Alma Mater Studiorum · Università di Bologna

Scuola di Scienze
Corso di Laurea Magistrale in Fisica del Sistema Terra

Commercial Microwave Links as opportunistic sensors for precipitation in northern Italy: building and validating an operational monitoring network

Relatore:
Prof. Federico Porcù

Presentata da:
Giacomo Roversi

Correlatore:
Dott. Pier Paolo Alberoni

Sessione III
Anno Accademico 2020/2021

Per aspera ad astra.

Abstract

There is a growing interest in emerging opportunistic sensors for precipitation, motivated by the need to improve its quantitative estimates at the ground. The scopes of this work are to discuss the effort of building a CML-based opportunistic sensing framework from scratch and to present preliminary assessment of the accuracy of CMLs retrieved rainfall rates in Northern Italy. The CML product, obtained by the open-source RAINLINK software package, is evaluated on different scales (single link, $5\,\mathrm{km} \times 5\,\mathrm{km}$ grid, river basin) against the precipitation products operationally used at Arpae-SIMC, the Regional Weather Service of Emilia-Romagna, in northern Italy. The results of the $15\,\mathrm{min}$ single-link validation with close-by rain gauges show high variability, which can be caused by the complex area physiography and precipitation patterns. Known sources of errors (e.g. the attenuation caused by the wetting of the antennas or random fluctuations in the baseline) are particularly hard to mitigate in these conditions without a specific calibration, which has not been implemented. However, hourly cumulated spatially interpolated CML rainfall maps, validated with respect to the established regional gauge-based reference, show similar performance (R^2 of 0.46 and CV of 0.78) to adjusted radar-based precipitation gridded products and better than satellite- ones. Performance improves when basin-scale total precipitation amounts are considered (R^2 of 0.83 and CV of 0.48). Avoiding regional-specific calibration therefore does not preclude the algorithm from working but has some limitations in POD and accuracy. A widespread underestimation is evident at both grid box (Mean Error of -0.26) and basin-scale (Multiplicative Bias of 0.7), while the number of false alarms is generally low and gets even lower as link coverage increases. Taking into account also delays in the availability of the data (latency of 0.33 hours for CML against 1 hour for the adjusted radar and 24h for the quality-controlled rain gauges), CML appears as a valuable data source in particular from a local operational framework perspective. Finally, results show complementary strengths for CMLs and radars, encouraging joint exploitation.

Sommario

La domanda di dati quantitativi precisi e rapidi di monitoraggio ambientale, ed in particolare sulla precipitazione, è in continua crescita. Per rispondere a questa domanda esiste una tecnologia che sta mostrando risultati molto promettenti ormai da anni, ma che ancora difficilmente viene integrata in ambito operativo: i link a microonde delle reti mobili cellulari (ovvero i *commercial microwave links*, abbreviato CML). Si tratta di una tecnologia "di opportunità", che sfrutta cioè strumenti già presenti sul territorio e già gestiti operativamente, ma per altri scopi, diversi dalla ricerca o dal monitoraggio.

L'obiettivo di questo lavoro è discutere di come meglio implementare un framework regionale di sensori di opportunità basato sui CML, poi di presentare i risultati della prima validazione a larga scala e su network operativo reale di un prodotto di stime di precipitazione derivato da CMLs.

La campagna sperimentale principale ha coinvolto due mesi di dati di attenuazione sulla rete Vodafone delle province di Bologna e Parma. Le analisi hanno mostrato come i link, anche senza specifica calibrazione degli algoritmi di retrieval, siano in grado di fornire autonomamente stime quantitative affidabili degli accumuli di precipitazione, soprattutto se considerati su scale temporali non eccessivamente brevi ($> 1\,\mathsf{h}$) e su aree estese ($> 100\,\mathsf{km}^2$) e ben attraversate da links ($> 0.2\,\mathsf{km}$ di links per km^2), e sprattutto se i fenomeni precipitativi sono di medio-alta intensità ed estesi spazialmente e temporalmente. In particolare, il prodotto interpolato di precipitazione cumulata oraria valutato rispetto al riferimento regionale ERG5 ha mostrato prestazioni complementari coi prodotti radar, ottenendo ottimi valori di FAR e buone correlazioni quantitative (entrambi punti deboli dei radar operativi) a scapito di un basso POD soprattutto per fenomeni brevi e circoscritti, situazione in cui invece il radar eccelle.

Confontate coi pluviometri, le reti CML sono superiori in risoluzione spaziale e temporale e forniscono anche una più rapida disponibiltà del dato (latenza). Possono essere quindi un ottimo strumento integrativo al servizio del previsore per il nowcasting oppure come servizio di allerta indipendente per eventi intensi o estremi (come sviluppato per il progetto RainBo LIFE15 CCA/IT/000035).

Contents

Abstract 3

Sommario 4

Introduction 8

1 Physical background 11

1.1 Precipitation . . . 11

1.1.1 Precipitation regimes . . . 12

1.1.2 Spatial autocorrelation . . . 14

1.1.3 Drop size distribution . . . 15

1.2 Raindrops and electromagnetic waves . . . 17

1.2.1 Extinction scattering geometry: microwave links . . . 17

1.2.2 Backscattering geometry: weather radars . . . 18

1.3 k-R and Z-R relations . . . 19

2 Experimental area 22

2.1 Orography and hydrology . . . 22

2.2 Local climatology . . . 23

3 Instrumentation 26

3.1 Commercial microwave links . . . 26

3.1.1 Opportunistic sensors . . . 27

3.1.2 Links characteristics . . . 28

3.1.3 Attenuation data . . . 28

3.1.4 ATPC . . . 30

3.1.5 CML networks in Emilia-Romagna . . . 31

3.2 Rain gauges . . . 32

3.3 ERG5 rainfall analysis . . . 32

3.4 Weather radars . . . 33

4 Methods **36**
4.1 RAINLINK rain retrieval algorithm 36
4.2 Error metrics 38

5 Building the framework **40**
5.1 The RainBo Life EU Project 40
5.2 Working with RAINLINK 41
5.2.1 Getting started: data formatting 41
5.2.2 Test runs with synthetic data 42
5.2.3 From a research software to a near real-time online service 42
5.2.4 New functions for RAINLINK 43
5.3 RAINLINK setup for Emilia-Romagna 44
5.3.1 Considerations about network characteristics 45
5.3.2 Considerations about climatology 45
5.3.3 Considerations about orography 46
5.3.4 Why avoiding a regional calibration 46
5.4 Early challenges 47
5.4.1 Dataset issues 47
5.4.2 Anomalous signals 49
5.5 First results from the complete Lepida network 53

6 Validating the products **56**
6.1 Data characteristics 56
6.1.1 CMLs 56
6.1.2 Reference fields 58
6.2 Coverage 58
6.3 Data quality, filters and outliers 59
6.4 Results 61
6.4.1 Single link verification 61
6.4.2 Gridded product verification 65
6.5 Discussion 75
6.6 Further insights 77
6.6.1 Boundary effects on total accumulations 77
6.6.2 Calibration attempts of *Aa* and *alpha* 77

Conclusions **81**

Introduction

Precipitation is, among geophysical observables, one of the most difficult to measure and monitor. This is because precipitation fields have a great deal of variability both in space and time. Accurate precipitation measurements would benefit a wide range of applications. Meteorology, hydrology, climatology, agriculture: precipitation plays a key role in all of them, and the list could be much longer. Recently, public interest in the topics of climate change and hydrogeological risk mitigation is finally growing, which gives to rainfall monitoring a particular social relevance too.

Precipitation rate can be measured (or estimated) either directly at the ground or using remote sensing approaches. Satellite estimates received a renewed boost in the last decade from the full exploitation of the Global Precipitation Measurement mission (GPM, Skofronick-Jackson et al. (2017)) that operationally releases a new suite of precipitation products with a high temporal and spatial resolution (Mugnai et al., 2013; Grecu et al., 2016). Despite the undoubted potentials of satellite products to provide estimates over open oceans and regions not equipped with ground instruments, their accuracy is difficult to assess at high spatial and temporal scales (Tang et al., 2020), and their latency hinders the use in real-time monitoring of rain patterns.

A relatively new and independent approach to the estimate of precipitation at the ground became available in the last decades thanks to the growing number of microwave links (or Commercial Microwave Links, CMLs) employed for the backhauling of the cellular communication networks. Growth which only recently and only in some densely populated areas seems to have come to a halt. Integrated precipitation content along a straight path between two antennas can be estimated by measuring the attenuation of the microwave signal travelling down the same path (Turner and Turner, 1970; Harden et al., 1978). Accurate experiments with dedicated hardware and numerical simulation were used to assess the capability of microwave links to measure average rainfall rates (Rahimi et al., 2003), drop size distribution (Rincon and Lang, 2002; van Leth et al.,

2019) and water content (Jameson, 1993). The possibility to have a spatially continuous rainfall field depends on the density and distribution of the links, making the CML approach of particular interest for urban areas (Upton et al., 2005; Overeem et al., 2011; Fenicia et al., 2012; Fencl et al., 2013; Rios Gaona et al., 2017; de Vos et al., 2018) with also direct hydrological use in combination with conventional instruments (Grum et al., 2005; Fencl et al., 2013). A further application of the CML approach could be in regions where other instruments are lacking or absent at all (Mulangu and Afullo, 2009; Abdulrahman et al., 2011; Doumounia et al., 2014). However, as it happens for conventional precipitation instruments, the quality of the retrieval is sensitive to several factors, which are often difficult to control (Leijnse et al., 2008), and to the precipitation microphysical structure (Berne and Uijlenhoet, 2007; Leijnse et al., 2010). Given these limitations intrinsic to the measurement geometry and to the nature of precipitation, possible synergistic approaches are considered, to minimize the uncertainties of the different instruments, suggesting the blending of CML measurements with conventional precipitation estimates, such as rain gauges (Fencl et al., 2017; Haese et al., 2017), radar (Cummings et al., 2009; de Vos et al., 2019), or both (Grum et al., 2005; Bianchi et al., 2013).

Even if the general relationship between signal attenuation and rain rate is already well established, the successful use of CML data to quantitatively monitor precipitation still depends on the quality and technical characteristics of the transmitted power data and on the fine-tuning of the algorithms. The somewhat standardised policies of acquisition and storage of the different companies in different countries make the use of CMLs feasible all around the world, but there is no standard way yet to access them as scientific data. As they consist mostly of confidential maintenance data, major obstacles to face are the widespread unwillingness of releasing them cost-free and the inadequate data-quality standards (Chwala and Kunstmann, 2019). Coordinated initiatives by the research institutions are just now slowly starting to create a common ground on which to build further.

In a world that lives and thrives on bigger and bigger quantity of data every day, opportunistic sensing is a key to the ground environmental and meteorological monitoring of the future, which has to be fast, data-rich and sustainable. The objective of this thesis work is to contribute toward an increased diffusion of opportunistic, open-source methods to help low-cost precipitation monitoring around the world. It tries to

pursue that objective by describing the challenges we have encountered to start up a CML opportunistic sensing framework in a new area, discussing issues and choices and presenting the validation results achieved by the retrieval algorithm.

Specifically, Chapter 1 gives the physical fundaments to the rain retrieval problem, Chapter 2 introduces the main characteristics of the experimental area and Chapter 3 does the same for the instrumentation. CMLs are thoroughly examined as opportunistic sensors, other conventional instruments are also briefly introduced. Chapter 4 contains a detailed description of the rain retrieval algorithm and a formal definition of the scores utilised in the evaluation metric. Chapter5 tackles all the preliminary considerations and analyses which led to build a consistent framework for a rigorous validation of the CML product, procedure which is presented in the following Chapter 6. Results of the validation are presented and discussed there.

1 Physical background

1.1 Precipitation

According to the AMS Glossary[1], precipitation consists of all hydrometeors formed in the atmosphere, including liquid, solid, or a combination of the two, that are large enough to fall as a result of gravity. Quantitatively, precipitation is the amount of the water substance that has fallen at a given point over a specified period of time, usually expressed in millimetres of liquid water depth (rain depth). The instantaneous rate at which rain is falling is instead the rain intensity, expressed in millimetres per hour (rain rate). If the accumulated precipitation is registered every hour, then the rain depth coincides with the average of the rain rate over that specific hour.

The rainfall intensity ($\mathrm{mm\,h^{-1}}$), assuming the density of water to be constant, can be calculated as follows:

$$R(t) = C_R \int_0^\infty V(D)\, w_T(D)\, N(D,t)\, dD \tag{1.1}$$

where D is the raindrop diameter (mm), $V(D)$ is the raindrop volume ($\mathrm{mm^3}$), $w_T(D)$ is its terminal falling velocity ($\mathrm{m\,s^{-1}}$), $N(D,t)$ is the numberof particles per cubic meter with diameter between D and $D+dD$ as a function of time t ($\mathrm{mm^{-1}\,m^{-3}}$) and $C_R = 3.6 \times 10^{-3}$ is a dimensionless unit conversion factor. $N(D,t)$ is called *drop size distribution* (DSD) and describes the precipitation at the microphysical level.

According to Atlas and Ulbrich (1977), the vertical terminal velocity of a raindrop depends on its diameter following the experimental relationship:

$$w_T(D) \simeq 3.87\, D^{0.67} \tag{1.2}$$

For spherical raindrops the volume is easily expressed as function of diameter. Defining

[1] `https://glossary.ametsoc.org/wiki/Precipitation`

a new coefficient $C_R' = C_R + 3.87 = 7.47 \times 10^{-3}$, the equation (1.1) becomes:

$$R(t) = C_R' \frac{\pi}{6} \int_0^\infty D^{3.67} N(D, t)\, dD \tag{1.3}$$

which is the 3.67th statistical moment of the drop size distribution $N(D, t)$.

1.1.1 Precipitation regimes

Precipitation at mid-latitudes may occur as a result of various mechanisms, prompted by displacements of air and water vapour masses through the atmosphere at both synoptic and regional scales. The characteristics of these motions vary quite regularly with the seasonal cycle (climate change aside).

Advection of warm moist air in warm frontal structures generates slow vertical ascending flows which produce widespread precipitation of weak intensity distributed in time over many hours. At our latitudes (northern Italy, 45°N) this happens particularly in late autumn and winter. Precipitation coming from this type of atmospheric structure is named *stratiform*. When analysing rainfall data, signals from stratiform precipitation are associated with low spatial variance and a time persistence of multiple hours over the same instrument or area.

Slow convection could also happen when an advective air mass is lifted up by some orographic forcing. In this case the precipitation field tends to be stationary with the terrain features, regardless the atmospheric motion. If the terrain features are complex, precipitation could gain spatial variance also in stratiform conditions.

Advecting cold fronts generate instead faster and clustered vertical motions which produce more intense and localised precipitating phenomena. The convection takes place at the edge and immediately behind the fast moving front. The cold front itself has a coherent lifetime of several hours or days, but the precipitation field extends only for a small distance along the direction of motion (perpendicular to the front line) forming typical rainfall bands. The signal over a specific area is therefore much more short-lived. Measurements at fixed locations show shorter persistence and spatial variance tends to be markedly anisotropic. Intense cold fronts are more frequent in spring, when intrusion of cold air from northern regions reach the warm air already present in the area. During warmest months, strong convection arises mainly from surface radiative heating. If particularly unstable conditions are present in the atmosphere column, a complete thunderstorm cell could develop. Precipitation will be most intense, very localised and swift. These phenomena are categorised as small-scale convection. Both cold-fronts

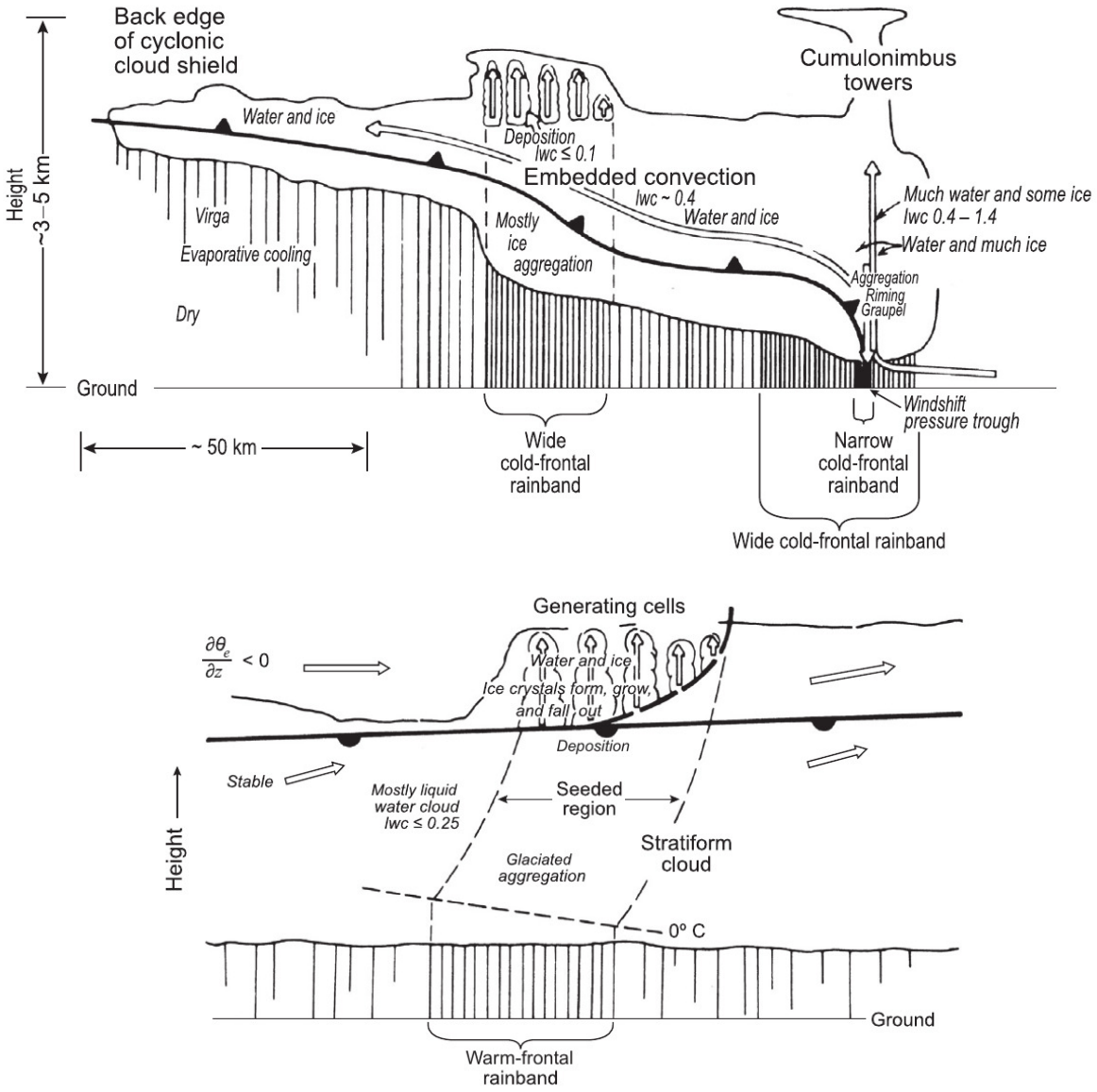

Figure 1.1: Formation of convective (up) and stratiform (down) precipitation: transverse sections of a cold and warm frontal systems respectively. From Matejka et al. (1980).

and small-scale convection are responsible of what is called *convective* precipitation. In Figure 1.1 conditions leading to convective and stratiform rain are represented schematically (Matejka et al., 1980). The tools to quantitatively discriminate between the two regimes are introduced in the next two sections.

1.1.2 Spatial autocorrelation

Measurements close to each other produce typically similar outcomes compared to measurements for points separated by larger distances. When analysing quantitative georeferenced rainfall data, spatial variability properties can be described by means of a semi-variogram.

The semi-variogram measures the degree of dissimilarity $\gamma(h)$ between data separated by a distance h. Calling $N(h)$ the number of data pairs $z(x_i)$ and $z(x_i + h)$ lying within a given distance h, the experimental semi-variogram is defined as the average squared difference between the components of the data pairs (Goovaerts, 1999) as in the following equation:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \Big[z(x_i) - z(x_i + h) \Big]^2 \tag{1.4}$$

The variance of the entire dataset is referred to as the *sill*. The distance at which the model semi-variogram equals the dataset variance is defined as the *range*. The range of a spherical variogram is also known as decorrelation distance. The variance of the sample at zero separation distance is called *nugget* and is commonly non zero for rainfall measurements.

Given a random spatial distribution of measures, semi-variograms can be used as a basis to obtain the spatial interpolation of the measured values. A variogram model is a mathematical function that is permitted for fitting the points on an experimental semi-variogram (positive definiteness is needed). The degree of spatial autocorrelation defined by a variogram model is used to assign weights in the Kriging interpolation. The spherical variogram model (Kigobe and Kizza, 2006) is defined as:

$$\gamma(h) = \begin{cases} n + c \left[1.5 \left(\frac{h}{a} \right) - 0.5 \left(\frac{h}{a} \right)^3 \right] & \text{if} \quad h \leq a \\ n + c & \text{if} \quad h > a \end{cases} \tag{1.5}$$

where a is the range, n is the nugget and $n + c$ is the sill.

Variograms for different periods of the year may differ quite a lot from each other. In Figure 1.2, sill, range and RMSE parameters are shown for hourly to daily accumulated precipitations over The Netherlands after a climatological analysis of 30 years of data (van de Beek et al., 2012). While the longer accumulations dampen the signal in the

variances (sill and RMSE, here presented instead of the nugget), the seasonal signal remains unaltered at every scale in the values of the range. Differences greater than 100% between winter and summer are evident.

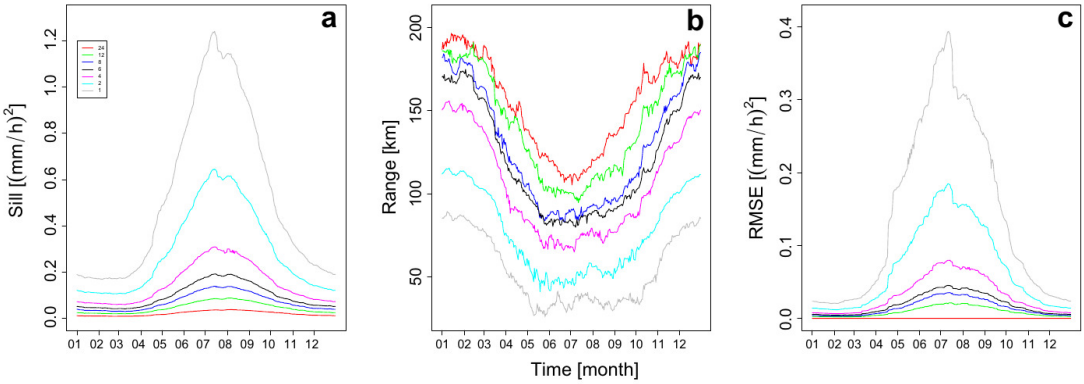

Figure 1.2: Variogram parameters resulting from a climatological analysis of 30 years of precipitation data over The Netherlands. (a) The seasonal fluctuation of the 90-day averaged sill between January 1st, 1979 and February 15th, 2009, averaged for each day of the year. The accumulation intervals indicated in the legend (colours) are given in hours. (b) The same for the range. (c) The same for the RMSE. From van de Beek et al. (2012).

These differences are caused by the different precipitation regimes examinated qualitatively in the previous section: during winter months, stratiform precipitation produces rainfall fields which are still autocorrelated even at distances of almost 100 km during a single hour, while summer convective events do not exceed ranges of 25 km and reach 100 km only at the daily scale (24 h).

1.1.3 Drop size distribution

The two precipitation regimes can be microphysically differentiated by looking at the size distribution of the hydrometeors (DSD) defined earlier. Some examples reported in Figure 1.3 from Caracciolo et al. (2006a) show variability in the population of the size classes ranging many orders of magnitude and a typical common shape: DSDs are skewed towards the bigger diameters and have a mode close to the small ones, while microphysical processes (respectively coalescence and break-up) avoid the presence of too many small and big droplets. Weak rainfall intensities are associated to particle diameters of a few millimetres. Moving from stratiform to convective precipitation (from black to red in the left panel) and consequently from small to high rainfall intensities

(following the legend in the right panel), the classes of small diameters are progressively depleted for the benefit of the bigger ones. The mode shifts accordingly to the right. Therefore, during convective precipitation not only a bigger number of raindrops, but also an increase in their average dimension has to be expected. Inside convective precipitation structures also the variance of the DSD is higher, both in space and time (see also the heterogeneity of the precipitation intensity depicted in Figure 1.1 with varying density of vertical lines under the cloud). During stratiform events the DSD is instead much more uniform throughout the rainfall field.

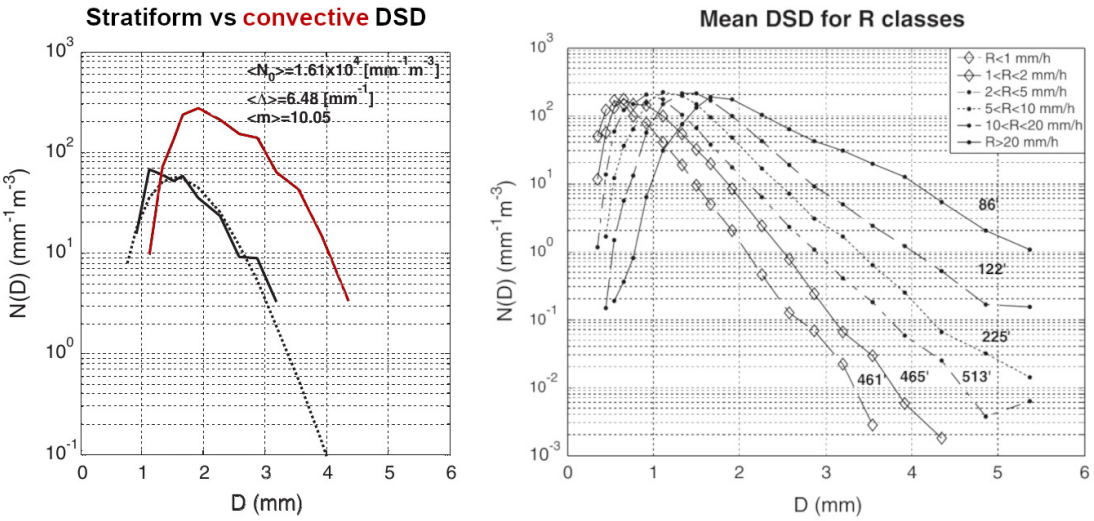

Figure 1.3: Different size distributions of raindrops. Left: averaged DSDs for two real events, one stratiform and one convective. Right: average DSDs for six rainfall rate categories. From Caracciolo et al. (2006a).

