Improving weather radar rainfall estimates by merging with commercial microwave link data: a fully reproducible, large-scale method intercomparison

Erlend Øydvin[1], Elia Covi[2], Maximilian Graf[3], and Christian Chwala[4]

[1]Faculty of Science and Technology, Norwegian University of Life Sciences, Ås, Norway
[2]Hydro-Meteo-Climate Structure, Arpae Emilia-Romagna, Bologna, Italy
[3]Department of Hydrometeorology, Deutscher Wetterdienst, Offenbach, Germany
[4]Institute of Meteorology and Climate Research (IMK-IFU), Karlsruhe Institute of Technology, Garmisch-Partenkirchen, Germany

Correspondence: Christian Chwala (christian.chwala@kit.edu)

Abstract.

Accurate rainfall estimation is essential for hydrometeorological applications, but capturing the fine spatiotemporal variability of rainfall remains challenging. In this study, we assess the impact of merging commercial microwave link (CML) data with weather radar for quantitative precipitation estimation (QPE) using two openly available datasets with contrasting observational densities. We compare multiple merging methods, including kriging with external drift (KED), and derive a block kriging interpolation method to account for the line-average nature of CMLs. The results show that merging CML data improves radar QPE, with reductions in mean absolute error (MAE) of up to 41% for multiplicative merging and up to 38% ~~on average~~ for KED, aligning well with similar studies using rain gauges. ~~However, the performance of merging methods varies with~~ Performance depends on rainfall intensity, distance~~to observations~~, and network density~~. In terms of Pearson correlation coefficient (PCC), additive methods outperform KED in data-dense networksand during extreme rainfall events, while in data-sparse regions, KED provides more consistent adjustments, particularly at medium ranges (up to 15 km) . At greater distances,~~ : in sparse networks, multiplicative methods yield the lowest RMSE and MAE, though KED maintains higher spatial correlation (PCC) over distance. In dense networks, KED and additive methods perform comparably. Overall, additive methods ~~again perform better by preserving radar variability. For RMSE and MAE, however, KED consistently outperforms additive methods across all settings~~perform best at capturing extremes. Despite accounting for line-averaging, block kriging performed similarly to point-based kriging. All merging methods reduced bias, RMSE, and MAE compared to unadjusted radar fields. The merging framework and intercomparison study are openly available, enabling reproducibility and further exploration by the scientific community.

1 Introduction

Accurate rainfall estimation is important for many hydrometeorlogical applications like flood forecasting, urban drainage design, or water-resources management. However, rainfall is highly intermitted both in space and time and capturing this variabil-

ity remains one of the main challenges in hydrometeorology, particularly during convective events. Even dense observational networks consisting of rain gauges and weather radars struggle to correctly resolve the often fine grained spatio-temporal variability of rainfall (Cristiano et al., 2017; Rombeek et al., 2025; Schleiss et al., 2020).

Rain gauges provide relatively precise point measurements and are considered as the traditional reference for quantitative precipitation estimation (QPE). Yet, their spatial density is typically insufficient to represent the full variability of rainfall at kilometer or sub-kilometer scales (O and Foelsche, 2019; Hänsler and Weiler, 2022). This limitation persists even in well-instrumented regions e.g. with one rain gauge for roughly every 200-400 km^2 in Central Europe, and it becomes critical in areas where the observational infrastructure is sparse. Consequently, radar observations have become more widely used for providing high-resolution areal rainfall information due to high resolution. Despite these advantages, weather radar rainfall estimates remain affected by multiple error sources, including attenuation, beam blockage, vertical profile effects, calibration drifts, and the presence of the melting layer which are partially but not completely resolved by the use of dual-pol radars (Aldana et al., 2025; Schleiss et al., 2020; Hänsler and Weiler, 2022). These uncertainties limit the accuracy of radar-only QPE and motivate the merging of radar data with ground-based observational data.

Commercial microwave links (CMLs) from the opportunistic use of attenuation data from cellular communication networks have emerged over the past decade as a robust source of rainfall information. Numerous studies have shown that path-averaged rain-induced attenuation along CMLs can be converted into accurate QPE, in particular for moderate and heavy precipitation (Chwala and Kunstmann, 2019; Uijlenhoet et al., 2018; Messer and Sendik, 2015). In Europe, CMLs are increasingly regarded as a complementary data source for rainfall (Olsson et al., 2025), while in data-sparse regions such as parts of Africa they can serve as a quasi-independent measurement system (Blettner et al., 2025; Djibo et al., 2023; Overeem et al., 2021). One particularity of CML data is that, unlike weather radar or rain gauge observations, it is not owned by national or local authorities. As a consequence, CML data is generally closed to individual research groups that obtained it and it does e.g. in the EU, not fall under open-data policies. Only recently have the first larger CML datasets become publicly available (Andersson et al., 2022; Covi and Roversi, 2025; Overeem et al., 2024c).

Merging has been extensively studied in the literature (Jewell and Gaussiat, 2015; Delrieu et al., 2014; Overeem et al., 2024b), and there are many methods for doing this. For instance, Goudenhoofdt and Delobbe (2009) merged weather radar with with a dense rain gauge network in Belgium. They found that merging methods that allowed the radar correction to vary spatially, performed better than applying a mean correction to the entire rainfall field. This finding was supported by Nielsen et al. (2024), who merged radar with personal weather station (PWS) and CML data, and found that a localized median correction dampened the effect of outliers. Sideris et al. (2014) added a temporal component to the merging process, allowing the previous timestep to influence the current time step. They found that the proposed method offered advantages over existing methods, particularly in areas where sparse rain-gauge coverage, measurement errors, or high variability compromise the stability of traditional geostatistical methods.

Although Nielsen et al. (2024) utilized CMLs in their study, these were mixed with many more PWSs, making it hard to assess the impact of CMLs on the weather radar data. Moreover, their merging method used the midpoint of the CMLs to represent its location, which is not optimal as the CMLs represent a path average.

In this work we assess the impact of merging data from commercial microwave links with weather radar data using two existing large openly available datasets. We also investigate different merging and interpolation methods from the literature, including ~~methods that can take~~ kriging with external drift (KED) and block kriging, which accounts for the line geometry of the CMLs ~~into account~~and provides uncertainty estimates as a byproduct of the interpolation process, aiming to better understand the strength and weaknesses of the different approaches. The merging methods and intercomparison study are implemented in `mergeplg` (Øydvin et al., 2025a) and made openly available and fully reproducible ~~(Covi et al., 2025; Øydvin et al., 2025a)~~ (Covi et al., 2025), allowing the scientific community to experiment and contribute. We seek to answer the following questions: How much can CML rainfall retrievals improve radar QPEs, and how do adjustment methods and network configurations influence this improvement?

2 Datasets

We use two open CML datasets that are accompanied by radar and rain gauge data as reference. The first is the OpenMRG dataset covering the Gothenburg region, Sweden, from June to August 2015 (Andersson et al., 2022). The datasets includes measurements from 364 CMLs, spanning about 2800 km2 (1 CML per 7.7 km^2), with the majority of the CMLs located in the city center (Fig. 1a). Each CML comprise two sublinks. Signal levels - both transmitted (TSL) and received (RSL) - were recorded every 10 seconds and resampled to 1-minute intervals by retaining the first value in each period. The signal levels have a quantization of approximately 0.1 dB. The CMLs are operated across a frequency range of 7 to 38 GHz, with the majority exceeding 25 GHz as shown in Fig. 1c). All variables exhibit high data availability, with 97.3% of CMLs, 99.1% of radar measurements, and 100% of rain gauge data present. The reference data consists of one gauge from SMHI and ten municipality rain gauges with 15 minute resolution, as well as a radar reflectivity composite from the three nearest SMHI weather radar with a 2 km and 5 minute resolution. The reflectivity values are transformed to rainfall values using the standard Marshall-Palmer Z-R relation (Marshall and Palmer, 1948).

