https://doi.org/10.5194/essd-2026-939
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.



Open Access Earth System Science Data Discussions

Operational microwave links for hydrometeorological sensing: a multisource dataset for the Czech Republic

Taoufiq Shit[1], Martin Fencl[1], Martin Mudroch[2], and Vojtěch Bareš[1]

[1]Department of Hydraulics and Hydrology, Czech Technical University in Prague, Prague, Czech Republic
[2]Department of Electromagnetic Field, Czech Technical University in Prague, Prague, Czech Republic

Correspondence: Taoufiq Shit (taoufiq.shit@fsv.cvut.cz)

Abstract. Commercial microwave links (CMLs) provide an opportunity to observe rainfall and atmospheric water vapour using existing telecommunication infrastructure. Their wider use remains limited by restricted access to operational signal data and by the processing needed to derive hydrometeorological information.

We present an open multisource dataset from a national operational CML network in the Czech Republic, covering 1 April to 31 October 2021. To our knowledge, this is the first CML dataset to be released openly that combines instantaneous signal observations recorded at 1 min intervals across a large spatial domain covering urban and rural regions, a large population of E-band links, and independent radar and meteorological station observations for the same period. This combination supports both rainfall estimation and water vapour sensing. The raw data cover 2574 CMLs with a wide range of frequencies and path lengths. After quality control, 1987 CMLs are retained, of which 810 operate in the E-band at 71–76 and 81–86 GHz. The dataset also provides 1 min rainfall estimates for all 1987 quality controlled CMLs, radar rainfall averaged along 1132 CML paths, and observations of precipitation from 424 stations, air temperature from 251 stations, and relative humidity from 247 stations.

Comparison with the nearest rain gauge within 5 km of each CML midpoint over gauge wet intervals gives a root mean square error (RMSE) of 3.63 $\mathrm{mm\,h^{-1}}$, a mean bias error of -0.91 $\mathrm{mm\,h^{-1}}$, and a Pearson correlation coefficient of 0.63. Water vapour density retrieved from a selected subset of 31 physical E-band CMLs and 45 directional sublinks gives an RMSE of 3.46 $\mathrm{g\,m^{-3}}$ and a Pearson correlation coefficient of 0.60 against station observations. These results illustrate the potential of the released observations for both rainfall estimation and water vapour sensing. The dataset is available at https://doi.org/10.5281/zenodo.23398309 (Shit et al., 2026).

1 Introduction

Reliable rainfall information is essential for flood warning, urban drainage, runoff modelling, water management, and weather research (Hingray et al., 2014; Olsson et al., 2025). Rain gauges provide direct observations at individual points but cannot resolve all rainfall variability between stations. Weather radar provides much wider spatial coverage, but its estimates are affected by beam geometry, terrain, clutter, attenuation, and uncertainty in the conversion from reflectivity to rain rate (Fabry, 2015). These limitations are especially important during short and spatially variable rainfall events.

https://doi.org/10.5194/essd-2026-939
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.



Commercial microwave links (CMLs) provide an additional source of rainfall information. Telecommunication operators use CMLs to connect sites within cellular communication networks. Rainfall along a link path attenuates the microwave signal, and this attenuation can be related to path averaged rain rate using the link frequency, polarization, and length (Messer et al., 2006; Overeem et al., 2011; Chwala and Kunstmann, 2019). Because the links are already deployed and maintained for communication, their signals can provide environmental observations without requiring the construction of a new dense meteorological network. The attenuation of the signals also depends on atmospheric water vapour, so high frequency links can be used to monitor water vapour near the ground (Fencl et al., 2020, 2021).

Wider use of CML observations is nevertheless limited by two related barriers. First, operational signal data are owned by telecommunication operators and are rarely released openly. This restricts reproducibility, comparison between methods, and evaluation across different climates and network designs. Second, recorded signal levels are not themselves rainfall estimates. Using them for rainfall applications generally requires quality checks and rainfall retrieval, whether with physical or data-driven methods. These steps can be an entry barrier for hydrologists, remote sensing researchers, water managers, and other users interested in rainfall applications rather than CML signal processing (Chwala and Kunstmann, 2019; Blettner et al., 2023).

Existing open CML datasets address important parts of these barriers but cover different experimental settings and frequency ranges. CoMMon provides detailed observations from a single 38 GHz research link with collocated meteorological instruments and is especially useful for process studies (Špačková et al., 2021). OpenMRG provides high resolution observations from a larger urban operational network at 7–38 GHz together with rain gauges and weather radar, creating an important benchmark for retrieval development (Andersson et al., 2022).

Longer and broader operational records are also available. The Dutch archive provides national observations mainly between about 13 and 40 GHz (Overeem et al., 2016a, 2024). OpenMesh provides observations at 5–6, 24, and 58–70 GHz from a community internet network in New York City (Jacoby et al., 2026), while OpenRainER combines operational links at 24.6 and 25.6 GHz with radar and gauge observations over contrasting terrain (Covi et al., 2026). These datasets provide valuable references, but users often still need to perform quality control, implement a rainfall retrieval, or assemble additional meteorological observations before the data can be used directly in hydrological applications.

The 71–76 and 81–86 GHz E-band ranges are still weakly represented in open rainfall datasets. These frequencies are increasingly used for high capacity cellular backhaul (Fencl et al., 2020). Their higher sensitivity to rainfall makes them useful for opportunistic sensing, but atmospheric gases, water vapour, hardware behaviour, and signal quantization can also contribute more strongly to the observed attenuation (Fencl et al., 2020, 2021). This creates both additional retrieval challenges and opportunities for atmospheric sensing.

The remaining gap is therefore an operational dataset that combines 1 min signals from a geographically and technically diverse national CML network, including a substantial E-band population, with documented quality control, rainfall estimates ready to use, and independent meteorological observations for the same period.

This study addresses this gap with an open multisource hydrometeorological dataset for the Czech Republic covering 1 April–31 October 2021. The network spans major cities and rural regions, with variation in link density, length, operating

https://doi.org/10.5194/essd-2026-939
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.



frequency, and local environment. The dataset contains 1 min raw signal observations from 2574 CMLs, quality controlled observations from 1987 CMLs, and 1 min rainfall estimates for the complete quality controlled population. Among these, 810 CMLs operate at 71–76 and 81–86 GHz.

The CML rainfall estimates are accompanied by path averaged radar rainfall and by precipitation observations from 424 stations that provide an independent point reference. The dataset also includes air temperature from 251 stations and relative humidity from 247 stations at 10 min resolution during the same April–October period.

The benchmark analyses show the main uses of the dataset. CML rain detection and rainfall estimates are evaluated against nearby rain gauges, which are not used in the retrieval, while path averaged radar rainfall provides a second independent reference in the event example. Rainfall field reconstruction illustrates spatial applications of the path observations. An E-band example explores water vapour density (WVD) retrieval using the same data and the accompanying meteorological records. These analyses describe the content, performance, and limitations of the released data rather than optimizing the retrieval methods.

2 Study area

The dataset covers the Czech Republic in Central Europe. The country has a temperate climate influenced by oceanic and continental conditions. Warm season precipitation often includes convective showers and thunderstorms, while longer stratiform events are commonly associated with frontal systems. Orographic effects produce additional variability between lowlands and mountain regions (Tolasz, 2007).

Figure 1 shows the observation networks. CMLs are concentrated mainly in populated regions, while rain gauges and weather stations are distributed across the country. The RADKLIM domain used here covers mainly the western part of the Czech Republic, so path averaged radar rainfall is available only for links that intersect valid radar pixels.

Figure 2 summarizes network wide availability during April–October 2021, the common period used for multisensor examples and benchmark comparisons. CML availability reflects both data gaps and changes in the set of links reporting at a given time. The lower values in May therefore do not indicate an outage of the whole network.

https://doi.org/10.5194/essd-2026-939
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.



