S1 Scope of the Supplement

The main paper describes the scientific motivation, processing logic, and role of each released product. This Supplement provides the technical information needed to inspect, join, and reuse the files: exact file names, dimensions, coordinates, variables, units, quality fields, acquisition limitations, retrieval settings, benchmark comparison, and repository organization.

Commercial microwave links are abbreviated as CMLs, the Czech Hydrometeorological Institute as CHMI, the German Weather Service as DWD, transmitted and received signal levels as TSL and RSL, wet antenna attenuation as WAA, water vapour density as WVD, root mean square error as RMSE, mean bias error as MBE, and the Pearson correlation coefficient as PCC. The released products cover 1 April to 31 October 2021.

S2 File inventory

Table S1: The seven released NetCDF files. All files cover 1 April to 31 October 2021 (UTC).

Product (number)	Filename	Step
Raw CML observations (2574 CMLs)	`CZ_CML_RAW_1MIN_20210401_20211031.nc`	1 min
Quality controlled CML observations (1987 CMLs)	`CZ_CML_PREPROCESSED_1MIN_20210401_20211031.nc`	1 min
Radar path averages (1132 paths)	`CZ_RADAR_CMLPATH_5MIN_20210401_20211031.nc`	5 min
CML rainfall estimates (1987 CMLs)	`CZ_CML_RAINRATE_GRAF_ITU2005_PASTOREK_1MIN_20210401_20211031.nc`	1 min
Ground precipitation (424 stations)	`CZ_RAIN_GAUGE_10MIN_20210401_20211031.nc`	10 min
Air temperature (251 stations)	`CHMI_10MIN_TEMPERATURE_20210401_20211031.nc`	10 min
Relative humidity (247 stations)	`CHMI_10MIN_HUMIDITY_20210401_20211031.nc`	10 min

The raw and quality controlled CML files use the dimensions `time`, `sublink_id`, and `cml_id`. The rainfall file uses `time_1min`, `sublink_id`, and `cml_id`. The radar file uses `time` and `cml_id`. The temperature and humidity files use `station` and `time`; the precipitation file uses `WSI` and `time`. The common interval contains 308160 time slots at 1 min, 61632 at 5 min, or 30816 at 10 min. The time coordinates use different reference dates in different files (for example minutes since 2020-01-01 or seconds since 1970-01-01), so they should always be decoded with their `units` attribute. The released rainfall file is the native 1 min product. The 5 min CML estimates used for the manuscript analyses are derived from it and are not released as an eighth NetCDF file.

S3 Repository organization

The data are distributed through the Zenodo dataset record, not through the GitHub code repository. The repository contains the Python scripts and notebooks used to process the data and produce the figures. The README explains how to install the required packages, set the data paths, and run the code. The simplified repository organization is:

```
essd-czech-cml-2021/
  README.md
  LICENSE               # MIT
  CITATION.cff
  THIRD_PARTY_NOTICES.md
  pyproject.toml
  requirements.txt
  config/               # user_paths.example.json
```

```
data/                # download instructions only
scripts/
src/essd_cz/
notebooks/
original/            # preserved original scripts
docs/
tests/
```

The seven NetCDF files can be stored in any local directory, with their paths specified once in `config/user_paths.json`. The notebooks use the configured paths. The WVD analysis can additionally use the three supporting files described in the data README.

S4 Commercial microwave link products

S4.1 Dimensions and coordinates

The raw and quality controlled CML products use the dimensions `time`, `sublink_id`, and `cml_id`. A physical CML generally contains two directional sublinks. Time is stored in UTC. The same `cml_id` values are used to join the radar path and rainfall products for the links retained in those files.

Table S2: Coordinates and link metadata in the raw and quality controlled CML files. The raw file stores the same quantities without units attributes; the units below apply to both files.