The second dataset is OpenRainER, covering the Emilia-Romagna region in Italy for the years 2021 and 2022 (Covi and Roversi, 2025; Covi et al., 2026). We use three months of data from June to August 2022. The dataset includes measurements from 153 CMLs, spanning about 21000 km^2 (1 CML per 137 km^2), with most CMLs located along the mountainous region in the south (Fig. 1b). TSL and RSL were recorded instantaneously with a resolution of 1 minute. The signal levels appear to have an underlying quantization of approximately 1 dB, with intermediate values in the data likely arising from temporal averaging. Frequencies range between 24.6 to 25.6 GHz (Fig. 1d). Compared to OpenMRG, the data availability is slightly lower, with 85 % of CMLs, and 80-100 % of radar measurements, and 88 % of rain gauge data present. The reference dataset consists of two components: 280 rain gauges distributed evenly across the study area and radar reflectivity data from a composite of two weather radars. The radar data is available in three formats: 5-minute radar reflectivity values, 15-minute accumulated rainfall maps derived using the standard Marshall-Palmer Z-R relationship (Marshall and Palmer, 1948), and gauge-adjusted radar data (Covi et al., 2026). For this study, we utilized the 15-minute accumulated rainfall maps derived from the Marshall-Palmer Z-R relationship, without applying gauge adjustment.

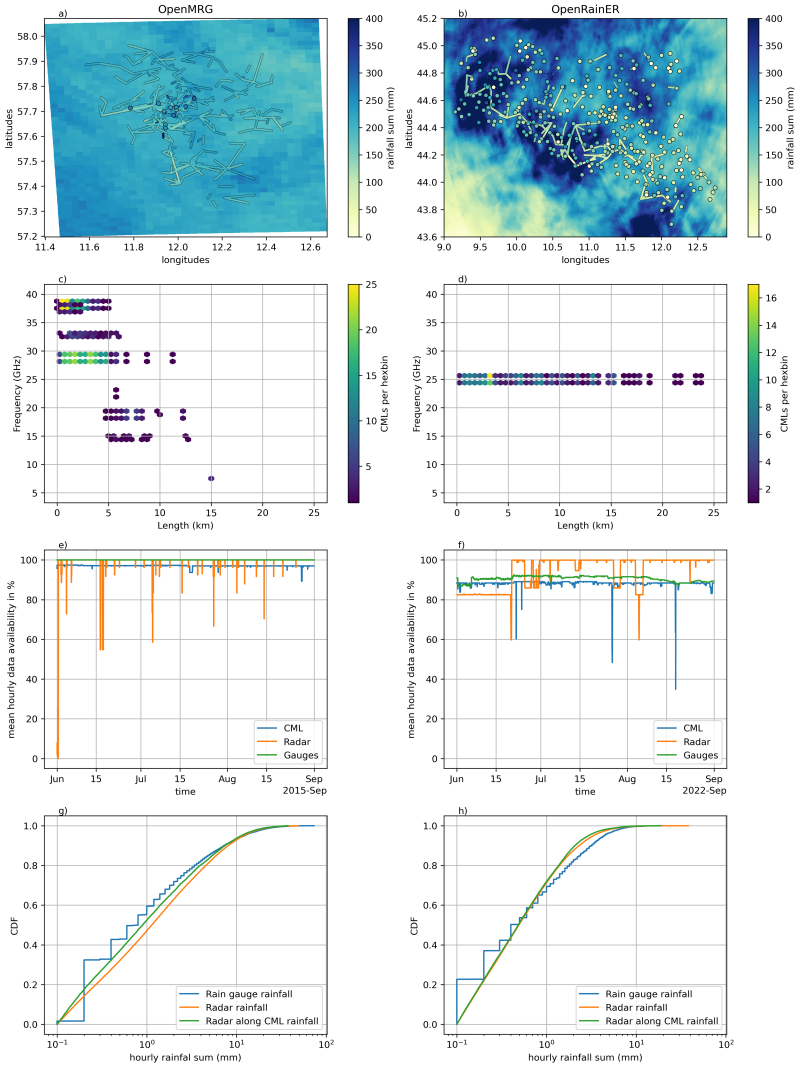


Figure 1. Maps and descriptive information on the OpenMRG (left column) OpenRainER (right column) dataset. a) and b) show the rainfall sum of the weather radar and the location of rain gauges and CMLs with their rainfall sum over the respective three month period. c) and d) show the distribution of CMLs over length and frequency. e) and f) show the mean hourly data availability. g) and h) show the cumulative distribution function over hourly rainfall sums.

The large dataset from the Netherlands (Overeem et al., 2024c) was not used, as its 15-minute min-max sampling would require different processing software (Overeem et al., 2024a), complicating the workflow. Furthermore, the two CML datasets used in this study have been extensively optimized within OpenSense, ensuring high confidence in their quality for CML QPE. Unlike the datasets used here, the Dutch dataset also lacks accompanying radar and gauge data, making its integration more complex.

The CML data was processed using a workflow following five steps: 1) the total signal loss (TL) is calculated by subtracting RSL from the TSL, 2) TL underwent quality control checks, 3) wet-dry classification is done using weather radar data, 4) baseline and wet antenna attenuation (WAA) is estimated and subtracted from the TL, 5) the attenuation caused by rainfall is converted to rainfall rates (Chwala and Kunstmann, 2019). For quality control, we compared the channels for each link and removed the intervals in which the differences between attenuation levels were greater than ~~30.~~ 30 dB. In addition, CMLs ~~with~~ were completely removed if they showed a strong diurnal cycle (5 h rolling standard deviation exceeding 2.0 dB for more than 10% of the time) or excessively noisy signals ~~, without fluctuations, or showing a strong diurnal cycle were completely removed~~(1 h rolling standard deviation exceeding 0.8 dB for more than 35% of the time), following Graf et al. (2020), or showed no fluctuations (1 h rolling standard deviation below 0.8 dB for more than 99% of the time). For the wet-dry classification using weather radar data, any radar timestep with more than 0.01 mm of rainfall per 15 minutes was marked as wet. To account for temporal variability, a 5-minute extension was added on each side of these wet periods. The baseline was estimated using the mean of the 5 minutes TL before any wet period. WAA was estimated using the KR-alt model from Pastorek et al. (2022), with $A_{max} = 6$, $\zeta = 0.7$ and $d = 0.15$. The attenuation caused by rainfall was converted to rainfall rates using the k-R relation defined in ITU (2005). All processing steps were implemented using pycomlink (Chwala et al., 2025a).

3 Merging methods

To improve radar rainfall maps using CML observations, various merging methods have been implemented. These methods rely on paired radar ($R(v)$) and CML ($\cancel{C(v)}R_{CML}(v)$) rainfall estimates $[\mathrm{mm/h}]$ at location v. The radar rainfall $R(v)$ is defined as the path-averaged radar rainfall along the CML, computed using the intersection weights between the CML path and the radar grid. This definition is the same for all merging methods. All methods use a local interpolation approach with a neighborhood of ~~12~~ 70 radar-CML pairs. The ~~12~~ choice of neighborhood size was based on a sensitivity analysis (see Fig. B1), which indicate that the results are not very sensitive to neighborhood size. The 70 nearest CMLs to each grid cell are identified based on the distance between the grid cell and the CML midpoints.

The merging process is applied independently for each time step, with both radar and CML data aggregated to an hourly resolution. Following the approach of previous studies (Jewell and Gaussiat, 2015; Goudenhoofdt and Delobbe, 2009; Velasco-Forero et al., 2009), the rainfall data is used in their original form rather than transforming them into Gaussian distribution. Table 1 shows an overview of the various merging methods used in the study.

3.1 Additive and multiplicative merging

Additive and multiplicative merging are methods for combining radar ($R(v)$) and CML ($\cancel{C(v)}R_{CML}(v)$) rainfall estimates. These methods assume that radar and CML data provide complementary information and that their differences or ratios can be interpolated to improve rainfall estimates. The ~~field being interpolated is denoted as $Z(v)$~~observed adjustment values at CML locations are denoted as $D(v)$ for additive and $F(v)$ for multiplicative merging, where:

- For additive merging: ~~$Z(v) = C(v) - R(v)$~~ $D(v) = R_{CML}(v) - R(v)$ (difference field),

– For multiplicative merging: ~~$Z(v) = C(v)/R(v)$~~ $F(v) = \ln(R_{CML}(v)/R(v) + \epsilon)$ (ratio field).

The small constant ϵ (e.g. 0.1) is added to the ratio to avoid issues with zero radar observations.

Range checks are applied to ensure reliable interpolation by excluding radar-CML pairs with large deviations. Unless otherwise stated, the following checks are used as the default:

– For Additive merging: ~~$|C(v) - R(v)| \leq 10$~~ $|R_{CML}(v) - R(v)| \leq 10$ mm/h,

– For Multiplicative merging: ~~$0.1 \leq C(v)/R(v) \leq 15$~~ $0.1 \leq R_{CML}(v)/R(v) \leq 15$ (dimensionless; thresholds are asymmetric by convention).

