

Reviewer#01 Comment:

The work presented in the manuscript address the possibility to improve the accuracy of the conversion from Zenith Wet Daley (ZWD), estimated from observations with a ground-based GNSS station, to Integrated Water Vapour (IWV) by adding data from a microwave radiometer capable of providing information about the temperature profile in the atmosphere. Although these results are of course unique for this specific station, the findings are primarily of interest for users that have (or are considering to buy) this type of radiometer. It is a much larger investment compared to the GNSS ground station itself. In any case, because as far as I know it is a new concept, and it makes sense that the study is made available for the entire scientific community.

My criticism is that when reading the manuscript it gives the impression that it is the uncertainties in the mean temperature that is the limiting factor for the quality of the IWV. I like to point out that even during extreme conditions (regarding the temperature profile) the uncertainty in the final estimate of the IWV, the Zenith Total Delay (ZTD) is at least as important. See e. g. Table 4 in: Ning et al. (2016). The uncertainty of the atmospheric integrated water vapour estimated from GNSS observations, Atmos. Meas. Tech., 9, pp. 79-92, <https://doi.org/10.5194/amt-9-79-2016>. Below I suggest alternative wordings in order to nuance this issue.

Response: We sincerely thank the reviewer for their encouraging assessment of our work and for recognizing the value of this synergistic concept for the broader scientific community. We also deeply appreciate the reviewer's constructive and highly accurate criticism regarding our initial framing of the error budget.

The reviewer is entirely correct. In our effort to highlight the specific thermodynamic errors associated with static climatological models, the original draft inadvertently gave the impression that the weighted mean temperature (T_m) is the sole or universally dominant limiting factor for Integrated Water Vapor (IWV) quality. We completely agree that the uncertainty originating from the Zenith Total Delay (ZTD) estimation—stemming from mapping functions, satellite orbits, clocks, and site-dependent electromagnetic effects—is at least as important, and frequently exceeds the error introduced by the T_m conversion factor.

We thank the reviewer for directing us to the comprehensive uncertainty framework presented in Ning et al. (2016). We have incorporated this vital reference into our introduction and uncertainty analysis sections to provide a much more balanced and scientifically accurate picture of the end-to-end GNSS-IWV error budget. Furthermore, we strongly agree with the reviewer's practical insight regarding the high capital investment required for radiometric hardware; we have added explicit commentary in our conclusion emphasizing that this approach is best targeted at specific climatological hotspots rather than universal deployment. Throughout the revised manuscript, we have eagerly adopted the reviewer's suggested wordings to thoroughly nuance our claims.

Rain is only mentioned in terms of extreme weather in the introduction, but it is not pointed out that the microwave radiometer algorithms more or less break down during rain. I assume that the station on Cyprus may not be exposed to rain during a large percentage of time, and perhaps the extreme temperature profiles (difficult to model based on ground temperature only) which the study is focused on, do never occur during rain? In any case, I think it is important to mention the poor accuracy of the radiometer retrievals during rain.

Response: We thank the reviewer for highlighting this critical operational limitation of microwave radiometry. The reviewer is absolutely correct that MWR retrieval algorithms degrade significantly during precipitation events due to liquid water emission and scattering on the radome. As the reviewer correctly hypothesized, the hydro-climate of Nicosia during our study period (the warm season, March to October 2025) is characterized by a prolonged dry season, meaning rainfall events were exceedingly rare and did not coincide with the extreme thermodynamic profiles under investigation. To ensure these wet-radome errors did not contaminate our dataset, we utilized the co-located Vaisala WXT536 weather transmitter (which records rain accumulation and duration) to identify and filter out any highly infrequent rainy epochs prior to our statistical analysis. We agree that this quality-control step and the inherent limitations of MWRs during rain should be explicitly stated for transparency. We have updated Section 2.2.1 to reflect this.

Specific comments

Line (L) 10: fundamentally limited --> significantly affected

Response: Corrected in the revised manuscript.

L 23: essential --> meaningful

Response: Corrected in the revised manuscript.

L 57: Here you state that "T_m is the primary source of uncertainty in GNSS meteorology after ZTD estimation". It will be appropriate to either (1) also here point out that the uncertainty in the ZTD is larger or comparable when it comes to the impact on the uncertainty of the final IWV, or (2) replace "the primary" with "one".

Response: The reviewer is entirely correct that the uncertainty originating from the ZTD estimation itself—particularly from mapping functions and the interpolation of surface pressure required for Zenith Hydrostatic Delay (ZHD)—frequently matches or exceeds the error introduced by the T_m conversion factor. While our specific CYGMEN site mitigates much of the ZHD error via a co-located pressure sensor (as detailed in Section 3.4.3), we agree that the generalized statement in the introduction needed to be refined for broader scientific accuracy. We have adopted the reviewer's suggestion by replacing "the primary" with "a primary" and adding a clause explicitly acknowledging that ZTD uncertainties are often comparable or larger.

