# 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. **Our replies are shown in blue. The related texts copied from the revised manuscript are shown in green.**

**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 instructive and we cited them. In the revised manuscript, we described the homogenisation approach briefly in Section 2.5 Homogenisation, and provided step-by-step details and examples at two stations (HERS and ERLA) in Appendix A: Homogenisation of GNSS IWV time series.

**In Section 2.5 Homogenisation**

The GPS IWV time series can be inhomogeneous due to changes in GPS data processing strategies and station-related changes like hardware changes or changes in the electromagnetic environment (Van Malderen et al., 2020; Nguyen et al., 2021 and references therein). We employed a homogenisation approach as described in Appendix A with step-by-step details and examples at two stations (HERS and ERLA).

Here, we only provide a brief introduction on the approach. We first avoided inhomogeneities due to changes in GPS data processing strategy by using the homogeneously reprocessed GPS ZTD product. We then homogenised the GPS IWV time series by using the RHtestsV4 software (Wang and Feng, 2013). This software is developed especially for the detection and adjustment of changepoints in climatic time series, and it has been used in the homogenisation of IWV time series in previous studies (Ning et al., 2016; Schröder et al., 2016; Van Malderen et al., 2020).

We took the IWV time series from all the six reanalyses as references for the GPS IWV homogenisation and used a strategy to avoid the impacts of possible changepoints in individual reanalyses. However, we did not homogenise the reanalyses IWV time series because they represent the native quality of the reanalyses that we would like to assess. In addition to matching the detected changepoints with inspecting GPS metadata information (GPS station log files and IGSMAIL), we also allowed for possible undocumented changepoints in the GPS IWV time series.

**In Appendix A: Homogenisation of GNSS IWV time series**

Long-term IWV time series often suffer from inhomogeneities due to changes in instrumentation, data processing methods, and local environmental conditions (Van Malderen et al., 2020; Nguyen et al., 2021 and references therein). These inhomogeneities can manifest themselves as changes in the mean of the time series ("biases") at specific epochs, i.e., breaks or changepoints. If such changepoints are not properly corrected, they can significantly modify the estimations of long-term linear trend and multi-temporal-scale variabilities (e.g., Ning et al., 2016; Van Malderen et al., 2020; Yuan et al., 2021). Therefore, the homogenisation of the IWV time series is essential for a sound understanding and proper interpretation of IWV variability under climate change.

In this work, we examined the homogeneity of the monthly GPS IWV time series by using the RHtestsV4 software (Wang and Feng, 2013). This software is developed especially for the detection and adjustment of changepoints in climatic time series, and it has been used in the homogenisation of IWV time series in previous studies (Ning et al., 2016; Schröder et al., 2016; Van Malderen et al., 2020). The software is based on a penalized maximal t test with the consideration of linear trend, annual cycle, and AR(1) noise in the time series (Wang et al., 2007; Wang 2008).

We took all the six reanalyses as references for the homogenisation of the GNSS IWV time series, meaning that we inspected the monthly IWV difference time series between the GNSS and each of the reanalysis IWV. It is noteworthy that the reanalyses may also contain changepoints (e.g., Ning et al., 2016; Schröder et al, 2016). However, we did not homogenise the reanalyses IWV time series because they represent the native quality of the reanalyses that we would like to assess in this work. Also, by taking all six reanalyses as references, we are confident to minimise the impact of inhomogeneities in either reanalysis on the homogenisation process of the GNSS IWV time series. Practically, we used the following strategy to avoid the impacts of changepoints in specific reanalyses on the homogenisation of the GPS IWV time series:

(1) We examined the GPS IWV and metadata (station log file and IGSMAIL) carefully. If there is an instrumentation change within the first (or the last) year, we removed the several months before (or after) the epoch of change. Moreover, we also inspected the station up-coordinate time series and excluded periods with quality problem, as it is well-known that they are strongly correlated with the GPS tropospheric delay estimates (Tregoning and Herring, 2006). An example is given in Fig. A1 and will be described later in this appendix.
(2) We used the *FindU.wRef* command of the RHtestsV4 software to identify all possible changepoints in each GPS-reanalysis IWV monthly mean difference time series, which can be significant at a confidence level of 99% no matter they are documented in metadata or not.
(3) If a changepoint is within three months before or after a documented change in instrumentation, we adjusted its epoch according to the metadata and set it as Type-0. The rest changepoints are set as Type-1.
(4) If identical Type-1 changepoints are reported within six months in at least four GPS-reanalysis IWV differences, but not supported by metadata, we recognised them as a single Type-1 changepoint at the median of the epochs.
(5) We estimated the amplitudes of the Type-0 and Type-1 changepoints and tested their amplitude significances at a confidence level of 99% with the *StepSize.wRef* command of the RHtestsV4 software.
(6) We calculated the amplitude of each changepoint as the average of all the significant amplitude estimates from the six GPS-reanalysis IWV differences.
(7) We removed a changepoint if its amplitude is only significant in less than four GPS-reanalysis comparisons or its amplitude is less than three times of its standard deviation. Then, we repeated the steps of (5) and (6) until all the rest changepoints are significant.

The changepoint identification is finished after one or two iterations for most stations. In the end, we identified 44 Type-0 and 9 Type-1 changepoints as listed in Table A1. The total number of 53 changepoints is consistent to a previous global GPS IWV homogenisation work carried out by Ning et al. (2016), which identified 45 changepoints in total at 101 stations.

As the changepoint detection was carried out on monthly level, the specific dates of the Type-0 changepoint are fixed as documented. However, for the Type-1 changepoints without supports from metadata, their time of occurrences are fixed to the 15th day of associated month. We adjusted the GPS IWV time series by adding the amplitude of each changepoint to the GPS IWV data points before its time of occurrence. Note that five out of the nine Type-1 changepoints are significant in all the six GPS-reanalyses comparisons. Although we tried to minimize the impacts of changepoints in specific reanalyses on the results here, we cannot completely rule out that identical changepoints appear in all the six reanalyses by ingesting the same observational datasets through data assimilation.

Figure A1 and A2 show the homogenisation results at two stations. Station HERS (Herstmonceux, UK, 0.34 °E, 50.87 °N) is characterised with abnormal variations in its up-coordinate time series before the changes of antenna and receiver on 1998-2-18 as shown in Fig. A1h, indicating low-quality observations related to the instrumentation. In addition, obvious abnormal variations can be seen in all the GPS-reanalysis comparisons before 2001 September. We checked IGSMAIL-3503 (https://lists.igs.org/pipermail/igsmail/2001/004876.html) which reported a repair of antenna at station HERS until 2001-9-3. Therefore, we excluded the GPS IWV data before the date of repair. Then, we used the RHtestsV4 software to identify the changepoints in the rest GPS IWV time series and found one on 2010-8-19, which is significant at a confidence level of 99%. It is a Type-0 changepoint due to an antenna and receiver changes as recorded in the station log file. After the homogenisation, the linear trend of the GPS IWV time series at station HERS has been reduced from 0.71 to 0.27 kg $m^{-2}$ $decade^{-1}$, which generally agrees better with the trend estimates from reanalyses, which are 0.15, 0.43, 0.33, 0.37, 0.10, and 0.59 kg $m^{-2}$ $decade^{-1}$ for CFSR, ERA5, ERAI, JRA55, MERRA2, and NCEP2, respectively.

Station ERLA (Erlangen, Germany, 11.01 °E, 49.59 °N) has a Type-1 changepoint in 2015 July for unknown reason, in addition to a Type-0 one due to antenna and radome changes on 2010-8-18 (Fig. A2). The date of the Type-1 changepoint was fixed to 2005-7-15. With the homogenisation, the linear trend of the GPS IWV time series at station ERLA has been increased from 0.14 to 0.40 kg $m^{-2}$ $decade^{-1}$, which is closer to the trend estimates from reanalyses, which are 0.43, 0.36, 0.41, 0.43, 0.52, and 0.68 kg $m^{-2}$ $decade^{-1}$ for CFSR, ERA5, ERAI, JRA55, MERRA2, and NCEP2, respectively. Moreover, we compared the GPS IWV trend to three nearby stations (KARL, KLOP, and WTZR) with values of 0.56, 0.32, and 0.66 kg $m^{-2}$ $decade^{-1}$ and their distances to ERLA of 198.6, 177.5, and 144.4 km, respectively. The results indicate an improved spatial consistency in the GPS IWV trends from the homogenised time series. Therefore, the homogenisation at station ERLA is considered to be reasonable.