Additionally, merged rainfall estimates exceeding 150 mm/h are capped to that value to prevent physically unrealistic output from propagating into the merged fields.

In additive merging, the interpolated field ~~$Z(v)$~~ $\hat{D}(v)$ is added back to the radar rainfall to produce the merged rainfall field:

$$R_{\text{merged}}(v) = R(v) + Z\hat{D}(v). \tag{1}$$

In multiplicative merging, the interpolated field ~~$Z(v)$~~ $\hat{F}(v)$ is applied as a multiplicative factor to the radar rainfall:

$$R_{\text{merged}}(v) = exp(R(v) \cdot Z\hat{F}(v)) - \epsilon. \tag{2}$$

3.2 Interpolation techniques

In both additive and multiplicative merging, the field $Z(v)$ (~~difference field or ratio field~~either $D(v)$ or $F(v)$) is interpolated to produce an estimate $\hat{Z}(v)$ at unobserved locations, using one of the following techniques:

Inverse Distance Weighting (IDW): IDW estimates ~~$Z(v_0)$~~ $\hat{Z}(v_0)$ at an unmeasured location v_0 as a weighted average of nearby observations $Z(v_i)$:

$$Z\hat{Z}(v_0) = \frac{\sum_{i=1}^{N} w_i Z(v_i)}{\sum_{i=1}^{N} w_i}, \quad w_i = \frac{1}{|v_i - v_0|^{\beta}}, \tag{3}$$

where w_i are the weights based on the inverse distance $|v_i - v_0|$ between the observation location v_i and the target location v_0, and β controls how strongly the weights decay with distance. In this study we set β equal to 2. IDW is computationally efficient and easy to implement, but it does not account for spatial correlations.

Ordinary Kriging (OK): OK estimates ~~$Z(v_0)$~~ $\hat{Z}(v_0)$ at a target location v_0 as a weighted sum of nearby observations $Z(v_i)$:

$$\hat{Z}(v_0) = \sum_{i=1}^{N} \lambda_i Z(v_i), \tag{4}$$

The weights λ_i are determined by solving a constrained minimization problem that seeks the best linear unbiased estimator (BLUE) of $Z(v_0)$. Specifically, the estimation variance

$$\sigma^2 = E\left[\left(Z(v_0) - \hat{Z}(v_0)\right)^2\right] \tag{5}$$

is minimized subject to the unbiasedness constraint $\sum_{i=1}^{N} \lambda_i = 1$. This yields the Kriging system of equations:

$$\sum_{j=1}^{N} \lambda_j C(d_{ij}) + \mu = C(d_{i0}), \quad \sum_{i=1}^{N} \lambda_i = 1. \tag{6}$$

Here, $C(d_{ij})$ is the covariance between locations v_i and v_j, $C(d_{i0})$ is the covariance between v_i and the target location v_0, and μ is the Lagrange multiplier ensuring that the weights sum to 1. ~~The covariance function $C(d)$ is related to the variogram $\gamma(d)$ by $C(d) = c - \gamma(d)$, where c is the sill parameter (Cecinati et al., 2018).~~

~~Our analysis indicates that the interpolation results are not particularly sensitive to variations in the variogram parameters, a finding that aligns with previous studies (Haberlandt, 2007; Goudenhoofdt and Delobbe, 2009). Consequently, we adopt a spherical variogram with a sill of 1 $(\mathrm{mm/h})^2$ and a range of 30 km, as suggested by van de Beek et al. (2012). In addition we used a nugget value of 0.3 $(\mathrm{mm/h})^2$, which gave better results when observations were very dense.~~

3.3 Kriging with External Drift (KED)

KED extends OK by incorporating radar rainfall $R(v)$ as an external drift to guide the interpolation~~of $Z(v)$. Unlike OK, which assumes a constant mean for the rainfall field, KED allows~~. Unlike the additive and multiplicative methods, KED interpolates the CML rainfall estimates $R_{CML}(v)$ directly, rather than a difference or ratio field. The radar field serves as a drift term, allowing the mean to vary spatially based on radar data. ~~This makes KED especially suited for merging radar and CML data in regions with large-scale spatial trends.~~

KED builds on the OK system of equations (Equation 6) with the addition of the drift constraint:

$$\sum_{i=1}^{N} \lambda_i R(v_i) = R(v_0), \tag{7}$$

where $R(v_i)$ is the radar rainfall at observation location v_i, $R(v_0)$ is the radar rainfall at the target location v_0, and λ_i are the kriging weights. The weights λ_i are determined by solving an augmented system of equations that accounts for both the variogram-based spatial correlation and the external drift (Jewell and Gaussiat, 2015; Sideris et al., 2014).

~~As with OK, we use a spherical variogram with a sill of 1 $(\mathrm{mm/h})^2$, a range of 30 km, and a nugget of 0.3 $(\mathrm{mm/h})^2$. Unless otherwise stated, the same range check as in additive merging is applied to exclude radar-CML pairs where $|C(v) - R(v)| > 10$.~~

3.4 Block Kriging

The methods described earlier assume point-based observations, which may not fully represent the spatial extent of line-averaged data such as CMLs. ~~To address~~ In this context, a block refers to the path of a CML link, a line segment over which the rainfall measurement is spatially averaged. To account for this, the kriging system is adapted to ~~estimate the mean over a spatial block~~ treat each CML as a block observation rather than a point. Block kriging minimizes the estimation variance of the block mean under the constraint of unbiasedness. Following the formulation in Journel and Huijbregts (1978, p. 306), the

Table 1. Overview of merging methods and interpolation approaches.

Merging Method	Interpolation Approach	Point (P)	Block (B)
Additive	Inverse Distance Weighting (IDW)	`add_p_idw`	-
	Ordinary Kriging (OK)	`add_p_ok`	`add_b_ok`
Multiplicative	Inverse Distance Weighting (IDW)	`mul_p_idw`	-
	Ordinary Kriging (OK)	`mul_p_ok`	`mul_b_ok`
KED	Kriging with External Drift (KED)	`ked_p`	`ked_b`

kriging system becomes:

$$\sum_{j=1}^{N} \lambda_j C_b(v_i, v_j) + \mu = C_b(v_i, v_0), \quad \sum_{j=1}^{N} \lambda_j = 1, \tag{8}$$

where $C_b(v_i, v_j)$ denotes the covariance between block-averaged values. To correct for the smoothing effect of spatial averaging, this covariance is computed from the underlying point covariance $C(h)$ as:

$$C_b(v_i, v_j) = C_b(0) + \bar{C}(v_i, v_j) - \frac{1}{2}\left[\bar{C}(v_i, v_i) + \bar{C}(v_j, v_j)\right], \tag{9}$$

Here, $\bar{C}(v_i, v_j)$ denotes the average point covariance between blocks v_i and v_j, while $\bar{C}(v_i, v_i)$ and $\bar{C}(v_j, v_j)$ represent the average point covariance within each block (see Appendix A for explicit definitions). The term $C_b(0)$ corresponds to the sill of the block-averaged process. ~~As with OK, we use a spherical variogram with a sill of 1 $(\mathrm{mm/h})^2$, a range of 30 km, and a nugget of 0.3 $(\mathrm{mm/h})^2$.~~

A more detailed justification for Equation 9 is given in Appendix A. Graf et al. (2021) and Goovaerts (2008) derive similar covariance expressions~~by estimating C_b directly from block-averaged data. In contrast, the approach developed here assumes~~. For all kriging-based merging methods, the spatial structure of the data must be characterized by a theoretical variogram. While Goovaerts (2008) developed a deconvolution method for estimating the point variogram ~~is known and analytically derives block covariance terms, explicitly accounting for block geometry and the smoothing effect of spatial averaging~~from block averaged data, Graf et al. (2021) used rain gauges to estimate the point variogram. In the next section, we describe a data-driven and a fixed approach for determining these parameters.

3.5 Variogram estimation

To address the uncertainty in variogram selection, we implement two distinct strategies for determining the parameters of a spherical variogram model (Haberlandt, 2007):

$$\gamma(d) = \begin{cases} c_0 + c_p \left[\frac{3}{2} \left(\frac{d}{h_r} \right) - \frac{1}{2} \left(\frac{d}{h_r} \right)^3 \right] & h \leq a \\ c_0 + c_p & h > a \end{cases} \tag{10}$$

where $\gamma(h)$ is the semivariance at lag distance h, c_0 is the nugget effect representing micro-scale variability and measurement error, c_p is the partial sill representing spatially correlated variance, and h_r is the range, beyond which observations are no longer spatially correlated. The absolute sill c is related to the nugget and partial sill through $c = c_0 + c_p$ and the covariance function $C(d)$ is related to the variogram $\gamma(d)$ by $C(d) = c - \gamma(d)$.

