We would like to thank all three Reviewers for their thoughtful comments and constructive feedback on our manuscript. Their input has helped us significantly improve the quality, clarity, and overall rigor of our work. In response to the reviewers' comments and during our internal review of the manuscript, we have made several changes, primarily focused on the Results section and the Appendix.

Before replying to the comments of the individual reviewers, we provide below is a summary of the principal updates integrated into the revised version:

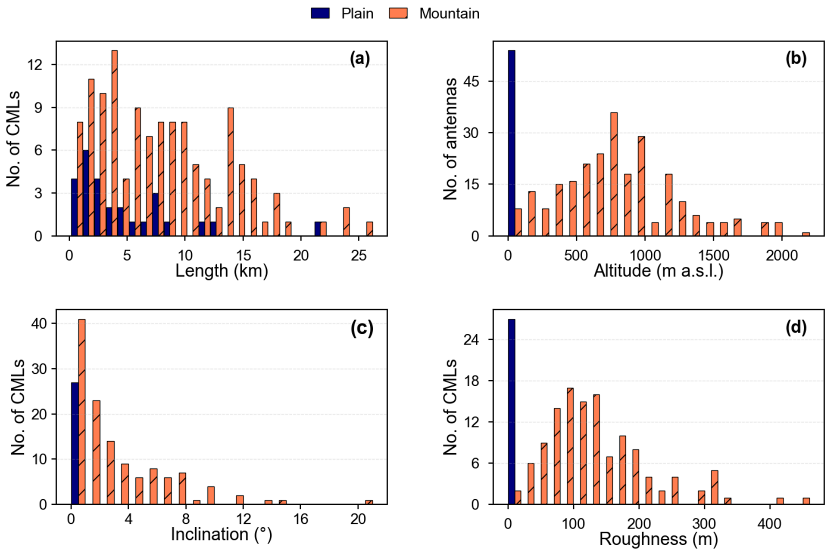
- **Correction and Refinement of Radar–CML Link Analysis (Section 4.2, former 3.2):** We identified and corrected a calculation error regarding the integration of radar information along the Commercial Microwave Link (CML) paths. Following this correction, we re-analyzed the case study. In doing so, we incorporated an additional rain gauge into the evaluation, added CML rainfall estimation and expanded the description of the experimental setup to better highlight the presence of complex terrain and local orography.
- **Stratification of Analysis by Link Characteristics (Section 4.3, former 3.3):** In accordance with the reviewers' recommendations, we categorized the performance analysis based on CML characteristics. The updated KPIs demonstrate a clear correlation with link length, aligning with established literature. Our long-term analysis revealed no significant performance discrepancies when comparing links deployed over complex terrain to those in flatter areas; therefore, a deeper investigation into this topic lies beyond the scope of this paper.
- **Clarification of the Appendix:** We have substantially revised the Appendix to clarify the primary objectives of the study and to improve overall readability and logical flow.
- **Abstract and introduction reviewed**.

In addition, the entire manuscript has been thoroughly polished to correct minor typographical errors, improve grammar, and enhance overall readability. Below, we outline the main modifications regarding the figures, tables and the structure of the paper:

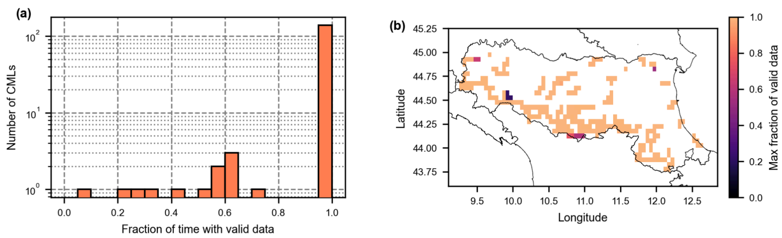
- Figure 2: Readability was improved by increasing the overall figure size and switching from stacked bar charts to side-by-side grouped bar charts.
- Figures 3 and 6: Added a 2D map displaying the spatial data availability of the instruments across the study region (panel (b)).
- Figure 7: Updated the color schemes in panels (b) and (c) and aligned their y-axis limits for direct comparability.
- Figure 8: Applied a discrete colormap to panels (b) and (c) to enhance data clarity.
- Figure 9: Completely revised to show the time series of CML attenuation alongside radar reflectivity.
- Figure 10: Added new panels, showing rainfall comparison and terrain orography for the analyzed event.
- Figures 11 and 12: Introduced as new figures replacing the former Figure 11, showing the long-term evaluation metrics between CML estimates and adjusted radar as a function of CML characteristics.
- Figure A1: enlarged.
- Table 1: Added a new table summarizing the open-source datasets available over the region Emilia-Romagna.
- Table 3: Added a new table to summarize OpenRainER variables.
- Table 5: former table 3. We added the statistics relatively to the rainfall amount. Now columns also display the numbers of stations with more than 25% and 75% of data.
- Manuscript Restructuring: The manuscript is now organized into 5 main sections:
    - 1) Introduction
    - 2) Study Area

- 3) Dataset (formerly Section 2)
- 4) Results (formerly Section 3)
- 5) Conclusions (formerly Section 4)

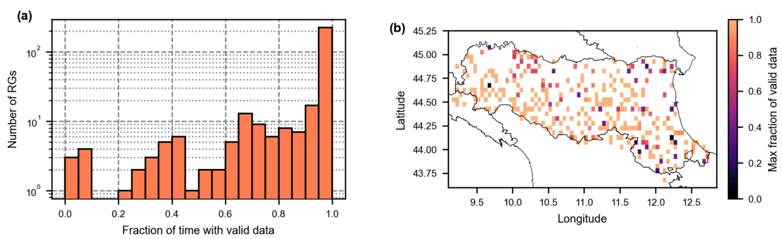
Finally an updated version of the README file was uploaded on Zenodo (v 2.0.2).