Average DSDs are modelled quite accurately by a three-parameter gamma function (Caracciolo et al., 2006a):

$$N(D) = N_0 \, D^m \, \exp(-\Lambda D) \tag{1.6}$$

where $N(D)$ is a time-independent DSD, $\Lambda = (3.67 + m)/D_0$ (mm^{-1}) is a scale parameter, N_0 $(\mathrm{mm}^{-1-m}/\mathrm{m}^3)$ is a concentration parameter, D_0 is the median volume diameter (mm) and m is the shape parameter. The stratiform DSD in the left panel of Figure 1.3 (black solid line) is indeed well approximated by a gamma distribution (black dashed line) with parameters as indicated in the top right corner of the left panel.

1.2 Raindrops and electromagnetic waves

1.2.1 Extinction scattering geometry: microwave links

In the scenario of a microwave transmission between two antennas, the signal is attenuated by obstacles encountered along the link path. The specific logarithmic attenuation $k(t)$ ($\mathsf{dB\,km^{-1}}$) of an electromagnetic wave signal is operationally defined as the ratio between received and transmitted power levels, thus the difference of their logarithms:

$$k(t) = \frac{Tx - Rx}{L} \tag{1.7}$$

where L is the link length (km) and Tx and Rx are the logarithmic transmitted and received power levels respectively, expressed in decibel milliwatt (dBm). For reference, logarithmic power level $P(\mathsf{dBm})$ is defined with respect to $1\,\mathsf{mW}$ as:

$$P(\mathsf{dBm}) = 10\log_{10}\left(\frac{P(\mathsf{W})}{0.001\,\mathsf{W}}\right) \tag{1.8}$$

being $P(\mathsf{W})$ the same power measured in watts.

Attenuation is modelled analytically by studying the scattering in extinction geometry. That is, everything which is scattered by the raindrops in any direction different from the incoming one does not reach the receiver. The parameter that controls the amount of radiation which gets to the receiver is the extinction cross-section $\sigma_{ext}(D, \lambda)$ ($\mathsf{mm^2}$), function of the diameter D of the scattering particle and the wavelength λ (mm) of the incident radiation. This means that also the scattering properties of the hydrometeors are related to their size distribution (van Leth et al., 2018).

The attenuation $k(t)$ will be expressed as:

$$k(t) = C_k \int_0^\infty \sigma_{ext}(D, \lambda)\, N(D, t)\, dD \tag{1.9}$$

Here we assume that the particle density is low enough such that multiple scattering can be neglected. $C_k = \frac{0.01}{\ln(10)}$ is a unit conversion factor.

The forward scattering amplitude matrix $\mathbf{S}(D)$ determines the amplitude of the outgoing (forward scattered) wave $\begin{pmatrix} E_{h0} \\ E_{v0} \end{pmatrix}$ as function of the incoming wave $\begin{pmatrix} E_h \\ E_v \end{pmatrix}$:

$$\begin{pmatrix} E_h \\ E_v \end{pmatrix} = \mathbf{S}(D) \begin{pmatrix} E_{h0} \\ E_{v0} \end{pmatrix} = \begin{pmatrix} S_{hh}(D) & S_{hv}(D) \\ S_{vh}(D) & S_{vv}(D) \end{pmatrix} \begin{pmatrix} E_{h0} \\ E_{v0} \end{pmatrix} \tag{1.10}$$

where Sij is the element of the scattering amplitude matrix for the component of radiation with incoming polarisation i and outgoing polarisation j, while v and h represent the vertically polarised and horizontally polarised components resp. (van der Hulst, 1957, 1981).

The raindrop extinction cross-section $\sigma_{ext}(D, \lambda)$ can be derived from $\mathbf{S}(D)$ from the following relation:

$$\begin{pmatrix} \sigma_{ext,h}(D) \\ \sigma_{ext,v}(D) \end{pmatrix} = \frac{\lambda^2}{\pi} \begin{pmatrix} S_{hh}(D) \\ S_{vv}(D) \end{pmatrix} \tag{1.11}$$

The elements of the forward scattering amplitude matrix can be calculated with the Mie scattering theory for spheres of arbitrary size and dielectric properties. For non-spherical drop shapes, they are instead calculated through the T-matrix approach. The relationship between the raindrop extinction cross-section and its diameter closely resembles a power law (Atlas and Ulbrich, 1977), similarly to the case of the terminal velocity parametrised by Equation (1.2). Equation (1.9) will then express some statistical moment of the drop size distribution too.

1.2.2 Backscattering geometry: weather radars

Weather radars work in backscattering geometry. A single antenna holds both transmitting and receiving roles. The electromagnetic pulses hit the hydrometeors and are scattered in all directions, but only the radiation coming back at the antenna is captured (scattering angle $\theta = 180°$). The ratio between transmitted and received power, leaving out all the spatial and angular dependencies, is proportional to a quantity called reflectivity η, which for hydrometeors small enough for the Rayleigh approximation to be valid, it is expressed as:

$$\eta = \frac{0.93\pi^5}{\lambda^4} Z \tag{1.12}$$

where 0.93 is the dielectric factor for liquid water. Reflectivity η is proportional to the reflectivity factor Z, which is the sixth statistical moment of the drop size distribution:

$$Z = \int_0^\infty D^6 \, N(D, t) \, dD \tag{1.13}$$

1.3 k-R and Z-R relations

Given what presented in the previous section, the rainfall intensity, the specific attenuation and the radar reflectivity factor are - with very few approximations - all statistical moments of the DSD. Therefore, they can be empirically related to each other by power laws:

$$k = \alpha R^{\beta} \tag{1.14}$$

$$Z = \alpha_z R^{\beta_z} \tag{1.15}$$

where α, α_z and β, β_z are fitted parameters.

By inverting (1.14) we obtain an expression of the rainfall intensity given the specific attenuation, again in a power-law form:

$$R = a k^{b} \qquad \text{with } b = \beta^{-1} \text{ and } a = \alpha^{-b} \tag{1.16}$$

Coefficients a ($\mathrm{mm\ h^{-1}\ dB^{-b}\ km^{b}}$) and b (dimensionless) all depend not only on the average DSD, but also on the temporal and spatial distribution of the DSD within the measured volume defined by the first Fresnel zone. Power-law parameters derived from a particular set of observations would then strictly speaking be valid only for links that have the exact temporal and spatial distribution and concentration of drop sizes within their path. This would mean that, even under the assumption of a uniform and unchanging DSD for a given climate, rainfall variability and intermittency within the link path volume would lead to inaccurate estimation of R with this relation, as a rain event evolves or passes over .

However, at the carrier wave frequencies typically employed in cellular communications links, the integrands in equations (1.3) and (1.9) are of similar magnitude (van Leth et al., 2018). As a result, the R–k relationship is almost independent of the DSD and the exponent b is close to 1 (Leinse, 2007). Therefore, a k-R relationship derived for this range of frequencies could be valid for a broad range of events.

Furthermore, because of the near-linearity of the relationship, parameters derived from either point measurements of the DSD or path-averages of the DSD can be used to derive path-average rainfall intensities from path-integrated attenuation in heterogeneous rain

fields:

$$\langle R \rangle = \frac{1}{L} \int_0^L a\,[k\,(s)]^b\; ds \approx \frac{1}{L}\, a \left[\int_0^L k\,(s)\; ds \right]^b = \frac{1}{L}\, a\, A^b \tag{1.17}$$

Here again L is the length of the link path, $k(s)$ is the specific attenuation at position s and A is the path-integrated attenuation. It is therefore demonstrated that it is possible to obtain path-averaged rainfall intensities from attenuation data of microwave radio links.

Leinse (2007) calculated a and b parameters specifically for the continental mid-latitude European climate and vertical polarisation. The Radiocommunication Sector of the International Telecommunication Union (ITU-R) provides a document with recommended coefficients for the k-R power-law relationship for both horizontal and vertical polarisations (ITU-R, 2005), meant for computing specific attenuation for given rain rates and for worldwide applications, Both coefficients are shown in Figure 1.4 as function of the carrier frequency of the link f (GHz). The b coefficient stays between 0.9 and 1.1 for frequencies between 20 and 35 GHz for all three models.

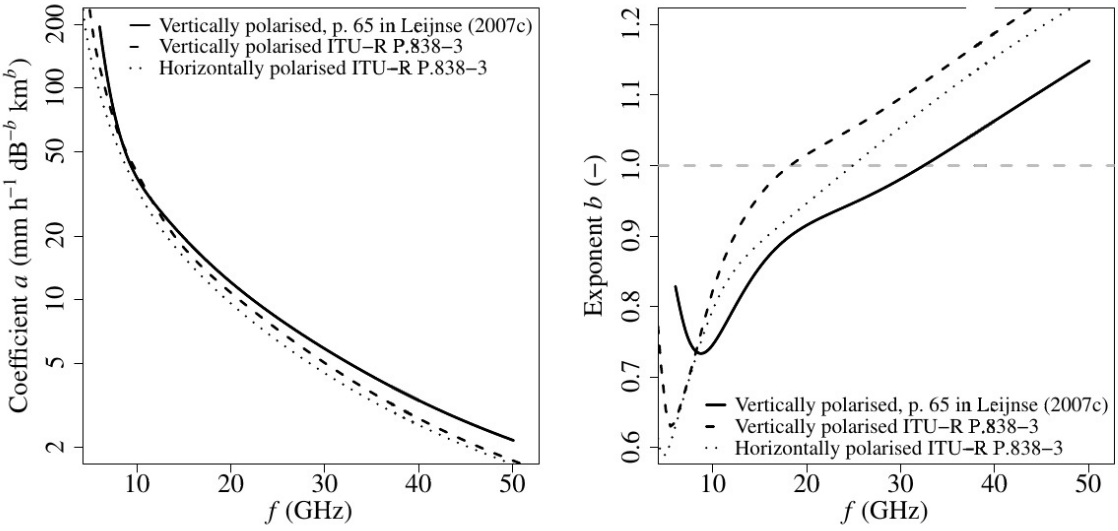

Figure 1.4: Values of coefficients a and b of the R-k relationship for frequencies ranging from 6 to 50 GHz. The grey-shaded area denotes the 37.0–40.0 GHz range. Note the logarithmic vertical scale in the left panel. Values have been computed from one dataset of measured drop size distributions (p. 65 in Leinse (2007); solid lines). The values recommended by the International Telecommunication Union (ITU-R, 2005) are also plotted (dashed and dotted lines for V and H polarised signals resp.). From Overeem et al. (2016a).

Low sensitivity of the k-R relationship to the DSD is a key factor to the exploitation of microwave links as meteorological instruments. Radar's Z-R relationship is instead markedly non-linear: usual standard α_Z and β_Z values are 200 and 1.6 respectively:

$$Z = 200\, R^{1.6} \tag{1.18}$$

Being β_Z greater than one, the sensitivity to the DSD is not negligible and all the assumptions made for the k-R relationship are not always valid. This makes it much more complicated to obtain consistent quantitative precipitation estimates from weather radars in different precipitation regimes.

2 Experimental area

2.1 Orography and hydrology

The physiography of Emilia-Romagna region is complex. The south-western part of the region consists of mountains: the Apennine ridge. The highest peaks (reaching about 2000 m a.s.l.) are located at the southern border. The north-eastern part is instead flat land: the alluvional plains of the Po river, which is the northern regional border. The eastern border is the Adriatic Sea. A map of the region is presented in Figure 2.1.

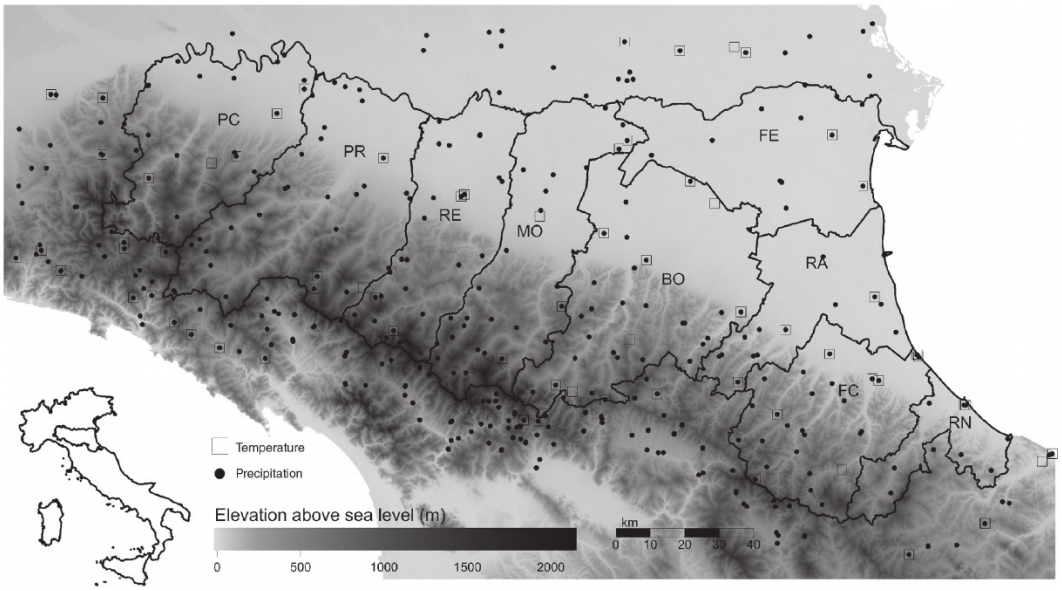

Figure 2.1: Map of Emilia-Romagna (Italy), administrative provinces (with codes) and orography from the regional high-resolution DEM. Dots and squares are the locations of the observational meteorological stations included in the 30 year climatological analysis of Antolini et al. (2016).

Rivers in Emilia-Romagna come from the Apennines and flow north-east toward the plains, where get into the Po river, after travelling suspended between embankments across the plains. Rivers east of Bologna reach the sea directly. River catchments are

located in the hilly region and have their closing sections short after entering the plains. Most of the cities and the densely populated and asset-rich areas lie near the closing sections of the mountain river basins (see Figure 2.2). The exposure of these areas to hydrogeological risks is therefore maximum and the importance of a precise and low-latency precipitation monitoring is evident. The flat land is vulnerable as-well: both the irrigation of the intensively cultivated fields and the drainage of the precipitation have to be obtained artificially, in a delicate equilibrium. Channels, pumps and water systems are maintained by various "Consorzi di Bonifica" (Land Reclamation Consortia).

Figure 2.2: Urban areas in Emilia-Romagna in 2008 in a scale from 0 (not at all urban) to 1 (completely urban). From Antolini et al. (2016).

2.2 Local climatology

Precipitation in the Po valley is well distributed along the year, with two peaks in spring and autumn, and relatively dry summers. Most of the analyses of this work are based on data collected in spring and summer months. During spring, precipitation is often related to cyclonic development with large frontal bands and moderate rainrates. Small scale convection can be triggered in many cases by orography. Frontal structures tend to come from west, carring the moisture of the Thyrrenian sea, or from north-east (strong katabatic winds). The maxima of the rainfall amounts are located on the Apennine ridge due to the orographic barrier they pose to the north-eastward moist fluxes. Annual cumulated precipitation (Figure 2.3) is deeply correlated with altitude.

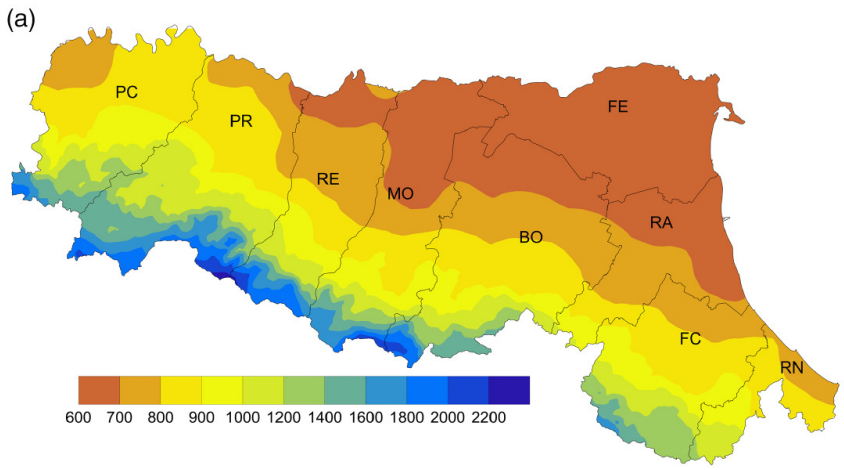

Figure 2.3: Mean annual cumulated precipitation (mm) averaged over 1961–2010. From Antolini et al. (2016).

The core of this work - the main validation described in Chapter 6 - is based on data coming from May and June 2016 and from the provinces of Bologna and Parma. A characterisation of the precipitation events that occurred over the target areas during the two months is shown in Fig. 2.4. An event is defined as any precipitation episode lasting for at least one hour (to grab also small scale thunderstorms) with at least two wet adjacent gridpoints of the $5\,\mathrm{km} \times 5\,\mathrm{km}$ grid that will be later described. Two consecutive dry hours are needed to separate subsequent events.

Most of the events occupy a very small area (below 10% of the target area, i.e. around $300\,\mathrm{km}^2$), and, in general, the average coverage of the events during their lifetime is below 60%. It has to be remarked that probably the events are not entirely contained in the target areas during their whole lifetime so that we can have only a partial view of the events. The maximum rainrate for each event has two peaks: one for very low rainrate below $5\,\mathrm{mm\,h}^{-1}$, probably due to stratiform precipitation, and a weaker one around $13\,\mathrm{mm}$, related to convective systems of moderate intensity. The duration of the events spans from one hour to one day, while most of the events produce small rain amounts, below $5\,\mathrm{mm}$ on the areal average. This analysis shows that the typical size of the events is small and, with some significant exception, characterised by low to moderate rainrate.

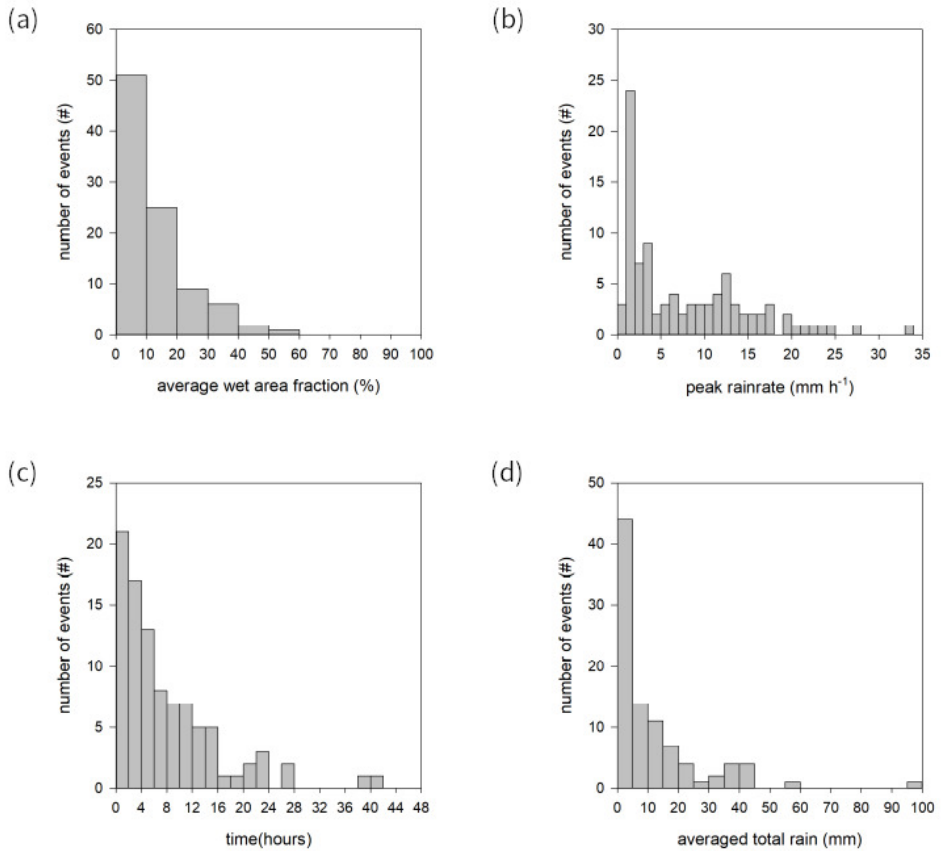

Figure 2.4: Summary indicators for precipitation: distributions for May and June 2016 in Emilia-Romagna.

3 Instrumentation

3.1 Commercial microwave links

Commercial microwave links (CMLs) are directional radio links that build the backhaul layer of the cellular communication network. The backhaul layer connects the backbone (mainly underground cables connecting data centers) to the edge of the network (the antennas which communicate with the mobile devices). An schematic idea of this hierarchy is given in Figure 3.1.

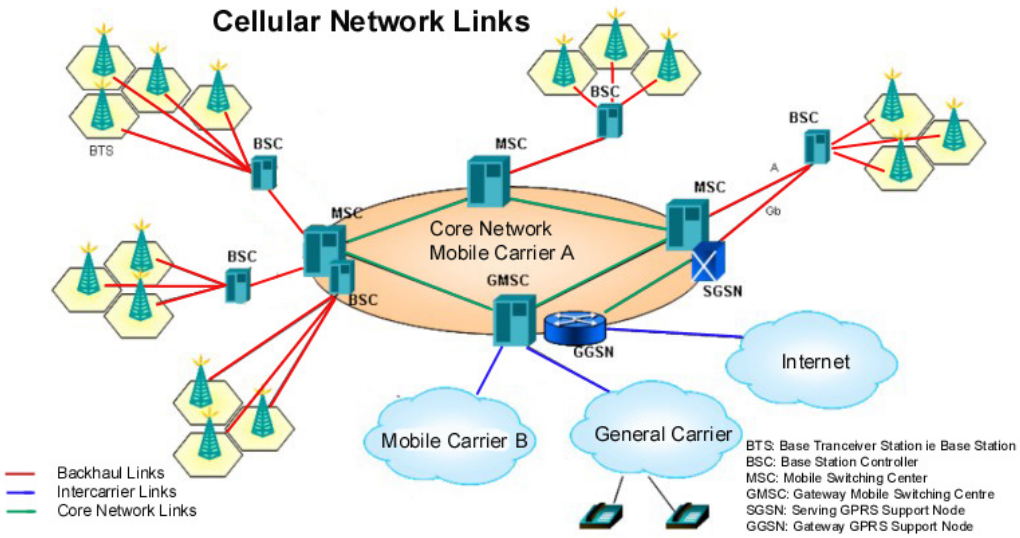

Figure 3.1: Network topology sketch, with backbone, backhaul and edge layers. CMLs are represented by the red lines. Credits: Oz MobileNet.

CMLs are a pair of parabolic antennas, often protected by a small radome (ideally water repellent), which are in free line of sight from one another. The antenna dishes have either 30 or 60 cm diameters and are installed at least 3 m above the ground, up to some dozens of meters when obstacles like buildings are present. They could either share the same pylons that host the edge layer of the network, represented by the rectangular 3G/4G antennas with high aspect ratio (Figure 3.2, left panel) or be found on dedicated

poles (Figure 3.2, right panel).

Figure 3.2: CML antennas over the Appennine crest, over M. Cimone (left) and Corno alle Scale (right), the two highest peaks of Emilia-Romagna. In yellow, an artistic impression of the electromagnetic beam around the line connecting the two antenna dishes (red line).

3.1.1 Opportunistic sensors

CMLs are not scientific instruments. Exploiting them to retrieve quantitative information about the precipitation is a so-called "opportunistic sensing" technique. Opportunistic sensing is possible when some valuable geophysical information is embedded in the hardware's operational data. In our case, information about the precipitation spatial and temporal patterns and intensity could be found in the attenuation which affects CML signals during rain events. Attenuation is actually an issue to deal with for telecommuication engineers. The opportunity of the sensing lies in transforming an issue to a source of information. Opportunistic also means that the hardware is already there for other purposes (in this case, telecommunication) and therefore installation and maintenance of the instruments are not responsibility and duty of the researchers.

Opportunistic sensing has advantages and disadvantages. Usually opportunistic geophysical sensors means a high quantity of data, a level of coverage of the territory that would be impossible with conventional instruments and low latency. Operational costs

are virtually zero, if the access to the data is granted cost-free (but we will see how this is not at all granted for now, at least in Italy). The main issue regards data quality, since the hardware has been certainly designed with purposes different from being an accurate scientific instrument. Not having direct control and ownership on your instruments exposes to reliability issues, like outages or abrupt setup changes. And it could also mean that you cannot have access to all the data you need or that you have to pay for it.

3.1.2 Links characteristics

CML path lengths span from $150\,\mathrm{m}$ to $30\,\mathrm{km}$; carrier frequencies approximately from 10 to $40\,\mathrm{GHz}$. Path length and carrier frequency tend to be anti-correlated in operative networks (Leijnse et al., 2008), because total path attenuation increases with both length and frequency (see Equation (1.7) and Figure 1.4). On the other hand, data bandwidth transmitted across the link increases with frequency and degrades with attenuation. CMLs are then usually operated at the maximum possible frequency before attenuation gets in the way. This translates to link sensitivities to rainfall between 0.1 and $1\,\mathrm{dB\,h\,mm^{-1}}$ (see Figure 3.3 and 3.4).

Transmitting powers are about $20\,\mathrm{dBm}$ ($0.1\,\mathrm{W}$), but - as will be later addressed - detailed information about transmitting powers has not been available for this work. CMLs frequently host two sublinks between the same antenna, back and forth. CMLs with two sublinks are called "duplex". Both horizontal and vertical polarisation are present, but usually sublinks of the same CML share the same polarisation and differ only in frequency by a small gap of around $1\,\mathrm{GHz}$. Multiple channels with the same carrier frequency but a different bandwidth could be present. Working with CMLs for meteorological purposes does not require by any means to have access to the data transmitted over the links themselves. All that is needed is the information of the power measured at the receiving end of the link and a time resolution sufficient enough to follow the rain-induced fluctuations.