Name	Dimensions	Unit	Definition
`time`	time	UTC	Observation time at 1 min resolution
`cml_id`	cml_id	none	Unique physical link identifier
`sublink_id`	sublink_id	none	Directional sublink index
`site_a_latitude`, `site_a_longitude`	cml_id	degrees	First link endpoint
`site_b_latitude`, `site_b_longitude`	cml_id	degrees	Second link endpoint
`length`	cml_id	km	Radio path length
`frequency`	sublink_id, cml_id	MHz	Directional carrier frequency
`polarization`	sublink_id, cml_id	none	Directional polarization (character array)
`qc_detection_limit_mmhr`	sublink_id, cml_id	mm h^{-1}	Minimum detectable rain rate, quality controlled file only; one value per directional sublink, calculated with ITU-R P.838-3 (2005) coefficients

S4.2 Signal variables

The raw file contains 2574 CMLs and the quality controlled file 1987 CMLs. In the quality controlled file, 3731 directional sublinks pass quality control; the other sublinks of these CMLs keep missing signal values, and `qc_sublink_status` gives the reason. The quality control covers 1 April to 31 October 2021 and follows Blettner et al. (2023), with the changes and parameters given in the main paper. All parameters are stored in the global attribute `qc_parameters_json` of the quality controlled file. The bits of `qc_flag` are: 1, invalid value removed; 2, minute with mixed valid and placeholder source values removed; 4, low RSL plateau removed; 8, total signal loss reconstructed during a blackout (the RSL is missing); 16, total signal loss interpolated over a short gap. At minutes with bit 8 or 16, `trsl` is not equal to `tx` minus `rx`.

Table S3: Signal and quality variables in the raw and quality controlled CML files. `trsl` and the quality variables are stored only in the quality controlled file.

Name	Dimensions	Unit	Definition
`tx`	time, sublink_id, cml_id	dBm	Transmitted signal level (TSL)
`rx`	time, sublink_id, cml_id	dBm	Received signal level (RSL), screened in the quality controlled file
`trsl`	time, sublink_id, cml_id	dB	Total signal loss, transmitted minus received signal level; reconstructed or interpolated where `qc_flag` bit 8 or 16 is set
`qc_flag`	time, sublink_id, cml_id	flag	Quality flag of each minute (bits listed in Sect. S4.2)
`qc_sublink_status`	sublink_id, cml_id	flag	0 when the sublink passes quality control, otherwise the reason for its exclusion

S5 Path averaged radar product

The radar path file contains the 1132 CMLs for which path averaged RADKLIM rainfall was generated. Of these paths, 1033 belong to quality controlled CMLs. In addition to the precipitation variable, it stores the CML identifiers, endpoint coordinates, and path lengths, so it can be opened and mapped on its own. Frequency and polarization are not stored in this file; `cml_id` is the key for joining it with the CML and rainfall files.

Missing values in `rr` indicate that no valid radar path average is available and must not be converted to zero. The native RADKLIM grid is not included. The source is the corrected 2021 RADKLIM YW NetCDF files provided by DWD, in which the precipitation variable is named `RR`; the released path averages are stored as `rr`. The stored `rr` (units attribute kg m^{-2}, equal to mm) is the precipitation *amount* per 5 min, not a rain rate in mm h^{-1}. For a 5 min equivalent mean rate, multiply the stored amount by 12.

Table S4: Selected coordinates, metadata and variables in the radar path product.

Name	Dimensions	Unit	Definition
`time`	time	UTC	Time slots at 5 min
`cml_id`	cml_id	none	Physical link identifier
`site_a_latitude`, `site_a_longitude`	cml_id	degrees	First link endpoint
`site_b_latitude`, `site_b_longitude`	cml_id	degrees	Second link endpoint
`length`	cml_id	km	Full path length
`rr`	time, cml_id	kg m^{-2} (mm per 5 min)	Path averaged RADKLIM precipitation amount

S6 Derived CML rainfall product

S6.1 Dimensions and principal variables

The released rainfall file contains 1 min estimates for each direction and the mean over valid directions. It uses the wet–dry classification of Graf et al. (2020), a constant dry baseline, the wet antenna correction of Pastorek et al. (2022), and the ITU-R P.838-3 coefficients from 2005. The file also stores the link metadata needed to use it on its own: `length` (km, per sublink), `frequency` (Hz),

`polarization`, and the endpoint coordinates. Frequency is stored in Hz in this file but in MHz in the signal files.

Table S5: Principal data variables in the published 1 min CML rainfall NetCDF. The two threshold variables have no units attribute; their values are in dB.

