Four-year commercial microwave link dataset for the Netherlands

Aart Overeem[1,2], Bas Walraven[2], Hidde Leijnse[1], Remko Uijlenhoet[2]

June 2024

[1]R&D Observations and Data Technology, Royal Netherlands Meteorological Institute, De Bilt, the Netherlands

[2]Department of Water Management, Delft University of Technology, Delft, the Netherlands

We present a dataset of commercial microwave link (CML) data for the Netherlands. This can be used to estimate path-average rainfall between telephone towers. It contains microwave frequency, end date & time of reading, minimum & maximum received power, path length, coordinates, link identifier, errored seconds, and severely errored seconds. The network consists of on average $\sim$3070 sub-links over $\sim$1818 unique link paths covering the Netherlands, having a temporal resolution of 15 min. The dataset spans the period 13 January 2011 up to and including 15 March 2015, although data gaps exist. The dataset contains part of the network from one of the three mobile network operators in the Netherlands during this 4-year period. The current network is expected to be very different and have a lower density, because in the Netherlands the CMLs have gradually been replaced by glass fiber-optic cables. The data have been provided by the mobile network operator (MNO) T-Mobile NL (since 5 September 2023 called Odido). Note that the transmitted signal levels were not available and are thus not provided. The MNO indicated that these were nearly constant (±0.2 dB) and no adaptive power control (ADPC) was used. The TSLs may therefore be considered constant for each individual link. For each CML one or two sub-links are available. Data from more than two sub-links were not provided by the mobile network operator, because these are only used for redundancy. Note that polarization was not provided, but that according to the MNO the majority of the links use vertically polarized signals.

The following datasets are provided:

- The **original raw data files** as obtained from mobile network operator T-Mobile NL. These are daily comma-separated values files that are gunzipped per day, each year having its own subfolder. Data were obtained from Nokia (1 dB power resolution) and NEC (0.1 dB power resolution; files after 12 August 2014 are empty) microwave links. Each vendor has its own daily file. All files together are provided in one zipped file **"RawCMLdata.zip"**. Description of variables of raw CML data:
  - LINK_ID: ID of the microwave link
  - SITE_ID: ID of location of equipment
  - SITE_LAT_SECS: latitude of the location in arc milliseconds (WGS84)
  - SITE_LON_SECS: longitude of the location in arc milliseconds (WGS84)
  - FAR_END_SITE_ID: ID of the other side of the microwave link
  - FAR_END_LAT_SECS: latitude of the other side of the microwave link in arc milliseconds (WGS84)
  - FAR_END_LON_SECS: longitude of the other side of the microwave link in arc milliseconds (WGS84)
  - FREQ: Transmit frequency in KHz
  - YYYYMMDDHHMMSS: textual representation of start time of measurement (local time, which is CET & CEST)
  - NEIDENTIFIER: Identifier of equipment in the management system
  - DURATION: Duration of the measurement period in seconds (for NEC) or minutes (for Nokia)
  - ES: Errored seconds according to ITU-T G.826
  - SES: Severely errored seconds according to IUT-T G.826
  - RXMIN_1 Lowest received power level of receiver 1 during the measurement period (dBm)
  - RXMAX_1 Highest received power level of receiver 1 during the measurement period (dBm)

- Using the R script “RawCML_to_RAINLINKFormat.R”, these original gunzipped data files are converted to one text file “IDRawCMLdata.dat”, that is provided as zip file “IDRawCMLdata.zip”. This file can be used with the open-source R package RAINLINK [RAINLINK, 2024, Overeem et al., 2024] to estimate path-averaged rainfall and to obtain interpolated rainfall maps, that can be visualized with Python script MapRAINLINK [Overeem, 2023a]. The data file “IDRawCMLdata.dat” could, likely after additional data formatting, also be employed by other rainfall retrieval packages. Each row provides the data from one individual microwave link and 15-min time interval.

  Variables in **“IDRawCMLdata.dat”** after processing raw data files by R script “RawCML_to_RAINLINKFormat.R” (**RAINLINK format**):

  - YStart: latitude of start point of link (degrees; WGS84)
  - XStart: longitude of start point of link (degrees; WGS84)
  - YEnd: latitude of end point of link (degrees; WGS84)
  - XEnd: longitude of end point of link (degrees; WGS84)
  - Frequency: microwave (carrier) frequency (GHz)
  - DateTime: date and end time (UTC; YYYYMMDDhhmm: year, month, day, hour, minute)
  - ES: errored seconds according to ITU-T G.826 (seconds; [1])
  - SES: severely errored seconds according to ITU-T G.826 (seconds; [2])
  - Pmin: minimum received signal level (dBm) over 15-min interval
  - Pmax: maximum received signal level (dBm) over 15-min interval
  - PathLength: distance of CML path (km)
  - Vendor: name of microwave link vendor (NOKIA or NEC)
  - ID: unique identifier of the sub-link

  The following processing steps are applied in “RawCML_to_RAINLINKFormat.R”:

  1. Rename variable names (header).
  2. Convert geographical coordinates (WGS84) from arc milliseconds to degrees.
  3. Calculate distance of each link (km). Afterwards, coordinates are rounded.
  4. Convert TX frequency from KHz to GHz.
  5. Only select CML data for which period duration is 15 minutes (NOKIA) or 900 seconds (NEC) and for which seconds are “00” to ensure 15-min sampling & add name of vendor of CML.
  6. Remove seconds from the start time.
  7. Convert local time (The Netherlands, CET) to UTC taking into account summer time (daylight saving time) and convert start time to end time of observation by adding 15 minutes & add name of vendor of CML
  8. Write relevant variables to output file which is readable by RAINLINK. Also errored and severely errored seconds are written as variable, although these are currently not used by RAINLINK.
  9. Assign unique ID by using the coordinates of start and end point of sub-links. The use of the IDs in the metadata is not advised due to possible errors, so the IDs are newly constructed based on the unique coordinates.
  10. Order data chronologically and write to output file.

Figures 1–12 provide an overview of the characteristics of the CML data regarding network topology, data availability, and coverage[3]. First, the following graphs are provided over the entire period: histograms (Figure 1) and time series with availability (Figure 2), histograms of path lengths, microwave frequency and link direction (Figure 3), scatter density plots of microwave frequency versus path length (Figure 4), and maps

[1] https://www.itu.int/rec/T-REC-G.826; probably https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-G.826-200212-I!!PDF-E&type=items

[2] https://www.itu.int/rec/T-REC-G.826; probably https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-G.826-200212-I!!PDF-E&type=items

[3] Figures were made with RAINLINK Version 1.31 functions ‘DataAvailability’ and ‘Topology’ and (the sometimes slightly modified) Python script MapRAINLINK [Overeem, 2023a].

Table 1: List of publications using subsets of the CML dataset. These use the same raw files, but the processing of raw files was slightly different compared to the processing in "RawCML_to_RAINLINKFormat.R".

Publication	Used data	Main objective
Overeem et al. [2013]	24 days from Nokia links	Evaluation of CML rainfall mapping
Rios Gaona et al. [2015]	12 days from Nokia links	CML measurement & interpolation uncertainties
Overeem et al. [2016a]	12 days from Nokia links	RAINLINK algorithm & sensitivity analyses
Overeem et al. [2016b]	2.5 years from Nokia + NEC links	Evaluation of CML rainfall mapping
Brauer et al. [2016]	$\sim$1 year from Nokia + NEC links	Hydrological simulation with CMLs
Van het Schip et al. [2017]	24 days from Nokia links	Wet-dry classification for CMLs with satellites
Rios Gaona et al. [2017]	$\sim$7 months of Nokia + $\sim$5 months of NEC links	CML dataset outperforms satellite products
Imhoff et al. [2020]	12 days from Nokia links	Nowcasting with CML rainfall maps
Wolff et al. [2022]	$\sim$3.5 months from Nokia links	Stochastic calibration of CML retrieval parameters
Overeem et al. [2023]	3 months of Nokia + NEC links	Evaluation of CML rainfall retrieval & radar-CML merging

with locations of link paths (Figure 5). Note that this holds for all figures displaying data availability: "data availability is computed over the number of time intervals in the dataset as obtained from DateTime. In case a time interval is missing, i.e. not present in DateTime, it is not included in the computations. Hence, the computed availability does only refer to the time intervals for which a DateTime entry is present in the dataset." [RAINLINK, 2024].