3.5.1 Reference variogram

First, we use a fixed spherical variogram as a reference baseline. van de Beek et al. (2012) show that variogram parameters for 1-hour Dutch summer rainfall have a range of approximately 30 km, though they also demonstrate that parameters vary substantially in time with rainfall characteristics. Since both datasets cover summer rainfall, we adopt the summer range as the fixed value $h_r = 30$ km, and retain it as a reference baseline against which data-estimated variograms are compared. The sill is set to $c = 1\ (\mathrm{mm/h})^2$, consistent with the variance-standardized scale used throughout this study. The nugget $c_0 = 0.3\ (\mathrm{mm/h})^2$ is chosen to be broadly consistent with the variogram estimates from our data (Fig. 3). Unless stated otherwise, the same parameters are used for all merging methods, as the variogram study shows no substantial differences between the difference and ratio fields (Fig. 3).

3.5.2 Variogram estimated from data

Second, we estimate variogram parameters from the data. Since the datasets span several months, we adapt the event-averaging approach of Haberlandt (2007) to a continuous time series.

Time steps are grouped into rainfall episodes: a time step is included in an episode if the radar exceeds 0.01 mm/h at any CML location, or is within 6 hours of a time step that does. For each episode, observations from all time steps are pooled. Following Haberlandt (2007), each time step's contribution is standardized by its variance s_t^2 before combining, giving the average experimental variogram:

$$\hat{\gamma}(d) = \frac{1}{2 \cdot n(d) \cdot s^2} \sum_{i=1}^{n(d)} [z(x_i) - z(x_i + d)]^2 \tag{11}$$

where $n(d)$ is the number of observation pairs at lag distance d across all time steps in the episode. A minimum of 100 pairs is required; if an episode contains fewer, neighbouring episodes are merged until the threshold is reached. The theoretical model is fitted by minimizing the weighted sum of squares between the experimental and theoretical variogram values, using only lag distances up to the 95th percentile of observed pair distances. This avoids the influence of poorly sampled long-range

lags, where few observation pairs cause high variability in the experimental variogram. Assuming $s_i^2 = c_0 + c$ reduces the estimation to two free parameters: the nugget c_0 and the range h_r.

This yields one set of variogram parameters per episode. To obtain a stable estimate that is representative for a longer period, the episode parameters are averaged over 14-day windows and applied to all time steps within that window. This also lets `mergeplg` reuse interpolation weights within each 14-period, reducing the computational time.

Unless stated otherwise, variograms for additive and KED-based methods were estimated from the radar–CML differences, and for multiplicative merging from the radar-to-CML ratios.

3.6 Evaluation and metrics

Each of the eight merging methods (Table 1) produces an adjusted rainfall field based on CML data at an hourly resolution. These adjusted fields, along with the original radar rainfall field, are evaluated by extracting rainfall values at each rain gauge location R_{g_i} from the nearest pixel in the rainfall field R_{f_i}, and computing metrics using `poligrain` (Chwala et al., 2024).

The root mean squared error (RMSE) evaluates the average magnitude of errors between observed (R_{g_i}) and estimated (R_{f_i}) rainfall, with larger errors penalized more heavily due to squaring:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(R_{g_i} - R_{f_i})^2} \tag{12}$$

The mean absolute error (MAE) provides the average magnitude of errors, treating all differences equally without emphasizing larger errors:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|R_{g_i} - R_{f_i}| \tag{13}$$

The percent bias (BIAS) quantifies the average bias as a percentage of the total observed rainfall, indicating whether the rainfall field systematically over- or underestimates rainfall:

$$BIAS = \frac{\sum_{i=1}^{N}(R_{g_i} - R_{f_i})}{\sum_{i=1}^{N} R_{g_i}} \times 100 \tag{14}$$

The Pearson correlation coefficient (PCC) measures the strength and direction of the linear relationship between observed and estimated rainfall. Values range from -1 (perfect negative correlation) to 1 (perfect positive correlation), with 0 indicating no correlation:

$$PCC = \frac{\sum_{i=1}^{N}(R_{g_i} - \bar{R}_g)(R_{f_i} - \bar{R}_f)}{\sqrt{\sum_{i=1}^{N}(R_{g_i} - \bar{R}_g)^2 \sum_{i=1}^{N}(R_{f_i} - \bar{R}_f)^2}} \tag{15}$$

3.7 Reproducibility

A significant contribution of this work is the development of `mergeplg`, an open-source Python package for interpolation and merging of ground-based and radar-derived rainfall data (Øydvin et al., 2025a). The package optimizes workflows by updating

CML coordinates and reusing kriging matrices across time steps. Built within the OpenSense data ecosystem (Chwala et al., 2025b), `mergeplg` supports data from CMLs, PWSs, and radar systems, provided they follow the standardized OpenSense data format (Fencl et al., 2024). The package also integrates with `poligrain` (Chwala et al., 2024), a Python library for geospatial operations and visualization, enabling efficient tasks such as spatial queries (e.g., finding points near lines), plotting georeferenced data and computing metrics. By releasing `mergeplg` as open source, we aim to promote transparency, reproducibility, and collaboration, fostering its use and adaptation within the broader scientific community.

4 Results

4.1 Two overview case studies

Figure 2 provides a case study of the application of different merging methods of two 24 km^2 patches of OpenMRG (2015-08-27 10:00) and OpenRainER (2022-08-18 10:00). In the OpenMRG case, the radar field underestimates rainfall compared to rain gauges, while all merging methods reduce this underestimation to varying degrees. When comparing the rainfall fields, KED produces smoother fields overall, while additive retains more variability, and multiplicative introduces the strongest variability and sharper features. These differences are also reflected in the difference fields, where additive methods show smoother corrections compared to multiplicative. The difference fields of KED, on the other hand, exhibits greater spatial variability.

In the case reported for OpenRainER, the lower density of the CMLs makes it difficult to capture precipitation variability; however, we can appreciate a reduction in radar overestimation compared to the reference rain gauges. Looking at the difference fields, we can observe stronger radial symmetry for the IDW based methods, while kriging kriging provides smoother corrections.

4.2 Variogram study

Fig. 3 shows the temporal evolution of variogram nugget and range parameters (with sill scaled to 1) for OpenMRG and OpenRainER, comparing CML (blue) and gauge (orange) data across four data configurations: Difference between radar and CML/gauge, ratio between radar and CML/gauge, CML/gauge alone and radar evaluated at CML midpoint or at gauge location.

In both datasets, CML data generally exhibit higher and more variable nugget and range values compared to gauge data. For radar-only fields (evaluated at the CML midpoint or gauge location), this variation is smaller, and the nugget is generally lower. Generally, the difference fields show larger variation than radar-only fields, followed by CML/gauge data; the ratio fields exhibit the largest variability in range and nugget estimates.

For OpenMRG, the range estimated directly from CML data is consistently larger than that from rain gauge data. This pattern also holds when using radar estimates at the CML midpoint or gauge location. This may be due to underestimation of range when rain gauges cover a small area (Western and Blöschl, 1999). In contrast, for OpenRainER, the range estimated from CML data is similar to the range estimated from rain gauge data and radar at the CML midpoint/gauge.

