Author’s response

We again thank the editor and reviewers for their constructive feedback on our manuscript. In this revision, we have focused on strengthening the physical and geostatistical rationale of our merging framework and refining technical descriptions. The principal updates include:

- **Physical Rationale for CML processing:** Expanded Section 1 to detail CML signal attenuation physics and processing.
- **Physical Rationale for the different merging methods:** Expanded Section 1 and Section 3 to describe why the different merging methods was selected.
- **Variogram Standardization and Sensitivity:** Clarified the variance-standardization procedure (c = 1) across all merging methods. Better justified the episode-level variogram pooling and the fixed variogram strategy.
- **Technical Refinements:** Adopted a mathematically robust log-transform back-transformation formula for multiplicative merging to handle low rain rates seamlessly, updated all figures with units where missing and explicit left-to-right bar descriptions.

Please find our point-by-point response below

Response to reviewer 3

This manuscript investigates the use of commercial microwave link (CML) data to improve weather radar rainfall estimates through data fusion. While the underlying idea is scientifically interesting and the study addresses an important topic, the manuscript still requires several major revisions before it can be considered for publication.

Major Comments
i) The manuscript should better explain the general context of Commercial Microwave Links (CMLs) and their relationship to rainfall estimation. While the methodology is described, the rationale for using CMLs is not sufficiently introduced. A clearer explanation of how rainfall-induced signal attenuation is exploited for precipitation retrieval, together with the advantages and limitations of CML-derived rainfall estimates, would help readers understand the motivation for this study.

We thank the reviewer for raising this point. Specific microwave link attenuation (k-R) is substantially less sensitive to variations in drop size distribution (DSD) than radar reflectivity (Z-R), particularly in the 15–35 GHz frequency range. This physical property occurs because the raindrop extinction cross-section at these frequencies scales similarly to drop terminal velocity, causing DSD variability to largely cancel out. We have updated the manuscript section 1. Introduction (lines 30 - 48) and 2. Dataset (lines 105 - 120) to clarify this point.

ii) The research questions require further clarification. The first research question ("How much can CML rainfall retrievals improve radar QPEs, and how do adjustment methods and network configurations influence this improvement?") is somewhat ambiguous. Is the

objective to improve radar rainfall estimates themselves, or to improve the accuracy of rainfall estimates with respect to ground truth (actual precipitation)?
The term adjustment methods should be clearly defined. Which quantities are adjusted? How many adjustment methods are compared , and why were these specific methods selected? A stronger justification for their selection would improve the scientific motivation.

We thank the reviewer for this constructive feedback. We have refined the wording of the research questions to make it explicitly clear that ground-level evaluation is conducted against independent rain gauge observations compared to unadjusted radar estimates (lines 70 - 75).

Furthermore, we have rewritten the introduction to the research questions (lines 61 - 69) as well as methods section to clearly define the merging paradigms (additive/multiplicative: lines 141-148, KED: lines 179-186), and the physical hypotheses motivating their selection.

Finally, we have changed the discussion in Section 5.2 to lean more on the physical and geostatistical rationale for these methods (lines 426-465).

iii) The assumptions underlying the KED implementation need to be discussed in greater detail. What type of relationship between the dependent variable and explanatory variable is assumed (e.g., linear or nonlinear)? More importantly, is the drift statistically tested before being incorporated into the interpolation? Several studies first evaluate the statistical significance of the relationship between the independent and dependent variables at each time step, and only include the drift when this relationship is statistically significant. Assuming a drift where no statistically significant relationship exists may bias the interpolation results.
The authors should discuss whether such statistical testing was considered and how the adopted approach may affect the results.

Garen, D. C. (1995). Estimation of spatially distributed values of daily precipitation in mountainous areas. In Mountain Hydrology: Peaks and Valleys in Research and Applications, pp. 237–242.
Garen, D. C., Johnson, G. L., & Hanson, C. L. (1994). Mean Areal Precipitation for Daily Hydrologic Modeling in Mountainous Regions. Journal of the American Water Resources Association, 30, 481–491.

We thank the reviewer for raising this important point. Our KED implementation assumes a linear relationship between CML observations and radar rain rates, and the external drift was applied unconditionally without prior statistical significance testing.

We agree that enforcing an unconditional drift when radar explanatory power is weak can introduce spatial bias or excessive smoothing. We have updated Section 5.2 to explicitly discuss these assumptions.

iv) The variogram modelling strategy requires stronger justification. Why is the variogram range fixed based on values reported in the literature instead of being estimated automatically together with the other semivariogram parameters?

We thank the reviewer for raising this important point regarding our variogram modeling strategy.

We evaluated both a fixed reference variogram (range h_r = 30 km) and dynamic data-estimated variograms specifically to evaluate the sensitivity of the merging methods to variogram parameter selection.