Open Access Earth System Science Data Discussions

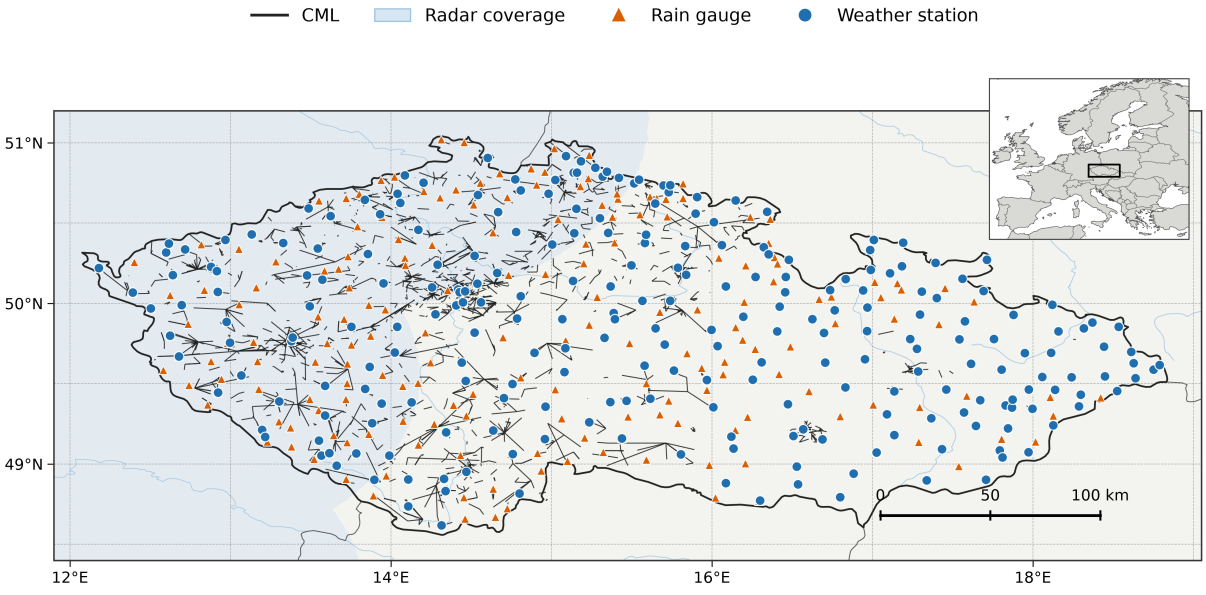

Figure 1. Observation networks in the Czech Republic. Black lines show the paths of the 2574 CMLs in the raw dataset. Orange triangles show rain gauges. Blue circles show weather stations with air temperature or relative humidity observations. The shaded area shows the part of the RADKLIM domain used to assign radar rainfall to link paths. The base map and the inset use Natural Earth data (public domain, https://www.naturalearthdata.com).

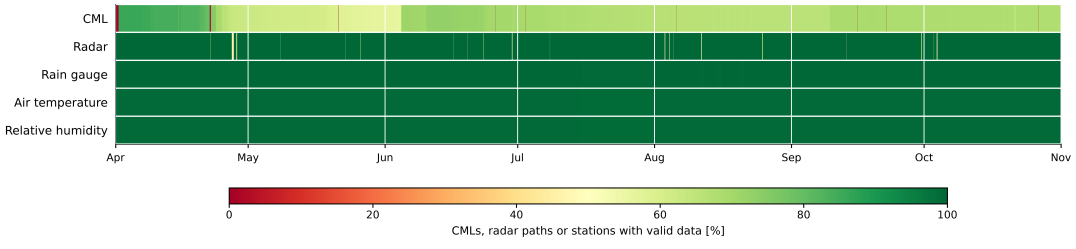

Figure 2. Availability during April–October 2021 for CML observations, path averaged radar rainfall, ground precipitation, air temperature, and relative humidity. Values show the percentage of CMLs, radar paths, or stations with valid data at each native time step (1 min for CMLs, 5 min for radar, and 10 min for stations). The CML row refers to all 2574 CMLs in the raw dataset. A CML is counted as available when at least one of its sublinks has both a transmitted and a received signal level.

https://doi.org/10.5194/essd-2026-939
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.



Open Access Earth System Science Data Discussions

Table 1. Overview of the released data products. All products cover the common period 1 April–31 October 2021. Detailed file structures and variable definitions are provided in the Supplement.

Data product	Spatial coverage	Time step	Measurement geometry	Primary use
Meteorological observations				
Ground precipitation	424 stations	10 min	Point	Independent point rainfall reference.
Air temperature	251 stations	10 min	Point	Atmospheric context and propagation studies.
Relative humidity	247 stations	10 min	Point	Atmospheric context and WVD reference.
Radar path averages	1132 CML paths	5 min	Path	Radar rainfall matched to CML geometry.
CML signal observations				
Raw CML observations	2574 CMLs	1 min	Path	Original signal data for independent processing.
Quality controlled CML observations	1987 CMLs	1 min	Path	Screened signals for retrieval and propagation studies.
Derived rainfall product				
CML rainfall	1987 CMLs	1 min	Path	Rainfall estimates ready to use.

3 Dataset

This section describes the provenance, scientific content, and intended role of each released product. Detailed file names, dimensions, coordinates, variables, units, quality fields, and retrieval configurations are documented in the Supplement and repository metadata. All products are available on Zenodo (Shit et al., 2026); access during the review is described in Sect. 8.

3.1 Ground station observations

The ground observations were obtained from the historical climatological open data archive of the Czech Hydrometeorological Institute (CHMI) (Czech Hydrometeorological Institute, 2025). The archive provides 10 min observations from automated stations and records observation times in UTC. Stations are identified by their WIGOS Station Identifier. Available station metadata, coordinates, elevation, measurement height, units, and CHMI quality information were retained during conversion to NetCDF.

Precipitation, air temperature, and relative humidity are released as separate products covering 1 April 00:00 UTC to 31 October 23:50 UTC 2021. The number of stations differs between the products because not every site measures every variable.

https://doi.org/10.5194/essd-2026-939
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.



The CHMI network contains different instrument types, and the public metadata do not provide a complete station specific mapping to instrument manufacturer or model. Detailed metadata are therefore preserved without assigning one instrument type to all stations (Czech Hydrometeorological Institute, 2016, 2017).

3.1.1 Precipitation

The precipitation product contains 424 automated stations. Its variable `rr` holds the CHMI element `SRA10M`, the precipitation amount in millimetres accumulated over each 10 min interval, together with the CHMI quality value and flag. The product provides an independent point reference for CML rain detection and CML rainfall. Differences are expected because a station samples one location, whereas a CML represents attenuation integrated along a path.

3.1.2 Air temperature

The air temperature product contains 251 automated stations. The principal variable, `T`, represents air temperature in degrees Celsius. The additional CHMI elements `TMA`, `TMI`, and `TPM`, corresponding to maximum, minimum, and near ground temperature, are retained where available together with their metadata and quality information. Air temperature provides meteorological context and is also used in the WVD retrieval.

3.1.3 Relative humidity

The relative humidity product contains 247 automated stations. Its principal variable, `H`, represents relative humidity in percent and is accompanied by the original CHMI quality information. Relative humidity is not used to calibrate or generate the released CML rainfall estimates. It is used only for the station reference in the WVD evaluation.

CHMI documentation lists Vaisala HMP45 and HMP155 combined temperature–humidity probes and PT100 resistance temperature sensors among the instruments used for temperature and relative humidity in this network (Czech Hydrometeorological Institute, 2016, 2017).

3.2 Weather radar observations

The radar reference is derived from the German Weather Service (DWD) RADKLIM YW Version 2017.002 product (Winterrath et al., 2018). For 2021, DWD documents that the originally distributed NetCDF files are incorrect; the corrected 2021 NetCDF files provided in the RADKLIM supplement were therefore used here. RADKLIM YW is provided at 5 min resolution on a 1 km grid and includes gauge adjustment information transferred from the hourly RADKLIM RW product. In the corrected NetCDF files, the native `RR` variable is a 5 min precipitation amount with units $\mathrm{kg\,m^{-2}}$, numerically equivalent to millimetres of water. The RADKLIM domain extends into the western part of the Czech Republic, while areas outside the valid radar composite remain missing.

To match the spatial support of the CML observations, radar pixels intersected by each link were weighted by the length of the link segment contained within each pixel. The path averaged 5 min precipitation amount is

https://doi.org/10.5194/essd-2026-939
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.



$$P_{t,i}^{\mathrm{radar}} = \frac{\sum_{k=1}^{N_i} l_{k,i} P_{k,t}}{L_i}, \tag{1}$$

where $P_{k,t}$ is the native 5 min precipitation amount of the DWD variable `RR` in pixel k at time t, $l_{k,i}$ is the length of link i within that pixel, L_i is the full link length, and N_i is the number of intersected pixels. If any intersected pixel has no valid radar value at a time step, the path value is missing for that step. For comparisons expressed as rain rate, the path averaged amount is converted to an equivalent 5 min mean rate as $R_{t,i}^{\mathrm{radar}} = 12 P_{t,i}^{\mathrm{radar}}$. This factor is used only for evaluation and plots of rain rates. The DWD source variable is named `RR`, while the released path averaged product stores $P_{t,i}^{\mathrm{radar}}$ as `rr`, also as a 5 min precipitation amount.

The released path averaged product contains precipitation amounts for 1132 CML paths at 5 min resolution from April through October 2021, 1033 of which belong to quality controlled CMLs. Link identifiers are preserved so that radar observations can be joined directly to the corresponding CML rainfall estimates. Missing radar values remain missing and must not be interpreted as zero rainfall or dry conditions. The native gridded RADKLIM product is not redistributed.

3.3 Commercial microwave link observations

The CML observations were provided by T-Mobile Czech Republic from an operational telecommunication network. Each physical link generally consists of two directional sublinks, for which transmitted and received signal levels were recorded at 1 min resolution. The metadata include endpoint coordinates, path length, carrier frequency, and polarization. The released observations represent the subset of operational links available through the research data acquisition system and should not be interpreted as the complete operator network. The released raw dataset contains all 2574 CMLs. Nineteen further CMLs of the acquisition are not included because their transmitted signal level was 0 dBm throughout 2021 or paired transmitted and received signal levels were missing. After quality control, 1987 CMLs were retained for rainfall processing. The quality controlled product contains the screened signal levels, the total signal loss calculated from them, and the estimated minimum detectable rain rate (Sect. 4.1).