L 60: Assuming that 1-2 % error corresponds to 1-2 mm in the PWV implies that the value of the PWV is 100 mm. I am not sure if such a high value has ever been observed. Why not just state the percentage uncertainty and let the reader figure out what it means in absolute value given the weather conditions.

Response: We sincerely thank the reviewer for catching this mathematical oversight. The reviewer is absolutely correct: equating a 1-2% relative error directly to a 1-2 mm absolute error implies an

unrealistic baseline PWV of 100 mm, which does not occur in nature. We have removed the erroneous absolute equivalent from the manuscript and adopted the reviewer's suggestion to state only the percentage uncertainty, allowing it to scale properly with the ambient meteorological conditions.

L 84: It will be interesting to know how many sites with co-located MWR and GNSS there are in the network?

Response: We thank the reviewer for pointing this out; we realize our original phrasing may have been ambiguous regarding the network's scale. While the CyMETEO network consists of a dense array of continuous GNSS stations distributed across the island, there is currently only one site equipped with a co-located Microwave Radiometer (MWR) and Radiosonde (RS) facility. Because MWRs are high-cost and operationally complex instruments, this single co-located setup at the Athalassa observatory serves as the project's central thermodynamic "supersite." We have revised the manuscript to explicitly clarify the number of MWR instruments and the exact architecture of the network.

Table 1: It will make sense to also state that the MWR "Role in Study" is also to provide the IWV.

Response: We appreciate the reviewer pointing this out. Given that Section 3.2 extensively utilizes the MWR-derived IWV as a high-quality comparative standard against both the radiosonde and GNSS retrievals, it is entirely appropriate to list IWV estimation as one of the instrument's primary roles. We have updated Table 1 accordingly to provide a more complete summary of the MWR's utility in this study. Table 1: You state that the humidity uncertainty of the RS humidity sensor is 4 %. This deserves some comment in the text when you use this sensor as "ground truth" given that you evaluate uncertainties at comparable and lower levels.

Response: The 4% uncertainty listed in Table 1 represents the manufacturer's specified uncertainty for individual, point-wise RH measurements. However, because our validation primarily relies on column-integrated quantities (IWV) and weighted vertical integrals (T_m), the random, uncorrelated measurement noise at discrete vertical levels is significantly suppressed during the mathematical integration process. While systematic sensor biases could theoretically persist, the Vaisala RS41 is widely established in the literature as the operational gold standard for in-situ profiling, with excellent systematic stability. We have added a brief discussion in Section 2.2.2 to explicitly address how error propagation during integration allows us to confidently utilize the RS as a "ground truth" benchmark for evaluating smaller relative uncertainties in the MWR and GNSS retrievals.

L 155: The IGS Ultra-Rapid products are not the optimal choice when aiming at the lowest possible uncertainty which should be the case for climate studies. Please comment on the impact of this choice and why it was done this way.

Response: The reviewer is entirely correct that for post-processed, long-term climate trend analysis, IGS Final products are the optimal and necessary choice. Our decision to utilize IGS Ultra-Rapid products was driven by our secondary objective to evaluate the proposed synergistic MWR-GNSS architecture under simulated real-time operational conditions. As discussed in our introduction, a major application of high-resolution PWV is the assimilation into NWP models for short-range precipitation "nowcasting." By using Ultra-Rapid products, we aimed to demonstrate that replacing static T_m models with real-time MWR data provides massive accuracy gains even when constrained by operational, near real-time orbital products. Regarding the impact: the degradation in ZTD accuracy when switching from Final to Ultra-Rapid products is typically on the order of a few millimeters, translating to less than 0.5 mm in PWV. As our uncertainty analysis demonstrates, this orbital/clock uncertainty is an order of magnitude smaller than the severe systematic biases (> 1.0 mm PWV) induced by the static empirical thermodynamic models during peak heating. Nevertheless, we agree this rationale and caveat must be explicitly stated. We have updated Section 2.2.3 to clarify this choice and its minimal impact on our thermodynamic error diagnosis.

L 182: What are the estimated values of alpha and beta?

Response: We thank the reviewer for pointing out this omission. The specific numerical values of the regression coefficients α (1.0623) and β (-15.6062) were inadvertently left out of the initial manuscript draft. We have now updated Section 2.3 to explicitly state the derived parameters used for the linear bias correction model.

Figure 4(b): Which estimate of T_m is used in this comparison?

Response: We thank the reviewer for pointing out this missing detail. For the GNSS IWV calculations shown in Figure 4(b) (and 4(c)), the weighted mean temperature (T_m) was derived using the static HGPT2 empirical model, representing our "Standard Retrieval" baseline. We apologize for not making this explicit in the figure's caption. We have updated the caption for Figure 4 to clearly specify the source of the T_m used for the GNSS data.