A central finding of our work, presented in Section 4.3 (Figure 4) and discussed in Section 5.3, is that the performance of the merging methods is remarkably robust and insensitive to whether the variogram range is dynamically estimated or set to a reasonable constant literature value, provided a non-zero nugget is used.

To make this rationale clearer to readers, we have rewritten Section 3.5 to better explain the different variogram strategies used and explicitly clarify that fixing the range serves as both a practical benchmark and a robust strategy backed by our sensitivity analysis in Sections 4.3 and 5.3.

Similarly, why is a spherical variogram selected a priori? Numerous variogram models are available in the literature (e.g., exponential, Gaussian, Matérn), and there is no clear justification for assuming that the spherical model is the most appropriate for this novel application. A more robust approach would be to test multiple variogram models, estimate their parameters automatically, and retain the model that provides the best fit to the experimental variogram.Please consider acknowledging and, if possible, quantifying the influence of this assumption on the results.

We thank the reviewer for highlighting the potential for evaluating alternative variogram models. We agree that formally comparing different model shapes (e.g., exponential vs. Gaussian) represents a comprehensive geostatistical approach.

In our study, we selected the spherical model a priori as it is standard practice in radar merging literature (e.g., Haberlandt, 2007; van de Beek et al., 2012). While we did not test alternative variogram functional forms, we conducted extensive sensitivity testing across different variogram parameters (ranges and nuggets). These tests revealed that merging performance is remarkably stable across different parameter estimation strategies. We attribute this low sensitivity primarily to two factors: (1) the spatial weighting structure of the merging algorithms, and (2) the inherent noise level of CML data, which necessitates a non-zero nugget that effectively dampens subtle structural differences between variogram shapes.

To ensure complete transparency regarding this scope, we have updated Section 5.3 to better reflect this discussion.

v) The manuscript states that it adopts the event-averaging approach proposed by Haberlandt (2007). However, the same study also investigated an automatic time-varying semivariogram fitting procedure, where a specific variogram is estimated at each time step. This approach may improve interpolation accuracy (e.g., lower RMSE in cross-validation of hourly rainfall). Why was this procedure not evaluated in the present study?

While Haberlandt (2007) evaluated dynamic hourly fitting, that study demonstrated that timestep-specific fitting yielded only negligible performance gains over pooled/averaged variograms. Given these minimal gains and the computational overhead of fitting dynamic variograms across multi-month continuous datasets, a window-pooled approach was adopted. We updated Section 3.5 to better reflect this decision (lines 208 - 209).

Furthermore, our sensitivity analysis (Fig. 4) confirms that overall merging performance remains largely invariant across a broad spectrum of variogram parameters. This indicates that timestep-by-timestep fitting plays a secondary role relative to the algorithm's spatial weighting structure and the inclusion of a non-zero nugget. However, we have expanded Section 5.3 to explicitly note that dynamic temporal fitting could still provide localized benefits (lines 504 - 508).

Minor Comments

Lines 96: The manuscript states that the TL data underwent quality control checks. Please specify: Which quality control procedures were applied? Were these procedures described elsewhere in the manuscript? If so, please refer the reader to the appropriate section.

Thank you for the comment. We restructured this section and added an explanation to each of the six-step overview list. Hence the quality control procedures (with channel-difference filtering, diurnal cycle removal, signal noise filtering, and flatline removal) were described immediately after they are mentioned (lines 106 - 111).

Line 98: The statement that rainfall-induced attenuation is converted into rainfall rates following Chwala and Kunstmann (2019) deserves additional explanation. Since this conversion is a fundamental step of the methodology, please briefly describe the retrieval procedure together with its main assumptions, advantages, and limitations.

Thank you for this suggestion. While the processing workflow was described in detail, we agree that explicitly stating the underlying physical conversion, its key assumptions, advantages, and limitations provides useful context for the reader.

We have expanded the description in Section 2, directly after the new overview list mentioned in the previous comment (lines 105 - 129).

Lines 100-108: To what extent can the adopted parameter values be assumed constant in space or transferable from previous studies? Please discuss whether these parameters are site-specific or generally applicable, and how sensitive the results may be to their values.

Thank you for raising this point regarding parameter transferability and sensitivity. We have clarified the transferability and sensitivity of the processing steps in the paragraph after the new overview list (lines 121 - 129).

Lines 114-115: Figure B1 is introduced without sufficient context. At this stage of the manuscript, neither the interpolation methods nor the underlying methodology have yet been presented, making the figure difficult to interpret. Please: introduce the figure more clearly in

the text; explain what is shown; define the quantities represented on the x- and y-axes; spell out all acronyms appearing in the labels; include a legend if needed; add units to the y-axis.