Name	Dimensions	Unit	Definition
`R_sublink_1min`	time_1min, sublink_id, cml_id	mm h^{-1}	Independent directional estimate
`R_mean_1min`	time_1min, cml_id	mm h^{-1}	Mean of valid directions at each minute
`graf_wet_1min`	time_1min, sublink_id, cml_id	flag	0 dry, 1 wet, 255 invalid
`graf_valid_1min`	time_1min, sublink_id, cml_id	flag	1 when the classification is valid and the rain rate is finite, else 0
`trsl_source_flag_1min`	time_1min, sublink_id, cml_id	flag	Bit 8: rain rate from total signal loss reconstructed during a blackout; bit 16: from interpolated total signal loss
`graf_threshold_db`	sublink_id, cml_id	dB	Graf threshold
`graf_rstd_q80_db`	sublink_id, cml_id	dB	Rolling standard deviation 80th percentile

S6.2 Retrieval configuration

Table S6 lists the settings used for the released rainfall product.

Table S6: Processing settings for the one published CML rainfall configuration.

Method or parameter	Value
Wet–dry classification	Graf et al. (2020); centred 60 min rolling standard deviation (RSTD); quantile 0.80 multiplied by 1.12
Valid observations per window	60 for threshold estimation; at least 45 for detection
Input signal	Quality controlled `trsl`, including reconstructed and interpolated values
Period for Graf thresholds and dry baseline	1 April–31 October 2021
Dry baseline	`pycomlink baseline_constant`, most recent valid dry observation
Rain coefficient set	ITU-R P.838-3, 2005
WAA correction	Pastorek, $A_{\max} = 14$ dB, $\zeta = 0.55$, $d = 0.1$
Temporal resolution of published rain rates	1 min
Reference observations used in retrieval	None (no radar or gauge input)

S6.3 Analysis at 5 min resolution and benchmark comparison

The main paper gauge comparison, event time series, and reconstructed rainfall fields use 5 min CML rainfall obtained from the released 1 min estimates. A 5 min directional value is the arithmetic mean of the five valid 1 min estimates in a clock aligned interval. If any of the five estimates is invalid, the 5 min value is missing. A 5 min wet flag is wet if any of its five valid minutes is wet. The mean for a physical CML is then calculated from whichever directional 5 min estimates are valid; it is not an average of the 1 min means of the CML.

The gauge validation in the main paper forms 10 min mean CML rain rates from two complete 5 min directional means and matches them to the 10 min CHMI precipitation amounts, which are labelled

at the end of each interval. Gauge precipitation in millimetres per ten minutes is multiplied by six for comparison in mm h^{-1}. The comparison uses nearest gauges within 5 km and gauge wet intervals. It is an independent evaluation, not a selection or calibration step.

Table S7: Verified 1 min release and its derived 5 min benchmark. All verified comparisons are reported in the repository evidence, not as a new sensitivity experiment.

Check	Verified result
Native released time series	308160 UTC slots at 1 min
CML population	1987 physical links; 3731 directional sublinks with rainfall
Finite directional 1 min rainfall values	631742984
Independent recomputation of the 1 min rainfall, Graf flags and thresholds	Identical
Derived 5 min directional rainfall and mean of each physical CML from the 1 min file	Identical
Gauge wet comparison in the main paper	$N = 1\,331\,687$; RMSE 3.63 mm h^{-1}; MBE -0.91 mm h^{-1}; PCC 0.63
Same comparison without intervals with reconstructed total signal loss	$N = 1\,327\,998$; RMSE 3.38 mm h^{-1}; MBE -0.89 mm h^{-1}; PCC 0.61

S6.4 Rain detection used by the rainfall product

The published rainfall file uses only CML signal variability to identify wet and dry periods with the method of Graf et al. (2020). For each directional sublink, the method uses a 60 min rolling standard deviation of total signal loss. Thresholds are estimated from the April–October 2021 record of each sublink. The resulting masks are used in dry baseline estimation and rainfall retrieval. Radar and gauges are not used for this classification and remain independent meteorological references. Missing observations are not reclassified as dry.

S7 CHMI ground station products

The temperature and humidity files use the dimensions `station` and `time`. The precipitation file uses `WSI` and `time`; `WSI` holds the same WIGOS station identifiers. Missing observations remain missing, and quality values and flags are kept as separate variables.

