Review of "Spectral analysis of GNSS zenith wet delay and microwave radiometer precipitable water during FESSTVaL Lindenberg 2021: a multi-scale observational test of horizontal heterogeneity sensitivity" by Gaël Kermarrec and co-authors.

I have read the RC1 and the authors' response AC1 alongside the original manuscript. The response reports useful additions, particularly the GNSS constraint-sensitivity experiments, the comparison with an independently produced solution, and clarification that the analysed quantity is the full wet delay. These address important concerns raised by RC1. However, the revised manuscript and the new Appendix A are not available to me, so I cannot assess their implementation or the full supporting results. Some requests below may already be covered there.

The study's strongest contribution is the reported association between GNSS spectral parameters and independent measures of variability in column-integrated water vapour. This is a worthwhile observational result. My remaining concerns mainly involve the reliability of those comparisons and the strength of the physical interpretation. I recommend major revision, but I do not consider new field observations or a comprehensive physical retrieval model necessary for publication if the claims are appropriately limited.

Major revision comments

1. Establish the quality of the radiometric reference data.

RC1 raised the reliability of MWR measurements during rain. AC1, G2 and S5 clarify that precipitation is not used to exclude GNSS observations, but this does not resolve how contaminated radiometric observations are handled. The validation depends on RAOHAT, FOGHAT and the intersite MWR differences, all of which require their own quality control.

Please document the flags and screening applied to each radiometer, including wet-radome recovery and treatment of gaps. The main heterogeneity correlations should be reported for a common set of valid radiometric windows, with retained sample counts.

Please explain the largest heterogeneity values in Figure 8 and demonstrate that they are retained after the radiometric quality-control procedures described above.

The GNSS record can remain continuous through precipitation. The conditions under which its atmospheric interpretation has been validated should nevertheless be stated separately.

2. Show that the spectral results are not primarily imposed by preprocessing.

The PPP sensitivity tests reported in AC1 are useful, but they retain the same detrending and spectral-fitting procedure. They therefore leave open the influence of the 25 min Savitzky–Golay filter and the frequency band used for fitting.

A focused check would be sufficient: specify the effective analysis band for both instruments, explain how filtering and sampling enter the fitted spectrum, show representative spectral fits, and test a small number of reasonable alternative detrending settings.

Please also report the occurrence and treatment of poorly identified or boundary-valued fits, given the apparent accumulation of large λ values in Figure 5. AC1, T6 addresses the figure's rendering, but not this numerical feature.

I am not requesting an exhaustive exploration of every processing choice. The purpose is to establish that the main associations survive reasonable choices and that the fitted scales are identifiable in the available band.

3. Carry the GNSS sensitivity tests through to the main scientific comparisons.

AC1, G1 reports high rank correlations between processing variants and a useful independent-solution comparison. This is encouraging. However, agreement between two estimates of λ does not demonstrate that all external correlations and regime contrasts are unchanged.

Please provide a compact table of the principal results for the existing processing variants, evaluated on identical matched windows. The key quantities are the heterogeneity correlations and the regime contrasts used to support the conclusions. This should use the reprocessing already performed rather than require another set of solutions.

Please also make the GNSS–MWR comparison symmetrical by reporting the MWR amplitude correlations alongside the GNSS amplitude correlations, as well as the corresponding scale correlations. Otherwise, the comparison between σ_GNSS and λ_MWR mixes differences between instruments with differences between statistical parameters.

Please verify the process-noise units and state the implemented time convention. If 0.0300 m $s^{-1/2}$ denotes the conventional random-walk amplitude coefficient q, the process-model increment standard deviation is $q\,\Delta t^{1/2}$, giving approximately 0.164 m over 30 s. This is a prior process-model increment uncertainty, not the posterior uncertainty or observed variability of the estimated ZWD.

4. Provide enough statistical detail to evaluate the reported confidence.

The manuscript already mentions block bootstrap and circular-shift tests, which are appropriate approaches. Please specify the block lengths and their selection, the permitted shifts, handling of gaps, sample counts and representation of the diurnal cycle in partial correlations. A check with non-overlapping windows would help establish robustness, given the approximately 92% overlap between adjacent windows.

These details are particularly important for the uncertainty of regime differences and external correlations. There is no need to repeat every analysis using several statistical methods, but the reported significance should be reproducible and account for serial dependence.