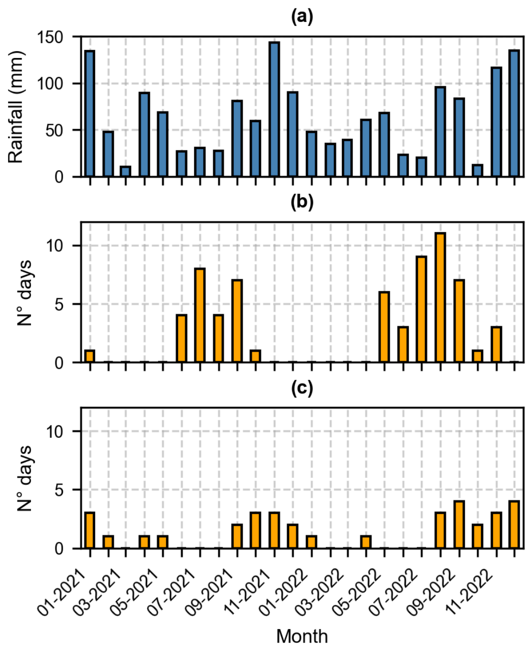
updated figure 2



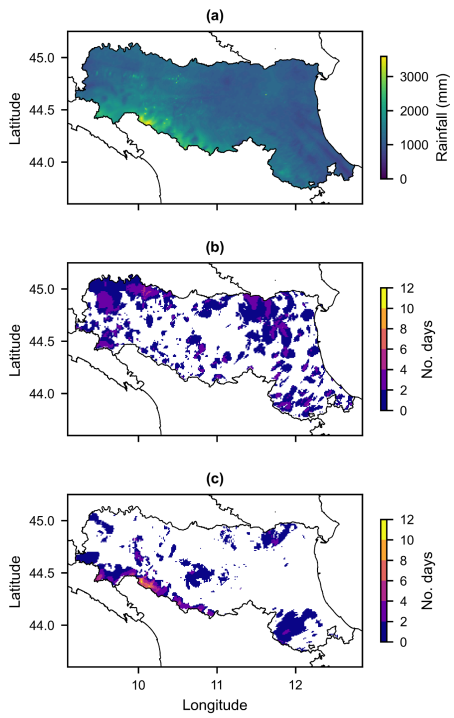
updated figure 3



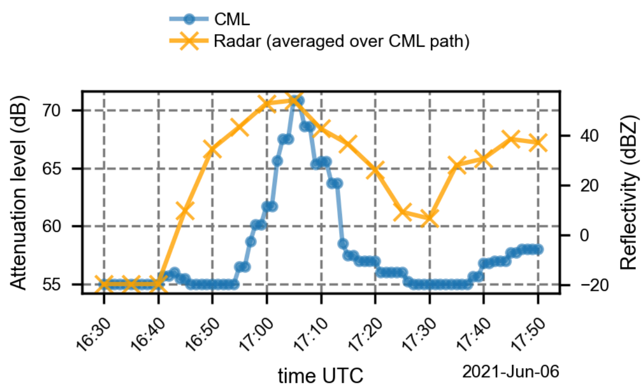
updated figure 6



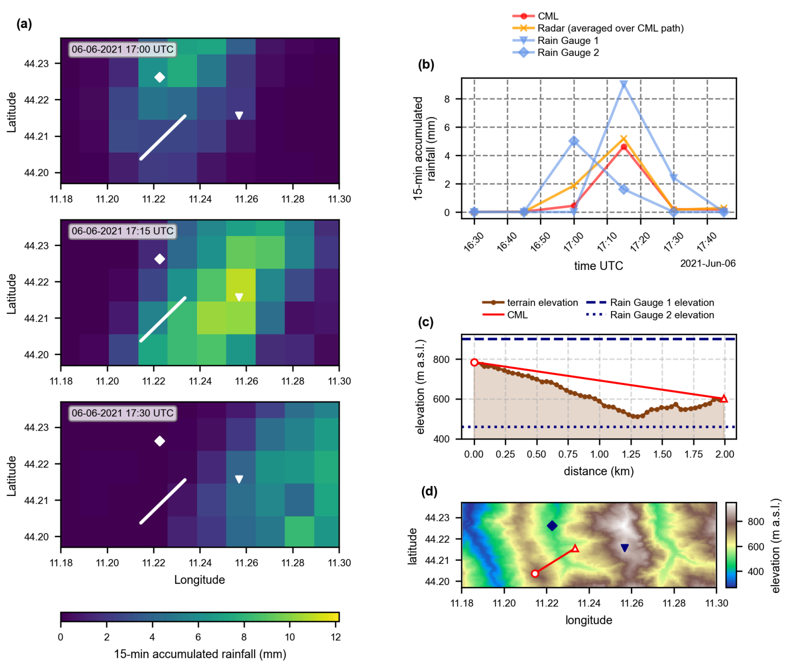
updated figure 7



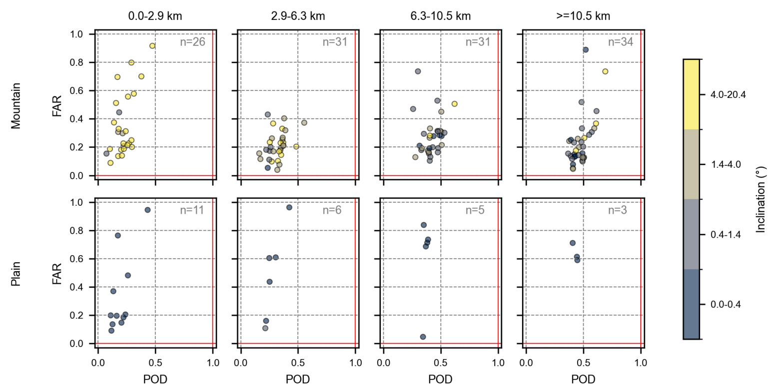
updated figure 8



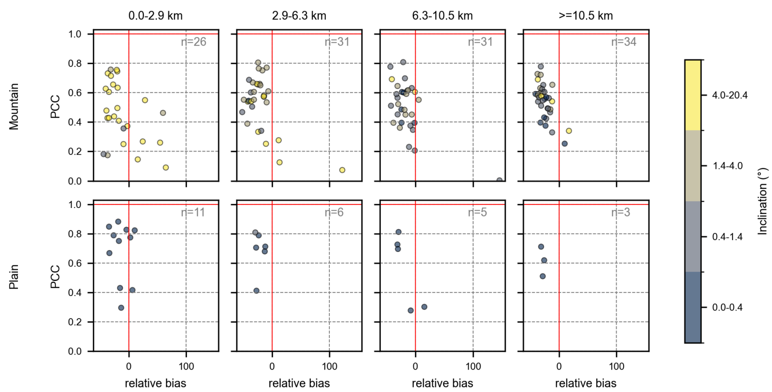
updated figure 9



updated figure 10



updated figure 11



updated figure 12

Table 1. Emilia-Romagna open-source precipitation-related datasets.

Name	Link	Description
OpenRainER (this work)	https://doi.org/10.5281/zenodo.10593848	Two years of precipitation data from CML, radar and weather stations.
ERG5 analysis	https://dati.arpae.it/dataset/erg5-interpolazione-su-griglia-di-dati-meteo	Complete set of environmental and atmospheric variables interpolated over the Emilia-Romagna region (Antolini et al., 2015).
Database of the Italian disdrometer network	https://doi.org/10.5281/zenodo.7708563	Laser disdrometer located inside the campus of the University of Bologna (Adirosi et al., 2023).
Meteonetwork	https://www.meteonetwork.it/	Personal Weather Stations (PWS) from the Meteonetwork association.

table 1 (added)

Table 5. Availability of the data collected by weather stations at ground in 2021-22.

