

Transboundary rainfall estimation using commercial microwave links

Nico Blettner, Martin Fencel, Vojtěch Bareš, Harald Kunstmann, Christian Chwala

Angaben zur Veröffentlichung / Publication details:

Blettner, Nico, Martin Fencel, Vojtěch Bareš, Harald Kunstmann, and Christian Chwala. 2023. "Transboundary rainfall estimation using commercial microwave links." *Earth and Space Science* 10 (8): e2023EA002869. <https://doi.org/10.1029/2023ea002869>.

Earth and Space Science



RESEARCH LETTER

10.1029/2023EA002869

Key Points:

- Transboundary rainfall maps based on independent networks of commercial microwave links (CMLs) are generated for the first time
- German and Czech data sets of CMLs differ significantly with respect to network characteristics
- Quality control is important for heterogeneous data of CMLs

Correspondence to:

N. Blettner,
nico.blettner@kit.edu

Citation:

Blettner, N., Fencel, M., Bareš, V., Kunstmann, H., & Chwala, C. (2023). Transboundary rainfall estimation using commercial microwave links. *Earth and Space Science*, 10, e2023EA002869. <https://doi.org/10.1029/2023EA002869>





Received 3 FEB 2023

Accepted 6 AUG 2023

Author Contributions:

Conceptualization: Nico Blettner, Martin Fencel, Vojtěch Bareš, Harald Kunstmann, Christian Chwala
Data curation: Nico Blettner, Martin Fencel, Christian Chwala
Formal analysis: Nico Blettner, Martin Fencel
Funding acquisition: Martin Fencel, Vojtěch Bareš, Harald Kunstmann, Christian Chwala
Investigation: Nico Blettner, Martin Fencel, Vojtěch Bareš, Christian Chwala
Methodology: Nico Blettner, Christian Chwala
Project Administration: Nico Blettner, Martin Fencel, Vojtěch Bareš, Harald Kunstmann, Christian Chwala
Resources: Martin Fencel, Vojtěch Bareš, Christian Chwala
Software: Nico Blettner, Christian Chwala

Transboundary Rainfall Estimation Using Commercial Microwave Links

Nico Blettner^{1,2} , Martin Fencel³ , Vojtěch Bareš³, Harald Kunstmann^{1,2} , and Christian Chwala¹ 

¹Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology, Garmisch-Partenkirchen, Germany, ²Institute of Geography, University of Augsburg, Augsburg, Germany, ³Department of Hydraulics and Hydrology, Czech Technical University in Prague, Prague, Czech Republic

Abstract Unlike actual rainfall, the spatial extent of rainfall maps is often determined by administrative and political boundaries. Similarly, data from commercial microwave links (CMLs) is usually acquired on a national basis and exchange among countries is limited. Up to now, this has prohibited the generation of transboundary CML-based rainfall maps despite the great extension of networks across the world. We present CML based transboundary rainfall maps for the first time, using independent CML data sets from Germany and the Czech Republic. We show that straightforward algorithms used for quality control strongly reduce anomalies in the results. We find that, after quality control, CML-based rainfall maps can be generated via joint and consistent processing, and that these maps allow to seamlessly visualize rainfall events traversing the German-Czech border. This demonstrates that quality control represents a crucial step for large-scale (e.g., continental) CML-based rainfall estimation.

Plain Language Summary Rainfall maps are usually based on gauge observations on the ground or radar. They are crucial for predicting or reconstructing flooding events. Commercial microwave links are special kinds of rainfall sensors. Their actual purpose is the signal propagation within a cellular network. However, since the signal is attenuated when it rains, they can also be exploited for rainfall estimation. To estimate rainfall from the observed attenuation requires careful data processing. Algorithms for that are usually adjusted to national data sets with their specific characteristics. In this study, we use, for the first time, two independent data sets of commercial microwave links (from Germany and the Czech Republic) with the aim of generating transboundary rainfall maps. As the data sets vary in many respects, several algorithms need to be adjusted and extended to allow processing them consistently. We show that it is possible to create meaningful rainfall maps of rain events that traverse the border between Germany and the Czech Republic. This can be considered a major step toward seamless rainfall maps on even larger, that is, continental scale.

1. Introduction

Rainfall estimation depends on the quality of observational networks, which are spatially constrained, for example, by administrative and political borders. Obviously, rainfall is not constrained in such ways and rainfall maps that comprise multiple countries are required. As an example, the Ahr flooding of July 2021, which was triggered by heavy rainfall, affected the border region between Belgium, Luxembourg, and Germany. Reliable spatial rainfall information on transboundary continental scale would be beneficial in such cases.

Transboundary rainfall maps usually require the exploitation of different independent observational networks. This is challenging due to differences in data acquisition and processing algorithms. Even for dedicated rainfall observations like weather radars, it is nontrivial to merge independent networks into a consistent homogeneous product (cf. the OPERA program, Huuskonen et al., 2014).

Opportunistic sensing via commercial microwave links (CMLs) provides an enormous potential for the generation of large scale rainfall maps. Rainfall products from CMLs have been derived since the past two decades (Leijnse et al., 2007; Messer, 2006) and high-quality rainfall maps were generated in many studies (e.g., D'Amico et al., 2016; Graf et al., 2020; Overeem et al., 2013, 2016b; Roversi et al., 2020; Zinevich et al., 2008). Moreover, CMLs are very abundant globally (Chwala & Kunstmann, 2019; Uijlenhoet et al., 2018) and hence, in theory provide a useful basis for transboundary or even continental scale maps. However, this goal has so far not been addressed. Up to now, all the reported studies considered various areas using different regional or at most national CML networks. In each case CML's from single network providers were deployed.

© 2023 The Authors. Earth and Space Science published by Wiley Periodicals LLC on behalf of American Geophysical Union.

