https://doi.org/10.5194/essd-2026-160
Preprint. Discussion started: 12 May 2026

Open Access Earth System Science Data Discussions

OpenRainER: an open-source dataset for studying the opportunistic sensing of rainfall in Emilia-Romagna, Italy

Elia Covi[1], Giacomo Roversi[2,3], and Roberto Nebuloni[4]

[1]Servizio Idro Meteo Clima (SIMC), Arpae, Bologna, Italy
[2]Department of Environmental Sciences, Informatics and Statistics (DAIS), Ca' Foscari University, Venice, Italy
[3]Institute for Atmospheric Sciences and Climate (ISAC), Italian National Research Council (CNR), Rome, Italy
[4]IEIIT, Italian National Research Council (CNR), Milan, Italy

Correspondence: Elia Covi (ecovi@arpae.it)

Abstract. We present the OpenRainER dataset of precipitation measurements, available open-source on Zenodo repository at https://doi.org/10.5281/zenodo.10593848. The dataset contains mainly precipitation related measurements over the region of Emilia-Romagna in northern Italy. Inside OpenRainER, measurement from the commercial microwave link network managed by Lepida S.c.p.A. are published, consisting in 1 min time series of the transmitted signal level and of the received signal level over 151 radio links. These data are primarily generated for link quality monitoring; however, they can be opportunistically exploited for weather monitoring as there is a well known direct relationship between rainfall intensity and decrease of the received signal level. The data are stored in NetCDF format and can be processed and converted into rainfall intensity time series along each radio link path using open-source tools developed within the COST Action OpenSense framework. We also provide concurrent data from the regional operational rain gauge network and two weather radars, as a reference for calibration and validation purposes. OpenRainER has peculiar characteristics with respect to similar open datasets: (1.) The links are distributed over very different types of terrain, including plains, hills, valleys, and mountain ridges (up to 2000 m a.s.l.), and both densely inhabited cities and rural areas. (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.

1 Introduction

The radio links used for wireless backhauling in mobile networks, hereinafter named Commercial Microwave Links (CMLs), can be exploited as opportunistic sensors of precipitation, by processing the data generated by the network management system for link quality control (Uijlenhoet et al., 2017). Opportunistic sensors may complement conventional sensors such as rain gauges and weather radars (Fencl et al., 2017; Bianchi et al., 2013), or they could be used in place of conventional meteorological networks, when the latter are scarce or entirely absent (David et al., 2021).

The physical principle behind the usage of radio waves to measure precipitation is summarized by a well-known relationship between the attenuation of a microwave signal traveling through a rainy layer and the rainfall intensity $R\,\mathrm{mm\,h^{-1}}$ (International

https://doi.org/10.5194/essd-2026-160
Preprint. Discussion started: 12 May 2026

Telecommunication Union, 2005):

$$R = a\gamma^b \tag{1}$$

where γ is rain attenuation per unit path length in units of $\mathrm{dB\,km^{-1}}$, while a and b are coefficients dependent on the frequency and on the polarization of the signal.

Access to data collected by CMLs is often limited by restrictive policies applied by mobile network operators, who are the owners of the network infrastructure. This limitation has fueled a growing interest in open-source CML databases within the research community. To the authors' knowledge, the largest dataset publicly available is one covering the entire Netherlands, which has been recently extended with four-year CML data (Overeem et al., 2024). It includes more than 1800 links with a temporal resolution of 15 min. The OpenMRG dataset instead includes 364 bi-directional CMLs located in the city area of Gothenburg, Sweden (Andersson et al., 2022), available during a three-month period and with a 10 s resolution. Concurrent rain gauge measurements and a radar composite are provided as well. The OpenMesh dataset features eight-month measurements at a 1 min resolution, collected by 103 CMLs from a community-run communication network in New York City (Jacoby et al., 2025). Another notable example of open CML dataset is OpenRainER (Covi and Roversi, 2024), which is presented in this paper.

OpenRainER stores the data collected by a CML network owned by the company Lepida S.c.p.A.. The network covers the Emilia-Romagna region (Italy). Since 2016, Lepida is collaborating with the Hydro-Meteo-Climate structure of the regional agency for environmental protection and energy of the Emilia-Romagna region (Arpae-SIMC) and with the university of Bologna in several activities. An important study was the RAINBO research project (2016-2019), in the frame of the EU LIFE Programme, aimed at testing the performance of CML rainfall retrieval over the region (Roversi et al., 2020). Since June 2020, Lepida's CML data are available to Arpae-SIMC in near-real time every 15 minutes. In 2022, Arpae-SIMC started to evaluate the operational use of CMLs in its weather service as part of the activities of the MODMET project, funded by the Department of Civil Protection of Italy. Arpae-SIMC is also involved in the CLIMAXPO project, where the fusion between CMLs and conventional meteorological sensors is investigated to enhance the detection of extreme weather events with the ultimate goal of improving adaptation strategies to climate change.

OpenRainER offers some unique features with respect to the other open datasets. Firstly, it provides an unprecedented heterogeneity and complexity of terrain and of precipitation regimes. Indeed, the Lepida network is mainly deployed in mountainous areas, where links travel over terrain at varying elevations and several link paths are slanted with respect to the horizontal direction. Secondly, all the CMLs operate at only two very similar frequencies (24.6 GHz and 25.6 GHz), which makes the compensation of frequency-dependent unwanted effects (e.g. wet antenna attenuation) more straightforward. Also, at these frequencies, the $R-\gamma$ relationship is not very sensitive to the raindrop size distribution (Chwala and Kunstmann, 2019), which reduces errors involved in the application of Eq. 1 in different precipitation regimes. On the other hand, there is no inverse relation between frequency and link length here, as seen instead usually in CML networks. This produces a wide range of sensitivities to rainfall, which can be a challenge for retrieval algorithms, hence providing an optimal testbed for e.g. rain/no-rain classification strategies.

https://doi.org/10.5194/essd-2026-160
Preprint. Discussion started: 12 May 2026

Open Access Earth System Science Data Discussions

Beside CML measurements, OpenRainER contains three different weather radar products, and also rainfall data collected by the operational network of the rain gauges used inside Arpae-SIMC. The dataset can then be integrated with other data publicly available over the region, e.g. the complete set of environmental and atmospheric variables of the ERG5 analysis from Arpae-SIMC (Antolini et al., 2015) (https://dati.arpae.it/dataset/erg5-interpolazione-su-griglia-di-dati-meteo), or data from a laser disdrometer located inside the campus of the University of Bologna (Adirosi et al., 2023) (https://doi.org/10.5281/zenodo.7708563) or the personal weather stations (PWS) from the Meteonetwork (https://www.meteonetwork.it/) association. The OpenRainER dataset is published under a Creative Commons license (CC-BY 4.0).

The data format follows the guidelines elaborated by the OpenSense (Opportunistic Precipitation Sensing Network) community for the standards and formats of opportunistic precipitation data (Fencl et al., 2023). OpenSense has been a COST Action project (https://opensenseaction.eu/), ended in October 2025, that brought together scientists, experts from national meteorological services, sensor network owners, and end-users of rainfall data, to advance the field of precipitation monitoring by opportunistic sensors. The scope of this paper is to present OpenRainER, improve the data usability, and provide important details that can help in interpreting and processing the data.