Observed variable	No. stations	No. stations $\geq$ 25% of data	No. stations $\geq$ 50% of data	No. stations $\geq$ 75% of data
rainfall	319	311	294	263
temperature	202	140	27	10
wind direction	37	19	9	2
wind velocity	37	19	9	2
relative humidity	92	35	20	9

updated table 5 (former table 3)

Table 3. OpenRainER dataset variables and descriptions inside the NetCDF files.

File name structure	Name	Longname / Description
CML_*Start*_*End*.nc	rsl	Received signal level [dBm]
	tsl	Transmitted signal level [dBm]
	time	UTC time [s]
	sublink_id	Sublink identifier
	cml_id	Commercial microwave link identifier
	length	Distance between pair of antennas [m]
	site_0_lat	Site 0 latitude [degrees in WGS84]
	site_0_lon	Site 0 longitude [degrees in WGS84]
	site_0_elev	Ground elevation above sea level at site 0 [m a.s.l.]
	site_1_lat	Site 1 latitude [degrees in WGS84]
	site_1_lon	Site 1 longitude [degrees in WGS84]
	site_1_elev	Ground elevation above sea level at site 1 [m a.s.l.]
	frequency	Sublink frequency [MHz]
	polarization	Sublink polarization
RADref_*YearMonth*.nc	reflectivity	Radar reflectivity factor [dBZ]
	time	UTC time [s]
	lon	Longitude [degrees WGS84]
	lat	Latitude [degrees WGS84]
	geo_dim	Geographical limits [yLL, xLL, yUR, xUR]
	mesh_dim	Grid mesh size [X_mesh_size, Y_mesh_size]
	mosaic	Mosaic description codes (1: Gattatico, 2: S. Pietro Capofiume, 3: Both)
RADrain_*YearMonth*.nc	rainfall_amount	Radar rainfall amount [mm]
	time	UTC time [s]
	lon	Longitude [degrees WGS84]
	lat	Latitude [degrees WGS84]
	geo_dim	Geographical limits [yLL, xLL, yUR, xUR]
	mesh_dim	Grid mesh size [X_mesh_size, Y_mesh_size]
	mosaic	Mosaic description codes (1: Gattatico, 2: S. Pietro Capofiume, 3: Both)
RADadj_*YearMonth*.nc	rainfall_amount	Adjusted radar rainfall amount [mm]
	time	UTC time [s]
	lon	Longitude [degrees WGS84]
	lat	Latitude [degrees WGS84]
	geo_dim	Geographical limits [yLL, xLL, yUR, xUR]
	mesh_dim	Grid mesh size [X_mesh_size, Y_mesh_size]
	mosaic	Mosaic description codes (1: Gattatico, 2: S. Pietro Capofiume, 3: Both)
AWS_*YearMonth*.nc	rainfall_amount	15min accumulated rainfall [mm]
	temperature	Measured temperature [°C]
	relative_humidity	Relative humidity [%]
	wind_velocity	Measured wind speed [m/s]
	wind_direction	Wind direction [degree]
	time	UTC time [s]
	id	Weather station identifier
	latitude	Weather station latitude [degrees WGS84]
	longiude	Weather station longitude [degrees WGS84]
	elevation	Ground elevation at site [m a.s.l.]

table 3 (added)

RC1: 'Comment on essd-2026-160', Anonymous Referee #1, 05 Jun 2026:
https://doi.org/10.5194/essd-2026-160-RC1

The authors present the OpenRainER data set. The data set consists of commercial microwave link (CML), weather radar, and rain gauge data with the main goal to improve the usability of the CML data for rainfall monitoring. The data set offers a significant contribution to the community involved in rainfall monitoring with CMLs. As an open data set in a field where the raw data is often non-disclosed, this CML data set can be of high value in benchmark studies. It is uniquely different from the other three open data sets mentioned in the paper, and the addition of weather radar and rain gauge data complete the data set. The data itself is of sufficiently high quality, though the limitations, especially of the CML data, should be elaborated on to improve the usability. The manuscript with the description and analyses, in my view, currently does not sufficiently support the data set, other than having it published on Zenodo. I have outlined my major concerns regarding these points below.

Preprocessing CML data

The TSL and RSL time series are processed in-house by the data owner. Even though the details are not disclosed, it would be valuable to be as precise as possible. Is it just the sampling frequency and quantization that are unknown? Because that is the only thing the appendix focuses on. If so, mention this concretely in the main text (L120-126). If the authors suspect other forms of preprocessing were done it would be helpful to know why the authors think so, and if they could share some of these processing steps they suspect have been done based on their experience with this data set. The appendix on the in-house processing is an important addition and I appreciate the authors have looked into this, however, in my view it is not sufficiently well written and difficult to understand at this point. Unfortunately the example in Line 259 does not make it much clearer to me. Maybe it is clear to the authors, but I suggest having a colleague (not involved in this exact field of research) to read through it and iterate together to make it more understandable.

Reply: We replaced this statement at the end of (L.120-126 in the original manuscript), with a new paragraph (last paragraph of Sec. 3.1, former 2.2). We clarified what is unknown, that is, the sampling frequency and the signal quantization. We also stated that our arguments are discussed in Appendix A and anticipated the outcomes of the appendix. In addition, we entirely revised Appendix A. We think that now it is understandable even to readers who are not familiar with signal format and data processing in telecommunications. Below, we address the reviewer's specific comments regarding the Appendix.

Some examples of what is unclear to me in the Appendix:

L251 -257 : How I understand it is within 1 minute there are N samples being recorded, and each of these samples has a quantization somewhere between RSL0 and RSLK . So K is simply the number of possible quantization values? Then shouldn't RSL0 + k in equation 1 actually be RSL0+k (same in L257)? If K is the quantization value itself, then I would expect equation A1 to be the sum of Nk(RSLk) and not Nk(RSL0 + k). Otherwise the range of RSL0 to RSLK mentioned in L254 does not make sense.

Reply: We assume that in an N-sample window, RSL can take one among K+1 possible quantized values. If the quantization is 1 dB and RSL_0 is the minimum RSL value within the window, then, the possible RSL values are $RSL_0 + 0$, $RSL_0 + 1$, …, $RSL_0 + K$ (dBm), assuming that the RSL takes all quantization levels from the minimum to the maximum, which is usually the case. Please note that, as all RSL values are integer numbers, we can write the generic sample as $RSL_0 + k$, where k=0, 1, …, K. Equation A1 is correct, it multiplies the k-th possible value of RSL, that is, $RSL_0 + k$, by the number of samples N_k with that value.