This is an open access article under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Supervision: Harald Kunstmann,
Christian Chwala

Validation: Nico Blettner

Visualization: Nico Blettner, Martin
Fencl, Vojtěch Bareš, Christian Chwala

Writing – original draft: Nico Blettner

Writing – review & editing: Nico
Blettner, Martin Fencl, Vojtěch Bareš,
Harald Kunstmann, Christian Chwala

The reasons for this gap in exploiting the potential of CMLs are legal constraints restricting the access to the data, and primarily the fact that combining two independent networks is particularly challenging for opportunistic rainfall sensors. They are not designed to measure rainfall and hence have no strict conventions and homogeneity of data quality as, for example, rain gauge networks. Deriving rainfall information from CMLs requires complex algorithms which have so far been developed for and adjusted to the specific data sets. These algorithms depend on network characteristics which can differ significantly between (national) data sets, for example, for historical reasons and because homogeneity is not a concern of the network providers. The differences affect the usability of the data as well as rainfall retrieval and mapping processes.

Hence, it is not clear and so far still an open question how transferable existing rainfall retrieval algorithms are. Can the algorithms that are established for one data set still be used when another independent data set is considered in addition? If so, does this enable the generation of seamless transboundary rainfall maps, and thereby open the pathway for large-scale continental CML-based maps?

According to these questions, the main goal in this study is to realize for the first time CML-based transboundary rainfall maps, exemplarily shown for Germany and the Czech Republic. We start with the application of the processing routines which were developed for the German data set (Graf et al., 2020) and transfer them to the combination of German and Czech data. We investigate to what extent the established methods are applicable, and propose an extension of their quality control methods where they are not applicable.

2. Data

This study covers a 1-month period (June 2021) and focuses on the border region between Germany and the Czech Republic. The CML data sets of the two countries are described in Section 2.1. As a reference, we use the gridded gauge-adjusted weather radar product RADOLAN-RW which is described in Section 2.2.

2.1. CML Data

Overall, the CML data sets comprises approximately 3,900 and 2,900 CMLs in Germany and the Czech Republic, respectively. In the analysis, we consider the border region with 1,167 (Germany) and 2,244 (Czech Republic) CMLs. Both data sets provide a 1-min temporal resolution.

The data sets vary significantly with respect to spatial distribution, frequencies, and lengths. The Czech data set has a higher CML density in populated regions (e.g., the city of Prague), whereas in Germany the CMLs are more evenly distributed. In Germany, CML frequencies essentially vary between 12 and 39 GHz while in the Czech Republic the data set comprises approximately 30% E-band CMLs with frequencies above 70 GHz. The CML length in Germany is above 1 km in 99% of the cases. In the Czech Republic, 26% of CMLs have a length below 1 km and 1% even below 0.1 km.

These differences have a strong effect on the sensitivity of the path attenuation on rainfall, that is, on the CML's detection limit, a parameter used throughout this study. The detection limit quantifies the minimum rain rate that is required to induce an integrated path attenuation of 0.33 dB, which is a standard signal quantization of the CMLs. Also, it roughly determines the precision of the retrieved rain rate. The detection limit is dependent on frequency and length which have a wider range of values in the Czech data. Hence, also the range of detection limits is larger in the Czech data. The Czech CMLs have detection limits from 0.04 to 132 mm/hr, whereas the detection limit of the German data lies mostly between 0.2 and 1 mm/hr, and only two German CMLs exceed a detection limit of 2 mm/hr. Note that a crucial difference in this regard is the presence (for the Czech Republic) and absence (for Germany) of E-band CMLs which are generally rather sensitive to rainfall. The E-band CMLs that are longer than a few kilometers have an exceptionally low detection limit below 0.1 mm/hr. This property is beneficial for sensing light rainfall but these highly sensitive CMLs are also more prone to experience very high attenuation. Strong rainfall can therefore even lead to a loss of connectivity along the CML as the receiver fails to record below a certain level (see Polz et al., 2023 and the description of blackout gaps in Section 3.1).

2.2. Reference Data RADOLAN-RW

We use RADOLAN-RW as a reference. It is a product of the German Weather Service (DWD) based on 17 C-band weather radar stations supplying gridded rainfall information on a 1-by-1 km spatial and hourly temporal

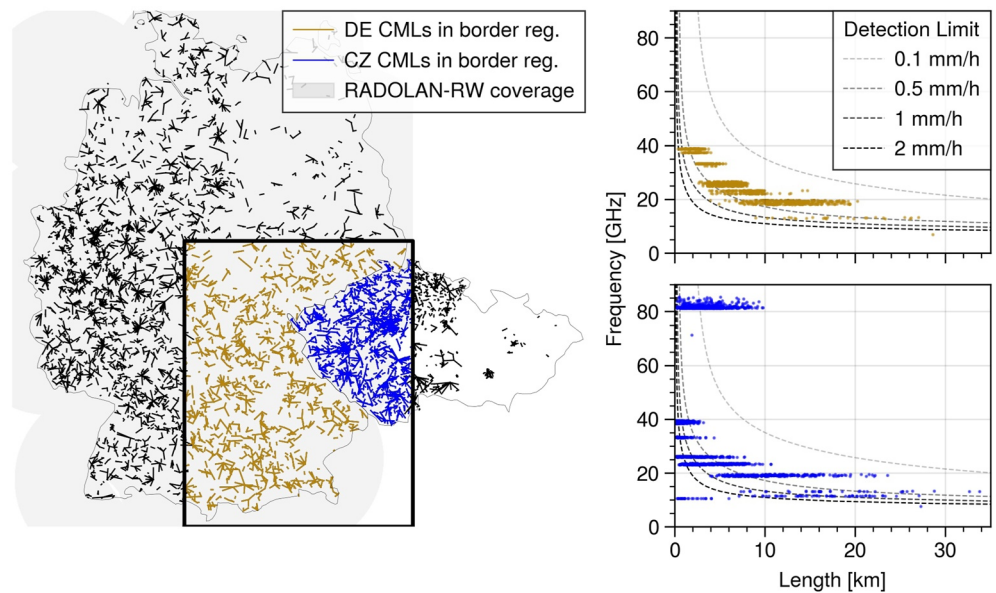


Figure 1. Data overview. Left: Sensor locations with the analyzed border region defined by the black box; the shaded background shows the coverage of RADOLAN-RW. Right: Distribution of frequency versus length of commercial microwave links (CMLs) within the analyzed region (German and Czech CMLs in upper and lower panel, respectively); dashed lines show levels of detection limit.

resolution. The radar information is adjusted to over 1,000 rain gauges by additive and multiplicative correction schemes (Bartels et al., 2004).