For the remainder of the figures results are presented per year. The spatial distribution and network density decline during the course of the period (Figure 6). The availability gradually declines in time (Figure 1), being related to network renewal and replacement by fiber-optic cables. Missing data were caused by data storage problems at the MNO and is not related to malfunctioning of CMLs. The average availability per 15 min time interval shows a stable delivery for Nokia sub-links (Figure 7), but a low(er) data availability for NEC sub-links from $\sim$15:00–$\sim$01:00 UTC caused by data storage problems at the MNO (Figure 8). Scatter density plots of microwave frequency versus path length are quite similar from year to year (Figure 9). The distribution of path length (Figure 10), microwave frequency (Figure 11) and direction (Figure 12) is also quite stable from year to year.

Note that for plotting all these figures 0.08% of the data has been removed, only keeping sub-links with a path length longer than 0 km and microwave frequency smaller than 60 GHz. This is because some sub-links had the same coordinates for the start point and end point resulting in a computed path length of 0 km, which is not physically possible. In addition, this removes the sub-links with a reported frequency of 254.1350 GHz, which is considered unrealistic and not useful for rainfall estimation. Hence, these data are removed to prevent error messages and also to make graphs more readable. Also note that the figures present the results for the raw CML data, so no processing with RAINLINK has been performed yet.

The mean country-wide link path density for the total dataset is $\sim$0.2 km km^2. This is computed by multiplying the mean number of unique link paths ($\sim$1818) by the mean path length ($\sim$3.95 km) and dividing by the area of the land surface of the Netherlands ($\sim 3.5 \times 10^4$ km^2).

Note that a somewhat differently processed dataset, but one that uses the same raw files, has been published before [Overeem, 2023b]. That dataset solely contains data from Nokia links. The raw data used to obtain that $\sim$4-month dataset are a subset of the current dataset. Moreover, for the current 4-year dataset the original, raw data files and the processing script are provided, making research more reproducible.

Acknowledgements

We gratefully acknowledge Ronald Kloeg and Ralph Koppelaar from T-Mobile NL (since 5 September 2023 called Odido) for the long-term pleasant cooperation since 2009, for providing the cellular telecommunication link data and for providing permission to make the CML data open.

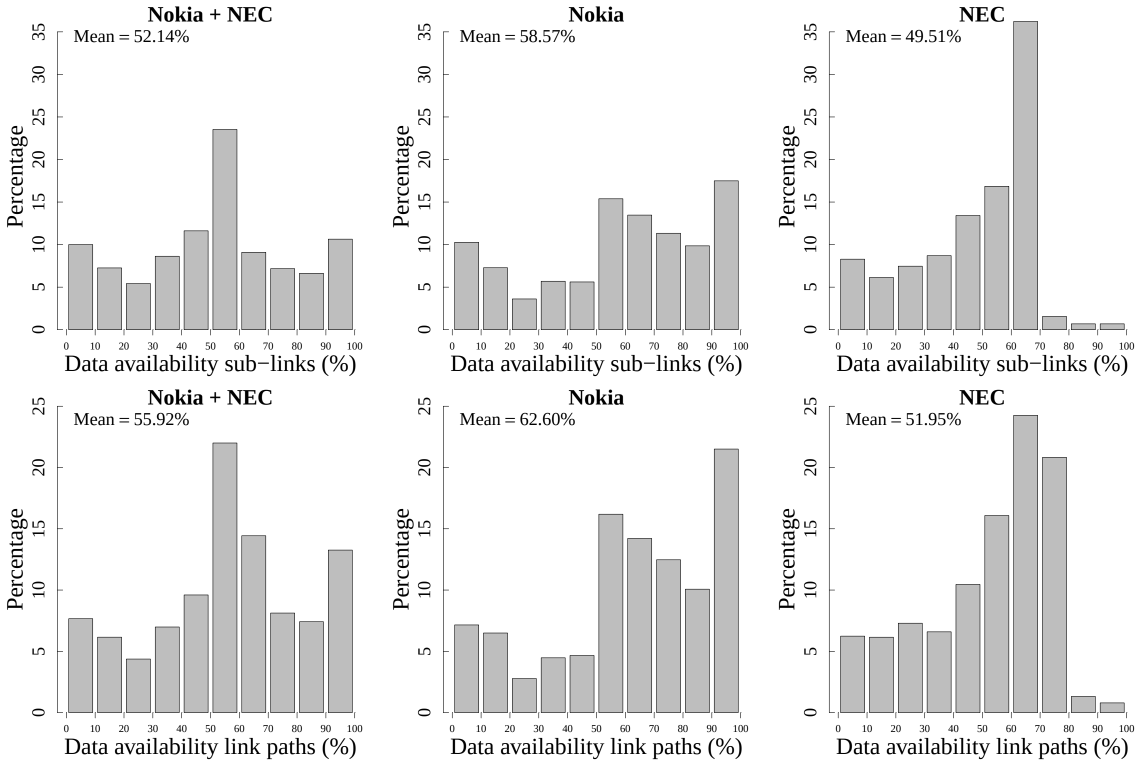

Figure 1: Histograms with availability for sub-links (first row) and for link paths (second row) for Nokia + NEC (left column), Nokia (middle column), and NEC (right column) links for the 2011–2015 period.