The remainder of the paper is structured as follows. Section 2 describes the characteristics of the study area, and the technical details of the CML data, the radar data and the rain gauge data, including data availability. Section 3 provides information about the rainfall features over the study-area in the two-year observation period and draws some preliminary comparison between the data collected by different sensors. Finally, Section 4 draws the conclusions.

2 DATASET

The OpenRainER dataset stores two consecutive years of continuous CML measurements (2021 and 2022), across the Italian region of Emilia-Romagna (Covi and Roversi, 2024). The CML network is composed of 151 links owned and managed by Lepida S.c.p.A.. The dataset also includes concurrent observations from operational sensors used inside the meteorological structure Arpae-SIMC, specifically two weather radars and 319 rain gauges belonging to the Hydro-Meteorological-Pluviometric Monitoring Network (RIRER). OpenRainER includes five data types (i.e. CML data, rain gauge data and three different radar products), stored as a collection of monthly NetCDF files. Table 1 summarizes the published datasets and the file naming convention used. NetCDF is a portable and scalable binary format suitable to store multi-dimensional data. Several scientific computing frameworks (e.g., Python, MATLAB, R) include toolboxes or libraries for reading and handling NetCDF files. Open-source Python libraries, developed within the OpenSense project and available at https://github.com/OpenSenseAction, offer free tools to read, process, and compare CML, radar and rain gauge data.

2.1 STUDY AREA

The Emilia-Romagna region is located in northern Italy, between latitudes 43.73 and 45.14 and between longitudes 9.20 and 12.76, respectively. The northern part of the region consists of a vast alluvial plain ("Pianura Padana", Padan Plain or Po Valley) enclosed between the Po river on its northern border and the line drawn by the ancient Roman road named "Via Aemilia", where

https://doi.org/10.5194/essd-2026-160
Preprint. Discussion started: 12 May 2026

Table 1. Overview of the OpenRainER dataset.

Source	Sensors	Archive	Size	Format	Description
Emilia-Romagna Radar Composite	2	`RADref.tar`	1.6 GB	NetCDF	5 min reflectivity maps
		`RADrain.tar`	1.3 GB	NetCDF	15 min accumulated rainfall maps
		`RADadj.tar`	1.6 GB	NetCDF	15 min rainfall maps adjusted by rain gauges
Lepida S.c.p.A. CML network	151	`CML.tar`	107.2 MB	NetCDF	RSL and TSL raw data with CML metadata
RIRER rain gauge network	319	`AWS.tar`	22.2 MB	NetCDF	15 min accumulated rain, temperature, wind (speed and direction), relative humidity, and metadata

all the major cities are located. The southern part of the region is occupied by the Apennine mountain range, where the terrain altitude steadily grows from the plain to the ridge, which constitutes the southern border. On the east side, the region confines with the Adriatic Sea. Altitudes are in the order of few tens of meters above sea level in the plains, with some areas near the coast even below the sea level, while the main peaks on the southern edge reach 2000 m. From an hydrological point of view, the western part of the territory lies almost entirely inside Po basin, with the northward flowing rivers being the last tributaries of it, while the rivers in the eastern half of the region flow directly to the sea. Most of the Po Plain has been affected by extensive land reclamations, while the Apennines area is prone to erosion and landslides.

For the purposes of this work, the region has been subdivided into two zones based on elevation: plain areas, defined as those below 100 m a.s.l., and mountain areas, encompassing all higher elevations. The mountainous zone accounts for approximately 50% of the total area, with an average elevation of 561 m. A Digital Terrain Map (DTM) of the region (Emilia-Romagna, 2016), resampled at 50 m horizontal resolution, is depicted in Fig. 1. The figure also shows the position of the CMLs (red lines) and of the rain gauges (circles).

2.2 COMMERCIAL MICROWAVE LINK

The CML data in OpenRainER consist of TSL and RSL time series at 1 min resolution alongside the corresponding time stamps, in UTC coordinates. The measured signal level is the power expressed in dBm. An overall 151 CMLs are available. As CMLs provide two-way radio communication, each link includes two sub-links, that is, two communication channels over the same physical path though in opposite directions. Most of the CMLs in OpenRainER have data available from both sub-links. The sub-link carrier frequency are 24.6 GHz and 25.6 GHz respectively. Time series at sub-link level provide redundancy and enable cross-checking individual link data. Finally, 82 out of 151 CMLs use vertical polarization, the other 69 use horizontal polarization. Figure 2 shows the histograms of the geometrical features of the CML network. CMLs are grouped according to the terrain type described in Sect. 2.1: orange bars represent CMLs with at least one antenna located in the mountainous area (n = 124), while blue bars indicate the remaining CMLs associated with the plain area (n = 27). The CMLs have path lengths ranging from a few hundred meters to 25 km, with a mean length of 7.4 km (panel (a)). Figure 2(b) displays the altitude of each CML terminal. Links deployed over mountainous terrain may have antennas positioned at markedly different

https://doi.org/10.5194/essd-2026-160
Preprint. Discussion started: 12 May 2026

Open Access Earth System Science Data Discussions

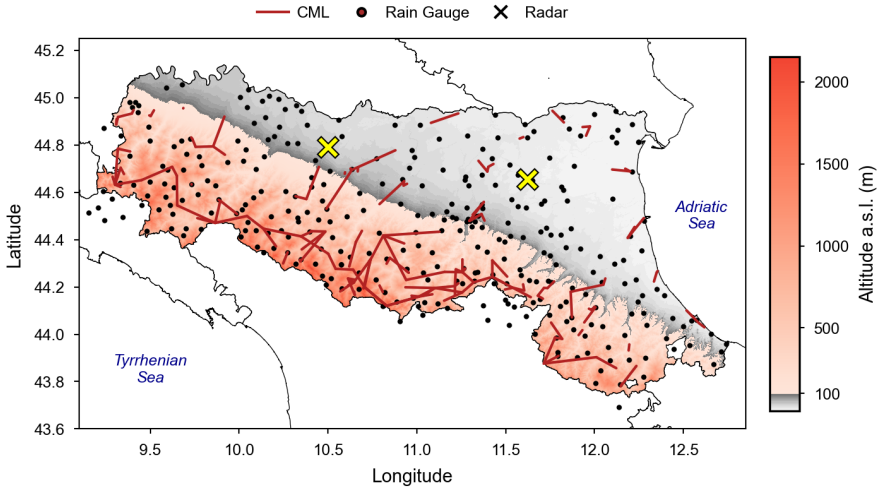


Figure 1. Spatial distribution of CMLs, rain gauges, and weather radars from the OpenRainER dataset across the Emilia-Romagna region. Red-shaded areas indicate elevations higher than 100 m (southern part of the region), while gray areas represent lower-lying terrain. DTM: (Emilia-Romagna, 2016); regional borders: (Istat, 2024).

altitudes, resulting in path inclination angles up to 20.4° (panel (c)). Consequently, the signals may propagate across different atmospheric layers, encountering different types and phases of precipitation. Furthermore, due to the complex orography of the region, the distance between the line-of-sight propagation path and the terrain is often non-uniform. To quantify this feature, we introduce the concept of terrain *roughness*, defined as the standard deviation of the altitude along the propagation path. The histogram of roughness is shown in Fig. 2(d). Some CMLs exhibit roughness values as high as $400\,\mathrm{m}$. This geometric characteristic complicates CML measurements, as it can enhance unwanted multipath effects as well as hinder the association of the measurement with a specific altitude.