The Czech network contains both conventional microwave links and a substantial E-band component. E-band is defined here by the paired allocations 71–76 and 81–86 GHz. The network is also characterized by short paths. Table 2 compares the E-band share and the short paths of the raw and quality controlled CMLs, and Fig. 3 shows the frequency and path length distributions of the raw dataset.

3.3.1 Data acquisition and signal characteristics

The links of the network are Ericsson MINI-LINK radios. The CML data were acquired by an application running at the T-Mobile Network Operations Centre, which queried each radio unit every minute for the transmitted signal level (TSL) and the received signal level (RSL) of each directional sublink via the Simple Network Management Protocol (SNMP) and stored the records in a PostgreSQL database. The acquisition was interrupted several times, for example after changes of the access to the operator network. These gaps remain missing in the data.

https://doi.org/10.5194/essd-2026-939
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

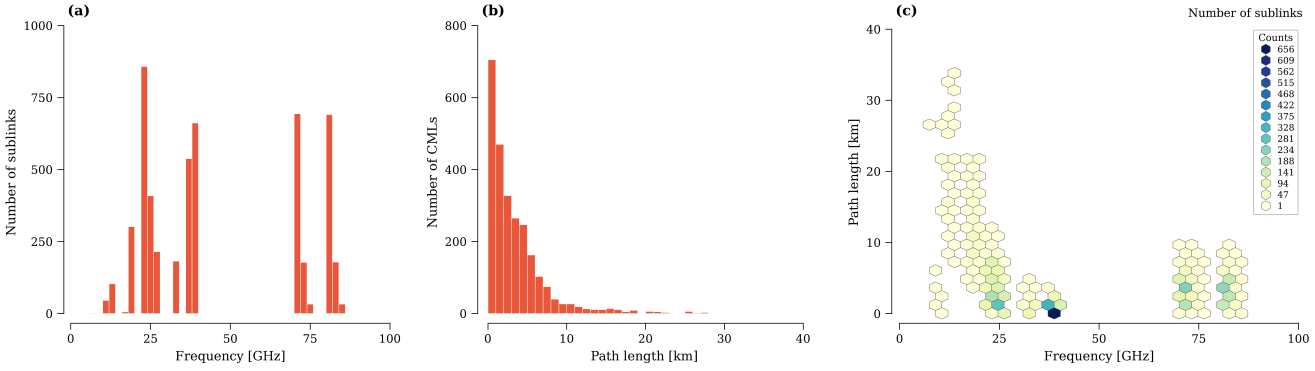

Figure 3. Characteristics of the commercial microwave link network in the raw dataset. Panel (a) shows the carrier frequencies of 5132 directional sublinks with available frequency metadata, panel (b) shows the path lengths of all 2574 CMLs, and panel (c) shows the relationship between frequency and path length for the 5132 sublinks.

Automatic transmit power control (ATPC) is enabled on all links. ATPC changes the transmitted power during operation to keep the received signal within a target range: when the attenuation along the path increases, for example during rain, the TSL is raised, and it is lowered again afterwards.

The recorded signal levels have a limited resolution, which differs between frequency bands. For links below 40 GHz, the TSL mostly changes in steps of 1 dB and the RSL in steps of about 0.33 dB, stored as values rounded to 0.1 dB. For most E-band links, both levels change in steps of 0.1 dB. These steps are the most common changes between consecutive minutes in the raw data, not manufacturer specifications, and a small part of the stored values lies between them. In the raw data, the TSL is constant during the period on 17 % of the sublinks below 40 GHz and on 5 % of the E-band sublinks. For links below 40 GHz, the TSL changes less often than the RSL, so the resolution of the total signal loss, TSL minus RSL, is mainly set by the RSL. The resolution limits the smallest attenuation that can be observed and thus also the resolution of the rain rate and water vapour density estimates: on short links one step corresponds to a high rain rate (Sect. 4.1), and for water vapour it determines the resolution $\Delta\rho_q$ (Sect. 4.4).

3.3.2 CML rainfall estimates

The dataset includes 1 min rainfall estimates for all 1987 quality controlled CMLs from April through October 2021. Wet and dry periods are identified using the method of Graf et al. (2020), and rainfall is estimated as described in Sect. 4.3. Estimates for each directional sublink that passes the quality control are retained separately. The released rainfall product has a time resolution of 1 min, while the 5 min values used in the comparisons are derived from these estimates.

The retrieval combines the wet–dry classification, a constant dry baseline, a correction for wet antenna attenuation (Pastorek et al., 2022), and the ITU-R P.838-3 attenuation–rainfall relation. It uses no radar or rain gauge observations. Rain rates calculated from reconstructed or interpolated total signal loss are marked in the rainfall file, so that users can exclude them.

https://doi.org/10.5194/essd-2026-939
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.



Open Access Earth System Science Data Discussions

4 Methods

This section describes quality control, rain detection and rainfall estimation from CMLs, water vapour retrieval from E-band CMLs, matching with and evaluation against independent meteorological observations, and rainfall field reconstruction.

4.1 Quality control

Operational CML records contain missing observations and disturbances related to network operation, receiver limits, maintenance, hardware behaviour, and data transfer. The quality control follows the processing logic developed for the Czech network by Blettner et al. (2023). In short, invalid values are removed, signal losses during strong attenuation are handled, and each directional sublink is then tested separately for excessive signal fluctuation and for a minimum detectable rain rate that is too high. Quality control was applied to the period 1 April–31 October 2021. Invalid TSL values at or below -99 dBm or equal to 255 and RSL values at or below -99.9 dBm or above -20 dBm were first set to missing. Low RSL plateaus, where the RSL stays below -85 dBm with a standard deviation below 0.5 dB, were also set to missing. The RSL can be lost during strong attenuation. For RSL gaps of up to 60 min next to an RSL below -65 dBm, the total signal loss was reconstructed as the highest TSL minus the lowest RSL of the sublink in the period, following Blettner et al. (2023). Unlike Blettner et al. (2023), the TSL and RSL values themselves were not filled, and gaps containing masked minutes were not reconstructed. Short gaps of up to 5 min in the recorded data were interpolated linearly when total signal loss was available on both sides. Reconstructed and interpolated minutes are marked in a quality flag.

Directional sublinks with persistent fluctuations not related to weather were excluded. A sublink failed this test when more than 10 % of valid 5 h windows had a signal standard deviation above 2 dB, or when more than 33 % of valid 1 h windows were above 0.8 dB. A window was valid only when all its minutes had total signal loss, and reconstructed minutes were not used in this test.

Sublinks with an estimated minimum detectable rain rate of at least 2 mm h^{-1} were also excluded. The stored detection limit is the rain rate that produces a path attenuation of 0.33 dB, the RSL step of the links below 40 GHz (Sect. 3.3.1), used for all sublinks and calculated for each directional sublink with the ITU-R P.838-3 (2005) attenuation–rainfall relation (International Telecommunication Union, 2005). Sublinks with a TSL of 0 dBm throughout the period were treated as having no usable signal. A physical CML was retained when at least one directional sublink passed these checks. The processing steps are listed in the metadata of the quality controlled file and in the Supplement.

Table 3 summarizes the outcome of the quality control, and Table 2 shows that it removes many of the shortest links. The 1987 retained CMLs contain 3731 directional sublinks that pass the quality control; the other sublinks of these CMLs keep missing signal values. Among the retained CMLs, 965 had at least one reconstructed signal blackout; in total, 6410 gaps covering 37414 directional minutes were reconstructed and flagged.

https://doi.org/10.5194/essd-2026-939
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.



Open Access Earth System Science Data Discussions

Table 2. Raw and quality controlled CML populations. E-band denotes the paired allocations 71–76 and 81–86 GHz. Percentages refer to all physical CMLs of each population; a dash marks a value that is not given.

	Raw	Quality controlled
Physical CMLs	2574	1987
with valid carrier frequency metadata	2566	1987
E-band CMLs	904	810 (40.8 %)
Paths shorter than 1 km	27 %	15 %
Paths shorter than 0.1 km	1 %	–

Table 3. Outcome of the quality control for the physical CMLs of the raw dataset. The exclusion categories are mutually exclusive.

	Physical CMLs
Raw dataset	2574
Without usable carrier frequency metadata, not assessed	8
Assessed by the quality control	2566
Excluded: minimum detectable rain rate of at least 2 mm h^{-1}	414
Excluded: signal fluctuation	99
Excluded: both criteria	3
Excluded: no usable observations in the period	63
Retained after quality control	1987

4.2 Rain detection

For each directional sublink, wet and dry minutes are identified from the variability of total signal loss, following Graf et al. (2020). The resulting classification is used for rainfall estimation and evaluated against rain gauges.