We thank the reviewer for these constructive suggestions regarding Figure B1. We have updated the figure to explicitly label the y-axis with units (RMSE [mm/h]) and directed readers to Table 1 for acronym definitions. To keep the Methods section streamlined, we removed references to this figure from Section 3, introducing the sensitivity analysis directly in Appendix B and referencing it in Section 5 (lines 421-422).

Finally, we respectfully prefer to maintain the layout without an embedded legend to preserve visual clarity across subplots. Adding a legend box would compress the panel height and risk obscuring data features near upper y-axis limits. Instead, the caption explicitly specifies that the three bars within each method group strictly follow a left-to-right order corresponding to neighborhood sizes of 12, 70, and 500 (all CMLs).

Lines 270-274: The discussion comparing IDW and kriging mainly reiterates well-established concepts from the literature. Either expand this section by discussing the specific findings obtained in this study, or consider removing these generic statements.

We thank the reviewer for this suggestion. We agree that the theoretical explanation comparing distance weighting and de-clustering was unnecessarily generic. We have simplified this paragraph to remove these textbook statements (Section 5.4).

Figure 2: In addition to the current results, it would be useful to present the relative differences with respect to the baseline scenarios (original radar estimates and rain gauges). This would make it easier to evaluate whether the different interpolation methods perform differently across rainfall magnitudes. Additionally, please spell out all acronyms in the figure caption.
We thank the reviewer for this suggestion. We understand the value of evaluating whether the interpolation methods perform differently across varying rainfall magnitudes.

To address the request for spelled-out acronyms, we have updated the caption of Figure 2 to refer to Table 1 where the method acronyms are explained.

Regarding the spatial maps, we have retained the absolute difference fields (mm/h) in Figure 2 rather than introducing relative (percentage) differences. Calculating relative spatial differences for 2D rainfall fields introduces extreme scaling artifacts and division-by-zero issues in areas with low or zero precipitation, which obscures the spatial patterns. However, to directly address the reviewer's underlying goal of evaluating performance across rainfall magnitudes, this exact analysis is systematically presented in Figure 6, where performance metrics (RMSE and BIAS) are broken down across specific rainfall intensity classes (light, moderate, heavy, and extreme).

Figure 4: Please include a legend explaining the meaning of the different hatch patterns used in the bar chart.

We thank the reviewer for this suggestion. However, we respectfully prefer to maintain the current layout without an embedded legend to preserve visual clarity.

A 5-item legend for the hatch patterns would require considerable space. Placing it inside the subplots risks obscuring critical data features, while placing it outside would significantly compress the y-axes. To ensure unambiguous interpretation without cluttering the plot area, the figure caption explicitly specifies that the five sub-bars within each method cluster strictly follow a left-to-right order, corresponding directly to strategies (1) through (5).

Line 293: The nugget values appear to be fixed throughout the analysis. Please justify this choice. Are these values derived from physical considerations, previous studies, or preliminary analyses ? Why is only the nugget fixed while the remaining semivariogram parameters are estimated? How might this assumption affect the interpolation results? A clearer explanation of this modelling choice would strengthen the methodological section.

We thank the reviewer for this comment and the opportunity to clarify our variogram modeling choices.

The dimensionless nugget ($c_0 = 0.3$) was derived directly from our empirical variogram fits across the event dataset. In our primary reference strategy, all parameters ($h_r = 30$km, $c = 1$, $c_0 = 0.3$) are held fixed, because our sensitivity analysis proved dynamic fitting yields no performance gain (Section 4.3). Interpolation results remain robust across parameter variations, provided a non-zero nugget is used.

To make the manuscript structure completely transparent, we have updated the text as follows:

1. **Methodology & Table 2 (Section 3.5):** We rewritten Section 3.5 and added Table 2 to explicitly define all five evaluated variogram strategies, detailing their physical rationales and parameter sources.
2. **Sensitivity Analysis & Decision Rule (Section 4.3):** We added text clarifying that dynamic parameter fitting yields no practical performance gain over the fixed reference variogram, explicitly establishing why the reference strategy was adopted for all main performance evaluations (lines 296–298).
3. **Contextualized Assumptions in the Discussion (Section 5.3):** We expanded the discussion on variogram assumptions, explaining why spatial weighting structures buffer against parameter variations and why a non-zero nugget is required for reliable merging (lines 477–508).

Response to reviewer 2

I’m very happy with how the authors took my comments on the previous version into account in the new version of the paper. I have some remaining points, and some remarks on the new parts of the manuscript that I would like to see addressed. Please note that line numbers below refer to the lines in the track-changes version of the revised manuscript.

- On line 126, the small constant of 0.1 will not help with zero radar observations (i.e., F(v) will still be infinitely large). I suggest using something like “F(v) = ln((R_CML(v)+eps) / (R(v) + eps))”.

