Community Comment

General Comments

1. Thank you for presenting this new GNSS dataset, which holds potential for climate studies. I found the manuscript well-structured and written, and the methodology sound. However, I would like to share a few comments and questions regarding some of the methodological choices and propose suggestions to further enhance the QC/QA process and overall quality of your dataset.

Response: Many thanks for your thorough and encouraging review/assessment of our work. We appreciate your comments and suggestions. Our point-by-point responses to each comment are provided below.

Major Comments

1. Section 2.2 GNSS data processing: While you mention adhering to the highest standards in your study, it is worth noting that the analysis was conducted using Bernese GNSS Software version 5.2. Since 2022, version 5.4 has been available, introducing several improvements. These include enhanced observation (RINEX) quality control and preprocessing, improved ambiguity resolution for PPP, and updated tropospheric models such as VMF3. Considering these advancements might further strengthen the robustness and quality of your dataset.

Response: Many thanks for this constructive suggestion. We prepared the reprocessing campaign in early 2020 and commenced it soon after the release of the IGS Repro3 products, using Bernese V5.2. The end-to-end campaign (data processing followed by QC/QA and validation) spanned several years, and at the outset we had not yet obtained a license for Bernese V5.4. This, in fact, is the main reason Bernese V5.2 was adopted for the present work.

We fully acknowledge the advances in V5.4, including enhanced observation quality control, improved ambiguity resolution for PPP, and updated tropospheric models e.g., VMF3, and agree these can further enhance the robustness and accuracy of tropospheric estimates. Hence, since late 2023, we have been working on processing operational datasets (Final and Rapid ZTD results) using Bernese V5.4 to take advantage of these improvements. In addition, we are also planning a new reprocessing campaign with Bernese V5.4 that will include multi-GNSS observations (this release is GPS-only) and expand network coverage via collaboration with regional data centres.

To address your concern (together with the other reviewers' related comments), we have added a few sentences in Section 7 noting a crucial limitation of this release, i.e., the dataset was produced with Bernese V5.2 and GPS-only data and illustrating our ongoing and planned upgrades.

Lines 690-697

> "*Despite these advancements, several key challenges and opportunities for improvement remain.* ***First, while this study mainly employed GPS observations, integrating multi-GNSS systems such as Galileo, GLONASS, and BeiDou could improve satellite visibility and geometry, and may enhance spatiotemporal availability and robustness, particularly in under-represented regions like polar areas and oceans. However, as noted in Section 2.2, introducing additional constellations can impose inter-system biases and calibration complexities that may induce shifts in the time series. In other words, the net benefit is context-dependent and not yet settled. Given this, our ongoing research is conducting a new reprocessing campaign that will incorporate***

multi-GNSS observations using the latest Bernese V5.4 and updated tropospheric models like VMF3, while managing inter-system and inter-frequency biases, harmonising antenna calibrations and metadata, and ensuring cross-system consistency."

2. Section 2.3 Retrieval of PWV: the calculation of ZHD from the numerical integration of refractivity is not recommended (Jones et al., 2020, chap. 5.4.2), especially when only 37 pressure levels are available such as with ERA5. Instead, Saastamoinen formula should be used with surface pressure which can still be computed from ERA5 (with adequate interpolation between levels). This is the approach actually used by Haase et al. (2003) for their GNSS ZTD to IWV conversion. Note that they use only Eq. (2) for the integration of radiosonde profiles, which have many more vertical levels.

Response: Thanks for pointing this out. ZHD can be obtained either via the Saastamoinen formula using surface pressure or by numerical integration of refractivity from reanalysis profiles. The Saastamoinen formula is reliable where accurate surface pressure is available. However, for GNSS sites without co-located pressure sensors, surface pressure must be interpolated or extrapolated from reanalysis fields (e.g., ERA5), depending on the site's position relative to the lowest pressure level of ERA5. In regions with complex topography, interpolation or extrapolation errors can propagate directly into ZHD when using the Saastamoinen formula, because ZHD is proportional to pressure errors. This is the main reason we adopted the integration method using ERA5 profiles. Specifically, by using multiple atmospheric layers, it reduces sensitivity to single-level pressure mismatches and helps maintain consistency across the large global network. From another aspect, the Saastamoinen formula assumes hydrostatic equilibrium and may not fully characterise vertical atmospheric variability, particularly in regions with complex topography or under severe weather conditions.

While Jones et al. (2020) recommend the Saastamoinen formula over integration approach due to potential integration errors, the "net/actual" impact of the two ZHD pathways on PWV, especially for long-term trends/behaviours, remains insufficiently quantified. To the best of our knowledge, there is no definitive experimental evidence/statistics showing that the Saastamoinen formula with interpolated or extrapolated pressure systematically outperforms ZHD integrated from ERA5's 37-level profiles. In the revised version, we clarify our rationale in Section 2.3 and acknowledge in Section 7 that the choice of ZHD method may influence PWV at some sites and times, while emphasising the need for further study.

Lines 181-184

"The retrieval of PWV from ZTD requires the inclusion of meteorological parameters, specifically temperature and pressure, at the locations of GNSS sites. ***However, the absence of meteorological sensors at most stations presents a significant challenge in obtaining these parameters. To address this and maintain consistency across the global network, this study used atmospheric data from the high-quality ERA5 dataset to provide the necessary meteorological inputs.****"*

Lines 701-704

"Thirdly, the refinement of data retrieval techniques is necessary to address challenges posed by complex topographies and high-altitude regions, thereby improving robustness in these environments. ***In particular, the ZHD estimation choice adopted in this study may influence PWV at some certain sites and times, in our next reprocessing, we will***

document and benchmark the differences between approaches."