L 299: This result is already stated in L 288.

Response: The reviewer is completely correct; the concluding sentence was an artifact of the drafting process and merely repeated the finding that had already been established earlier in the section. We have deleted the redundant sentence to improve the flow and conciseness of the paragraph.

L 341: Here again it is implicitly assumed that the error in the ZTD is zero, which is not true. Please rewrite.

Response: We appreciate the reviewer's in catching this phrasing. Similar to the earlier comment regarding the introduction, our text inadvertently implied a flawless ZTD estimation by focusing solely on the thermodynamic conversion step. We completely agree that the total PWV error is a combination of both ZTD estimation errors and T_m inaccuracies. We have rewritten this section to explicitly state that the linear dependency on T_m applies specifically to the conversion factor (Π), under the assumption of a given ZTD, rather than representing the absolute entirety of the PWV error budget.

Figure 9: This figure does not give the whole picture. Even if you do not include additional errors in the figure, I think it is necessary to mention them. I am thinking of orbit and clock errors, as well as mapping function, and site dependent (the electromagnetic environment such as reflections/multipath) errors in the GNSS processing.

Response: The reviewer is absolutely correct that Figure 9 isolates only the uncertainties associated strictly with the conversion factor (Π) and does not represent the full, end-to-end error budget of the GNSS PWV retrieval. While the specific focus of this component-wise analysis was to contrast thermodynamic assumptions against physics constants, we completely agree that it is necessary to explicitly acknowledge the geodetic and site-dependent errors that affect the Zenith Total Delay (ZTD) estimation prior to this conversion step. We have revised Section 3.4.3 to explicitly mention orbit and clock errors, mapping function uncertainties, and site-dependent electromagnetic effects (such as multipath), providing the reader with a fully transparent overview of the total error budget.

L 475: Give a reference for the "known artifact".

Response: Provided the references revised manuscript

L 510 - 514: "This correction effectively halves the uncertainty in the final GNSS water vapor product compared to standard climatological approaches." is not true, and it is not "the dominant error source in GNSS meteorology". These statements ignore the errors in the ZTD from the GNSS processing. When

rewriting this text I suggest that you instead use words such as "may be comparable to", "are significant", or something similar. In this context I think you also have to point out that there are many sites world-wide where the use of microwave radiometry may not be motivated (given the high cost, all the other uncertainties, and that models for T_m is not that poor everywhere).

Response: We sincerely thank the reviewer for this crucial final check on our concluding statements. The reviewer is absolutely right that our original concluding phrasing inadvertently dismissed the substantial errors inherent in the ZTD estimation itself (orbits, clocks, mapping functions), thus overstating the relative impact of the T_m conversion step on the *total* error budget. We have revised the conclusion to adopt the reviewer's more precise terminology, stating that T_m errors are "significant" and "comparable to" ZTD uncertainties. Furthermore, we strongly agree with the reviewer's point regarding the global applicability of this method. We have added a dedicated caveat to the conclusion explicitly stating that given the high operational cost of MWRs, and the fact that empirical T_m models perform perfectly adequately in many regions without complex coastal thermodynamics, universal deployment of this synergistic architecture is not motivated. We thank the reviewer for helping us ground our conclusions with this important operational context.

Technical Corrections

L 27 + more: You use the American spelling of vapour, although ACP is a European journal?

Response: We have performed a comprehensive find-and-replace throughout the entire manuscript, including the title, to ensure the British spelling ("vapour") is used consistently.

L 30: Here you introduce PWV (unit mm). Later (L 118) you also use IWV (unit kg/m^2). You should use only one of these in order not to confuse the reader. I suggest to chose IWV because the delays from GNSS is expressed in unit of length and it makes it easier to see the difference.

Response: Our rationale is that in the meteorological, climatological, and severe-weather forecasting communities—which are the primary end-users of this GNSS data—PWV in millimeters is the universally recognized standard, as it intuitively corresponds to precipitation depth. Furthermore, the foundational literature in GNSS meteorology (e.g., Bevis et al., 1992) established the convention of mapping Zenith Wet Delay (in mm) directly to PWV (in mm) via the dimensionless conversion factor (II). However, we fully agree with the reviewer that introducing both IWV (kg m^{-2}) from the radiometer/radiosonde integrations and PWV (mm) from the GNSS could confuse readers unfamiliar with the terms. To resolve this, we have added an explicit clarifying statement in the methodology section confirming that the two terms describe the exact same physical quantity and are numerically equivalent ($1 \text{ kg m}^{-2} = 1 \text{ mm}$), allowing us to maintain meteorological conventions while completely eliminating any ambiguity.

L 34: Here the acronym EM is defined, but it is defined again on L 464. At many other places you do not use the acronym. The acronym is only used 9 times, so I suggest not to use it at all. I think it will make the reading better.