RADOLAN-RW was chosen as it is the official real-time product for quantitative precipitation estimation of DWD and has been used in other studies like, for example, Graf et al. (2020). Despite the fact that two countries are considered in this study, RADOLAN-RW serves as the only reference. This is done to avoid additional potential error sources stemming from combining independent reference data sets. Unfortunately, there is no gridded data set of comparable resolution and quality available that covers both Germany and the Czech Republic completely. Nevertheless, the border region considered in this study is covered by RADOLAN-RW to a large extent (see Figure 1), and hence RADOLAN-RW can be considered a suitable reference.

We compare the estimated rainfall maps with RADOLAN-RW directly (Section 4.3), but also, we evaluate the rainfall retrieval based on the CML paths (Section 4.2). For this, we calculate the weighted sum of pixel values of RADOLAN-RW taking into account the length of the pixels' intersections with the CML paths.

3. Methods

The CML processing algorithms are primarily based on those applied in Graf et al. (2020), which originally were adjusted to a purely German CML data set and a different period than considered in this study. The processing can be subdivided into two aspects: (a) dealing with erroneous data, i.e., quality control, which is particularly relevant for opportunistic data with their potentially high number of error sources associated with engineering details rather than atmospheric aspects; and (b) rain rate retrieval and mapping which involve steps that are related to well understood challenges but nonetheless are associated with considerable uncertainties. The difference between the German and Czech CML data sets required some adaptations and extensions to the quality control part of the established algorithms.

3.1. Quality Control

Investigating the data sets reveals the necessity for quality control algorithms. Before describing the individual algorithms conducted in this regard (Steps 1–6 below), we summarize the main observations that justify them. The justification is largely based on the physical limits of rainfall, its statistics and the knowledge of how rainfall can and cannot be reflected in CML observations. For most described patterns, an example is given in Figure 2.

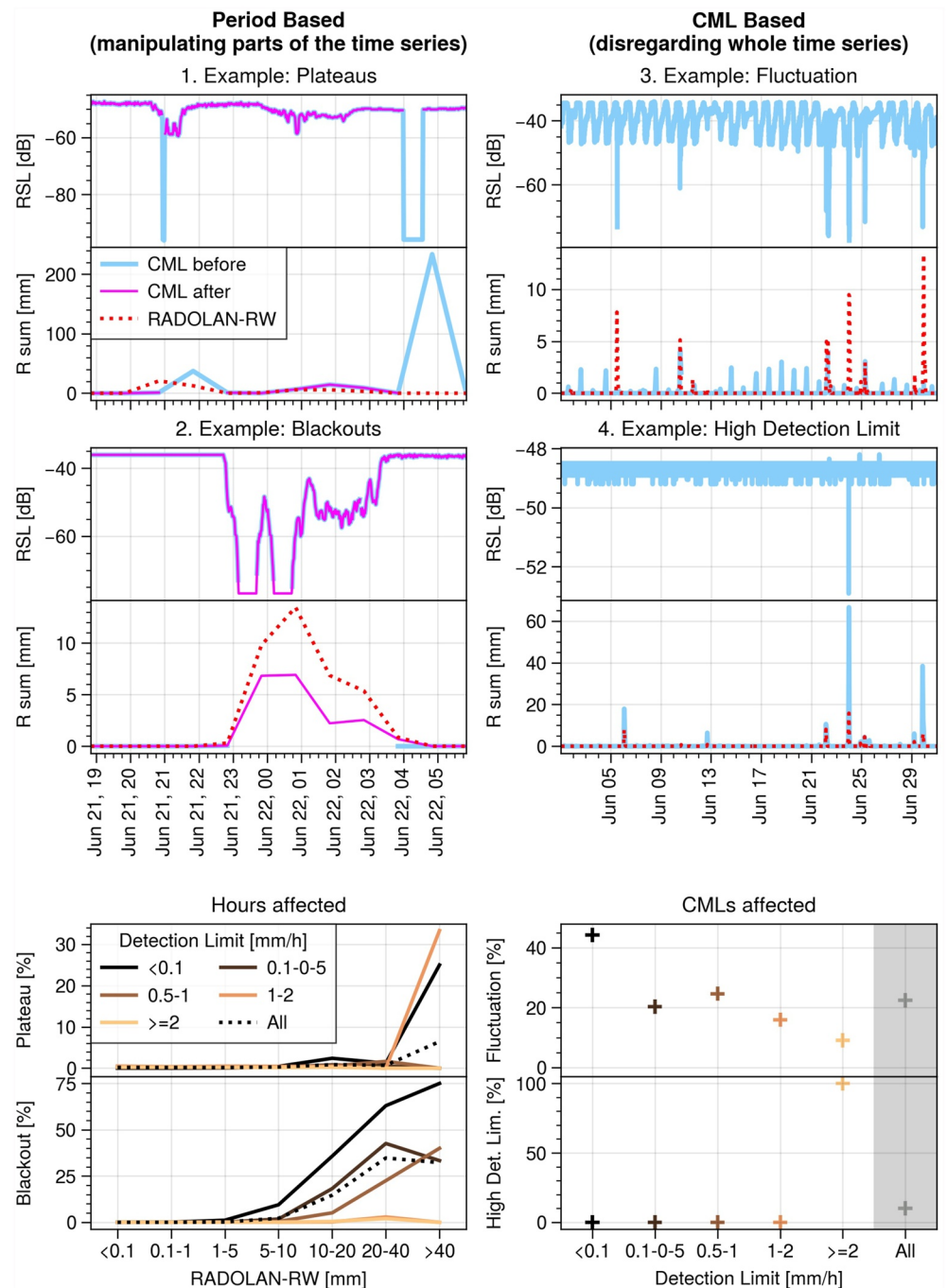


Figure 2. The effects of quality control algorithms. Four exemplary time series that have been treated by one of the steps of Section 3.1 are shown in the upper part. Statistics on the abundance of similar occurrences are presented in the lower part. The left column treats period based steps, and the right column steps that affect commercial microwave links (CMLs) as a whole. In the left column affected means that at least 10 min per hour are either filtered (*plateau filter*) or filled (*blackout gap filling*). In the right column, the percentage of CMLs affected by either fluctuation or high detection limit is shown. Note that this analysis relates to the processing line *Full* in which all the steps of quality control are conducted.