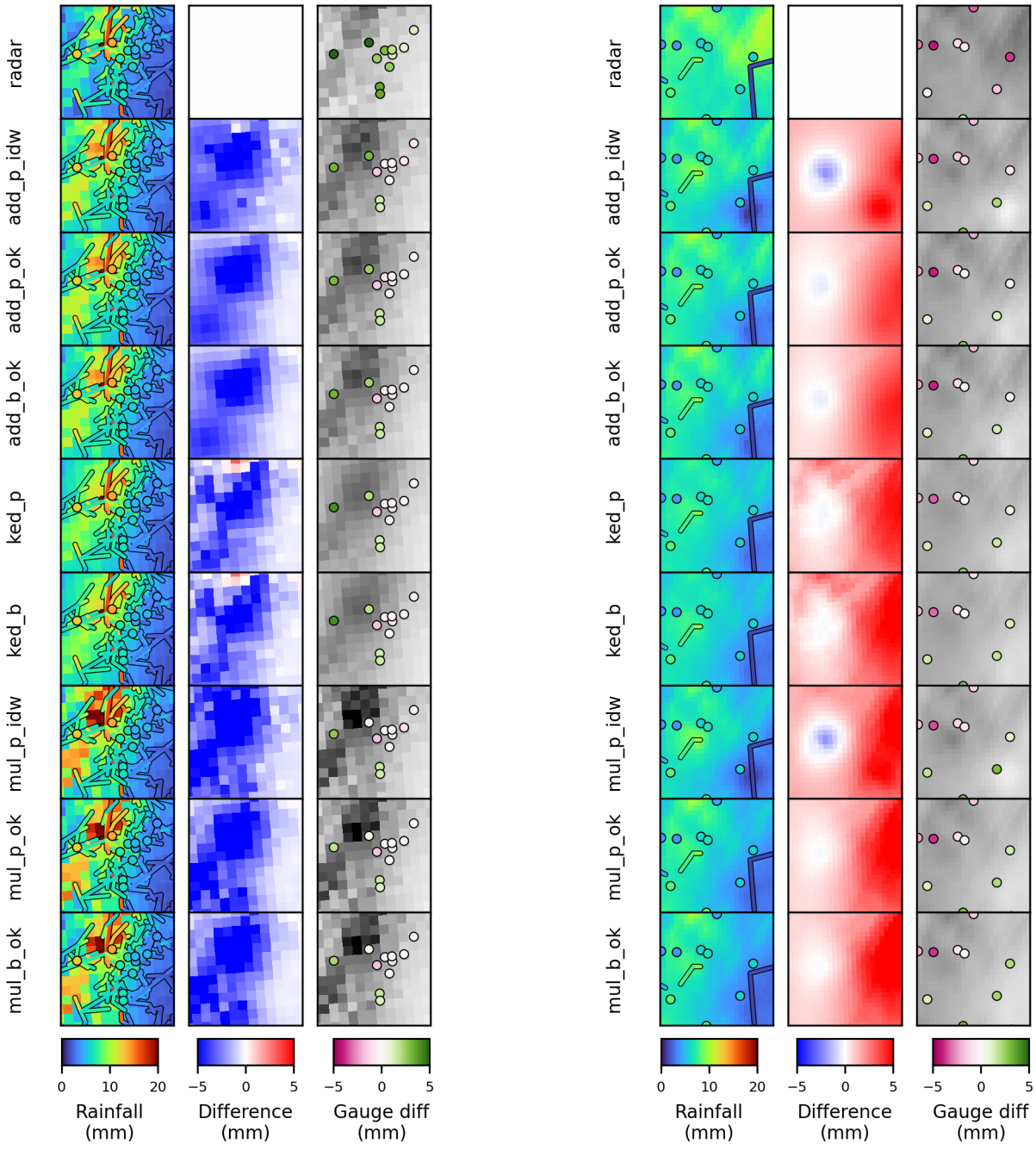


Figure 2. Case studies for both datasets. Left: OpenMRG, 2015-08-27T10:00; Right: OpenRainER, 2022-08-18T10:00. Each map shows an area of 24 km^2. The first column shows the 1-hour rainfall maps, the second column shows the differences between the merged field and the original radar field, and the third column shows the differences between the fields and the rain gauges, with the rainfall field from each method shown as a gray-scale background for spatial context.

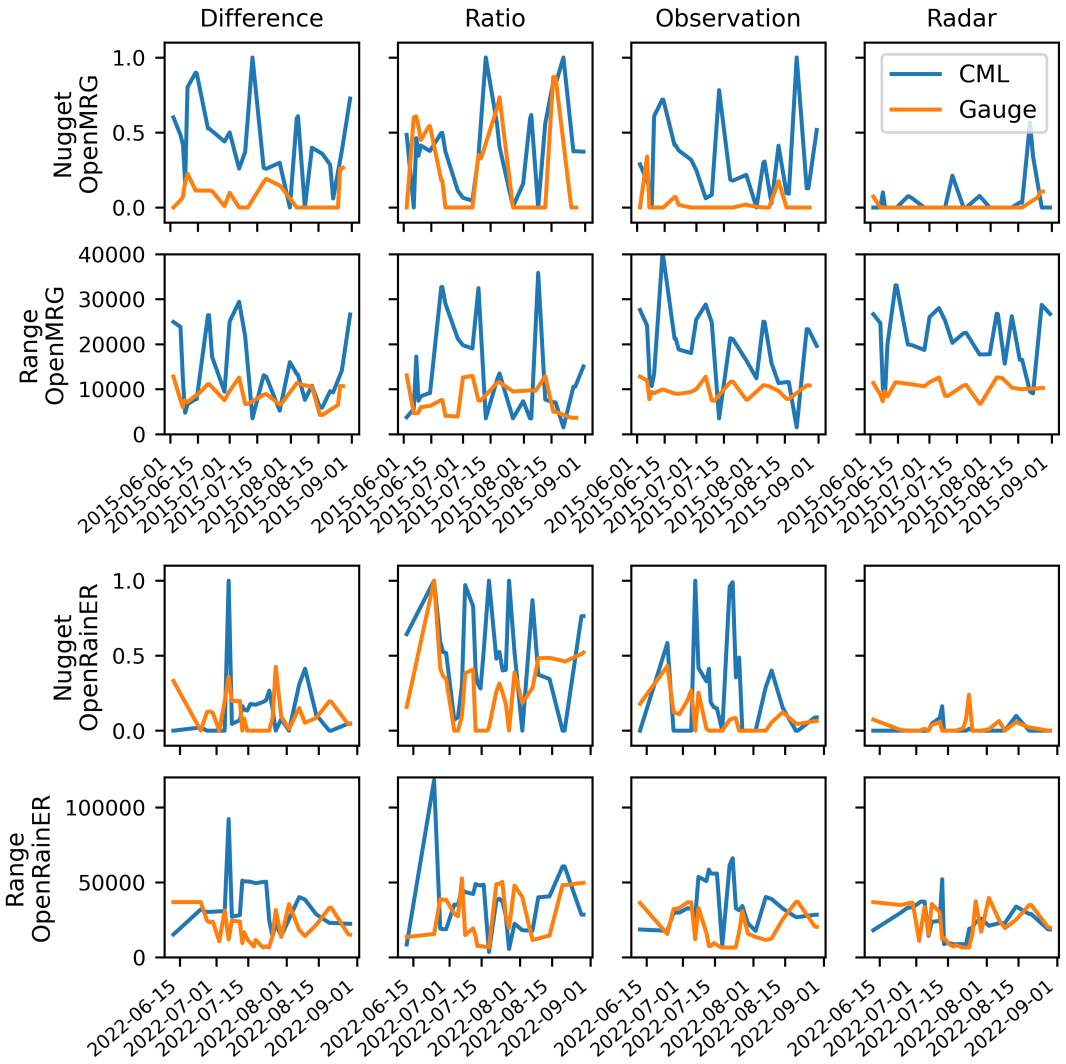


Figure 3. Variogram parameters (nugget and range; sill scaled to 1) are shown for OpenMRG (top two rows) and OpenRainER (bottom two rows). Columns represent: (1) difference between radar and CML/gauge; (2) ratio between radar and CML/gauge; (3) CML/gauge data alone; and (4) radar evaluated at CML midpoint (CML) or at gauge location (gauge). Colors distinguish CML (blue) and gauge (orange).

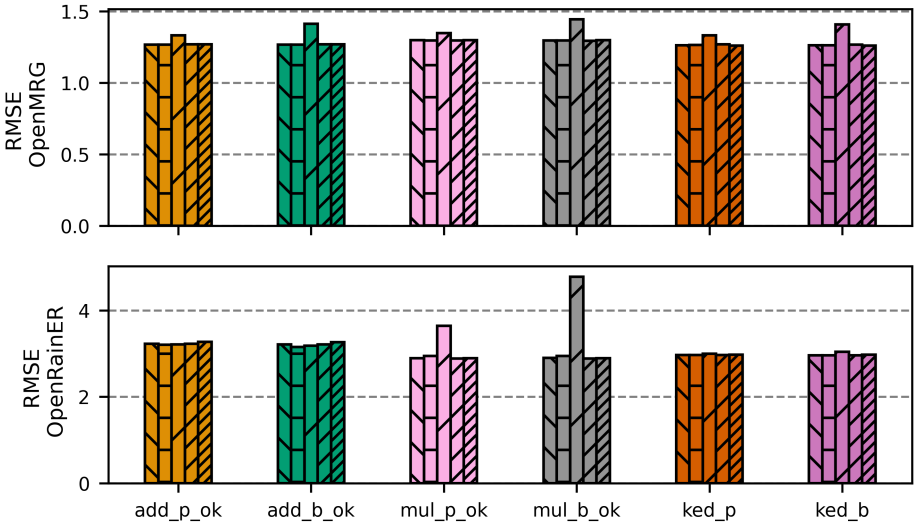


Figure 4. Root Mean Square Error (RMSE) for different radar–CML merging methods—additive, multiplicative, and kriging with external drift (KED)—applied to the OpenMRG (top) and OpenRainER (bottom) datasets. For each method, five variogram estimation strategies are compared: (1) constant range and nugget from van de Beek et al. (2012), (2) variogram fitted to data, (3) variogram fitted to data with nugget fixed at 0, (4) variogram fitted to data with nugget fixed at 0.3, and (5) variogram fitted to data with nugget fixed at 0.6. Bars show mean RMSE values for each combination.