Response: We completely agree that low-frequency acronyms can disrupt the flow of the text, and the inconsistent use and double-definition of "EM" was an oversight on our part. We have adopted the reviewer's suggestion and removed the "EM" acronym entirely. We have performed a global replace throughout the manuscript to spell out "Eastern Mediterranean" in all instances.

L 60 + more: 2% --> 2 % (SI recommendation)

Response: Corrected in the revised manuscript.

Table 1: 1 sec --> 1 s (SI recommendation)

Response: Corrected in the revised manuscript.

L 114: Training Set --> training set

Response: Corrected in the revised manuscript.

L 115: Validation Set --> validation set

Response: Corrected in the revised manuscript.

L 132: delete "total column" (it is already in the definition of the IWV)

Response: Corrected in the revised manuscript.

L242: 6.7°C --> 6.7 °C

Response: Corrected in the revised manuscript.

L 284: whereas the MWR exhibits --> whereas it relative to the MWR exhibits

Response: Corrected in the revised manuscript.

L 560, 567: doi missing

Response: Corrected in the revised manuscript.

L 589: the doi link is not correct, shall be:

Response: Corrected in the revised manuscript.

<https://link.springer.com/book/10.1007/978-3-030-13901-8>, and the author list is not correct.

(Note that I have not checked all the doi links, so it may be a good idea for you to carry out?)

Response: Corrected throughout in the revised manuscript.

L 621: add [https://doi.org/10.1175/1520-0442\(1996\)009%3C3561:TWVCAT%3E2.0.CO;2](https://doi.org/10.1175/1520-0442(1996)009%3C3561:TWVCAT%3E2.0.CO;2)

Response: Corrected in the revised manuscript.

Reviewer#02 Comment:

Review of the Manuscript

“Enhancing GNSS Water Vapor Retrieval via Synergistic Microwave Radiometry: Thermodynamic Error Diagnosis and Bias Correction”

by Avinash N. Parde, Christina Oikonomou, and Haris Haralambous.

General Comments

GNSS is a well-established technique for observing precipitable water vapor (PWV). However, GNSS provides zenith total delay (ZTD) data, which must be combined with auxiliary data, specifically, zenith hydrostatic delay (ZHD) and weighted mean temperature (T_m), to derive PWV. Various methods and data sources have been used to estimate T_m . Empirical regression models (e.g., Bevis et al., 1992) relate T_m to surface temperature (T_s) and thus require T_s as auxiliary data. Static climatological models (e.g., GPT2w and HGPT2) are standalone and cannot capture short-term variations in T_m . Only numerical weather prediction (NWP) models can provide real-time T_m estimates, albeit with limited spatiotemporal resolution. This manuscript proposes to use ground-based MWR temperature- and moisture-profile measurements as an alternative data source for deriving T_m . The idea is original and the approach is worth exploring. However, the study from methodological and interpretative deficiencies that must be addressed before publication.

Response: We sincerely thank the reviewer for their time, their highly constructive feedback, and their excellent, concise summary of the current landscape of GNSS meteorology. We are especially grateful for the reviewer's acknowledgement that our core approach—utilizing ground-based Microwave Radiometer (MWR) profiles as an alternative data source to diagnose and mitigate T_m conversion errors—is original and a valuable contribution to the field. We also deeply appreciate the reviewer's rigorous identification of the methodological, statistical, and interpretative deficiencies present in the original draft. We have taken these critiques very seriously. Guided by the reviewer's excellent literature recommendations (e.g., Ning et al., 2016; Healy, 2011; Wang, 2005), we have performed a comprehensive revision of the manuscript to elevate its metrological rigor.

As detailed in our point-by-point responses below, the major improvements to the manuscript include:

- **Rigorous Error Propagation:** We have corrected previous mathematical misstatements, integrated formal partial derivatives for the uncertainty budget, and adjusted our framing to correctly identify T_m as a dominant *conversion-related* error, rather than the absolute largest GNSS error.
- **Metrological Clarity:** We have strictly differentiated between static systematic biases (e.g., refractivity constants) and dynamic compound errors, and corrected the visualization scaling in our uncertainty analysis.
- **Contextualizing Limitations:** We have explicitly clarified the diagnostic intent behind our scale height (H_v) analysis to expose the MWR's physical "smoothing errors", and we have better justified the use of the static HGPT2 model as our "blind" geodetic baseline.
- **Structural Reorganization & Terminology:** We have refined the manuscript's flow to ensure mathematical derivations are introduced prior to the results, and we have standardized all phrasing to adhere strictly to meteorological conventions (transitioning to IWV terminology and eliminating non-standard jargon).

We believe that the reviewer's expert, meticulous feedback has fundamentally strengthened the scientific validity of this paper. Please find our detailed, point-by-point responses and corresponding text revisions below.