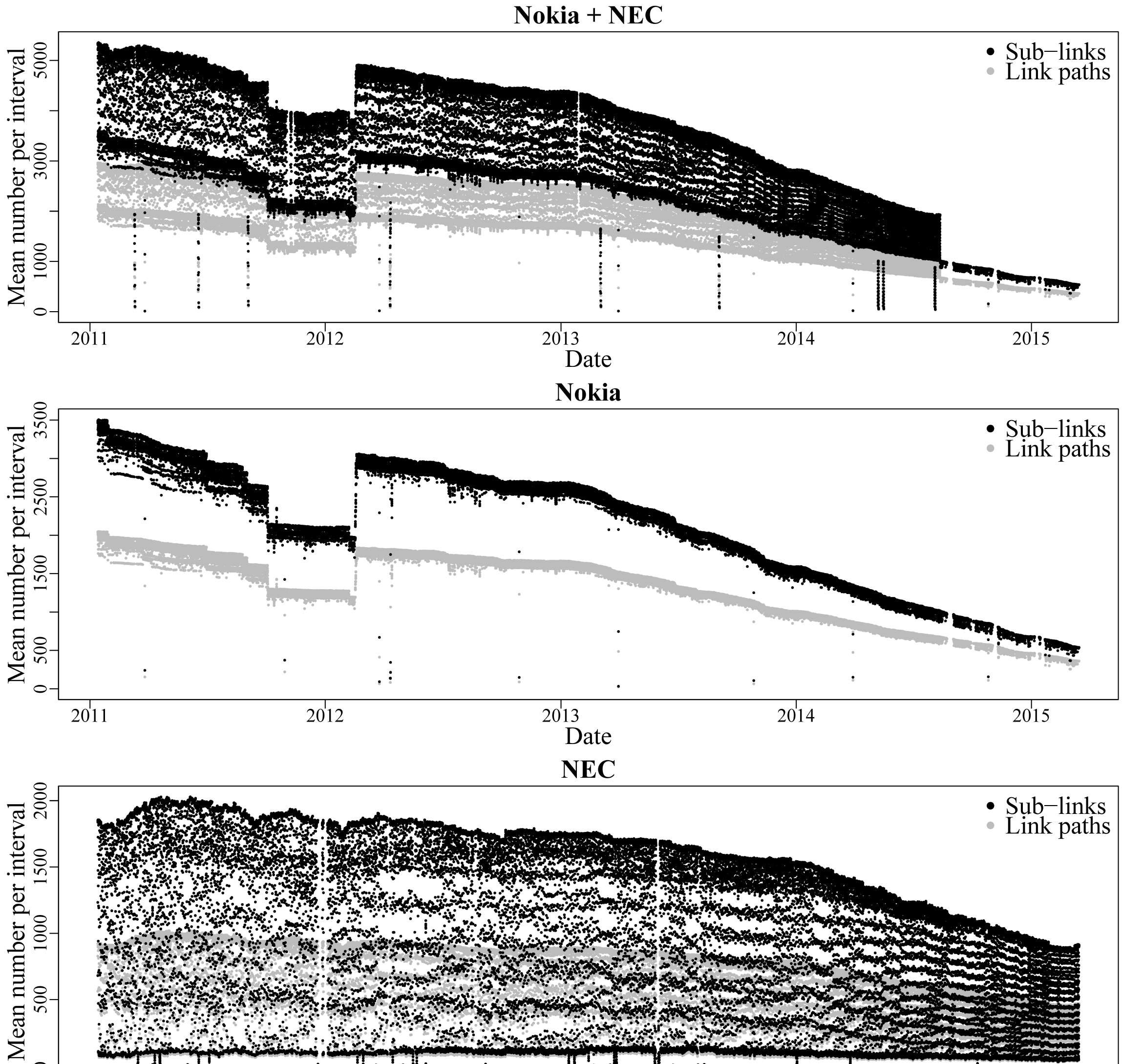

Figure 2: Numbers of available sub-links and link paths per 15 min time interval over the entire period for Nokia + NEC (top), Nokia (middle), and NEC (bottom) links for the 2011–2015 period.

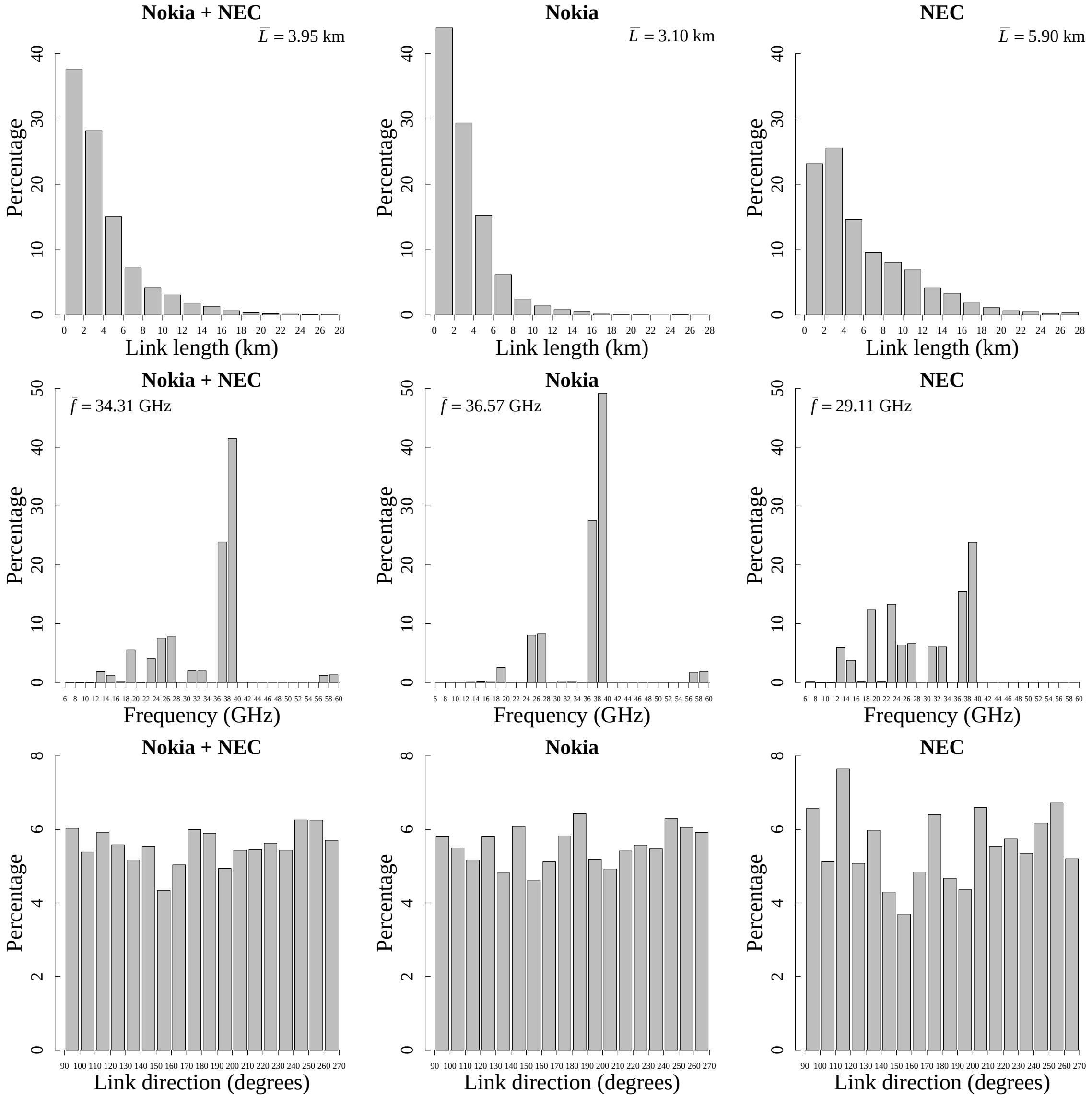

Figure 3: Percentages of links for a range of (first row) link path length L (km), (second row) microwave frequency f, and (third row) link direction classes for Nokia + NEC (left column), Nokia (middle column), and NEC (right column) links for the 2011–2015 period.