4.3 Impact of variogram parameters on interpolation

Fig. 4 presents the RMSE values for the OpenMRG and OpenRainER datasets across various radar–CML merging methods and variogram estimation strategies. Each merging method - additive, multiplicative, and KED - is evaluated in both point kriging (p) and block kriging (b) variants, giving six method groups in total. For each group, five variogram estimation strategies are compared: using constant range and nugget values from van de Beek et al. (2012), fitting the variogram to data with either a free nugget or a fixed nugget value (0, 0.3, or 0.6). For additive and KED, the variogram is estimated from the difference between the radar and CML, while for multiplicative the variogram is estimated from the ratio between CML and radar. The data is scaled so that the sill is equal to 1.

In OpenMRG, a nugget of 0 produces moderately worse performance for all methods (RMSE increase of 0.07–0.15 mm), while on OpenRainER additive and KED methods are robust even to a zero nugget. Multiplicative methods show the strongest sensitivity to the nugget=0 case on OpenRainER, where block kriging RMSE increases from approximately 2.90 to 4.78 mm. Among non-zero nugget strategies, the free nugget, fixed nugget at 0.3, and fixed nugget at 0.6 perform similarly across all methods, and the reference variogram parameters perform at least as well as any data-estimated strategy. Overall, the choice of variogram estimation strategy has little practical importance provided a non-zero nugget is used, with multiplicative block kriging on sparse networks being the most sensitive case.

4.4 Overall performance of adjustment methods

The performance of radar and merged radar-CML rainfall fields was evaluated at an hourly resolution using rain gauges as ground truth (Fig. 5). Overall, the OpenRainER dataset shows higher RMSE and MAE compared to the OpenMRG dataset, indicating greater variability and challenges in accurately estimating rainfall in this dataset.

Focusing on the OpenMRG dataset, all merging methods improve the radar field, with relatively similar performance across methods. On average, MAE is reduced from 0.84 mm to ~~0.69~~ 0.67 mm (~~-18~~20%), RMSE is reduced from 1.44 mm to ~~1.29~~ 1.28 mm (-11%), and PCC is increased from 0.55 to 0.69. ~~The BIAS shows some variation with best improvement from Multiplicative methods.~~

In contrast, the OpenRainER dataset exhibits more variation between merging methods. The radar has a RMSE of about 3.6 mm, which is reduced to 3.2 mm (-10%) for additive, ~~and 3.1~~ 3.0 mm (~~-14~~17%) for KED, ~~while multiplicativeshow no improvement~~and 2.9 mm (19%) for multiplicative. In terms of MAE, the radar has a MAE of about 2.0 mm, which is reduced to 1.6 mm (~~-23~~21%) for ~~Additive, 1.26~~ for additive, 1.25 mm (-38%) for KED, and ~~1.35~~ 1.19 mm (~~-33~~41%) for multiplicative.

In terms of PCC, the radar has a PCC of about 0.75, which is increased to 0.76 for additive ~~, decreased to 0.73 for KED~~and KED, and decreased to ~~0.65~~ 0.74 for multiplicative. In terms of BIAS, the radar has a BIAS of about 107 %, which is reduced to ~~32~~ 35 % for additive, ~~-7~~ -6 % for KED and ~~26~~ -2 % for multiplicative. Thus, ~~KED show~~ multiplicative shows the smallest absolute BIAS and the best RMSE ~~, MAE and BIAS, while additive show slightly better~~ and MAE, while KED achieves a slightly higher PCC.

The ~~block kriging correction provides a small improvement for the multiplicative merging methods, reducing RMSE from 3.6 mm to 3.5 mm. For the additive and KED methods, the~~ effect of the block kriging correction is less pronounced, but compared to point kriging, it consistently reduces the relative BIAS, which can result in negative BIAS values becoming more negative.

4.5 Intensity-based analysis

To evaluate performance under varying rainfall intensities, an intensity-based analysis was performed (Fig. 6). Results are presented for four rainfall intensity bins, based on the rain gauge values: light rainfall $[0.2, 2.5]$ mm/h, moderate rainfall $[2.5, 10]$ mm/h, heavy rainfall $[10, 50]$ mm/h, and extreme rainfall (> 50 mm/h). Light rainfall constitutes 63.3% of rainfall in OpenMRG (with 22.7% with less than 0.2 mm/h) and 73.2% in OpenRainER (with 1.6% with less than 0.2 mm/h), therefore the light rainfall class reflects the results from overall performance analysis most. Generally, OpenRainER exhibits larger errors for light and moderate rainfall, but its radar outperforms the OpenMRG radar for heavy rainfall. Additionally, OpenRainER includes extreme rainfall events, which are absent in OpenMRG.

For OpenMRG, all merging methods consistently outperform the radar across all rainfall intensity bins, showing similar performance overall. The ~~main exception is for~~ exception is heavy rainfall, where multiplicative ~~IDW performs slightly worse in terms of RMSE, while multiplicative kriging achieves better results for both RMSE and bias.~~ methods perform slightly