Specific comments

1. Overemphasis on T_m as the Main Source of Uncertainty

The manuscript claims that T_m is the primary source of uncertainty in GNSS PWV data. While reducing T_m errors would improve PWV retrieval, it is well established that T_m is not the main error source (e.g., Ning et al., 2016). The exaggerated emphasis on T_m should be revised throughout the manuscript.

Response: The reviewer is completely correct; our original phrasing inadvertently and inaccurately elevated T_m to the primary source of total uncertainty, contradicting established geodetic literature which demonstrates that Zenith Total Delay (ZTD) estimation errors generally dominate the absolute error budget. We have conducted a comprehensive review of the entire manuscript to eliminate this overemphasis. We have systematically replaced terms such as "main," "primary," and "dominant error source" when referring to T_m . Instead, we now correctly frame T_m as a "significant, secondary source of uncertainty," specifically isolating it as the dominant error *within the thermodynamic conversion step* (II), while explicitly acknowledging that ZTD and ZHD uncertainties (driven by orbits, clocks, mapping functions, and pressure sensors) represent the bulk of the total end-to-end GNSS retrieval error. We believe these global revisions align the manuscript with the rigorous metrological standards established by Ning et al. (2016).

2. Error Propagation Analysis

The discussion of error propagation from ZTD to PWV is fragmented and inconsistent (Introduction, Sections 2.4, 3.3, 3.4, and 3.5). The quoted error values often do not align with standard error propagation rules and overestimate the impact of T_m errors on PWV. For

Example: A 1 K error in T_m would propagate to $\sim 0.37\%$ in PWV, or ~ 0.18 mm if $PWV = 50$ mm (an exceptionally high value for the Mediterranean region), not 0.3–0.5 mm (lines 54–60). A 1–2% error in T_m cannot translate to a 1–2 mm error in PWV. A comprehensive formulation using partial derivatives should be introduced, with derivatives quantified. This is essential for assessing the proportional contributions of each error source.

Response: We sincerely thank the reviewer for this meticulous mathematical check. The reviewer is absolutely correct; our original text inadvertently inflated the absolute impact of T_m errors by misapplying the "1% rule of thumb" and confusing relative percentage errors with absolute numerical errors. We deeply appreciate the reviewer catching this overestimation. We have thoroughly audited and corrected all quoted error values throughout the manuscript to align perfectly with standard propagation rules (e.g., explicitly stating that a 1 K error yields a $\sim 0.36\%$ relative error, or ~ 0.18 kg m^{-2} at a 50 kg m^{-2} IWV baseline). Furthermore, we completely agree that the error propagation discussion was too fragmented. To resolve this, we have consolidated the theoretical error framework into a single, dedicated subsection in the Methodology (Section 2.3). As requested, we have introduced the comprehensive formal error propagation equations utilizing partial derivatives. This provides a rigorous mathematical foundation to assess the exact proportional contributions of ZWD, T_m , and the refractivity constants to the final IWV uncertainty budget.

3. Systematic vs. Random Errors

The manuscript must clearly distinguish between systematic and random errors. For example: Errors in refractivity constants introduce a bias in derived PWV. Their associated uncertainty cannot be directly compared to T_m -related uncertainty, which likely combines both systematic and random components (Figure 9a). Figure 9b may contain a factor-of-10 error. The discussion also omits the uncertainty due to the refractivity constant k_1 in ZHD calculations.

Recommendation: Revisit the seminal works by Bevis (1992, 1994), Wang (2005), Healy (2011), and Bock (2021) on T_m modeling and refractivity constants.

Response: First, we must sincerely thank the reviewer for their sharp eye regarding Figure 9b; they correctly identified a factor-of-10 visualization error. Upon reviewing our plotting code, we discovered that while the total uncertainty budget (Figure 9a) correctly calculated the maximum bounding error between the three constant sets (~ 0.2 mm), the scatter plot in Figure 9b inadvertently plotted the difference between Rieger and Davis. Because the Davis and Rieger curves mathematically intersect near standard atmospheric temperatures ($T_m \approx 280$ K), their difference is artificially small (~ 0.02 mm). We have corrected Figure 9b to plot the true bounding difference (Rieger minus Bevis), which correctly reflects the ~ 0.2 mm magnitude shown in the boxplot. Furthermore, we completely agree that our previous framing blurred the critical statistical distinction between static systematic biases and dynamic uncertainties. We have revised Section 3.4 to explicitly clarify that refractivity constant variations (k_2, k_3) represent static, systematic biases, whereas T_m -related errors represent a compound uncertainty featuring both a systematic offset and a dynamic random component. We clarify that Figure 9 juxtaposes these strictly to illustrate their relative magnitude on the final error budget. We also thank the reviewer for pointing out the omission of k_1 . We have added a dedicated acknowledgement of k_1 's uncertainty into the Zenith Hydrostatic Delay (ZHD) calculation section. Finally, we have integrated the seminal works of Bevis (1992, 1994) and Healy (2011) to rigorously ground our discussion of refractivity constants and their statistical behaviors.