L259: Doesn't Nk already vary between 0 and K, instead of between 0 and N?

Reply: No, N_k is the number of samples with RSL value equal to RSL_0 + k within a set of *N* samples.

EqA.3: follow the same format as in Eq. A1 and write it out completely. E.g. if *K=1* I would expect Eq. A3 to be 1/N * (N_0 * RSL_0 + N_1*(RSL_0+1)) based on the format in Eq. A1

Reply: Exactly. By further noting that $N_1=N-N_0$, we get the expression in Eqn. A.3

Having someone outside of your direct field of research read through this appendix will hopefully improve the readability and make it more accessible to those outside the direct CML community too. Finally I would suggest making figures A1 and possibly A2 a lot larger (w)ider) as now it is very difficult to deduce anything from these time series or follow along when the authors refer to these figures. For example, the bulk 1 dB quantization step in L251 is not deducible from Fig. A1 based on the current vertical axis.

Reply: We thank the reviewer for suggesting this test. We revised the Appendix and presented it to readers who are not experts in signal processing. They were able to understand the topic discussed. We increased the size of Figure A1, adding a denser set of labels on the y-axis. Ticks are now every 1 dB and labels every 2 dB. Now the range of values of the time series can be deduced from the figure. Instead, we did not change Figure A2, because this figure aims at showing, in a qualitative way, that the RSL values fill every bin of the histogram, where bin centres have been calculated according to the procedure explained in the text.

Additional analyses

To gain more trust in this opportunistic data source some additional analyses would be recommendable. Section 3.3. on rainfall comparison is rather limited at this moment. The authors state that the data set has some peculiarities compared to other open CML data sets, such as the presence of orography, and the use of nearly identical CML operating frequencies. However, nothing is done with these features at this moment. It would be fairly straight-forward to split the results in Figure 11 by orography / elevation, for example. The same for link length. Since the frequency is predominantly the same, how does the accuracy of the CML rainfall estimates vary with link length? This would quickly give a user some insights into the quality of the data that makes this data set unique.

Reply: while most of the CMLs tend to underestimate precipitation compared to the selected reference (negative relative bias), some of the shorter links show an overestimation. We replaced Figure 11 with 2 figures (11 and 12), where the selected KPIs are shown for different categories of links length and elevation.

Similarly, in L162 you state gauges are "the most accurate ground-truth reference against which CML-based rainfall retrievals can be compared in our case". However no such analysis is currently performed. To gain confidence in the (reference) data a short comparison between CMLs and the nearest gauge, similar to what is done for radar would be useful. Also a comparison between gauges themselves, a simple corrollelogram or double mass curve could already yield some insights into the quality of the reference data itself and potentially spot outliers.

Reply: Thank you for your valuable suggestions. The analysis section has been thoroughly revised, and the results are now stratified to highlight the impact of specific link characteristics (most notably link length) on data performance. In the comparative analysis, we chose to use the gauge-adjusted radar dataset as the reference rather than relying solely on rain gauge observations. This approach allowed us to include both ground-based precipitation data sources, without restricting the evaluation exclusively to nearby rain gauges, which would have limited the number of feasible comparisons. Finally, we decided not to include a dedicated rain gauge precipitation analysis in order to keep the results section concise and sharply focused on the evaluation of CML performance.

Added value of split between convective and stratiform – Section 3.1

The split between stratiform and convective is interesting to get an idea of the local climatology, but is not really used in any analysis at the moment. So what is the exact added value to this description of the dataset? A quick addition could be to split the results from Fig 11 in time and show the performance of CMLs vs. weather radar in stratiform and convective conditions.

Reply: The primary goal of defining these two classes is to provide a regional climatological context, showing where and how frequently these phenomena occur. This characterization serves as a practical baseline for data users to locate specific extreme events and select high-impact case studies within the dataset. We agree that this intent was not sufficiently clear in the original text. Therefore, we have revised the subsection to explicitly frame these definitions as a descriptive tool for case-study identification. Furthermore, conducting a long-term comparative analysis based solely on these events (as defined) would be particularly complex, given the spatial distribution of the instruments and the variable number of available observations.

In general I am not suggesting to do many different analyses, as that is not the goal of such a data paper, but simply to split the analyses currently done by some of the characteristics mentioned in the paper to get an idea of how the CML data performs in different conditions.

Specific comments on content:

L42: So you have 1 minute data but receive it every 15 minutes?

Reply: Correct. Text files are received every 15 minutes, containing data collected at 1-minute intervals.

L47-56: I would switch the order of this paragraph, and put it before L37-46, so that after describing the other data sets you first describe what makes the OpenRainER data set unique.

Reply: the history paragraph (L37-46) was mainly removed, following RC2 suggestion.

L51: arguably “wet-antenna attenuation” is not the most frequency dependent effect, so I would give another example. A much larger effect could be the presumed uniform attenuation along the path.

Reply: We revised the paragraph. “*OpenRainER stores the data collected by a CML network owned by the company Lepida S.c.p.A. Emilia-Romagna provides a great complexity of terrain, from vast plains to steep valleys, and of precipitation regimes, from extensive stratiform systems to deep convective thunderstorms. Within this context, the Lepida CML network itself exhibits marked spatial heterogeneity, comprising both sparse, isolated link groups and high-density spatial clusters that collectively provide higher coverage over some areas, notably along the Apennine ridge in the southern part of the region. The links span plains, urban areas, valley bottoms, and ridgelines. Links deployed in complex terrain span a wide range of elevations and exhibit varying tilt angles relative to the horizontal plane. Link lengths range from 155 m to 25 km, yet only two carrier frequencies are used across all link lengths, and they are very close to each other: 24.6 GHz and 25.6 GHz. At these frequencies, the $R-\gamma$ relationship is relatively insensitive to the raindrop size distribution (Chwala and Kunstmann, 2019), which reduces the errors involved in applying Eq. (1) across different precipitation regimes. However, the absence of the typical inverse relationship between frequency and link length, commonly observed in conventional commercial microwave networks, results in a wide range of* *rainfall sensitivities (expressed in dB per mm h−1; Roversi et al. (2020)). This presents a challenge for retrieval algorithms. On the other hand, these characteristics represent an optimal testbed for assessing rain/no-rain classification strategies and wet antenna correction algorithms."*

L57: Nice to mention other publicly available datasets! Hyperlinks in the text don't read nicely though. A small table would fit this better. You can then add some extra columns like whether these data are freely available which could benefit the community too.