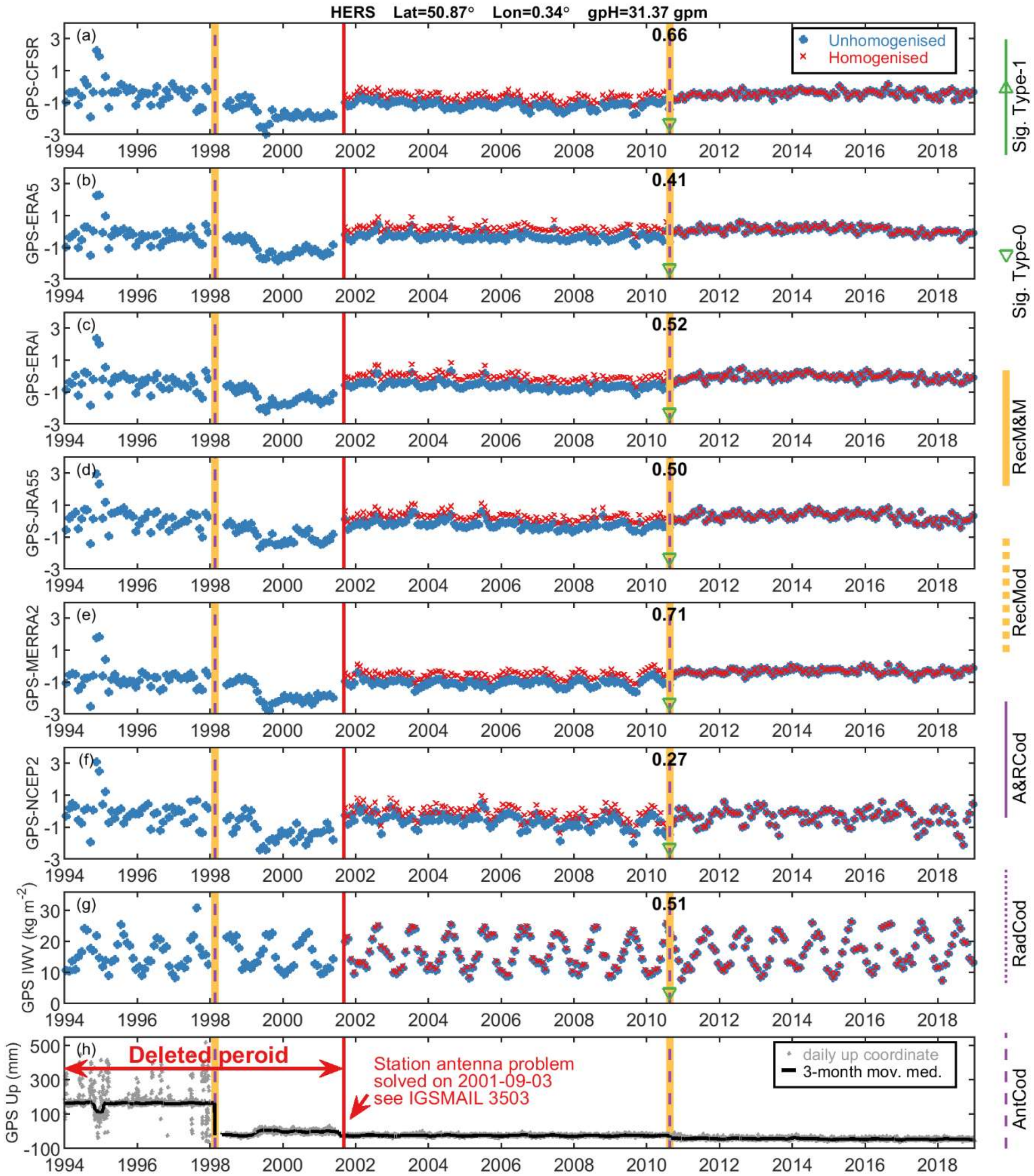


**Figure A1.** Monthly GPS-reanalysis IWV difference (a–f), monthly GPS IWV (g), and daily up-coordinate time series at station HERS. The GPS data before 2001-9-3 were deleted due to a problem in station antenna (see IGSMAIL 3503). The IWV time series before and after the homogenisation are labelled as blue dots and red crosses in (a)–(g). The types of instrumentation changes are listed in Table A2. The Type-0 changepoint on 2010-8-19 is significant at a confidence level of 99% (green downward-pointing triangle), and its value was calculated as the average of the values estimated in each GPS-reanalysis comparison as shown in (a–f).

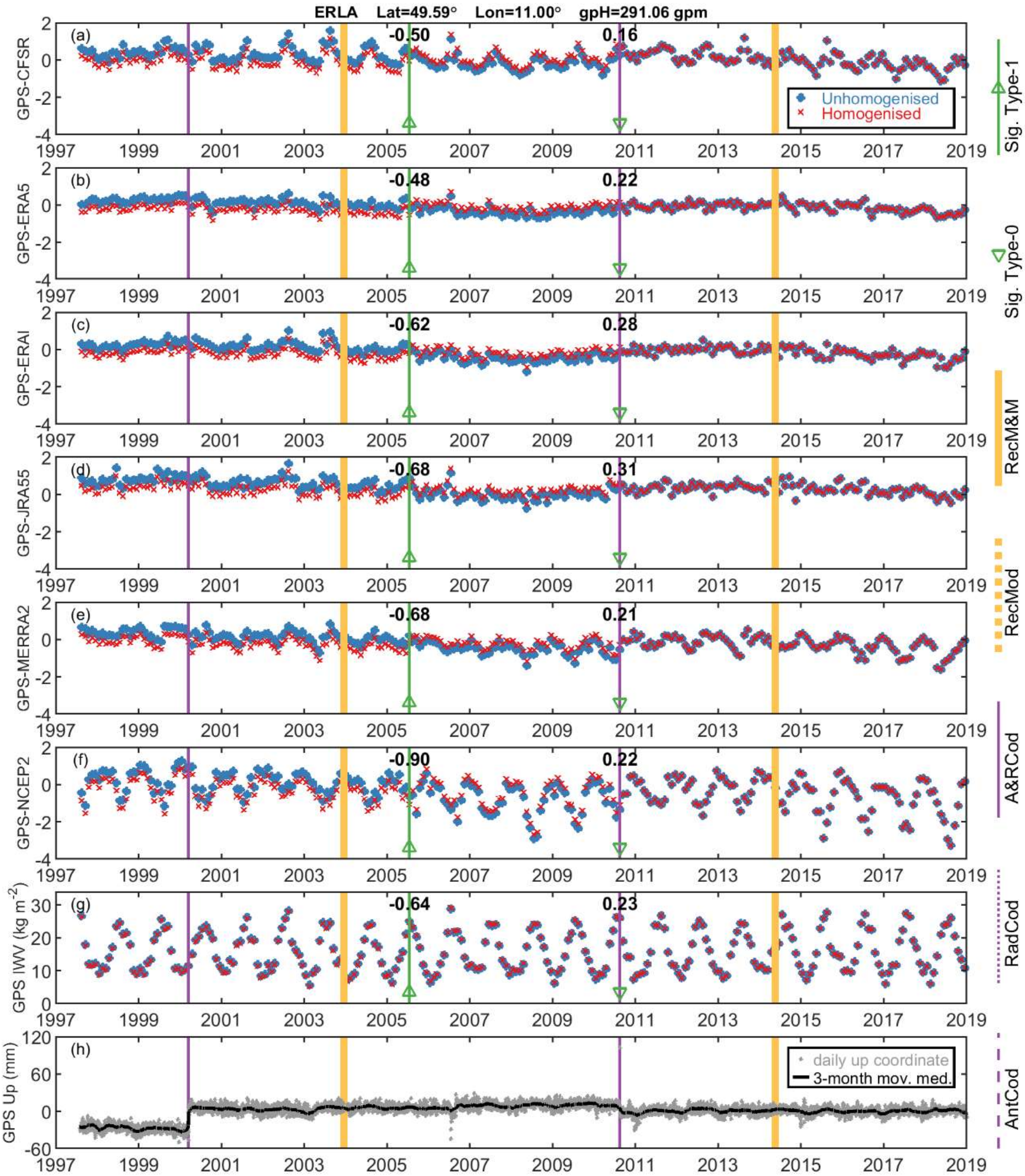


**Figure A2.** The same as Fig. A1 but for station ERLA. A Type-0 changepoint is significant on 2010-8-18 (green triangle) due to changes in antenna and radome at the station. A Type-1 changepoint in 2005 July (green vertical line and upward-pointing triangle) is significant but without support from metadata, and hence its date was fixed to 2005-7-15.

**Table A1.** The identified changepoints in GPS IWV time series. The Type-0 and Type-1 are changepoints with and without changes in instrumentation as documented in GPS station log files, respectively. The full names of the Type-0 changepoint events are shown in Table A2, and the Type-1 changepoints are labelled as Unknown. The G-C, G-E, G-I, G-J, G-M, and G-N indicate the GPS IWV changepoints in kg $m^{-2}$ estimated by compared to CFSR, ERA5, ERAI, JRA55, MERRA2, and NCEP2, respectively. The NaN indicates that the changepoint in the specific GPS-reanalysis comparison is insignificant at a confidence level of 99%. The mean and SD are the mean value and standard deviation of each changepoint.