We thank the reviewer for this excellent point. In our original pipeline, we bypassed the division-by-zero issue by explicitly masking observation pairs where R(v) = 0. The small constant eps in our approach was intended solely to avoid ln(0) when R_cml = 0. Consequently, the original text inaccurately stated the purpose of epsilon.

However, we completely agree that the reviewer's suggested formulation, F(v) = ln((R_CML(v)+eps) / (R(v) + eps)), is mathematically better and provides more stable ratios during light rainfall events. We have adopted this formulation in both the manuscript and our codebase (line 148).

- The new Eq. (2) on line 140 is incorrect: it should be “R_merged(v) = exp(F_hat(v)) * R(v) – eps” (or, if my suggestion from the previous remark is used, is should be “R_merged(v) = (R(v) + eps) * exp(F_hat(v)) – eps”)

We thank the reviewer for pointing out the typed error in eq 2. This error was not reflected in the codebase and the results would stay the same. However, as we agree to change the transformation function (see comment above) the backtransformation also changes.

In the updated manuscript and codebase we now use “R_merged(v) = (R(v) + eps) * exp(F_hat(v)) – eps” as backtransformation function, as suggested by the reviewer (line 157).

This improved the multiplicative PCC and RMSE score, especially when CMLs more than 20km away from rain gauges are included (see updated fig 7).

- On lines 214-217, the nugget and sill for the reference variograms are stated to be the same for all methods. However, by definition, they cannot be the same for the multiplicative and the additive methods because one is unitless, and the other has units of (mm/h)^2. The values can, by pure coincidence, be the same (e.g., 1 (mm/h)^2 and 1), but they represent physically very different things. The same holds for the nuggets discussed in Section 4.3. Please adapt the manuscript to reflect this.

We thank the reviewer for this helpful comment. We clarify that all experimental variograms are standardized by dividing by the total spatial variance of the event prior to model fitting. This division removes the physical units, transforming the semivariance values into a strictly dimensionless space where the total sill is fixed to c = 1. Consequently, the fitted nugget and sill parameters represent unitless proportions of the total variance rather than absolute physical values, making them mathematically compatible and directly comparable across all merging methods.

We acknowledge that this standardization procedure was not clearly articulated in our original draft. In the revised manuscript, we have updated Sections 3.5 and 4.3 to explicitly define this dimensionless standardization, removed any conflicting physical units for nugget

and sill, and noted that the scale-invariance of the kriging equations preserves the exact same interpolation weights and merged estimates.

- On line 223, how is a time step's variance defined? Is it simply the variance of all of the data points available for that time step?

We thank the reviewer for raising this question. The statement regarding standardizing each time step by its individual variance s_t^2 was an error in our text description that did not accurately reflect our numerical implementation.

In our algorithm, data pairs are not standardized per individual time step. Instead, observation pairs across all time steps within a rainfall episode are pooled together to calculate the empirical semivariances. Standardization is then applied at the episode level by dividing each bin's semivariance by the episode's total spatial variance, defined as the mean semivariance of all bins between the 5th and 95th distance percentiles.

We have revised Section 3.5 by removing the s_t^2 term and equation, updating the text to describe the episode-level standardization correctly (lines 210 - 222).

- On lines 226-231 the time aggregation before variogram estimation is discussed. It is not clear to me how the sill is determined from the individual standard deviations of the underlying time steps. Please add this to the paper.

We thank the reviewer for this comment. Individual time step standard deviations are not used to calculate the sill. Instead, observation pairs from all time steps within an episode are pooled together, and the empirical semivariances are calculated for the entire episode. The total sill is then fixed to c = 1 by dividing these binned semivariances by the episode's overall variance (the mean semivariance of bins between the 5th and 95th distance percentiles). Section 3.5 has been updated to clarify this episode-level standardization process (lines 216 - 222) .

- On lines 235-236 it is stated that the variogram parameters are estimated from the radar-to CML ratios (multiplicative merging) or from the radar-CML differences (additive merging and KED). For KED, it would make more sense to me that the variogram would be based on either the CML or radar rainfall estimates, because that's what you're interpolating. And given that, radar data would give you much more information about the spatial structure of precipitation, so that would make radar more suitable for variogram parameter estimation. On the other hand, the nugget may also include CML measurement errors, so you would need to include those errors separately.

We thank the reviewer for this thoughtful observation. In Kriging with External Drift (KED), theoretical consistency requires modeling the variogram of the residuals rather than the raw primary variable or external trend field.