S7.1 Source and instrument provenance

The observations originate from the CHMI historical climatological open data archive (Czech Hydrometeorological Institute, 2025). CHMI operates a heterogeneous automated network that includes electronic weighing and tipping-bucket precipitation gauges, combined temperature–humidity probes, and resistance temperature sensors.

Official documentation lists instruments used in the wider network, including MRW500 weighing gauges, MR3H tipping-bucket gauges, HMP45 and HMP155 temperature–humidity probes, and PT100 temperature sensors. The public metadata do not link these models to the stations in the released 2021 products, so no instrument model is assigned to individual stations. Available sensor heights, descriptions of the measured elements, and the source quality information are kept.

S7.2 Coordinates shared by station products

Table S8 lists the coordinates and station metadata of the temperature and humidity files.

Table S8: Coordinates in the CHMI temperature and humidity files. The precipitation file uses `WSI` instead of `station` and `FULL_NAME` instead of `station_name` (Table S9).

Name	Dimensions	Unit	Definition
`station`	station	none	WIGOS station identifier
`time`	time	UTC	30816 time slots at 10 min, 1 April–31 October 2021
`lat`, `lon`	station	degrees	Station position
`elevation`	station	m	Elevation above mean sea level
`station_name`	station	none	CHMI station name
`GH_ID`	station	none	CHMI source identifier

S7.3 Precipitation product

The released precipitation product contains 424 stations and 30816 time slots at 10 min. Its variable `rr` is the CHMI element SRA10M, the precipitation amount over 10 min in millimetres, not mm h^{-1}. The variable has no units attribute in the file. CHMI quality information is stored in the variables `quality` and `flag`. In the source system, numeric quality codes 0, 1, 2, 3, 4 and 5 denote good, suspect, poor quality, estimated, missing and unknown quality, respectively. The CHMI flag Z denotes use of a backup precipitation gauge.

S7.4 Air temperature product

The temperature file contains 251 stations and 30816 time slots at 10 min. The source variables `T`, `TMA`, `TMI`, and `TPM` and their quality metadata are retained as supplied in the released temperature file.

S7.5 Relative humidity product

The relative humidity file contains 247 stations and 30816 time slots at 10 min. Relative humidity `H` is reported in percent with the original CHMI quality metadata.

Table S9: Variables in `CZ_RAIN_GAUGE_10MIN_20210401_20211031.nc`.

Name	Dimensions	Unit	Definition
`rr`	WSI, time	mm	Precipitation amount over 10 min (no units attribute in the file)
`quality`	WSI, time	source code	CHMI quality value (codes 0 to 5, Sect. S7.3)
`flag`	WSI, time	source code	CHMI flag
`WSI`	WSI	none	WIGOS station identifier
`lat`, `lon`, `elevation`	WSI	degrees, m	Station position and elevation
`FULL_NAME`, `GH_ID`	WSI	none	Station name and CHMI identifier

S8 Missing values and quality fields

Missing numerical observations in the station and radar path products remain missing and must not be interpreted as zero. CHMI quality values and flags are kept as separate variables instead of being replaced by one acceptance filter for the whole dataset. Users should define accepted quality states according to their application and report that selection.

Table S10: Selected variables of the CHMI temperature product.

Name	Dimensions	Unit	Definition
`T`	station, time	°C	Air temperature
`TMA`, `TMI`, `TPM`	station, time	°C	Supplementary CHMI temperature elements
`T_quality`, `T_flag`	station, time	source code	Quality information; the same variables exist for TMA, TMI, and TPM
`T_sensor_height`	station	m	Sensor height where supplied; also for TMA, TMI, and TPM

Table S11: Selected variables of the CHMI relative humidity product.

Name	Dimensions	Unit	Definition
`H`	station, time	%	Relative humidity
`H_quality`	station, time	source code	Original CHMI quality value
`H_flag`	station, time	none	Original CHMI flag
`H_sensor_height`	station	m	Sensor height where available

The raw and quality controlled CML products are separate files so users can select the processing level appropriate to their work. Users developing a new retrieval should state whether raw or quality controlled signals are used and whether directional sublinks are processed separately. In the rainfall file, a valid minute classified as dry can legitimately have zero estimated rainfall; missing or invalid minutes should not be treated as zero rainfall.