For directional sublink s, the total signal loss TL_s is calculated as transmitted signal level minus received signal level and includes the reconstructed and interpolated values described in Sect. 4.1. For each sublink, the classification is defined by

$$\sigma_{60,s}(t) = \mathrm{std}\{TL_s(\tau) : \tau \in \mathcal{W}_{60}(t)\}, \qquad \theta_s = 1.12\, Q_{0.80,s}, \tag{2}$$

where $\mathcal{W}_{60}(t)$ is the centred 60 min window and $Q_{0.80,s}$ is the empirical 0.80 quantile of $\sigma_{60,s}$ calculated from complete windows of 60 samples in April–October 2021. Classification at a given minute requires at least 45 valid observations within the window, and a valid 1 min observation is wet when $\sigma_{60,s}(t) > \theta_s$.

For comparison with the 10 min gauge reference, the 1 min classifications are aggregated to the 10 min gauge intervals using the CML–gauge matching described in Sect. 4.5. An interval is wet when at least one valid minute is wet, and only complete

https://doi.org/10.5194/essd-2026-939
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.



Open Access Earth System Science Data Discussions

intervals are retained. Missing observations are not filled. Gauge intervals with an equivalent rain rate of at least 0.1 mm h^{-1} are treated as wet. The detection evaluation uses the same CML–gauge pairs as the rainfall comparison but also includes gauge dry intervals. Using gauges rather than radar for this evaluation avoids restricting it to the western RADKLIM domain.

4.3 Rainfall estimation

The released rainfall product uses the wet–dry classification described above. Neither radar nor gauge observations are used in the retrieval.

For each directional sublink, the dry baseline is estimated with the constant baseline method using the most recent valid dry observation. The observed excess attenuation is

$$A_{\mathrm{obs},s}(t) = \max[TL_s(t) - B_s(t), 0], \tag{3}$$

where $B_s(t)$ is the dry baseline of the rainfall retrieval. Wet antenna attenuation is calculated as

$$A_{\mathrm{waa},s}(R) = A_{\max}\left[1 - \exp\left(-dR^{\zeta}\right)\right], \tag{4}$$

with $A_{\max} = 14$ dB, $\zeta = 0.55$ and $d = 0.1$ (Pastorek et al., 2022), and the rain rate $R_s(t)$ in mm h^{-1} is obtained from

$$R_s(t) = \left[\frac{\max(A_{\mathrm{obs},s}(t) - A_{\mathrm{waa},s}(R_s(t)), 0)}{L_i\, a_s}\right]^{1/b_s}, \tag{5}$$

where L_i is the path length of link i, to which sublink s belongs, and a_s and b_s are the ITU-R P.838-3 coefficients for the sublink frequency and polarization (International Telecommunication Union, 2005). The 2005 coefficient set is used, and the implicit equation, which includes wet antenna attenuation, is solved within the retrieval. Rainfall is retrieved at 1 min resolution for each directional sublink, and these estimates are released at their native time step. For analyses requiring 5 min rain rates, five consecutive 1 min directional estimates are averaged only when all five are valid; otherwise the 5 min value is missing. At either time step, the rain rate of a physical CML is the mean of its valid directional estimates, calculated after any temporal aggregation. The released NetCDF retains both directional estimates and also provides their valid mean at 1 min resolution.

4.4 Water vapour density retrieval

The E-band observations are also used to retrieve WVD as an example of atmospheric sensing with the released data. The aim is not to propose or optimize a new retrieval method. The physical basis follows Fencl et al. (2020, 2021).

For each directional sublink, the dry air attenuation is calculated using ITU-R P.676-12 (International Telecommunication Union, 2019) through ITU-Rpy (version 0.4.0) (del Portillo, 2017) and subtracted from the measured total signal loss. The calculation uses the link frequency, path length, a fixed pressure of 980 hPa, and air temperature from the nearest suitable

https://doi.org/10.5194/essd-2026-939
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.



Open Access Earth System Science Data Discussions

CHMI station within 20 km. Relative humidity is not used in the CML retrieval. Wet periods and nearby unstable periods are excluded from the baseline estimation using CML based screening, including a 6 h buffer around detected wet periods. The remaining dry observations are used to define a lower dry signal level, which is updated from the preceding days within each identified hardware state. Here a hardware state denotes a contiguous signal regime between persistent level shifts of the signal. In the equations below, $\widetilde{TL}_s(t)$ denotes the total signal loss calculated from the raw TSL and RSL after subtraction of the dry air attenuation, and $B_s^{\mathrm{WVD}}(t)$ denotes this lower dry signal level. Both differ from $TL_s(t)$ and $B_s(t)$ of the rainfall retrieval (Sects. 4.2 and 4.3). The attenuation anomaly $\Delta A_s(t)$ in dB, the specific attenuation anomaly $\Delta\gamma_s(t)$ in dB km^{-1}, and the P.676 inversion for the WVD $\rho_{v,s}(t)$ are

$$\Delta A_s(t) = \widetilde{TL}_s(t) - B_s^{\mathrm{WVD}}(t), \tag{6}$$

$$\Delta\gamma_s(t) = \frac{\Delta A_s(t)}{L_i}, \tag{7}$$

$$\gamma_{\mathrm{P.676}}(f_s, T(t), p, \rho_{v,s}(t)) - \gamma_{\mathrm{P.676}}(f_s, T(t), p, \rho_{v,0,s}) = \Delta\gamma_s(t). \tag{8}$$

Here $\gamma_{\mathrm{P.676}}$ is the specific attenuation of ITU-R P.676-12 in dB km^{-1}, $p = 980$ hPa, f_s is the sublink frequency in GHz, $T(t)$ is the matched air temperature in °C, which is converted to kelvin for the ITU-Rpy calculation, and $\rho_{v,0,s}$ denotes the reference condition of the hardware state. In the processing code this reference condition is a fixed value, $\rho_{v,0,s} = 0$ g m^{-3}: the lower dry signal level $B_s^{\mathrm{WVD}}(t)$ is assumed to correspond to zero water vapour density. Equation (8) is solved numerically for $\rho_{v,s}(t)$ in g m^{-3}; no station humidity is used to set the CML baseline.

The station reference used for validation is calculated from the relative humidity $RH(t)$ in percent and the air temperature $T_{\mathrm{obs}}(t)$ in °C of the matched station (Sect. 4.5) as

$$e(t) = \frac{RH(t)}{100}\, e_{\mathrm{sat}}[T_{\mathrm{obs}}(t)], \tag{9}$$

$$\rho_{v,\mathrm{obs}}(t) = 216.7\, \frac{e(t)}{T_{\mathrm{obs}}(t) + 273.15}, \tag{10}$$

where e is the water vapour pressure in hPa, $e_{\mathrm{sat}}(T) = 6.112\, \exp[17.67T/(243.5+T)]$ hPa is the saturation vapour pressure over water with T in °C, and $\rho_{v,\mathrm{obs}}$ is in g m^{-3}. The CHMI quality codes and flags of the temperature and humidity records are not used; all finite values are retained, except relative humidity values outside 0–105 %, which are excluded from the station reference.

Because WVD observability depends strongly on path length and on the effective resolution of the stored signal levels, the quantization resolution expressed in WVD, $\Delta\rho_q$, is calculated for each directional sublink. The quantization of total signal loss, $q_{\mathrm{TL},s}$, is estimated empirically from discrete level changes in the raw 1 min TSL and RSL within the stable hardware states identified in the WVD processing. It therefore describes the effective resolution of the stored observations, not a specification of the radio hardware. The conversion is

https://doi.org/10.5194/essd-2026-939
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.



$$q_{\gamma,s} = \frac{q_{\mathrm{TL},s}}{L_i}, \qquad \Delta\rho_{q,s} = \frac{q_{\gamma,s}}{\partial\gamma_{\mathrm{P.676}}/\partial\rho_v}. \tag{11}$$

The P.676 sensitivity in Eq. (11) is evaluated for each sublink frequency at $T = 20\,^{\circ}\mathrm{C}$, $p = 980$ hPa, and $\rho_v = 10\ \mathrm{g\,m^{-3}}$.

The main comparison combines the 71–76 and 81–86 GHz allocations and uses CMLs between 4 and 10 km long and directional sublinks with $\Delta\rho_q \leq 0.20\ \mathrm{g\,m^{-3}}$. The threshold is specific to this dataset and is not intended as a universal criterion. After these criteria are applied, the comparison contains 31 physical CMLs and 45 directional sublinks.

4.5 Matching with reference observations

Radar rainfall is assigned to each CML path by the length weighted average of Eq. (1) (Sect. 3.2).

For the gauge comparison in Fig. 6, each physical CML is assigned to the nearest CHMI rain gauge within 5 km of its midpoint. The CHMI element `SRA10M`, stored as `rr` in the released NetCDF, contains 10 min precipitation amounts in millimetres, which are converted to equivalent mean rain rates by multiplying by 6. The CHMI quality codes and flags of the precipitation record are not used; all finite amounts are retained. Each timestamp marks the end of the preceding 10 min interval. For this benchmark, the released 1 min directional rain rates are first aggregated into 5 min means, which require five valid samples (Sect. 4.3). The corresponding 10 min CML value is the mean of the two derived 5 min estimates beginning at $t - 10$ and $t - 5$. Both 5 min estimates must be finite, and the interval must contain a complete valid 10 min wet–dry classification. No temporal interpolation, filling, or empirical time lag is applied. The two directional sublinks are evaluated independently. Only gauge wet intervals with an equivalent rain rate of at least 0.1 $\mathrm{mm\,h^{-1}}$ are included in the rainfall comparison, and rain rates are limited to 0–150 $\mathrm{mm\,h^{-1}}$; gauge dry intervals enter only the detection evaluation in Fig. 5. The gauge amounts have a resolution of 0.1 mm per 10 min, so this wet criterion selects positive recorded precipitation amounts; the smallest positive equivalent rate is 0.6 $\mathrm{mm\,h^{-1}}$.