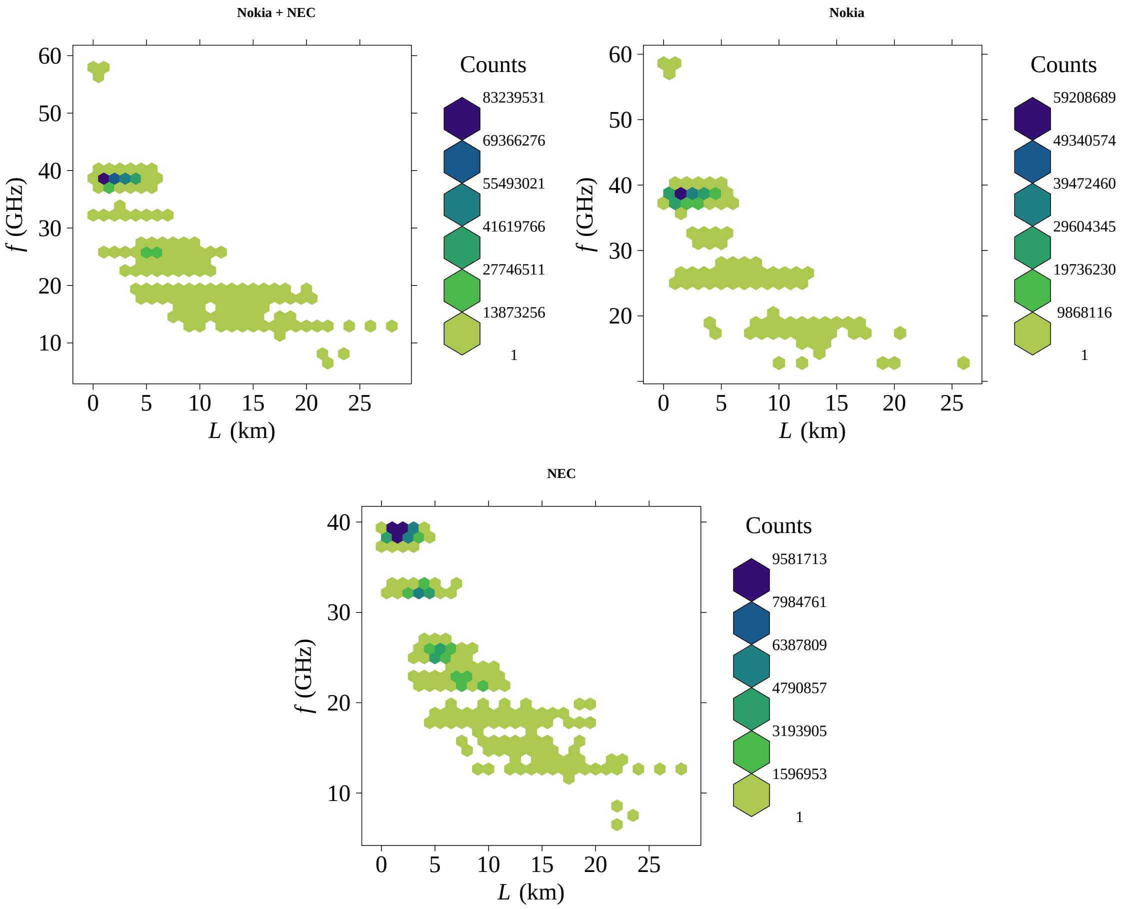

Figure 4: Scatter density plots of microwave frequency f (GHz) against path length L (km) for Nokia + NEC (top left), Nokia (top right), and NEC (bottom) links for the 2011–2015 period.

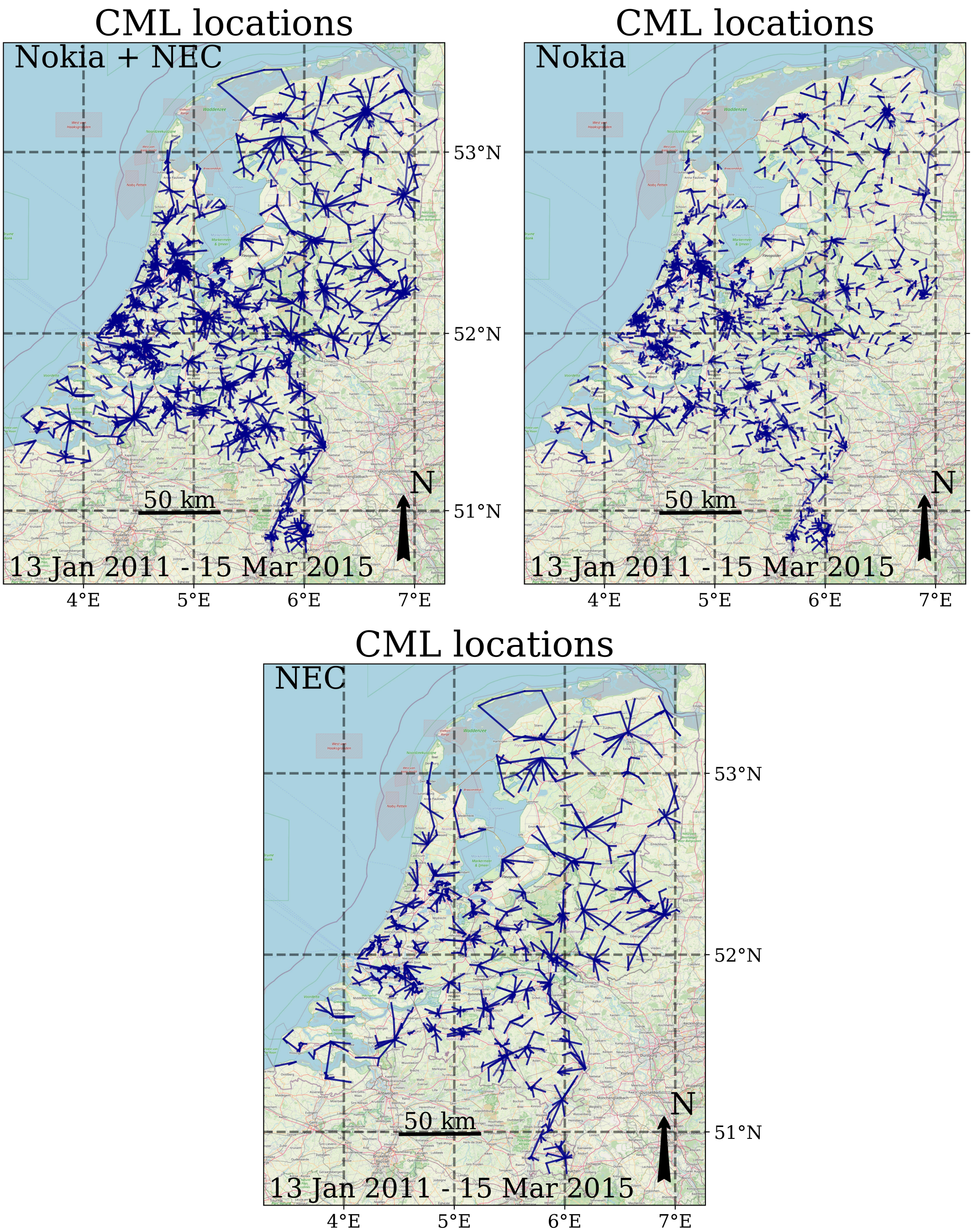

Figure 5: Maps of the Netherlands with locations of the used link paths from the cellular telecommunication network for Nokia + NEC (top left), Nokia (top right), and NEC (bottom) links for the 2011–2015 period.