4. MWR Retrieval Bias

The initial MWR retrievals used in this study exhibit a large bias, suggesting inadequate neural network training. This issue is common with MWR profilers and is acknowledged later in the manuscript (lines 247 and 268), where bias correction is introduced. While bias correction improves accuracy, the conventional approach would be to refine the MWR retrieval algorithm through additional training with radiosonde data.

Response: We thank the reviewer for this perceptive comment and fully agree with the fundamental premise: refining the MWR's native Neural Network (NN) retrieval algorithm using local radiosonde data is indeed the conventional and ideal long-term solution to mitigate these biases. However, we opted for a post-retrieval bias correction due to strict data-volume limitations. Robust NN training for passive radiometry requires a comprehensive, multi-year historical archive of high-resolution local radiosonde profiles to adequately represent the full climatological envelope (capturing inter-annual variability, distinct seasonal transitions, and extreme anomalies). Because the CYGMEN infrastructure at the Athalassa observatory was recently established, our current dataset (covering only the 2025 warm season) is statistically insufficient to train a custom NN from scratch without incurring a severe risk of overfitting to this specific seasonal window. Consequently, relying on the manufacturer's standard NN (trained on broader regional historical priors) and applying a site-specific linear bias correction was the most

scientifically sound and operationally feasible approach available to immediately improve the GNSS-PWV conversions. We have added text to the manuscript to explicitly clarify this data-volume constraint, explaining why full algorithmic retraining is deferred until a multi-year radiosonde climatology is established.

5. Water Scale Height (Hv) Analysis

The analysis of the water scale height (Hv) is imprecise and irrelevant to this study for several reasons: Hv is a crude representation of the moisture profile. The discussion on upper cutoff altitude highlights the ambiguity in estimating Hv from MWR data, as reflected in the large scatter in MWR results (Figure 5). The uncertainty in estimated Hv values (quantifiable via least-squares regression) is not considered when comparing MWR and radiosonde (RS) estimates.

Response: We appreciate the reviewer's critical evaluation of the Hv analysis. We actually completely agree with the reviewer's physical assessment: Hv is indeed a mathematically crude representation of the true moisture profile, and the MWR estimates of it are highly ambiguous and characterized by severe scatter. However, we respectfully disagree that this makes the analysis irrelevant to the study. The primary objective of Section 3.2 was not to advocate for Hv as a high-quality retrieval product, but rather to explicitly use it as a diagnostic tool to expose the severe vertical "smoothing errors" inherent to passive microwave radiometry. By demonstrating the MWR's total failure to capture Hv (yielding an artificial, upwardly stretched distribution), we sought to provide a transparent, balanced assessment of the instrument: proving that while the MWR is highly accurate for total column mass (IWV) and thermal profiling (Tm), it completely breaks down when trying to resolve vertical moisture structure. We apologize if this diagnostic intent was not clear in the original draft. We have revised Section 3.2 to explicitly clarify that Hv is used solely as a demonstrative tool to expose the MWR's structural limitations. Furthermore, we concede the reviewer's excellent statistical point regarding the individual least-squares fit uncertainties. We have added commentary noting that the structural failure (smoothing error) completely dwarfs the statistical fit uncertainty, rendering the MWR's Hv mathematically invalid.

6. Comparison with HGPT2

The manuscript argues that MWR-derived Tm is superior to HGPT2, which is unsurprising. HGPT2 is a static climatological model and cannot compete with real-time data sources like MWR (Figures 6 and 9). Due to this intrinsic limitation, static models are less relevant for meteorological applications compared to empirical and NWP models. Additionally, the Tm variable in HGPT2 is derived via regression on Ts, which is known to introduce spurious diurnal cycles into Tm (Wang, 2005; Bock, 2021).

Recommendation: Remove HGPT2 from the study and instead compare MWR-based Tm to operational NWP-derived Tm estimates (e.g., ERA5 with 1-hourly resolution).