For the WVD comparison, each CML is matched to the nearest CHMI station with both relative humidity and air temperature within 5 km of its path. Two consecutive valid 5 min CML estimates are averaged and compared with the station reference of Eq. (10) at the end of the same 10 min interval.

4.6 Evaluation metrics

Rain detection is evaluated directly from the contingency table of hits, false alarms, misses, and correct negatives shown in Fig. 5. Rainfall estimation and WVD are evaluated using root mean square error (RMSE), mean bias error (MBE), and the Pearson correlation coefficient (PCC),

https://doi.org/10.5194/essd-2026-939
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.



$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(\hat{x}_j - x_j)^2}, \tag{12}$$

$$\mathrm{MBE} = \frac{1}{N}\sum_{j=1}^{N}(\hat{x}_j - x_j), \tag{13}$$

$$\mathrm{PCC} = \frac{\sum_j(\hat{x}_j - \overline{\hat{x}})(x_j - \bar{x})}{\sqrt{\sum_j(\hat{x}_j - \overline{\hat{x}})^2 \sum_j(x_j - \bar{x})^2}}, \tag{14}$$

where $\hat{x}_j$ is the CML estimate and x_j the reference observation of pair j, N is the number of pairs, and $\overline{\hat{x}}$ and $\bar{x}$ are the means of the CML estimates and of the reference observations. Thus, the reported bias is defined as CML minus reference.

4.7 Rainfall field reconstruction

For one event, CML path rainfall estimates were assigned to link midpoints and interpolated by inverse distance weighting with power $\alpha = 2$. The event figure uses rolling 1 h rainfall accumulations formed from 12 consecutive 5 min values derived from the 1 min rainfall estimates, with complete 5 min support required. The CML reconstruction is compared with the native two-dimensional RADKLIM YW field accumulated over the same 1 h windows; the radar field is not reconstructed from the path averaged radar product.

5 Results

5.1 Multisensor event time series

Figure 4 shows the response of a representative E-band link during a rainfall event. Before the event, both directional sublinks remain close to a stable dry level of about 35 dB. Around the onset of rainfall, the total signal loss rises by more than 30 dB in both directions, reaching about 70–80 dB in three successive peaks. This shows how strongly an E-band link responds to intense rain, with retrieved CML rain rates reaching about 100 mm h^{-1}. The two sublinks show very similar timing, although their peak magnitudes differ slightly.

The CML rain rate follows the main timing of the radar and gauge observations, with the strongest attenuation occurring during the period of highest rain rate. The CML peak rates exceed the radar and gauge values, which may result from the different spatial and temporal sampling of a short path, a 1 km radar pixel and a point gauge.

Relative humidity at the nearest station rises from about 55–60 % before the event to above 95 % during the first rainfall period and remains near saturation afterwards, so panel (d) marks the onset of the event but gives no information on the later rain periods. Air temperature falls from about 29 to 21 °C at the onset of rainfall and continues to decrease to about 17 °C overnight, including during the second, weaker rain period. The station observations place the event in its atmospheric context but do not change the interpretation of the signal.

https://doi.org/10.5194/essd-2026-939
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

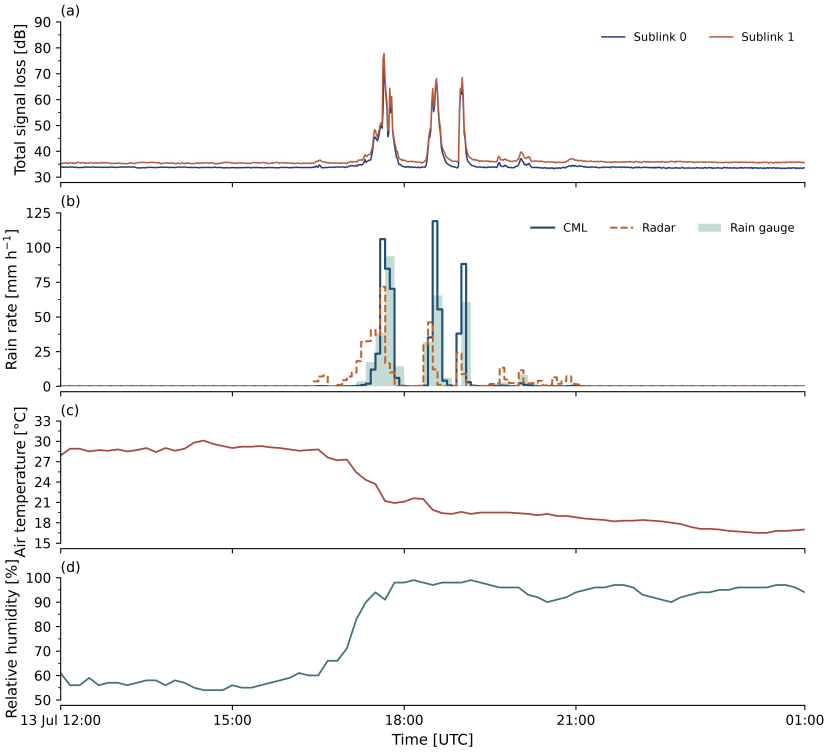

Figure 4. Multisensor time series for a representative E-band CML during the 13–14 July 2021 rainfall event. Panel (a) shows 1 min total signal loss for the two directional sublinks. Panel (b) compares the 5 min CML rain rates aggregated from the released 1 min estimates and path averaged radar rain rates with the rain rate of the nearest rain gauge (10 min precipitation amount). Panels (c) and (d) show air temperature and relative humidity from the nearest available stations. Times are in UTC.

5.2 CML rain detection

Figure 5 shows that the detector agrees with the gauge for most dry intervals (88.6 % correct negatives) and finds about three quarters (74.6 %) of the gauge wet intervals. The missed intervals are expected mainly in light rain. The smallest positive gauge amount corresponds to 0.6 mm h^{-1}, which can still fall below the minimum detectable rain rate of short links and contribute to missed detections. False alarms affect 11.4 % of the gauge dry intervals and are about four times more numerous than hits. The detector can also respond to signal variability unrelated to rain, for example on short and strongly quantized E-band links, and a point gauge can record no rain while rain affects another part of the link path. The large number of dry intervals in the April to October record also raises the absolute number of false alarms. Because gauge dry intervals are excluded from Fig. 6, these false alarms do not appear in the rainfall comparison, but the same mechanisms contribute to the CML values at very low gauge rates discussed below.

https://doi.org/10.5194/essd-2026-939
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

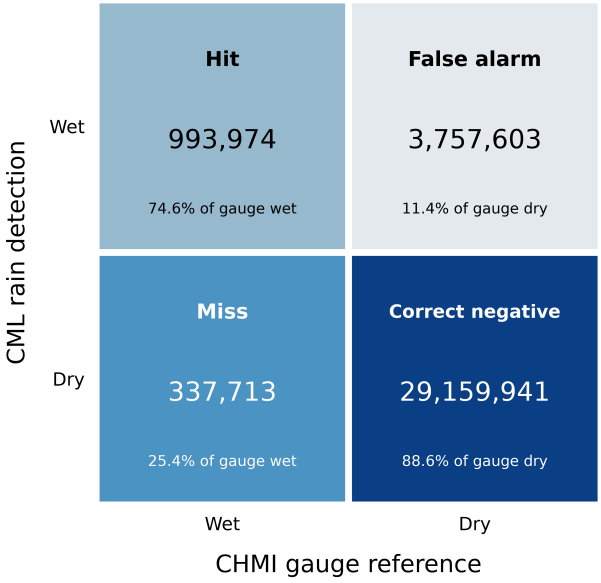

Figure 5. Contingency table for rain detection from CML data only against the nearest CHMI rain gauge for April–October 2021 after aggregation of the 1 min wet–dry flags to the 10 min gauge intervals. Each CML is matched to the nearest gauge within 5 km of its midpoint, as in Fig. 6. Gauge intervals with an equivalent rain rate of at least 0.1 mm h^{-1} are treated as wet. Directional sublinks are evaluated independently and only paired valid intervals are included. Percentages are given relative to the gauge wet and gauge dry totals.

The method of Graf et al. (2020) is therefore useful as a reproducible baseline that uses only CML data, rather than as a final classifier for every link. In the released rainfall product it is used consistently across the quality controlled CML population, while the gauges and, where available, radar are retained as independent references.