Reply: We moved the hyperlinks in a dedicated table. All mentioned dataset are open-source.

L85: I would make "Study Area" Section 2, and "Datasets" a separate section 3.

Reply: We agree with the reviewer's suggestion and have updated the manuscript accordingly.

L126: To make the main text self-contained, explicitly mention the interesting characteristics you find in your Appendix here.

Reply: We revised the paragraph including this line, adding also the outcomes of the analysis in Appendix A. See also our reply to the reviewer's comment under the title **"**Preprocessing CML data**"**.

Fig 2: This and all other Figures, make them a lot larger as they are difficult to read now. Moreover, stacked bar charts are generally not a good idea since the reader is never sure if one of the bars covering the other. Instead put the orange and blue bars next to each other. If you must really have a single bar per bin, use different widths so the height of both bars are always visible. Finally, it would be useful to quickly define the term roughness in the caption as well.

Reply: We thank the reviewer for suggesting this change. Fig 2 enlarged and improved. Figures in the results part have also been revised.

Fig3: To enhance the usability, this data would better fit on a map so it becomes clear which links have less data, which would especially be of importance if they are all in the same regions.

Reply: We added this information by aggregating the data availability of the instruments over a regular grid. Figures were added as panel (b) for both CMLs and rain gauges (figures 3 and 6).

L142: Should it be "the one only covered by SPC"? Would be good to mention the mosaic codes in your README / dataset here.

Reply: We revised section 3.2 (former 2.3) by adding an exhaustive description of the mosaic codes (the variables are already reported in the README). Both radars can cover the region, but as the radar beam goes further from the radar the quality of the measure gets lower (beam broadening, higher sampling altitudes). We improved the description: "Radar products inside OpenRainER are derived from reflectivity measurements collected by two C-band weather radars operated by Arpae-SIMC. One is located in the western sector of the region, near the town of Gattatico (GAT), while the other one is near San Pietro Capofiume (SPC), in the eastern sector of the region. To identify which radar contributed to the composite, a 'mosaic' variable is included within the radar products. This variable can take three values: 1 if the map is generated from GAT measurements, 2 from SPC, and 3 from both radars combined"

L145: Would be good to add the resolution to the README on Zenodo too.

Reply: README on Zenodo has been reviewed and published (v 2.0.2) following the suggestions.

L147-154: If there is a publication describing the entire processing chain from the Arpae, or some comparison / validation studies on the radar data it would be valuable to mention that here for reference.

Reply: Unfortunately there are no dedicated publications available.

L158: What did these minor modifications do? Try to be as specific as possible to give the reader/user confidence in the data.

Reply: The modifications concerned the method used to clean radar data of unwanted signals. We expanded the sentence: ''Note that the radar processing chain was slightly modified over the two-year period covered by the dataset, specifically in the radar echo filtering process. As of 20 June 2022, a fuzzy logic-based method replaced the decision tree approach, resulting in a final product with fewer unwanted radar echoes and a better representation of rainfall structures.''

Fig6: Similar to Fig3, a map with the gauge locations colored by availability would be useful.

Reply: revised as figure 3 (see above).

Table 3: For completeness adding precipitation to this table would nice

Reply: We added the precipitation to table 5 (former table 3).

L174: Would be useful to know which stations or at least how many stations operate at temporal resolutions below 15 minutes. Possibly add this to table 3 or again, on a map.

Reply: After further investigation only one station showed temporal sampling below 15 minutes, but this was probably due to an error in the acquisition process. We added every detail in the paragraph. "*It should be noted that during April 2022, the station with ID S. Pietro Capofiume_1162264_4465378 recorded measurements at minutes 0, 6, 15, 21, 30, 36, 45, and 51, likely due to errors in the acquisition process.*"

L198: Explain shortly how the averaging of all radar pixels is done. Weighting the length of the CML in the radar pixel?

Reply: Yes, the averaging of the radar pixels is weighted by the length of the CML segment intersecting each pixel. Specifically, the weight assigned to each radar pixel is proportional to the fraction of the total link length that crosses that pixel. This interpolation was implemented using the open-source Python library poligrain (https://github.com/OpenSenseAction/poligrain).

L201-203: This hypothesis can be fairly quickly checked by coarsening the CML resolution to 15 minutes too, and/or shifting the timestamp to the beginning/end of each interval.

Reply: the paragraph was completely revised. See "Correction and Refinement of Radar–CML Link Analysis (Section 4.2, former 3.2)" above.

L204: "which reflect the different spatial sampling performed by rain gauges, CML and radars". Not necessarily true, or in any case not only true! It could be due to rainfall variability in time, differences in altitude, etc.

Reply: the paragraph was completely revised. See "Correction and Refinement of Radar–CML Link Analysis (Section 4.2, former 3.2)" above.

L210: " however it captures the temporal dynamics of rainfall better than the rain gauge". How so? The timing of the rain gauge seems to agree more with timing of the radar..

Reply: the paragraph was completely revised. See "Correction and Refinement of Radar–CML Link Analysis (Section 4.2, former 3.2)" above.

Fig8: Panel b and c show number of days, this a discrete variable, hence a discrete rather than a continuous colormap would be more useful.

Reply: We agree with this suggestion and changed the figure accordingly.

L223-224: In my opinion this conclusion is too easily drawn. To give the user of this data set some more confidence in the data it would be very helpful to simply stratify these results, as mentioned above in the general comments, by mountains/plains, link length, stratiform/convective.

Reply: the paragraph was completely revised. See "Stratification of Analysis by Link Characteristics (Section 4.3, former 3.3)" above.

L241: Lepida S.c.p.A. (the company that owns the CML network). -> mobile network operate Lepida S.c.p.A.

L 249: by the CML network -> in a network management system

L250: What is basically constant? Can you mention a range, i.e. -0.1 to +0.1 dB?

Textual comments:

L3: measurement -> measurements

L4: in 1 min -> of 1 min

L10: I would use "unique" instead of "peculiar"

L10: datasets: -> datasets.

L16: hereinafter -> hereafter

L17: by the -> by a, or in a

L21: "well known" is subjective, remove

L65: OpenSense has been -> OpenSense, a COST Action project that ended in October 2025, brought together..

L138: Joss and Waldvogel (1988) -> (Joss and Waldvogel, 1988)

L182: lapse? Do you mean interval?

L213: to to -> to

L214: estimate -> estimates

Reply: Thank you, we have corrected the reported errors.

Comments on the README on Zenodo:

- Change temporal frequency to temporal resolution.
- Write out RG/WR when using it for the first time
- If available, cite a publication that describes the radar processing chain used by Arpae.