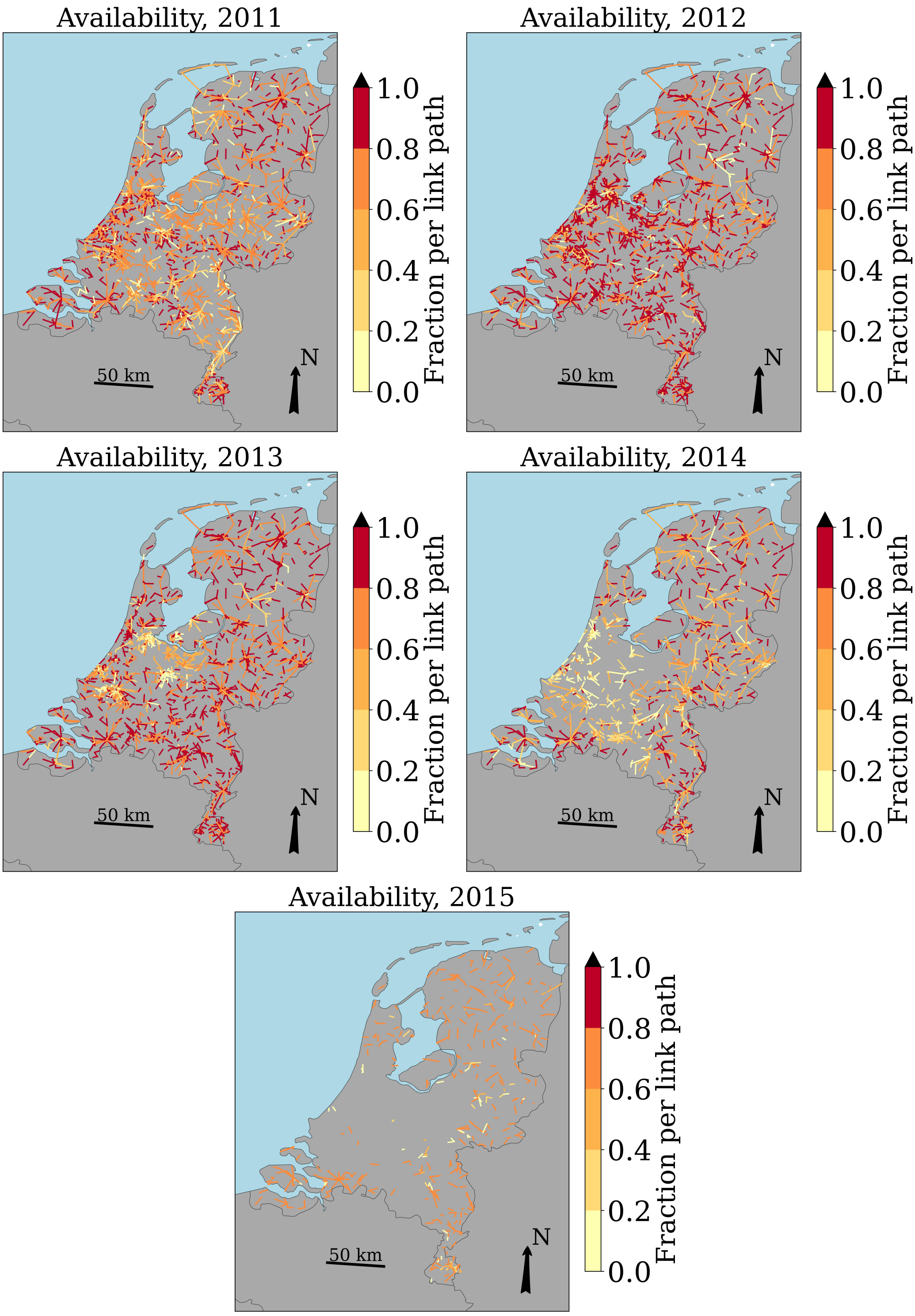

Figure 6: Maps of the Netherlands with locations of the used link paths from the cellular telecommunication network for Nokia + NEC for each year individually. The colors denote the availability of link data used to estimate rainfall. This shows that in the later years the spatial distribution and network density have declined.

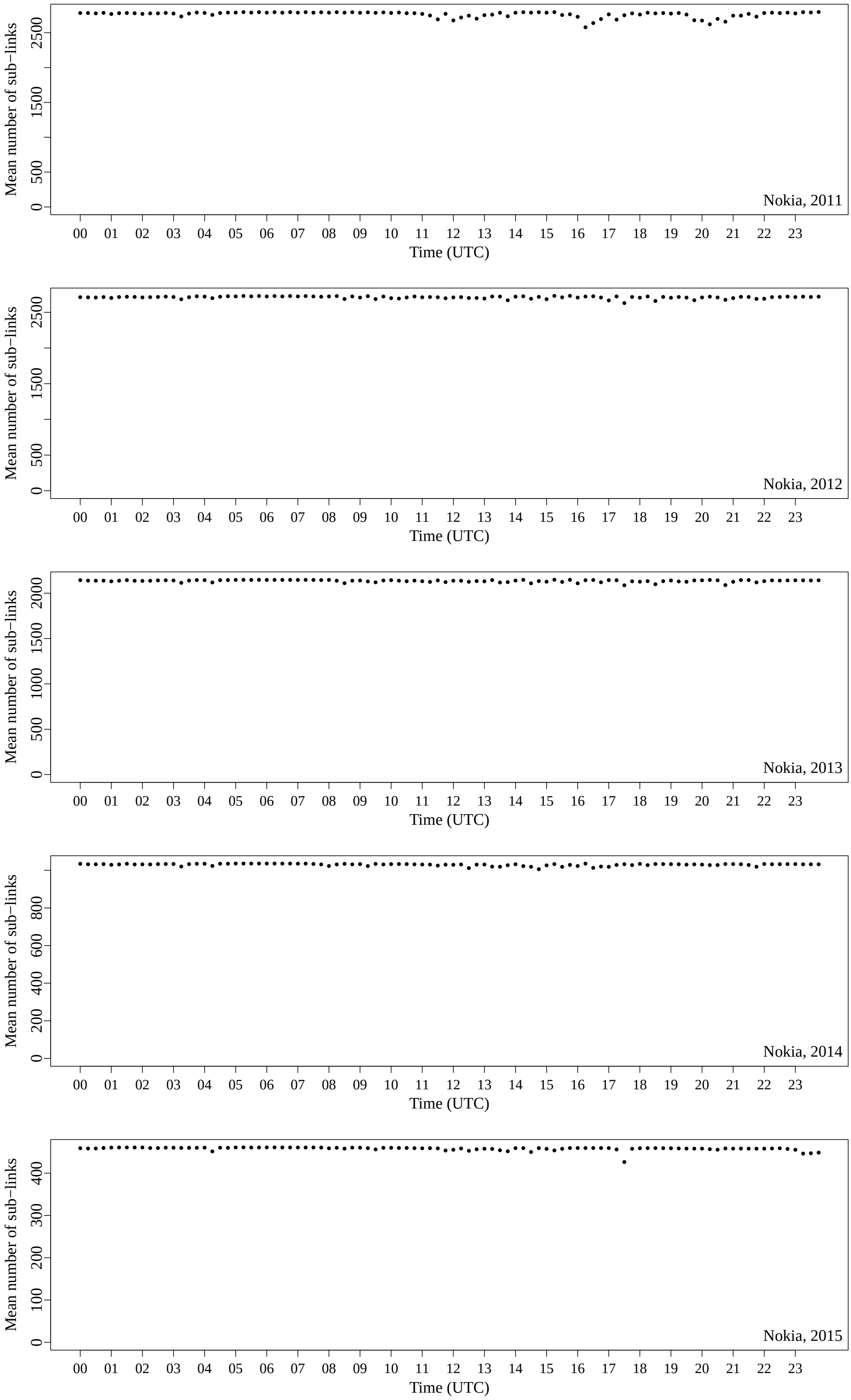

Figure 7: Average availability per 15 min time interval over an entire year for the Nokia sub-links for the years 2011–2015.