To support community assessment, we hope to release the data, including ZTD, ZHD, ZWD, PWV, and Tm), enabling users to recompute PWV with a Saastamoinen-based ZHD if preferred. We also plan to implement a side-by-side comparison of Saastamoinen- and integration-based ZHD within our next reprocessing campaign. If this is of interest, we would warmly welcome you (appreciate any guidance from you) and any interested colleagues to join this systematic evaluation.

3. Section 3.2 Screening based on GNSS-ZTD results only: proper credit should be given to Bock, 2020, who introduced the general approach and the methodology to choose range-check and outlier check limits for GNSS ZTD and formal errors which are actually followed here.
To clarify the purpose and usage of the formal errors for the screening process, it may be judicious to more section 4.1 here.

Response: Many thanks for your suggestion.

First, throughout the entire processing campaign over the past few years, we have indeed read a lot of publications from Prof. Bock. The range check and outlier detection framework we follow also builds on (Bock, 2020). In the revised version, we now cite the reference where the screening limits and methodologies are introduced to acknowledge this great contribution.

Second, after careful consideration and our internal discussion, we prefer to stick to the status quo and retain Section 4.1 within Section 4. Our rationale is to preserve a clear separation of roles and improve overall readability.

Specifically, Section 3 (data screening) documents how we process daily-solution screening based on coordinate repeatability (Section 3.1); ZTD outlier detection using temporal variability and formal errors (Section 3.2); and PWV screening (Section 3.3). In short, it presents methodologies and screening criteria. Section 4 (quality assessment) reports the distributions of formal errors and their temporal variation (Section 4.1); cross-comparison of PWV with external references (Section 4.2); and offset detection (Section 4.3). Despite the shared adoption of formal errors, Section 4 is designed to present results and interpretation. Moreover, keeping these statistics together avoids duplication and maintains a coherent narrative flow.

In response to your suggestions, we have made several modifications in Section 3.2 in the revised version. First, we now cite (Bock, 2020) alongside the description of the general screening method and the rationale for selecting range-check and outlier thresholds; Second, we add a sentence with an explicit cross-reference to Section 4.1; Lastly, and also based on your following Comment#12, we include a clarifying sentence on the screening limit:

Lines 230-235

"Following the coordinate repeatability evaluation, ZTD values underwent further screening utilising range checks and outlier detection, ***following the standardised approach outlined in (Bock, 2020)****. As the first step, ZTDs outside the range of 1–3 m (Bock et al., 2014) and those with formal errors (σ_ztd) exceeding 10 mm were excluded. Please note that* ***formal errors in the ZTD estimates are an important indicator of the quality of GNSS atmospheric parameters and are therefore widely used in screening. For context, it shows that 99.996% of formal errors are ≤ 10 mm in this work. More details about the analysis of formal errors are provided in Section 4.1."***

4. Line 225 to 260: referring to systematic biases may be misleading here. Referring to the

"consistency of ZTD estimated from collocated GNSS sites" as you do later on (Line 256) seems more precise.

Response: Thank you for this valuable clarification. In the revised version, we have gone through the relevant sentences and replaced the misleading phrases throughout:

Lines 242-243

> *"While this step ensures a refined ZTD dataset for PWV retrieval without requiring external reference models, e.g., ERA5, it still has several limitations, **particularly in detecting inconsistencies within ZTD time series**."*

Lines 274-276

> *"**Similar inconsistencies (biases exceeding 20 mm) were also identified** at four additional stations (LRA1, UTK1, UTK2, and CLS6) when compared to co-located stations and the ERA5 dataset."*

5. I have one concern here with the impact of height differences. As you mention on line 238, "discrepancies … are often attributed to height differences". This effect is actually expected. A simple rule of thumb approach predicts bias in ZHD around 10 mm and a few mm from ZWD as well based on a 50 m height difference. Such biases can be avoided by applying a proper vertical correction such as described in Bock et al., 2022. Following this approach can be very valuable in detecting station-specific biases when several nearby stations are available.

Response: Thanks for raising this concern. In this work, the ZTD comparison between co-located stations serves only as a "preliminary" screening/diagnostic to flag potential issues, with no sites excluded at this stage on the basis of co-located differences. The final assessment of data quality relies on the comprehensive comparison of GNSS-PWV against ERA5-PWV, which provides an independent external reference.

Regarding height effects, we totally agree that an appropriate vertical correction is valuable when several nearby sites are available and height differences are non-negligible. In this work, however, 93% of the co-located pairs of stations have height differences within 10 m, which corresponds to an expected ZTD difference of less than ~2 mm. This is well below the typical uncertainty of ZTD estimates. In addition, we also note that applying a vertical correction might introduce additional error, particularly under complex atmospheric conditions and in rugged topography. Hence, on this basis, we did not apply a vertical correction at this preliminary step.

To address your concern, we make some revisions in the manuscript: First, we explicitly state that no vertical correction was applied in the co-located station comparison and explain the rationale. Second, we recommend that users working with specific local networks apply vertical normalisation (Bock et al., 2022) where appropriate, or explore other vertical correction methods:

Lines 255-262

> *"After a detailed evaluation, discrepancies in ZTD between co-located GNSS sites are often attributed to height differences. **It should be noted that no vertical correction was applied in this study, as the co-location comparison serves only as a preliminary step to identify potential problematic stations and to provide a general indication of the quality of ZTD estimates. In addition, in this network, 93% of station pairs have a height difference within 10 m, corresponding to an expected ZTD difference of ~2 mm. A vertical correction procedure may also introduce errors of comparable or larger magnitude under certain conditions like in complex atmospheric conditions or over rugged topography. Nevertheless, for local analyses where height differences are***

non-negligible, we recommend applying vertical corrections following methods described in (Bock et al., 2022) or exploring alternative approaches."