Figure 3 shows a histogram of the availability of CML data. Data availability (x-axis) is defined as the fraction of the two-year observation period during which both TSL and RSL values are valid for at least one sub-link. The analysis shows that 139 out of 151 CMLs have a data availability equal to or higher than 95%. The time series of the TSL and of the RSL were preprocessed in-house by Lepida before being made available to Arpae-SIMC. Unfortunately, the details of the pre-processing procedure were not disclosed. The Appendix of this paper reports the outcomes of an analysis over CML raw data in dry conditions, that is, when the effects of the atmosphere on signal variations are at a minimum. This analysis unveils some interesting characteristics of the signal.

2.3 WEATHER RADAR

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

https://doi.org/10.5194/essd-2026-160
Preprint. Discussion started: 12 May 2026

Open Access Earth System Science Data Discussions

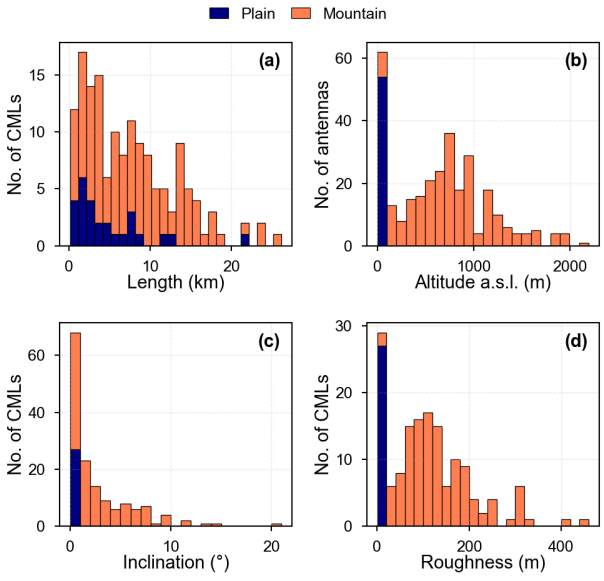


Figure 2. CML geometrical features. (a): Histogram of path lengths; (b): Histogram of antenna elevations; (c): Histogram of the inclination angles of the line-of-sight propagation paths; (d): Histogram of the terrain roughness. Blue (orange) bars indicate CMLs with both antennas positioned at an altitude below (above) 100 m.

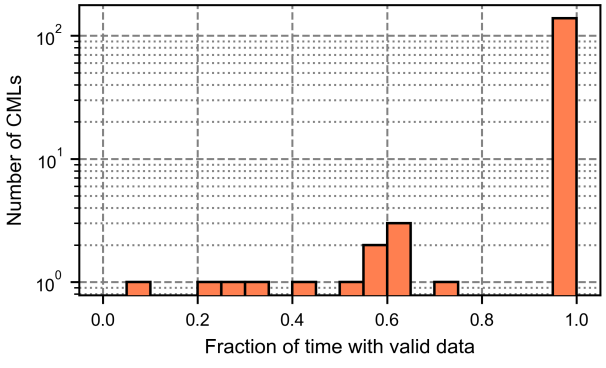


Figure 3. Histogram of CML data availability. Total number of CMLs: 151.

https://doi.org/10.5194/essd-2026-160
Preprint. Discussion started: 12 May 2026

Open Access Earth System Science Data Discussions

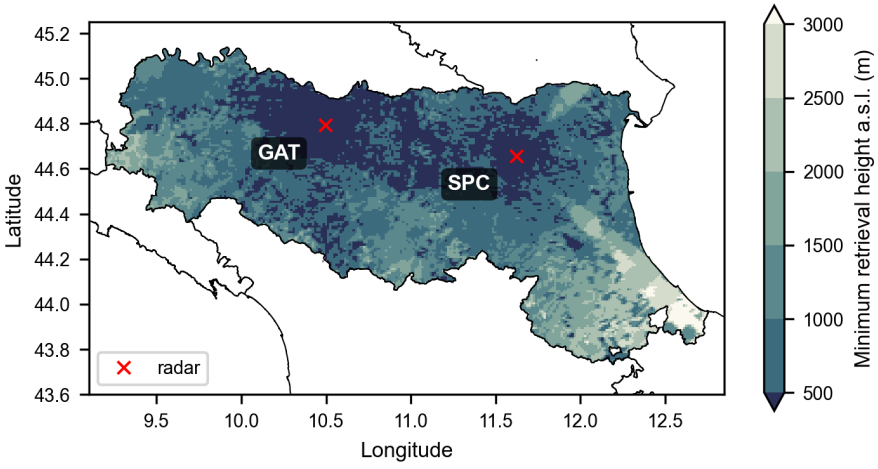


Figure 4. Minimum height above ground of the radar beam that can be used for retrieving precipitation. The red crosses mark the location of the two radars. Regional borders: (Istat, 2024).

.

one is near San Pietro Capofiume (SPC), in the eastern sector of the region. The radar characteristics are detailed in Table 2. The composite of the two radars covers the whole region, but both radars are located in the flat portion of the region. This complicates the reconstruction of the precipitation field over the Apennines, where beam blockage and ground clutter can significantly affect radar echoes. To minimize echo contamination, radar scans must therefore be performed at relatively high elevation angles. Figure 4 illustrates an example of the minimum height above ground level of the radar beam used to reconstruct the precipitation field. As expected, the minimum height markedly increases toward the southern part of the region. In addition, strategies of obstacle avoidance in the vicinity of the radar end up in radial sectors with even higher heights (e.g. the two angular sectors in light colors North-East and South-East of SPC). Since retrieving surface precipitation from aloft measurements is inherently challenging Joss and Waldvogel (1988), quantitative precipitation estimates derived from radar products are less reliable in the mountainous areas of the region. Figure 5 shows radar data availability during the observation period: measurements from the SPC radar are missing between April and July 2021. As reflectivity maps can be generated from one radar or from the composite of the two radars, a lower degree of precision is expected in the eastern part of the region (i.e. the one covered by SPC) during this period.

Three different radar products are included in OpenRainER: 5 min maps of the ground reflectivity field, 15 min accumulated precipitation field, and 15 min accumulated rainfall adjusted by rain gauge measurements. The three radar products are provided on a lat-lon coordinate grid (EPSG:4326 reference system), with a resolution of 0.01265 x 0.009 degrees ($\sim 1\,\mathrm{km}^2$). The radar grid covers a rectangular domain between latitudes [43.4, 46.0] and longitudes [8.5, 13.2]. The radar data processing chain consists of:

https://doi.org/10.5194/essd-2026-160
Preprint. Discussion started: 12 May 2026

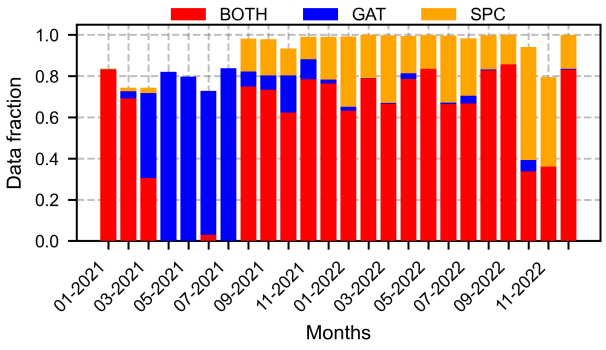


Figure 5. Availability of radar data, expressed as fraction of valid data on a monthly basis. Three different bar colors are used: blue and orange bars refer to the data collected from GAT and SPC respectively, whereas red bars refer to the composite maps.