3.1.3 Attenuation data

Opportunistic precipitation estimates can be derived from specific attenuation data (see Section 1.3). Specific attenuation can be derived from transmitting and receiving power levels according to Equation (1.7). Attenuation can also be derived from receiving power levels only, provided that the transmitting level remains constant over time and some

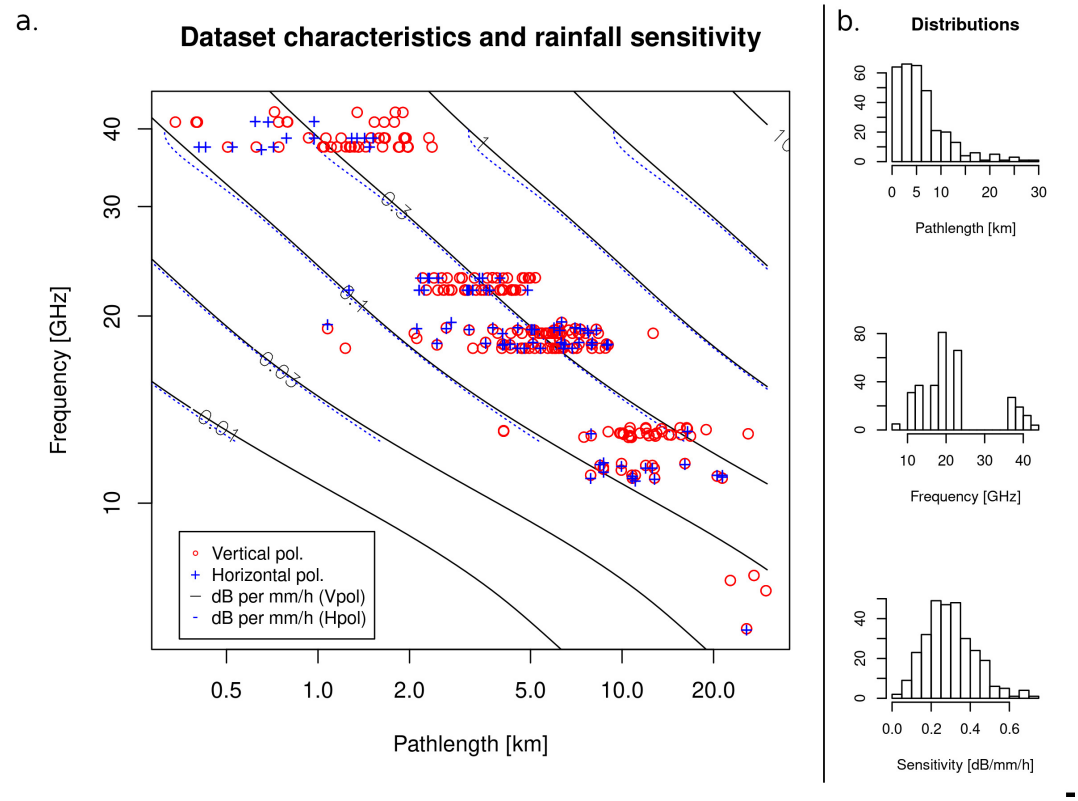

Figure 3.3: Path lengths vs frequency of Vodafone CMLs in the Bologna and Parma areas. Both axes are logarithmic. Sensitivities to rainfall for horizontal and vertical pol. are indicated as contour lines.

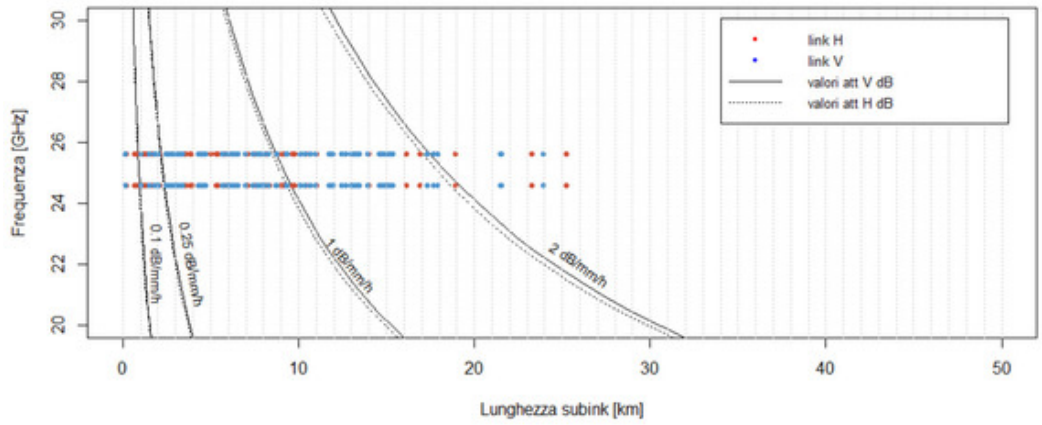

Figure 3.4: Path lengths vs frequency of Lepida CMLs in Emilia-Romagna. Sensitivities to rainfall for horizontal and vertical pol. are indicated as contour lines. Credits: Elia Covi, UniBo

reference baseline can be determined during clear sky periods. A method to determine dry (free path) and wet (attenuated) periods is then also necessary.

Tx and Rx power levels are logged and stored by the telecommunications providers for maintenance purposes. Received powers are measured by the antenna electronics with a resolution of $1\,\mathrm{dBm}$ at a cadence of $10\,\mathrm{Hz}$. But such dense information is apparently too redundant for engineering needs, since it is actually stored only at longer intervals or in an aggregated form. In our experience (and also looking at the literature) we have found two types of logged data: "instantaneous" readings and "MinMax" aggregated data.

Instantaneous data is built by arbitrarily sampling the $10\,\mathrm{Hz}$ original signal at a constant frequency, which could be 1 or $15\,\mathrm{min}$. Arbitrarily means that there is no connection between the signal and when it is sampled, it is just the value that happens to be read when the time interval expires and carries no information about what was the signal like between two subsequent readings. Power levels could then rise or sink between two consecutive instantaneous measurements without leaving any trace in the system.

MinMax data is instead the actual measurement of the dynamic range of the $10\,\mathrm{Hz}$ signal, sampled at a given time interval. Here all information is processed, but the information about the mean signal level is lost. MinMax data is in the format of $[P_{min}, P_{max}]$ pairs per time interval. It has been demonstrated from Ostrometzky and Messer (2020) that, once fixed the sampling frequencies, the MinMax format carries more information than the instantaneous one[1]. When the sampling frequency also changes and happens to be higher in the instantaneous data, it is not clear how much the two opposite contributions to information content will balance.

In some cases, the same CMLs under the same provider changed data logging format over time. This is the case for example for the T-Mobile network in The Netherlands, which moved from $15\,\mathrm{min}$ MinMax data (Overeem et al., 2016b) to $15\,\mathrm{min}$ instantaneous data (de Vos et al., 2019), thus actually losing information content in the process.

3.1.4 ATPC

As mentioned, signal attenuation during rainfall is actually an issue for the telecommunication providers. To avoid signal loss due to attenuation, CMLs are usually equipped with Automatic Transmit Power Control devices (ATPC) which modulate the transmit power to guarantee the most constant power level at the receiving end of the link, cancelling minor fluctuations of the total attenuation along the path. ATPC can boost the trans-

[1]This became clear to us in June 2019 during the CML Symposium in Garmisch

mitting signal up to 6 dB. If both Tx and Rx are known, ATPC should pose no threat to rainfall sensing, since attenuation is linear in the logarithm of the power and the attenuation estimate is conserved. In case of receiving powers only, an active ATPC disrupts completely the sensing capabilities, because the relation between attenuation and Rx is no longer unambiguous (Overeem et al., 2016a).

3.1.5 CML networks in Emilia-Romagna

In this work, two different networks by two providers were analysed. The two providers are Lepida S.c.p.A. and Vodafone Italia S.p.A.; both use microwave devices produced by SIAE Microelettronica S.p.A., although we have no information about the exact models.

Vodafone network data covers the administrative provinces of Bologna (BO) and Parma (PR). It consists of four months of 15 min MinMax receiving power levels from 357 duplex links with ATPC, accompained by data about the applied ATPC correction. A confidentiality agreement keeps us from showing the exact locations of the Vodafone links, so we have developed an algorithm to estimate how much of CML paths is travelling over a specific area ($\mathrm{km\,km^{-2}}$). If the area is a single interpolation pixel, we call that link length density "Local Coverage", or LC. The spatial distribution of the Vodafone network in terms of LC over $5\,\mathrm{km} \times 5\,\mathrm{km}$ pixels is shown in Figure 3.7. It can be seen that the greatest CML coverage follows the human presence. Highest density is achieved around the two capital cities and along the main communication routes. The hilly region of the province of Parma and the rural plains of the province of Bologna have instead the least covered grid boxes.

Lepida network serves the main purpose of supplying mobile connectivity to areas where commercial providers have no convenience reaching, as well as connecting public administrations and institutions. Its spatial distribution is therefore most unusual, being concentrated mainly in the hilly territories. Lepida data initially was 1 min instantaneous Rx readings from only 37 duplex CMLs with ATPC over Bologna and Modena (MO) provinces. Data from CMLs from the entire region were gathered, counting 346 duplex links, not always active at the same time (see Figure 3.5). Nevertheless, Tx power levels became available only by the end of 2020, after the conclusion of this preliminary work. A peculiarity of this network is that all the links work on just two frequencies: 24.6 and 25.6 GHz (see Figure 3.4).

Regarding this work, the preliminary analyses and the software developement were performed using both Lepida and Vodafone networks. The rigorous validation of the retrieval algorithm instead was carried out on the Vodafone network alone.

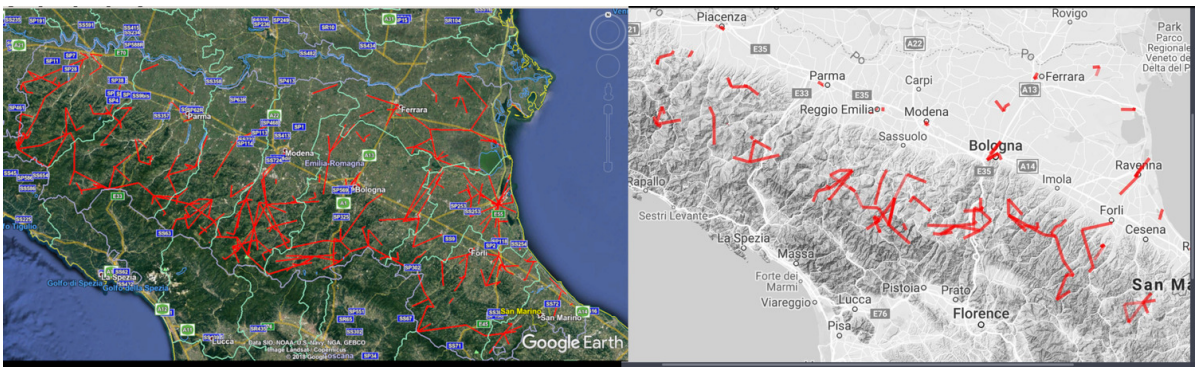

Figure 3.5: Lepida CML network: complete link catalog (left), including various links still to be installed, and links active on August 15th 2018 (right).

3.2 Rain gauges

Rain gauge networks provide spot measurements of the amount of rain fallen within the instrument's sampling area, which is some decimeters squared at most. Rain amounts are cumulated over time at intervals going from one minute to one day. They suffer known instrumental contraints, e.g. underestimation of high rainfall rates and low accuracy at low ones for a series of different reasons (Lanza and Stagi, 2012). Moreover, the point-like nature of their sampling, combined with the relatively low spatial density of the instruments, causes representativeness limitations of the rainfall fields derived from rain gauges (Porcù et al., 2014). Rain gauges are the reference instrument for precipitation monitoring in Emilia-Romagna.

Arpae RIRER (regional hydro-meteorological network) is a rain gauge network established in 2001 by bringing together existing hydrological and meteorological station networks, managed at the time by various public institutions and local authorities. The network of the whole Region is composed of 285 stations, equipped with tipping bucket rain gauges: 110 of them are divided between the provinces of Bologna (54) and Parma (56). Rain gauges have different sampling intervals (from 10 to 60 minutes), they undergo a process of homogenisation and quality control and are released as an hourly point-like product.

3.3 ERG5 rainfall analysis

The ERG5 gridded meteorological dataset has been developed by Arpae-SIMC, to support agricultural activities in the region of Emilia-Romagna. ERG5 data are opera-

tionally produced since 2001, interpolating the hourly station measurements of the main meteorological variables (air temperature, relative humidity, precipitation, wind, solar irradiance) onto a $5\,\mathrm{km} \times 5\,\mathrm{km}$ grid covering Emilia-Romagna.

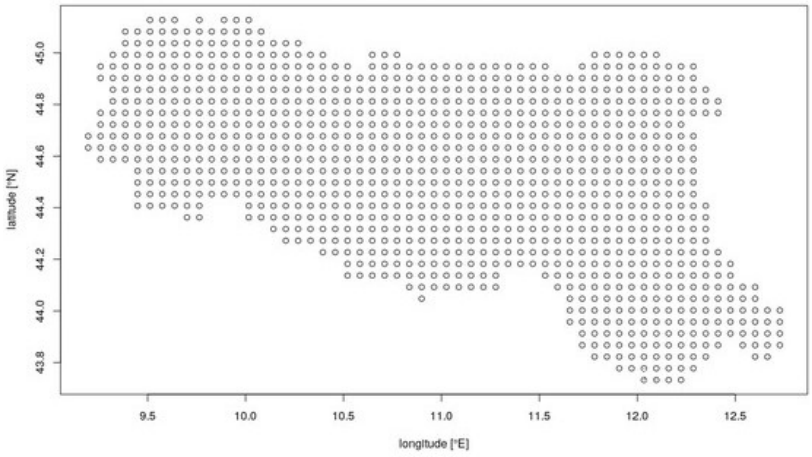

Figure 3.6: Arpae ERG5 interpolation grid. The grid step is $5\,\mathrm{km}$ in both directions.

The interpolation method used for hourly precipitation consists of a Shepard (1968) modified scheme using topographic distances instead of Cartesian distances. This allows the interpolation to take into account the influence of topography on precipitation, by making locations separated by orographic obstacles more distant than they would be if Cartesian distances were used (Antolini et al., 2016). Data are stored and distributed freely in the form of GRIB2 files, which were imported in an R environment thanks to the rNOMADS package (Bowman and Lees, 2015).

3.4 Weather radars

Ground-based weather radars are widely used by hydro-meteorological services to quantitatively monitor precipitation fields. Often deployed in large scale networks (Serafin and Wilson, 2000; Huuskonen et al., 2014; Saltikoff et al., 2019), radars are an effective trade off between spatial-temporal coverage and accuracy in the measurements. However, radar estimates are affected by several errors, which the last generation polarimetric systems have only partially mitigated (Figueras i Ventura et al., 2012; Gou et al., 2019).

The regional weather radar network is composed of two C-Band systems, located in San Pietro Capofiume and in Gattatico (easternmost and westernmost red crosses in Figure 3.7, respectively). For each instrument the equivalent radar reflectivity factor

close to the ground is extracted and interpolated from polar coordinates to a 256×256 Cartesian grid of $1\,\mathrm{km}\times1\,\mathrm{km}$ resolution, then merged to obtain a composite of both radars.

Raw radar images are affected by various non-meteorological echoes that are removed before computing the Quantitative Precipitation Estimation (QPE). The current scheme used at Arpae-SIMC during operational service includes many steps: the ground clutter is removed at first statically through the map of signal-free elevations recorded in dry conditions, then dynamically by combining a beam trajectory simulation at the current atmospheric state (as measured by radio soundings) and a Digital Elevation Model (Fornasiero et al., 2006). The beam blocking reduction and correction is performed based on a geometric optic approach (Bech et al., 2003), while anomalous propagation is detected after the analysis of the echo coherence in the vertical direction (Alberoni et al., 2001). The final conversion between reflectivity and rainfall rate is performed on the corrected dataset using the classic relationship $Z = aR^b$, with $a = 200$ and $b = 1.6$.

Rain rates are obtained every 5 minutes and hourly total rain amount is computed by an advection algorithm which takes into account the movement of the precipitating systems. The algorithm is based on the computation of the maximum cross-correlation between consecutive maps, leading to the estimate of the displacement vector for each precipitating system. The rainfall field is then reconstructed every minute between the observations and cumulated over each hour. Finally, radar QPE is adjusted with rain gauges data, via the spatial analysis of the ratio G/R between rain gauges (G) and radar (R) rainfall rates over the station locations. The spatial analysis is obtained as the weighted mean of the G/R values where the weight is a function of both the distance of the grid point from the station and the mean spacing between 5 observations (Koistinen and Puhakka, 1981; Amorati et al., 2012). In this work we will compare the CML product with both adjusted and unadjusted radar QPEs.

Instruments location can be seen in Figure 3.7: rain gauges as blue dots, weather radars as red crosses (with $100\,\mathrm{km}$ radius circles around them to give a scale of their range, even though they could reach more than double distances in some conditions). Coloured pixels are LC for the Vodafone network, calculated over the ERG5 $5\,\mathrm{km}\times5\,\mathrm{km}$ grid. Contour lines represent the orography.

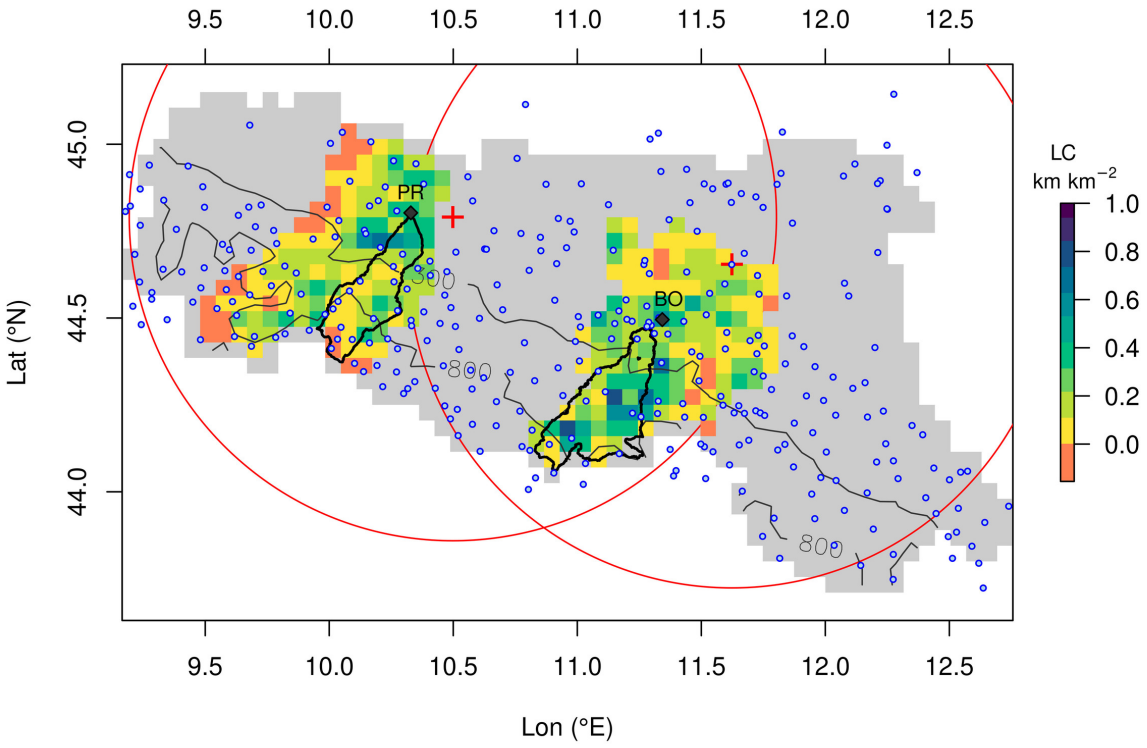

Figure 3.7: Map of the Emilia-Romagna region in Northern Italy (grey area) with instrumentation locations. The coloured areas are the administrative provinces of Bologna (BO) and Parma (PR). The colour scale represents the Link Coverage, LC, where orange is not a negative value but exactly zero. Black thick lines delimit the two river basins (Parma, to the east, and Reno). Blue dots and red crosses indicate operational raingauge and weather radar locations, respectively, while red circles are the 100 km radar coverage. Thin black lines show two elevation contours (300 and 800 m a.s.l.). The capital cities of the two areas (Bologna and Parma, resp. BO and PR) are indicated with the black diamonds.

4 Methods

4.1 RAINLINK rain retrieval algorithm

The open-source RAINLINK algorithm (Overeem et al., 2016a) is published on GitHub (`https://github.com/overeem11/RAINLINK`) in the form of an R package. We used the 1.14 version of the RAINLINK algorithm, available online from July 2019. We added some minor modifications and optimisations to the original version. They are presented in Section 5.2.4. The forked version of the software is available here: `https://github.com/giacom0rovers1/RAINLINK`.

The algorithm works for both instantaneous power measurements and $[P_{min}, P_{max}]$ pairs: for the present work we use the latter, on 15 minutes intervals. The algorithm treats P_{min} and P_{max} separately (we will then use P_i to refer to both alternatively). Two separate rain estimates R_{min} and R_{max} will thus be obtained. The retrieval process is summarised below, while we show more details of the data filtering in Section 6.3.

1. **Preprocessing**: the raw input goes through three consistency checks concerning data formatting and labelling. Any multiple observations for the same LinkID and DateTime are discarded, each LinkID is verified to maintain the same metadata throughout the whole dataset (Frequency, PathLength and antenna coordinates) and rows with NA values in any of the columns except for Polarisation (which is supposed vertical if not indicated) are discarded as well.

2. **Wet-Dry Classification**: the samples are discriminated in wet and dry periods by assuming that rainfall is correlated in space, through the so-called Nearby Links Approach (NLA), which works as follows. For each link, a time interval with a decrease in the received power is labelled as wet if at least half of the links in the vicinity (within $15\,\mathrm{km}$ radius) experience a comparable reduction, i.e. if the medians of the attenuation and the specific attenuation of the nearby links are below $-1.4\,\mathrm{dB}$ and $-0.7\,\mathrm{dB\,km^{-1}}$ respectively. This is the second most computationally time-consuming step of the algorithm.

3. **Baseline determination**: a $24\,\mathrm{h}$ moving-window median of the quantity $\frac{1}{2}\left(P_{min} + P_{max}\right)$ over the dry time intervals defines a reference level P_{ref} (baseline). Computationally, this is the most time-consuming step of the algorithm.

4. **Outliers filter and power correction**: outliers due to malfunctioning links can be removed again by assuming that rainfall is correlated in space. The filter discards a time interval of a link for which the cumulative difference between its specific attenuation and that of the surrounding links over the previous $24\,\mathrm{h}$ (including the current time interval) becomes lower than the outlier filter threshold, which is fixed at $-32.5\,\mathrm{dB\,km^{-1}\,h}$. After removing the outliers, the classification information is used to clean the receiving powers from the noise over the dry periods. The corrected powers P_i^{Cor} will be equal to P_{ref} on dry periods and P_i on wet ones.

5. **Rainrate retrieval**: attenuation A_i is computed as $A_i = P_{ref} - P_i^{Cor}$. A fixed quantity $Aa = 2.3\,\mathrm{dB}$ is subtracted from the attenuation A_i in order to take into account the wet-antenna effect, which is independent of the path length and assumed independent also of frequency and rain intensity. If $A_i - Aa > 0$ then the specific attenuation k_i ($\mathrm{dB\,km^{-1}}$) is calculated as $k_i = \left(A_i - Aa\right)/L$, otherwise 0 is returned. Path-averaged mean rain intensities R_i ($\mathrm{mm\,h^{-1}}$) are finally calculated using the $k - R$ relationship $R_i = a\left(k_i\right)^b$, where the coefficients a and b were from Leinse (2007) and Leijnse et al. (2010) for vertical and horizontal polarisation, respectively.

6. **Path-averaged rainfall depth**: to obtain a single path averaged rain depth, R_i are combined through a weighted mean: $R = \frac{1}{4}\left[\alpha R_{min} + \left(1 - \alpha\right) R_{max}\right]$. The factor $\frac{1}{4}$ transforms rain rates into 15 min rain depths. The weight α varies between 0 (estimate derived from P_{max} only) and 1 (estimate derived from P_{min} only); we adopted the default value ($\alpha = 0.33$). We specify that, unlike Overeem et al. (2016a), we chose to keep the subscripts related to the original receiving powers: in our notation the rain rate R_{min} is higher than R_{max} because it is obtained from the most attenuated signal P_{min}.

7. **Interpolation**: CML path averaged precipitation estimates are assigned to the mid points of the links like point measurements ("virtual rain gauges"). Interpolation of the point-like measurements is performed at hourly scale with ordinary kriging on a spherical semi-variogram on the ERG5 grid. Sill and range parame-

ters are estimated from the available rain gauge stations of three consecutive years. Nugget parameter is set as 1/10 of the sill, as in Overeem et al. (2016a). The interpolated field is truncated if it gets smaller than $0.05\,\mathrm{mm}$, which is half of the minimum detectable rain from a rain gauge.

4.2 Error metrics

Two sets of classical skill indicators, broadly used by the validation community (Nurmi, 2003), have been used to evaluate the products performances. The first set consists of four categorical indicators. They assess the capability of the product to detect rainfall occurrence. They are computed after a definition of a confusion matrix by counting the number of samples where both estimate and observation agree on classifying wet (hits, H), or dry (correct negatives, CN) samples, and where there are misses (M, observed wet and estimated dry) or false alarms (F, observed dry and estimated wet). Namely, Probability of Detection, False Alarm Ratio, Multiplicative Bias and Equitable Threat Score are defined respectively as:

$$POD = \frac{H}{H + M} \tag{4.1}$$

$$FAR = \frac{F}{H + F} \tag{4.2}$$

$$MB = \frac{H + F}{H + M} \tag{4.3}$$

$$ETS = \frac{H - H_{rnd}}{H + M + F} \tag{4.4}$$

where H_{rnd} represents the number of hits obtained by chance.

The second one consists of four continuous indicators. They evaluate the skill in correctly estimating the quantitative precipitation rate. Given e_i and o_i as estimated and observed values respectively, continuous indicators are the normalised Mean Error

and the normalised Mean Absolute Error, defined as:

$$ME = \frac{\sum_i (e_i - o_i)}{\bar{o}} \tag{4.5}$$

$$MAE = \frac{\sum_i \|e_i - o_i\|}{\bar{o}} \tag{4.6}$$

and the Coefficient of Variation (CV), defined as the root mean square error divided by the mean of the observed values $\bar{o}$, and the Pearsons' Correlation Coefficient (CC), as the covariance of observed o_i and estimated values e_i divided by the product of the two standard deviations (Nurmi, 2003; Overeem et al., 2016b).

Both the interpolated CML and the reference field have a large number of very low positive values (below $0.1\,\mathrm{mm\,h^{-1}}$) that have very limited physical relevance, but which are potentially very influential in normalised error metrics. Thus we have set a wet-dry threshold equal to the minimum rain quantity detected by the tipping bucket rain gauge, i.e. $0.1\,\mathrm{mm\,h^{-1}}$, for both estimate and reference. Categorical indicators are calculated with respect to this threshold for the whole dataset, while all the continuous indicators are computed only for the product-reference pairs where both values exceed the threshold (i.e. wet-wet). ME, MAE and CV are normalised with the averaged reference rain depth.

5 Building the framework

5.1 The RainBo Life EU Project

A European project co-financed by the LIFE Program, called BLUEAP, assessed in 2015 a particular vulnerability of the city of Bologna to severe rainfall events and flash flooding of small urban watercourses. In September 2016 the Life EU Project RainBO[1] (LIFE15 CCA/IT/000035) started as BLUEAP follow-up, with the goal to develop and implement solutions to mitigate hydrological risk and increase urban resilience to meteorological extremes. Lepida S.c.p.A. was the project coordinator; the involved partners were Arpae Emilia-Romagna- SIMC, Nier Ingegneria S.p.A., MEEO S.r.l. and the Municipality of Bologna. The project ended on 7 June 2019.

Figure 5.1: Logos of RainBO (LIFE15 CCA/IT/000035) project partners.

