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

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).

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?

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.

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

5. Please explain how you compute the trends.

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.

7. What is MERRA2' in Figure 5?

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.