Table 2. Characteristics of the two weather radar operated by Arpae-SIMC.

Radar Location	Gattatico (GAT)	San Pietro Capofiume (SPC)
Coordinates	44.790 10.498	44.655 11.623
Elevation [masl]	35	8
Frequency [GHz]	5.602	5.620
Antenna Height [m]	25	23
Radial Resolution [m]	250	
Beamwidth [deg]	0.9	
Model	GPM500	

1. an initial cleaning process including static and dynamic radar clutter identification, beam blockage correction and anomalous propagation removal (Fornasiero et al., 2006; Alberoni et al., 2001; Bech et al., 2003);

2. extrapolation of reflectivity profiles at ground level (Germann and Joss, 2002);

3. convective/stratiform classification and reflectivity-to-precipitation intensity conversion (Steiner et al., 1995; Franco et al., 2005).

Radar reflectivity values for rain estimation are capped at 50 dBZ to avoid overestimation due to hail. Rainfall intensity is calculated every 5 min, while 15 min accumulation is obtained as the weighted average of the fields reconstructed every 1 min (i.e. the duration of a full radar scan) via an advection algorithm that takes into account the motion of precipitating systems (Dotzek et al., 2002). The radar precipitation product adjusted with rain gauges is obtained by multiplying the original data by a spatial correction mask (Ochoa-Rodríguez et al., 2019). It should be noted that the radar processing chain underwent minor modifications over the two-year period covered by the dataset.

https://doi.org/10.5194/essd-2026-160
Preprint. Discussion started: 12 May 2026

Open Access Earth System Science Data Discussions

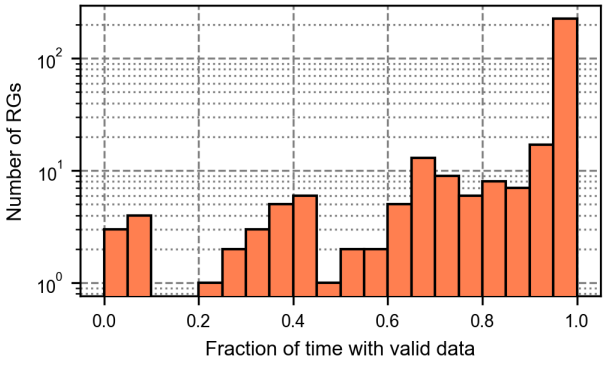


Figure 6. Histogram of rain gauge data availability. Total number of rain gauges: 319.

Table 3. Availability of the data collected by weather stations at ground in 2021-22.

Observed variable	No. of stations	No. of stations with at least 50% of data
temperature	202	27
wind direction	37	9
wind velocity	37	9
relative humidity	92	20

2.4 WEATHER STATION

OpenRainER provides data from the ground network of 319 regional hydro-meteorological weather stations managed by Arpae-SIMC. The data include measurements of 15 min accumulated precipitation, temperature, relative humidity, wind speed and wind direction. The 15 min accumulated precipitation from Arpae-SIMC station is an operational product, which represents the most accurate ground-truth reference against which CML-based rainfall retrievals can be compared in our case.

The ground observations in OpenRainER are part of the open-source data of Arpae-SIMC (ARPAE Emilia-Romagna, 2023). All Arpae-SIMC Open Data are uploaded after quality control checks, based on the guidelines defined by the National System for Environmental Protection (SNPA) (Barbero et al., 2017). The locations of the weather stations are shown in Fig. 1: 127 stations are located in the plains (altitude ≤ 100 m), the remaining 192 are in the mountainous areas.

Figure 6 shows the data availability of the 15 min accumulated rainfall measurements collected by rain gauges during the two-year observation period. The analysis indicates that 225 out of 319 stations exhibit a data availability $\geq 95\%$. Ancillary meteorological variables show instead a markedly lower availability, as not all weather stations are equipped with the same set of sensors. Table 3 provides an overview of data availability for the different variables, reporting the number of stations with at least one observation and the number of stations with more than 50% data completeness, assuming a nominal sampling frequency of 15 min (although some stations may operate at a lower temporal resolution).

https://doi.org/10.5194/essd-2026-160
Preprint. Discussion started: 12 May 2026

Open Access Earth System Science Data Discussions

3 RESULTS

To facilitate the use of the dataset, this section provides the following information: an overview of the characteristics of precipitation (type and intensity) within the Emilia-Romagna region as observed by rain gauges and radars in 2021-22 (Sect. 3.1); a qualitative comparison between the time series of CML raw data and the ones of rainfall accumulation produced by rain gauges and radars (Sect. 3.2) and a preliminary comparison between between CML rainfall estimates and adjusted radar (Sect. 3.3).

3.1 Rainfall characteristics

To highlight the different precipitation regimes occurring in the region, two classes of rainfall events are defined:

1. *strong convective events*: cases in which a rain gauge or a $1\,\mathrm{km}^2$ radar pixel record an accumulated precipitation higher than $15\,\mathrm{mm}$ in the lapse of $15\,\mathrm{min}$;

2. *persistent stratiform events*: cases in which a rain gauge or a $1\,\mathrm{km}^2$ radar pixel record a daily accumulated precipitation higher than $80\,\mathrm{mm}$.

The above thresholds are operationally used by Arpae-SIMC to identify significant precipitation events, such as thunderstorms or prolonged rainfall episodes. Figure 7 (a) shows the monthly average rainfall accumulation measured by the rain gauges. Figure 7 (b) and (c) report the occurrence of convective and stratiform events from January 2021 to December 2022. Days are accounted if at least one rain gauge detects one of the two aforementioned event types. Strong convective events occur almost exclusively during the summer months and were more frequent in 2022 than in 2021. In contrast, persistent stratiform events are considerably more frequent during the winter months, reflecting the typical precipitation regime of a mid-latitude climate. A similar analysis was also carried out in the spatial domain, using the information extracted from the adjusted radar product, as shown in Fig. 8. The results indicate that significant convective events affected different areas of the region over the two-year observation period. In contrast, persistent stratiform events were mainly concentrated across the mountain area of the Apennines in the southern part of the region.