5.3 Rainfall estimation

The gauge comparison is independent because gauges are not used in the retrieval. Figure 6 shows the 1 331 687 matched gauge wet intervals of the directional sublinks. Gauge dry intervals are excluded, so the RMSE of 3.63 mm h^{-1}, the bias of -0.91 mm h^{-1}, and the PCC of 0.63 describe rainfall periods rather than the whole season. The densest part of the cloud lies at low rain rates, where most of the wet record is found. Below about 10 mm h^{-1} the pairs lie close to the 1:1 line, so light and moderate rain is retrieved without a clear bias. Above about 40 mm h^{-1} the pairs lie mainly below the 1:1 line, so the highest gauge rain rates are underestimated, consistent with the negative overall bias. This may result from intense rain that covers only part of the link path and from loss of the received signal during heavy rain. Two further features are visible. A vertical band of pairs at very low gauge rates with CML values up to more than 100 mm h^{-1} may result from non rainfall signal fluctuations and from wet antenna attenuation that remains after rain has stopped at the gauge. A horizontal band of CML values near zero extends across the full gauge range; these are the misses of the detector in Fig. 5, for which the retrieval returns zero rainfall

https://doi.org/10.5194/essd-2026-939
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.



Open Access Earth System Science Data Discussions

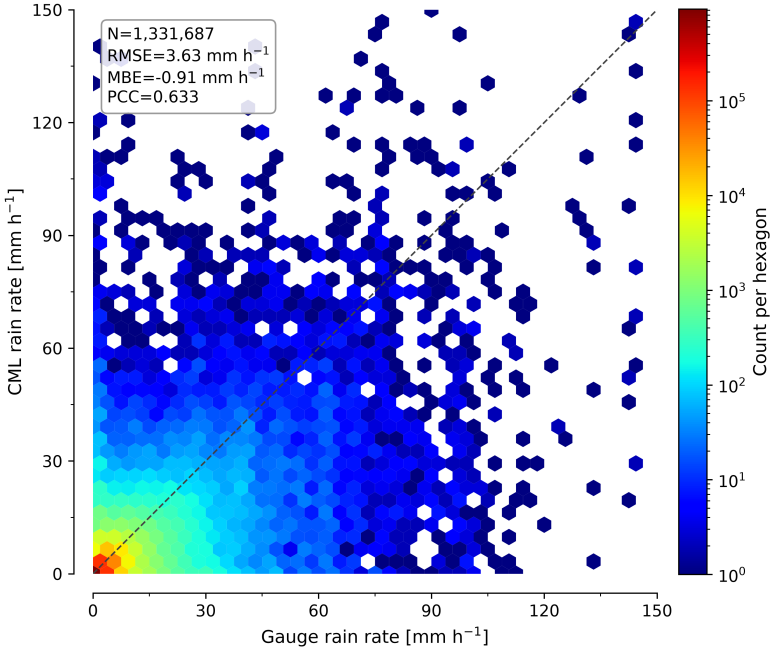

Figure 6. Validation of CML rain rates against the rain rates of CHMI rain gauges for April–October 2021. Each CML was matched to the nearest gauge within 5 km of its midpoint. The released 1 min CML estimates were aggregated to complete 5 min blocks, and two valid 5 min means were averaged to match each 10 min gauge interval, requiring a valid wet–dry classification. Directional sublinks were evaluated independently. Only gauge wet intervals ($\geq$ 0.1 mm h^{-1}, corresponding to positive 10 min precipitation amounts) are shown. The dashed line is the 1:1 line.

while the gauge reports rain. The spread in both directions partly reflects the difference between a point gauge and rainfall integrated along a path.

When the 3689 matched gauge wet intervals that contain at least one minute with reconstructed total signal loss, that is, a loss of the received signal during heavy rain (Sect. 4.1), are excluded, 1 327 998 intervals remain. The RMSE then decreases noticeably from 3.63 to 3.38 mm h^{-1}, while the bias (-0.89 instead of -0.91 mm h^{-1}) and the PCC (0.61 instead of 0.63) change only slightly. The reconstructed intervals are mostly periods of heavy rain at the gauge. The reconstruction keeps these periods in the record, but the CML rain rates have large errors there, and because the rain rates are high, these few intervals increase the RMSE noticeably.

5.4 Rainfall fields

Figure 7 shows the rainfall system developing from fragmented rainfall into a broader event. The native RADKLIM field identifies a more spatially continuous structure, whereas the CML reconstruction is more localized and reflects the uneven geometry of the telecommunication network. At 18:05 and 19:05 UTC, the CML field shows heavy rainfall in the same parts of the domain as the radar, but as separate maxima around single link midpoints, with higher peak rainfall accumulations

https://doi.org/10.5194/essd-2026-939
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

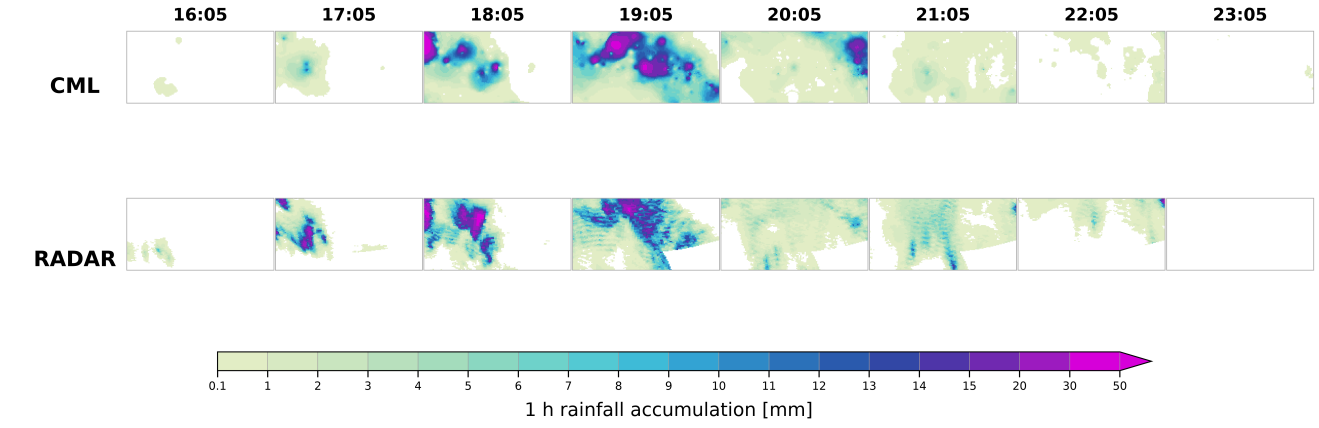

Figure 7. Rainfall fields on 13 July 2021, with panels from 16:05 to 23:05 UTC. The upper row shows the CML field reconstructed from rainfall estimates at link midpoints (aggregated from the released 1 min series) by inverse distance weighting, and the lower row shows the native RADKLIM YW field. Columns show rolling 1 h rainfall accumulations ending at the labelled hourly panel times.

than the radar. The CML field also leaves blank areas where no links exist. The broad timing and movement agree, but the difference in spatial continuity shows why CML rainfall fields are best treated as complementary observations or as input to merging rather than as direct replacements for radar.

5.5 Water vapour density evaluation

Figure 8 shows the station comparison for the selected E-band subset. The densest part of the cloud follows the 1:1 line between about 4 and 12 g m^{-3}, so the retrieval tracks the station reference in the most common conditions. Above about 12 g m^{-3} the CML values spread upward to 25 g m^{-3} while the station reference stays below about 18 g m^{-3}, so moist conditions are overestimated by some sublinks. A band of CML values near zero across the whole station range is also visible. In these periods the signal remains at its dry baseline level although the station reports a change in humidity, which may be related to the baseline estimation. The station reference covers only about 3 to 18 g m^{-3}, so the comparison says nothing about very dry conditions. The result is therefore a demonstration of the water vapour information contained in the selected E-band observations, not an optimized operational retrieval. A pressure sensitivity check using 930 and 1030 hPa changed the mean retrieved WVD by approximately -0.33 and $+0.32$ g m^{-3}, respectively, relative to 980 hPa. Table S13 in the Supplement compares WVD estimates by frequency, path length, and quantization resolution.

Figure 9 shows the temporal behaviour of the two observations for one sublink of the selected population over 7 days. The example was selected for continuity and temporal variation, not for agreement with the station: of all windows of 7 days of the 45 sublinks with a CML estimate at no less than 90 % of the 10 min steps and no gap longer than 24 h, it has the largest variability of the station WVD. The daily mean of the station reference rises from about 6.6 g m^{-3} on 13 June to about 11 g m^{-3} on 16–18 June, and the CML estimate follows this change, with a mean difference of 0.1 g m^{-3} over the 7 days. Within single

https://doi.org/10.5194/essd-2026-939
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.