| Station | Type | Date | Mean | SD | G-C | G-E | G-I | G-J | G-M | G-N | Events |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **APEL** | 1 | 2006-09-15 | -0.31 | 0.10 | -0.26 | -0.30 | -0.32 | -0.22 | -0.26 | -0.50 | Unknown |
| **APEL** | 0 | 2013-06-19 | 0.36 | 0.10 | 0.36 | 0.32 | 0.40 | 0.44 | 0.47 | 0.20 | RecM&M |
| **APEL** | 1 | 2016-04-15 | -0.33 | 0.04 | -0.37 | -0.29 | NaN | -0.32 | -0.33 | NaN | Unknown |
| **BELL** | 0 | 2012-01-26 | 1.22 | 0.30 | 0.67 | 1.43 | 1.30 | 1.34 | 1.47 | 1.11 | AntCod |
| **BELL** | 0 | 2014-09-22 | -0.47 | 0.13 | -0.36 | -0.49 | -0.49 | -0.32 | -0.46 | -0.69 | AntCod |
| **BELL** | 0 | 2016-05-12 | -1.61 | 0.05 | -1.56 | -1.57 | -1.59 | -1.71 | -1.63 | -1.63 | AntCod and RecM&M |
| **BRST** | 0 | 2011-10-26 | 0.36 | 0.08 | 0.45 | 0.28 | 0.39 | 0.41 | 0.37 | 0.25 | AntCod and RecM&M |
| **CREU** | 0 | 2016-05-17 | -1.03 | 0.17 | -0.81 | -1.11 | -1.08 | -1.04 | -0.87 | -1.28 | AntCod and RecM&M |
| **DELF** | 0 | 2000-07-23 | 0.23 | 0.05 | NaN | 0.26 | 0.27 | 0.20 | 0.18 | NaN | A&RCod and RecM&M |
| **DOUR** | 0 | 2015-03-02 | 0.37 | 0.03 | 0.40 | 0.35 | 0.35 | 0.34 | 0.39 | NaN | AntCod |
| **EIJS** | 0 | 2000-04-28 | 0.22 | 0.07 | 0.33 | 0.25 | 0.19 | 0.18 | 0.15 | NaN | A&RCod and RecM&M |
| **ERLA** | 1 | 2005-07-15 | -0.64 | 0.15 | -0.50 | -0.48 | -0.62 | -0.68 | -0.68 | -0.90 | Unknown |
| **ERLA** | 0 | 2010-08-18 | 0.23 | 0.05 | 0.16 | 0.22 | 0.28 | 0.31 | 0.21 | 0.22 | A&RCod |
| **EUSK** | 0 | 2001-05-09 | -0.74 | 0.13 | -0.56 | -0.74 | -0.76 | -0.73 | -0.71 | -0.97 | A&RCod |
| **GOPE** | 0 | 1999-11-04 | 0.27 | 0.03 | 0.25 | 0.24 | NaN | NaN | 0.31 | 0.28 | A&RCod and RecM&M |
| **GOPE** | 0 | 2000-07-24 | -0.96 | 0.07 | -1.06 | -0.86 | -0.97 | -1.00 | -0.98 | -0.92 | A&RCod and RecM&M |
| **GOPE** | 1 | 2001-09-15 | -0.51 | 0.10 | -0.52 | -0.57 | -0.34 | -0.51 | -0.65 | -0.49 | Unknown |
| **GOPE** | 0 | 2006-07-14 | -0.39 | 0.09 | -0.39 | -0.30 | -0.32 | -0.33 | -0.47 | -0.52 | A&RCod |
| **GOPE** | 0 | 2009-12-14 | 0.35 | 0.10 | 0.17 | 0.33 | 0.40 | 0.43 | 0.33 | 0.42 | A&RCod and RecM&M |
| **GOPE** | 1 | 2016-05-15 | -0.36 | 0.06 | NaN | -0.29 | -0.36 | -0.33 | -0.35 | -0.46 | Unknown |
| **HELG** | 0 | 2008-09-02 | 0.19 | 0.06 | 0.23 | 0.15 | 0.11 | 0.19 | 0.27 | 0.18 | A&RCod and RecM&M |
| **HELG** | 0 | 2014-09-09 | -0.14 | 0.04 | NaN | -0.19 | -0.10 | -0.11 | NaN | -0.17 | A&RCod and RecM&M |
| **HERS** | 0 | 2010-08-19 | 0.51 | 0.16 | 0.66 | 0.41 | 0.52 | 0.50 | 0.71 | 0.27 | AntCod and RecM&M |
| **HOBU** | 0 | 2007-02-28 | -0.31 | 0.08 | -0.23 | -0.24 | -0.33 | -0.40 | -0.24 | -0.40 | A&RCod and RecM&M |
| **HOBU** | 0 | 2010-11-22 | 0.40 | 0.08 | 0.34 | 0.28 | 0.44 | 0.50 | 0.40 | 0.45 | A&RCod |
| **HOBU** | 0 | 2015-05-27 | -0.22 | 0.06 | -0.27 | -0.19 | -0.20 | -0.13 | -0.21 | -0.30 | RecM&M |
| **HOFN** | 0 | 2001-09-21 | -0.95 | 0.09 | -1.11 | -0.91 | -0.92 | -0.85 | -0.98 | -0.91 | A&RCod |
| **KARL** | 0 | 2001-05-10 | -0.64 | 0.14 | -0.64 | -0.47 | -0.54 | -0.65 | -0.68 | -0.89 | A&RCod |
| **KLOP** | 0 | 2001-05-08 | -0.60 | 0.11 | -0.40 | -0.55 | -0.65 | -0.67 | -0.62 | -0.72 | A&RCod |
| **LAMP** | 1 | 2013-03-15 | 0.99 | 0.17 | 1.05 | 0.70 | NaN | 1.00 | 1.10 | 1.10 | Unknown |
| **LAMP** | 0 | 2014-04-11 | -1.75 | 0.18 | -1.53 | -1.60 | -1.94 | -1.82 | -1.66 | -1.95 | AntCod and RecM&M |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **LAMP** | 0 | 2017-09-26 | 0.52 | 0.13 | 0.48 | 0.36 | 0.69 | 0.47 | 0.59 | NaN | AntCod |
| **LEED** | 0 | 2008-11-11 | 0.38 | 0.11 | 0.24 | 0.38 | 0.39 | 0.46 | 0.53 | 0.29 | A&RCod |
| **LEIJ** | 0 | 2010-07-01 | 0.25 | 0.06 | NaN | 0.23 | 0.31 | 0.29 | NaN | 0.17 | A&RCod |
| **MAN2** | 0 | 2008-01-23 | 0.29 | 0.04 | NaN | 0.24 | 0.32 | 0.32 | 0.28 | NaN | AntCod and RecMod |
| **MODA** | 0 | 2007-07-12 | 1.54 | 0.08 | NaN | 1.45 | 1.64 | 1.60 | 1.56 | 1.47 | RecM&M |
| **MODA** | 0 | 2008-01-08 | -1.60 | 0.28 | -1.11 | -1.53 | -1.69 | -1.68 | -1.92 | -1.70 | RecM&M |
| **MOPI** | 1 | 2004-08-15 | -0.53 | 0.09 | -0.43 | -0.54 | -0.51 | -0.44 | -0.65 | -0.62 | Unknown |
| **OSNA** | 0 | 2004-04-22 | 0.38 | 0.07 | 0.47 | 0.41 | 0.32 | 0.31 | 0.37 | NaN | AntCod |
| **OSNA** | 0 | 2007-04-23 | -0.61 | 0.06 | -0.60 | -0.57 | -0.62 | -0.59 | -0.57 | -0.73 | A&RCod and RecM&M |
| **OSNA** | 0 | 2011-04-05 | 0.23 | 0.06 | 0.18 | 0.14 | 0.28 | 0.29 | 0.24 | 0.26 | A&RCod |
| **OSNA** | 0 | 2015-06-11 | -0.18 | 0.06 | -0.17 | -0.20 | -0.12 | -0.16 | -0.18 | -0.28 | RecM&M |
| **PENC** | 0 | 2003-05-22 | 0.55 | 0.10 | 0.54 | 0.57 | 0.56 | 0.59 | 0.37 | 0.68 | AntCod |
| **PENC** | 0 | 2007-06-26 | -0.32 | 0.06 | -0.29 | -0.30 | -0.29 | -0.29 | -0.34 | -0.43 | RecM&M |
| **PTBB** | 1 | 2014-06-15 | -0.50 | 0.11 | -0.57 | -0.49 | -0.39 | -0.36 | -0.53 | -0.65 | Unknown |
| **REYK** | 1 | 2003-03-15 | 0.31 | 0.03 | NaN | 0.35 | 0.34 | 0.28 | 0.29 | NaN | Unknown |
| **REYK** | 0 | 2008-03-13 | -0.30 | 0.08 | -0.19 | -0.28 | -0.36 | -0.31 | -0.25 | -0.42 | A&RCod |
| **REYK** | 0 | 2013-05-02 | 0.27 | 0.09 | 0.24 | 0.13 | 0.34 | 0.33 | 0.28 | NaN | A&RCod and RecM&M |
| **SULD** | 0 | 2005-06-14 | -0.53 | 0.11 | -0.54 | -0.52 | -0.56 | -0.55 | -0.32 | -0.67 | AntCod |
| **TERS** | 0 | 2008-09-16 | -0.20 | 0.06 | -0.29 | -0.18 | -0.20 | -0.14 | -0.13 | -0.26 | RecM&M |
| **TERS** | 0 | 2013-08-29 | 0.35 | 0.05 | 0.30 | 0.37 | 0.39 | 0.36 | 0.41 | 0.28 | RecM&M |
| **TRDS** | 0 | 2007-05-07 | 0.25 | 0.06 | 0.27 | 0.20 | 0.27 | 0.18 | 0.32 | NaN | A&RCod |
| **WSRA** | 0 | 2000-01-06 | 0.14 | 0.02 | NaN | 0.15 | 0.16 | 0.11 | 0.13 | NaN | RecM&M |

**Table A2.** Types of changes in GPS instrumentation.

| | Abbreviation | Type of change |
|---|---|---|
| 1 | AntCod | Antenna Code |
| 2 | RadCod | Radome Code |
| 3 | A&RCod | Antenna and Radome Code |
| 4 | RecMod | Receiver Model |
| 5 | RecM&M | Receiver Make and Model |

**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 first evaluated all the reanalyses with respect to GPS at 1-hour temporal resolution. For the reanalyses with coarser resolutions (3-hour and 6-hour), we interpolated their time series to 1-hour by using cubic spline, which was found to be slightly superior to linear interpolation (see our reply to *Reviewer#1 Comment #4*).