3.2 Qualitative CML product comparison

Figure 9 presents an example of concurrent measurements from CML, radar, and a rain gauge sensor for a convective event occurred on 6 June 2021. The blue line with circle markers is the $1\,\mathrm{min}$ total path attenuation (i.e. the difference between TSL and RSL) measured by a $2\,\mathrm{km}$ CML operating at $24.6\,\mathrm{GHz}$ with horizontal polarization. The yellow line with cross markers is the $15\,\mathrm{min}$ accumulated precipitation calculated averaging all the radar pixels intersected by the CML. Finally, the red line with triangle markers is the $15\,\mathrm{min}$ accumulated precipitation measured by the closest available rain gauge (located at a distance of $1.8\,\mathrm{km}$ from the CML). The CML peak appears shifted relative to the maxima observed by the rain gauge and the radar, likely due to the different temporal resolutions of the measurements ($1\,\mathrm{min}$ versus $15\,\mathrm{min}$) and to the fact that the rain gauge and radar time stamps correspond to the end of each $15\,\mathrm{min}$ accumulation interval. The width of the rainfall peak is different in the three cases, which reflect the different spatial sampling performed by rain gauges, CML and radars (i.e. point, line, and volume

https://doi.org/10.5194/essd-2026-160
Preprint. Discussion started: 12 May 2026

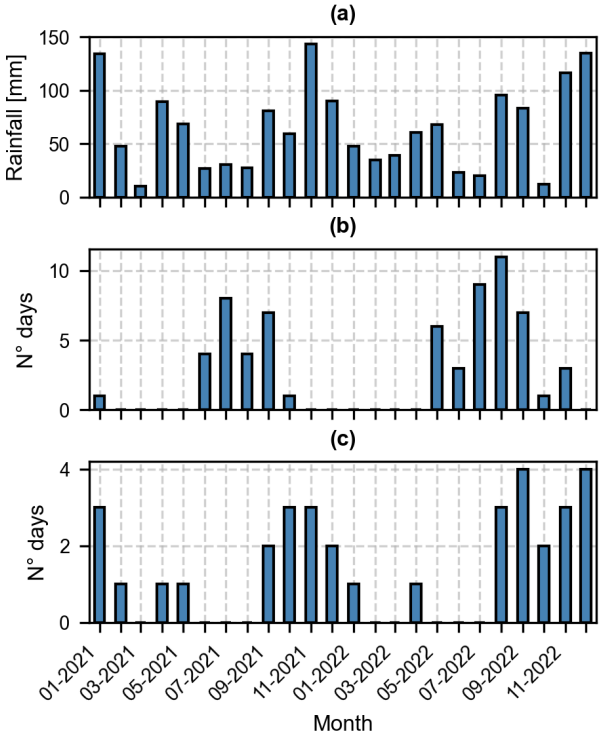


Figure 7. (a): Monthly average accumulated precipitation within the Emilia-Romagna region as measured by a network of 319 rain gauges; (b): number of days with heavy convective events (c) number of days with persistent stratiform events. Observation period: 2021-22.

sampling). Figure 10 complements Fig. 9 by illustrating the 2D radar-derived rainfall field over the CML–rain gauge area at three time steps centered on the passage of the precipitation cell. The selected case features intense and localized precipitation developing in the southern part of the region, within the Apennine area. The rainfall cell passes directly over the rain gauge and it moves very fast beyond it. The spatiotemporal characteristics of the cell as well as the different spatial and temporal sampling of the sensors explains why the maximum accumulated precipitation by the CML and by the rain gauge are similar. The CML is not touched by the core of the cell, however it captures the temporal dynamics of rainfall better than the rain gauge.

3.3 Rainfall comparison

A preliminary assessment of CML-derived 15 min accumulated rainfall estimates is conducted by comparing them with other sensors. Specifically, gauge-adjusted radar data is used as a reference to to include both radar and rain gauge networks. To address the different spatial geometries, CML estimate were compared with the average rainfall from the radar cells intersecting the corresponding link path. CML rainfall estimates were retrieved using the pycomlink library through the following steps: (1) total attenuation processing; (2) rainfall detection via rolling standard deviation (Schleiss and Berne, 2010); (3) constant baseline determination; (4) wet antenna attenuation correction (Pastorek et al., 2022); and (5) derivation of specific attenuation

https://doi.org/10.5194/essd-2026-160
Preprint. Discussion started: 12 May 2026

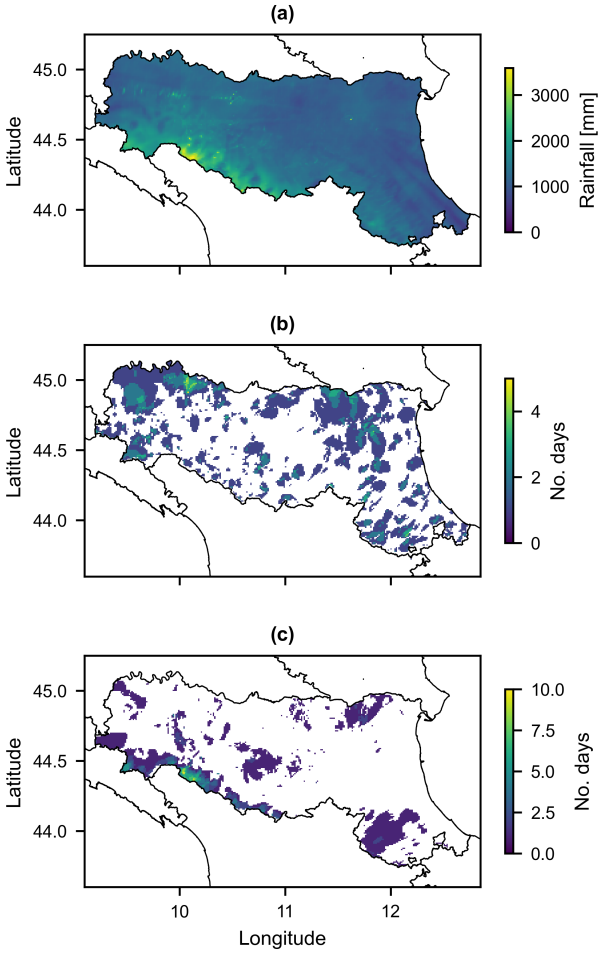


Figure 8. (a): Total accumulated precipitation in Emilia-Romagna as measured by the two regional weather radars; (b): number of days with heavy convective events (c) number of days with persistent stratiform events. Analysis based on adjusted radar field.

https://doi.org/10.5194/essd-2026-160
Preprint. Discussion started: 12 May 2026

Open Access Earth System Science Data Discussions

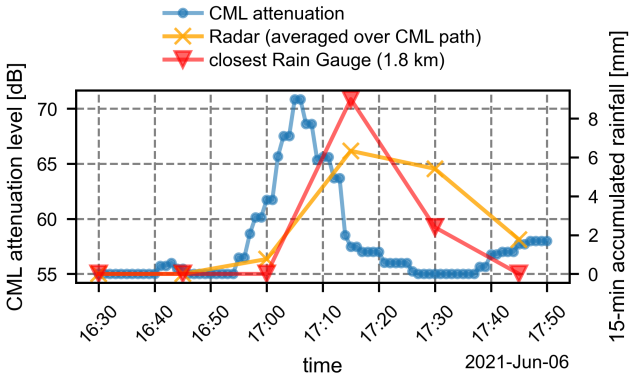


Figure 9. Comparison of rainfall-related measurements derived from CML signal attenuation (TSL–RSL), weather radar estimates, and rain gauge observations (ID: Madonna dei Fornelli_1125681_4421571), together with the attenuation of the CML link (ID: 307). CML signal attenuation is shown by the blue line with circle markers and refers to the left y-axis. Radar estimates and rain gauge measurements are represented by the yellow line with cross markers and the red line with triangle markers, respectively, and refer to the right y-axis.