Response: We thank the reviewer for this insightful comment and for providing the excellent references regarding the spurious diurnal cycles introduced by Ts regressions. We enthusiastically agree with the reviewer's physical assessment: the reliance on Ts perfectly explains the severe midday bias amplification we diagnosed in Section 3.4.1. We have added the suggested references (Wang, 2005; Bock, 2021) to this section to strengthen our physical explanation of the model's structural failure. Regarding the recommendation to remove HGPT2: while we completely agree that dynamic NWP models (like hourly ERA5) are vastly superior for meteorological applications, our study specifically addresses the geodetic

community where "blind" empirical models (like GPT2w/HGPT2) remain the operational standard for users who lack real-time access to dense NWP fields or in-situ sensors. In this context, HGPT2 is employed as our representative baseline. Our objective was to quantify the exact penalty (in millimeters of IWV) of relying on these static models during severe thermodynamic events in the Eastern Mediterranean, and to demonstrate how local MWR synergy easily mitigates this. A direct comparison with real-time hourly NWP products (e.g., ERA5) is a highly relevant next step for evaluating data assimilation frameworks, but it falls beyond the scope of the current hardware-centric study. We have updated the manuscript (in both the Methodology and Discussion) to better clarify why HGPT2 was retained as the "blind geodetic baseline" and to explicitly acknowledge that NWP comparisons are the logical subject of future work.

7. Terminology and Clarity

The manuscript contains awkward phrasing and non-standard terminology. Examples include:

“Critical error culmination”

“Thermodynamic severity”

“Noon ballooning”

“Thermodynamic path delays”

“Thermodynamic errors”

Concepts such as uncertainty, error, and bias are used inconsistently and confusingly throughout the manuscript.

Response: We sincerely thank the reviewer for highlighting these terminological inconsistencies. We completely agree that the use of non-standard, overly descriptive phrasing detracted from the scientific rigor of the manuscript. We have performed a comprehensive review of the text to align all terminology with standard atmospheric and metrological conventions.

Specifically, we have made the following global replacements:

- "Critical error culmination" has been replaced with standard threshold terminology (e.g., "Systematic Bias Threshold").
- "Thermodynamic severity" has been replaced with precise physical descriptors, such as "moisture abundance".
- "Noon ballooning" has been replaced with "diurnal bias peak".
- "Thermodynamic path delays" and "Thermodynamic errors" have been corrected to specify exactly what is being measured (e.g., "Tm parameterization errors" or "thermodynamic conversion uncertainty").

Furthermore, we have rigorously audited our use of the terms *error*, *bias*, and *uncertainty* throughout the manuscript to ensure they strictly adhere to standard metrological definitions (where *bias* denotes systematic offset, *uncertainty* denotes statistical dispersion/RMSE, and *error* refers generally to the residual against the reference). We believe these changes have significantly improved the clarity and professionalism of the paper.

8. Scientific Rigor in Sections 3.5, 4, and 5

Section 3.5: Focuses excessively on the deficiencies of HGPT2 and static models, which are well known. These issues could be avoided by excluding HGPT2 from the analysis. Section 4: The discussion on planetary boundary layer (PBL) and free-troposphere “decoupling” lacks evidence. Radiosonde (RS) data could help document this climatic feature, but it is unclear how MWR profiles would capture it effectively. Section 5 (Conclusions): Requires revision based on the above comments.

Response: We appreciate the reviewer's rigorous assessment of our discussion and concluding sections.

Regarding Section 3.5: As noted in our earlier response, we respectfully elect to retain HGPT2 because "blind" empirical models remain the operational standard for GNSS geodetic processing where real-time NWP data is unavailable. However, we agree with the reviewer that the fundamental deficiencies of static models are already well known in the atmospheric community. We have streamlined Section 3.5 to remove overly generalized critiques of static models, focusing strictly on quantifying the specific penalty ($>1.0 \text{ kg m}^{-2}$ IWV error) they incur during severe events in the Eastern Mediterranean.

Regarding Section 4: The reviewer raises a highly astute physical point. We completely agree that the MWR, due to the broad weighting functions of its K-band channels, cannot effectively capture the sharp inversion layer indicating PBL decoupling (which is exactly why the MWR exhibits the "smoothing errors" documented in our Hv analysis). Following the reviewer's excellent suggestion, we have updated Section 4 to explicitly state that the high-resolution Radiosonde (RS) profiles serve as the physical evidence documenting this decoupling, and we clarify that the MWR's inability to natively resolve this boundary is precisely what drives the need for our site-specific bias corrections.

Regarding Section 5: The Conclusions have been comprehensively revised to align with these nuances, specifically dialing back the HGPT2 critique and emphasizing the physical limitations and necessary corrections of the MWR hardware.

9. Manuscript Organization

The manuscript's structure is confusing. For example: The capacity of MWR to retrieve PWV directly is not mentioned until Section 3.2. The derivation of GNSS PWV in Section 3.2 is unclear. T_m estimates are evaluated in Section 3.3, after the first GNSS PWV estimates are assessed (Section 3.2). MWR calibration is introduced in Section 3.4, after MWR T_m data are used.

