Community #1: https://doi.org/10.5194/acp-2021-797-CC1

Dear Peng Yuan and co-authors,

Thank you for releasing this interesting study. I am happy to see that you used the representativeness statistic that we proposed in a previous publication and that you confirm and extend our results to other reanalyses.

Below I submit a few questions and comments about your manuscript. Thank you in advance for your answers.

Best regards,

Olivier BOCK

Reply: Thanks a lot for your comments.

1. Please comment on the choice and on the quality of the used GPS data set (NGL), as other data sets exist for Europe (e.g. the EPN repro2, Pacione et al., 2017).

Reply: Pacione et al. (2017) developed a GNSS tropospheric dataset of about 18 years (1996–2014) in the framework of EPN-Repro2. The authors also compared the results obtained with different GNSS data processing schemes. In this work, we selected the ZTD product developed by NGL as it has a much longer time series from 1994 Jan. to 2018 Dec. (25 years). As IWV is the meteorological parameter that we are interested in, we directly evaluated GPS IWV instead of ZTD. Our comparisons show that the IWV estimates from NGL's ZTD and various reanalyses are in good agreement, indicating that the ZTD product has good quality.

2. Please provide more details on the homogenization method and results (e.g. the number and magnitude of detected breaks) and comment on their uncertainty. Explain also how the offsets in the GPS series are corrected, knowing that the breaks are detected in the GPS – reanalysis series and not in the GPS series directly.

Regarding the homogenization method, I checked your earlier paper (Yuan et al., 2021), and was wondering why you used a manual segmentation method when many statistical methods exist, which have been assessed by Van Malderen et al., 2020. Can you comment on that choice?

I also understand that in your segmentation method, you select only breaks which are confirmed by known equipment changes from the IGS log files. As you may have experienced: i) not all breaks are easy to detect (the example illustrated in Yuan et al., 2021, is a very optimistic case); ii) the IGS metadata may be incomplete and iii) the reanalysis may also have breaks. These limitations should be acknowledged in the paper.

Moreover, regarding the first two points, I think the manual approach is very subjective and also probably too conservative. You mention in the former paper that you detected 21 breaks from 108 stations over 21 years, i.e. an average of 1 break per station every 108 years. This number is very small compared to other studies, e.g. Ning et al., 2016, and Nguyen et al., 2021, using statistical methods. Overall, Nguyen et al., 2021, detected 1 break per station every 5.8 years (after screening) considering all breaks, among which the validated cases represent 1 break per station every 16 years. Both studies also show some obvious examples of undocumented breaks (namely for HERS) and breaks attributed to the reanalysis. Regarding the last point, you write that no obvious breaks were found in the reanalysis. What are your criteria to detect breaks in the reanalysis?

Reply: Thank you very much for the comments. The publications you mentioned are very nice and we will cite them. In the revised manuscript, we used all the six atmospheric reanalyses for the detection. 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.

The performance of different statistical changepoint detection tools has been assessed by Van Malderen et al., (2020), based on synthetic time series with inserted breakpoints. The authors concluded that the combination of different statistical detection tools, together with the use of the available metadata information on GPS instrumental changes, would be the most valid approach of homogenizing GPS IWV time series. In this study, we did combine the use of a statistical detection tool with the use of metadata. In addition, there are disputes on whether the breaks due to extreme climate events or unknown reason should be identified as changepoints. In the revised manuscript, we accepted these changepoints, but with caution.

As for the reason why fewer changepoints were reported in Yuan et al. (2021) than your paper is because we found that Power-Low noise are more suitable than the commonly-used AR(1) for the daily GPS–ERA5 IWV time series over Europe. The assumption of AR(1) noise underestimated the uncertainty of the break amplitude, and thus identify more breaks. Moreover, in that paper, we excluded the first several years in the time series if there is or quality problem. For instance, the data at station HERS before 2001 September was excluded because its quality had been poor until the antenna was repaired at that time (IGSMAIL-3503). However, three changepoints were identified at station HERS during 1998−2021 in Nguyen et al. (2021). In addition, we did not find report of changepoints at station HERS in Ning et al. (2016).

Regarding potential changepoints in reanalyses, we will remind the readers as suggested. 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 will be mentioned. However, we do not want to homogenise the reanalyses in this work. This is because they represent the native quality of the reanalyses that we would like to assess.

Reference

Ning, T., Wickert, J., Deng, Z., Heise, S., Dick, G., Vey, S., and Schöne, T.: Homogenized Time Series of the Atmospheric Water Vapor Content Obtained from the GNSS Reprocessed Data, Journal of Climate, 29, 2443–2456, https://doi.org/10.1175/JCLI-D-15-0158.1, 2016.

Van Malderen, R., Pottiaux, E., Klos, A., Bock, O., Bogusz, J., Chimani, B., Elias, M., Gruszczynska, M., Guijarro, J., Kazancı, S. Z., and Ning, T.: Homogenizing Gps Integrated Water Vapour Time Series: Methodology and Benchmarking the Algorithms on Synthetic Datasets, 13, 2020.

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.

Yuan, P., Hunegnaw, A., Alshawaf, F., Awange, J., Klos, A., Teferle, F. N., and Kutterer, H.: Feasibility of ERA5 integrated water vapor trends for climate change analysis in continental Europe: An evaluation with GPS (1994–2019) by considering statistical significance, Remote Sensing of Environment, 260, 112416, https://doi.org/10.1016/j.rse.2021.112416, 2021.

3. The analysis of the diurnal cycle is interesting. However, to make a fair intercomparison, the reanalyses should be analysed at the smaller common resolution which is 6-hourly, and not interpolated to a higher resolution (1-hourly). For the two reanalyses which have higher resolution (ERA5 and MERRA-2), you may show both the native and under-sampled (6-hourly) results.

Reply: Thank you for your constructive suggestion. We modified the comparisons accordingly. We compared the six reanalyses to GPS IWV at the least common temporal resolution (6-hourly). For ERA5 and MERRA2, we also carried out the comparisons at their respective native temporal resolutions. In addition, we would like to evaluate the potential benefit of ERA5 from its improvement in temporal resolution. Therefore, we compared the 1-hourly ERA5 to the other 3-/6-hourly reanalyses in modelling 1-hourly diurnal cycle and intraday variations of IWV.

4. In section 3.2, you may mention that the moist bias of ERAI over Europe was also reported by Parracho et al. 2018.

Reply: Added as suggested.

5. Please explain how you compute the trends.

Reply: Added as suggested.

6. In Section 6, you may mention that the trend results are also in line with the findings of Parracho et al. 2018, and Nguyen et al., 2021.

Reply: Added as suggested.

7. What is MERRA2' in Figure 5?

Reply: It was a typo and was removed.

Nguyen KN, Quarello A, Bock O, Lebarbier E. Sensitivity of Change-Point Detection and Trend Estimates to GNSS IWV Time Series Properties. Atmosphere. 2021; 12(9):1102. https://doi.org/10.3390/atmos12091102

Pacione, R., Araszkiewicz, A., Brockmann, E., and Dousa, J.: EPN-Repro2: A reference GNSS tropospheric data set over Europe, Atmos. Meas. Tech., 10, 1689–1705, https://doi.org/10.5194/amt-10-1689-2017, 2017.