For a fairer intercomparison, we also evaluated ERA5 and the other reanalyses at their respective native resolutions. Accordingly, we extracted the ERA5 IWV every 3 and 6 hours. Statistics of the evaluation results are listed in Table 2.

**Table 2.** Statistics of the consistencies in diurnal IWV anomalies from reanalyses compared to GPS.

| | Temporal resolution | CFSR | ERA5 | ERAI | JRA55 | MERRA2 | NCEP2 |
|---|---|---|---|---|---|---|---|
| $\gamma$ | 1 h | 0.98±0.04 | 0.98±0.03 | 0.91±0.04 | 0.86±0.04 | 1.02±0.04 | 0.75±0.05 |
| | 3 h | – | 0.97±0.03 | – | – | 1.02±0.04 | – |
| | 6 h | 0.99±0.03 | 0.96±0.03 | 0.92±0.04 | 0.87±0.04 | – | 0.76±0.05 |
| $r$ | 1 h | 0.85±0.05 | 0.89±0.05 | 0.85±0.06 | 0.83±0.06 | 0.86±0.06 | 0.69±0.08 |
| | 3 h | – | 0.89±0.05 | – | – | 0.86±0.06 | – |
| | 6 h | 0.87±0.05 | 0.90±0.06 | 0.86±0.07 | 0.84±0.07 | – | 0.69±0.08 |
| KGE | 1 h | 0.85±0.05 | 0.88±0.06 | 0.82±0.07 | 0.78±0.07 | 0.85±0.06 | 0.60±0.09 |
| | 3 h | – | 0.89±0.06 | – | – | 0.86±0.06 | – |
| | 6 h | 0.86±0.06 | 0.89±0.06 | 0.83±0.07 | 0.79±0.07 | – | 0.60±0.08 |
| $RMS_\Delta$ (kg $m^{-2}$) | 1 h | 1.14±0.19 | 0.97±0.21 | 1.14±0.22 | 1.20±0.21 | 1.14±0.22 | 1.57±0.24 |
| | 3 h | – | 0.97±0.21 | – | – | 1.15±0.22 | – |
| | 6 h | 1.09±0.21 | 0.94±0.21 | 1.09±0.24 | 1.16±0.22 | – | 1.59±0.24 |

**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:** We redrafted Section 6 Assessments of linear trends and provided more details on the calculation of IWV trends.

**In Section 6 Assessments of linear trends**

We carried out a homogenisation of the GPS IWV time series by using the RHtestsV4 software (Wang and Feng, 2013) before the analysis. The software is dedicated to the homogenisation of climatic time series. In addition, we adopted a homogenisation strategy which allows for changepoints with and without support from metadata. We took all the six reanalyses as references and attempted to avoid the impacts of possible changepoints in specific reanalyses. However, we did not homogenise the reanalyses IWV time series because they represent the native quality of the reanalyses that we would like to assess. The homogenisation approach is detailed in Appendix A.

We then estimated the linear IWV trends from all the six reanalyses and the homogenised GPS IWV time series after the removal of annual cycle. In order to obtain realistic uncertainties of the trend estimates, we analysed the time series by using the Hector software version 1.7.2 (Bos et al.,2019). We tested four commonly used noise models, namely White Noise (WN), first-order AutoRegressive AR(1), AutoRegressive Moving Average ARMA(1,1), and Power-Law noise (PL). We then selected the optimal model of each time series by using Bayesian information criterion (BIC; Schwarz, 1978). Readers are referred to Yuan et al. (2021) for more details. The IWV trend estimates, associated uncertainties, and specific optimal noise models are listed in Table S3.

**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.

**In Section 6 Assessments of linear trends**

the geographical patterns of the IWV trends are consistent to previous studies (Parracho et al., 2018; Nguyen et al., 2021; Yuan et al., 2021).

**7.** What is MERRA2' in Figure 5?
**Reply:** It was a typo and was removed.

## 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. **Our replies are shown in blue. The related texts copied from the revised manuscript are shown in green.**

### 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, the geopotential height of the station should be used. We added the following texts.

**In section 2.3 IWV retrievals**

It is noteworthy that geopotential height system is employed in the reanalyses. Accordingly, the geopotential heights ($H_{gp}$) of the GPS stations rather than their ellipsoidal ($H_{el}$) or orthometric heights ($H_{or}$) are used in the above calculations. The conversion of height systems is carried out as follows (Dirksen et al., 2014; Wang et al., 2016; World Meteorological Organization, 2018):

$$H_{or} = H_{el} - N,$$

$$H_{gp} = \frac{\gamma_s(\varphi_s)}{9.80665} \cdot \frac{R(\varphi_s) \cdot H_{or}}{R(\varphi_s) + H_{or}}, \quad (5)$$

$$\gamma_s(\varphi_s) = 9.780325 \frac{1 + 1.93185 \times 10^{-3} \cdot \sin^2(\varphi_s)}{(1 - 6.69435 \times 10^{-3} \cdot \sin^2(\varphi_s))^{0.5}}, \quad (6)$$

$$R(\varphi_s) = \frac{6.378137 \times 10^6}{1.006803 - 6.706 \times 10^{-3} \cdot \sin(\varphi_s)^2}, \quad (7)$$

where the $N$ is the geoid heights in meter from the Earth Gravitational Model 2008 (EGM2008; Pavlis et al., 2012).

**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), J. Geophys. Res.: 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, J. Geophys. Res.: Atmos., 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. Moreover, we described the homogenisation approach briefly in Section 2.5 Homogenisation and provided step-by-step details and examples at two stations (HERS and ERLA) in Appendix A: Homogenisation of GNSS IWV time series. You can also find the copied texts in our *reply to Community #1 Comment #2*.

**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:** Thank you for the constructive suggestion. 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".

We therefore added the following words to Section 4.3 Diurnal anomalies:

For the reanalyses with coarser resolutions (3-hour and 6-hour), we interpolated their time series to 1-hour by using cubic spline, which is slightly superior to linear interpolation.

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

**Reply:** Corrected.

# Reviewer #2: https://doi.org/10.5194/acp-2021-797-RC2

## General Comments

The work presented in the manuscript gives an overall summary of applications of ground-based GPS observations in Europe of estimated time series of integrated water vapour (IWV), which to my knowledge is unique. It is broad in the sense that it deals with temporal scales from sub-daily to decades, while many previously published results often focus on one particular "signal", e.g. diurnal, annual, trends. As far as I can tell there are no new results in the manuscript, i.e. results that are different from what is already published. Three times it is stated that the results are "in line" with previously published results (lines 222, 323, and 399). Of course, it is also an important part of research to verify earlier findings, but if possible, I would appreciate if there was more emphasis on noted differences compared to earlier results. I am afraid I cannot help with the details. It is an impressive reference list and for me it is impossible to get a reasonably complete overall knowledge during the time allowed for the review.

**Reply:** Thank you very much for your affirmation. **Our replies are shown in blue. The related texts copied from the revised manuscript are shown in green.**

Ground-based GPS is a unique technique to evaluate the quality of IWV from atmospheric reanalyses. The evaluation can provide information on how to improve their performances in retrieving IWV. However, most previous studies in Europe only evaluated the IWV from ERA-Interim produced by ECMWF, which has been superseded by ERA5 since 2019 August. The time lengths of those studies are also relatively short (<20 years).

To our knowledge, this is **the first study** which used 25 years of 1-hourly GPS IWV in Europe to evaluate the performances of the newly released ERA5 in modelling multiple temporal scale variations of IWV from intraday to decades. In addition to the ERA5 and ERA-Interim produced by ECMWF, this study also evaluated the IWV from four commonly used products developed by USA and Japan, which have rarely been evaluated in Europe. An advantage of this comprehensive evaluation is that it is capable to avoid impacts due to differences in reference GPS IWV data and evaluation methods, so that the evaluations on the performances of the various reanalyses are more comparable.

There is an interesting finding which **has never been reported in previous studies** as far as we know. That is, the mismatches in the diurnal cycle of ERA5 IWV from 09 to 10 UTC and from 21 to 22 UTC. The artificial shifts are most likely due to the edge effect in each ERA5 assimilation cycle. We carried out an elaborate analysis on the spatiotemporal characterisations of the artificial shifts in ERA5 IWV diurnal cycles. Results show that they are -0.08 and 0.19 kg $m^{-2}$ at the two epochs and cannot be ignored. We added a specific section for this finding, see Section 4.2 Mismatches in diurnal ERA5 IWV cycle.

We also modified Section 4.1 Diurnal GPS IWV cycle as suggested by your comment #4. We modelled the diurnal IWV cycle by using diurnal and semidiurnal harmonics. Moreover, we analysed the characteristics of the diurnal IWV cycles by considering each station's altitude, and distance-to-sea, and climate zones. We also discussed different mechanisms, such as solar heating, land-sea breeze, and orographic circulation. This comprehensive analysis **is rare in Europe** as far as we know.