While radar data capture spatial structure well, estimating the variogram directly from the smooth radar field yields a nugget value near zero because it lacks point-scale observational noise (as shown in Fig. 3). Estimating the variogram from radar–CML differences ensures that both the residual spatial correlation and the CML-specific measurement

noise/micro-scale variability are directly captured in the nugget effect. This prevents the KED interpolation from overfitting to noisy CML observations without requiring an arbitrary, manually specified error term.

We have added a brief clarification on this rationale to Section 3.5 and cited Haberland 2007 (who also uses difference between radar and raingauge for estimating variogram used in KED) as support for using difference-based variogram estimators for KED (line 212).

- In Figure 3, and in Section 4.2, it would be interesting to also see results for the variogram parameter estimates if the entire radar field were used (see my previous point), and not just the points at the CML/gauge locations.

We thank the reviewer for this insightful comment. In response, we performed a synthetic sampling sensitivity analysis (restricted to rain gauges in the OpenRainER dataset) to evaluate whether restricting radar evaluations to sensor locations introduces spatial sampling bias or misses fine-scale variability compared to denser uniform spatial coverage.

1. Experimental Setup & Domain Bounding To test the effect of spatial density within the study domain, we generated synthetic gauge networks bounded strictly within the geographic envelope (latitude/longitude extent) of the real rain gauge network. We evaluated empirical variograms across 29 rainfall events for three synthetic density tiers (100, 300, and 500 gauges randomly distributed over the radar grid) and compared their estimated range and nugget parameters against the actual rain gauge locations (~280 gauges).

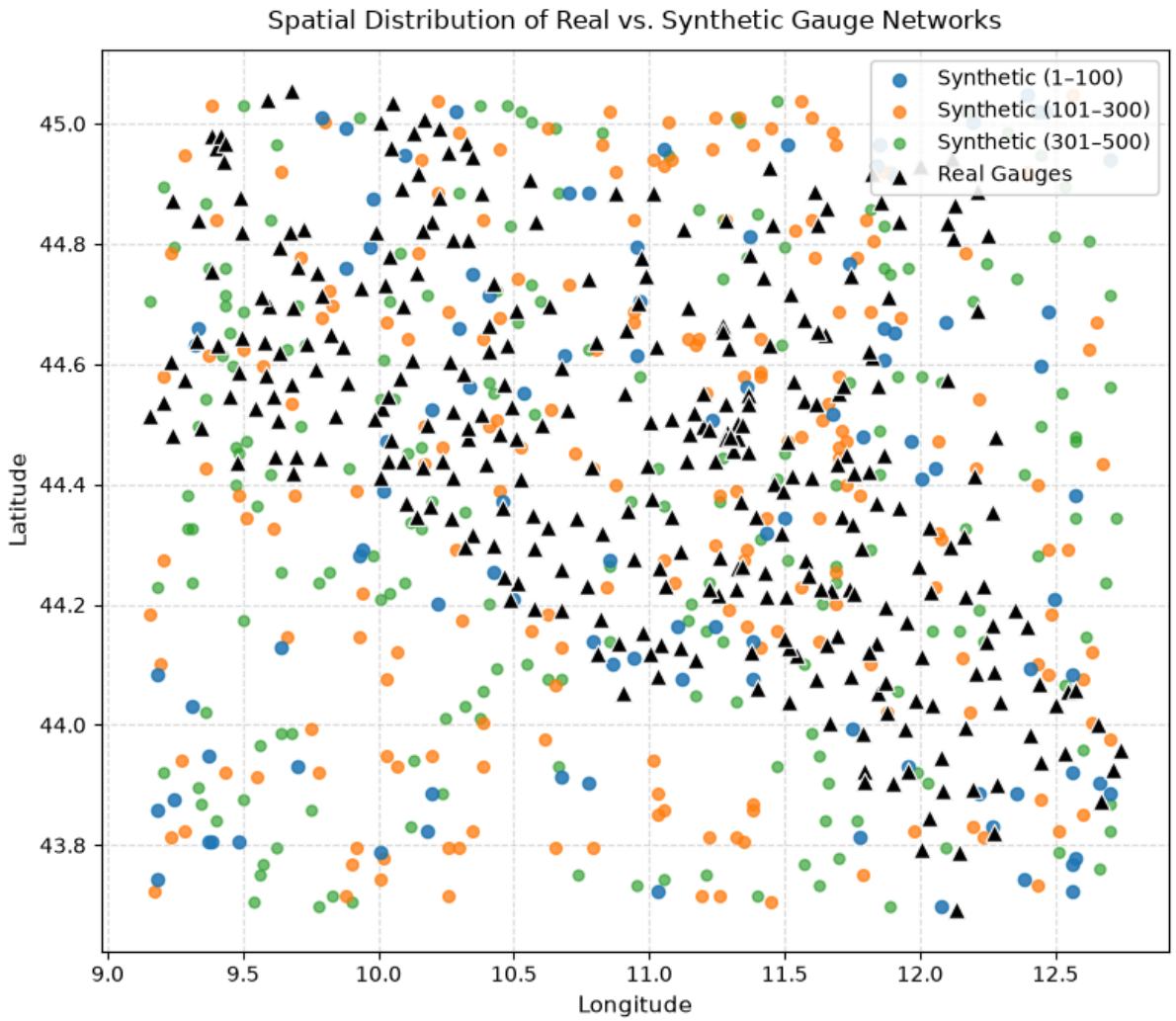


Figure R1: *Spatial distribution of real rain gauges compared against synthetic gauge networks (100, 300, and 500 locations) sampled strictly within the spatial envelope of the study domain.*