6. Another concern is with the impact of equipment changes which can mask short-term site-specific biases when computed over long periods. This should mentioned here and also underlines the importance of offset detection that is discussed later.

Response: Many thanks for the valuable suggestion. In the revised manuscript, we now add a few sentences to note this explicitly and underline the importance of offset detection:

Lines 277-279

*"In addition, **equipment changes can introduce offsets, leading to inconsistencies at co-located sites and biasing long-term trend analyses by masking short-term, site-specific effects when statistics are aggregated over long periods. More details regarding the offset detection procedure adopted in our study is described in Section 4.3."***

7. Line 235 to 250: the discussion employs the terms "bias, deviations and differences", are these referring to mean values? Please clarify. Separating the mean and standard deviation of differences rather than using the RMS (which mixes both) would also give more insight into the nature of the differences.

Response: We appreciate this valuable suggestion. We have double-checked all the terminologies and now report standard deviation and bias separately throughout the revised version.

8. Line 260: "additional quality control" would better fit here in place of "additional screening".

Response: Amended in the revised version:

Lines 289-291

*"Therefore, to address these limitations, **additional quality control** of the dataset is crucial. This can be achieved by comparing ZTD values with an independent reference dataset, such as ERA5, to validate and enhance the overall quality of the results"*

9. Section 3.3 Screening based on comparison with reference PWV data: eliminating the negative IWV values may not be sufficient and may actually be hiding a more general bias between GNSS and ERA5 (possibly with seasonal variation). To avoid this caveat, it is preferable to compare ZTD values (as also recommended in Jones et al., 2020, chap. 5.4.1). Then you would probably notice a bias and decide to remove the flawed stations or find that representativeness errors in ERA5 are the limitation.

Response: Thank you for this constructive advice. We agree that removing negative PWV alone is not sufficient and may obscure broader GNSS–ERA5 biases. To address this and accommodate other use cases (also based on the other reviewers' comments), we made the following updates:

First, we now provide two versions of the final dataset (supplied to PANGAEA and uploaded to our online portal. *Note that the new datasets will become publicly available after the completion of the curation/review process, which may take a short while depending on the queue*):

(1) an "unfiltered" dataset containing all GNSS-PWV estimates after internal quality control only, without any ERA5-based outlier exclusion

(2) an ERA5-screened/filtered set in which ERA5 is used to remove outliers, see Section 3.3.

This allows users to select the version most appropriate for their applications, for example, when benchmarking GNSS against ERA5, we recommend the unfiltered "raw" dataset.

Second, we now add a few sentences to advise that ZTD comparisons are preferable for screening as they avoid additional conversion uncertainties and potential GNSS–ERA5 representativeness differences embedded in PWV, as per Jones et al. (2020). As indicated in our previous responses, the released products include ZTD, ZHD, ZWD, and PWV, enabling data users to conduct direct further checks/analyses.

Lastly, a similar ZTD comparison study is also under our consideration for our future work.

In general, as data contributors, these updates strike a balance between ensuring the integrity and accessibility of the dataset and encouraging data users to perform further in-depth and innovative analyses based on these data.

Lines 327-331

> "***Moreover, while we only adopt PWV for screening, we encourage ZTD-level comparisons in future analyses, as they avoid additional conversion uncertainties and potential GNSS–ERA5 representativeness differences (Jones et al., 2020). Accordingly, to accommodate various use cases, we provide two versions of the PWV dataset: an unfiltered product that contains all GNSS-derived PWV estimates after internal quality control, and an ERA5-screened product in which ERA5 is used only to flag and optionally remove gross outliers.***"

10. Line 274-292: I don't understand the rationale of comparing GNSS PWV at a target station to ERA5 PWV at nearby stations. Here you mix to types of errors: GNSS vs. ERA5 (different data sources) and target vs. nearby (difference due to distance). If this procedure is inspired by Nguyen et al., 2024, it is actually not relevant here. Please clarify or correct.

Response: Thanks for this comment. We definitely agree that mixing cross-source and cross-site differences would be problematic. However, our main intention, perhaps insufficiently clear in the previous statements, was to use only PWV differences at the target site as the decision variable, and to use nearby stations to stabilise the monthly dispersion estimates, thus avoiding the mixing of error types. In the revised version, we rephrase these statements and add some explanations:

Lines 303-313

> ***"Following the removal of negative PWV values, a robust outlier detection and elimination method was applied. This method comprises two steps: (1) identifying nearby sites and (2) establishing monthly, site-specific thresholds. First, for each station, nearby stations were identified within 2° in latitude and longitude and with a vertical separation less than 500 m. Next, for the target and each nearby station, we computed the differences between the GNSS-PWVs and ERA5-PWVs. For each month, these differences were pooled to estimate the distribution and define the monthly thresholds of the target station using the aforementioned IQR-based method, i.e.,*** $Q_1 - 3 \times IQR,\ Q_3 + 3 \times IQR]$***, where*** $IQR = Q_3 - Q_1$***, and*** Q_1 ***and*** Q_3 ***represent the 25th and 75th percentiles, respectively. The resulting monthly thresholds were then applied to the PWV time series of the target station to flag and remove outliers. This procedure was applied to all stations, yielding site-specific thresholds that account for local spatiotemporal variability, and repeated iteratively until no additional outliers were identified. This method provides more robust, locally representative thresholds than using only the PWV differences at the target station,***

which may fail to detect problematic results when large system inconsistencies exist, such as the PUB2 case in Section 3.1."

11. Figure 8: this PWV difference series is really suspect. It looks like the ERA5 values in your GNSS – ERA5 differences are very close to zero. Please check.