Squared Spearman correlations should not be described directly as fractions of physical variance explained.

5. Correct interpretations that are not supported by the present construction.

Several statements can be resolved primarily through clearer definitions and more restrained wording.

The quadratic difference used to define σ_extra is a constructed variance contrast. It does not isolate an additional atmospheric component without assumptions about covariance, observation kernels, noise and spectral bandwidth. Please describe it accordingly and provide a limited sensitivity check for the conversion factor and the treatment of negative variance differences.

AC1, S4 correctly notes that multiplying an amplitude by a positive constant preserves its ranks. This does not generally hold for σ_extra, because the conversion coefficient enters a subtraction and determines which values are floored to zero.

Similarly, the reported 7% identifies windows where the fitted MWR amplitude exceeds the fitted GNSS amplitude; it is not an independent estimate of the relative instrumental noise.

Please reconcile the definition, units and numerical value of the conversion coefficient in Eq. (1). If Π is defined by PRW=Π ZWD, the conversion from PRW to ZWD uses 1/Π. State explicitly which coefficient is applied and how the water-vapour-weighted mean temperature Tm is obtained.

State whether the PRW-to-ZWD conversion uses a constant or time-dependent coefficient. If a constant is used, justify its representativeness and assess whether plausible Tm variability materially affects the reported results, particularly σ_extra.

The LIN0 regression should be described as an association with GNSS interstation or processing-related variability, rather than a quantitative attribution of roughly one third of the variance to constellation geometry. AC1, S3 improves the description of the stations, but does not establish that their difference isolates geometry. The sidereal test likewise supports only the statement that no enhanced recurrence was detected by that test, but does not exclude multipath.

The manuscript would benefit from a clearer explanation of why a persistent coherence scale is meteorologically relevant and what information it provides beyond existing BLH diagnostics. In particular, the authors should clarify whether the persistence reflects a genuine atmospheric structure (e.g. residual-layer or entrainment-zone moisture organization) or partly the vertically integrated GNSS observation geometry.

Acceptable alternatives to further physical validation

Several stronger interpretations can be retained as hypotheses without requiring substantial new analyses for this paper.

The FOGHAT and intersite MWR comparisons concern vertically integrated water vapour. They support sensitivity to horizontal column variability, but do not uniquely locate it at the boundary-layer top. The correlation with retrieved humidity variability near 2 km is suggestive, especially given the changing BLH, rather than a definitive localization. The authors can either strengthen that attribution using existing suitable data or qualify it consistently in the abstract, discussion and conclusions.

The same choice applies to the surface-layer result. Near-zero contemporaneous correlations do not exclude delayed transport, footprint differences or surface contributions subsequently observed aloft. Lagged and wind-conditioned analyses would be useful, but are not essential if the conclusion is restricted to the absence of a strong zero-lag association with the tested proxies.

Finally, λ is an inverse temporal scale. Without a validated temporal-to-spatial relationship, I suggest describing it as a temporal spectral parameter associated with horizontal water-vapour variability.

Nocturnal persistence can be discussed as potentially consistent with residual-layer or other elevated moisture organization, while acknowledging that the mechanism is unresolved. Its lower coefficient of variation than BLH demonstrates different temporal behaviour, not superior reliability or established meteorological value.

These changes would preserve the interesting observational result while bringing the interpretation into line with the evidence.

Additional technical comments

Clarify that the ratio of 3.7 in Section 5.1 refers to the fitted Matérn high-frequency amplitude parameters after detrending, not the standard deviations of the raw observations.

Clarify that the PRW to ZWD factor is not a simple constant because Tm is not a constant.

Clarify what makes the quality of the LIN0 data inferior to LDB2.

I suggest adopting technique-based notation throughout the manuscript: ZWD_GNSS for the GNSS-derived estimates and ZWD_MWR for the MWR-derived ZWD equivalents, with corresponding spectral parameters (λ_GNSS,σ_GNSS) and (λ_MWR,σ_MWR). The current subscripts ZWD and PRW mix the distinction between physical quantities with the distinction between measurement techniques. PRW should remain the designation for the original MWR water-vapour retrieval before conversion.

Revision requirements

My major-revision recommendation therefore rests on a bounded set of requirements:

- establish the quality of the radiometric comparisons,
- demonstrate robustness to the main processing choices using the available data and existing reprocessing,
- document the statistical inference,
- correct unsupported attribution.