



Supplement of

Commercial microwave links as a tool for operational rainfall monitoring in Northern Italy

Giacomo Roversi et al.

Correspondence to: Federico Porcù (federico.porcu@unibo.it)

The copyright of individual parts of the supplement might differ from the CC BY 4.0 License.

1 Characteristics of the rejected data

The whole set of available CMLs is represented in Fig. S1a, according to the pairs of path length (x-axis) and frequency (yaxis). Theoretical sensitivity is calculated for every length-frequency pair through the inversion of the kR relationship at a fixed 5 1 mmh^{-1} rain rate. It has to be remembered here that the manipulations within the algorithm (especially the Aa threshold) do not allow a direct translation from the theoretical sensitivities to actual instrumental uncertainties or error bands. The theoretical sensitivity field is shown as contour lines of equal sensitivity with small differences for the two polarizations.



Figure S1. Dataset characteristics over the field of the theoretical sensitivity. Data in red are below the 0.1 dB per mmh⁻¹.

Table S1. Percentages of the data rejected after the different steps of the algorithm.

	Low Sens	NAs	Outliers	Valid data	Total
% of Total	4.13	10.9	3.87	81.1	100

Starting from an initial dataset of 357 duplex CMLs, 4.13 % of the data is rejected as it has a theoretical sensitivity lower than 0.1 dB per mmh⁻¹ (red selection in Fig. S1). This involves 15 CMLs (see Fig. S2a). Another 10.9 % is lost during the Preprocessing, mainly due to missing values ("NA" in the R language) alternatively on Pmin or Pmax. Some mitigation 10 technique was attempted, exploiting the remaining observation of the Pmin-Pmax pair, but did not lead to any consistent result. The median of the number of CMLs involved here is 37, with around 9 % variance between different days and an isolated peak on June 3 (see Fig. S2b). Both low sensitivity and NAs are issues intrinsic to the dataset we received, and their mitigation is out of our control, so we had no alternatives but to discard them all. Rejected percentages of the total available data are listed

15 in Tab. S1.



Figure S2. Top left: number of low sensitivity data rows. Top right: number of rows with NAs, Bottom Left: daily numbers of CML with outliers. Bottom right: hourly numbers of CML with outliers.



Figure S3. Distributions of the hardware characteristics of the CML reported as outliers compared with the ones of the whole dataset.

Lastly, 3.87 % of the data is classified as "outlier" from the outlier filter of the RAINLINK algorithm. The resulting valid dataset is 81.1 % of the initially available one. Outliers affect a median of 14 CML per time interval (15 min), but in this case, the corresponding variance between different days is a higher 27 % (see Fig. S2c and S2d). Outliers happen mostly on high-frequency links (> 35 GHz) and on the shortest lengths (< 4 km) with a secondary peak on the longest (> 17 km), as

shown in Fig. S3. No calibration of the -32.5 dBkm⁻¹h threshold was attempted. Please note that LC, BC, and the other 20 dataset descriptors of Tab. 1 of the main manuscript are all computed for the valid dataset, not on the total one.

2 Characteristics of precipitation in the study areas



Figure S4. Comparison of estimated and gauged PDF at 15 min single link (left) and hourly interpolated scales (right). Sampling is performed on 1 mm bins.

In the southern part of the Po Valley, where our target areas are located, precipitation is well distributed along the year, with two peaks in spring and autumn, and relatively dry summer. During spring, a transition season for the Mediterranean area, precipitation is often related to cyclonic development, typical of colder months, with large frontal bands and moderate rain 25 rates, but also small scale convection triggered in many cases by orography. The PDF of rain rates in the region during the two months of our research are reported in Fig. S4 (dashed lines) for the single 15 minutes raingauges (left) and the hourly ERG5 product at $5x5 \text{ km}^2$ resolution (right). In the same plots, are also reported the PDF of rain rates as computed by RAINLINK (solid line), 15 min, single link (left), and hourly interpolated on the same ERG5 5x5 km² grid (right).