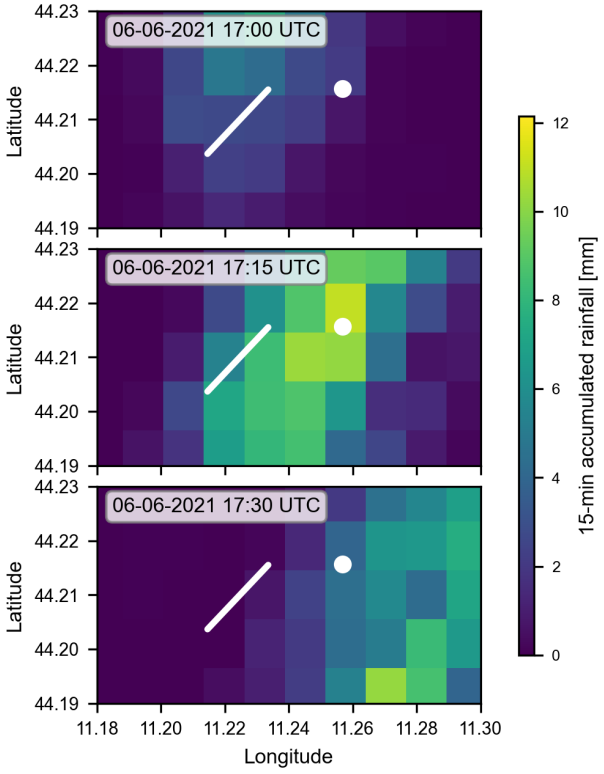
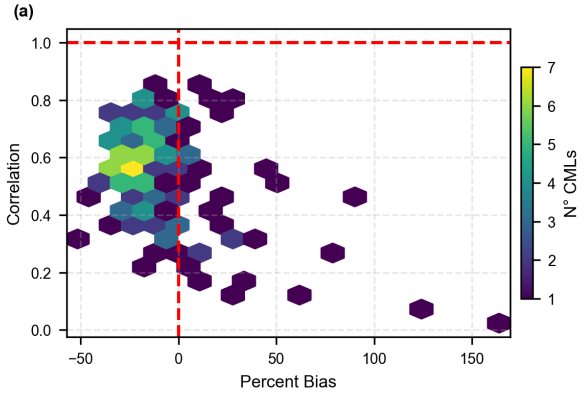


Figure 10. Radar-derived rainfall maps at three consecutive instants during the event of previous Fig. 9. The white line is the position of the CML path (ID: *307*), while the white circle is the position of the rain gauge (ID: *Madonna dei Fornelli_1125681_4421571*).

https://doi.org/10.5194/essd-2026-160
Preprint. Discussion started: 12 May 2026

Open Access Earth System Science Data Discussions



(b)

	Adj Radar: Yes	Adj Radar: No
CML: Yes	88917 (0.9%)	87025 (0.9%)
CML: No	134667 (1.4%)	9603705 (96.9%)

Figure 11. Comparison between CML rainfall estimates and rain gauge-adjusted radar data: (a) Percent bias versus Pearson correlation coefficient; (b) contingency table of detection performance.

and conversion to rainfall intensity. Figure 11 presents the comparative results. The metrics in panel (a) were calculated exclusively for samples where both gauge-adjusted radar and CML estimates recorded an accumulated precipitation ≥ 0.2 mm, corresponding to the 'True Hits' defined in the contingency table in panel (b). Results shows that 76% of 147 analyzed CMLs underestimate precipitation with respect to the adjusted radar product and 39% shows correlations higher than 0.6. Based on the contingency table in panel (b), the Probability of Detection (POD) is estimated at 0.4, with a False Alarm Ratio (FAR) of 0.5. The observed performance issues are primarily attributed to the lack of site-specific calibration for the retrieval algorithms, especially regarding rain detection. However, the presence of strong correlations even in this initial assessment confirms the promising potential of the CML network.

4 Conclusions

The two-year OpenRainER dataset presented in this work includes TSL and RSL data (power levels in dBm) collected from a network of 151 CMLs located in the Emilia-Romagna region (Northern Italy), which can be used to retrieve precipitation intensity. In addition, the dataset stores concurrent rainfall observations from the regional rain gauge network, three weather radar products, and meteorological observations by ground stations. The availability of different types of precipitation data enables

https://doi.org/10.5194/essd-2026-160
Preprint. Discussion started: 12 May 2026

the evaluation of different processing methods for deriving rainfall from CML data. The Emilia-Romagna region represents a particularly challenging and informative test case, as its mountainous areas are characterized by complex orography. In this context, the reconstruction of near-surface precipitation fields is especially demanding, since complex terrain reduces the representativeness of rain gauge measurements and partially obstructs the radar beam. As shown by the analysis of precipitation fields, the mountainous area is characterized by large accumulations; therefore, the presence of CMLs in this area can help assess their actual impact on precipitation monitoring. In this context, the dataset allows for the assessment of spatialization techniques for CML-derived rainfall estimates and their integration with radar data, intended as a correction of radar-based precipitation fields. Additionally, it provides an opportunity to evaluate the potential contribution of CML coverage in areas where operational sensors (rain gauges or radars) are missing or non-operational, as well as to compare their performance in areas where multiple observational systems are co-located. The OpenRainER dataset was made available through a scientific collaboration between ARPAE Hydro-Meteo-Climate Service and Lepida S.c.p.A. (the company that owns the CML network). The existence of this ongoing collaboration leaves the door open for future extensions and updates of the dataset.

Data availability. The dataset is available open-source on Zenodo under CC-BY 4.0 license, https://doi.org/10.5281/zenodo.10593848 (Covi and Roversi, 2024).

Appendix A

Information about the quantization step of the signal and the average over time of TSL and RSL data were not disclosed by the data owner. Signal quantization introduces a noise term superimposed to the signal, which sets a bound on the minimum detectable rainfall intensity, dependent on link characteristics (path length, frequency and polarization). To better understand the nature of the CML raw data sampled by the CML network and available in OpenRainER, time series of RSL under precipitation-free conditions were analyzed (in contrast, TSL in dry conditions is basically constant).