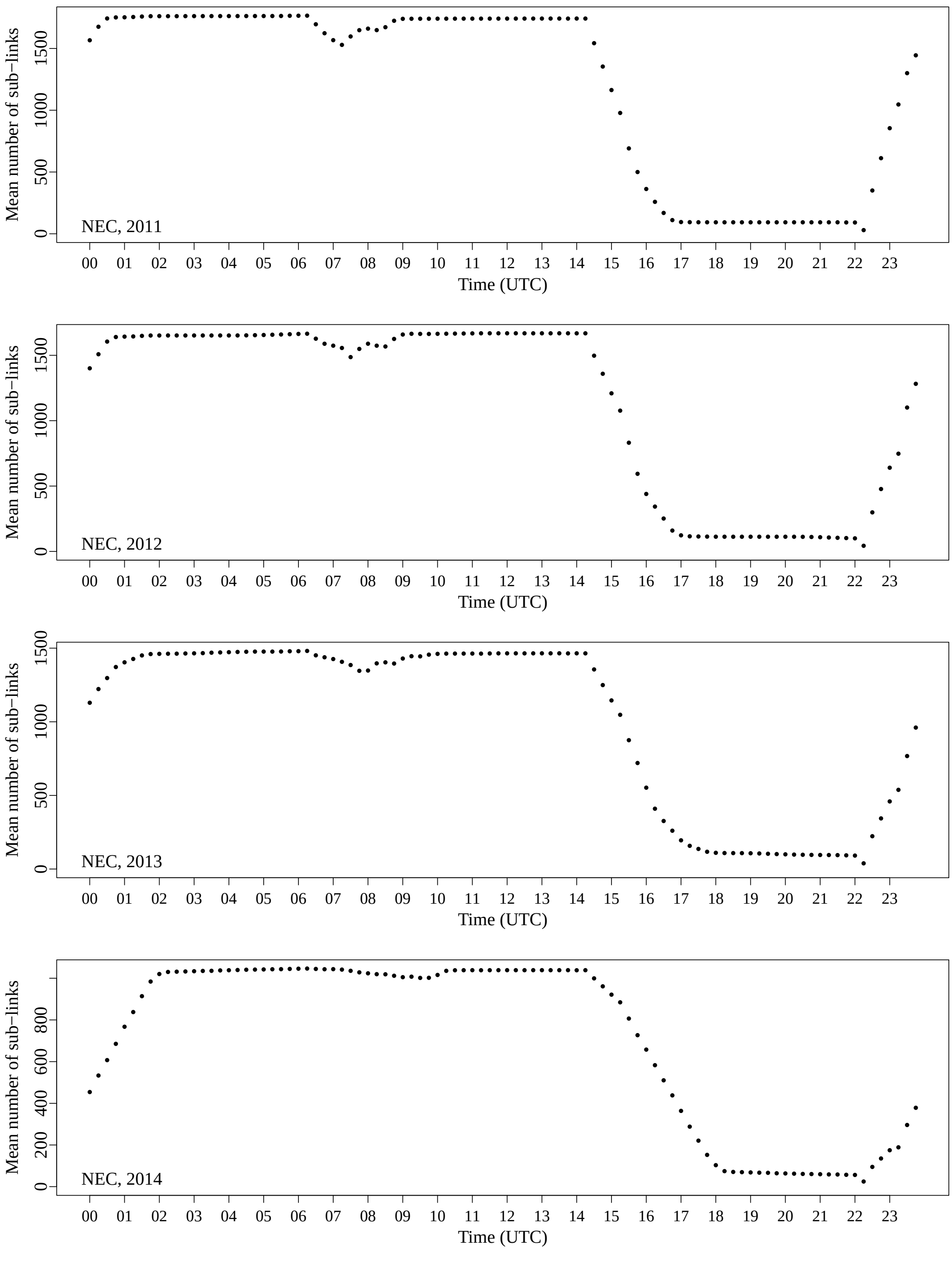

Figure 8: Average availability per 15 min time interval over an entire year for the NEC sub-links for the years 2011–2014.

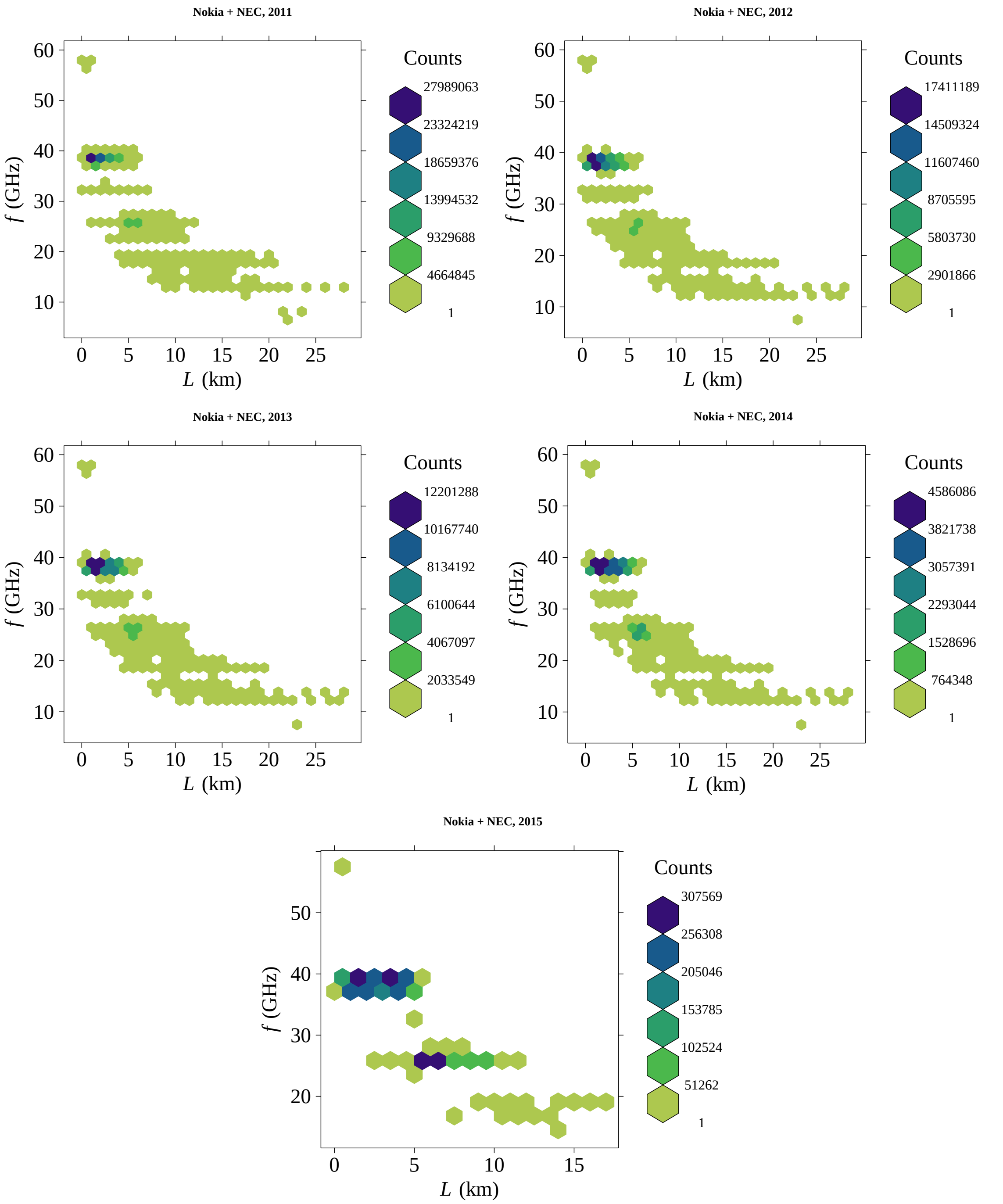

Figure 9: Scatter density plots of microwave frequency f (GHz) against path length L (km) for Nokia + NEC links for every year in the 2011–2015 period.