We observe anomalous data points and periods in several CMLs. For instance, a limited number of unreasonably low or high values leads to spikes in the time series. Moreover, the time series of some Czech CMLs show periods in which the baseline of received signal levels (RSL) drops to values far below the median and then stays at approximately (but not exactly) this level for several minutes, hours or even days before it leaps up again. We

refer to these patterns by *plateaus*. Furthermore, we encounter gaps in the time series of the RSL at presumably rainy periods, when the signal before and/or after the gap is significantly lower than the median of the whole time series. Those gaps are considered *blackouts*, that is, they are gaps caused by a failure of the receiver to process RSL values below a certain threshold in the case of heavy rainfall (Polz et al., 2023). Moreover, we observe short gaps in the time series, due to outages in the acquisition or other technical aspects. These are not considered blackouts if the RSL is not particularly low before or after the gaps.

In addition to the period-based observations, there are issues that affect CMLs as a whole. Most prominently, several CMLs show high fluctuations throughout the raw signal time series. These fluctuations may occur in daily or random patterns and are often clearly stronger and affect more time steps than the fluctuations induced by rainfall. Moreover, the Czech data include CMLs with very high detection limits. These CMLs are, by definition, not capable of measuring weak rainfall. Furthermore, they are less precise even if the rainfall exceeds the detection limit. A reason for this is the quantization of the recorded signal which does only allow a coarse estimation.

The above mentioned observations led to the definition of the following steps that are applied to improve data quality. Of those, the first four affect only single data points of the time series, while the latter two affect CMLs as a whole.

3.1.1. Step 1: Removing Specific Fill Values

Missing values are often given as numerical *fill* values, for which there is not a strict convention. We set signal levels to missing values if they have any of the following values: -99.9 , -99 , 255 , or approximately $1e37$. Of course, other fill values might occur in other data sets. Nevertheless, this step is defined in a specific way as each fill value might have a different unknown reason and meaning, which makes it useful to identify them and to address them directly.

3.1.2. Step 2: Filtering Plateaus

The *plateau filter* applies to data points that fulfill both of the following conditions: (a) the centered rolling maximum RSL of three data points is below -85 dB, and (b) the centered rolling standard deviation of RSL is smaller than 0.5 dB. Additionally, data points that are adjacent (next and second to next) to such plateaus are filtered. The threshold of -85 dB was chosen as the distribution of RSL exhibits a peak for lower values, which is not explicable by rainfall induced attenuation.

3.1.3. Step 3: Filling Blackout Gaps

Following the approach of Polz et al. (2023), we consider a period of missing values to be a blackout gap if the last RSL value before the gap or the first value after the gap is below -65 dB. In this step, these gaps are filled by the lowest RSL recorded by the CML over the whole month. Note that the maximum period that can be filled does not exceed 1 hr, i.e., at maximum half an hour after a gap starts and half an hour before the end of a gap. The gap is not filled at all if its length exceeds 1 hour.

3.1.4. Step 4: Filling 5-Min Gaps in the Time Series

The steps above depend on RSL. Step 4, in contrast, is based on the total loss, i.e., the difference between transmitted and RSL. If there are gaps in the total loss time series and if they do not exceed 5 min, we interpolate them linearly. If they exceed 5 min they will remain unaffected by this step.

3.1.5. Step 5: Filter Due To Fluctuations in the Time Series

This filter comprises two tests: (a) the 5 -hr rolling standard deviation of the total loss exceeds 2 dB at least 10% of the time; (b) the 1 -hr rolling standard deviation of the total loss exceeds 0.8 dB at least 33% of the time. All CMLs that fulfill at least one of these conditions are removed in this step.

3.1.6. Step 6: Filter by Detection Limit

The detection limit is defined as the minimum rainfall required to induce an observable change in the signal of a CML. It is calculated via the frequency and length of the CML, a fixed quantization of 0.33 dB, and the k - R relation with parameters defined by ITU-R (2005). Step 6 removes CMLs with a detection limit of at least 2 mm/hr. This threshold was chosen heuristically but based on the fact that a large proportion of the rainfall amount in the Central European climate can be attributed to rain rates below this value. Hence, CMLs that cannot sense such low intensity rainfall are neglected.

3.2. Rain Rate Retrieval and Spatial Interpolation

We calculate rain rates via a procedure of four steps based on the total loss time series of individual CMLs: (a) a classification of wet and dry periods based on the rolling standard deviation over 1 hr, (b) a subtraction of the baseline that is assumed to be constant during a rain event, (c) correction for signal loss induced by wet antennas based on the method proposed by Leijnse et al. (2008), (d) calculation of rainfall based on the k-R-relation with parameters defined by ITU-R (2005). The minutely rain rates are aggregated to hourly rainfall amounts which are then used to estimate rainfall maps via inverse distance weighting (IDW). For further details we refer to Graf et al. (2020), who explain the rain rate retrieval and interpolation in more detail. We follow their approach with the following two exceptions. First, we always classify the blackout gaps that have been filled according to Step 3 (see Section 3.1) as wet periods; second, we only consider neighbors in a maximum distance of 30 km in the IDW algorithm.