An example showing five daily time series relative to as many CMLs is reported in Fig. A1. The data unveil a $1\,\mathrm{dB}$ bulk quantization step, even though there are occurrences of intermediate values between the expected RSL integer values (in dBm). We investigated the reason of this behavior, assuming that the available $1\,\mathrm{min}$ RSL data are the result of an average over time of N raw data samples having a $1\,\mathrm{dB}$ quantization step. Named RSL_0, RSL_1, ..., RSL_K the $K+1$ possible integer values of the RSL in the raw time series over the average window, sorted in ascending order, then the average RSL value is

$$RSL_N = \frac{1}{N}\sum_{k=0}^{K} N_k(RSL_0 + k) \tag{A1}$$

where N_k are the occurrences of the signal level $RSL_0 + k$, the integer index k ranges from 0 to K, and

$$N = \sum_{k=0}^{K} N_k \tag{A2}$$

https://doi.org/10.5194/essd-2026-160
Preprint. Discussion started: 12 May 2026

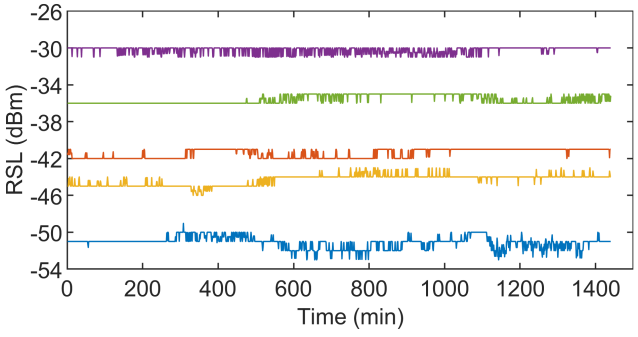


Figure A1. Example of five daily time series of the RSL during dry atmospheric conditions.

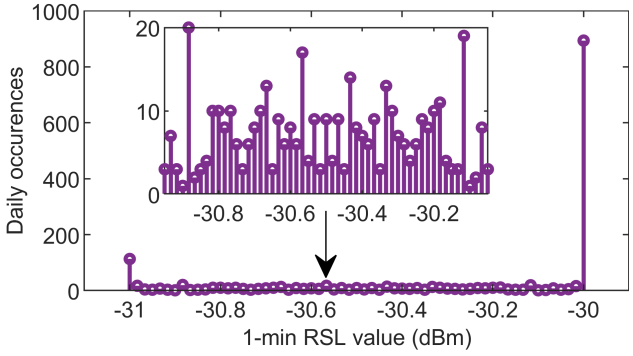


Figure A2. Counts of matching of the RSL values with the expected values (x-axis) of a 60-sample average over raw data digitized with a 1 dB quantization step between -31 and -30 dBm. The RSL series on top of previous Fig. A1 is considered.

Now, letting N_k vary between 0 and N, we find all the possible values of RSL_N. For instance, in the simple case $K = 1$, as in the time series on top of Fig. A1 where RSL_0=-31 dBm and RSL_1=-30 dB, (A1) becomes

$$RSL_N = \frac{1}{N}\left(N \cdot RSL_0 + N - N_0\right) \tag{A3}$$

and there are $N + 1$ possible values for RSL_N In particular, if $N = 60$, the 61 possible values of RSL_N coming out of (A3) are are -31, -30.9833, -30.9667, -30.9500, -30.9333, ..., -30. If we take the 1440 samples of the RSL time series on top of the figure and calculate the counts of matching with the above set of 61 possible values, it turns out that each of the 1440 samples matches one of the axis values, as shown in Fig. A2. If $K > 1$, as is the case for the last two time series in Fig. A1, the number of possible values of RSL_N increases, however the outcome is the same as in the simple case $K = 1$.

We conclude that the available 1 min RSL data are the result of an arithmetic average (on a dBm scale) of raw data sampled every 1 s and with a 1 dB quantization step. This data format is not ideal in the presence of rain where the signal is expected to undergo much faster fluctuations than in a dry atmosphere. In fact, averaging $N = 60$ RSL samples on a logarithmic scale (i.e. the dBm values) introduces a bias with respect to a linear average, which increases with the variance of the N-element time series. Overall, the logarithmic average introduces an additional error which overlaps to the 1 dB quantization error. In

https://doi.org/10.5194/essd-2026-160
Preprint. Discussion started: 12 May 2026

principle, this error could be compensated if we are able to retrieve the $K+1$ values of the coefficient N_k in (A3) from (A1) and (A2). If $K=1$, (A1) and (A2) are a system of two equations and two variables, hence the occurrences of the two quantization levels within each 1 min window can be worked out. If $K>1$, the number of variables increases, therefore several combinations of the possible $K+1$ RSL values return the same RSL average value.

Author contributions. CE reached the agreement with data owners to openly share the dataset. CE and RG conceived the idea, selected the data to be published and curated and reformatted the dataset according to the required specifications. CE, RG, and NR contributed to the description of the dataset, performed the analyses presented in the paper, and contributed to writing and revising the manuscript. All authors reviewed and approved the final version of the manuscript.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Acknowledgements. This study was conducted in the context of the project CLIMate Adaptation for the PO River Basin District – LIFE CLIMAX PO (LIFE21-IPC-IT-LIFE CLIMAX PO; CUP J49H22000030002; ID 101069928). We acknowledge the indispensable support of Lepida S.c.p.A., which provided access to their CML network. The research was also carried out within the framework of COST Action OPENSENSE, CA20136, "Opportunistic Precipitation Sensing Network," supported by COST (European Cooperation in Science and Technology). The authors used ChatGPT (OpenAI) for grammatical revision and language improvement of parts of the manuscript. The authors reviewed and edited the output and take full responsibility for the content.

https://doi.org/10.5194/essd-2026-160
Preprint. Discussion started: 12 May 2026

References

Adirosi, E., Porcu, F., Montopoli, M., Baldini, L., Bracci, A., Capozzi, V., Annella, C., Budillon, G., Bucchignani, E., Zollo, A., Cazzuli, O., Camisani, G., Bechini, R., Cremonini, R., Antonini, A., Ortolani, A., Melani, S., Valisa, P., and Scapin, S.: Database of the Italian disdrometer network, Earth System Science Data, 15, 2417–2429, https://doi.org/10.5194/essd-15-2417-2023, 2023.

Alberoni, P. P., Andersson, T., Mezzasalma, P., Michelson, D., and Nanni, S.: Use of the vertical reflectivity profile for identification of anomalous propagation, Meteorological Applications, 8, 257 – 266, https://doi.org/10.1017/S1350482701003012, 2001.

Andersson, J., Olsson, J., Beek, C., and Hansryd, J.: OpenMRG: Open data from Microwave links, Radar, and Gauges for rainfall quantification in Gothenburg, Sweden, Earth System Science Data, 14, 5411–5426, https://doi.org/10.5194/essd-14-5411-2022, 2022.

Antolini, G., Auteri, L., Pavan, V., Tomei, F., Tomozeiu, R., and Marletto, V.: A daily high-resolution gridded climatic data set for Emilia-Romagna, Italy, during 1961–2010, International Journal of Climatology, https://doi.org/10.1002/joc.4473, 2015.

ARPAE Emilia-Romagna: Dati dalle stazioni meteo locali della rete idrometeorologica regionale, https://dati.arpae.it/dataset/dati-dalle-stazioni-meteo-locali-della-rete-idrometeorologica-regionale, dataset open data, 2023.

Barbero, S., Zaccagnino, M., Mariani, S., Lastoria, B., Braca, G., Bussettini, M., Casaioli, M., Marsico, L., Rotundo, R., Pavan, V., Ricciardi, G., Zenoni, E., Cicogna, A., Micheletti, S., Cazzuli, O., Di Priolo, S., Ranci, M., Rondanini, C., Bianco, G., Egiatti, G., Montanini, P., Saccardo, I., Campione, E., Pupillo, S., Iocca, F., Lazzeri, M., Tedeschini, M., Marzano, V., Schena, P., Licciardello, A., Manzella, F., Brunier, F., and Ratto, S.: Linee Guida per il controllo di valità dei dati idro-meteorologici, Tech. rep., ISPRA, 2017.