Opportunistic sensing of rainfall through CMLs was chosen as a technologically advanced and economically sustainable method for constant environmental monitoring by public administrations and services. Specific attention was given to real-time and early warning capabilities. The task of developing a framework for CML low-latency rain estimates was assigned to MEEO S.r.l., an IT company based in Ferrara (Italy), focused on satellite data and Earth monitoring, where I spent four months of internship to familiarize with CML data and rain retrieval software.

RAINLINK (Overeem et al., 2016a) was selected as retrieval algorithm (see Section 4.1). In the early months of 2017, two different CML data sources became available: Vodafone S.p.A. and Lepida itself. Lepida S.c.p.A. began logging *Rx* data from its CML network shortly after the start of the project, starting a collaboration with the University

[1] `www.rainbolife.eu`

of Bologna which still continues to date. However, the initially gathered data were not many and the CMLs were scattered over a wide area, thus another batch of CML data was purchased from Vodafone S.p.A. by MEEO S.r.l. with project funds, specifically over the areas of major interest: the cities of Bologna and Parma and the river basins of the rivers flowing through them.

5.2 Working with RAINLINK

5.2.1 Getting started: data formatting

RAINLINK needs the interpreter for the R programming language to be installed on the machine. It also needs some dependency packages to be installed both inside and outside the R environment. I wrote a script to automate these steps, in order to have fast deployment capabilities of the system. After the installation of the dependencies, R works out of the box with default settings tailored for The Netherlands. In order for the interpolation to work in other areas of the world, a grid of coordinates specific to the area of interest should be provided. For Emilia-Romagna we used the grid of the ERG5 interpolated rainfall product (see Figure 3.6).

The RAINLINK software requires a specific redundant data formatting, in which every row is self sufficient in itself and contains both data and metadata. Specifically, every pair of MinMax power level readings is associated to a date-time string, a unique link ID, but also the coordinates of the two antennas of the link and its path length and frequency (an example from the sample dataset provided with the program is given in Table 5.1.).

Table 5.1: Data format required by the algorithm and first rows of the sample data provided with the RAINLINK package and coming from The Netherlands, relative to three days of September 2011.

Frequency (GHz)	DateTime YYYYmmddHHMM	Pmin (dBm)	Pmax (dBm)	PathLength (km)	XStart (°E)	YStart (°N)	XEnd (°E)	YEnd (°N)	ID
38.0	"201109090815"	-57	-55	2.7	5.598	52.136	5.566	52.122	68
37.9	"201109090815"	-47	-47	3.2	5.454	52.168	5.460	52.196	81
25.4	"201109090815"	-52	-51	6.2	5.612	52.185	5.662	52.233	73
38.0	"201109090815"	-44	-43	3.8	5.514	52.208	5.460	52.196	74

Usually raw CML data comes instead in power readings associated to timestamps and IDs only. Building the metadata database, verifying its correctness and consistency and

standardising spatial and temporal coordinate systems takes most of the initial setup effort.

5.2.2 Test runs with synthetic data

Initially there was no data available over the region. A simulation tool was developed to test the software package over the new area. A pseudo-randomised Monte Carlo process was used to generate an imaginary network of links over the target area. The total number of links was set equal to the one in the NL 2011 sample, with roughly the same path length and frequency distribution. At that time, it was not taken into account that the province of Bologna is roughly one fifth of The Netherlands in surface area. Average link density was therefore overestimated by a factor of five. Outputs in fact show no coverage related issues in the reconstructed rain fields.

Synthetic precipitation was then generated. Equation (5.1) produced a series of advecting rain bands, while Equation (5.2) produced instead an advecting convective cell.

$$R(t) = \begin{cases} R_0 \sin\left[2\pi\left(\kappa r - \omega t\right)\right] + \xi & \text{if} \quad \sin\left[2\pi\left(\kappa r - \omega t\right)\right] \geq 0 \\ 0 & \text{if} \quad \sin\left[2\pi\left(\kappa r - \omega t\right)\right] < 0 \end{cases} \tag{5.1}$$

$$R(t) = R_0\, e^{-\frac{1}{2}(\kappa r - \omega t)^2} + \xi \tag{5.2}$$

where r is the distance from the lower left corner, k is the wave number, ω is the frequency, R_0 is the maximum amplitude and ξ is a small random noise. The rain fields advects north-west at speed ω/κ.

5.2.3 From a research software to a near real-time online service

RAINLINK was born as research software. To make it run in automatically in a real-time service perspective, many tools had to be built around it. Specifically: a data collection service, a formatting and preprocessing unit and a suite of visualisation tools. A preliminary analysis tool and an output validation module were also planned. The software was called Rainlink4EMR (see Figure 5.3).

A more advanced and definitive edition of Rainlink4EMR (2.0) was written in Python by Moris Pozzati (MEEO S.r.l.) based on version 1.0, featuring daemon services running

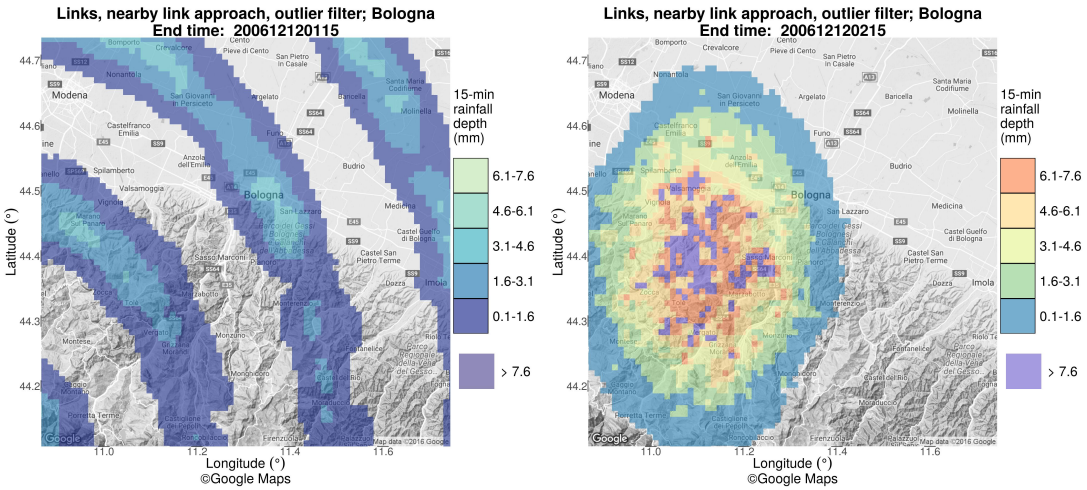

Figure 5.2: Simulated detection of synthetic rainfall patterns over a random CML network in the Bologna area with default RAINLINK parameters.

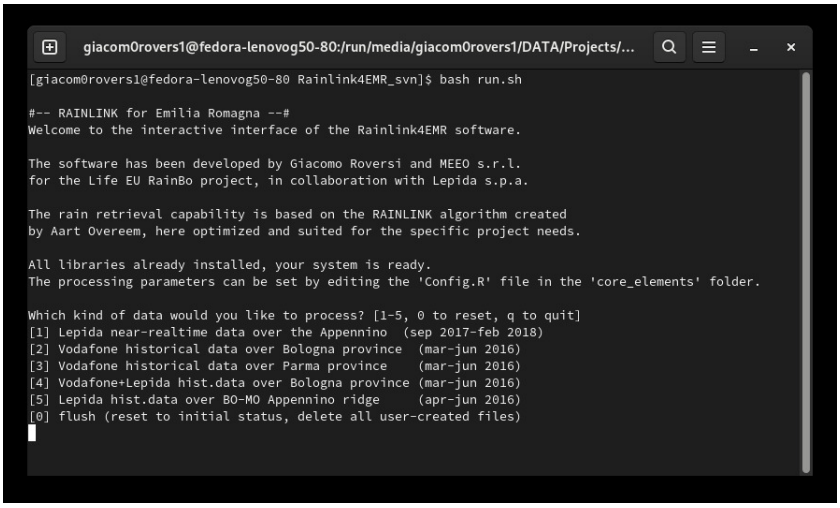

Figure 5.3: Main menu of the user interface of Rainlink4EMR (1.0), written in Bash.

in background to automatically collect and process data every 15 min. Version 2.0 is now running on Lepida S.p.a servers.

5.2.4 New functions for RAINLINK

The RAINLINK algorithm takes a long time to run on large datasets. With the aim of speeding the workflow up, new functions have been added to the original RAINLINK software package. For example, the approach to rainfall cumulation (hourly and daily) through `if()`s and `for()` cycles was abandoned in favour of a more computa-

tionally efficient one that exploits large data matrices. The custom functions are called `fast50x_accu1hr()` and `fast50x_accu24hr()` because tests showed faster cumulation times by a factor of 50 with respect to the original method.

The R packages that were added are the following.

The `tidyverse` suite provides functions for faster database pivoting, modern data analysis and manipulation tools and new plotting capabilities.

The `raster` package is specifically designed for georeferenced data fields. It allows stacking of multiple layers, performs mathematical operations both horizontally between pixels and vertically through the layers stack, pixel by pixel comparison of different products, summaries and statistics and so on. Maps plotting with colour scales and contours is also much more straightforward. The old graphics output can be viewed in Figure 5.2, while the new one in Figure 3.7 , 6.9 and 6.12.

The `txtProgressBar` package provides useful live information about the progress of the processing directly in the system terminal.

All new packages have been added to the dependencies list.

The original method for data visualisation required a file specifying each map pixel as edge coordinates of a polygon centred over the interpolation grid. Creating a new set of polygons each time the interpolation grid changed was seen as a limiting factor, thus a script was developed to automate the process (see the `PolyGridGen()` function). Data visualisation via polygons proved also to be the source of some coding bugs that required some time and the direct collaboration with RAINLINK's author Aart Overeem on GitHub to be solved. It was then decided to drop all the original visualisation code in favour of the new approach with rasters introduced above. To date, a polygons grid file is nevertheless still required in order for the software not to crash.

5.3 RAINLINK setup for Emilia-Romagna

The implementation of RAINLINK in Emilia-Romagna required some technical and conceptual considerations, based on either the differences or the similarities between northern Italy and The Netherlands regarding climatology, orography and features of the CML networks.

5.3.1 Considerations about network characteristics

The CMLs' operational frequency in our Vodafone dataset spans between 5.0 and 45.0 GHz. The default frequency allowance window of the RAINLINK algorithm is 12.5 - 40.5 GHz instead. We decided to extend it to 10.0 - 45.0 GHz but no further. Five CMLs belonging to the 5 to 10 GHz interval were left out. We also removed ten other links which had higher frequencies but whose sensitivities were below $0.1\,\mathrm{dB\,h\,mm^{-1}}$ (more details in Section 6.3) to avoid contamination from coarse, low sensitivity signals.

The average link density of the Vodafone network ($0.043\,\mathrm{km\,km^{-2}}$) is the same as the one of the network used for the original setup of the algorithm (see Table 6.1).

Lepida links have all the same carrier frequency of around 25 GHz, but very different path lengths (from a few hundreds of meters to over 20 km). This produces very different sensitivities to rainfall, even though almost all links are above the $0.1\,\mathrm{dB\,h\,mm^{-1}}$ threshold. Average link density was not calculated for the Lepida network because of its irregular layout. Integrated rainfall fields coming from that network showed clearly that Lepida CMLs are too sparse to attempt a spatial interpolation at the regional scale (see Section 5.5).

5.3.2 Considerations about climatology

Spherical variogram parameters (see Section 4.1, point 7., "Interpolation" and Section 1.1.2) were provided by Arpae-SIMC for three years from a pool of RIRER validated raingauges coming from the entire region. Range and Sill are 36.12 km and $1.12\,\mathrm{mm^2}$ respectively. These values resemble very much the median for May and June of the outputs of the "ClimVarParam" sub-function of Overeem et al. (2016a), which approximates thirty years of Dutch climate (van de Beek et al., 2012). Accordingly, it is expected that both the network structure and the rainfall spatial patterns are similar between the Italian and Dutch sites. This assumption drives the choice for the correct value of the NLA radius of the wet-dry classification algorithm.

The default k-R relationship from Overeem et al. (2016a) is also maintained unchanged, since northern Italy and The Netherlands share a similar climate. The average drop size distributions (DSD) differences between the two countries are expected to be negligible (Caracciolo et al., 2006b) and certainly lower than the expected variations of DSD along the link paths and during the 15 minutes time intervals (Tokay et al., 2017).

5.3.3 Considerations about orography

The differences from The Netherlands regarding orography are more relevant. We expect that rainfall patterns could deviate from the average behaviours described by the variograms when interacting with the complex orography of the hilly part of the region. However, we do not have enough data to calibrate the NLA radius at a small scale or considering geographical sub-samples. Moreover, a shorter NLA radius could theoretically improve the consistency with the expected decorrelation length, but, since the network in the hilly region mostly consists of medium to long links, candidates which will fall inside the NLA radius could be too few to ensure a statistical significance of the samples. Thus, we left its value unaltered at 15 km, but we are expecting that some issues could possibly arise in the areas characterised by the most heterogeneous terrain.

5.3.4 Why avoiding a regional calibration

All the other algorithm's parameters were not specifically calibrated. The reasons behind this out-of-the box approach are numerous:

- As suggested by their authors (Overeem et al., 2016a), a solid calibration of the RAINLINK retrieval algorithm should be implemented exploiting numerous instruments along the link paths and organising dedicated measurement campaigns, which were not feasible for us.
- The overall temporal span should also allow the dataset to be split into two non-overlapping datasets for calibration and validation, but the total wet hours available to us were not enough to grant a statistical significance for both sub sets.
- The gauge-adjusted radar product (which is commonly exploited in most CML studies) is not the one currently selected by the regional weather agency Arpae-SIMC as their quantitative reference, a choice that went in favour of the interpolated rain gauges product ERG5 (see Section 3.3). The spatial and temporal resolution of ERG5, however, is too low to perform an effective calibration.

Therefore, we analysed some CML - rain gauge pairs only where the gauges were already in the vicinity of the links (Section 6.4.1), while we validated the rest of the dataset against the reference only through its interpolated product (Section 6.4.2 and 6.4.2).

We considered RAINLINK's ability to function as a standalone system – while other approaches rely on gauges or radars for Wet-Dry classification – as one of its key features. However, since RAINLINK did not include any standardised algorithm or procedure for calibration, performing it would lead to a huge increase in the set up efforts, which would make other algorithms (where adaptation to local characteristics is naturally present, e.g. neural networks) much more competitive. It is a very recent and welcome news the publication of a specific stochastic tool for a standardised RAINLINK calibration (Wolff et al., 2022). This tool could maybe lead to different choices and evaluations about calibration in future similar works.

5.4 Early challenges

Working with CML data was not plain and straightforward at the beginning. Mostly all the people involved were new to the challenge of the scientific (meteorological) exploitation of those data. It took a great effort and the work had to be restarted from scratch a few times. In this Section are described the main issues we have encountered.

5.4.1 Dataset issues

Invalid or missing data

The Vodafone Italia S.p.A. dataset originally consists in four months of data: from March to June 2016. First statistics showed a constant number of CMLs through the whole four months. Lower performance for the colder months was thought to be coming from unfavourable climatological conditions (light rain). But later deeper analyses on the intermediate products showed that around half of the links were not producing valid estimates but were rejected by the preprocessing of the algorithm. They were in fact invalid or null data (see Figure 5.4). March and April 2016 had to be rejected. Initially, due to non-standardised practices of data acquisition, even May 2016 for Parma was missing, but fortunately it had been later recovered.

Time shifts

Vodafone timestamps' time zone was initially unknown (sic). It was noted that a gap associated to the change from standard time to daylight saving time was weirdly not present. Later (autumn 2017), during synchronisation attempts with CMLs of the Lepida

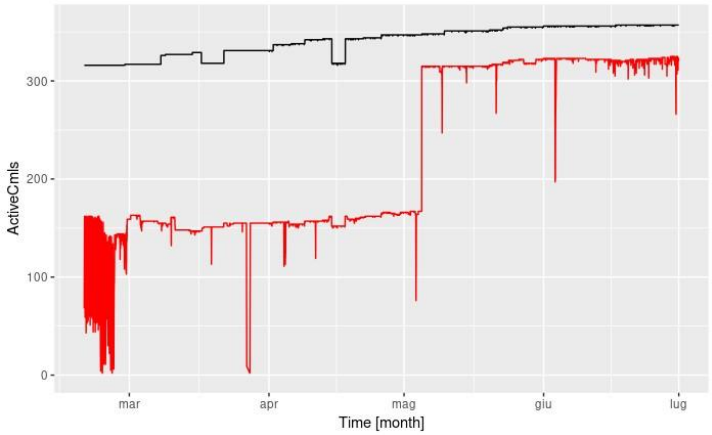

Figure 5.4: Number of CMLs in the dataset for every timestamp: total number of links (black) and accepted as valid by the algorithm (red).

network, which have their timestamps in non-ambiguous UNIX time[2], it was established a constant UTC+2 time zone for the Vodafone dataset throughout the entire year.

Merging attempts

In the early project stages, the Lepida network (counting then only 37 duplex links) was considered a good integration of the Vodafone one, since it filled its gaps in the least covered areas by construction (see Section 3.1.5). It was therefore developed a way to merge the two datasets together. The effort was successful and also led to the individuation (and solving) of the time-shift issue just described.

The merging was carried out aggregating the instantaneous power levels of the Lepida network in batches 15 min long, same as the time resolution of the Vodafone network. From those batches the minimum and maximum *Rx* powers were extracted and considered MinMax pairs. In July 2019 - after discussing the matter also with international experts - it became evident that that was not the correct approach: instantaneous measurements do not contain any information about the signal intensity between two consecutive readings, which means that the minimum and maximum instantaneous values are not truly representative of the real dynamic range inside the time interval. Results were nevertheless promising, but given such a flaw already in the preliminary hypotheses, any further attempt to merge the two different datasets was abandoned.

[2]Or UNIX Epoch or Posix time, number of seconds elapsed since 00:00:00 UTC January 1, 1970

ATPC devices

The CMLs analysed in this work are all equipped with ATPC (see Section 3.1.4). Unfortunately we did not know the transmitting powers (Tx) of the links. In the Vodafone case, it was because of confidentiality restrictions; as for Lepida, Tx logging and sharing was not yet implemented.

But Vodafone engineers were able to provide us some ATPC data. Specifically: the maximum modulation offsets (in dB) that had been applied during each time interval. Through that information we are able to correct the receiving power levels, compensating for the power modulation effects and simulating CML data with constant transmitting powers and allowing RAINLINK to estimate attenuations from receiving powers only.

The correction we have conceived intervenes only on minimum received powers (P_{min}) because they are undoubtedly affected by the ATPC. They are manually lowered by the maximum ATPC modulation applied within the respective $15\,\mathsf{min}$ time window. Maximum receiving powers (P_{max}) instead are left untouched. The ATPC working frequency and the $15\,\mathsf{min}^{-1}$ sampling frequency do not correlate and there was no way to predict the real P_{max} level and infer a reasonable compensation. This could result in a broader gap between P_{min} and P_{max}.

During the processing we discovered that Lepida CMLs were also equipped with ATPC devices. That forces us to abandon the Lepida dataset, which just showed promising results (see Section 5.5), until 2021 when also Tx power levels became available.

5.4.2 Anomalous signals

Two case studies of anomalous attenuation signals deserve to be documented, even though an effective rain retrieval algorithm should reject them as outliers.

Sunny morning near Budrio

In the data from early days of March 2016 a big attenuation signal showed up in one link. It was more intense than other signals caused by rainfall in the following days. It endured various hours and showed a less irregular profile than usual rainfall attenuations (see Figure 5.5). Interestingly, the discrepancy between that link and its neighbours disappeared completely during rain events, when all links behaved similarly.

The anomalous attenuation produced a spot-like artefact over the town of Budrio (BO), clearly visible during dry periods but also noticeable in daily accumulations of days with variable weather (see Figure 5.6). Further inspections found an evident corre-

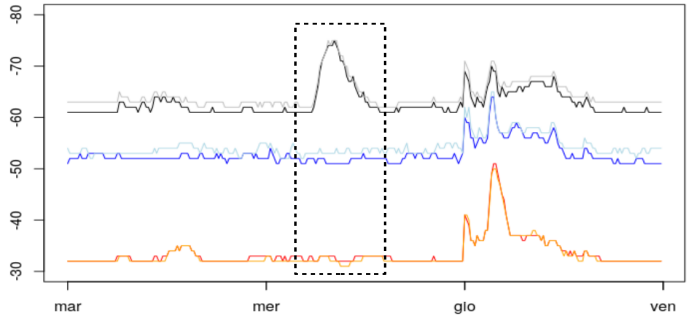

Figure 5.5: *Rx* power levels from three neighbouring CMLs near Budrio (BO). Only the two sublinks of the upper CML (black and grey lines) experience an anomalous attenuation. Vertical axis is inverted.

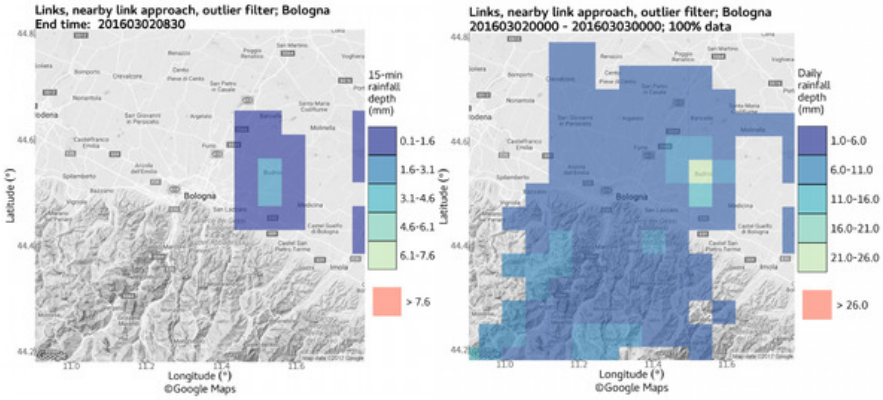

Figure 5.6: 15 min and daily interpolated rain depth over Budrio showing the signature of the anomalous attenuation.

lation with the signal of the solar radiation, particularly in the first hours after sunrise (see Figure 5.7). Many hypotheses were formulated: dew deposits in mornings with clear-skies and low temperatures, anomalous propagation of the radio signal due to stratification and inversion of the vertical temperature profile, bending of the antenna's mount due to direct solar radiation heating and consequent misalignment of the beam, misalignment due to winds, heating of the electronic. But more thorough analyses were not tried and the faulty CML was removed from the pool of links.

Certainly, the problem had to do with something related to that specific link, since neighbours did not show any indication of similar behaviours. It probably was affected by some low misalignment tolerance or damaged radome or unstable structure that enhanced its sensitivity to noise and disturbances.

During the CELLENMON online meeting of October 14, 2020, hosted by the Tel Aviv University, we had the chance to verify that other groups found the same issue and

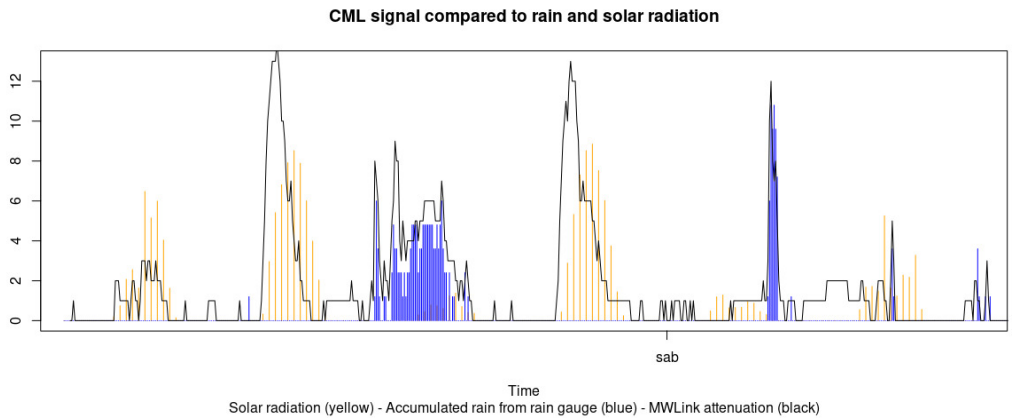

Figure 5.7: Anomalous attenuation (black line) repeatedly coupled to the solar radiation signal measured by a nearby station (yellow). Regular attenuation coupled with rainfall signal from a rain gauge of the same station (blue).

proposed similar explanations, but no one came to a definitive conclusion.

This signal was not filtered because of the very low link density around it. In fact, it showed up only in the first two months of data (March and April 2016, the ones when the active valid links were only 50% of total). Once back at normal coverage levels, the algorithm probably manages to correctly filter the anomalous signal, since no other occurrence are registered in May and June.

Cloud melting layer

We report here a particular result of the preliminary validation conducted with Arpae-SIMC. This validation appeared in our poster presentation at the 10th ERAD[3] in Ede-Wageningen (The Netherlands, July 2018).

On March 2, 2016 a huge overestimation occurs for the CML product over Appennines near Bologna. The hourly precipitation analysis shows the passage of a meteorological system heading west from east. Precipitation is detected and recorded in all products (unadjusted radar, gauge adjusted radar, ERG5 interpolation of the rain gauges and CMLs).

Figure 5.8 shows all four products for two representative hours. It is evident that CMLs (lower right) strongly overestimate the precipitation in the Apennines area.

Local hourly maximum values reach 40 mm for CMLs whereas the other estimates are below 5 mm (see Figure 5.9). In this area links are normally placed on top of hills

[3]European Conference on Radar in Meteorology and Hydrology

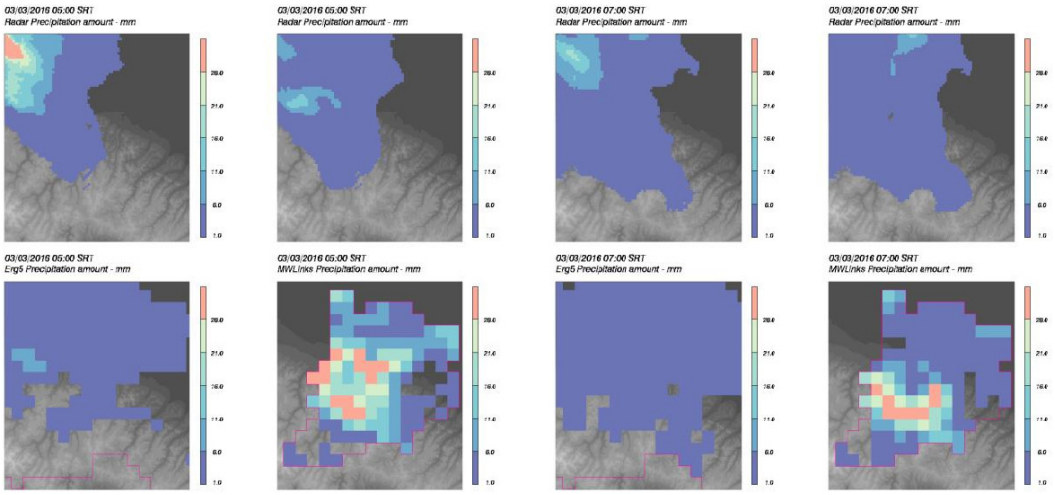

Figure 5.8: Quantitative cumulated precipitation estimates at 05.00 UTC (Left) and 07.00 UTC (Right) on 10/03/2016. Each group is composed by four panels : (Top Left) Radar, (Top Right) Radar adjusted; (Bottom Left) ERG5 ; (Bottom Right) CMLs, here indicated as MWL for MicroWave Links.