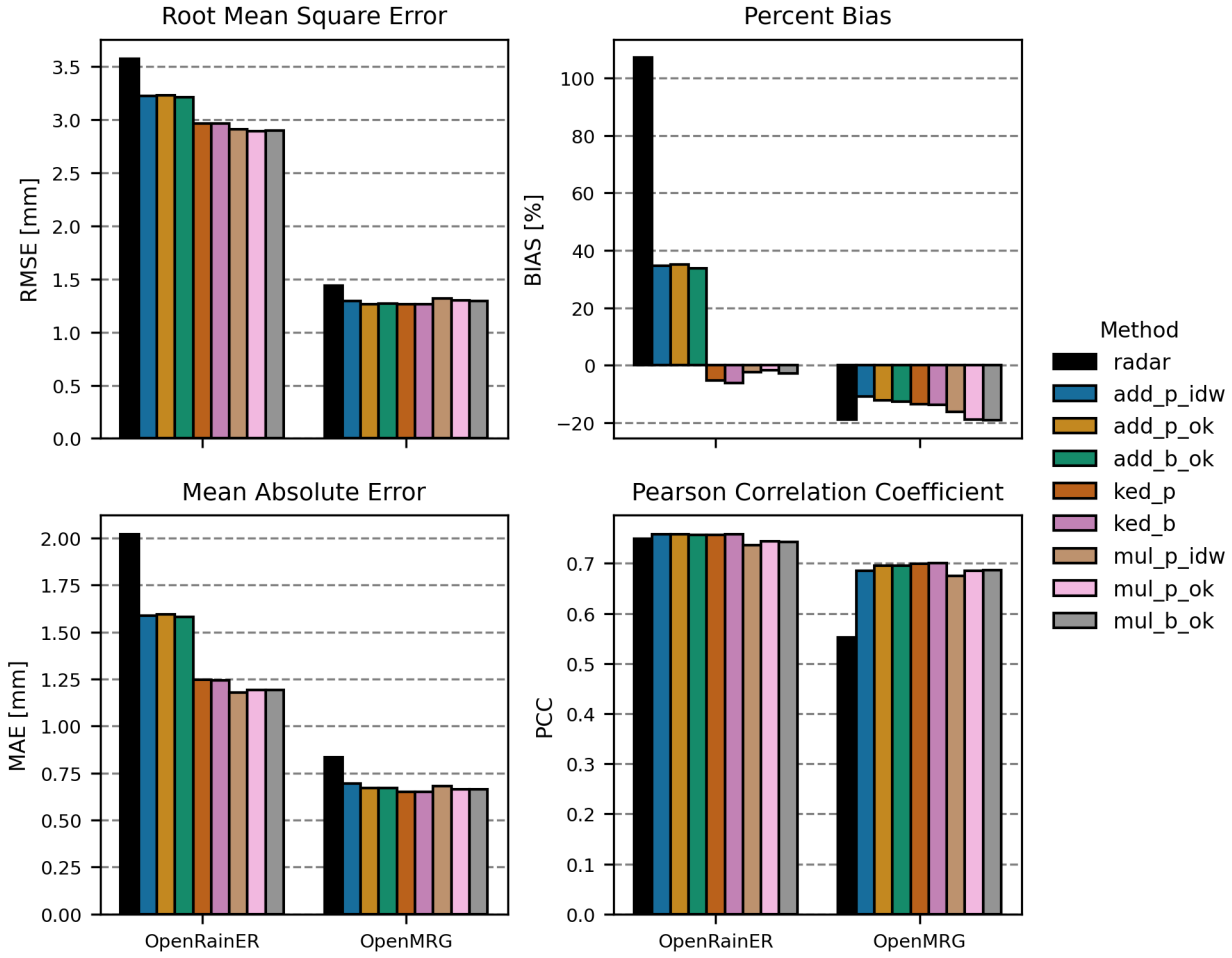


Figure 5. Root Mean Square Error (RMSE), Percent Bias (BIAS), Mean Absolute Error (MAE) and Pearson Correlation Coefficient (PCC) for all merging methods applied to the OpenRainER and OpenMRG datasets.

worse than additive and KED, with multiplicative IDW being the least good among them — though all methods still improve on the radar.

In contrast, OpenRainER shows greater variability between merging methods. ~~KED performs best for~~ For light and moderate rainfall, ~~while radar and additive performs best for~~ multiplicative and KED outperform additive, with multiplicative achieving the lowest RMSE overall. For heavy and extreme rainfall~~. Multiplicative show moderate performance for light rainfall but perform less good at higher intensities~~, radar and additive perform best, while KED and multiplicative show substantially higher RMSE. In terms of BIAS, OpenRainER shows a strong positive BIAS for low rainfall rates, which decreases with intensity, eventually becoming negative. OpenMRG exhibits a near-zero BIAS for low rainfall rates, which gradually decreases at higher intensities.

4.6 Distance-based analysis

A distance-based analysis was performed on the OpenRainER dataset (Fig. 7). We grouped rain gauges into 2 km distance increments (e.g., 0–2 km, 0–4 km, etc.), with each group including all gauges within the specified range from a CML. The OpenMRG dataset was excluded here because its dense CML and rain gauge network caused all rain gauges to fall into the first distance bin, reducing the relevance of a distance-based analysis.

Generally, the distance-based analysis showed that all methods improve as the distance from CML decreases. In terms of RMSE and MAE, ~~KED performed best for most~~ multiplicative kriging performed best at all distances, with RMSE approaching ~~2.7~~ 2.6 mm/h and MAE approaching ~~1.1~~ 1.0 mm/h ~~. Additive methods follow a trend similar to KED , but perform worse~~at the closest distances. KED performed similarly, with RMSE approaching ~~2.85 mm/h, MAE approaching 1.3 mm/h. Multiplicative methods generally perform least good~~2.7 mm and MAE approaching 1.1 mm. Additive methods performed worse, with RMSE approaching 2.9 ~~mm/h~~ mm and MAE approaching ~~1.1 mm/h, though Kriging-based Multiplicative methodsoutperform all other methods at around 5 km~~1.4 mm. Notably, multiplicative IDW performs substantially worse than multiplicative kriging at close distances, approaching the performance of additive methods. In contrast, additive IDW performs similarly to additive kriging across all distances. Block kriging shows no clear improvement over point kriging at any distance.

In terms of BIAS, all merging methods ~~improve the radar , which has a strong positive BIAS of~~ reduce the radar BIAS of approximately 100%. Additive methods ~~improve slightly~~ have a positive BIAS that decreases as distance from CMLs decreases, approaching 20% ~~. Multiplicative methods show stronger reductions, with BIAS approaching 0 % . KED maintains the lowest BIAS overall, starting at a moderate negative value~~ at the closest distances. Both KED and multiplicative methods have a negative BIAS that increases in magnitude as distance decreases, reaching about -10% for KED and -13% for multiplicative kriging at the closest distances. At far distances, multiplicative BIAS is closest to zero (-2%), while KED is around -5%~~and decreasing further to -10%. For PCC, at distances greater than 20 km, additivemethods perform slightly better than radar, while KED perform slightly worse. However, for distances less than 20 km, .~~

In terms of PCC, additive and KED outperform radar at all distances. Compared to additive, KED ~~improve fastest, with KED ultimately outperforming~~ improves most rapidly with decreasing distance, achieving the highest PCC (0.83) near CMLs. Close

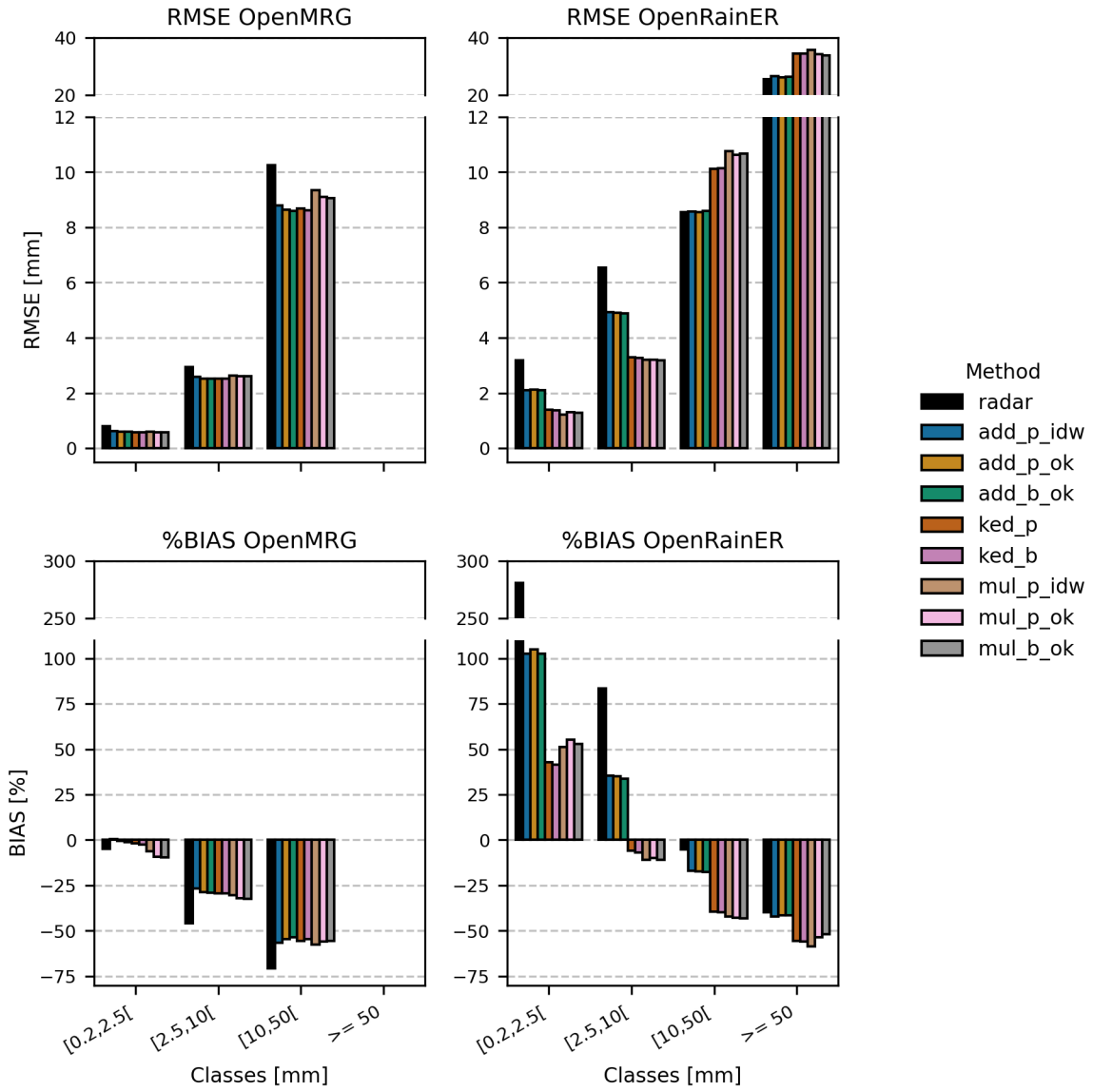


Figure 6. Root Mean Squared Error (RMSE) and Percent Bias (BIAS) for light rainfall $[0.2, 2.5]$ mm/h, moderate rainfall $[2.5, 10]$ mm/h, heavy rainfall $[10, 50]$ mm/h, and extreme rainfall (> 50 mm/h) for all merging methods applied to the OpenRainER and OpenMRG datasets. The intensity classes are based on rain gauges values.