Reply: README on Zenodo has been reviewed and published (v 2.0.2) following the suggestions.

RC2: 'Comment on essd-2026-160', Anonymous Referee #2, 09 Jun 2026 Citation: **https://doi.org/10.5194/essd-2026-160-RC2**

The paper describes a dataset of microwave link data for rainfall estimation completed by an extensive reference dataset. Two-year dataset of 151 microwave links sampled at 1-min resolution represent a valuable contribution to the reproductive research of opportunistic sensing. The paper is well structured and clearly written. It contains relevant descriptions of available data as well as rainfall activity during the covered period. All the figures are well designed with clear legends and titles. I recommend the paper for publication in ESSD. My suggestions for improvements are only minor and rather technical.

Reply: We thank the reviewer for recognizing the relevance of the proposed dataset. We have revised the manuscript to address the technical points highlighted in her/his general comments. Our point-by-point responses to the reviewer's comments are provided below.

L37-46 The whole history of collaboration is not directly related to the paper and too lengthy.

Reply: We agree with the comment and have removed the paragraph from the revised manuscript.

L54-56 different lengths with the same frequencies are useful not only for testing dry-wet strategies but also e.g. for wet antenna correction algorithms.

Reply: We thank the reviewer for the suggestion; the example has been added to the paragraph.

Section 2.2.: Please provide at least an approximate quantization of TSL and RSL as calculated in Appendix A already here. To my understanding of Appendix A, the quantization of RSL is around 1 dB, however, I am not sure what it is the quantization of TSL.

Reply: We are afraid it is not possible to provide an approximate quantization step for RSL. In fact, in the Appendix, we heuristically demonstrated that the 1-min RSL values in the dataset are averages over 60-samples of raw data with a 1-dB quantization. If all the 60 samples were equal (though different from one 60-sample window to another), the quantization step of the output RSL time series would be 1 dB. On the other hand, if the RSL undergoes variations in the order of several dB within the 1-min window (which may happen during rain events), then the output RSL will have a modulus after division, which is a rationale number (in dBm), dependent on the occurrences of each raw RSL value within the window, according to Eq. (A2). For instance, if 10 samples are -31 dBm, 40 are -30 dBm and 10 are -29 dBm, then the 60-sample average will be -30 dBm. Instead, if 15 samples are equal to -31 dBm, 40 are -30 dBm and 5 are -29 dBm, then the average would be -30.1667, with a modulus equal to 1/6. As many of such combinations are possible, the modulus can take really many values. Therefore, in our view, it would be misleading to provide an approximate quantization step for the output RSL. The TSL data were pre-processed by the data owner in a similar manner. The TSL is orders of magnitude higher than the RSL and it is usually an extremely stable signal. Therefore, the 1-min TSL average is obtained in many cases from 60 equal raw values, resulting in flat TSL time series exhibiting integer values (in dBm). An exception occurs when ATPC (Automatic Transmit Power Control) is activated. When the RSL decreases below a certain threshold (e.g. during heavy rain), the TSL is increased, typically by a constant dB amount, to compensate for the RSL loss. This process is basically instantaneous with respect to the time scales considered here. The same occurs when ATPC is switched off. However, if the 1-min window encompasses one of the above TSL step-like variations, the output TSL will show a non-integer dBm value, which depends on the position of the TSL step within the window.

L121: consider rephrasing such as it is clear you do not mean at least one sublink in the whole dataset but at least on sublink of a CML.

Reply: Thanks for the suggestion, we rephrased the sentence: "*Data availability is defined for each CML as the fraction of the two-year observation period during which both TSL and RSL values are valid for at least one of its sub-links*."

Table 3 - For completeness consider also providing statistics for the rainfall variable in this table. A reader going through the paper just briefly might miss that weather stations are actually also rain gauges.

Reply: We revised table 5 (former table 3) as suggested.

Figure 7: For better comparison consider using the same range of y-axis in panel (b) and (c)

Reply: Figure 7 has been changed accordingly.

Figure 9 title: According to guidelines the figure title should be brief and does not need (should not) to describe what it is apparent from the legend and titles of axes.

Reply: Figure 9 has been replaced as part of the overall reframing of this section.

Section 3.3: PLease provide information on selection of parameters in the CML processing chain. Have you calibrated them, or used some values from literature?

Reply: We used values from literature. We added a description of the whole process at the beginning of Sec. 4 (former section 3) of the revised manuscript (second paragraph). "*To derive rainfall estimates from raw CML signal levels, the open-source pycomlink library was utilized (Chwala et al., 2025). Dry and wet periods were classified using a rolling standard deviation window of 60 min with a threshold of 0.8, following the methodology of Schleiss and Berne (2010). The baseline attenuation during dry weather conditions was determined using a constant baseline approach, which averages the signal from the five dry timesteps immediately preceding rainfall onset. To compensate for wet antenna attenuation (WAA), the empirical model formulation proposed by Pastorek et al. (2022) was implemented with pycomlink's default parameters. Finally, specific rain attenuation was converted to rainfall intensity using the power-law relationship described in Eq. 1*."

L215: provide reference to pycomlink

Reply: We have added the correct reference (Sec. 4, 2nd paragraph). Thank you for pointing this out.

Appendix A:

Appendix A is substantially more difficult to read than the main sections of the paper and it is unclear what exactly is being investigated ("this behavior"). Stating clearly the objective of this analysis at the beginning would help. Do you try to reconstruct the quantization of original raw data (you do not have access to)? Or do you try to quantify the quantization error of 1 min RSL data from this dataset? Why is TSL data quantization not considered?

Reply: We have revised Appendix A also according to the comments of Reviewer #1. The objective of the analysis conducted in the appendix is clarified in its first statement. We tried to reconstruct the in-house data processing conducted by the data owner over the raw RSL data from the patterns of the available RLS time series. As for the TSL, inspection of the time series shows that the TSL was processed as the RSL, therefore the same considerations hold. The difference between TSL and RSL is that the TSL is orders of magnitude higher than the RSL and that it is a stable signal, hence the observed time series of TSL are usually flat showing constant integer dBm values. In the revised appendix, we added a few statements about the TSL quantization. See also our reply to the previous comment of the reviewer on Section 2.2.

L256-L256,I believe ”0+k“ should be both lower index of RSL. This applies for the equation A1 and also the following description on L256. Also provide units.

Reply: These indices refer to the raw RSL data, which we assumed have a 1 dB quantization. Therefore all RSL raw samples are integer numbers (on a dBm scale). By consequence, writing RSL_{0+k} is equivalent to writing RSL_0 + k, when we assume that the RSL takes all the possible (integer) values from RSL_0 to RSL_{K-1} with k=0,1,...,K-1, as it is usually the case.