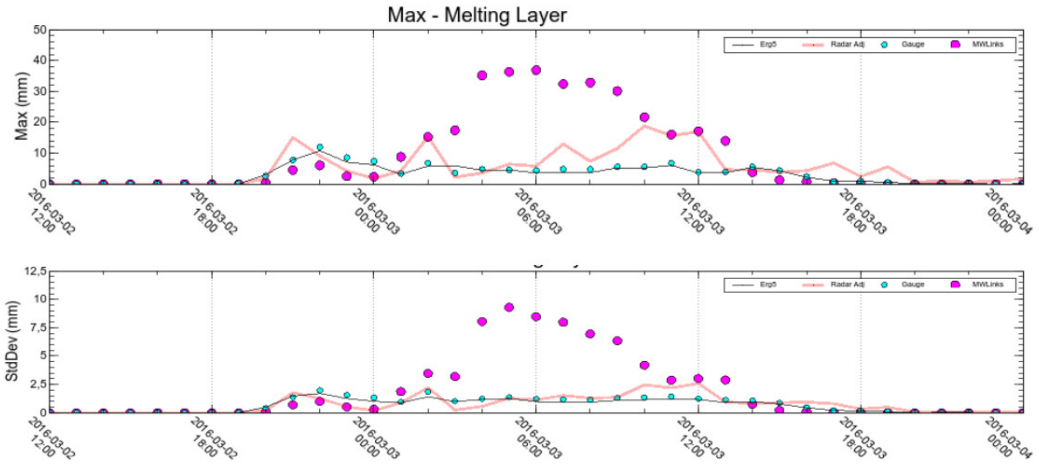

Figure 5.9: Time series of hourly maxima (Top Panel) and Standard Deviation (Bottom Panel) from 02/03/2016 12:00 UTC to 04/03/2016 00 UTC for the four products. CMLs are the magenta dots. Credits: P.P. Alberoni

and mountains; in this way their path is free of any obstacle. This means that they even pass several hundred meters above the ground when crossing the valleys. CMLs therefore could detect hydrometeors which are different (in size composition or in phase) with respect to the ones collected by gauges on the ground. The radar will sample even higher atmospheric volume, in order not to be contaminated by ground clutter and echo.

Figure 5.10 shows instantaneous temperature and hourly precipitation for two weather

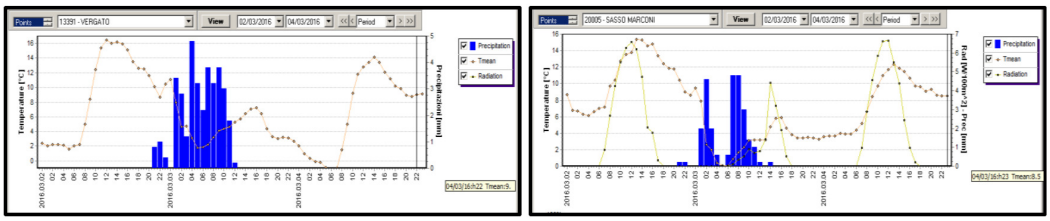

Figure 5.10: Time series of temperature (line and dots) and hourly precipitation (blue bars) for the stations of Sasso Marconi (Left) and Vergato (Right), from 02/03/2016 00:00 UTC to 05/03/2016 00 UTC. Credits: P.P. Alberoni

stations in the Apennines area (Sasso Marconi and Vergato). It can be seen that the temperature close to the ground falls around 0°C during the precipitation event. We can infer that there was rain at the ground, there was mixed-phase precipitation along microwave link path (melting snowflakes) and there was snow at higher levels, where radar samples are taken.

Ice particles have large dimensions with respect to raindrops, especially if in the form of snowflakes and aggregates. Yet they have a weak extinction cross section compared to liquid water. When the snowflakes start to melt, however, water surrounds the ice particle, which now is still almost the original size but comparable to a water drop from an electromagnetic point of view. Being the attenuation dependent on about the third power of the particle diameter (see Section 1.2), melting snowflakes act as strong microwave attenuators. Hence the CML overestimation. This effect is similar to the well known "bright-band" in radar meteorology.

5.5 First results from the complete Lepida network

The first rainfall event caught by the Lepida network extended to the whole region came at the middle of August 2018. The Arpae-SIMC analysis of the event reported intense phenomena from August 13th, starting east in the provinces of Parma (PR) and Piacenza (PC) and then moving west until the adriatic sea, with very bad weather hitting the city of Ravenna (RA) and the coast on August 15th.

The CML network caught some of the thunderstorm cells, when they happen to intercept the link paths. Looking at subsequent interpolated maps, the westward advection is clearly present, even though the system moved mainly over the plains, where the Lepida is very low where not absent. And then, also quantitative estimates over the city of Ravenna between 05.30 and 07.00 UTC of August 15, 2018 were remarkably consistent

with raingauge measurements (see Table 5.2 and Figure 5.11). At the same time it became clear that the network was not able to reproduce the rainfall fields in most of the areas due to very low coverage. Localised phenomena fell completely undetected between the sparse links.

Table 5.2: 15 min-cumulated precipitation over Marina di Ravenna (RA) during the morning of August 15, 2018: rain gauges measurements and CML estimates.

UTC time	Rain gauge (mm)	CML (mm)
05.30	12.7	14
06.00	4.2	5
06.30	10.1	16
07.00	24.8	24

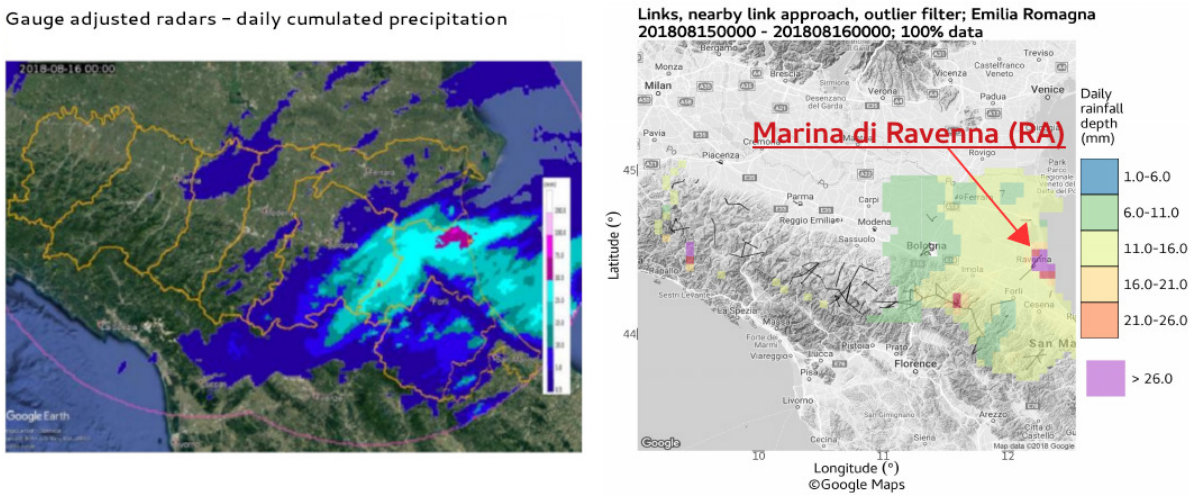

Figure 5.11: Daily cumulated rainfall for gauge adjusted radars (left) and interpolated CMLs (right), August 15th 2018.

The Lepida network is uniquely complex and hard to exploit, but potentially most valuable. The location of Lepida's CMLs gives the opportunity to monitor areas of major hydrological interest. In fact, the Apennines ridge is where most of the precipitation usually falls and where main river's basins have their most high and steep parts (see Section 2.1). But it poses critical challenges to the opportunistic rain retrieval. Having long links over a complex orography means that the height of the link above the ground is constantly changing and the probability to find variable meteorological conditions along the path gets higher.

The low link density not only increases the chances of missing small scale precipitating structures entirely, but also cause trouble to the classification algorithm (Wet-Dry) which

in RAINLINK's case relies on spatial autocorrelation of the rain field (NLA approach, see Section 4.1).

6 Validating the products

This is a quantitative validation of the CML estimates limited only to the months of May and June 2016 for the sole Vodafone network over the administrative provinces of Bologna and Parma and the two river basins of Reno and Parma inside them. These results appeared in Roversi et al. (2020).

6.1 Data characteristics

We have considered 57 days from 5 May to 30 June 2016 and two target areas: the provinces of Bologna (BO, $3702\,\mathrm{km}^2$) and Parma (PR, $3447\,\mathrm{km}^2$). Both are located in the Po Valley, Emilia-Romagna, northern Italy (coloured areas in Figure 3.7). The two river basins are indicated as thick lines in Figure 3.7.

Target areas are divided into square pixels of $5\,\mathrm{km}\times5\,\mathrm{km}$ (see also Section 3.3) and this grid will be used to carry out rainfall interpolation and products intercomparison.

The validation has been carried out comparing, at different spatial and temporal scales, the rain amount obtained by CMLs, through the RAINLINK algorithm (Overeem et al., 2016a) with all other rainfall estimates operationally available over the target domain. In particular, CML product has been compared with radar rain rates, both raw and gauge-adjusted, rain gauges measurements and the operational precipitation analysis (ERG5) made available by Arpae-SIMC.

6.1.1 CMLs

Microwave attenuation data and metadata were purchased from Vodafone Italia S.p.A. within the Life EU project RainBO LIFE15 CCA/IT/000035 (Alberoni et al., 2018).

All 357 CMLs are "duplex" links, which means that two sub-links (back and forth) are present for the same link (although they could also be not simultaneously active). Signal polarisation is vertical for 259 CMLs, horizontal for the remaining 98, while carrier-signal frequencies span from 6 to $42.6\,\mathrm{GHz}$, with an average frequency ($\overline{f}$) of $22.1\,\mathrm{GHz}$.

Path lengths of the links vary from 162 m to 30 km, the interquartile range extends between 2.4 and 8 km, and the average length is 6 km.

The whole set of available CMLs is represented in Fig. 6.1a, according to the pairs of path length (x-axis) and frequency (y-axis). Theoretical sensitivity is calculated for every length-frequency pair through the inversion of the kR relationship at a fixed $1\,\mathrm{mm\,h^{-1}}$ rainrate. It has to be remembered here that the manipulations within the algorithm (especially the *Aa* threshold) do not allow a direct translation from the theoretical sensitivities to actual instrumental uncertainties or error bands. The theoretical sensitivity field is shown as contour lines of equal sensitivity with small differences for the two polarisations.

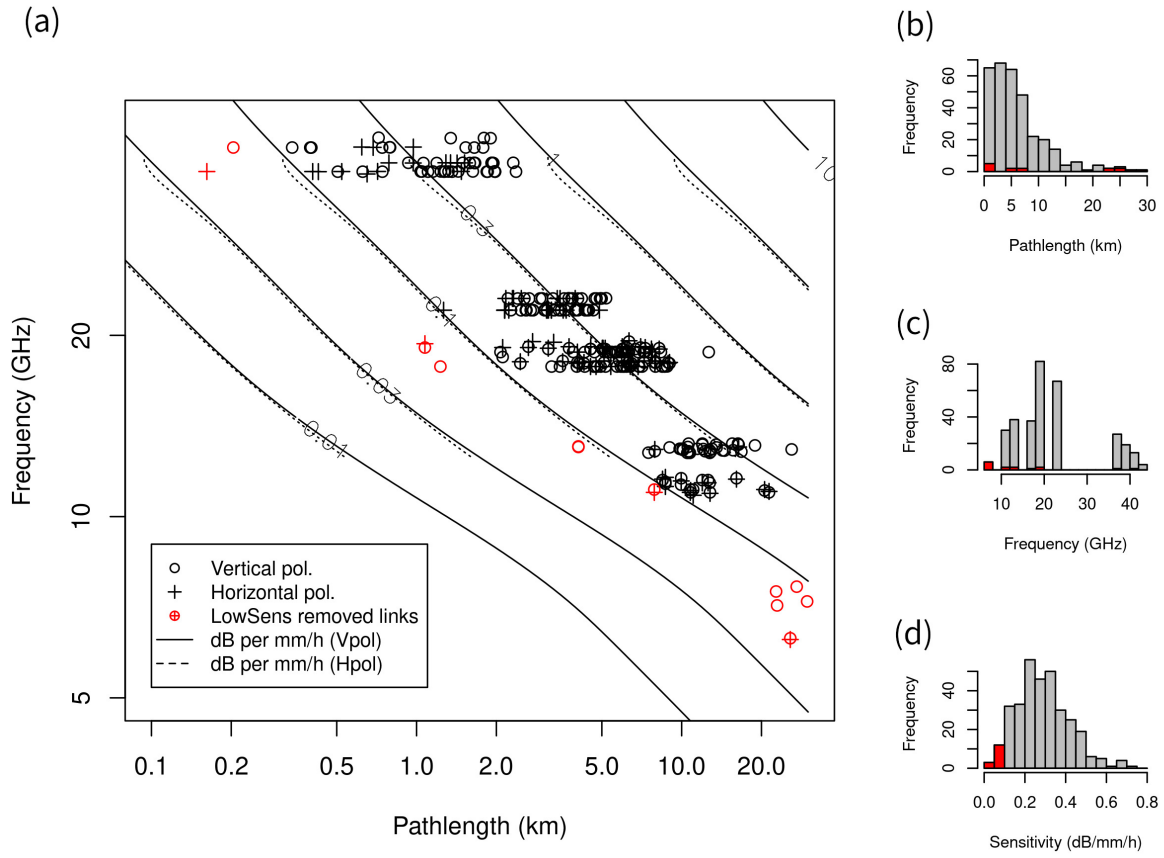

Figure 6.1: Dataset characteristics over the field of the theoretical sensitivity. Data points in red are below the $0.1\,\mathrm{dB\,mm^{-1}\,h}$

As expected, the carrier frequency is anti-correlated with path length since high frequencies, while allowing a wider transmission band, are more prone to attenuation compared to the lower frequencies (Leijnse et al., 2008).

6.1.2 Reference fields

Hourly and 15 min rain gauge data are provided by Arpae RIRER (see Section 3.2).

Among all the variables included the in ERG5 analysis product, we consider here only the hourly accumulated precipitation. Its input is based on the same rain gauges of the RIRER network, onlyin this case no longer limited to the two target areas, but rather extended to the whole Region. Some discrepancies are therefore to be expected between the two products, mainly near the borders and in areas where the distribution of the instruments is less uniform.

The radar dataset is based on hourly precipitation estimates obtained from the composite of the regional radar network managed by Arpae-SIMC (see Section 3.4).

6.2 Coverage

The number of working CMLs varies slightly over the months: it grows from 348 at the beginning of May to the maximum of 357 in June. The number of valid CMLs for rain retrieval is lower because of the quality checks and sensitivity filtering performed by the pre-processor of the algorithm (see Section 4.1, point 1), resulting in a median number of 308 valid CMLs with very small fluctuations. Most of the rejected data is empty or incomplete (P_{min} or P_{max} missing), probably due to failures in reading or storing raw data. More details on the rejected data are presented in the following Section 6.3.

Table 6.1: CML datasets comparison.

Variable	Unit	ER	NL	Overeem et al. (2013)	Overeem et al. (2016b)
Total area	km^2	7149	35500	35500	35500
CMLs	counts	308	1527	1514	2044
sub-links	counts	606	2473	2902	3383
LD	km^2	0.043	0.043	0.043	0.058
LL	km	5.8	2.9	3.1	3.6
BC	$\mathrm{km\,km}^{-2}$	0.25	0.13	0.13	0.21
$\overline{f}$	GHz	22.1	37.11	37-40	37-40

We introduce four parameters to summarize the topological structure of the CML network: the link density LD (defined as the total number of link paths divided by the whole area, in km^{-2}), the average link length LL (in km), the bulk link coverage BC (defined as the sum of the link path lengths divided by the total area, in $\mathrm{km\,km}^{-2}$) and

the local link coverage LC (calculated as BC but for each gridbox, in $\mathrm{km\,km^{-2}}$). Due to Vodafone confidentiality restrictions, we are not allowed to show the exact location of the available links, so we show, instead, in Figure 3.7 the spatial distribution of LC.

Since the RAINLINK original settings depend on the network characteristics, we compared the Emilia-Romagna network (ER) with the one from The Netherlands (NL), which is included in the RAINLINK software package as test sample (Overeem et al., 2016a), and with other datasets on which the algorithm was employed (Overeem et al., 2013, 2016b). The datasets properties are summarised and compared in Table 6.1. ER has comparable link density and higher average link length, resulting in a higher bulk coverage with respect to the NL network. The province of Bologna hosts more than half of the links (195 against 113) and thus has a higher LD.

6.3 Data quality, filters and outliers

Starting from an initial dataset of 357 duplex CMLs, 4.13 % of the data is rejected as it has a theoretical sensitivity lower than $0.1\,\mathrm{dB\,mm^{-1}\,h}$ (red selection in Fig. 6.1). This involves 15 CMLs (see Fig. 6.2a). Another 10.9 % is lost during the Preprocessing, mainly due to missing values (`NA` in the R language) alternatively on Pmin or Pmax. Some mitigation technique was attempted, exploiting the remaining observation of the Pmin-Pmax pair, but did not lead to any consistent result. The median of the number of CMLs involved here is 37, with around 9 % variance between different days and an isolated peak on June 3rd (see Fig. 6.2b). Both low sensitivity and NAs are issues intrinsic to the dataset we received, and their mitigation is out of our control, so we had no alternatives but to discard them all. Rejected percentages of the total available data are listed in Tab. 6.2.

Table 6.2: Percentages of the data rejected after the different steps of the algorithm.

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

Lastly, 3.87 % of the data is classified as "outlier" from the outlier filter of the RAINLINK algorithm. The resulting valid dataset is 81.1 % of the initially available one. Outliers affect a median of 14 CML per time interval ($15\,\mathrm{min}$), but in this case, the corresponding variance between different days is a higher 27 % (see Fig. 6.2c and 6.2d). Outliers happen mostly on high-frequency links ($> 35\,\mathrm{GHz}$) and on the shortest lengths

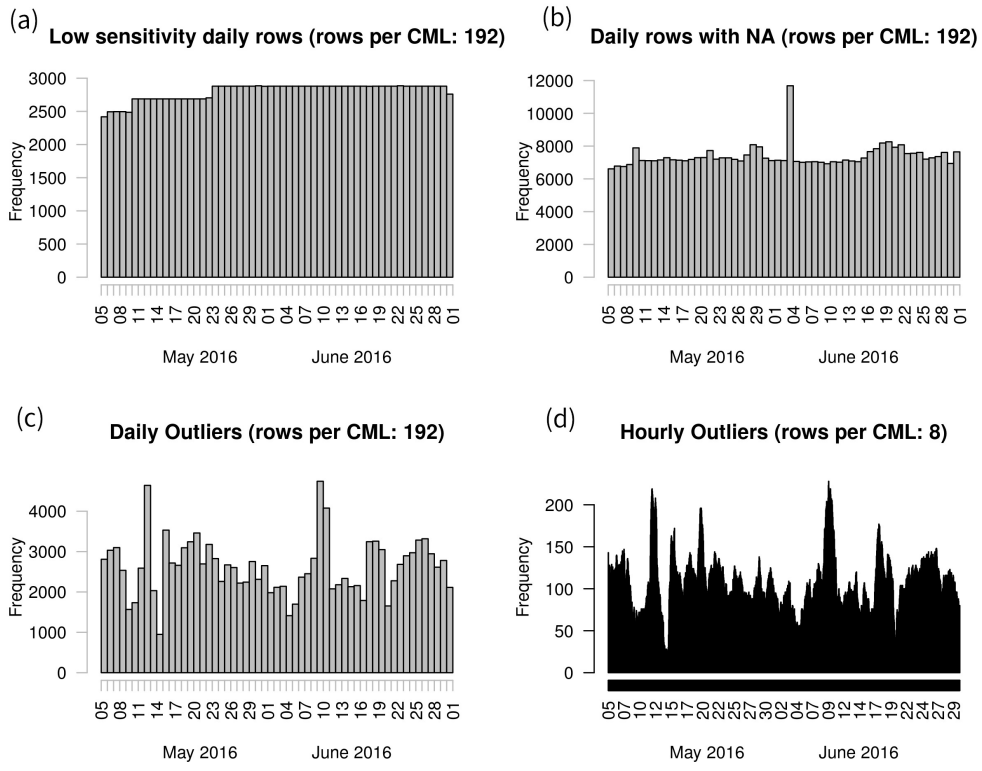

Figure 6.2: (a) Number of low sensitivity data rows. (b) Number of rows with NAs. (c) Daily numbers of CMLs with outliers. (d) Hourly numbers of CMLs with outliers. Frequency has to be divided by the number of rows per CML.

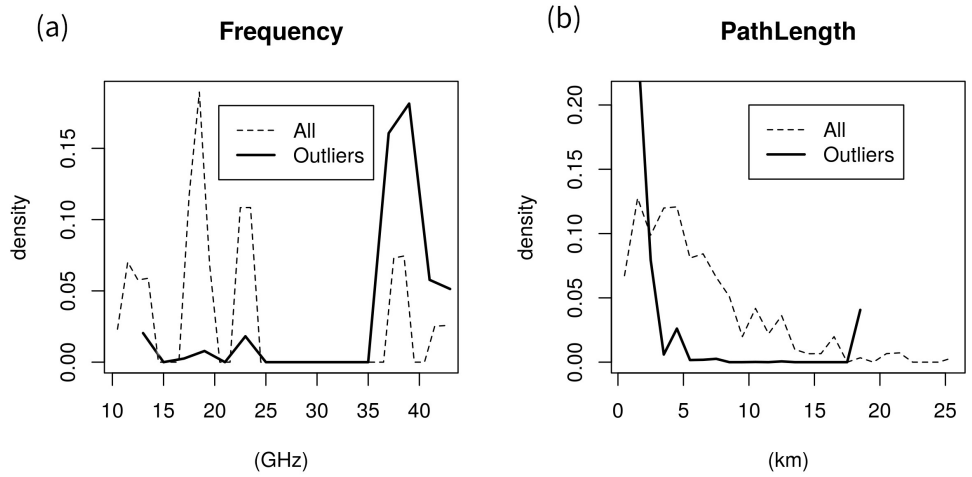

Figure 6.3: Distributions of the hardware characteristics of the CML reported as outliers compared with the ones of the whole dataset.

($< 4\,\mathrm{km}$) with a secondary peak on the longest ($> 17\,\mathrm{km}$), as shown in Fig. 6.3. No calibration of the $-32.5\,\mathrm{dB\,km^{-1}\,h}$ threshold was attempted. Please note that LC, BC, and the other dataset descriptors of Table 6.1 are all computed for the Valid dataset, not for the Total one.

6.4 Results

We carried out the validation of the CML product at three different levels. First, we compared single link estimates with the measurements of nearby rain gauges. This was made at the shortest temporal scale available (which is 15 minutes). Success and failure cases are discussed, explanations for the latter are advanced. Secondly, we compared the interpolated $5\,\mathrm{km} \times 5\,\mathrm{km}$ CML hourly rainfall maps versus the ERG5 product at grid box scale, also analysing three case studies. In the third step, the map comparison is carried out at a basin scale in a multisensor perspective, i.e. including also all the other precipitation products available at Arpae-SIMC.

6.4.1 Single link verification

We have manually selected links coming from a set of different terrains and looking for an active rain gaugein their proximity. The distance between CML and respective rain gauge is reported between round brackets for every link in Figure 6.4. It is always below $3\,\mathrm{km}$, which is significantly lower than the correlation distance of precipitation in Italy (Puca et al., 2014) and it is also always lower than the length of the link itself. In general, no dependence of the link performance on the distance from the rain gauge is found.

Selected links had to be active for all the analysed period. In many cases more than one link was selected for one rain gauge. Temporal sampling is kept at the highest frequency, which is a measurement every 15 minutes for both the CML and the rain gauges. 12 rain gauges and 26 CMLs have been chosen, 14 of which are in the northern part of the domain and the other 12 on the hilly region at elevations between $193\,\mathrm{m}$ and $960\,\mathrm{m}$ a.s.l..

The rain depths of the 26 CMLs are reported in Figure 6.4 for the whole study period, grouped accordingly to the closest rain gauge and ranked by its altitude. A large variability is found (ranging from near-perfect agreement to discrepancy of a factor of 2 or 3 in the worst cases). 75% of the 26 links' CCs are between 0.5 and 0.88, with overall median value 0.68, proving an acceptable overall skill. We relate this variability to the heterogeneity of CML sensitivity, the small scale of the meteorological events and different site exposure and elevation. In most cases, CMLs underestimate the rain gauge values: the links located in the lowlands (Figure 6.4a, 6.4b, 6.4d and 6.4e) show a better correspondence than those in the hilly regions, where underestimation is more significant.

In some cases (Figure 6.4f, 6.4k and 6.4l) the discrepancies between CMLs close to the same rain gauge (but different in location, frequency and length) are much lower than the CML-rain gauge differences: all these CMLs are in good mutual agreement and share the same classification issues, resulting in a systematic underestimation which therefore seems to be caused by the algorithm setup. In other cases (Figure 6.4b, 6.4d and 6.4g) some links clearly outperform other members of the same group. This second kind of discrepancies is more likely related to real differences, like inhomogeneous rainy structures which crossed the link paths or different hardware setups, while there is no evidence of a correlation with frequency or path length. The difference between the two directions of the same link is generally below 10%, except for the Ostia Parmense site (see Figure 6.4g).

To gain a deeper understanding of better and worse performance of the single links, we performed a more detailed analysis of case studies at the rain-event scale (Figure 6.5). We show a case when the link retrievals accurately match the measurements of the close-by rain gauge, and a case with markedly low performance. In Figure 6.5, graph panels are organised in columns by CML and in rows by sub-link. In the top panels are shown all the signals managed by the algorithm: the reference power P_{ref}, the raw received powers P_{min} and P_{max} and the filtered received powers P_{min}^{Cor} and P_{max}^{Cor}. In the middle panel rain gauge measurements are compared with CML estimates and also the minimum and maximum attenuation signals are plotted (A_{max} and A_{min} respectively). The grey background indicates when the classification detects a dry period. The pink background indicates the band inside which attenuation is considered as caused by a wet antenna (Aa parameter) and is discarded for rain retrieval. The bottom panels show the cumulated rainfall depths in the same time frame.

