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 (Tm), to derive PWV. Various methods and data sources have been used to estimate Tm. Empirical regression models (e.g., Bevis et al., 1992) relate Tm to surface temperature (Ts) and thus require Ts as auxiliary data. Static climatological models (e.g., GPT2w and HGPT2) are standalone and cannot capture short-term variations in Tm. Only numerical weather prediction (NWP) models can provide real-time Tm 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 Tm. 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 Tm 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 Tm 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 (Hv) 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 Tm as the Main Source of Uncertainty

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

Response: The reviewer is completely correct; our original phrasing inadvertently and inaccurately elevated Tm 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 Tm. Instead, we now correctly frame Tm as a "significant, secondary source of uncertainty," specifically isolating it as the dominant error *within the thermodynamic conversion step* (Π), 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 Tm errors on PWV. For

Example: A 1 K error in Tm would propagate to ~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 Tm 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 Tm 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 ~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, Tm, 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 Tm-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 k1 in ZHD calculations.

Recommendation: Revisit the seminal works by Bevis (1992, 1994), Wang (2005), Healy (2011), and Bock (2021) on Tm 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 Rüeger and Davis. Because the Davis and Rüeger curves mathematically intersect near standard atmospheric temperatures (Tm≈280 K), their difference is artificially small (~0.02 mm). We have corrected Figure 9b to plot the true bounding difference (Rüeger 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 (k2′,k3) represent static, systematic biases, whereas Tm-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 k1. We have added a dedicated acknowledgement of k1'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 cutoƯ 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 eU̔ectively. 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 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. Tm 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 Tm 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 Tm 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 (Tm, 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 (Tm) 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 Tm 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 Tm source. Compare MWR-derived Tm to both NWP and RS Tm.

Response: We appreciate the suggestion and agree that high-resolution NWP models (like hourly ERA5) provide superior dynamic Tm 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/m², which serves as the equivalent mass column metric to PWV).*

Comparison	Correlation (r)	RMSE (kg/m²)	Mean Bias (kg/m²)
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 Hv discussion, as it is not relevant to the study.

Response: We appreciate the reviewer's perspective, but we respectfully elect to retain the Hv discussion. Our intention is not to present MWR-derived Hv 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 Tm 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.