Open Access Earth System Science Data Discussions

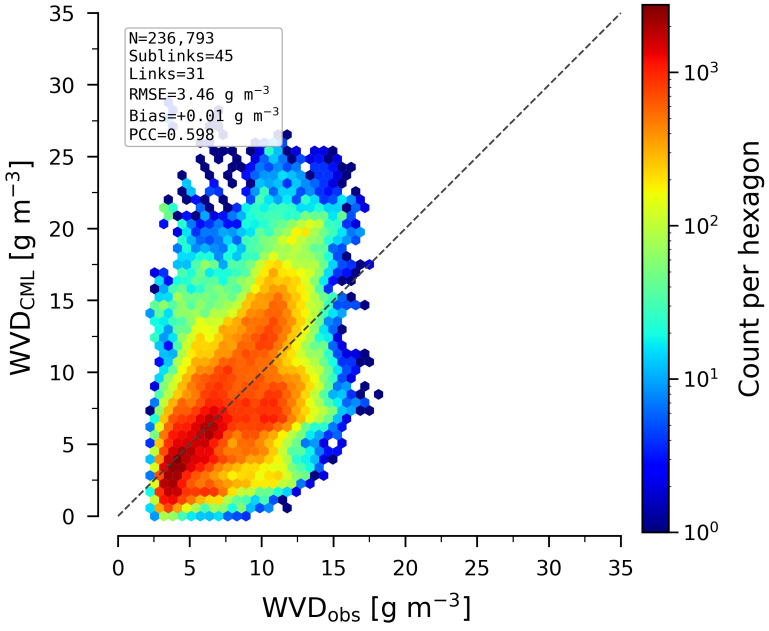

Figure 8. CML water vapour density compared with the station reference for the selected subset of 31 physical CMLs and 45 directional sublinks. The comparison combines the 71–76 and 81–86 GHz allocations and applies the path length and quantization criteria described in Sect. 4.4.

380 days, the two series differ by up to about 4 g m^{-3}, for example on 14 June. Over all its pairs, this sublink agrees better with the station than most of the 45 sublinks (RMSE 1.64 g m^{-3}; median of the 45 sublinks 2.74 g m^{-3}). The gaps are mostly periods classified as wet, including the 6 h buffer around them, in which no WVD is retrieved (Sect. 4.4). One sublink and one period do not represent the whole E-band population.

https://doi.org/10.5194/essd-2026-939
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

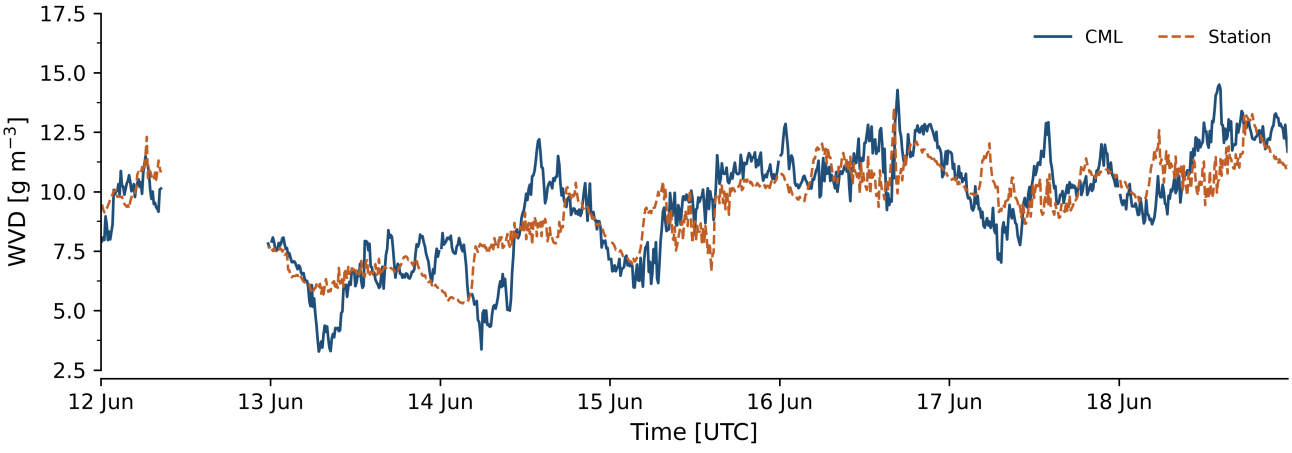

Figure 9. Water vapour density estimated from the CML and the matched CHMI station reference from 12 to 18 June 2021 (CML `85108_88054_3061`, sublink 0). Both series show matched 10 min observations. Gaps mainly reflect excluded wet periods and their 6 h buffer, with some missing signal observations.

6 Discussion

The results show both the value and the limits of rainfall retrieval from a heterogeneous operational CML network. Rain detection with the method of Graf et al. (2020) captures a substantial part of gauge observed wet periods but also produces misses and false alarms, so the detected wet periods should not be used without further screening. False alarms can add rainfall in periods that are dry. Other methods, for example those using nearby CMLs or weather radar, may improve the identification of wet and dry periods (Overeem et al., 2011, 2016b). The rainfall estimates agree with nearby gauges at light and moderate rain rates, while the largest gauge peaks are more often underestimated. Part of the spread is expected because gauges are point observations whereas CMLs integrate attenuation along a path. Strong attenuation can also lead to loss of the received signal during intense rainfall, removing observations when rainfall is highest (Chwala and Kunstmann, 2019; Polz et al., 2023). The reconstruction of total signal loss during such gaps keeps heavy rain periods in the record, but it assumes the highest TSL and the lowest RSL of the sublink, so the reconstructed rain rates are uncertain and can also appear when the gauge records no rain. Very short links can give unrealistically high rain rates, above 200 mm h^{-1} at 1 min resolution for some sublinks, most of them on CMLs shorter than about 1 km and often when the received signal is close to the receiver noise floor. These values are kept in the product and are not filtered. Depending on the application, users may therefore apply a stricter quality control.

The dataset connects the original operational signals with their downstream use. Raw and quality controlled signal layers allow users to test alternative screening and retrieval methods, while the released rainfall product provides a direct starting point for users mainly interested in rainfall mapping, merging, or hydrological applications. The rainfall retrieval uses only CML signal information for wet–dry classification, so both the gauges and the radar remain independent, and the radar is available for evaluation where the RADKLIM domain overlaps the network. This matters because wet–dry classification,

https://doi.org/10.5194/essd-2026-939
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.



baseline estimation, wet antenna correction, and the attenuation–rainfall relation remain major sources of uncertainty in CML rainfall retrieval (Schleiss and Berne, 2010; Overeem et al., 2011; Polz et al., 2020; Graf et al., 2020). Publishing the processing levels together keeps these choices visible instead of hiding them inside one final rainfall product.

The E-band population adds another use of the dataset. Earlier studies showed that operational E-band links are sensitive to atmospheric water vapour as well as rainfall (Fencl et al., 2020, 2021). Here, the WVD comparison is restricted to 31 physical CMLs and 45 directional sublinks that satisfy path length and quantization based observability criteria. The agreement with the station reference shows that useful water vapour information is present in this selected population, but the result should not be generalized to all E-band links in the dataset. Signal quantization, path length, hardware stability, baseline estimation, and the representativeness of nearby stations remain important limitations. The $\Delta\rho_q$ threshold used here is therefore specific to this dataset.

The dataset also has spatial limitations. Radar coverage is concentrated mainly in the western part of the Czech Republic, and CML density and orientation follow communication needs and the part of the operator network available to the research data acquisition, rather than meteorological sampling. The network contains a wide range of frequencies, path lengths, hardware types, and local environments, so one processing configuration cannot be expected to perform equally well for every link (Graf et al., 2020; Roversi et al., 2020). These limitations reduce uniformity across the dataset, but they are typical of real operational telecommunication networks. The dataset allows users to work with the same network from raw signals through quality control to rainfall estimates, with independent reference observations for testing and further method development.

7 Conclusions

This paper presents an open multisource hydrometeorological dataset for the Czech Republic covering 1 April–31 October 2021. The dataset contains 2574 raw CMLs, 1987 quality controlled CMLs, path averaged radar rainfall for 1132 CML paths, CML rainfall for all 1987 quality controlled links, precipitation from 424 stations, air temperature from 251 stations, and relative humidity from 247 stations.

The dataset supports two complementary uses. Raw and quality controlled signals allow independent development of CML processing methods, while the rainfall product provides application ready path rainfall without requiring users to implement the full retrieval chain. The large E-band component and the accompanying meteorological observations also support high frequency propagation studies, shown here by the WVD retrieval for a selected subset of 31 physical CMLs and 45 directional sublinks.

The dataset therefore brings together operational telecommunication signals, transparent quality control, matched reference observations, and rainfall estimates in one reproducible release.

https://doi.org/10.5194/essd-2026-939
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.



8 Code and data availability

The dataset is available on Zenodo at https://doi.org/10.5281/zenodo.23398309 (Shit et al., 2026). During the review, the record is an unpublished draft. Access to the data has been granted to reviewers and editors through the anonymous Zenodo review link. For members of the public interested in participating in the public discussion, the data are available upon request (contact: taoufiq.shit@fsv.cvut.cz) during the review period. The record contains seven NetCDF products covering 1 April to 31 October 2021, three supporting files for the WVD analysis, and the data documentation, each stored as a separate file. The seven products contain raw CML observations, quality controlled CML observations, 1 min CML rainfall estimates, path averaged RADKLIM rainfall, and CHMI observations of precipitation, air temperature, and relative humidity. The 5 min CML rain rates used in the benchmark analyses are derived from the released 1 min rainfall estimates and are not provided as a separate data product.