We compared the intraday IWV variations quantified as diurnal anomalies and confirmed that ERA5 is superior to the other reanalyse at all the 1-, 3-, and 6-hour temporal resolutions. We believe the results are very new and interesting to the community. This is because one of the most important advantages of ERA5 is its much higher temporal resolution compared to the other products (1-hourly versus. 3-, 6-hourly), but its possible improvement **has not been evaluated in Europe**.

Europe is known as the continent with the most significant warming speed. Evaluations of the long-term IWV trends from the atmospheric reanalyses with the 25 years of GPS IWV are also conducive to a better understanding of climate change. In the revised manuscript, we modified the homogenisation of GPS IWV time series and carried out fairer intercomparisons and evaluations of the long-term IWV trends from six reanalyses in Europe. We concluded that ERA5 is also superior to the others in modelling the linear IWV trends. However, due to significant discrepancies with respect to GPS, CFSR and NCEP2 are not recommend for the analysis of IWV trends over Southern Europe and the whole Europe, respectively. To our knowledge, these intercomparisons and evaluations are **reported for the first time**.

## Specific comments

**1.** L108: I do not understand the meaning of "integration rate of 95 %"? Can you explain what is being integrated?

**Reply:** We defined the integration rate of the daily IWV series at a GPS station as follows:

$$rate = \frac{N}{MJD_{last} - MJD_{first} + 1} \times 100\% \quad (1)$$

where $N$ is the number of daily IWV estimates of the GPS station. $MJD_{first}$ and $MJD_{last}$ are the Modified Julian Dates of the first and last daily IWV estimates, respectively. We added the following sentence to Section 2.1 GPS data:

The integration rate is the ratio between the number of available daily IWV data points and the theoretical number of all possible observations in the time range.

**2.** L112: You report that the observations were weighted based on the elevation angle. Is it not important how the weighting was done (a weighting function including sine and cosine terms)?

**Reply:** We added the following sentence to Section 2.1 GPS data:

The observations were weighted based on elevation ($e$) dependent function of $\sin e$. The cutoff elevation angle was set to 7°.

**3.** L192: It is mentioned that homogenisation was done as described by Yuan et al. (2021). I think such a process is critical and it deserves some more detail in your paper instead of having to go through the reference. For example, do you allow breaks to be inserted in the GPS IWV time series at a specific time epoch even if there has been no change noted in the log file for the hardware or the environment at the site?

**Reply:** We modified the homogenisation approach in the revised manuscript by using a statistical changepoint detection tool, the RHtestsV4 software (Wang and Feng, 2013). The software is based on a penalized maximal *t* test with the consideration of linear trend, annual cycle, and AR(1) noise in the time series (Wang et al., 2007; Wang 2008).

In addition to the documented changepoints in metadata (GPS station log files and IGSMAIL), we also allowed for undocumented changepoints in the GPS IWV time series without support from the metadata. We described the homogenisation approach briefly in Section 2.5 Homogenisation and provided step-by-step details and examples at two stations (HERS and ERLA) in Appendix A: Homogenisation of GNSS IWV time series. You can also find the copied texts in our *reply to Community #1 Comment #2*.

**Reference**


Wang, X. L. and Y. Feng.: RHtestsV4 User Manual. Climate Research Division, Atmospheric Science and Technology Directorate, Science and Technology Branch, Environment Canada, 2013.

Wang, X. L., Wen, Q. H., and Wu, Y.: Penalized Maximal t Test for Detecting Undocumented Mean Change in Climate Data Series, J. Climatol. Appl. Meteorol., 46, 916–931, https://doi.org/10.1175/JAM2504.1, 2007.

Wang, X. L.: Accounting for Autocorrelation in Detecting Mean Shifts in Climate Data Series Using the Penalized Maximal t or F Test, J. Appl. Meteor. Climatol., 47, 2423–2444, https://doi.org/10.1175/2008JAMC1741.1, 2008.


**4.** L262: My interpretation is that you determine the amplitudes of the diurnal signal as the peak-to-peak value regardless of when the peaks occur. This makes me wonder if the results will be different if instead the phase and amplitud of the sine wave with a 24 h period is estimated, e.g, through the method of least squares. (In some studies also a semidiurnal term, a period of 12 h, is estimated.) It will be of interest if you comment on this, at least for a couple of sites in different climate zones?

**Reply:** We carried out the analysis in a way as you suggested and redrafted "Section 4 Assessments of diurnal variations". We modelled the diurnal IWV cycle by using diurnal and semidiurnal harmonics.

In Section 4.1 Diurnal GPS IWV cycle, we analysed the characteristics of the diurnal IWV cycles by considering each station's altitude, and distance-to-sea (SeaDist), and climate zones.

In Section 4.2 Mismatch in diurnal ERA5 IWV cycle, we found and evaluated the mismatches in the diurnal cycle of ERA5 IWV from 09 to 10 UTC and from 21 to 22 UTC. There is an important finding which **has never been reported in previous studies** as far as we know.

### 4.1 Diurnal GPS IWV cycle

Starting with the all-time averaged amplitudes of the diurnal GPS IWV harmonic shown in Fig. 6e, two remarkable values can be first noted at stations NICO (Nicosia, Cyprus, 33.14 °N, 33.40 °E, 161.9 m) and ZECK (Zelenchuksky, Russia, 43.79 °N, 41.57 °E, 1143.4m) with values of 1.2 and 1.0 kg $m^{-2}$, respectively. Moreover, the diurnal harmonics at the Mediterranean Coast are generally stronger than the other regions in Europe (0.5−0.8 versus 0−0.5 kg $m^{-2}$). Obvious seasonal differences can also be seen in their diurnal harmonics, with significantly larger amplitudes in summer (June, July, and August; JJA) than the other seasons due to the stronger solar heating effect with minimal cloud coverage in Mediterranean summer (Enriquez-Alonso et al., 2016). However, the semidiurnal harmonics are much weaker, and their seasonal variations are less significant (Fig. 6f−j). The all-time averaged semidiurnal amplitudes are lower than 0.22 kg $m^{-2}$, except the two stations NICO and ZECK with values of 0.4 and 0.3 kg $m^{-2}$, respectively. The ratios between the all-time averaged semidiurnal and diurnal amplitudes are lower than 30% at 88% of the stations (95/108). As for the phases, the diurnal and semidiurnal terms are generally consistent over seasons (Fig. k−t), and their all-time averaged peaks are within 15−21 LST and 01−06 LST at 90% of the stations (97/108), respectively.

In order to compare the characteristics of diurnal IWV amplitudes at different stations, we calculated each station's relative amplitude as the ratio between their respective (semi-) diurnal amplitudes and mean IWV as displayed in Fig. 7a and 7d. We classified the GPS stations into three types according to their geographical characteristics (Fig. 7a) and analysed their relationships with each station's altitude and distance-to-sea (SeaDist). Firstly, we divided the stations with a limit of 20 km on their SeaDist. We further separated the stations located at Mediterranean Coast (MedCoast; SeaDist<20 km, 32 °N <Lat<46 °N, 5 °W <Lon<45 °E) from the other coastal (OtherCoast) stations, because their characteristics are quite different as can be seen from Fig. 7a. Consequently, the 108 stations are classified into 62 Inland stations, 12 MedCoast stations, and 34 OtherCoast stations.

As can be seen from Fig. 7a−c, all the OtherCoast stations are lower than 300 m and their relative diurnal IWV amplitudes are the weakest, with a range from 0.3% to 1.9% and a median of 1.1%. Within the altitude limit of 300 m, the Inland stations are characterised with moderately larger diurnal amplitudes (1.5%−2.5%). The results indicate the effect of land-sea breeze circulation on mitigating the intensity of diurnal IWV cycle at the Atlantic coasts of Europe with respect to the Inland Europe. However, the land-sea breeze effect can be less significant for the MedCoast stations, because their diurnal amplitudes are significantly larger (1.1%−4.2%) than the OtherCoast stations. The stronger diurnal IWV cycles at the MedCoast stations can be explained by the stronger solar heating effect at the Mediterranean Coast than the other European coasts, especially during summer daytime under stable and clear sky weather condition. In addition, it can be seen from Fig. 7b and 7e that the relative diurnal and semidiurnal IWV amplitudes are well correlated with altitudes, with correlation coeffieicents of 0.66 and 0.67, respectively. This relationship indicates the effect of orographic circulation, which can enhance the diurnal range of temperature at higher altitudes (Diedrich et al., 2016).