S9 Quality considerations for each product

Table S12 summarizes the main limitation and the recommended use of each product.

Table S12: Principal quality considerations for reuse of the released products.

Product	Main limitation	Recommended use
Raw CML signals	Receiver limits, gaps, plateaus, link disturbances	Alternative quality control and retrieval development
Quality controlled CML signals	Dependence on screening, gap treatment, and blackout reconstruction	Starting point for retrieval and propagation studies
Path averaged radar	Radar uncertainty, western coverage and missing paths	Independent spatial reference; missing is not zero
CML rainfall	Wet–dry classification, dry baseline, WAA, quantization, loss of signal in heavy rain, uncertain reconstructed values, and very high rain rates on some short links	Rainfall applications; keep directional and quality information; use `trsl_source_flag_1min` to exclude reconstructed values if needed
Ground precipitation	Point support, instrument types and source quality states	Independent rain gauge reference
Air temperature / relative humidity	Point rather than path measurement	Atmospheric context and reference for WVD studies

S9.1 WVD comparison by frequency, path length and quantization

Table S13 compares the WVD estimates with CHMI observations by frequency, path length, and

quantization resolution. From 1352 directional series, 587 have matched station observations, representing 295 physical CMLs and 2,834,230 pairs. The selected 31 physical CMLs and 45 directional sublinks used in Fig. 8 are shown separately. These groups differ in more than one property, so the results do not isolate the effect of frequency, length, or quantization.

Table S13: WVD compared with CHMI stations for the matched E-band population, with wet periods screened by the method of Graf et al. (2020). L is the path length. RMSE and bias are in g m^{-3}. Bias is CML minus CHMI. The results are calculated from the matched observations.

Group / class	Physical CMLs	Directional sublinks	Pairs	RMSE	Bias	PCC
All matched E-band data						
All eligible pairs	295	587	2,834,230	6.38	+2.41	0.413
Frequency allocation						
71–76 GHz	293	293	1,425,254	6.62	+2.55	0.402
81–86 GHz	294	294	1,408,976	6.12	+2.27	0.426
Path length						
$2 \leq L < 3$ km	74	148	745,692	8.02	+4.14	0.363
$3 \leq L < 4$ km	64	128	590,035	6.07	+2.99	0.433
$4 \leq L < 5$ km	79	157	703,283	6.26	+1.92	0.367
$5 \leq L < 6$ km	36	72	337,246	6.25	+2.05	0.450
$6 \leq L < 10$ km	42	82	457,974	3.26	-0.14	0.625
Quantization resolution $\Delta\rho_q$ (g m^{-3})						
< 0.20	31	45	236,793	3.46	+0.01	0.598
$0.20 \leq \Delta\rho_q < 0.25$	85	85	383,400	6.80	+2.07	0.398
$0.25 \leq \Delta\rho_q < 0.30$	114	114	475,884	6.12	+1.97	0.413
$0.30 \leq \Delta\rho_q < 0.40$	163	172	888,105	5.30	+1.85	0.472
$0.40 \leq \Delta\rho_q < 0.50$	95	95	485,488	6.61	+3.62	0.415
≥ 0.50	61	75	363,798	9.20	+4.64	0.325
Subset used in the main paper ($4 \leq L \leq 10$ km, $\Delta\rho_q \leq 0.20$ g m^{-3})						
Selected subset	**31**	**45**	**236,793**	**3.46**	**+0.01**	**0.598**

The metrics are calculated from all matched observations in each group. Physical CML counts cannot be added across frequency or quantization classes because the two directions can belong to different classes. The overall row includes 762 observations from one direction with no quantization estimate; this direction is omitted from the quantization groups.

References cited in the Supplement

- Blettner, N., Fencl, M., Bareš, V., Kunstmann, H., and Chwala, C.: Transboundary Rainfall Estimation Using Commercial Microwave Links, Earth and Space Science, 10, e2023EA002869, https://doi.org/10.1029/2023EA002869, 2023.
- 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.
- 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.
- 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.