type) and Wave 2D spectra are interpolated by nearest-neighbour. For more information about our grids and interpolations see in this presentation https://confluence.ecmwf.int/download/attachments/55122669/intro-interpolation-2016.pdf?api=v2

4. Line 202: "The 3- and 6-hourly IWVs are linearly interpolated into 1-hourly time series." Have the authors assessed the accuracy of the interpolated IWV? For IWV which changes in a high frequency, linear interpolation seems to be not a good choice.

Reply: This is an interesting question. We evaluated four interpolation approaches provided by MATLAB (https://ww2.mathworks.cn/help/matlab/ref/interp1.html?lang=en), including "linear", "spline", "pchip", and "makima". We took the 1-hourly GPS IWV of the 108 stations in Europe as reference series. We then selected two subsets of 3- and 6-hourly GPS IWV series from the 1-hourly series. After that, we interpolated the 3- and 6-hourly GPS IWV series into 1-hourly series by using the four approaches. The average Root-Mean-Square (RMS) estimates of the IWV differences between the original 1-hourly GPS IWV and those interpolated from 3- and 6-hourly series are as follows:

Table. Average RMS of differences between the original 1-hourly GPS IWV (kg m^{-2}) and those interpolated from 3- and 6-hourly time series

	linear	spline	pchip	makima
3-hourly	0.32	0.27	0.29	0.28
6-hourly	0.72	0.70	0.70	0.70

The Table shows that "spline" has the lowest average RMS of IWV differences, and we therefore selected "spline" for the temporal interpolation of IWV instead of "linear".

5. Line 157: There seems to be a missing full stop between "reanalyses" and "Compared". Please check it.

Reply: Corrected.