2. Parameter Convergence and Network Densities As shown in Figure R2, variogram range and nugget estimates exhibit clear convergence:

- **Under-sampling at 100 gauges:** A sparser network of 100 gauges introduces noticeable instability, showing elevated nugget estimates (e.g., reaching 1.0 during low-signal event periods in mid-July) and greater variance in estimated range.
- **Convergence at 300–500 gauges:** The parameter trajectories for 300 and 500 synthetic gauges align closely with the estimates obtained from the real gauge locations (~280 gauges) across almost all events.

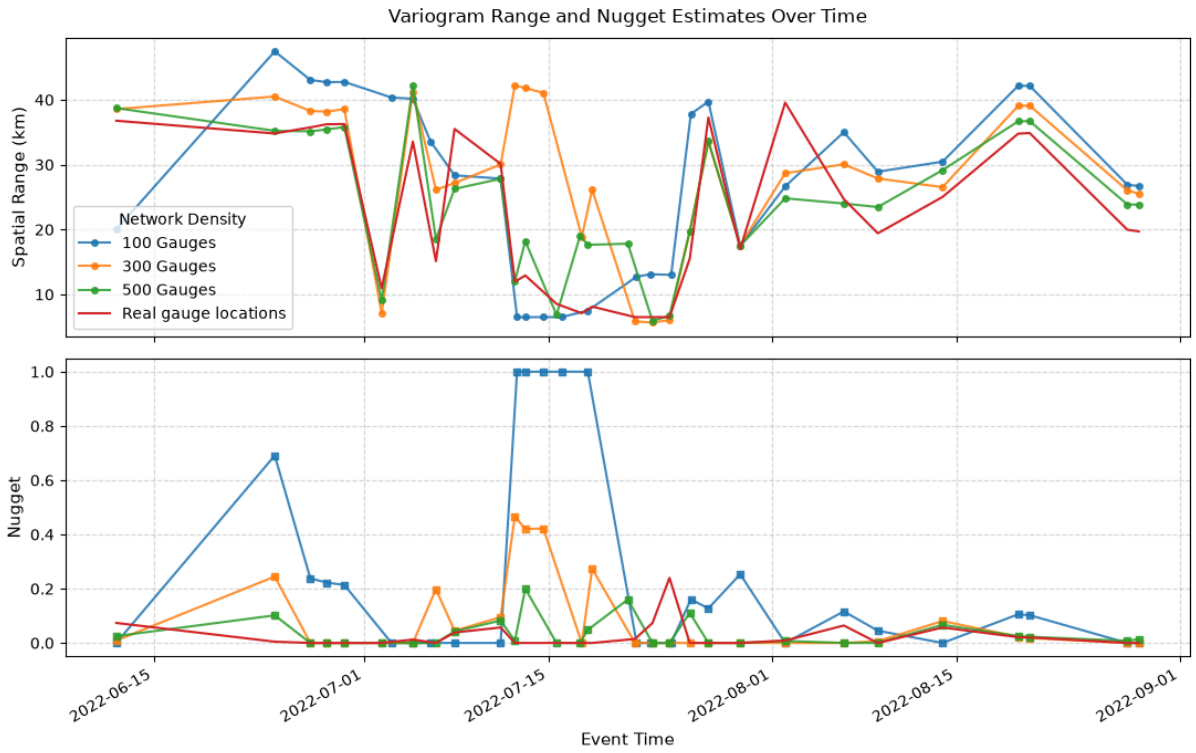


Figure R2: *Time series of spatial range (km) and nugget estimates across 29 rainfall events for 100, 300, and 500 synthetic gauge configurations vs. actual rain gauge locations.*

3. Isolating Network Topology and Sampling Geometry *Beyond parameter convergence, our primary objective in sampling the underlying radar field at sensor coordinates (CMLs vs. rain gauges) is to isolate the effect of spatial network geometry (line-integrated links vs. point gauges) from physical measurement error and domain extent differences. Using identical spatial envelopes and bounded reference points ensures that observed parameter discrepancies stem directly from sensor topology rather than differing domain bounds.*

Manuscript Revision:*For conciseness, we have provided the complete sensitivity analysis and figures above within this response letter for your reference. In the manuscript, we have updated Section 4.2 (lines 284 - 286) to clarify that ~150 sampling locations are sufficient to resolve the spatial correlation scale in OpenRainER:*

- In Figure 3, I think that the sill should be shown as well.