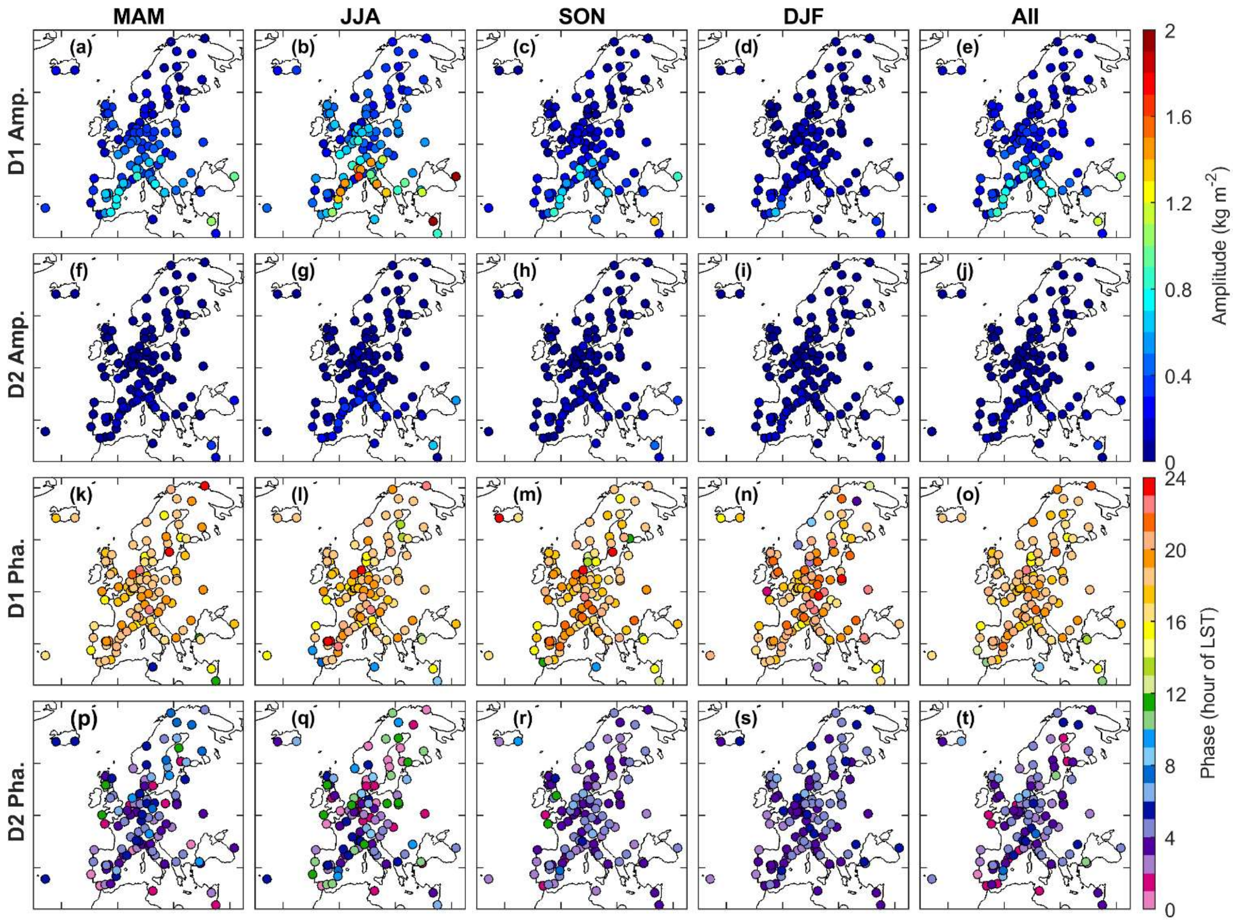


**Figure 6.** Plots of the amplitudes (a and f) and phases (k and p) for the first (D1) and second (D2) harmonics of diurnal GPS IWV cycle averaged in MAM (spring) at each station, respectively. The other subplots are for JJA (summer), SON (autumn), DJF (winter), and annual, from left to right, respectively. The phases are in Local Solar Time (LST) at the peak of associated harmonics.

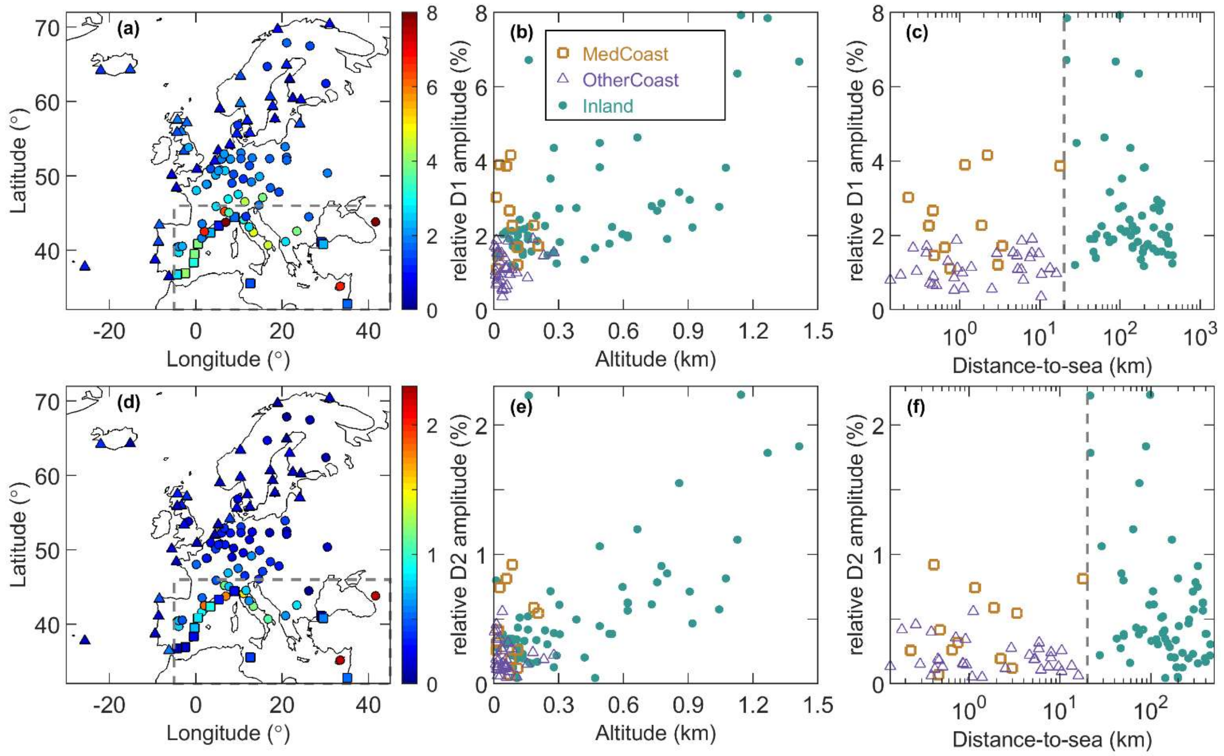


**Figure 7.** (a) relative amplitudes of the first harmonic ($D_1$) of diurnal GPS IWV cycle. (b) and (c) are the variations of the relative $D_1$ amplitudes with respect to station altitude and distance-to-sea, respectively. (d−f) are the same as (a−c) but for the second harmonic ($D_2$). The stations are classified into three types, namely Inland, MedCoast and OtherCoast. The type of Inland includes 62 stations with their distance to sea (SeaDist) no shorter than 20 km. The type of MedCoast contains 12 stations located at the coastal region of Mediterranean (SeaDist<20 km, 32 °N <Lat<46 °N, 5 °W<Lon<45 °E). The rest 34 stations are classified as OtherCoast.

In addition, we selected six stations with various altitudes, SeaDist, and climates to illustrate the diversity of diurnal IWV cycles in Europe as shown in Fig. 8. The climate zones of the GPS stations are classified according to Köppen Climate Classification (Beck et al., 2018) and the properties of the stations are listed in Table S2.

Station NICO is located on Cyprus Island with hot semi-arid climate (BSh). Despite only 21.5 km far away from coastline, NICO has the largest diurnal IWV amplitude with a value of 1.2 kg m$^{-2}$, equivalent to a relative amplitude of 6.7%. This is mainly due to the large diurnal temperature range in Cyprus, especially in summer (Price et al., 1999).

Both the MedCoast station VALE (Valencia, Spain, 39.48 °N, 0.34 °W, 27.0 m) and the OtherCoast station NEWL (Newlyn, UK, 50.10 °N, 5.54 °W, 11.0 m) are very close to coastline, with SeaDist values of 1.2 and 0.5 km, respectively. However, their diurnal amplitudes are quite different (absolutely 0.8 versus 0.1 kg m$^{-2}$, relatively 3.9% versus 0.7%). As explained earlier for the difference between the two station types, their weather conditions are different. VALE is located at Mediterranean Coast with cold semi-arid (BSk) climate. Its strong diurnal cycle, especially in summer with an amplitude of 1.4 kg m$^{-2}$ (4.6%), is attributed to intense solar heating under minimal cloud coverage weather condition. In contrast, NEWL is located at the coast of the English Channel with temperate oceanic (Cfb) climate, and its smaller diurnal amplitude can be due to the weaker solar heating effect under the unstable and cloudy weather condition, in addition to the land-sea breeze effect on mitigating diurnal temperature range.

Station ZECK is in the Greater Caucasus with humid continental (Dfb) climate. Its diurnal amplitude is much stronger than PENC (Penc, Hungary, 47.79 °N, 19.28 °E, 248.3 m) with the same climate (1.0 kg m$^{-2}$ and 7.9% versus 0.2 kg m$^{-2}$ and 1.5%). As ZECK is much higher than PENC, their difference in amplitude is consistent with the pattern for most Inland stations, which can be explained by the effect of orographic circulation (Diedrich et al., 2016). In addition, KIR0 (Kiruna, Sweden, 67.88 °N, 21.06 °E, 469.3 m) is typical for many stations in North Europe. Although its diurnal cycle is quite weak with an amplitude of only 0.1 kg m$^{-2}$ (1.7%), it is well fitted by the sinusoidal harmonic curve fitting.