Response: Thanks for raising this concern. We have double-checked the computation and plotting of the differences between GNSS-PWV and ERA5-PWV and confirm that the ERA5 estimates are correctly used in the calculation, and there is no unintended zeroing. The "near-zero" phenomenon in this figure, in fact, mainly results from the scale being dominated by an anomalous surge in the GNSS-PWV time series at AC30 during 2018.

To further clarify this, Fig. 1 shows the time series of GNSS-PWV and ERA5-PWV at AC30. Very low PWV values are actually expected during the winter season at this site, given that it is located at a latitude of ~60° and an elevation of ~750 m (see Fig.2 for the picture of the site). ERA5-PWV generally remains below 10 mm. In 2018, however, it can be found in Fig.1 that GNSS-PWV rises rapidly to ~70 mm, well beyond what is climatologically plausible given its location/elevation.

For comparison, Fig. 3 shows results for a nearby site AC79, located ~25 km from AC30 but at lower elevation (~290 m), where GNSS-PWV and ERA5-PWV agree well and PWV typically remains below 30 mm. This comparison supports the interpretation that the very low ERA5-PWV at AC30 is reasonable, while the rapid increase of GNSS-PWV from ~10 mm to ~70 mm at AC30 is abnormal.

Although we have not yet identified the cause of this abnormal variation in GNSS-PWV, the ZTD results at AC30 provided by UNAVCO (Fig. 4) also exhibit a similar anomalous increase.

Accordingly, we have flagged AC30 as problematic in the text and will continue investigating the underlying cause in our future work.

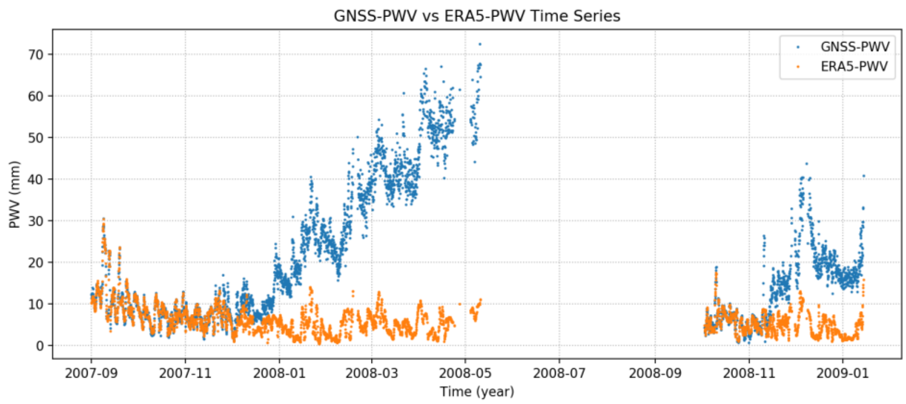

Fig. 1 PWV time series over AC30

Fig. 2 Picture of the AC30 station

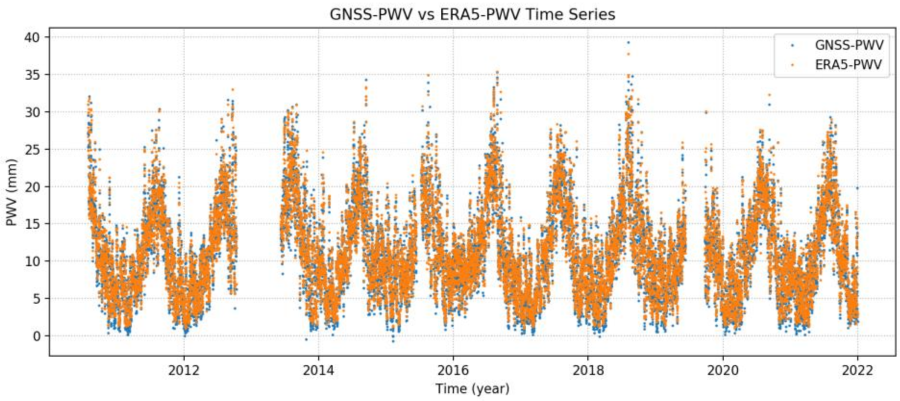

Fig. 3 PWV time series over AC79

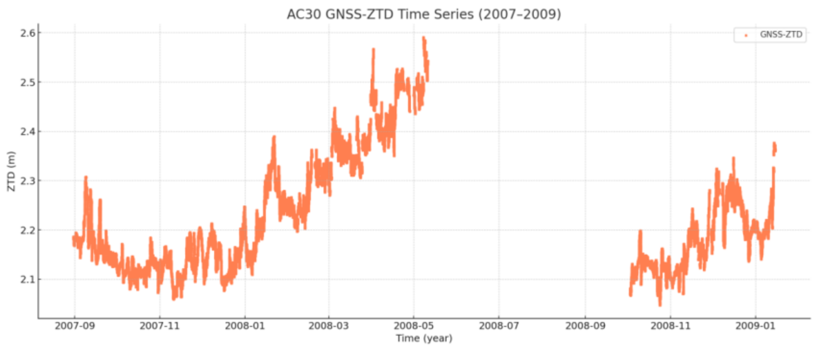

Fig 4 ZTD time series provided by UNAVCO

12. Section 4.1 Formal errors in ZTD estimations: consider moving this Section to Section 3.2. Give the % of the CDF corresponding to a formal error of 10 mm, which is the limit used for the range check in Section 3.2

Response: Many thanks for your suggestion.