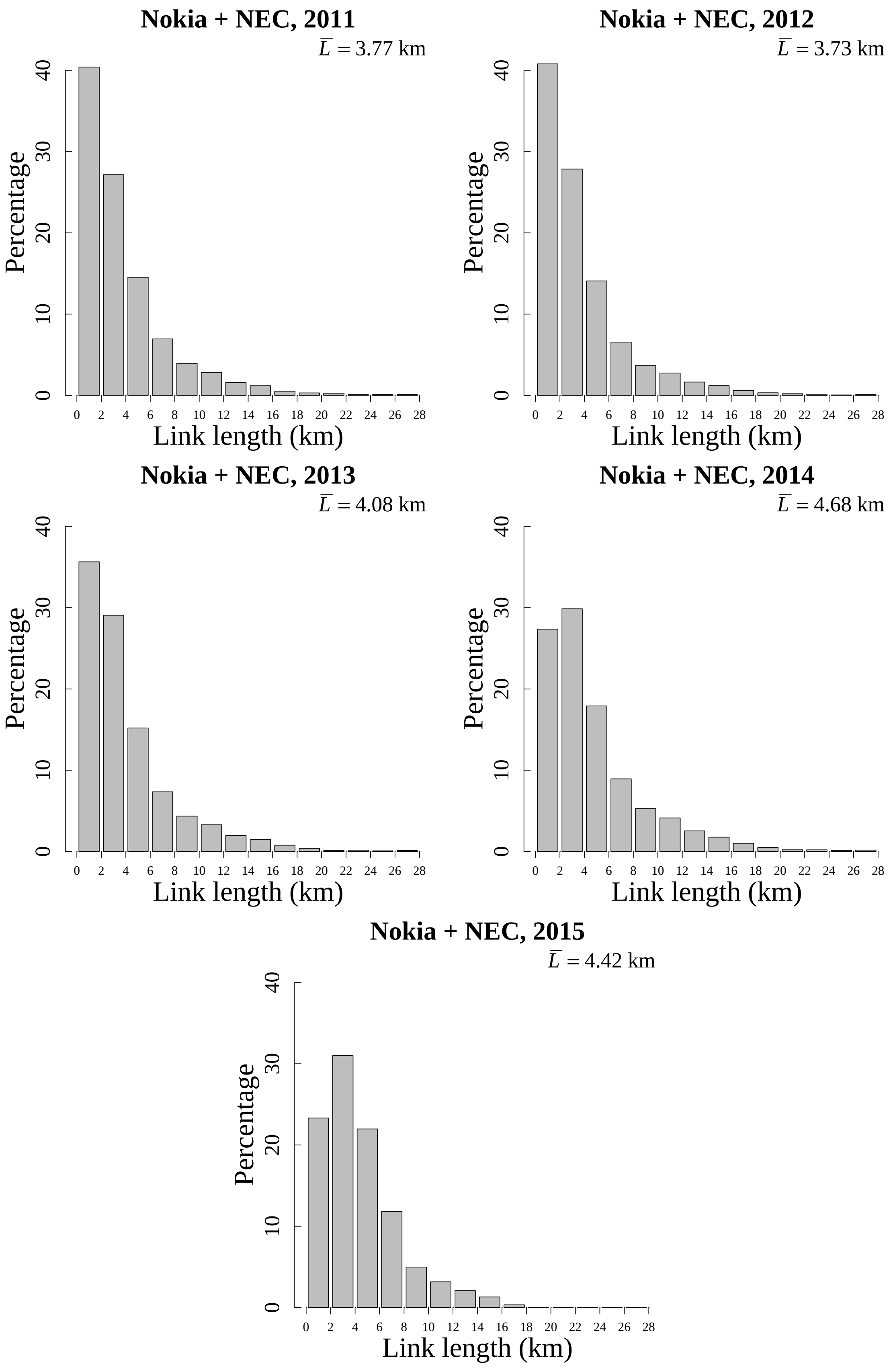

Figure 10: Percentages of links for a range of path length L for Nokia + NEC links for every year in the 2011–2015 period.

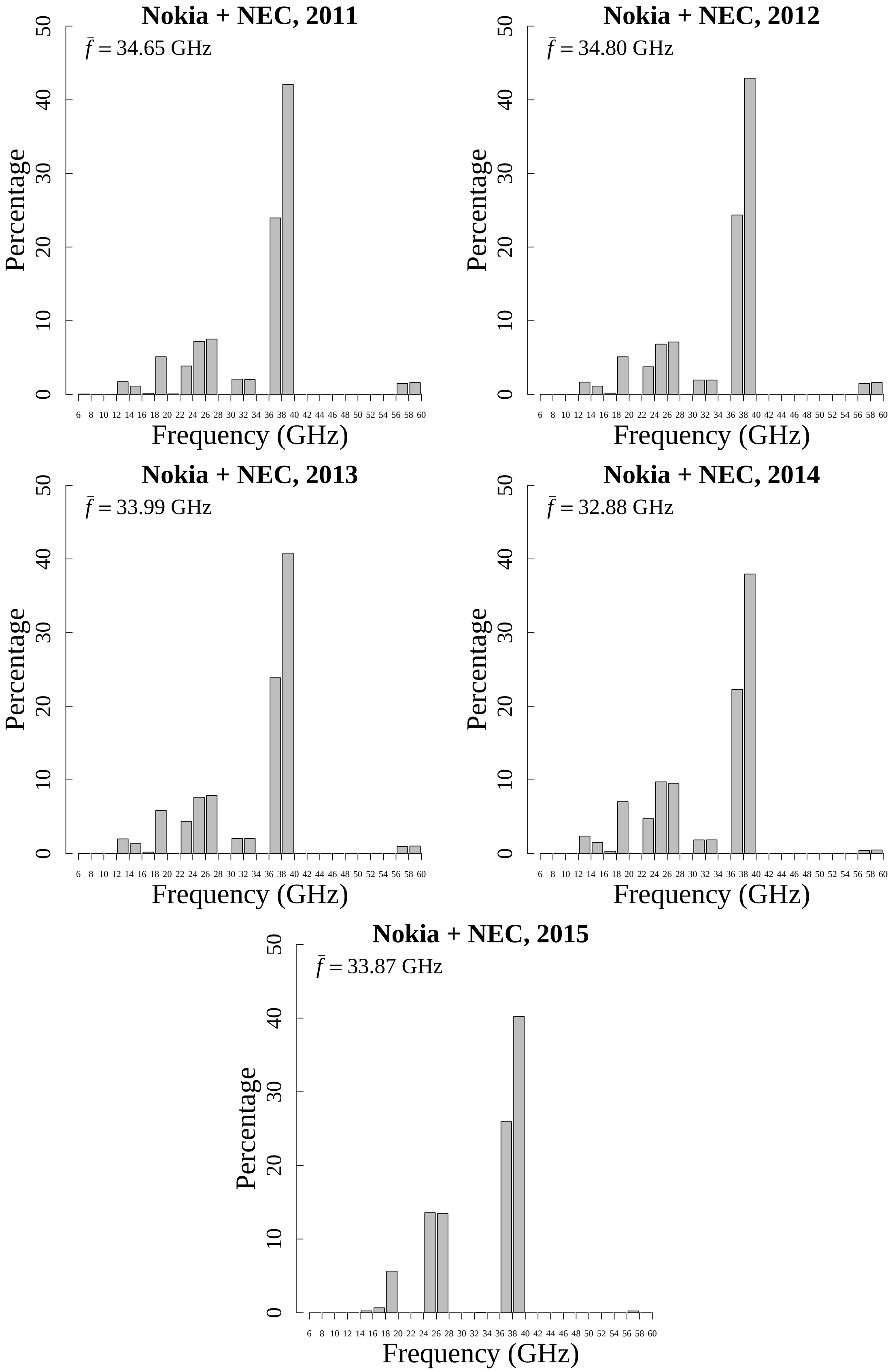

Figure 11: Percentages of links for a range of microwave frequency f for Nokia + NEC links for every year in the 2011–2015 period.