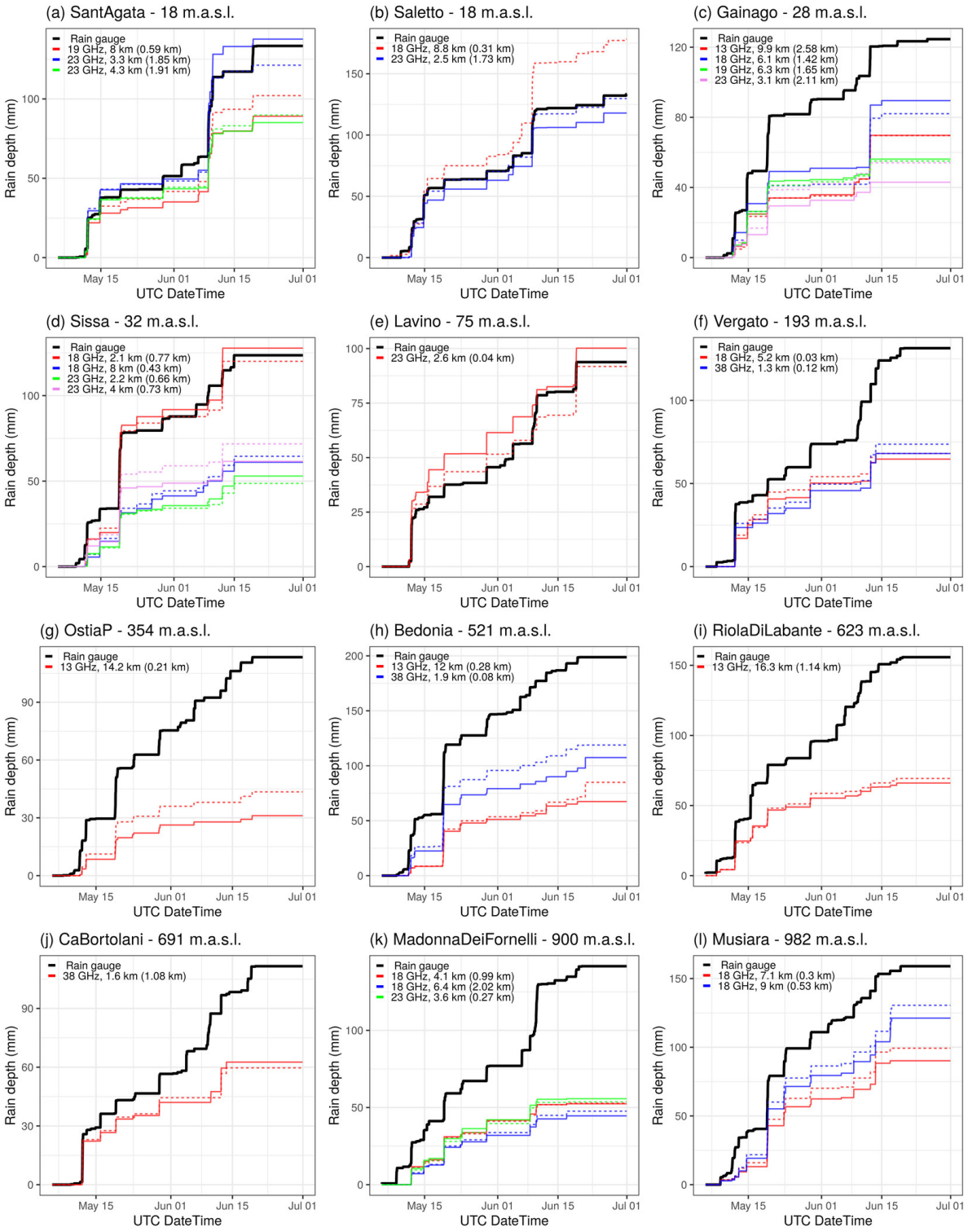

Figure 6.4: Accumulated rain depths over the entire period for the 26 CMLs selected for the single-link analysis. Each tile is named by the corresponding rain gauge, whose accumulated rain depth is shown by the black thick line. Solid and dashed lines represent the two directions (if both active) for every CML (distinguished by colour). CMLs are identified in the legend by carrier frequency, link length and (between brackets) distance from the rain gauge.

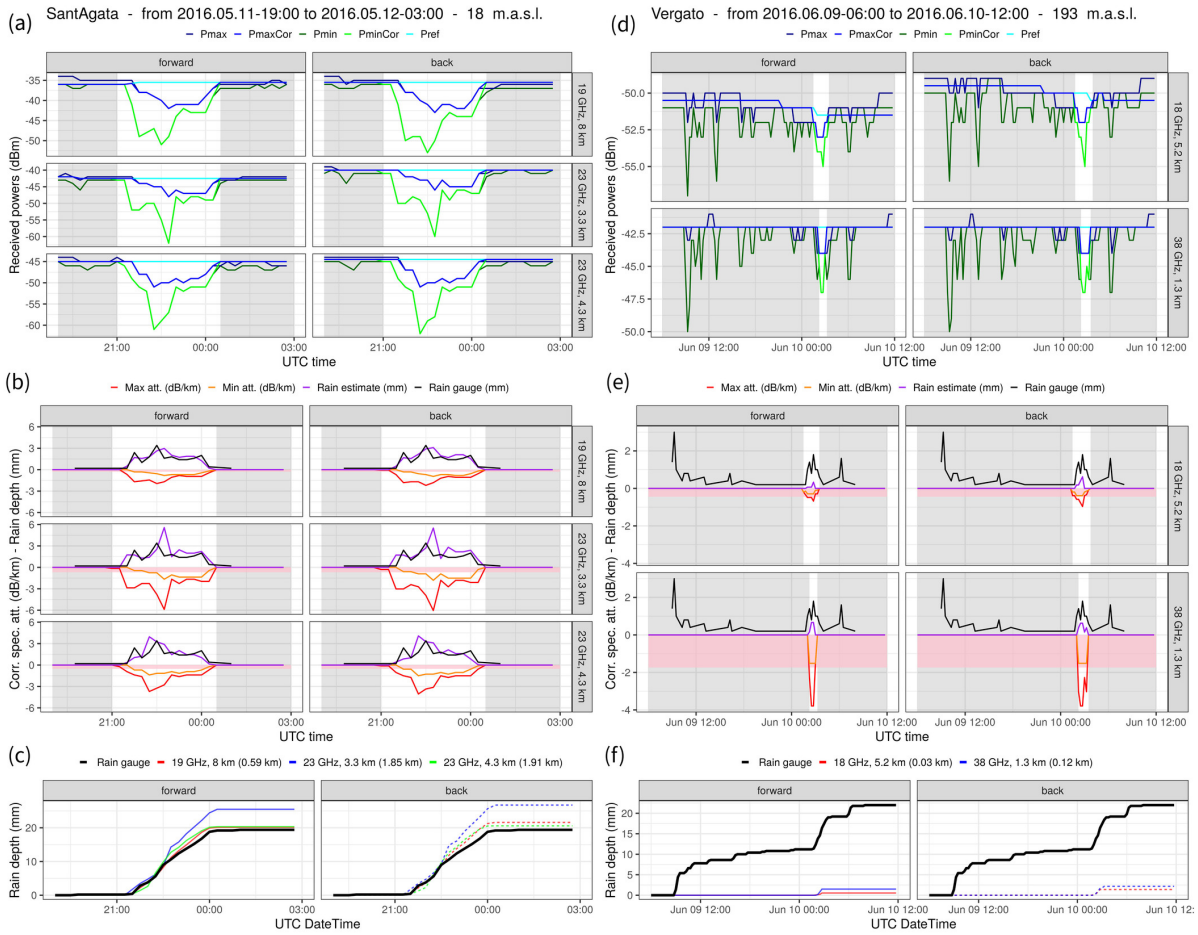

Figure 6.5: Single link analysis for Sant'Agata (from 11.05-19:00UTC to 12.05-03:00UTC) and Vergato (from 09.06-06:00UTC to 10.06-12:00UTC): (a) and (d) show the received signals (P_{max}, blue; P^{Cor}_{max}, light blue; P_{min}, green; P^{Cor}_{min}, light green; P_{ref}, cyan); (b) and (e) show maximum attenuations (red), minimum attenuation (orange), estimated rainrate (purple), and gauge measurements (black); in (c) and (f) the cumulated raingauge rainrate (black) is plotted with the link estimates. Grey vertical bands correspond to intervals labelled as dry by the NLA classification, pink horizontal bands correspond to the threshold in $\mathrm{dB\,km^{-1}}$ of the Wet Antenna correction of $2.3\,\mathrm{dB}$. Y-axes ranges are specific for each CML as received powers differ between different path lengths.

Best cases example

Between 11 and 12 May 2016 an extensive convective system covered the Bologna Province area almost entirely, with a maximum rainrate of $23\,\mathrm{mm\,h^{-1}}$, and widespread precipitation around. For this case, the NLA classification on the three links near

Sant'Agata (Bologna Province, 18 m a.s.l.) works properly: in Figure 6.5b most of the measured rain is on white background. In Figure 6.5a, after the attenuation event, the noisy signal is correctly filtered, and a very small amount of rain (just above the gauge threshold) is neglected. The agreement is qualitatively very high between each pair of sub-links and good among the different links, in terms of specific attenuations and retrieved quantities (see Figure 6.5b). Quantitative retrievals give some overestimation for one of the CMLs, whose effect is evident on the accumulation plot (Figure 6.5c) where the total rain depths are compared. During the two months the Sant'Agata links are generally in good agreement with the close-by rain gauge, with CC ranging between 0.66 and 0.88 and CV between 0.47 and 0.96.

Worst cases example

Between 8 and 10 June 2016 an event hit the Vergato site (Bologna Province, 193 m a.s.l.). It was characterised by intense rainfall peaks (rainrate up to 14.6 mm h^{-1}) and iterated moderate scattered precipitation. Many wet intervals are missed due to wet-dry misclassification (Figure 6.5e), leading to a 20 mm loss in the rain accumulation (Figure 6.5f). The POD over the entire period for these two links is between 0.22 and 0.29.

In the case when the NLA classification correctly identifies some rain occurrence, there is still a general quantitative underestimation. It could be seen that half of the signal is hidden from the wet antenna attenuation threshold. The continuous scores for the wet-wet sample on the entire period show a good correlation with gauges but are poor in statistical relevance because of the high number of misses. They nevertheless confirm the tendency to underestimate, by around 40% (ME$=-0.40$).

6.4.2 Gridded product verification

The verification of the RAINLINK gridded product (1 h cumulated on the 5 km$\times$ 5 km grid) with respect to the ERG5 product is first performed at the highest available resolution (grid box by grid box), since the two products intentionally share the same interpolation grid. Secondly, the comparison is carried out at the basin scale by matching spatially averaged time series over areas of different size, in parallel with other operational precipitation products available at Arpae-SIMC.

Overall PDFs

The PDF of rainrates in the region during the two months of our research are reported in Fig. 6.6 (dashed lines) for the single 15 minutes raingauges (left) and the hourly ERG5 product at $5\times5\,\mathrm{km}^2$ resolution (right). In the same plots, are also reported the PDF of rainrates as computed by RAINLINK (solid line), 15 min, single link (left), and hourly interpolated on the same ERG5 $5\times5\,\mathrm{km}^2$ grid (right).

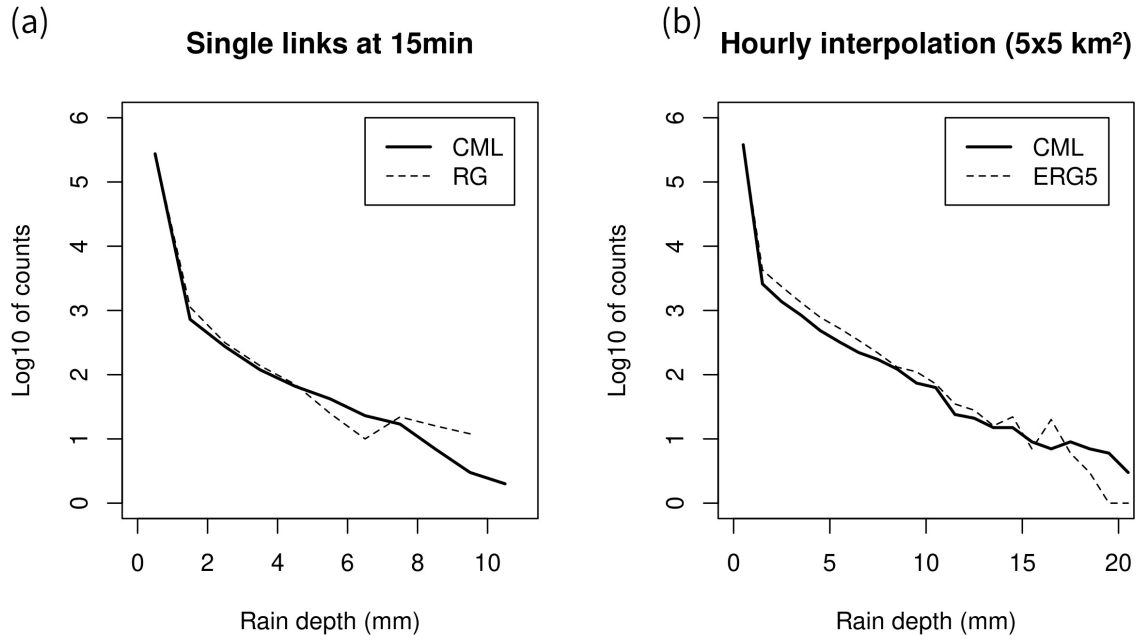

Figure 6.6: Comparison of estimated and gauged PDF at 15 min single link (left) and hourly interpolated scales (right). Sampling is performed on 1 mm bins.

The plots show that single-CML estimates are rather good for low to moderate rainrate (around 5 mm), while for higher, more discrepancies are found. Moving to the interpolated version, we can note that interpolation leads to a general underestimation widespread along with the whole rainrate range.

Highest resolution matching

Figure 6.7 shows a scatter density plot for the whole dataset over the entire period. CML estimates from RAINLINK in northern Italy over uneven ground have an overall underestimating performance of -26 % on the accumulated rain over the two months. The CV is 0.78 and R^2 (the square of the Pearson's correlation coefficient CC) is 0.46, based on a sample of 10672 total wet hours. To make the comparison with past works

easier, we computed continuous indicators with the filter set as Reference >0.1 mm and with no filtering at all. Results with the first setting yield worse indicators, increasing the ME to -0.41 and the CV to 0.95, with a second digit increase for R^2, around 0.5. The no-filter run shows values of $ME = -0.33$ and $R^2 = 0.53$ which are aligned with our most filtered results, while $CV = 4.6$ is greatly affected by very small rainrates. These results are in agreement with similar studies (Overeem et al., 2013, 2016b) despite the differences in the products involved: comparison between our results, with both filters, and the ones presented in the mentioned works are shown in Table 6.3.

Table 6.3: Comparison with previous studies: *Ref* is the reference, *Pr* is the product. The filter, when active, is always an high-pass filter over the 0.1 mm threshold.

	ER (this work)			Overeem et al. (2013)	Overeem et al. (2016b)
Time span	2 months			3 months	2.5 years
Time scale	60 min			15 min	60 min
Pixel area	25 km^2			81 km^2	74 km^2
Reference	intp. gauges			Gauge-adj. radar	Gauge-adj. radar
Active filter	Ref AND Pr	Ref	none	Ref OR Pr	Ref
ME	-0.26	-0.41	-0.33	0.02	-0.16
CV	0.77	0.95	4.6	1.13	0.64
R^2	0.47	0.50	0.53	0.49	0.49

The performance of the rain detection capabilities with respect to the 0.1 mm threshold is evaluated by the set of categorical scores defined in Section 4.2. Quantitative continuous indicators from now on are computed only for the grid boxes where both CML and ERG5 reported more than 0.1 mm at the same time. Categorical and continuous indicators are evaluated for five areas, with a different extension (S) and average Link Coverage ($\overline{LC}$). They are reported in Table 6.4, ranked according to the $\overline{LC}$ value: Parma Province (PP), Total Area (TA), Parma River Basin (PRB), Bologna Province (BP), Reno River Basin (RRB). The total area and the two provinces do not have any specific hydrological meaning, but could be seen as a good foretype of larger river basins with heterogeneous terrain (see Figure 3.7). All normalised indicators are relative to the average reference (ERG5) rain rate. Numbers in bold (italics) are the best (worst) value in the column.

We found ETS values ranging from 0.38 to 0.43, which are comparable with the ones obtained from satellite observations (Puca et al., 2014; Feidas et al., 2018) in similar regions. For four out of five areas (excluding RRB for now) the RAINLINK product

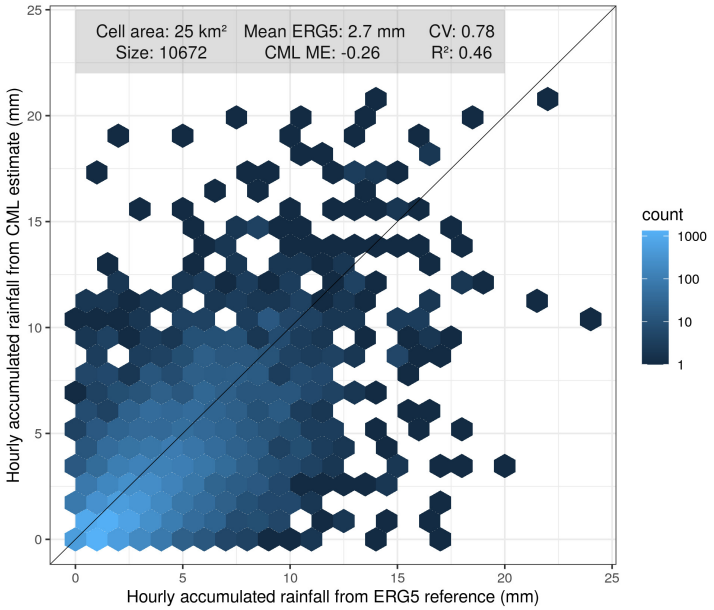

Figure 6.7: Hourly validation of link rainfall maps against ERG5 rainfall maps at grid box scale (highest resolution). Only the rainfall depths in which both CMLs and ERG5 measured $>0.1\,\mathrm{mm}$ were used. The black line is the $y = x$ line.

Table 6.4: Statistical indicators for each considered area, considering the highest resolution information (grid box scale), shown in ascending order of $\overline{LC}$. Continuous indicators are normalised and fractional. Values in bold (italics) are the best (worst) values in the column.

Area	$\overline{LC}$ ($\mathrm{km\,km^{-2}}$)	S ($\mathrm{km^2}$)	FAR	POD	ETS	MB	ME	MAE	CV	CC
PP	0.17	3447	0.28	0.51	0.41	0.71	*-0.34*	*0.55*	*0.80*	*0.62*
TA	0.18	7149	0.30	0.54	0.42	0.77	-0.26	0.52	0.77	0.68
PRB	0.19	624	0.30	0.48	0.38	0.69	-0.31	0.50	0.76	0.67
BP	0.19	3702	*0.32*	**0.57**	**0.43**	**0.83**	**-0.18**	0.48	0.73	0.74
RRB	0.29	828	**0.16**	*0.39*	*0.35*	*0.47*	-0.31	**0.45**	**0.62**	**0.80**

underestimates the rain occurrence (MB < 1), with a relatively low value of POD (0.48 to 0.57). The FAR is also rather small, (0.28 to 0.32), resulting in ETS values (0.38 to 0.43). Mean Error confirms the underestimation of rain amount (ME between -0.18 and -0.34), CV ranges between 0.73 and 0.80, CC between 0.62 and 0.74. For comparison, Petracca et al. (2018) analysed over Italy the instantaneous estimate of the Global Precipitation Measurement - Dual-frequency Precipitation Radar (GPM-DPR), considered as the most reliable and accurate instrument to measure precipitation from space.

Over a footprint of a size comparable to the one used in this paper, the best value of CC is 0.57, while the CV was between 1 and 2. Other validation studies of GPM-DPR products in the alpine region (Speirs et al., 2017) obtained relatively good POD (up to 0.78), FAR (below 0.08) and CC (up to 0.63) over flat terrain, with a dramatic drop of the skill indicators when areas with complex topography are considered.

The averages over the Reno River Basin stand out for all the indicators, either positively or negatively; therefore they need a separate description. As highlighted in Table 6.4 in bold and italics fonts, RRB has half the FAR the other samples have (0.16), almost ten points less CV (0.62) and nearly fifteen points better CC (0.8, which is unexpectedly high), with the mean errors aligned to the other samples. The higher accuracy in the estimates is reached at the expense of POD, ETS and MB: around 50% of the rainfall duration is lost in this area. The main peculiarity of the RRB area is the high $\overline{LC}$, which is 50% higher than the rest of the regions.

The marked improvement of continuous indicators for RRB suggests that the quantitative matching between estimated and reference could be positively related to $\overline{LC}$. Thus, we further investigate its effect on scores by grouping each grid box by LC quartiles, regardless of the actual geographical location, and reported the results in Figure 6.8. Five out of six indicators improve as LC increases (FAR, MAE, ETS, CC and CV), among which the most striking is the FAR, while POD remains mostly unchanged, allowing the ETS improvement.

Case studies

To assess the performance of RAINLINK with respect to the structure of rainfall fields we focused the analysis on three one-day-long events with different characteristics, for which RAINLINK provided results of varying quality.

The best performance was achieved on May 19 (see Figure 6.9, left), when an intense event was characterised by a few convective episodes on the Apennines, in the Parma Province. Precipitation peaks were around $90\,\mathrm{mm\,d^{-1}}$ (see Figure 6.9c), maximum and mean hourly rainrates were about 24 and $2.6\,\mathrm{mm\,h^{-1}}$, respectively (see Table 6.5). A large area of widespread moderate precipitation over the Bologna Province (Figure 6.9a) is also present. RAINLINK is able to localize precipitation local maxima (Fig. 6.9b), even if it occurred in areas where link coverage is relatively poor (see Figure 3.7), providing also accuracy in the peaks intensity. Estimated PDF matches closely the ERG5 curve, indicating that all rainrates are represented in the estimates (see Figure 6.9d).

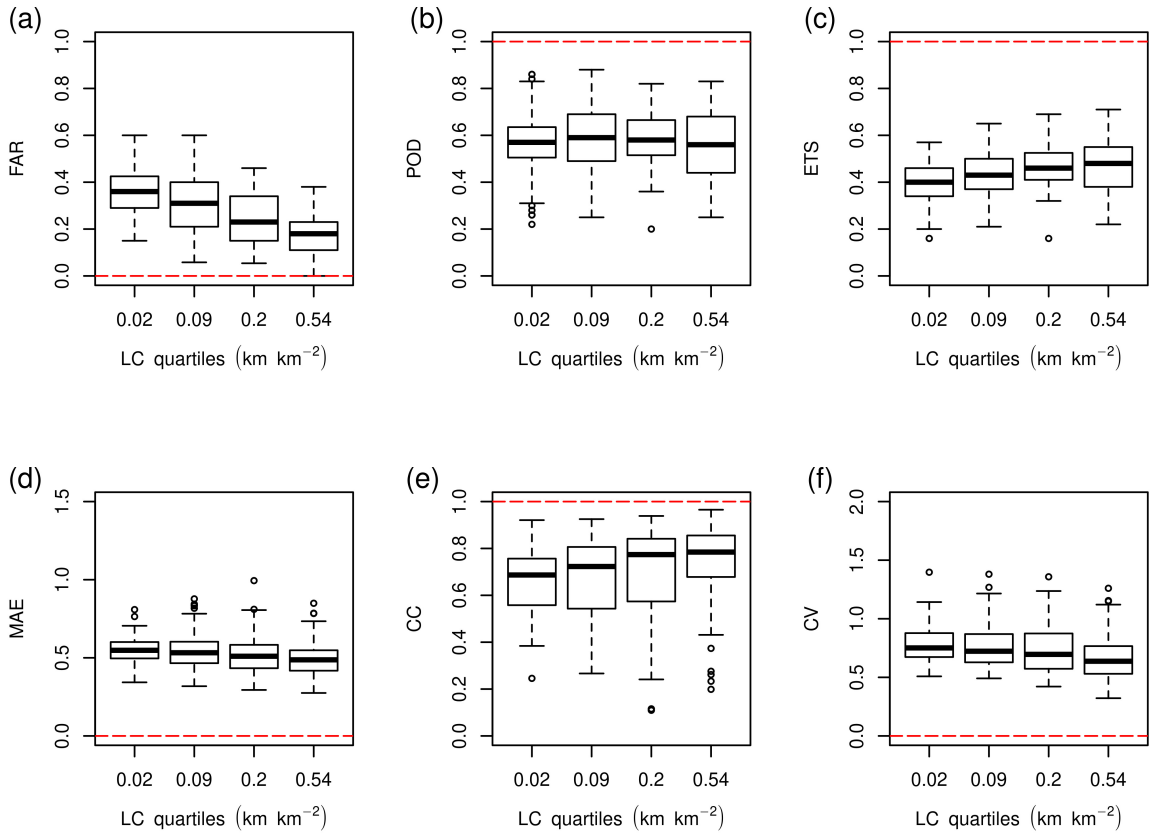

Figure 6.8: Distributions of four statistical indicators computed for every grid box and grouped in boxplots by quartiles of the link coverage LC (labelled by the quartiles centre). Red dashed lines are the optimal values for each score.

Though, underestimation is present at all ranges, more markedly at the highest rain-rates. Numerical indicators confirm the goodness of the estimate, in terms of wet area detection (ETS=0.59) and relative error (CV=0.69), while the fractional amount of rain lost by the estimate is low (ME=−0.29).

Table 6.5: Rainfall characteristics and performance indicators for the three one-day case studies

Date	mean R (mm)	max R (mm)	wet fraction	FAR	POD	ETS	ME	CV	CC
19.05	2.60	24.0	0.37	0.10	0.77	0.59	-0.29	0.69	0.78
11.05	2.50	21.0	0.35	0.10	0.66	0.49	-0.40	0.76	0.82
12.05	1.80	14.0	0.16	0.20	0.58	0.46	-0.65	1.10	0.46

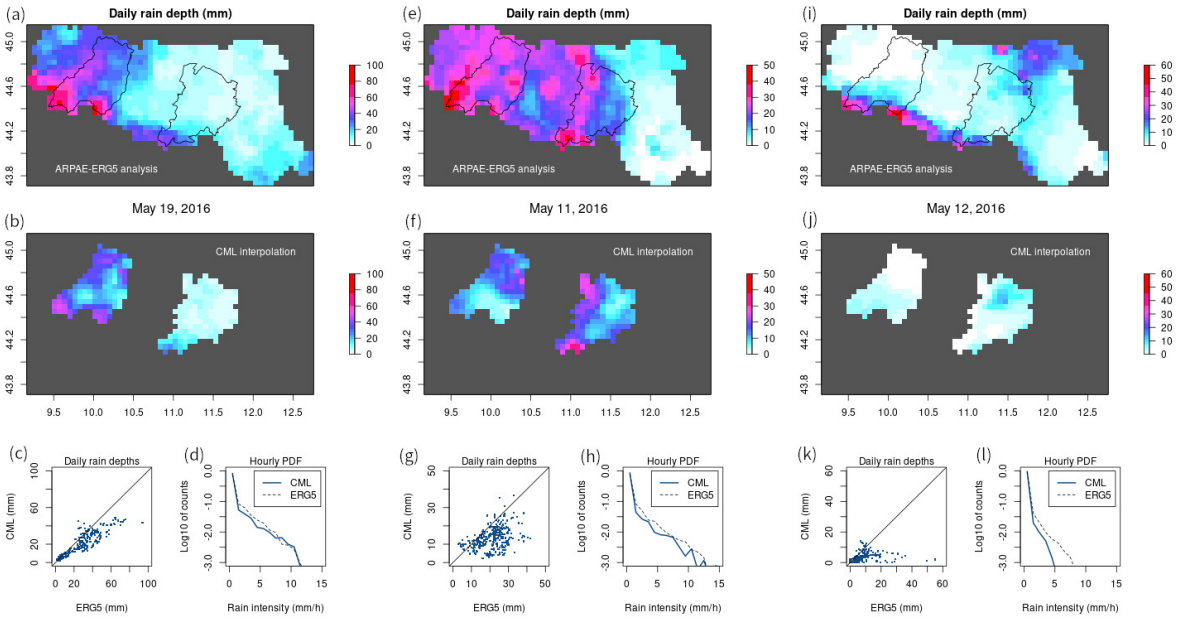

Figure 6.9: Analysis three one-day case studies (May 19, left; May 11, center; May 12, right): (a), (e) and (i) daily cumulated ERG5 precipitation; (b), (f) and (j) daily cumulated RAINLINK precipitation; (c), (g) and (k) scatterplot between the two daily precipitation; (d), (h) and (l) PDF of hourly rain rates.