3.3. Analysis Setup

In the first part of the analyses (Section 4.1), we show the effect of all the steps considered for quality control. In Section 4.2, we quantify the quality of path-averaged rain rates in a comparison to RADOLAN-RW along those path using the performance indices mean absolute error (MAE), bias (BIAS), and Pearson correlation coefficient (PCC) defined as follows:

$$MAE = \mu(|R_{CML} - R_{RAD}|) \quad (1)$$

$$BIAS = \frac{\mu(R_{CML} - R_{RAD})}{\mu(R_{RAD})} \quad (2)$$

$$PCC = \frac{cov(R_{CML}, R_{RAD})}{\sigma(R_{CML})\sigma(R_{RAD})} \quad (3)$$

where $\mu(\cdot)$, $cov(\cdot)$, $\sigma(\cdot)$ are the mean, the covariance, and the standard deviation over time, respectively. R_{CML} and R_{RAD} are the CML and the RADOLAN-RW hourly rainfall amounts, respectively. Finally, in Section 4.3, the rainfall maps are evaluated qualitatively.

For the latter part of the analyses, we distinguish three processing lines which differ in the selection of steps used for quality control (see Section 3.1). We refer to these processing lines by the terms *No Filter*, *Graf 2020*, and *Full*. In the *No Filter* case, only the basic Step 1 is performed. In the *Graf 2020* case, additionally Steps 4 and 5 are performed. These steps have been adopted from Graf et al. (2020) and hence, this processing line represents a current standard approach of dealing with data quality adjusted to a purely German data set. In the processing line *Full*, all steps defined in Section 3.1 are performed. The steps are always conducted in the order used above. We analyze the effect of the different processing lines in Sections 4.2 and 4.3. In Section 4.1, we analyze all the mentioned steps conducted for quality control and do not distinguish between the processing lines.

4. Results

4.1. Effect of Quality Control Algorithms

Figure 2 showcases the effect of steps introduced in Section 3.1 and provides statistics on the amount of data that is affected. The first example shows that there would be extremely high rain rates toward the end of the shown period, if the *plateau filter* was not active. The second example shows how *blackout gap filling* can help to capture a rain event that otherwise would have been missed. In the third example, a strongly fluctuating CML yields rain events far too often and without correlation to the reference. The fourth example presents a CML with a high detection limit; although it captures most of the rain events the amount is generally far too high and even minor changes in RSL suggest strong rain. In the latter two examples, the respective CMLs are removed completely from the analysis when considering the *Full* processing line.

The lower part of Figure 2 shows that only a small amount of data is affected by the *plateau filter* and the *blackout gap filling*. Affected hours are defined as hours in which at least 10 min are labeled either as a *plateau* or as a *blackout gap*. Only for the class of data points that are associated with high reference rain rates and either very low or very high detection limits, the *plateau filter* affects a larger share. For the *blackout gap filling* there is a

clear positive correlation between the amount of affected hours and reference rain rates. Moreover, CMLs with lower detection limit are affected more often. Similarly, mostly the CMLs with low detection limits are affected by high *fluctuation*. By definition, the filter based on the detection limit affects only the class with the highest detection limit.

4.2. Path-Based Quantitative Analysis

We analyzed path-averaged rain rates of the CMLs in the border region for 1 month by comparison to RADOLAN-RW along the CML paths and analyzing the performance metrics. Figure 3 shows CML quantities and the performance metrics dependent on detection limit, the kind of quality control algorithms, and the country. The boxplots indicate the spread over the CML dimension.

The different range of detection limits of the two data sets can be seen in the upper row of Figure 3. A reduction of the number of CMLs can be observed in the figure, and explained by the filtering involved in the quality control steps. While this reduction affects the data sets of both countries, it is clearly more pronounced for the Czech data set. Starting from *No Filter* the additional filtering of *Graf 2020* affects CMLs of all classes of detection limits. The additional filters (*Full*) almost only affect the Czech CMLs and primarily the ones of high detection limits.

The performance metrics depend on the detection limit. All three metrics deteriorate toward the classes of high detection limits. This can be seen by worse median values and, in a more pronounced manner, by the worse mean values, and the marked skewness of the distributions, especially for the detection limit classes >0.5 mm/hr. The BIAS additionally shows a general increase with detection limit: While CMLs with very low detection limit (e.g., the ones of E-band frequency) tend to underestimate the rainfall amounts, the ones with high detection limits tend to overestimate. The effect of the detection limit can mainly be seen in the Czech data where each class contains a considerable number of CMLs. For Germany, the effect is less clear due to the small number of CMLs with detection limits above 1 mm/hr in the German data set. Nonetheless, outliers in MAE and BIAS are more prevalent for the detection limit class of 0.5–1 mm/hr compared to the class 0.1–0.5 mm/hr, also in the German data. Note that for readability not all outliers of MAE and BIAS are shown in Figure 3.

The performance metrics also depend on the three processing lines and the two countries. Considering the effect independent of the detection limit, that is, focusing on the shaded parts of Figure 3, the following observations can be made. A reduction of outliers with extended processing can be observed throughout. While the median of the MAE varies only little for the different processing lines, the number of outliers is clearly reduced by the enhanced processing. This can particularly be seen for the Czech data by the improved mean values: for example, the MAE of the Czech CMLs has the values 0.19, 0.19, 0.11 mm, for the *No Filter*, the *Graf 2020*, and the *Full* processing lines, respectively. A similar observation can be made for the BIAS, where the medians are very close to zero for all processing lines, but where the mean values of *No Filter* (0.46) and *Graf 2020* (0.46) are clearly higher than those of the processing line *Full* (0.02). Independent of the country, the PCC improves with increasingly effective processing both in terms of the median (from 0.89 to 0.91 to 0.93 for the German data, and from 0.84 over 0.84 to 0.86 for the Czech data) as well as by a reduction of the number of outliers and increasing mean values.