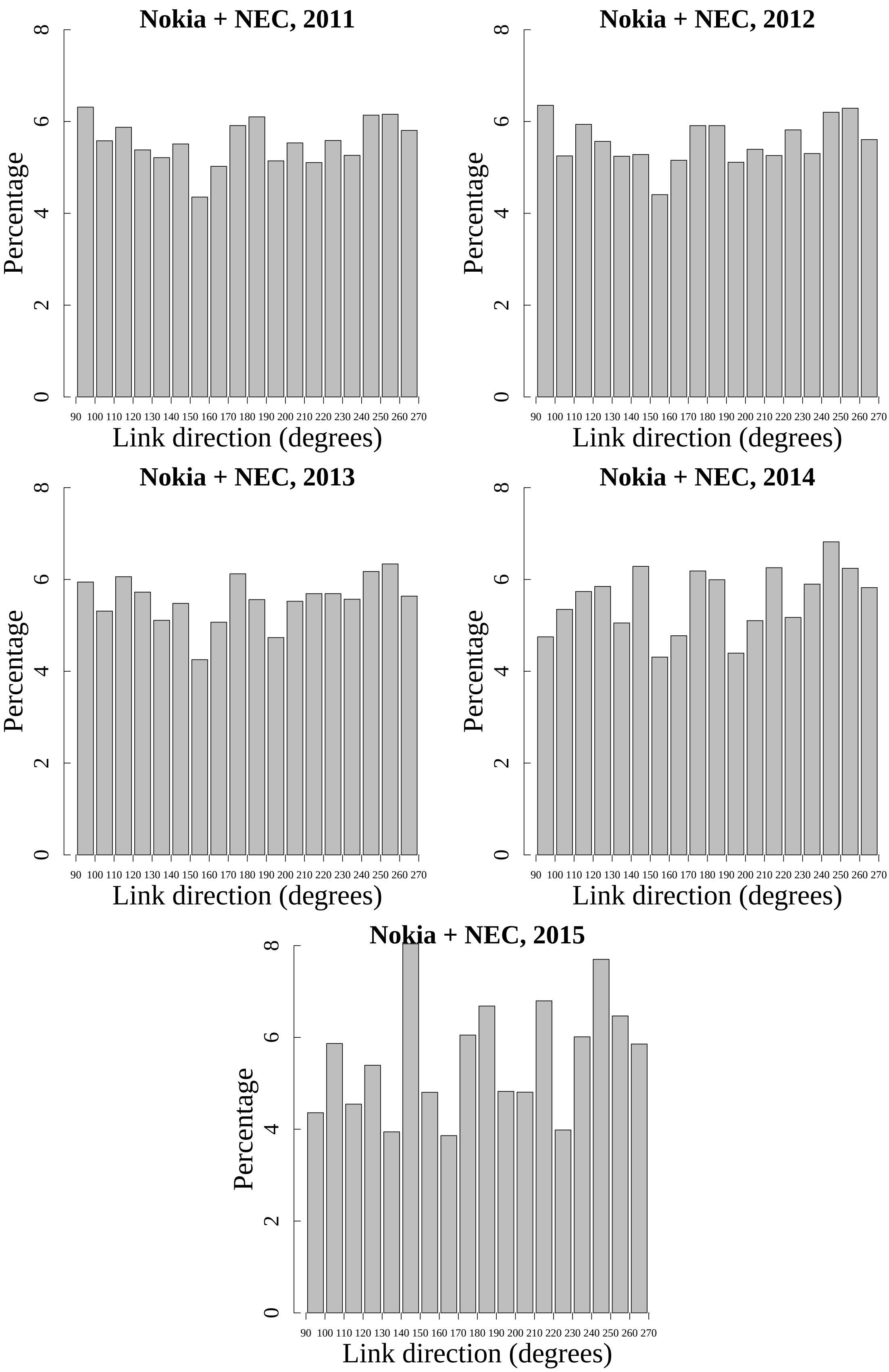

Figure 12: Percentages of links for a range of link direction classes for Nokia + NEC links for every year in the 2011–2015 period.

References

C. C. Brauer, A. Overeem, H. Leijnse, and R. Uijlenhoet. The effect of differences between rainfall measurement techniques on groundwater and discharge simulations in a lowland catchment. *Hydrol. Process.*, 30: 3885–3900, 2016. doi: 10.1002/hyp.10898.

R. O. Imhoff, A. Overeem, C. C. Brauer, H. Leijnse, A. H. Weerts, and R. Uijlenhoet. Rainfall nowcasting using commercial microwave links. *Geophys. Res. Lett.*, 47:n/a–n/a, 2020. doi: 10.1029/2020GL089365. e2020GL089365.

A. Overeem. MapRAINLINK (v1.02), 2023a. URL https://doi.org/10.5281/zenodo.7611398.

A. Overeem. Commercial microwave link data for rainfall monitoring. version 2., 2023b. URL https://doi.org/10.4121/7a692e36-c32f-4916-813b-c62d2566e8d8.v2. 4TU.ResearchData. Dataset.

A. Overeem, H. Leijnse, and R. Uijlenhoet. Country-wide rainfall maps from cellular communication networks. *P. Natl. Acad. Sci. U.S.A.*, 110:2741–2745, 2013. doi: 10.1073/pnas.1217961110.

A. Overeem, H. Leijnse, and R. Uijlenhoet. Retrieval algorithm for rainfall mapping from microwave links in a cellular communication network. *Atmos. Meas. Tech.*, 9:2425–2444, 2016a. doi: 10.5194/amt-9-2425-2016.

A. Overeem, H. Leijnse, and R. Uijlenhoet. Two and a half years of country‐wide rainfall maps using radio links from commercial cellular telecommunication networks. *Water Resour. Res.*, 52:8039–8065, 2016b. doi: 10.1002/2016WR019412.

A. Overeem, R. Uijlenhoet, and H. Leijnse. *Quantitative precipitation estimation from weather radars, personal weather stations and commercial microwave links*, volume 3, pages 27–68. Institution of Engineering and Technology, United Kingdom, 2023. ISBN 9781839536267. doi: 10.1049/SBRA557H_ch2. URL https://research.tudelft.nl/en/publications/quantitative-precipitation-estimation-from-weather-radars-persona. Eds.: Bringi, V.N. and Mishra, Kumar Vijay and Merhala Thurai.

A. Overeem, H. Leijnse, L.W. De Vos, and M. Silver. RAINLINK (v.1.31), 2024. URL https://doi.org/10.5281/zenodo.12211069.

RAINLINK. Github - overeem11/rainlink: Retrieval algorithm for rainfall mapping from microwave links in a cellular communication network, 2024. Available via https://github.com/overeem11/RAINLINK, accessed 21 June 2024.

M.F. Rios Gaona, A. Overeem, H. Leijnse, and R. Uijlenhoet. Measurement and interpolation uncertainties in rainfall maps from cellular communication networks. *Hydrol. Earth. Syst. Sc.*, 19:3571–3584, 2015. doi: 10.5194/hess-19-3571-2015.

M.F. Rios Gaona, A. Overeem, A. M. Brasjen, J.F. Meirink, H. Leijnse, and R. Uijlenhoet. Evaluation of rainfall products derived from satellites and microwave links for the Netherlands. *IEEE Transactions on Geoscience and Remote Sensing*, 55:6849–6859, 2017. doi: 10.1109/TGRS.2017.2735439.

T. I. Van het Schip, A. Overeem, H. Leijnse, R. Uijlenhoet, J. F. Meirink, and A. J. van Delden. Rainfall measurement using cell phone links: Classification of wet and dry periods using geostationary satellites. *Hydrolog. Sci. J.*, 62:1343–1353, 2017. doi: 10.1080/02626667.2017.1329588.

W. Wolff, A. Overeem, H. Leijnse, and R. Uijlenhoet. Rainfall retrieval algorithm for commercial microwave links: Stochastic calibration. *Atmos. Meas. Tech.*, 15:485–502, 2022. doi: 10.5194/amt-15-485-2022.