First, the exact percentage value corresponding to a formal error of 10 mm is 99.996%. We have added the following sentences in the revised version:

Lines 340-343

"The majority of formal errors range between 0.5 mm and 2 mm, peaking at about 1 mm. The cumulative percentage curve (orange line) rises steeply, reaching 90 % at 2 mm and 99.73 % at 5 mm. The mean and median values of these errors are 1.38 mm and 1.23

mm, respectively. ***Beyond the X-axis range shown in Fig. 9, as indicated in Section 3.2, this curve attains 99.996% at 10 mm."***

Second, as suggested, 10 mm is the determined screening limit used in Section 3.2, hence we have added a contextual sentence there for clarity:

Lines 230-235

> *"Following the coordinate repeatability evaluation, ZTD values underwent further screening utilising range checks and outlier detection, following the standardised approach outlined in (Bock, 2020). As the first step, ZTDs outside the range of 1–3 m (Bock et al., 2014) and those with formal errors (σ_ztd) exceeding 10 mm were excluded. Please note that formal errors in the ZTD estimates are an important indicator of the quality of GNSS atmospheric parameters and are therefore widely used in screening.* ***For context, it shows that 99.996% of formal errors are ≤ 10 mm in this work. More details about the analysis of formal errors are provided in Section 4.1."***

Regarding the structure of the manuscript, as also explained in our responses to Comment#3, after careful consideration and our internal discussion, we decided to still retain the results and statistics in Section 4 (Quality assessment). This placement preserves a clear division of roles, i.e., Section 3 documents methods and screening criteria and Section 4 illustrates outcomes and interpretation, as well as improves readability and narrative flow. To address your suggestion and aid navigation, Section 3.2 now explicitly cross-references this subsection.

13. Line 299: "The formal errors of the estimated ZTD are known to play a key role in analysing the quality of GNSS". Although formal errors may help in the QC/QA of GNSS ZTD estimates (Bock, 2020), it is an overstatement to say that they play a key role. Please moderate.

Response: Thanks for your reminder. We now revise this sentence to:

Lines 338-339

> *"The formal errors of the estimated ZTD are* ***a useful indicator*** *for analysing the quality of GNSS atmospheric parameters (Bock et al., 2020)"*

14. Section 4.2 Cross-comparison of PWV with external references.
Line 232: the description of ERA5 (number of pressure levels, horizontal resolution, etc.) should be given earlier, e.g. in Section 2.3 when ERA5 is first introduced and used.

Response: We are grateful for this constructive advice. These descriptions have been moved to Section 2.3 in the revised version.

15. Line 383 and 405 + Line 229 (Section 3.2): explain why you chose three different collocation limits for these comparisons (GNSS vs GNSS, GNSS vs. VLBI, and GNSS vs. RS).

Response: The use of different collocation limits is due to these comparisons have distinct sensing geometries and data-availability constraints, with diverse implications for representativeness error.

First, regarding the two comparisons GNSS vs. GNSS and GNSS vs. VLBI, both techniques sense ZTD at fixed ground-based stations with comparable theory. We therefore apply a tight collocation criterion (≤1 km horizontal and ≤50 m vertical separation) to minimise representativeness errors.

Second, regarding GNSS vs RS, RS-derived ZTD/PWV are determined from vertical atmospheric profiles along a balloon ascent that typically drifts tens of kilometres over 1–2 hours. Imposing a very small (such as ≤1 km) horizontal limit adds little value for representativeness here and would severely constrain coverage, e.g., only 22 pairs of co-located GNSS and radiosonde stations would remain with such horizontal limit. To balance representativeness with adequate coverage, we use a horizontal limit of 50 km, yielding 402 pairs of co-located GNSS-RS stations globally. Moreover, this threshold is consistent with many previous studies conducting such intercomparisons (often 50–100 km).

In summary, we use a tight limit when both sensors are fixed, zenith-looking instruments, and a broader, yet standard, limit for GNSS–RS to reflect the balloon's moving sampling volume and to maintain adequate spatial coverage.

In the revised version, we add a few sentences in Section 4.2.3 to clearly illustrate this:

Lines 439-445

> *"It is worth noting that **we adopt a different collocation limit here than for the GNSS–GNSS and GNSS–VLBI comparisons as radiosonde-derived PWV is obtained along a balloon ascent that typically drifts tens of kilometres over 1–2 hours. Imposing a very small horizontal limit would add little value for representativeness and would severely constrain coverage (e.g., with the same horizontal limit, only 22 GNSS–radiosonde pairs would remain). To balance representativeness and sampling, we therefore use a horizontal limit of 50 km.** Under these criteria, we identified 402 GNSS–radiosonde pairs, with the number of paired PWV samples ranging from 888 to 23,749 (with an average of 7283 samples), equivalent to ~10 years of observations per station."*

16. Units for ZTD and PWV comparisons: use mm for ZTD and kg/m2 for PWV to avoid confusion. I am wondering whether some of the published GNSS vs. VLBI comparisons are not cited in the wrong unit (e.g. large biases and RMS values).

Response: Thanks for your suggestion.

We acknowledge that IWV/PWV can be expressed in kg/m² or mm, as per many WMO guidelines (https://library.wmo.int/records/item/68695-guide-to-instruments-and-methods-of-observation?offset=3) and also some official resources (https://www.meteo.be/en/research-themes/water-vapour/). On account of the unit of mm is widely used in a substantial number of existing publications, we decide to still report PWV in mm in this study to maintain consistency with much of the literature (including those cited in this work) and aid interpretability. To address your concern and ensure clarity, we now state this equivalence explicitly at first mention in Section 2.3.

Lines 207-208

> *"Note that, in this study, PWV values are reported in mm, numerically equivalent to kg/m^2, i.e., 1 mm = 1 kg/m^2, for readability and consistency."*

Second, regarding your concern about the units in the GNSS–VLBI citations (if we understand it correctly), we have double-checked all referenced studies. We identified an issue, i.e., the statistics reported by Steigenberger et al. (2007) should be ZWD, instead of ZTD. This has been amended in the revised manuscript:

Lines 423-424

> *"For example, Steigenberger et al. (2007) reported a **ZWD** bias of 7.2 mm and an RMS of 14.1 mm, ..."*

17. Section 4.3 Offset detection: my main concern here is that you applied the detection to a

subset of stations only (less than 50% of all your stations). Considering the use of this dataset for climate studies (e.g. trend analysis), this is a serious drawback. Please explain why not all stations have been checked and whether you intend to further complete this.