The second case (11 May) shows a more patchy rainfall field (Figure 6.9e), which resulted from a series of storms that occurred in the area during the day. Maximum and mean rates are lower with respect to the first case (Figures 6.9g, 6.9h), as well as the wet fraction of overall samples (see Table 6.5). Some local peaks are correctly located (especially inside the Bologna Province), as shown in Figure 6.9f, and some others, in Parma Province and particularly on the Apennines, are missing. In this case the underestimation is marked for all rainrates, resulting in higher ME (-0.40) and lower POD (0.66).

A completely different scenario is represented by case three (May 12), when ERG5 measured light to moderate precipitation (see Figure 6.9i), with maxima on the Apennines, and a much lower fraction of wet samples. RAINLINK (Figure 6.9j) is not able to estimate the highest rainrates, neither to locate the area with the highest intensity. Moreover, it find a spurious peak in the northern area of the Bologna Province, which is not detected by ERG5. Here the fractional amount of rain loss is -65%, the POD is low, and an increase of FAR is also to be remarked, indicating that underestimation again dominates throughout the whole range of rainrates (see Figures 6.9k, 6.9l), but in

case of light rain, overestimation could also take place.

Areal averages matching

In this Section, the matching between estimate and reference field is performed at basin (and Province) scales, comparing hourly rain amounts averaged over areas of different sizes. The areas selected for this evaluation are the ones introduced in the previous Section: two of them are chosen because of direct hydrological interest (RRB and PRB), while the other three (BP, PP and TA) are selected to assess the impact of the increasing target area.

In Table 6.6 we present the categorical indicators calculated around the $0.1\,\mathrm{mm\,h^{-1}}$ threshold and the continuous indicators calculated on wet-wet occurrences only, for the five mentioned areas listed this time in order of increasing area size. In general, best performance is found for the largest areas (BP and TA), while the smallest ones (PRB and RRB) show the worst values. CML product underestimates precipitation occurrence (MB between 0.41 and 0.70) and amount (ME between -0.18 and -0.34) at all scales. Due to the areal averaging, CC is markedly higher than the high-resolution values reported in Table 6.4. The characteristic behaviour of RRB (lowest FAR and POD, highest CC) also remains in this case.

Table 6.6: Values of the statistical indicators for the mean rain amounts over each considered area, shown in ascending order of surface area S. Values in bold (italics) are the best (worst) values in the column.

Area	S ($\mathrm{km^2}$)	$\overline{LC}$ ($\mathrm{km\,km^{-2}}$)	FAR	POD	ETS	MB	ME	MAE	CV	CC
PRB	624	0.19	*0.18*	0.51	0.43	0.63	-0.34	*0.45*	*0.61*	*0.84*
RRB	828	0.29	**0.03**	*0.40*	*0.36*	*0.41*	*-0.34*	0.40	0.52	**0.93**
PP	3447	0.17	0.14	0.57	0.48	0.66	-0.34	0.48	0.56	0.98
BP	3702	0.19	0.14	0.60	0.51	**0.70**	**-0.18**	**0.33**	0.49	0.91
TA	7149	0.18	0.10	**0.64**	**0.55**	**0.70**	-0.26	0.35	**0.48**	0.91

The same areal-averaged statistical indicators have also been computed for all the operational products available at Arpae-SIMC for routine use and described in Section 6.1.2, reported to an hourly scale and compared with the ERG5 product. We show in Figure 6.10 the values of the statistical indicators as a function of the target area.

The rain gauge product, obtained by averaging the measurements of the rain gauges in the area, performs similarly to its interpolated version ERG5, as expected, and diverges

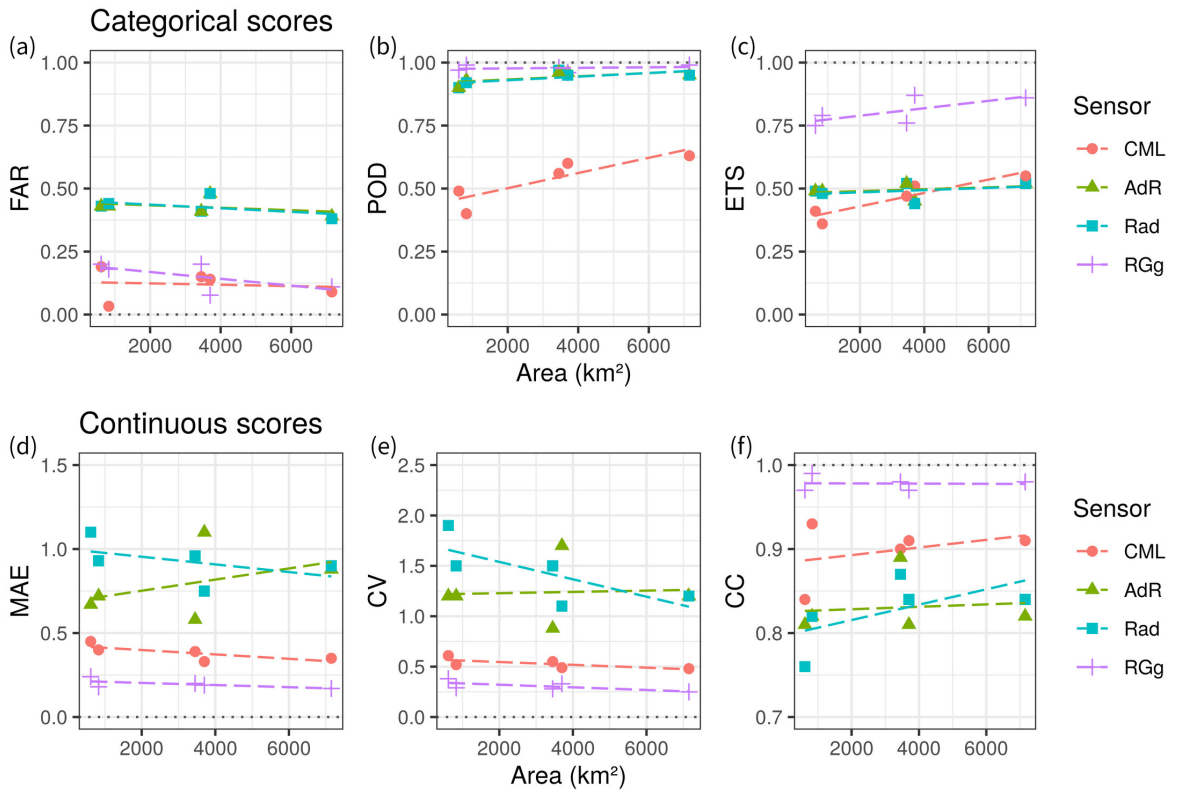

Figure 6.10: Scores of the areal-averaged rainfall amounts grouped per sensor and plotted against basin area. Linear fits are highlighted with dashed lines. The CML scores are also indicated numerically in Table 6.6.

only for small areas, where the impact of a single sensor in disagreement with neighbours is the highest.

Radar product shows, in this metric, almost the same performance both with and without the gauge adjustment[1]. Both have very good detection capabilities (POD is almost 1) but high rates of false alarms (FAR around 0.5) and marked quantitative discrepancies (MAE around 0.9, CV between 0.75 and 2).

The CML product outperforms both radar products in terms of CC, CV, MAE and FAR, while it lacks in detection capability (CMLs POD between 0.4 and 0.6). Figure 6.11 shows that the overestimating and underestimating behaviours, of radar and CML products respectively, can be seen as complementary. For radars, the spread is more relevant than for CML, but it has to be remarked that the latter has a smaller sample

[1]This is to be expected since the radar adjustment acts only at the rain gauges' locations and does not guarantee the consistency of the areal average of the entire rain field. Furthermore, the adjustment mainly affects rainfall rates higher than our threshold of 0.1 mm and has lower performance as spatial variance increases, e.g. in case of small scale convection.

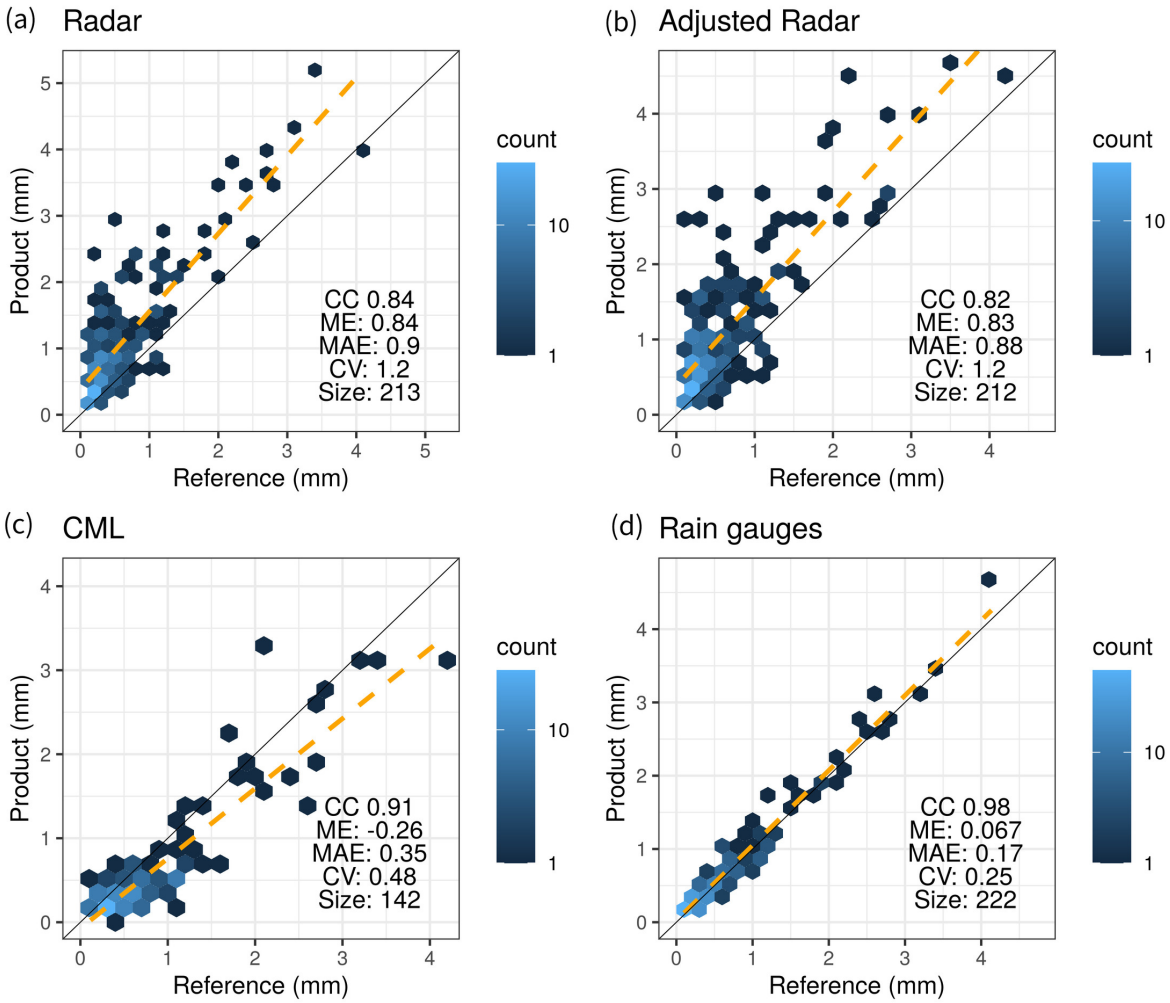

Figure 6.11: Comparison of hourly areal-averaged rainfall depths from the four products against the ERG5 reference. The total area (TA) wet-wet hours are considered.

size due to the already mentioned low POD issues. It has also to be said that part of the radar's high FAR and overestimation could represent real rain from small precipitating structures, often observed between meteorological spring and summer in Italy (see Supplement), that are randomly missed by the rain gauges (and therefore by the ERG5 reference product as well).

In Table 6.7 the latency and sampling characteristics of the four precipitation products we took for comparison are reported, along with CML product. CMLs operational specifications refer to an implementation of the RAINLINK algorithm as part of a real-time service (see Section 5.2.3), tested in 2019 by MEEO S.r.l. within the RainBO project (LIFE15 CCA/IT/000035).

Table 6.7: Latency and spatial and temporal sampling of the considered precipitation products.

Product	Reference time step (min)	Latency (min)	Spatial resolution (km)
CML	15	20	5
Radar raw	5	15	1
Radar adj.	60	60	1
Rain gauges raw	60	60	-
ERG5	60	1440	5

6.5 Discussion

The underestimating behaviour that emerged in the single-link-vs-gauge analysis (Section 6.4.1) seems to be largely imputable to a wrong wet-dry classification. Though we do not have a dataset large enough to support general statements, looking at Figure 6.5d we could gain some insights about what goes wrong in two of the most problematic CMLs of our population, the Vergato ones.

Most of the rain which is sensed by the gauge is falling in intervals that the NLA reports as "Dry" (grey background). P_{min} in fact clearly experiences some decrease, which is coupled to the missed rainfall, but P_{max} does not. This behaviour of P_{max} is not an issue in itself, as the NLA classification relies on P_{min} only. It indicates, however, that there are power fluctuations which happen faster than 15 min, otherwise P_{max} would have decreased too. Rapid fluctuations, in turn, suggest irregular, rapidly varying or scattered precipitation patterns. These are actually elements that could affect the correct classification, since NLA relies on the spatial correlation of the rainfall field in a range of 15 km (see Section 5.3). Therefore, a P_{max} signal which remains always near the baseline could be a precursor of local NLA issues. Classification errors are likely the best explanation for the low POD scores.

Given how we filtered the data (Product >0.1 and Reference >0.1), we need a source of error other than the misclassification as responsible for the quantitative underestimation measured by the set of continuous indicators (see Section 4.2 for reference). We saw that half of the signal in the correctly classified interval (Figure 6.5e, white background) remains under the wet antenna attenuation threshold (pink horizontal band). We can presume that the antenna is actually dry, so the Aa threshold in this case is reasonably too high (as also noted by de Vos et al. (2019)). However, a simple sensitivity test, carried out to assess the impact of a decrement in the Aa threshold on the single-link-vs-gauge

scores, did not lead to any substantial improvement, especially if the new value is used to process the whole dataset. More information is provided in the Supplementary Material.

Comparing the average performance of the interpolated product (Section 6.4.2) above the different sub-areas, particularly with respect to the RRB one, we can infer that higher mean coverage ($\overline{LC}$) leads to a more selective NLA classification, which reduced FAR and POD. When grouping the single grid boxes based on their coverage (see Figure 6.8), it seems however that the sensitivity to LC could explain only the improvement in FAR, but not the sharp decline in POD, suggesting that LC was probably not the only variable at play in the Reno basin. These results integrate the findings of Overeem et al. (2016b), that highlighted the positive impact of higher $\overline{LC}$ on CV and CC at lower spatial resolution. Other studies will be conducted in the future to gain more insights on these topics.

Looking at a daily scale (Section 6.4.2), the interpolated output of RAINLINK is undoubtedly able to resolve small size, short-living events, and even providing quantitatively accurate estimates. In case of widespread, moderate precipitation the overall rain pattern is still effectively represented, but some underestimation of the numerical values appears. When in presence of light and intermittent rainfall, instead, we see the consequences of the issues emerged during the single-link analysis. The rainfall maps in the panels i and j of Fig. 6.9 reveal that the discontinuity of the link distribution across the borders of the considered areas could be another possible source of discrepancies. We give more insight about this issue examining the maps of the total rainfall accumulation in the Supplement.

Even knowing that the limitations we have just discussed are not negligible, we can still compare the CML interpolated product's performance against the traditional ones, to see whether some overall sensing skill is present or not. We used the areal-averaged hourly rainfall accumulations (see Section 6.4.2) to compare products with different spatial resolutions. The comparison between radar and CML is particularly interesting as they appear to be rather complementary data sources. CML product in this setup clearly lacks the detection capability (POD) of the radar. CML retrieval process however, being based on electromagnetic attenuation instead of back-scattering, does not share the radar's high sensitivity to the drop size distribution (Leijnse et al., 2008). This could make the CML a more robust sensor, in the sense that the same coefficients can

be applied regardless the different types of rain (convective, stratiform, mixed), and the values of the continuous indicators seem to endorse that.

Alongside the considerations on the sensing skills, it is valuable to a forecaster in an operational context also the time which will elapse from the acquisition of primary data (ideally, the occurrence of the event) to the actual delivery of the product ready for use. We referred to it as "latency". It can be seen from Table 6.7 that the combination of short-latency and high resolution provided by CMLs is unmatched by all the other products except the raw radar, which although lacks the required quantitative accuracy. It is left to the operators' preference, based on products' error structure, current meteorological conditions, and customers requirements, to make use of the most suitable product or of a combination of them. CMLs are valuably able to widen the range of available options.

6.6 Further insights

Other aspects were addressed only briefly and are presented below in the hope that they will be interesting hints for further research efforts.

6.6.1 Boundary effects on total accumulations

The cumulated rain over the entire period on every $5 \times 5\,\mathrm{km}^2$ gridbox is analysed with the following maps and scatterplot. Most of the points show a cumulated depth of around $100\,\mathrm{mm}$ over two months. Most of the CML product shows a 10-50 % underestimation, while two different branches gather the highest positive and negative discrepancies. When looking at the map, the points of the two branches are not randomly placed but are grouped in specific zones. It is reasonable to expect that the presence of the boundaries of the domain is probably affecting both the CML product skill and the reference reliability. Further studies could work on subsampled areas where raingauges are the most uniform and where CMLs are present not only inside but also outside the interpolation region.

6.6.2 Calibration attempts of *Aa* and *alpha*

We performed some sensitivity analysis for *Aa* and *alpha*, as recommended by Overeem et al. (2016a). However, it is our feeling that the reference data we had available (which are used daily in operational offices) are not ideal to be used as a calibrator, in terms

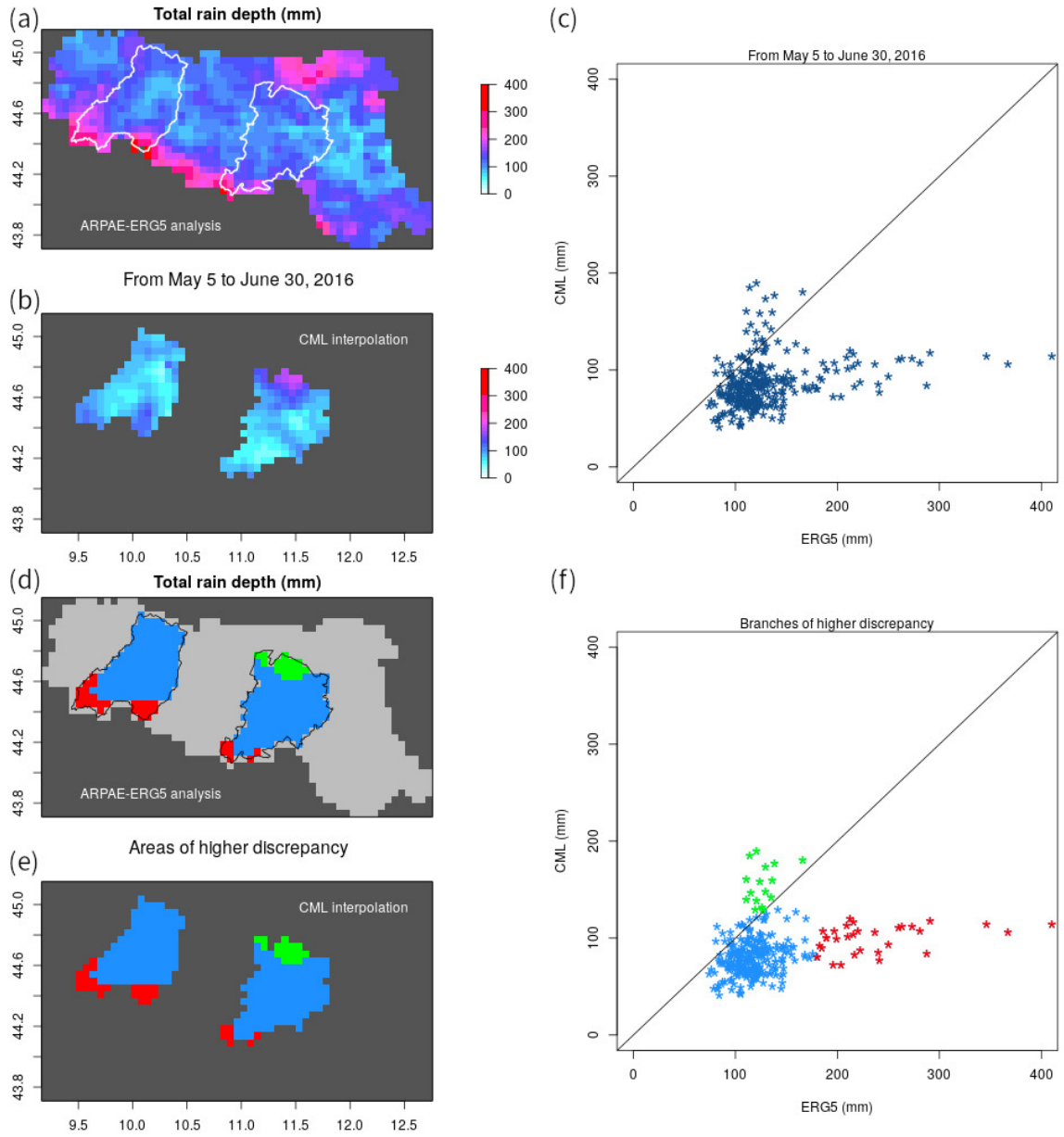

Figure 6.12: Top: total accumulated rain depth over May and June 2016. Bottom: branches of high discrepancy are highlighted on the map.

of quality and spatial and temporal characteristics. For this reason, we eventually chose not to use the calibrated parameters inside the primary validation process. We checked for the 27 links with a close-by 15 min raingauge in which *Aa* and *alpha* values were producing the best overall performances, in terms of Pearson's Correlation Coefficient (CC) and normalised coefficient of variation (CV). The results are summarised in the following Fig. 6.13.

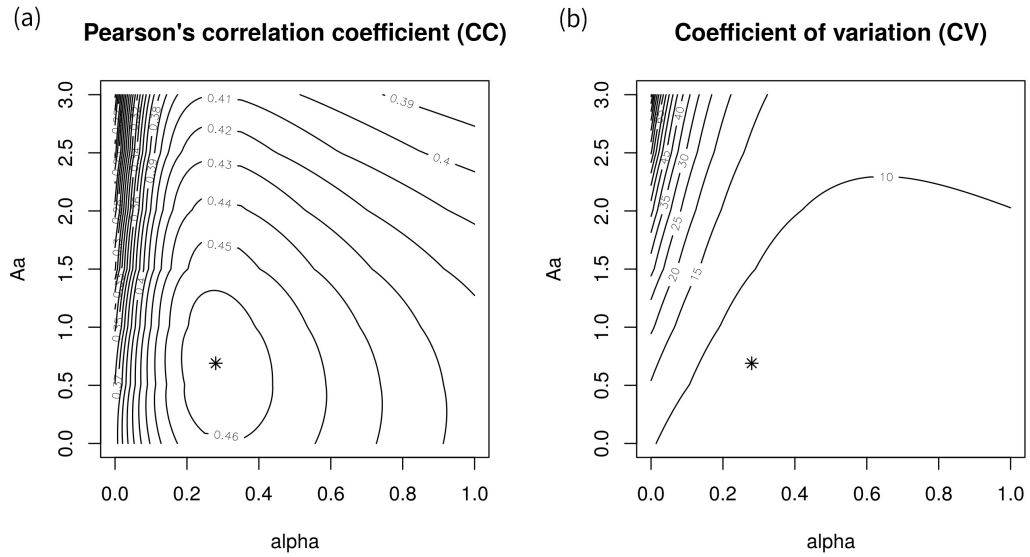

Figure 6.13: Sensitivity analyses to two coupled retrieval parameters of the RAINLINK algorithm. Loss functions CC and CV.

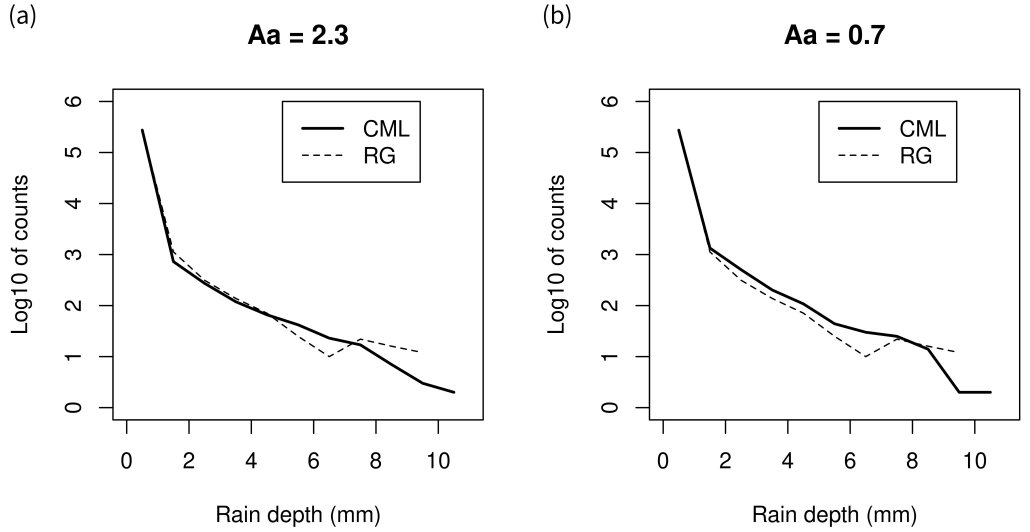

Figure 6.14: PDF of the single-CML estimates against the raingauges for two different values of *Aa*.