Equation A3: I am convinced that the units are wrong in this equation. As RSL is not unitless the equation has to result in adding different units no matter what units of N are.

Reply: This equation is a special case of (A2) when K=1. The derivation is in the following (we also provide also units between brackets):

$$RSL_N = \frac{1}{N}\sum_{k=0}^{1} N_k(RSL_0[dBm] + k[dB]) = \frac{1}{N}\left(N_0 RSL_0[dBm] + N_1 RSL_0[dBm] + N_1[dB]\right) =$$

$$\frac{1}{N}\left(N_0 RSL_0[dBm] + \left(N - N_0\right)RSL_0[dBm] + N[dB] - N_0[dB]\right) = \frac{1}{N}\left(NRSL_0[dBm] + N[dB] - N_0[dB]\right)$$

In the above equation, we added up dBm and dB, which is correct. We think that this was the equation implemented during in-house processing. We pointed out that this equation is conceptually not correct, as the N-sample average in dBm is not the average of the N-samples each expressed in dBm. However, from a dimensional point of view, the above equation works.

RC3: 'Comment on essd-2026-160', Anonymous Referee #3, 09 Jun 2026 Citation: **https://doi.org/10.5194/essd-2026-160-RC3**

In this manuscript, the authors present OpenRainER, an open dataset of continuous precipitation measurements across the Emilia-Romagna region from 2021 to 2022. The dataset integrates 1-minute raw transmitted and received signal levels from 151 commercial microwave links, alongside concurrent observations from 319 rain gauges and three weather radar products. Overall, the manuscript describes a dataset that is highly relevant and worth publishing. Because raw CML data is typically restricted by network operators, the authors' effort to publish this collection openly on Zenodo is a valuable effort. It provides the research community with a practical, much-needed resource to independently test and validate rainfall estimation algorithms. Still, the manuscript's current presentation requires significant improvement to fully realize the dataset's value and ensure its practical accessibility. Specifically, the authors must refine their data visualizations, clarify the mathematical notation in the appendix detailing the proprietary pre-processing steps, and address several compliance issues regarding ESSD formatting and visual accessibility guidelines prior to publication.

Reply: We thank the reviewer for recognizing the relevance of the proposed dataset. We have revised the manuscript to address the weaknesses highlighted in her/his general comments. Our point-by-point responses to the reviewer's comments are provided below.

Dataset and Usability:

The dataset successfully fulfills its primary objective of providing a highly useful, novel, and unique resource benchmarking precipitation retrievals within the Earth system sciences. It contains, in addition to the main focus (opportunistic sensors with CMLs), highly valuable reference data, and it seems to be reproducible and already integrated into the OpenSense repository tools. The time span (2 years), the resolution (1-min), and the number of CMLs provide a comprehensive benchmark, which is strengthened even further by the radar and gauge data, providing a complete benchmark for the opportunistic sensing community.

- Comprehensive Concurrent Data: The dataset provides an excellent, highly complete "trifecta" of concurrent measurements. It pairs 1-minute data from 151 CMLs with measurements from 319 operational rain gauges and three different radar products.
- High Data Availability: The dataset is rich and robust for a two-year continuous period. 139 of the 151 CMLs and 225 of the 319 rain gauges boast a data availability of 95% or higher
- The published data complies with the format of the specified OpenSense community standards, and already have been employed for applications.

Issues

- The authors state two main specific differences from other datasets, noting, (1) “The links are distributed over very different types of terrain,” which seems to be correct and is agreed upon as a valuable contribution. This is specifically even mentioned in the abstract. However, if there is value here, it is expected to be referred to more in the paper (besides a general representation of the metadata).

Reply: We agree with the reviewer that the original manuscript did not sufficiently emphasize the distinctive characteristics of our dataset compared with other publicly available CML datasets. In particular, our dataset includes CMLs deployed across highly heterogeneous terrain, ranging from flat lowlands to mountainous regions. To better highlight this key feature, we revised the Results section (Section 4, former section 3) by providing more explicit information on the geographical setting of the analyzed links. Specifically:

- As Figure 10 presents a CML deployed in a mountainous area, we added the altitude profile along the propagation path together with an orographic map of the surrounding region.
- We added Figure 11 and Figure 12 (as a replacement of figure 11), which report KPIs for wet/dry classification and of CML-based rainfall estimation, RESPECTIVELY, divided by terrain type (flat and mountainous).

These new figures are discussed in detail in the revised manuscript to better illustrate the impact of terrain characteristics on the dataset and the proposed methodology.

- For the CML network in Section 2 (Figures 1 and 2), the authors are only showing an aggregation/summary of results over all links. Although it is a data paper and does not require deep analysis, providing an example of the statistics and specific events can help the reader justify using this dataset and appreciate its uniqueness.

Reply: The purpose of Figures 1 and 2 is to show the metadata. Figure 1 shows the geographical location of all the sensors, while Figure 2 reports distributions of CML metadata (path length, antenna elevation, inclination angles of the path and terrain roughness along the path). Statistics of the KPIs for wet/dry classification and rainfall estimation based on CML data and divided by terrain type are shown in the new Figures 11 and 12. Finally, a specific event is illustrated in Figures 9 and 10 and commented on in the text. In conclusion, we think that the new format of Section 4 (former section 3) highlights the relevance of this dataset. On the other hand, we do not understand well what the reviewer means by “example of statistics.”

- The claim (2)“While conventional CML networks show a frequency distribution inversely proportional to the link path length, here links operate at 24.6 or 25.6 GHz. This offers a wider range of sensitivities for testing classification and retrieval algorithms.” I am not sure this is worth mentioning as main uniqueness with respect to the other published datasets and wireless networks in general. Usually, operators work in specific frequency ranges and from specific source CML’s data usually comes with few bands. Also, from a quick look at the other datasets, the frequency bands do contain the same scale of CMLs, with also diverse lengths (OpenMRG, OpenMesh), such that I would not mark it as a core value of the presented dataset, but rather as a feature.

Reply: we accept the suggestion of the reviewer and correspondingly rephrased the abstract.

If there are specific characteristics that are novel and can be derived with the data's specific setup/topology regarding the frequency vs. length, it should be emphasized in the analysis part. Still, I think the variety of the dataset as mentioned, in addition to the new region and the terrain (which should be exemplified), are enough to emphasize the value of the data.

Reply: we didnot conduct a specific analysis that emphasizes the frequency/path length characteristics of the dataset. On the other hand, we illustrated (Figure 8) that the dataset contains a variety of precipitation events including strong convective events and l persistent stratiform events (as defined in Sec. 4.1). Therefore, we decided to highlight in the revised abstract also this aspect together with the one relative to the terrain characteristics.