Response: As stated in the manuscript, "*2485 sites with observation periods exceeding 10 years and data missing rates below 20 % were selected*". That is said we applied two eligibility criteria (record length and data integrity) before conducting offset detection. This choice also reflects the requirements of the adopted methodology. Specifically, sufficient length and continuity are needed to robustly distinguish changes from seasonal variability and serially correlated noise. When time series are short or gappy false-positive rates increase and breakpoint magnitudes become poorly constrained, which risks over-flagging and over-correction. Hence, long and relatively complete records are the primary candidates for trend estimation and homogenisation. In addition, a further practical consideration is that documented equipment/firmware changes (from available log files) are uneven across the global network, and sites with sparse metadata reduce our ability to validate breaks and risk spurious classifications

We recognise the importance of broader coverage for climate applications and, as records lengthen and metadata improve, in our ongoing work, we intend to extend the detection to additional sites and evaluate methods better suited to shorter time series.

18. Another couple of questions concern the confidence that can be placed in the PMTred method and the validation of the detected change-points. First, the number of change points with this method seems a little underestimated compared to studies using other methods (Venema et al., 2012; Van Malderen et al., 2020; Nguyen et al., 2021). Second, the validation with GNSS metadata is not robust as some equipment changes may be missing and not only equipment changes are suspected to produce offsets, but also environment changes. In addition, the numerous firmware changes are not expected to have a significant impact and instead lead to accepting many false detections. To overcome these limitations, it may be advisable to cross-compare the results from different detection methods and to implement a more robust validation method, e.g. based on multiple pairwise comparisons (Caussinus and Mestre, 2004; Menne and Williams, 2009; Nguyen et al., 2024).

Response: Thank you for these thoughtful suggestions. We agree that both the choice of method and validation strategy can affect the number of detected changepoints.

First, as contributors in a data description paper, our implementation is intentionally conservative. We analyse monthly PWV differences and require strong statistical evidence: a 95% critical value for changes corroborated by metadata and 99.9% for unrecorded changes. Thie approach reduces false positives in short or noisy records but can achieve fewer detections than methods tuned for higher sensitivity.

Second, in this study, metadata are adopted only for corroboration, not as a substitute for statistical evidence. A recorded change is tagged "documented" only if a statistically significant break is also detected. Conversely, metadata entries do not, by themselves, trigger a break. In addition, we also acknowledge that environmental changes may cause offsets and that site logs can be incomplete.

As in the current release, all detected changepoints are provided as flags, and we do not adjust the data. This allows users to apply their homogenisation strategies, while we expand the method and coverage in our subsequent versions/updates.

To address your concern, we have added these sentences in the revised version:

Lines 515-523

"For clarity, the detected changepoints are provided as flags alongside the PWV series, and the archived PWV time series are not modified based on these detections. Although ERA5 is used as a reference to aid detection, it is not the actual "truth", as previous studies suggest that both reanalysis and GNSS data may contain inhomogeneities (Bock and Parracho, 2019; Zhang et al., 2019; Yuan et al., 2025). Moreover, in this study, we did not perform a separate changepoint search on ZTD, ZHD or Tm. Since PWV is derived from these parameters, any discontinuity in them can induce a corresponding offset in PWV, and we will further examine these in detail in future updates. ***Additionally, to strengthen robustness, our future releases will cross-compare multiple detection methods (Van Malderen et al., 2020; Quarello et al., 2022; Nguyen et al., 2025) and adopt relative-homogeneity checks based on multiple pairwise comparisons (Caussinus and Mestre, 2004; Menne and Williams, 2009; Nguyen et al., 2024)."***

19. Regarding alternative methods, Van Malderen et al., 2020, evaluated several of them in the similar context as the present study. Their benchmark study showed that PMTred, which is based on tests, is not performing well with the type of data used here (GNSS minus reanalysis differences). The best methods are indeed based on penalized maximum likelihood. One of them was first published under the name GNSSseg by Quarello et al., 2022, and recently updated and renamed PMLseg by Nguyen et al., 2025. This method uses penalized maximum likelihood and was specially developed for the segmentation of GNSS minus reanalysis differences. The authors may find it interesting as an alternative or simply for cross-checking their PMTred results.

Response: Many thanks for highlighting alternative offset-detection approaches. We are aware of the study by Van Malderen et al. (2020), which shows limited performance of PMTred for GNSS–reanalysis differences and indicates that penalised maximum-likelihood methods perform best. In line with this, we will cross-check our results using the GNSSseg/PMLseg (Quarello et al., 2022; Nguyen et al., 2025) and report the level of agreement (e.g., precision/recall of detected offsets and timing differences) in future releases.

As noted in our response above, we now add text acknowledging method sensitivity and outlining our plan to cross-compare multiple methods and adopt relative-homogeneity, multi-pair validation for corroboration. The current offset flags are provided alongside the PWV time series and we will update these files and document agreement metrics as the homogenisation effort progresses.

20. Figures 20 and 21: the interpolation of 2D fields over the ocean looks very unrealistic. It may be preferable to mask the oceans in these figures.