Response: We thank the reviewer for highlighting these areas of structural confusion. We completely agree that the methodological foundations—specifically the MWR's native retrieval capacity and the mathematical derivation of GNSS IWV—should be established prior to the results. We have moved these descriptions to Section 2 (Methodology), significantly expanding the GNSS derivation equations to ensure clarity. Regarding the sequence of Section 3 (Results), we respectfully elect to maintain the current order (assessing IWV in 3.2, evaluating T_m in 3.3, and introducing the calibration in 3.4) as it follows a deliberate, "top-down diagnostic" narrative. Our rationale is to first establish the macroscopic baseline performance of the final operational products (Section 3.2). Once the discrepancies in the end-products are presented, we subsequently isolate and diagnose the underlying intermediate physical variable causing them (T_m , Section 3.3). Finally, having demonstrated the specific systematic biases in the native data, we introduce the site-specific MWR calibration (Section 3.4) as the targeted solution. We recognize, however, that without proper signposting, this diagnostic flow can be confusing. To resolve this without disrupting the manuscript's core structure, we have added explicit transition sentences at the beginning of Section 3 and within Section 3.2. These forward-references clearly map out this top-down approach,

guiding the reader through the diagnostic process. We believe these strategic text additions fully resolve the clarity issues while preserving the narrative arc.

Recommendations for Revision

Before publication, I recommend the following:

1. Improve MWR retrievals through complementary neural network training.

Response: The objective of this paper was to diagnose the impact of thermodynamic errors (T_m) on GNSS-PWV retrievals and to demonstrate a computationally efficient, easily implementable mitigation strategy. As our results show, applying a simple, site-specific linear regression to the final T_m product successfully reduced the RMSE from 2.32 K to 1.43 K, aligning it with the theoretical accuracy limits of ground-based radiometry. Because custom NN training requires a multi-year climatological dataset that is still being built under the new CYGMEN infrastructure, our linear approach provides an immediate, highly effective solution for operational GNSS networks. Nevertheless, we recognize the scientific value of the reviewer's suggestion. We have added a statement in the Discussion section acknowledging that customized NN training represents the logical next step for future work as the local radiosonde archive expands.

2. Replace HGPT2 with a high-resolution NWP model (e.g., ERA5) as an alternative T_m source. Compare MWR-derived T_m to both NWP and RS T_m .

Response: We appreciate the suggestion and agree that high-resolution NWP models (like hourly ERA5) provide superior dynamic T_m estimates. However, we respectfully elect to retain HGPT2 as our baseline. Our study is specifically designed to evaluate the "blind" empirical climatologies that remain the operational standard for GNSS geodetic users who lack real-time access to NWP data streams. Our objective is to quantify the specific penalty of relying on these static models during severe local events and to demonstrate how local MWR hardware mitigates this. While benchmarking MWR retrievals against hourly NWP fields is a highly relevant next step, it represents a distinct data assimilation exercise that falls outside the intended scope of this current hardware-focused validation study.

3. Assess GNSS PWV against MWR and RS PWV.

Response: Based on the manuscript's validation analysis in Section 3.2, here is the assessment of the water vapor retrievals across the three observational methods. (*Note: The manuscript conducts this specific statistical intercomparison using Integrated Water Vapor (IWV) in kg/m^2 , which serves as the equivalent mass column metric to PWV.*)

Comparison	Correlation (r)	RMSE (kg/m^2)	Mean Bias (kg/m^2)
MWR vs. RS	0.971 +1	1.72 +1	+0.45 +1
GNSS vs. RS	0.955	1.98 +1	-0.45 +1
MWR vs. GNSS	0.964	2.05	+0.89 +1

4. Remove the H_v discussion, as it is not relevant to the study.

Response: We appreciate the reviewer's perspective, but we respectfully elect to retain the H_v discussion. Our intention is not to present MWR-derived H_v as a reliable meteorological product, but rather to use its deliberate failure as a transparent diagnostic tool. The large scatter and ambiguity shown in the analysis explicitly demonstrate the "smoothing error" inherent to passive microwave radiometry, proving to the

reader that the instrument cannot resolve vertical moisture compactness. Rather than removing the section, we have added a brief clarification to Section 3.2 to ensure this diagnostic intent is unmistakable.

5. Conduct a proper error propagation analysis (Ning, 2016) to contextualize T_m errors relative to other sources in the ZTD-to-PWV processing chain.

Response: Revised the manuscript accordingly.

6. Adhere to standard scientific terminology and ensure consistent use of terms like uncertainty, error, and bias.

Response: Adhered to standard scientific terminology in the revised manuscript.

References

Wang 2005: doi:10.1029/2005JD006215

Healy 2011: doi:10.1029/2010JD014013

Ning 2016: doi:10.5194/amt-9-79-2016

Bock 2021: <https://doi.org/10.5194/essd-13-2407-2021>

Response: Added these references in the text as well as in the reference list.

Final Assessment

I do not recommend publishing this manuscript in its current form. The authors should address the methodological, organizational, and terminological issues outlined above. A revised version, incorporating these suggestions, would significantly improve the manuscript's scientific rigor and clarity.