We thank the reviewer for this comment. Throughout this study, all experimental and theoretical variograms are variance-standardized so that the total sill is normalized to c = 1. This standardization isolates the spatial correlation structure (nugget c_0 and range h_r) and allows direct comparison across different rainfall events and sensor configurations, regardless of differences in absolute rainfall variance. Since the psill is just 1 - nugget, we do not think its necessary to show psill either. See also reply above.

- For interpolation, the part of the variogram that is mostly used is the part close to d=0. In that range of distance values, the spherical variogram becomes close to linear, with

two parameters: the nugget (intercept) and the sill-to-range ratio (slope). It would be interesting to see how that ratio behaves over time (e.g., does it vary less wildly than the range itself?). So this is another variable that could be added to Fig. 3 (possibly instead of the sill and/or range; see my previous remark). This would also help in the discussion on lines 548-558.

We thank the reviewer for this excellent point regarding the variogram behavior near the origin. Because all variograms in our study are standardized to a total sill of c = 1, the initial slope is determined by the ratio of the partial sill (psill) to the range: (1 - c_0) / h_r.

To test this hypothesis, that this combined ratio might vary less wildly than the range itself, we calculated and plotted this metric for our time series (see **Figure R3** below).

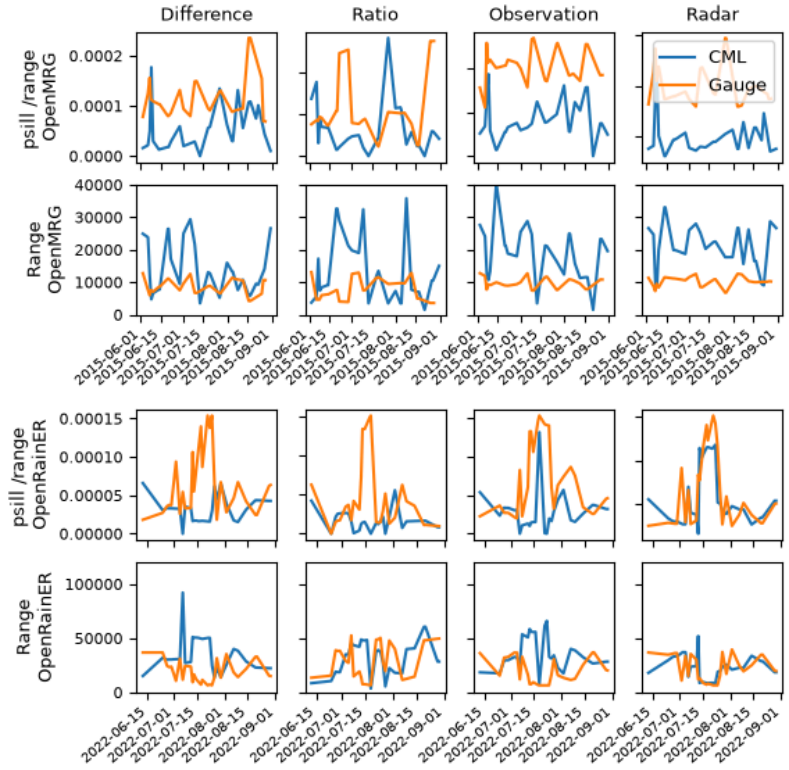


Figure R3: Time series of the variogram slope/ratio (1 - c_0) / h_r (top rows) compared to the original range h_r (bottom rows).

As shown in Figure R3, the ratio actually show slighly greater temporal volatility than the range itself. Because h_r acts as the denominator, minor dips in range mathematically magnify the ratio, creating severe spikes (e.g., the gauge estimates in OpenMRG). Furthermore, the resulting units (variance per km) yield small fractional values that are difficult to interpret physically. For these reasons, we have opted to retain the range (h_r) in the main manuscript's Figure 3, as distance in kilometers provides a cleaner, more intuitive metric for storm scales.

However, your point prompted a valuable addition to our discussion. Because this near-origin gradient inherently fluctuates along with c_0 and h_r, it confirms that the stability of our merging performance is not due to stable (hidden in the ratio) variogram parameters,

but rather the inherent robustness of the merging algorithms. We have updated Section 5.3 to clarify this connection (lines 477 - 498)

- Can a legend be added to Fig. 4? The caption has the necessary information, but it would help the reader to have a legend in the figure itself as well.