Response: Thank you for the constructive suggestion. Based on your (*mask ocean areas*), the two figures have been refined accordingly:

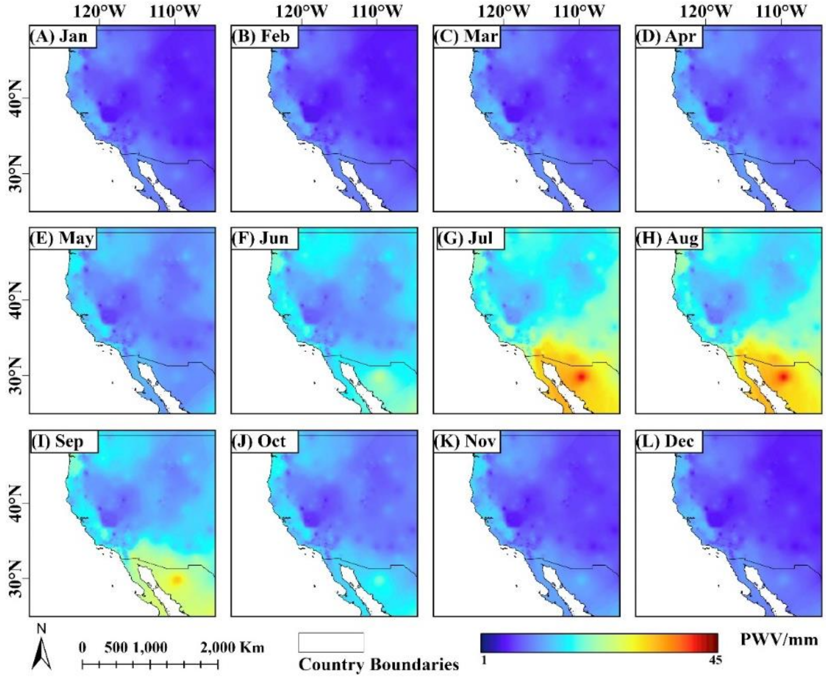

Figure 20. Rendered images of the 16-year climatological monthly mean PWV values for each month, derived from 590 GNSS sites located on the West Coast of the United States.

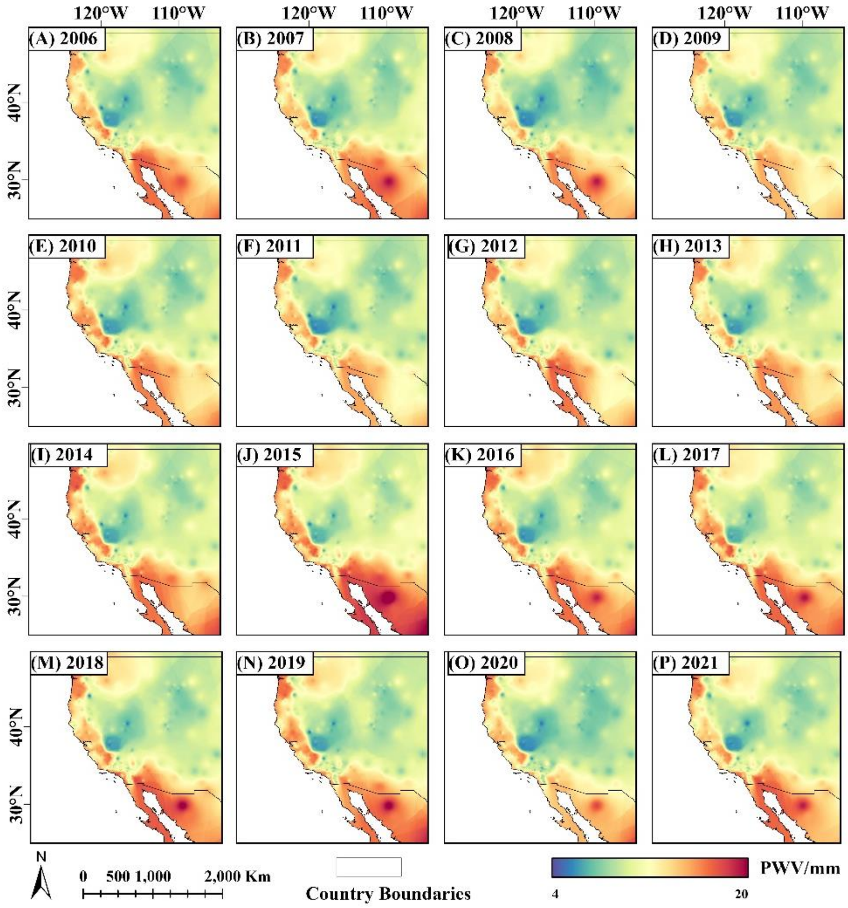

Figure 21. Rendered images of the annual mean PWV values for each year over the 16-year period 2006–2021, calculated from 590 GNSS sites located on the West Coast of the United States

21. Line 596: I was wondering if ENSO was not a more likely explanation for the interannual variability observed in these figures.

Response: Thank you for this insightful suggestion. The 2015–16 El Niño likely also contributed to the interannual variability. As noted in the manuscript, the "Blob" is an extratropical North Pacific marine heatwave and is distinct from (but interacting with) the ENSO, which is a tropical Pacific mode. Because these events co-occurred in 2015, and we cannot isolate their contributions to the phenomenon with the present analysis, we have moderated the wording (e.g., using "likely") and now state explicitly that both processes probably contributed to the observed peak. In addition, we have also added some supporting references here:

Lines 646-649

> *"This peak likely represents the co-occurrences of the strong 2015/16 El Niño (L'Heureux et al., 2017) and the North Pacific marine heatwave known as "the Blob", a significant mass of relatively warm water in the northeast Pacific Ocean off the coast of the United States (Bond et al., 2015; Di Lorenzo and Mantua, 2016; Peterson et al., 2015). Together, these phenomena generated positive temperature anomalies exceeding 2.5 °C, with the warm ocean surface heating the overlying atmosphere."*

22. Section 7: Summary and Outlook: the text of this section seems to overstate a little the quality of the dataset given the mentioned limitations.