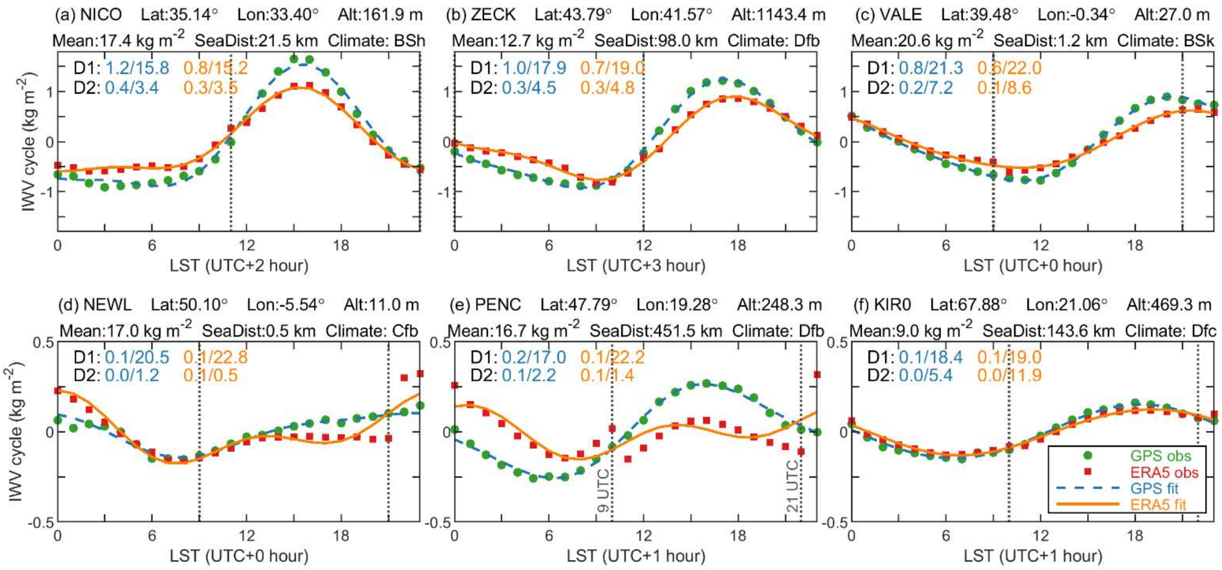


**Figure 8.** Diurnal IWV cycles at selected six stations obtained from 1-hourly GPS (green dots) and ERA5 (red squares). The stations are selected with the consideration of different altitudes, distance-to-sea (SeaDist), and climate zones classified according to Köppen Climate Classification (Beck et al., 2018). The data points are fitted with diurnal (D1) and semidiurnal (D2) harmonics (blue dashed curve for GPS and orange curve for ERA5). The amplitudes and phases of the D1 and D2 harmonics are also given. The phases are shown as the Local Solar Time (LST) at the peak of associated harmonics. For instance, the D1 amplitude and phase of GPS IWV at station NICO are 1.2 kg m$^{-2}$ and 15.8 LST, respectively. By comparison, the values of its ERA5 IWV are 0.8 kg m$^{-2}$ and 15.2 LST, respectively. The vertical black dotted lines at 09 and 12 UTC indicate the time of possible mismatches in the ERA5 IWV cycle.

### 4.2 Mismatches in diurnal ERA5 IWV cycle

Only the diurnal IWV cycle from the 1-hourly ERA5 time series was evaluated with respect to GPS, as the temporal resolutions of the other reanalyses are too coarse to characterise the diurnal cycle. However, we found significant shifts in the diurnal IWV anomalies between 09 and 10 (10-09) UTC as well as between 21 and 22 (22-

21) UTC at part of the stations, such as NEWL and PENC displayed in Fig. 8. The ERA5 developers have noticed such mismatches in the diurnal cycles of individual meteorological variables, such as its near surface wind, temperature, and humidity products (see Known Issues 8 and 9 in https://confluence.ecmwf.int/display/CKB/ERA5%3A+data+documentation#ERA5:datadocumentation-Knownissues), and the problem is attributed to the edge effect in each ERA5 assimilation cycle (from 10 to 21 UTC and from 22 to 09 UTC +1 day). However, according to our knowledge, the magnitude of the mismatch in the diurnal cycle of ERA5 IWV and its spatiotemporal characterisations have not been investigated yet. Therefore, we will quantify and analyse the mismatch in the ERA5 IWV over Europe in this section.

Figure 9 compares the diurnal IWV cycle from ERA5 and GPS for each season at station PENC. From this figure, we can derive that there are no mismatches in the diurnal GPS IWV cycle, although the GPS IWV (slightly) relies on $T_m$ (hence humidity and temperature) from ERA5 for the ZTD to IWV conversion, as shown in Eq. (2). GPS IWV estimates are therefore regarded as reference data to evaluate the mismatches in the diurnal cycle of ERA5 IWV. At station PENC, the mismatches in ERA5 are seasonal dependent, which are strongest in summer (JJA) but weakest in winter (DJF).

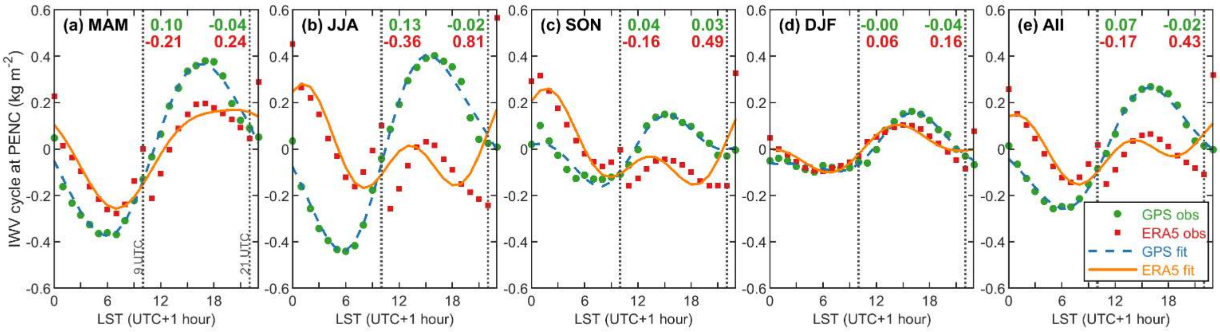


**Figure 9.** (a) Similar to Fig. 8 but for the seasonal and all-time averaged diurnal IWV cycles at station PENC. The green and red numbers are the IWV shifts from ERA5 and GPS, respectively. The numbers on the left and right are the IWV shifts from 09 to 10 UTC and from 21 to 22 UTC, respectively.

Figure 10 compares the shifts at 10-09 UTC and 22-21 UTC from ERA5 and GPS at all the stations, respectively. The shifts in the GPS IWV cycle are regarded as reference, representing the natural IWV changes at the two epochs. As can be seen from Fig. 10a–e and 10k–o, the ERA5 IWV series generally drop from 09 to 10 UTC and then jump from 21 to 22 UTC. The ERA5 artificial shifts at the two epochs are most significant in summer, with average values of -0.23 and 0.35 kg m$^{-2}$, respectively. In contrast, the average natural shifts in summer estimated from GPS IWV are only 0.11 and -0.08 kg m$^{-2}$ respectively. Moreover, the all-time averaged natural shifts in GPS IWV are only 0.05 and -0.05 kg m$^{-2}$, respectively. However, the artificial shifts in ERA5 IWV are -0.08 and 0.19, respectively. As can be seen from the geographic distributions of the shifts shown in Fig. 10e and 10o, the ERA5 shifts at 10-09 UTC are most significant at the Alps and Eastern Europe, whereas the shifts at 22-21 UTC are more widespread in Central Europe. The reasons for their geographical patterns are unknown and needs further investigation. Since the average diurnal amplitude of the reference GPS IWV is only 0.32 kg m$^{-2}$, the artificial shifts in ERA5 IWV cannot be ignored when analysing the diurnal IWV cycle in these regions.