We appreciate the reviewer's suggestion to improve the readability of Figure 4. However, we respectfully prefer to keep the figure in its current layout without an embedded legend for two main reasons:

1. A 5-item legend displaying the distinct hatch patterns requires considerable vertical/horizontal space. Placing it inside either subplot risks obscuring data features—most notably the pronounced performance spike in `mul_b_ok` in the bottom panel (OpenRainER), which extends close to the upper axis limit. Placing a legend outside the subplots would significantly reduce the vertical panel height and compress the y-axes.
2. Within every method cluster, the five sub-bars strictly follow the left-to-right order corresponding directly to strategies (1) through (5) listed in the caption.

To make this mapping immediately clear to the reader without cluttering the visual space, we have clarified the figure caption to explicitly highlight this left-to-right sequence.

- Can a legend be added to Fig. 8? The caption has the necessary information, but it would help the reader to have a legend in the figure itself as well.

As with Figure 4, we respectfully prefer to retain the current layout for Figure 8 without an embedded legend for two key reasons:

1. Adding a legend box within either panel risks obscuring notable performance peaks—such as the prominent error spikes for `ked_p` and `ked_b` (exceeding 4.0) in the OpenRainER dataset (bottom panel). Placing the legend outside the subplots would compress the panel height and reduce the vertical resolution of the bars.
2. Within each method cluster, the three bars strictly follow the left-to-right sequence corresponding to the three range-check configurations: (1) no check, (2) standard, and (3) strict.

To ensure immediate clarity for the reader without compromising the plot's visual space, we have updated the caption of Figure 8 to explicitly state this left-to-right orde.

- On line 448, “MMB” is referred to without an explanation of what it is, so please add “moving median bias adjustment” to the paper.

We have updated this line to explicitly define MMB alongside the other bias adjustment acronyms (MFB and IDWB) upon their first mention in the manuscript (line 419-420 in the updated manuscript).

- On line 560 it is stated that “the adjustment or residual fields vary more gradually than raw rainfall”. If that were the case, we should see significantly longer ranges for the two

right-hand panels than for the two left-hand panels in Fig. 3 (if I understand well what is plotted there). That doesn't seem to be the case, so please rephrase this statement.

We thank the reviewer for this observation. You are entirely correct that empirical variograms of residual and ratio fields in Figure 3 do not display systematically longer correlation ranges than raw rainfall. Further this could be explained by physics, as key radar errors, such as path attenuation and Z–R conversion uncertainties, scale directly with rainfall intensity, causing spatial error structures to inherit the correlation structures of the storm events.

In the updated manuscript we have removed the statement regarding residual field smoothness and updated the paragraph to clarify that algorithm stability likely stems from the spatial weighting and smoothing mechanics of the merging methods rather than an increased spatial range in the residual fields (lines 485 - 499).

- On lines 560-561, it is stated that "specific range and sill values have a limited practical impact". This could be true, but I would argue that the sill-to-range ratio is practically most important because most of the interpolation is done where the distance is much smaller than the range. For the same reason, the difference between using a spherical or a linear variogram (as in Haberlandt, 2007) would be limited. I suggest to add this to the discussion.

We thank the reviewer for this insightful comment regarding near-origin variogram behavior.

To test whether the sill-to-range ratio (1 - c_0) / h_r acts as a stable driver of performance, we plotted this metric over time across our datasets (Figure R3 above). As shown, the ratio exhibits substantial temporal volatility alongside the individual parameters, confirming that merging stability stems from algorithmic spatial weighting rather than a stable near-origin slope.

We also fully agree regarding model geometry: because spatial interpolation relies primarily on short inter-sensor distances where linear and spherical models behave nearly identically, choice of variogram parameters has minimal practical impact. This insight strengthens our discussion of (Haberland 2007), indicating that the high errors in their linear model were driven by omitting a nugget (zero nugget) rather than the linear shape itself.

We have updated Section 5.3 to incorporate both the near-origin ratio evaluation and this short-distance model equivalence when re-interpreting the findings of Haberland (2007), (lines 494 - 498).

- On line 659, it is stated that "geostatistical merging assumes that ground-based observations represent the truth". This may be the case, but at least Kriging gives the possibility to include uncertainty information into the nugget parameter (which represents either extremely small-scale variability or measurement error). This could be included in the discussion.

We thank the reviewer for this insightful comment. We agree that Kriging frameworks can incorporate measurement error and micro scale variability through the nugget parameter or extensions like Kriging for Uncertain Data (also mentioned in Appendix A1, line 625) . We have updated the manuscript to explicitly acknowledge this capability (lines 561). We also

clarify that while adjusting the nugget parameter filters zero mean random noise, it cannot inherently resolve a persistent network wide bias (see lines 561 - 565).