Response: We have moderated the wording in the revised version to avoid overstatement and have made the key limitations explicit:

Lines 679-689

> *"This study has produced a global GNSS climate data record to help address data gaps in existing climate observing networks. Spanning up to a 22-year period from 2000 to 2021, the dataset includes hourly ZTD and PWV estimates from 5085 sites, providing broad spatiotemporal coverage and good overall accuracy globally. Advanced data reprocessing strategies, aligned with the IGS standards, were used to promote the consistency and accuracy of the generated atmospheric parameters, enhancing their suitability for climate applications. The quality of the dataset was evaluated via a rigorous quality assessment framework and cross-comparisons with various external references, including ERA5 reanalysis dataset, sounding profiles, and VLBI measurements. Generally good agreement across these datasets was demonstrated, with consistent water vapour estimates across diverse geographic and climate conditions. The dataset represents a critical step in GNSS climatology, offering valuable insights into the spatiotemporal variability of atmospheric water vapour. Further analyses of diurnal, monthly, seasonal, and annual variations in ZTD and PWV highlighted their importance in understanding climate variability, including responses to weather extremes and long-term climate trends.*
>
> *……*

Lines 711-715

> *Overall, the generated dataset represents a meaningful step toward fully harnessing the transformative potential of GNSS atmospheric monitoring techniques for advancing climate and atmospheric studies. By addressing critical challenges and leveraging cutting-edge methods, this dataset provides a reference for GNSS climatology, offering a foundation for future research and operational applications across this interdisciplinary field. These contributions may enhance our understanding of atmospheric dynamics,*

supporting sustainable development and facilitating informed decision-making."

23. While the dataset could benefit from further homogenization and the use of the latest GNSS processing software, this work represents a valuable contribution that will enrich discussions and collaborations within the IAG community (e.g. within the ICCC joint working group C.8 on Optimal processing and homogenization of GNSS-PW climate data records).

Response: We sincerely appreciate your encouraging and constructive comments. They are indeed helpful in further improving the quality of the manuscript and the generated dataset. We note the relevance to the ICCC joint working group C.8 and, in parallel, we are participating in a related working group (4.2.5) under the IAG frame. We are truly looking forward to future collaboration opportunities and hope to make our humble contributions to the community.

24. References

Bock, O. (2020) Standardization of ZTD screening and IWV conversion, in: Advanced GNSS Tropospheric Products for Monitoring SevereWeather Events and Climate: COST Action ES1206 Final Action Dissemination Report, edited by Jones, J., Guerova, G., Douša, J., Dick, G., de Haan, S., Pottiaux, E., Bock, O., Pacione, R., and van Malderen, R., chap. 5, pp. 314–324, Springer International Publishing, https://doi.org/10.1007/978-3-030-13901-8_5.

Bock, O., Bosser, P., and Mears, C. (2022) An improved vertical correction method for the inter-comparison and inter-validation of integrated water vapour measurements, Atmos. Meas. Tech., 15, 5643–5665, https://doi.org/10.5194/amt-15-5643-2022

Caussinus, H. & Mestre, O. (2004) Detection and correction of artificial shifts in climate series. Journal of the Royal Statistical Society Series C: Applied Statistics, 53(3), 405–425. Available from: https://doi.org/10.1111/j.1467-9876.2004.05155.x

Jones, J., Guerova, G., Douša, J., Dick, G., de Haan, S., Pottiaux, E., Bock, O., Pacione, R., van Malderen, R. (Eds.) ; Advanced GNSS Tropospheric Products for Monitoring Severe Weather Events and Climate: COST Action ES1206 Final Action Dissemination Report (2020), Springer International Publishing: Cham, Switzerland, 2020 ; https://doi.org:10.1007/978-3-030-13901-8.

Menne, M.J. & Williams, C.N. (2009) Homogenization of temperature series via pairwise comparisons. Journal of Climate, 22(7), 1700–1717. Available from: https://doi.org/10.1175/2008jcli2263.1.

Nguyen, K. N., Bock, O., & Lebarbier, E. (2024). A statistical method for the attribution of change-points in segmented Integrated Water Vapor difference time series. International Journal of Climatology, 44(6), 2069–2086. https://doi.org/10.1002/joc.8441

Nguyen, K. N., Bock, O., & Lebarbier, E. (2025), PMLseg: a R package for the segmentation of univariate time series based on Penalized Maximum Likelihood. https://github.com/khanhninhnguyen/PMLSeg/tree/main

Quarello, A., Bock, O., Lebarbier (2022) GNSSseg, a statistical method for the segmentation of daily GNSS IWV time series, Remote Sensing, 14, 3379, https://doi.org/10.3390/rs14143379

Van Malderen, R., E. Pottiaux, A. Klos, P. Domonkos, M. Elias, T. Ning, O. Bock, J. Guijarro, F. Alshawaf, M. Hoseini, A. Quarello, E. Lebarbier, B. Chimani, V. Tornatore, S. Zengin Kazancı, J. Bogusz (2020). Homogenizing GPS integrated water vapor time series: benchmarking break detection methods on synthetic datasets. Earth and Space Science, 7, e2020EA001121. https://doi.org/10.1029/2020EA001121

Venema, V., Mestre, O., Aguilar, E., Auer, I., Guijarro, J. A., Domonkos, P., et al. (2012). Benchmarking monthly homogenization algorithms. Climate of the Past, 8, 89–115. https://doi.org/10.5194/cp-8-89-2012.

Response: Many thanks for providing the detailed reference list. Most of these references are now being appropriately cited in the revised version.