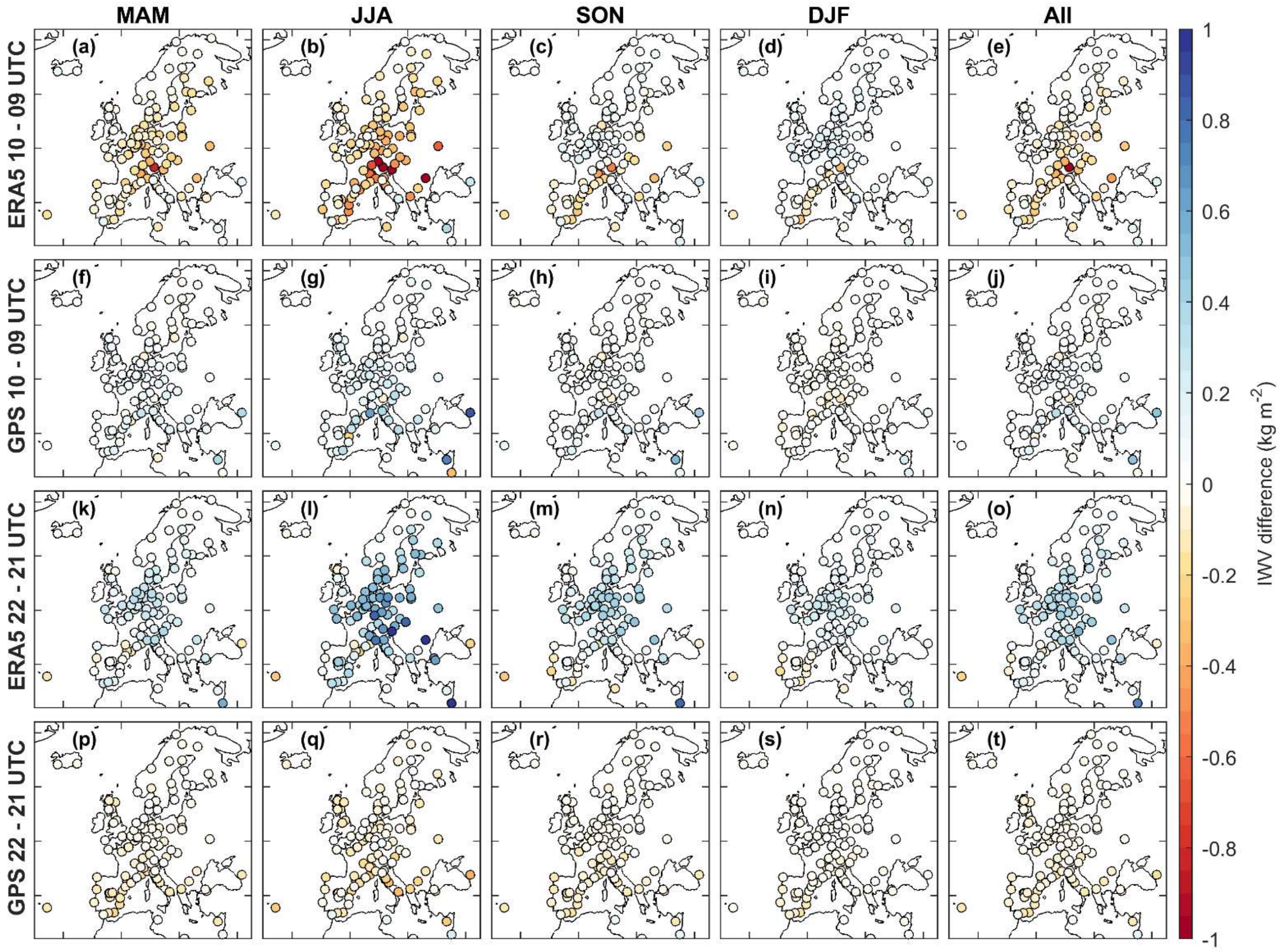


**Figure 10.** Seasonal and all-time averaged IWV shifts in the GPS and ERA5 diurnal cycles from 09 to 10 UTC and from 21 to 22 UTC.

**5.** L268: You find a correlation between the diurnal amplitude and the station height. Since station height (I guess) correlate with the site's distance to the ocean, another approach would be to correlate the amplitude with this distance. It is well known that the ocean (as long as there is no ice) acts like a low pass filter on daily variations in temperature and humidity.

**Reply:** Thank you for the constructive suggestion. According to your suggestion, in Section 4.1 Diurnal GPS IWV cycle, we classified the stations into three types, namely Inland, MedCoast and OtherCoast as shown in Fig. 7. We also analysed the characteristics of the diurnal IWV cycles by considering each station's altitude, and distance-to-sea, and climate zones. See our reply to your Comment #4.

**6.** L315: This whole section seems questionable if it is worth to be published? Do the GPS IWV data yield any new findings? Given the very high correlation between IWV from GPS and from the reanalyses, it seems as all the reported patterns, and their time dependences, will be seen by using reanalyses data only?

**Reply:** We agree with you that this part is out of the scope of this work, and thus it was removed. We will address the issues related to interannual variations of IWV and its relationship with extreme weather events in a separate study in the future.

## Technical Corrections

**7.** Line (L)1+: You use the American spelling of vapour, although ACP is a European journal?

**Reply:** Thank you for the suggestion. We used "vapour" and the style of English (UK) in the revised manuscript.

**8.** L97: ... IWV -using ... ?

**Reply:** Replaced with "by using".

**9.** L17: 2%-18% --> 2 %–18 % (similar changes to be carried out many times in the manuscript)

**Reply:** Replaced hyphen with en dash.

**10.** L154: IWVs --> The IWV values ?

**Reply:** Replaced as suggested.

**11.** L157: reanalyses Compared --> reanalyses. Compared

**Reply:** Modified as suggested.

**12.** L203: IWVs are --> IWV for all sites and days are ?

**Reply:** Modified as follows in Section 2.4 Pre-processing:

The daily mean IWV time series at each station is further aggregated into monthly mean IWV series if there are at least 15 data points available in a month.

**13.** L398: (29.5°E, 40.8°N), --> (29.5 °E, 40.8 °N), (see also L447-448)

**Reply:** Modified as suggested.

**14.** L444: 0-0,4 --> 0.0 – 0.4 ?

**Reply:** Replaced hyphen with en dash.

**15.** L446: 0,4-1 --> 0.4 – 1.0 ?

**Reply:** Replaced hyphen with en dash.

**16.** L480+: doi links are missing for almost all references and the established standard acronyms for journals are not used.

**Reply:** Added the doi links and used standard acronyms of the journals.

**17.** Figure 2: The yellow colour is not ideal. I suggest to use cyan or magenta instead. You may also consider to use darker colours in Figures 5 and 8. Different colours in these figures are not really needed for clarity, although it may look nicer compared to have it all in black.

**Reply:** We modified the colour scheme of Fig. 2. We also used darker colours in Figure 5.

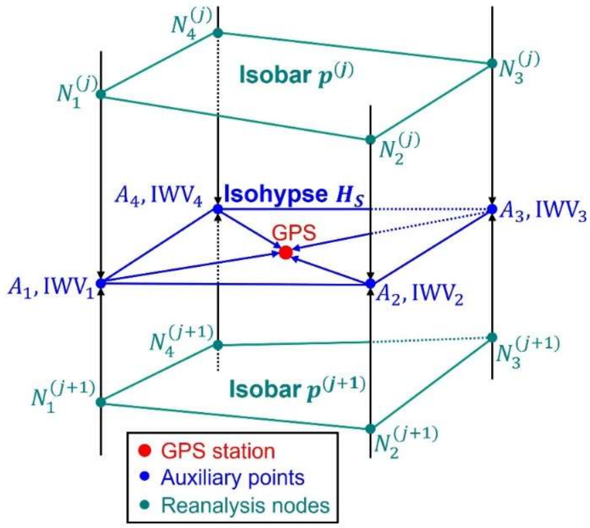


**Figure 2.** Schematic plot of the vertical and horizontal interpolation of the reanalysis pressure level products. The IWV at each auxiliary point ($A_i$; orange dots) is calculated with vertical interpolation or extrapolation of the adjacent reanalysis nodes ($N_i^{(j)}$ and $N_i^{(j+1)}$; green dots). The IWV at the GPS station (red dot) is then estimated with horizontal interpolation of the auxiliary points.

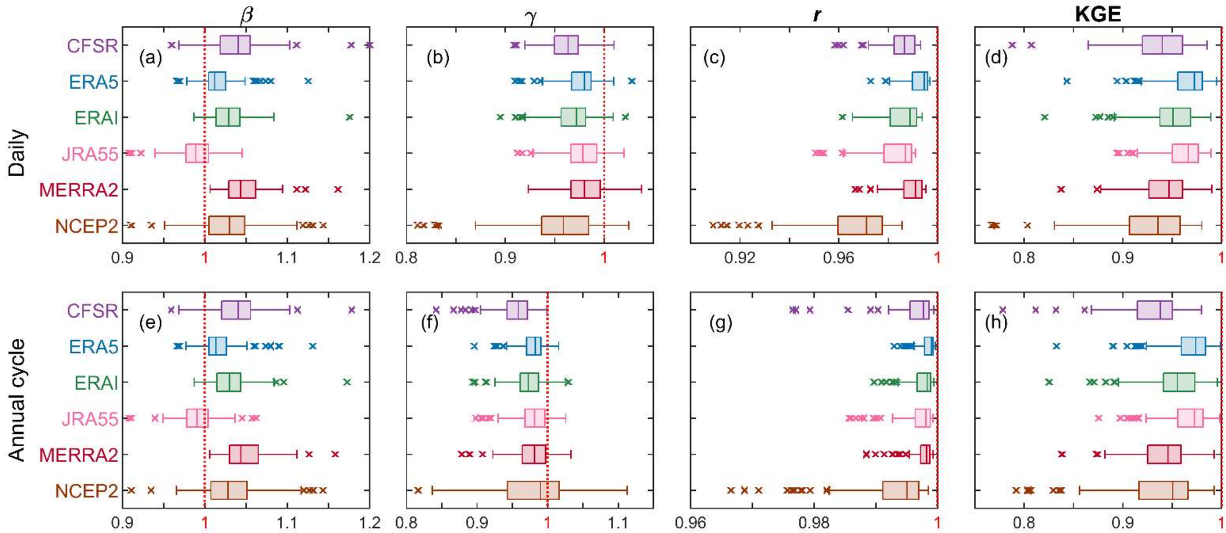


**Figure 5.** Box-whisker plots of the KGE parameters for the daily time series (a–d) and monthly annual cycle of IWV (e–h) from the reanalyses compared to GPS for the 108 stations.