The CC surface (left) shows a clear maximum at $alpha = 0.3$, $Aa = 0.7$, while CV (right) reaches no local minimum in the examined domain, but has a plateau-like area of good performance, in which falls the best match for CC. This analysis suggests an *Aa* value much smaller than the one chosen by Overeem et al. (2016a) (and kept for our

work), which was 2.3 dB, while the *alpha* parameter remains almost unchanged from the default value of 0.33.

However, looking at the PDF of the estimated rainfall with the two different *Aa*, (see Fig. 6.14), emerges that the physical representativeness drops down with the lower *Aa*. Applying the calibrated parameters on the whole dataset also leads to worse overall results: CV worsens of 0.05 with no sensible improvement on R2, while the same happens for FAR and ETS, which change of 0.08 and 0.01 respectively.

Conclusions

An assessment of the rainfall retrieval capability of CML opportunistic sensors over complex terrain in northern Italy is conducted at different spatial and temporal scales for two months of data. We implemented the open-source RAINLINK algorithm in a new area and data, where no regional CML studies had previously been performed. We evaluated its performance through a complete validation scheme which utilises operational precipitation products as reference, gauging at the same time also the implementation efforts and identifying major strengths and weaknesses to facilitate profitable use of CML products.

First, 26 CMLs are compared with the closest rain gauges at a 15 min scale. Overestimation and underestimation of rain amount are both present, though the latter appears dominant. A marked variability among different links does not prevent to achieve a generally acceptable skill (CC from 0.50 to 0.88). The wet-dry classification approach (NLA, based on the spatial correlation of the rain) and the value of the wet antenna correction (Aa) may produce some misses in both rainfall occurrence and amount, particularly in case of small scale or intermittent episodes. Finally, CMLs located at higher elevation generally show worse performance.

Interpolated products obtained from 308 links confirm that a non-negligible quantity of rain is missed (normalised Mean Error is -0.26, overall CC is 0.68, and overall CV is 0.78), but also show that the rain retrieval capability is suitable for the operational application, especially if the product is integrated over large areas (CC rises to 0.92). Higher link densities increase the quality of the CML estimates at both grid box and basin scales, mostly in terms of decreased FAR.

Performance at the daily scale shows enhanced skill in case of heavy precipitations, even in the case of localised episodes. Problems arise instead during light to moderate rainfalls, when the limitations emerged during the single-link analysis become evident. Negative impact on the overall results comes from areas with poor sensor coverage, especially near the borders of the studied areas, but it should be considered that also reference rainfall fields can be affected by shortcomings of the same nature.

Furthermore, when compared to other products currently available for real-time operational exploitation, the RAINLINK output shows similar or better abilities, especially if low FAR is valued more than high POD and if latency is also taken into account. The integration of a CML-based product into an operational weather service appears worthwhile, even in a plug-in implementation that omits specific local calibration.

The performances of this technology will increase the more it spreads and the more data is made available. Further project could find valuable information in this work about both good strategies to replicate and errors to avoid. For this purpose, a list of technical specs and requirements is also added in the Appendix.

Acknowledgements

I would first like to thank Aart Overeem and his colleagues for having released open source the RAINLINK algorithm he developed and for the kind feedback and support they provided to this research without hesitation.

I will dearly thank Stefania Pasetti, Marco Folegani, Simone Mantovani, Stefano Natali, Mario Cavicchi, Moris Pozzati, Sergio Ferraresi, Damiano Barboni, Laura Vettorello, Lorenzo Gardenghi and Maria Luisa Quarta of MEEO S.r.l. for their support and friendliness starting this project.

I am very grateful to D. Vecchiato and A. Viaro of Vodafone Italia S.p.A. and to Stefania Nanni and Stefano Giannandrea of Lepida S.c.p.A. for the technical assistance with the data.

I would like to deeply thank Pier Paolo Alberoni and Anna Fornasiero of ARPAE-SIMC for the invaluable contribution they gave to this work and to my personal growth.

And I am very thankful to Prof. Federico Porcù, without whom this journey would not even have begun, for having believed in my strengths and for having endured my weaknesses with great patience.

Arianna, my family and my friends know how happy and grateful I am to have them.

Suggested specifications and requirements for future projects

These are our suggested data requirements and technical specifications to build a reliable CML rain retrieval system (updated May 3, 2020). The ideal CML candidate is a duplex microwave link, namely a radio relay between two antennas (A and B) with two back and forth sub-links: from A to B and from B to A.

Data

A unambiguous date-time stamp is needed, preferably in UTC time zone. Two univocal ID are required for the two sublinks, consistently indicated also in the metadata sheet. Receiving (Rx) and transmitting (Tx) instantaneous power levels are needed, expressed in dBm. For given duplex CML, four readings are needed: Tx_A, Tx_B, Rx_A, Rx_B. Power resolution should be at least $1\,\mathsf{dBm}$, better if $0.3\,\mathsf{dBm}$ or even $0.1\,\mathsf{dBm}$. Temporal resolution should be of $1\,\mathsf{min}$ or higher ($10\,\mathsf{s}$, $1\,\mathsf{s}$).

Metadata

Carrier frequencies should be indicated for both sublinks, with resolution of at least $100\,\mathsf{MHz}$ or better. Carrier polarisation should be specified (H or V). Antenna coordinates (four per CML: LatA, LonA, LatB, LonB) should have at least 0.0001° resolution.

Ground height: at least for the antennas but better if a regional Digital Elevation Model (DEM) is available. Antenna tower heights in meters (both A and B) is useful information too. Link path-length (km is not explicitly required as it is easily calculated from antenna coordinates.

Periods of outage should be clearly specified by `NA`s other reference value (e.g. $-99\,\mathsf{dBm}$). Minimum power step, i.e. the actual power resolution (dBm), has to be specified because it could differ from the number of decimals provided for the Tx and Rx power levels.

Technical details

Moreover, information about the technical system are valuable for the data analysis: the hardware actual logging frequency (usually 10 Hz or higher), the antenna manufacturer, the antenna model and type, the antenna's dish dimensions, programmed servicing and maintenance, date of installation. For wet-antenna attenuation studies, information about the presence, material and status of antenna protective radomes is vital.

Data access and acquisition

Manual data transfer should be discouraged (spreadsheets, zipped folders), favouring instead automated standardised solutions. Data could be transferred periodically via FTP by the provider to the researchers and operators, or access to the server logging the data and hosted by the provider could be given to the researchers with SSH authentication (better management and lower latency), or even the researchers could perform direct SNMP interrogations to the cellular towers to retrieve power informations (Chwala et al., 2016)

Policy on data disclosure should be as open-source as possible, at least for some sample dataset, in order to encourage international joint collaborations and allow algorithm calibrations from different teams.

Bibliography

A.Y. Abdulrahman, T. Bin Abdulrahman, S.K. Bin Abdulrahim, and U. Kesavan. Comparison of measured rain attenuation and itu-r predictions on experimental microwave links in malaysia. *International Journal of Microwave and Wireless Technologies*, 3:477–483, 08 2011. doi: 10.1017/S1759078711000171.

Pier Paolo Alberoni, T Andersson, P Mezzasalma, D B Michelson, and S Nanni. Use of the vertical reflectivity profile for identification of anomalous propagation. *Meteorological Applications*, 8(3):257–266, 2001. doi: 10.1017/S1350482701003012.

Pier Paolo Alberoni, Anna Fornasiero, Giacomo Roversi, Stefania Pasetti, Marco Folegani, and Federico Porcù. Comparison between different qpe based on: Microwave links, radar adjusted and gauges. In *10th European Conference on Radar in Meteorology and Hydrology*, pages 851–860. KNMI, 2018. doi: 10.18174/454537.

R. Amorati, P.P. Alberoni, and A. Fornasiero. Operational bias correction of hourly radar precipitation estimate using rain gauges. In *Proceedings of the Seventh Eropean Conference on Radar in meteorology and Hydrology*, 2012. URL `http://www.meteo.fr/cic/meetings/2012/ERAD/extended_abs/QPE_007_ext_abs.pdf`.

G. Antolini, L. Auteri, V. Pavan, F. Tomei, R. Tomozeiu, and V. Marletto. A daily high-resolution gridded climatic data set for emilia romagna, italy, during 1961-2010. *International Journal of Climatology*, 36:1970–1986, 2016.

David Atlas and Carlton W. Ulbrich. Path- and area-integrated rainfall measurement by microwave attenuation in the 1–3 cm band. *Journal of Applied Meteorology and Climatology*, 16(12):1322 – 1331, 1977. doi: 10.1175/1520-0450(1977)016<1322:PAAIRM>2.0.CO;2.

Joan Bech, Bernat Codina, Jeroni Lorente, and David Bebbington. The sensitivity of single polarization weather radar beam blockage correction to variability in the vertical refractivity gradient. *Journal of Atmospheric and Oceanic Technology*, 20(6):845–855, 2003. doi: 10.1175/1520-0426(2003)020<0845:TSOSPW>2.0.CO;2.

A. Berne and R. Uijlenhoet. Path-averaged rainfall estimation using microwave links: Uncertainty due to spatial rainfall variability. *Geophysical Research Letters*, 34(7), 2007. doi: 10.1029/2007GL029409.

Blandine Bianchi, Peter Jan van Leeuwen, Robin J. Hogan, and Alexis Berne. A variational approach to retrieve rain rate by combining information from rain gauges, radars, and microwave links. *Journal of Hydrometeorology*, 14(6):1897–1909, 2013. doi: 10.1175/JHM-D-12-094.1.

D.C. Bowman and J.M. Lees. Near real time weather and ocean model data access with rnomads. *Computers and Geosciences*, 78:88 – 95, 2015. ISSN 0098-3004. doi: https://doi.org/10.1016/j.cageo.2015.02.013.

C. Caracciolo, F. Prodi, A. Battaglia, and F. Porcù. Analysis of the moments and parameters of a gamma dsd to infer precipitation properties: A convective stratiform discrimination algorithm. *Atmospheric Research*, 80:165–186, 2006a. doi: https://doi.org/10.1016/j.atmosres.2005.07.003.

Clelia Caracciolo, Franco Prodi, and Remko Uijlenhoet. Comparison between pludix and impact/optical disdrometers during rainfall measurement campaigns. *Atmospheric Research*, 82 (1):137 – 163, 2006b. ISSN 0169-8095. doi: https://doi.org/10.1016/j.atmosres.2005.09.007. 14th International Conference on Clouds and Precipitation.

C. Chwala, F. Keis, and H. Kunstmann. Real-time data acquisition of commercial microwave link networks for hydrometeorological applications. *Atmospheric Measurement Techniques*, 9(3):991–999, 2016. doi: 10.5194/amt-9-991-2016.

Christian Chwala and Harald Kunstmann. Commercial microwave link networks for rainfall observation: Assessment of the current status and future challenges. *WIREs Water*, 6(2): e1337, 2019. doi: 10.1002/wat2.1337.

R.J. Cummings, G.J.G. Upton, A.R. Holt, and M. Kitchen. Using microwave links to adjust the radar rainfall field. *Advances in Water Resources*, 32(7):1003 – 1010, 2009. ISSN 0309-1708. doi: https://doi.org/10.1016/j.advwatres.2008.08.010. Weather Radar and Hydrology.

L. W. de Vos, A. Overeem, H. Leijnse, and R. Uijlenhoet. Rainfall estimation accuracy of a nationwide instantaneously sampling commercial microwave link network: Error dependency on known characteristics. *Journal of Atmospheric and Oceanic Technology*, 36(7):1267–1283, 2019. doi: 10.1175/JTECH-D-18-0197.1.

L.W. de Vos, T.H. Raupach, H. Leijnse, A. Overeem, A. Berne, and R. Uijlenhoet. High-resolution simulation study exploring the potential of radars, crowdsourced personal weather

stations, and commercial microwave links to monitor small-scale urban rainfall. *Water Resources Research*, 54:10293–10312, 2018. doi: 10.1029/2018WR023393.

A. Doumounia, M. Gosset, F. Cazenave, M. Kacou, and F. Zougmore. Rainfall monitoring based on microwave links from cellular telecommunication networks: First results from a west african test bed. *Geophysical Research Letters*, 41:6015–6021, 2014. doi: 10.1002/2014GL060724.

Haralambos Feidas, Federico Porcu, S. Puca, Angelo Rinollo, C. Lagouvardos, and Vassiliki Kotroni. Validation of the h-saf precipitation product h03 over greece using rain gauge data. *Theoretical and Applied Climatology*, 131:377–398, 11 2018. doi: 10.1007/s00704-016-1981-9.

M. Fencl, M. Dohnal, J. Rieckermann, and V. Bareš. Gauge-adjusted rainfall estimates from commercial microwave links. *Hydrology and Earth System Sciences*, 21(1):617–634, 2017. doi: https://doi.org/10.5194/hess-21-617-2017.

Martin Fencl, Jörg Rieckermann, Marc Schleiss, David Stránský, and Vojtěch Bareš. Assessing the potential of using telecommunication microwave links in urban drainage modelling. *Water science and technology : a journal of the International Association on Water Pollution Research*, 68:1810–8, 11 2013. doi: 10.2166/wst.2013.429.

Fabrizio Fenicia, Laurent Pfister, Dmitri Kavetski, Patrick Matgen, Jean-François Iffly, Lucien Hoffmann, and Remko Uijlenhoet. Microwave links for rainfall estimation in an urban environment: Insights from an experimental setup in luxembourg-city. *Journal of Hydrology*, 464-465:69 – 78, 2012. ISSN 0022-1694. doi: https://doi.org/10.1016/j.jhydrol.2012.06.047.

Jordi Figueras i Ventura, AA Boumahmoud, B Fradon, P Dupuy, and P Tabary. Long-term monitoring of french polarimetric radar data quality and evaluation of several polarimetric quantitative precipitation estimators in ideal conditions for operational implementation at c-band. *Quarterly Journal of the Royal Meteorological Society*, 138:2212–2228, 2012. doi: 10.1002/qj.1934.

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

P. Goovaerts. Geostatistics in soil science: state-of-the-art and perspectives. *Geoderma*, 89:1–45, 1999. doi: https://doi.org/10.1016/S0016-7061(98)00078-0.

Y. Gou, H. Chen, and J. Zheng. Polarimetric radar signatures and performance of various radar rainfall estimators during an extreme precipitation event over the thousand-island lake area in eastern china. *Remote Sensing*, 11:2335, 2019. doi: 10.3390/rs11202335.

Mircea Grecu, William Olson, s Munchak, Sarah Ringerud, Liang Liao, Ziad Haddad, Bartie Kelley, and Steven McLaughlin. The gpm combined algorithm. *Journal of Atmospheric and Oceanic Technology*, 33:2225–2245, 2016. doi: 10.1175/JTECH-D-16-0019.1.

Morten Grum, Stefan Kraemer, Hans-Reinhard Verworn, and Axel Redder. Combined use of point rain gauges, radar, microwave link and level measurements in urban hydrological modelling. *Atmospheric Research*, 77(1):313 – 321, 2005. ISSN 0169-8095. doi: https://doi.org/10.1016/j.atmosres.2004.10.013. Precipitation in Urban Areas.

B. Haese, S. Hörning, C. Chwala, A. Bardossy, B. Schalge, , and H. Kunstmann. Stochastic reconstruction and interpolation of precipitation fields using combined information of commercial microwave links and rain gauges. *Water Resources Research*, 53:10740–10756, 2017. doi: https://doi.org/10.1002/2017WR021015.

B.N. Harden, J.R. Norbury, and W.J.K. White. Attenuation/rain-rate relationships on terrestrial microwave links in the frequency range 10- -40 ghz. *Electronics Letters*, 14(5):154 – 155, 1978. doi: 10.1049/el:19780103.

Asko Huuskonen, E. Saltikoff, and Iwan Holleman. The operational weather radar network in europe. *Bulletin of the American Meteorological Society*, 95:897–907, 06 2014. doi: 10.1175/BAMS-D-12-00216.1.

ITU-R. Specific attenuation model for rain for use in prediction methods, 2005. URL `https://www.itu.int/dms_pubrec/itu-r/rec/p/R-REC-P.838-3-200503-I!!PDF-E.pdf`.

A. R. Jameson. Estimating the path-average rainwater content and updraft speed along a microwave link. *Journal of Atmospheric and Oceanic Technology*, 10(4):478–485, 1993. doi: 10.1175/1520-0426(1993)010<0478:ETPARC>2.0.CO;2.

M. Kigobe and M. Kizza. Dealing with spatial variability under limited hydrogeological data. case study: Hydrological parameter estimation in mpigi-wakiso. In *Proceedings from the International Conference on Advances in Engineering and Technology, Elsevier Science Ltd*, pages 211–220, 2006. doi: https://doi.org/10.1016/B978-008045312-5/50024-8.

J. Koistinen and T. Puhakka. An improved spatial gauge-radar adjustment technique. In *Preprints of the 20th Conference on Radar Meteorology*, page 179–186. Am. Meteorol. Soc., 1981.

Luca Lanza and L Stagi. Non-parametric error distribution analysis from the laboratory calibration of various rainfall intensity gauges. *Water science and technology : a journal of the International Association on Water Pollution Research*, 65:1745–52, 04 2012. doi: 10.2166/wst.2012.075.

H. Leijnse, Remko Uijlenhoet, and J.N.M. Stricker. Microwave link rainfall estimation: Effects of link length and frequency, temporal sampling, power resolution, and wet antenna attenuation. *Advances in Water Resources*, 31:1481–1493, 11 2008. doi: 10.1016/j.advwatres.2008.03.004.

Hidde Leijnse, Remko Uijlenhoet, and Alexis Berne. Errors and uncertainties in microwave link rainfall estimation explored using drop size measurements and high-resolution radar data. *Journal of Hydrometeorology*, 11(6):1330–1344, 2010. doi: 10.1175/2010JHM1243.1.

H Leinse. *Hydro-meteorological application of microwave links - Measurement of evaporation and precipitation.* PhD thesis, Wageningen University, 2007.

T.J. Matejka, Jr Houze, R.A., and P.V. Hobbs. Microphysics and dynamics of clouds associated with mesoscale rainbands in extratropical cyclones. *Quarterly Journal of the Royal Meteorological Society*, 106:29–56, 1980. doi: https://doi.org/10.1002/qj.49710644704.

A. Mugnai, D. Casella, E. Cattani, S. Dietrich, S. Laviola, V. Levizzani, G. Panegrossi, M. Petracca, P. Sanò, F. Di Paola, D. Biron, L. De Leonibus, D. Melfi, P. Rosci, A. Vocino, F. Zauli, P. Pagliara, S. Puca, A. Rinollo, L. Milani, F. Porcù, and F. Gattari. Precipitation products from the hydrology saf. *Natural Hazards and Earth System Sciences*, 13(8):1959–1981, 2013. doi: 10.5194/nhess-13-1959-2013.

Chrispin Mulangu and Thomas Afullo. Variability of the propagation coefficients due to rain for microwave links in southern africa. *Radio Science - RADIO SCI*, 44, 06 2009. doi: 10.1029/2008RS003912.

Pertti Nurmi. Recommendations on the verification of local weather forecasts. *ECMWF Technical Memorandum*, 430, 01 2003.

Jonatan Ostrometzky and Hagit Messer. On the information in extreme measurements for parameter estimation. *Journal of the Franklin Institute*, 357(1):748–771, January 2020. doi: 10.1016/j.jfranklin.2019.11.039.

A. Overeem, H. Leijnse, and R. Uijlenhoet. Measuring urban rainfall using microwave links from commercial cellular communication networks. *Water Resources Research*, 47(12), 2011. doi: 10.1029/2010WR010350.

Aart Overeem, Hidde Leijnse, and Remko Uijlenhoet. Country-wide rainfall maps from cellular communication networks. *Proceedings of the National Academy of Sciences*, 110(8):2741–2745, 2013. ISSN 0027-8424. doi: 10.1073/pnas.1217961110.

Aart Overeem, Hidde Leijnse, and Remko Uijlenhoet. Retrieval algorithm for rainfall mapping from microwave links in a cellular communication network. *Atmospheric Measurement Techniques*, 9:2425–2444, 06 2016a. doi: 10.5194/amt-9-2425-2016.

Aart Overeem, Hidde Leijnse, and Remko Uijlenhoet. Two and a half years of country-wide rainfall maps using radio links from commercial cellular telecommunication networks. *Water Resources Research*, 52:8039–8065, 10 2016b. doi: 10.1002/2016WR019412.

Marco Petracca, Leo D'Adderio, Federico Porcu, Gianfranco Vulpiani, Sebastianelli Stefano, and S. Puca. Validation of gpm dual-frequency precipitation radar (dpr) rainfall products over italy. *Journal of Hydrometeorology*, 19:907–925, 2018. doi: 10.1175/JHM-D-17-0144.1.

Federico Porcù, Lisa Milani, and Marco Petracca. On the uncertainties in validating satellite instantaneous rainfall estimates with raingauge operational network. *Atmospheric Research*, 144:73–81, 07 2014. doi: 10.1016/j.atmosres.2013.12.007.

S. Puca, Federico Porcu, Angelo Rinollo, Gianfranco Vulpiani, Pierre Baguis, Snezhanka Balabanova, E. Campione, A. Erturk, Simone Gabellani, R. Iwanski, Marian Jurašek, Jan Kaňák, Judit Kerényi, Georgy Koshinchanov, G. Kozinarova, Peter Krahe, B. Lapeta, Eszter Labo, Lisa Milani, and F. Gattari. The validation service of the hydrological saf geostationary and polar satellite precipitation products. *Natural hazards and earth system sciences*, 14:871–889, 04 2014. doi: 10.5194/nhess-14-871-2014.

A. R. Rahimi, A. R. Holt, G. J. G. Upton, and R. J. Cummings. Use of dual-frequency microwave links for measuring path-averaged rainfall. *Journal of Geophysical Research: Atmospheres*, 108(D15), 2003. doi: 10.1029/2002JD003202.

Rafael Rincon and Roger Lang. Microwave link dual-wavelength measurements of path-average attenuation for the estimation of drop size distributions and rainfall. *Geoscience and Remote Sensing, IEEE Transactions on*, 40:760 – 770, 05 2002. doi: 10.1109/TGRS.2002.1006324.

Manuel Rios Gaona, Aart Overeem, Tim Raupach, Hidde Leijnse, and Remko Uijlenhoet. Rainfall retrieval with commercial microwave links in são paulo, brazil. *Atmospheric Measurement Techniques Discussions*, pages 1–21, 09 2017. doi: 10.5194/amt-2017-287.

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

E. Saltikoff, Günther Haase, Laurent Delobbe, Nicolas Gaussiat, Maud Martet, Daniel Idziorek, Hidde Leijnse, Petr Novák, Maryna Lukach, and Klaus Stephan. Opera the radar project. *Atmosphere*, 10:320, 06 2019. doi: 10.3390/atmos10060320.

Robert J. Serafin and James W. Wilson. Operational weather radar in the united states: Progress and opportunity. *Bulletin of the American Meteorological Society*, 81(3):501–518, 2000. doi: 10.1175/1520-0477(2000)081<0501:OWRITU>2.3.CO;2.

Donald Shepard. A two-dimensional interpolation function for irregularly-spaced data. In *Proceedings of the 1968 23rd ACM National Conference*, ACM '68, page 517–524, New York, NY, USA, 1968. Association for Computing Machinery. ISBN 9781450374866. doi: 10.1145/800186.810616.

Gail Skofronick-Jackson, Walter A. Petersen, Wesley Berg, Chris Kidd, Erich F. Stocker, Dalia B. Kirschbaum, Ramesh Kakar, Scott A. Braun, George J. Huffman, Toshio Iguchi, Pierre E. Kirstetter, Christian Kummerow, Robert Meneghini, Riko Oki, William S. Olson, Yukari N. Takayabu, Kinji Furukawa, and Thomas Wilheit. The global precipitation measurement (gpm) mission for science and society. *Bulletin of the American Meteorological Society*, 98(8):1679–1695, 2017. doi: 10.1175/BAMS-D-15-00306.1.

Peter Speirs, Marco Gabella, and Alexis Berne. A comparison between the gpm dual-frequency precipitation radar and ground-based radar precipitation rate estimates in the swiss alps and plateau. *Journal of Hydrometeorology*, 18:1247–1269, 2017. doi: 10.1175/JHM-D-16-0085.1.

G. Tang, M. P. Clark, S. M. Papalexiou, Z. Ma, and Y. Hong. Hydro-meteorological assessment of three gpm satellite precipitation products in the kelantan river basin, malaysia. *Remote Sensing of Environment*, 240:11697, 2020. doi: 10.1016/j.rse.2020.111697.

Ali Tokay, Leo Pio D'Adderio, Federico Porcù, David B. Wolff, and Walter A. Petersen. A field study of footprint-scale variability of raindrop size distribution. *Journal of Hydrometeorology*, 18(12):3165–3179, 2017. doi: 10.1175/JHM-D-17-0003.1.

D. J. W. Turner and D. Turner. Attenuation due to rainfall on a 24 km microwave link working at 11, 18 and 36 ghz. *Electronics Letters*, 6(10):297–298, May 1970. ISSN 0013-5194. doi: 10.1049/el:19700208.

Graham Upton, A.R. Holt, R.J. Cummings, A.R. Rahimi, and J.W.F. Goddard. Microwave links: The future for urban rainfall measurement? *Atmospheric Research*, 77:300–312, 09 2005. doi: 10.1016/j.atmosres.2004.10.009.

C.Z. van de Beek, H. Leijnse, P.J.J.F. Torfs, and R. Uijlenhoet. Seasonal semi-variance of dutch rainfall at hourly to daily scales. *Advances in Water Resources*, 45:76–85, 2012. doi: https://doi.org/10.1016/j.advwatres.2012.03.023. Space-Time Precipitation from Urban Scale to Global Change.

H. C. van der Hulst. *Light scattering by small particles*. John Wiley & Sons, New York, USA, 1957, 1981.

T. C. van Leth, H. Leijnse, A. Overeem, and R. Uijlenhoet. Estimating raindrop size distributions using microwave link measurements. *Atmospheric Measurement Techniques Discussions*, 2019:1–27, 2019. doi: 10.5194/amt-2019-51.

Thomas van Leth, Aart Overeem, Hidde Leijnse, and Remko Uijlenhoet. A measurement campaign to assess sources of error in microwave link rainfall estimation. *Atmospheric Measurement Techniques*, 11:4645–4669, 08 2018. doi: 10.5194/amt-11-4645-2018.

W. Wolff, A. Overeem, H. Leijnse, and R. Uijlenhoet. Rainfall retrieval algorithm for commercial microwave links: stochastic calibration. *Atmospheric Measurement Techniques*, 15(2): 485–502, 2022. doi: 10.5194/amt-15-485-2022.