The effect of the processing lines can also be seen within individual classes of detection limit. Especially for the category of CMLs with detection limits in the range 0.5–1 mm/h, the extended processing affects the mean of MAE and BIAS strongly, while the effects on the quantiles depicted in the boxplots are small. This shows that the extended processing mainly reduces the number of outliers and thereby their influence on the metrics.

4.3. Rainfall Maps

Figure 4 shows rainfall maps for an event (21 June, 21:50 to 22 June, 4:50) that traverses the German-Czech border. The CML-derived maps capture the event well. That is, via all of the processing lines (first three rows in Figure 4) it is possible to generate maps that reproduce the overall pattern of the event.

Nevertheless, particularly for the *No Filter* processing line several shortcomings can be observed. For example, there are spots of overestimation. These appear most prominently in the cities and towns in the Czech Republic where the CML networks are dense (e.g., in Prague located within the red square in the upper left map of Figure 4, and Strakonice encapsulated by the orange square in the panel of the last time step). Moreover, there

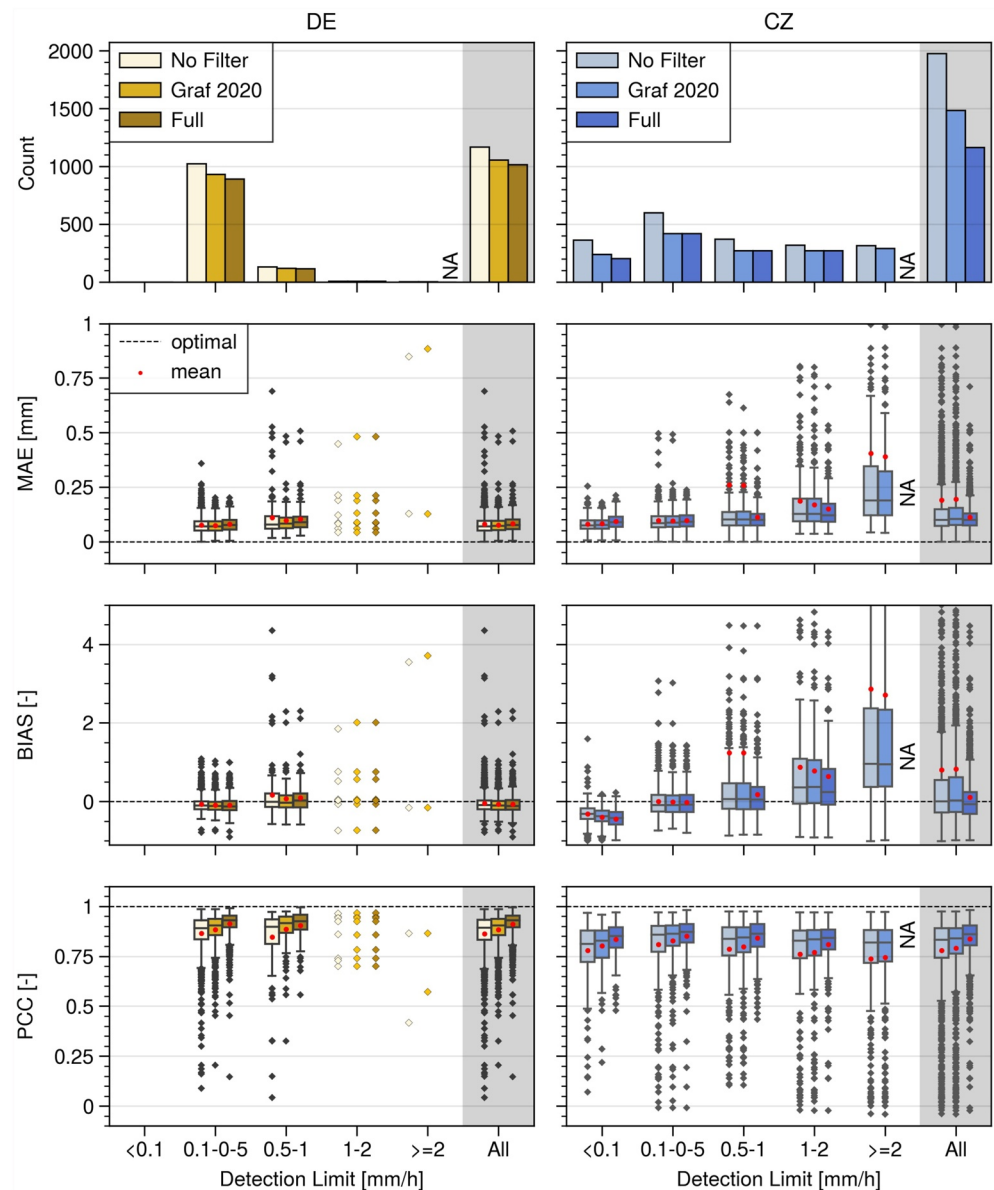


Figure 3. A path-based quantitative analysis for the whole month (June 2021). The commercial microwave links (CMLs) are categorized into classes of detection limit and three different processing lines are shown by different color intensities of bars and boxplots. By definition, CMLs are not available in the *Full* processing for the highest class of detection limits. The first row shows the amount of CMLs in each class. The latter three rows show the mean absolute error (MAE), the BIAS, and the Pearson correlation coefficient, respectively. The shaded part of the figures considers all CMLs independent of their detection limits. The left and right column consider the German and Czech CMLs, respectively. For Germany, two classes of detection limit contain very few CMLs and for those the metrics are shown as individual points for each CML instead of boxplots. Note that not all data points lie within the presented range of values for the MAE and the BIAS.

are white spots of underestimation within the rainfall field, particularly, at the time stamp 01:50 (highlighted by a purple square). Furthermore, for the first two time steps in which the rainfall is mostly located over Germany (region highlighted by magenta square), the high spatial variability as well as the high amounts observable in the reference is only weakly represented in the CML-derived maps.