Bech, J., Codina, B., Lorente, J., and Bebbington, D.: The Sensitivity of Single Polarization Weather Radar Beam Blockage Correction to Variability in the Vertical Refractivity Gradient, Journal of Atmospheric and Oceanic Technology - J ATMOS OCEAN TECHNOL, 20, 845–855, https://doi.org/10.1175/1520-0426(2003)020<0845:TSOSPW>2.0.CO;2, 2003.

Bianchi, B., Van Leeuwen, P. J., Hogan, R., and Berne, A.: A Variational Approach to Retrieve Rain Rate by Combining Information from Rain Gauges, Radars, and Microwave Links, Journal of Hydrometeorology, 14, 1897–1909, https://doi.org/10.1175/JHM-D-12-094.1, 2013.

Chwala, C. and Kunstmann, H.: Commercial microwave link networks for rainfall observation: Assessment of the current status and future challenges, Wiley Interdisciplinary Reviews: Water, 6, https://doi.org/10.1002/wat2.1337, 2019.

Covi, E. and Roversi, G.: OpenRainER, Zenodo dataset, https://doi.org/10.5281/zenodo.10593848, licensed under CC-BY 4.0, 2024.

David, N., Liu, Y., Kumah, K., Hoedjes, J., Su, B., and Gao, H.: On the Power of Microwave Communication Data to Monitor Rain for Agricultural Needs in Africa, Water, 13, https://doi.org/10.3390/w13050730, 2021.

Dotzek, N., Gourley, J. J., and Fehr, T.: Vertical profiles of radar reflectivity and rain rate: observational and modeling results, in: Proceedings of the European Conference on Radar in Meteorology and Hydrology (ERAD), vol. 1, pp. 243–249, ERAD Publications Series, location not specified, https://www.essl.org/dotzek/papers.htm, observational and modeling results, 2002.

Emilia-Romagna, R.: Digital Terrain Model (DTM) – 5 m resolution, https://metasfera.regione.emilia-romagna.it/ricerca_metadato?uuid=r_emiro:2016-08-08T155835, licensed under CC-BY 4.0. Accessed: 2026, 2016.

Fencl, M., Dohnal, M., Rieckermann, J., and Bareš, V.: Gauge-adjusted rainfall estimates from commercial microwave links, Hydrology and Earth System Sciences, 21, 617–634, https://doi.org/10.5194/hess-21-617-2017, 2017.

Fencl, M., Nebuloni, R., Andersson, J., Bareš, V., Blettner, N., Cazzaniga, G., Chwala, C., Colli, M., Vos, L., El Hachem, A., Charles, G., Giannetti, F., Graf, M., Jacoby, D., Habi, H., Musil, P., Ostrometzky, J., Roversi, G., Sapienza, F., and Zheng, X.: Data formats and standards for opportunistic rainfall sensors, Open Research Europe, 3, 169, https://doi.org/10.12688/openreseurope.16068.1, 2023.

https://doi.org/10.5194/essd-2026-160
Preprint. Discussion started: 12 May 2026

Fornasiero, A., Bech, J., and Alberoni, P. P.: Enhanced radar precipitation estimates using a combined clutter and beam blockage correction technique, Natural Hazards and Earth System Sciences, 6, 697–710, https://doi.org/10.5194/nhess-6-697-2006, 2006.

Franco, M., Sánchez-Diezma, R., Sempere-Torres, D., and Zawadzki, I.: An Improved Methodology for Classifying Convective and Stratiform Rain, in: 32nd Conference on Radar Meteorology, Albuquerque, NM, USA, https://ams.confex.com/ams/32Rad11Meso/techprogram/paper_96469.htm, extended abstract, Session 7R.4, 2005.

Germann, U. and Joss, J.: Mesobeta Profiles to Extrapolate Radar Precipitation Measurements above the Alps to the Ground Level, Journal of Applied Meteorology - J APPL METEOROL, 41, 542–557, https://doi.org/10.1175/1520-0450(2002)041<0542:MPTERP>2.0.CO;2, 2002.

International Telecommunication Union: Recommendation ITU-R P.838-3: Specific attenuation model for rain for use in prediction methods, ITU-R Recommendation, geneva, Switzerland, 2005.

Istat: Borders of administrative units, https://www.istat.it/notizia/confini-delle-unita-amministrative-a-fini-statistici-al-1-gennaio-2018-2/, licensed under CC-BY 4.0, 2024.

Jacoby, D., Yu, S., Hu, Q., Hine, Z., Johnson, R., Ostrometzky, J., Kadota, I., Zussman, G., and Messer, H.: OpenMesh: Wireless Signal Dataset for Opportunistic Urban Weather Sensing in New York City, Earth System Science Data Discussions, 2025, 1–27, https://doi.org/10.5194/essd-2025-238, 2025.

Joss, J. and Waldvogel, A.: Operational use of radar for precipitation, Tech. rep., MeteoSwiss, technical report, 1988.

Ochoa-Rodríguez, S., Wang, L., Willems, P., and Onof, C.: A Review of Radar-Rain Gauge Data Merging Methods and Their Potential for Urban Hydrological Applications, Water Resources Research, 55, 6356–6391, https://doi.org/10.1109/TGRS.2021.3110004, 2019.

Overeem, A., Walraven, B., Leijnse, H., and Uijlenhoet, R.: Four-year commercial microwave link dataset for the Netherlands, Version 1, 4TU.ResearchData dataset, https://doi.org/10.4121/be252844-b672-471e-8d69-27269a862ec1.v1, dataset, CC BY 4.0, 2024.

Pastorek, J., Fencl, M., Rieckermann, J., and Bareš, V.: Precipitation Estimates From Commercial Microwave Links: Practical Approaches to Wet-Antenna Correction, IEEE Transactions on Geoscience and Remote Sensing, 60, 1–9, https://doi.org/10.1109/TGRS.2021.3110004, 2022.

Roversi, G., Alberoni, P. P., Fornasiero, A., and Porcu, F.: Commercial microwave links as a tool for operational rainfall monitoring in Northern Italy, Atmospheric Measurement Techniques, 13, 5779–5797, https://doi.org/10.5194/amt-13-5779-2020, 2020.

Schleiss, M. and Berne, A.: Identification of Dry and Rainy Periods Using Telecommunication Microwave Links, Geoscience and Remote Sensing Letters, IEEE, 7, 611 – 615, https://doi.org/10.1109/LGRS.2010.2043052, 2010.

Steiner, M., Houze, R., and Yuter, S.: Climatological characterization of three-dimensional storm structure from operational radar and rain gauge data, Journal of Applied Meteorology, 34, 1978–2007, https://doi.org/10.1175/1520-0450(1995)034<1978:CCOTDS>2.0.CO;2, 1995.

Uijlenhoet, R., Overeem, A., and Leijnse, H.: Opportunistic remote sensing of rainfall using microwave links from cellular communication networks, Wiley Interdisciplinary Reviews: Water, 5, https://doi.org/10.1002/wat2.1289, 2017.