-The code reference for GitHub is only for the main library (https://github.com/OpenSenseAction), which is not specific to the OpenRainER dataset. It may contain a few examples, but I would suggest providing more focused notebooks to demonstrate the specific published data instead of the entire library. Specifically, such examples are already available at https://github.com/OpenSenseAction/opensense_example_data/tree/main/OpenRainER, which can be a solid base for data visualization, downloading, and usage.

Reply: Thank you for the suggestion. We have added the link to the example. However, we decided to keep the link to the OpenSense repository as well; although it is not specific to our dataset, it provides a comprehensive suite of highly useful tools and libraries for working with CML data.

Data Presentation

The data presentation and visualization visualizations do a great job showing the dataset's physical setup and data availability. The CML metadata histograms (Figure 2) clearly split plain vs. mountain conditions across path length, altitude, inclination, and roughness. Temporal coverage is also handled well, histograms for CMLs and rain gauges (Figures 3 & 6) and a monthly bar chart for weather radars (Figure 5) together give users a clear, honest picture of the network's reliability over the two years.

Issues:

- CML metadata presentation: The NetCDF variables are never listed anywhere in the paper. A simple table showing exactly what's included would save users from having to download the data just to find out. This is especially important for a data journal like ESSD.

Reply: we added a dedicated table in the revised manuscript. See Table 3 and the corresponding description in the first paragraph of Section 3.

- CML data visualization (Figure 2): The histograms show statistics about mountain links, but it can be more extended to describe the data. Figure 2 is too small to read comfortably and needs to be enlarged (see Reviewer 1 comments). Also, authors can show what those links look like physically, or by visual what (even I would suggest on adding a 2D elevation profile for one extreme mountain link would make the steep inclination). For Figure 3, the logarithmic y-axis is misleading, with only 151 CML links total, a linear scale would show the availability distribution much more honestly.

Reply: Figure 2 has been modified to improve readability. In the example of Figure 10, we have added the 2D elevation profile of a CML, as suggested (The inclination angle of this CML is 5.3°, that is, the 76.6% percentile of the distribution in Figure 2(c)). Regarding Figure 3, we believe that the log scale is better here than the linear scale because it allows the reader to quantify relatively small and large numbers on the same graph. Our purpose here is to highlight the number of CMLs with a low fraction of valid data.

- Radar Modifications: The manuscript only says the radar processing chain had "minor modifications" over two years, please spell out exactly what changed and how it might affect the data. A citation to the Arpae processing chain or a validation study would also help (See Reviewer 1).

Reply: The minor modifications concern the method used to remove unwanted signals from the radar data. To address the reviewers' comment, we expanded the sentence as follows: ''*Note that the radar processing chain was slightly modified over the two-year period covered by the dataset, specifically in the radar echo filtering process. As of 20 June 2022, a fuzzy logic-based method replaced the decision tree approach, resulting in a final product with fewer unwanted radar echoes and a better representation of rainfall structures.”* To date, there are no published studies available concerning the validation of this specific radar product or processing chain.

- Terminology & Rain Gauges: "Rain Gauges" and "Weather Stations" refer to the same 319 sites, please pick one term and use it consistently. Add the 15-minute precipitation availability stats to Table 3 so readers don't miss the dataset's core variable (See Reviewers 1 & 2). Also, it's unclear why 50% was chosen as the completeness threshold, is this a standard value in the field? Please justify or cite it. Finally, Figures 7 and 9 need some visual improvement (See Reviewer 2).

Reply: Thank you for bringing this to our attention. We have revised the text to use the term "rain gauges" only when we strictly refer to direct rainfall measurements. Furthermore, we have added precipitation statistics to table 5 (former table 3); to provide a clearer view of the data distribution, we included the 25th and 75th percentiles alongside the 50% completeness threshold. 25th, 50th, 75th percentile were used to give an idea of the data availability distributions.

Analysis:

The authors established a regional rainfall baseline using the reference RGs and weather radars, then evaluated the CMLs through a qualitative storm case study and a preliminary quantitative comparison against the adjusted radar data. This comparison demonstrated the CML data's potential to track rainfall, but also revealed significant underestimations and false alarms when the raw data is processed without complex, site-specific calibration. Overall, for a dataset paper, the analysis should not be too long, and the authors' approach of providing one high-level aggregated result and a specific CML case study is a good start. However, the current analysis does not add much new information compared to existing works. To truly demonstrate the value and uniqueness of this dataset, I strongly suggest adding an analysis that leverages the authors' claimed contributions, i.e., stratifying the performance by dataset features like complex terrain and varying link lengths operating at the same frequencies. Furthermore, since the dataset is already two years old and the `pycomlink` package is already being utilized, these tools can be directly applied without too much effort to demonstrate usability for real-world products. For example, generating 2D CML-derived rainfall maps to compare against the radar rainfall maps would be a standard addition that significantly raises the value of the data.

Reply: We have followed these suggestions to thoroughly revise the analysis section. Given the sparse nature of the CML network, we believe that a spatial map generated solely through CML interpolation would not be reliable.

3. Additional Comments/Written Quality

The manuscript is generally readable but would benefit from a careful revision pass — there are a moderate number of typos, grammar issues, and repeated words throughout. Previous reviews already flag several of them, and I strongly recommend a thorough proofread before the next submission.

Just from going through the abstract alone:

CML is never formally defined, on first mention write "commercial microwave link (CML) network."

L2: "on Zenodo repository" → "on the Zenodo repository" L3: "precipitation related" → "precipitation-related" L4: "measurement from" → "measurements from" L4: "consisting in 1 min time series" → "consisting of 1-min time series" L5: "radio 5 links" → "radio links" — stray line number L7: "well known" → "well-known" L7: "decrease of the received" → "decrease in the received" L8: "peculiar" → " sounds like something is strange or odd. Better to use "distinctive" or "unique characteristics. Point (2) here + ‘,’.

It is of course not the reviewer's responsibility to catch every issue, but even from a quick read there are noticeable word repetitions and some phrasing that could simply be cleaner. The other reviewers point out additional mistakes as well.

L106: "frequency are" → "frequencies are" L178: "between between" → "between" L196: "event occurred" → "event that occurred" L208: "explains" → "explain" L219: "Results shows" → "Results show" LXX: "Days are accounted if" → "Days are accounted for if"

As mentioned, I recommend the authors do a full end-to-end proofread, ideally with a native English speaker or grammar tool, before resubmission.

Reply: the document has been proofread with a grammar tool.