Positive effects of extended quality control algorithms can be observed by comparing the different processing lines. The spots of overestimation in the Prague region are present in all time steps for the *No Filter* case, and also in the *Graf 2020* case, but not anymore when applying the *Full* processing. The local false rainfall in Strakonice is already removed via the *Graf 2020* processing. The extended processing also helps to reduce some of the white

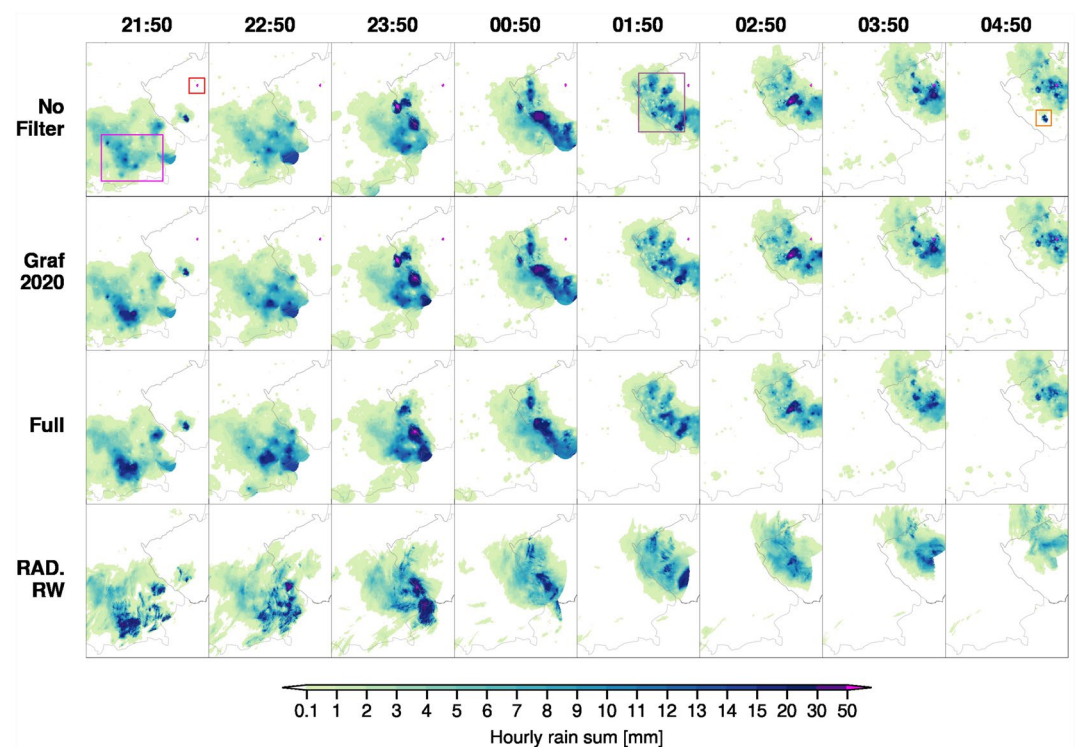


Figure 4. Maps of a rainfall event (21 June, 21:50 to 22 June, 4:50) (time progressing from left to right). The first three rows are interpolations based on commercial microwave links for the different processing lines. The bottom row is the reference RADOLAN-RW. A comparison of the *Full* processing and RADOLAN-RW in a movie sequence can be found in Blettner (2023).

spots that appear while the rain event is located over the westernmost part of the Czech Republic (e.g., time step 01:50), though several of these spots persist. The underestimation in the time steps 21:50 and 22:50 is reduced from the *No Filter* to the *Full* processing lines, albeit the representation of the spatial variability remains limited.

5. Discussion and Conclusion

We found that two individual CML data sets can be processed consistently with acceptable results even when applying algorithms that had been adjusted to only one of them and for a different period. However, while this holds for many CMLs and over most periods, it produces unrealistic rainfall amounts in some situations, which, despite their rarity can have strong influence on the maps.

Thereby, this study confirmed that it is crucial to deal with quality control when using CML data for rainfall estimation. Not only the frequency and the length distributions that determine the detection limit of the CMLs, but also unreliable periods or gaps in the time series of individual CMLs need to be considered. Some issues such as blackout gaps and CMLs with high fluctuations in the signal exist in both data sets. Others, like the periods we refer to as *plateaus* are only observable in the Czech data set. Global processing algorithms are required that address the individual characteristics but still allow a consistent treatment of all available data. The need to extend the set of algorithms developed for one data set when applied to a different independent data set, shows precisely the degree to which established routines are transferable, and where they are insufficient.

We applied and analyzed quality control algorithms which we partly adopted from Graf et al. (2020) and partly developed in this study. These algorithms involve filtering, that is, a reduction of the amount of data, which is generally not desirable. However, filtering is less problematic for generating rainfall maps if the sensor density is high in relation to the resolution of the map. In this study, the majority of filtered CMLs is in the Czech Republic and often in the cities where the network is dense enough so that the loss of several devices with questionable observations is justifiable.

Note that we describe and analyze a set of steps to improve data quality, which we do not necessarily consider optimal and exhaustive, but which help to deal with the major challenges we encountered. There is room to improve the algorithms and to add additional ones. For example, the classification of wet and dry events may be improved by including a spatial consistency check similar to what was done by Overeem et al. (2016a), instead of basing it purely on the time series of individual CMLs.

In this study, we show the feasibility of combining heterogeneous CML data sets to generate transboundary rainfall maps. Furthermore, we present straightforward algorithms that can help to deal with issues of data quality. Thereby, we made another step toward minimizing erroneous data points, which need to be expected, given that CMLs are opportunistic rainfall sensors. The persistence of some minor issues with data quality does not prevent the generation of reasonable transboundary maps. This shows the potential of CMLs providing a basis for even larger-scale (e.g., continental) rainfall maps, which is a challenge even for dedicated sensors.