The processing code and notebooks are available at https://github.com/toufikshit/essd-czech-cml-2021. The repository includes the workflows for CML data preparation and quality control, rainfall retrieval, aggregation from 1 to 5 min, evaluation against rain gauges and radar, rainfall field reconstruction, WVD analysis, and figure generation. Instructions for using the code and setting the data paths are provided in the repository README and notebooks. The file structures, variables, units, and quality fields are described in the Supplement and data dictionary.

The CML data products are distributed under the Creative Commons Attribution–NonCommercial 4.0 International licence (CC BY-NC 4.0), with permission from T-Mobile Czech Republic a.s. The licence requires attribution and restricts commercial reuse. The CHMI and DWD derived products are distributed under CC BY 4.0 with the appropriate source attribution. The licence for the code is stated separately in the GitHub repository. The native RADKLIM grids and original external source archives are not redistributed.

Author contributions. TS prepared the data, code, analysis, figures, and first manuscript draft. MF and VB supervised the work and contributed to dataset design, scientific interpretation, and manuscript revision. MM contributed to data management.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The authors thank T-Mobile Czech Republic a.s. for providing the commercial microwave link observations and for permitting their release for non-commercial use. The Czech Hydrometeorological Institute is acknowledged for providing precipitation, air temperature, and relative humidity observations through its open data system. The German Weather Service is thanked for providing the RADKLIM YW precipitation product. This work was supported by the Czech Science Foundation (GAČR), Czech Republic, under Grant No. 24-13677L (MERGOSAT).

https://doi.org/10.5194/essd-2026-939
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.



References

Andersson, J. C. M., Olsson, J., van de Beek, R. C. Z., 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.

Blettner, N., Fencl, M., Bareš, V., Kunstmann, H., and Chwala, C.: Transboundary Rainfall Estimation Using Commercial Microwave Links, Earth and Space Science, 10, e2023EA002 869, https://doi.org/10.1029/2023EA002869, 2023.

Chwala, C. and Kunstmann, H.: Commercial microwave link networks for rainfall observation: Assessment of the current status and future challenges, WIREs Water, 6, e1337, https://doi.org/10.1002/wat2.1337, 2019.

Covi, E., Roversi, G., and Nebuloni, R.: OpenRainER: an open-source dataset for studying the opportunistic sensing of rainfall in Emilia-Romagna, Italy, Earth System Science Data Discussions, https://doi.org/10.5194/essd-2026-160, preprint, accepted for publication in Earth System Science Data, 2026.

Czech Hydrometeorological Institute: Annual Report of the Czech Hydrometeorological Institute 2015, Tech. rep., Czech Hydrometeorological Institute, Prague, Czech Republic, ISBN 978-80-87577-59-2, https://www.chmi.cz/w/vyrocni-zprava-chmu-2015, 2016.

Czech Hydrometeorological Institute: Annual Report of the Czech Hydrometeorological Institute 2016, Tech. rep., Czech Hydrometeorological Institute, Prague, Czech Republic, ISBN 978-80-87577-70-7, https://www.chmi.cz/w/vyrocni-zprava-chmu-2016, 2017.

Czech Hydrometeorological Institute: Klimatologická data: popis dat na serveru opendata.chmi.cz, Technical documentation, https://opendata.chmi.cz/meteorology/climate/Klimatologicka_data_popis.pdf, 2025.

del Portillo, I.: ITU-Rpy: A Python implementation of the ITU-R P. Recommendations to compute atmospheric attenuation in slant and horizontal paths, https://github.com/inigodelportillo/ITU-Rpy, accessed: 12 October 2026, 2017.

Fabry, F.: Radar Meteorology: Principles and Practice, Cambridge University Press, https://doi.org/10.1017/CBO9781107707405, 2015.

Fencl, M., Dohnal, M., Rieckermann, J., and Bareš, V.: Atmospheric observations with E-band microwave links: challenges and opportunities, Atmospheric Measurement Techniques, 13, 6559–6578, https://doi.org/10.5194/amt-13-6559-2020, 2020.

Fencl, M., Dohnal, M., and Barěs, V.: Retrieving Water Vapor From an E-Band Microwave Link With an Empirical Model Not Requiring In Situ Calibration, Earth and Space Science, 8, e2021EA001 911, https://doi.org/10.1029/2021EA001911, 2021.

Graf, M., Chwala, C., Polz, J., and Kunstmann, H.: Rainfall estimation from a German-wide commercial microwave link network: optimized processing and validation for 1 year of data, Hydrology and Earth System Sciences, 24, 2931–2950, https://doi.org/10.5194/hess-24-2931-2020, 2020.

Hingray, B., Picouet, C., and Musy, A.: Hydrology: A Science for Engineers, CRC Press, Taylor & Francis Group, https://doi.org/10.1201/b17169, 2014.

International Telecommunication Union: Specific attenuation model for rain for use in prediction methods, Recommendation ITU-R P.838-3, ITU Radiocommunication Sector, Geneva, Switzerland, https://www.itu.int/rec/R-REC-P.838-3-200503-I/en, 2005.

International Telecommunication Union: Attenuation by atmospheric gases and related effects, Recommendation ITU-R P.676-12, ITU Radiocommunication Sector, Geneva, Switzerland, https://www.itu.int/rec/R-REC-P.676-12-201908-S/en, 2019.

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, 18, 5817–5836, https://doi.org/10.5194/essd-18-5817-2026, 2026.

https://doi.org/10.5194/essd-2026-939
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.



Messer, H., Zinevich, A., and Alpert, P.: Environmental Monitoring by Wireless Communication Networks, Science, 312, 713, https://doi.org/10.1126/science.1120034, 2006.

Olsson, J., Horváth-Varga, L., van de Beek, R., Graf, M., Overeem, A., Szaton, M., Bareš, V., Bezak, N., Chwala, C., De Michele, C., Fencl, M., Seidel, J., and Todorović, A.: How close are opportunistic rainfall observations to providing societal benefit?, Journal of Hydrometeorology, 26, 1585–1602, https://doi.org/10.1175/JHM-D-25-0043.1, 2025.

Overeem, A., Leijnse, H., and Uijlenhoet, R.: Measuring urban rainfall using microwave links from commercial cellular communication networks, Water Resources Research, 47, W12 505, https://doi.org/10.1029/2010WR010350, 2011.

Overeem, A., Leijnse, H., and Uijlenhoet, R.: Two and a half years of country-wide rainfall maps using radio links from commercial cellular telecommunication networks, Water Resources Research, 52, 8039–8065, https://doi.org/10.1002/2016WR019412, 2016a.

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, 2016b.

Overeem, A., Walraven, B., Leijnse, H., and Uijlenhoet, R.: Four year commercial microwave link dataset for the Netherlands, 4TU.ResearchData [data set], https://doi.org/10.4121/be252844-b672-471e-8d69-27269a862ec1.v1, 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.

Polz, J., Chwala, C., Graf, M., and Kunstmann, H.: Rain event detection in commercial microwave link attenuation data using convolutional neural networks, Atmospheric Measurement Techniques, 13, 3835–3853, https://doi.org/10.5194/amt-13-3835-2020, 2020.

Polz, J., Graf, M., and Chwala, C.: Missing Rainfall Extremes in Commercial Microwave Link Data Due To Complete Loss of Signal, Earth and Space Science, 10, e2022EA002 456, https://doi.org/10.1029/2022EA002456, 2023.

Roversi, G., Alberoni, P. P., Fornasiero, A., and Porcù, 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, IEEE Geoscience and Remote Sensing Letters, 7, 611–615, https://doi.org/10.1109/LGRS.2010.2043052, 2010.

Shit, T., Fencl, M., Mudroch, M., and Bareš, V.: Operational microwave links for hydrometeorological sensing: a multisource dataset for the Czech Republic (April to October 2021), Zenodo [data set], https://doi.org/10.5281/zenodo.23398309, 2026.

Špačková, A., Bareš, V., Fencl, M., Schleiss, M., Jaffrain, J., Berne, A., and Rieckermann, J.: A year of attenuation data from a commercial dual-polarized duplex microwave link with concurrent disdrometer, rain gauge, and weather observations, Earth System Science Data, 13, 4219–4240, https://doi.org/10.5194/essd-13-4219-2021, 2021.

Tolasz, R., ed.: Climate Atlas of Czechia, Czech Hydrometeorological Institute, Prague, ISBN 978-80-86690-26-1, 2007.

Winterrath, T., Brendel, C., Hafer, M., Junghänel, T., Klameth, A., Lengfeld, K., Walawender, E., Weigl, E., and Becker, A.: RADKLIM Version 2017.002: Reprocessed quasi gauge-adjusted radar data, 5-minute precipitation rates (YW), Deutscher Wetterdienst [data set], https://doi.org/10.5676/DWD/RADKLIM_YW_V2017.002, corrected 2021 NetCDF files from the DWD supplement were used, 2018.