- 30 The plots show that single-CML estimates are rather good for low to moderate rain rate (around 5 mm), while for higher, more discrepancies are found. Moving to the interpolated version, we can note that interpolation leads to a general underestimation widespread along with the whole rain rate range. A characterisation of the events that occurred in the two areas in the two months is shown in Fig. S5. An event is defined as any precipitation episode lasting for at least one hour (to grab also small scale thunderstorms) with at least two wet adjacent gridpoints. Two consecutive hours are needed to separate subsequent events. 35

Most of the events occupy a very small area (below 10%, i.e. around 300 km^2), and in general, the average coverage of the events during their lifetime is below 60 %. It has to be remarked that very likely the events are not entirely contained in the target areas during the whole lifetime so that we can have only a partial view of the events. The maximum rain rate for each event has two peaks, one for very small rain rate below 5 mmh⁻¹, probably due to stratiform precipitation, and a smaller one

around 13 mm, related to convective systems of moderate intensity. The duration of the events spans from one hour to one day, 40 while most of the events produce small rain amounts, below 5 mm on the areal average. This analysis shows that the typical size of the events is small and, with some significant exception, characterized by low to moderate rain rate. This is certainly a challenging situation to test RAINLINK algorithm, since low-moderate rain rate makes the wetting of the antennas more



Figure S5. Distributions of summary indicators for precipitation for May and June 2016 in Emilia-Romagna.

influential, and small scale precipitation areas can be poorly defined by sparse sensors, such as both CML and the raingauges used for verification.

3 The total accumulation of interpolated products

The cumulated rain over the entire period on every 5x5 km² grid box is analysed with maps and scatter plots in Fig. S6. Most of the points show a cumulated depth of around 100 mm over two months. Most of the CML product shows a 10-50 % underestimation, while two different branches gather the highest positive and negative discrepancies. When looking at the map, the points of the two branches are not randomly placed but are grouped in specific zones. It is reasonable to expect that the presence of the boundaries of the domain is probably affecting both the CML product skill and the reference reliability. Further studies could work on subsampled areas where raingauges are the most uniform and where CMLs are present not only inside but also outside the interpolation region.

4 Calibration of Aa and alpha

55 We performed some sensitivity analysis for *Aa* and *alpha*, as recommended by Overeem et al. (2016). However, it is our feeling that the reference data we had available (which are used daily in operational offices) are not ideal to be used as a calibrator, in terms of quality and spatial and temporal characteristics. For this reason, we eventually chose not to use the calibrated parameters inside the primary validation process. We checked for the 27 links with a close-by 15 min raingauge in



Figure S6. Top: total accumulated rain depth over May and June 2016. Bottom: branches of high discrepancy are highlighted on the map.

which Aa and alpha values were producing the best overall performances, in terms of Pearson's Correlation Coefficient (CC) and normalized coefficient of variation (CV). The results are summarized in the following Fig. S7.

The CC surface (left) shows a clear maximum at alpha = 0.3, Aa = 0.7, while CV (right) reaches no local minimum in the examined domain, but has a plateau-like area of good performance, in which falls the best match for CC. This analysis suggests an Aa value much smaller than the one chosen by Overeem et al. (2016) (and kept for our work), which was 2.3 dB, while the alpha parameter remains almost unchanged from the default value of 0.33. However, looking at the PDF of the estimated rainfall with the two different Aa, (see Fig. S8), emerges that the physical representativeness drops down with the lower Aa.

 65 rainfall with the two different Aa, (see Fig. S8), emerges that the physical representativeness drops down with the lower Aa. Applying the calibrated parameters on the whole dataset also leads to worse overall results: CV worsens of +0.05 with no sensible improvement on R2, while the same happens for FAR and ETS, which change of +0.08 and +0.01 respectively.



Figure S7. Sensitivity analyses to two coupled retrieval parameters of the RAINLINK algorithm. Loss functions CC and CV.



Figure S8. PDF of the single-CML estimates against the raingauges for two different values of Aa.

References

Overeem, A., Leijnse, H., and Uijlenhoet, R.: Retrieval algorithm for rainfall mapping from microwave links in a cellular communication
network, Atmospheric Measurement Techniques, 9, 2425–2444, https://doi.org/10.5194/amt-9-2425-2016, 2016.