Data Availability Statement

CML data were provided by Ericsson Germany and T-Mobile Czech Republic. For the processing we use the *pycomlink* software (Chwala et al., 2021) available under BSD-3-Clause License at <https://doi.org/10.5281/zenodo.4810169>. The *pycomlink* repository contains an exemplary CML data set which enables reproducing the general processing. Additional new processing steps that have been applied in this study will be added to *pycomlink* soon. They can be reproduced by the code and an exemplary data set which we provide at (Blettner, 2023), <https://doi.org/10.5281/zenodo.7973736>. Also the data shown in the maps is available under the aforementioned link. The full CML data sets, however, are not publicly available because of the network providers' interest in safeguarding the coordinates of site locations. An individual agreement with the network provider needs to be established to obtain the CML data for research purposes. RADOLAN-RW data (Bartels et al., 2004) are available from the Climate Data Center of the German Weather Service at https://opendata.dwd.de/climate_environment/CDC.

Acknowledgments

The authors thank Ericsson Germany and T-Mobile Czech Republic for the access to CML data, and acknowledge the open access to the RADOLAN-RW data archive from the German Weather Service. This work is funded by the German Research Foundation and the Czech Science Foundation via the SpratLINK project (Grants CH-1785/2-1, 20-14151J). Open Access funding is enabled and organized by Projekt DEAL.

References

- Bartels, H., Weigl, E., Reich, D. T., Lang, P., Wagner, A., Kohler, O., & Gerlach, N. (2004). Routineverfahren zur Online-Aneichung der Radarniederschlagsdaten mit Hilfe von automatischen Bodenniederschlagsstationen (Ombrometer).
- Blettner, N. (2023). nblettner/transboundary_cml_rainfall_maps: V0.0.1. *Zenodo*. <https://doi.org/10.5281/zenodo.7973736>
- Chwala, C., & Kunstmann, H. (2019). Commercial microwave link networks for rainfall observation: Assessment of the current status and future challenges. *WIREs Water*, 6(2), e1337. <https://doi.org/10.1002/wat2.1337>
- Chwala, C., Polz, J., Graf, M., Sereb, D., Blettner, N., Keis, F., & Boose, Y. (2021). *pycomlink/pycomlink*: V0.3.2. *Zenodo*. <https://doi.org/10.5281/zenodo.4810169>
- D'Amico, M., Manzoni, A., & Solazzi, G. L. (2016). Use of operational microwave link measurements for the tomographic reconstruction of 2-D maps of accumulated rainfall. *IEEE Geoscience and Remote Sensing Letters*, 13(12), 1827–1831. <https://doi.org/10.1109/LGRS.2016.2614326>
- Graf, M., Chwala, C., Polz, J., & Kunstmann, H. (2020). Rainfall estimation from a German-wide commercial microwave link network: Optimized processing and validation for 1 year of data. *Hydrology and Earth System Sciences*, 24(6), 2931–2950. <https://doi.org/10.5194/hess-24-2931-2020>
- Huuskonen, A., Saltikoff, E., & Holleman, I. (2014). The operational weather radar network in Europe. *Bulletin of the American Meteorological Society*, 95(6), 897–907. <https://doi.org/10.1175/bams-d-12-00216.1>
- ITU-R. (2005). Specific attenuation model for rain for use in prediction methods (recommendation) (p. 838-3). Retrieved from August 14, 2023, https://www.itu.int/dms_pubrec/itu-r/rec/p/R-REC-P.838-3-200503-I!!PDF-E.pdf
- Leijnse, H., Uijlenhoet, R., & Stricker, J. (2008). Microwave link rainfall estimation: Effects of link length and frequency, temporal sampling, power resolution, and wet antenna attenuation. *Advances in Water Resources*, 31(11), 1481–1493. <https://doi.org/10.1016/j.advwatres.2008.03.004>
- Leijnse, H., Uijlenhoet, R., & Stricker, J. N. M. (2007). Rainfall measurement using radio links from cellular communication networks. *Water Resources Research*, 43(3), W03201. <https://doi.org/10.1029/2006WR005631>
- Messer, H., Zinevich, A., & Alpert, P. (2006). Environmental monitoring by wireless communication networks. *Science*, 312(5774), 713. <https://doi.org/10.1126/science.1120034>
- Overeem, A., Leijnse, H., & Uijlenhoet, R. (2013). Country-wide rainfall maps from cellular communication networks. *Proceedings of the National Academy of Sciences*, 110(8), 2741–2745. <https://doi.org/10.1073/pnas.1217961110>
- Overeem, A., Leijnse, H., & Uijlenhoet, R. (2016a). Retrieval algorithm for rainfall mapping from microwave links in a cellular communication network. *Atmospheric Measurement Techniques*, 9(5), 2425–2444. <https://doi.org/10.5194/amt-9-2425-2016>
- Overeem, A., Leijnse, H., & Uijlenhoet, R. (2016b). Two and a half years of country-wide rainfall maps using radio links from commercial cellular telecommunication networks. *Water Resources Research*, 52(10), 8039–8065. <https://doi.org/10.1002/2016WR019412>
- Polz, J., Graf, M., & Chwala, C. (2023). Missing rainfall extremes in commercial microwave link data due to total loss of signal. *Earth and Space Science*, 10(2), e2022EA002456. <https://doi.org/10.1029/2022EA002456>
- Roversi, G., Alberoni, P. P., Fornasiero, A., & Porcù, F. (2020). Commercial microwave links as a tool for operational rainfall monitoring in Northern Italy. *Atmospheric Measurement Techniques*, 13(11), 5779–5797. <https://doi.org/10.5194/amt-13-5779-2020>
- Uijlenhoet, R., Overeem, A., & Leijnse, H. (2018). Opportunistic remote sensing of rainfall using microwave links from cellular communication networks. *WIREs Water*, 5(4), e1289. <https://doi.org/10.1002/wat2.1289>
- Zinevich, A., Alpert, P., & Messer, H. (2008). Estimation of rainfall fields using commercial microwave communication networks of variable density. *Advances in Water Resources*, 31(11), 1470–1480. <https://doi.org/10.1016/j.advwatres.2008.03.003>