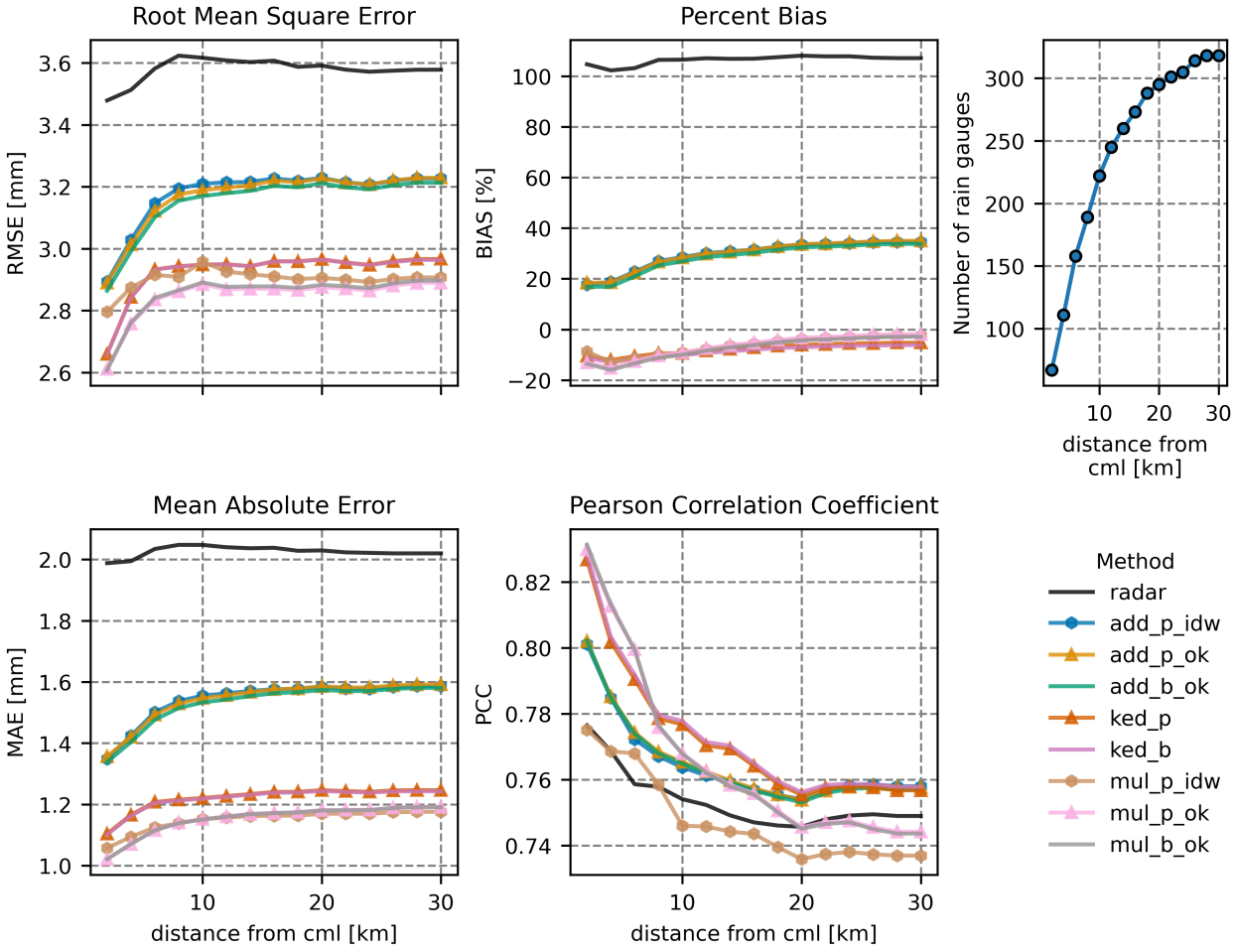


Figure 7. Root Mean Square Error (RMSE), Percent Bias (BIAS), Mean Absolute Error (MAE), and Pearson Correlation Coefficient (PCC) as functions of the maximum allowed distance between reference rain gauges and CMLs for all merging methods applied to the OpenRainER dataset. Rain gauges are grouped into 2 km distance increments, with each group including all gauges within the specified range from a CML.

to observations, multiplicative kriging performs similarly good as KED, but as distance from CML increases the performance drops below all other methods~~near rain gauges.~~. Multiplicative IDW performs worse.

Interestingly, radar performance (RMSE and PCC) also improves slightly when rain gauges are near CMLs, potentially due to better radar coverage in those areas.

4.7 Sensitivity analysis of adjustment value range checks

A range check sensitivity analysis was performed to evaluate the impact of different range check thresholds on the merging methods (Fig. 8~~and Fig. ??~~). Three range check configurations were tested: no check (no range check applied), standard,

and strict. The thresholds for additive and KED methods were difference = 10 mm/h (standard) and difference = 5 mm/h (strict), while for Multiplicative methods they were ratio = [0.1, 15] (standard) and ratio = [0.2, 8] (strict). Under standard settings, the additive range check removed fewer than 2% of valid CML-radar pairs in both datasets (0.0% in OpenMRG, 1.4% in OpenRainER). In contrast, the multiplicative range check removed 37.4% of valid pairs in OpenMRG and 53.2% in OpenRainER.

For the OpenMRG dataset, range checks had little impact on RMSE, and all combinations of merging methods and range checks improved the rainfall fields compared to radar alone. In ~~terms of BIAS, stricter range checks reduced the BIAS for additive and KED methods (e.g., additive IDW decreases from -10 % to -12.5 %). For multiplicative methods, stricter range checks improved the BIAS (e.g., multiplictive IDW increase from -10 % to 6 %)~~

~~In~~ the OpenRainER dataset, range checks play a more critical role~~. Without range checks, the merged fields often~~ : without them, merged fields can perform worse than radar alone~~. For example, RMSE for radar is about 4 mm, but increases to 5 mm for KED without range checks. Applying range checks improves RMSE for all methods: standard range checks perform best~~ , with KED RMSE exceeding the radar baseline of about 3.6 mm. Standard range checks are the preferred setting for additive and KED methods~~, while stricter range checks are optimal for multiplicative methods. In terms of BIAS, standard range checks generally result in the lowest BIAS. In terms of MAE, the standard range check was optimal for all methods~~. For multiplicative methods, performance is similar between standard and strict settings, suggesting that any reasonable threshold is sufficient provided a check is applied.

4.8 Detailed case-studies

To better understand the strengths and limitations of the various merging methods, two case studies were performed (Fig. 9 and Fig. 10). The case studies span an area of approximately the same size, with the ~~OpenMRG~~ OpenRainER dataset covering longitude $(9.6, 10)$ and latitude $(44.7, 45)$, and the ~~OpenRainER~~ OpenMRG dataset covering longitude $(11.8, 12.2)$ and latitude $(57.5, 57.8)$.

The radar fields reveal differences in rainfall patterns between the two case studies. For OpenRainER , rainfall is more scattered, with several localized extremes, while for OpenMRG, rainfall is less scattered and concentrated along a single line. Additionally, the density of CMLs is much higher in the OpenMRG dataset, resulting in interpolated fields that are very similar across merging methods due to the dense observation network. In the OpenRainER dataset, differences between merging methods are more pronounced. At location $(9.7, 44.95)$, a CML observes approximately 40 mm, while the radar observes $10\ \mathrm{mm}$. In the additive and KED merging methods, this CML observation is excluded by the range check (difference > 10 mm). However, in the multiplicative merging methods, the ratio check (ratio = 4) does not flag the observation, and it is included. Since there are no nearby CMLs, this results in a large rainfall estimate in the merged field. While there are no nearby rain gauges to validate the observation, the resulting rainfall estimate appears overly high. At another location $(9.75, 44.8)$, the additive merging method excludes a CML observation due to the range check, while the multiplicative merging method retains it. A nearby rain gauge indicates that the CML measurement is likely correct, suggesting that this observation should have been included.

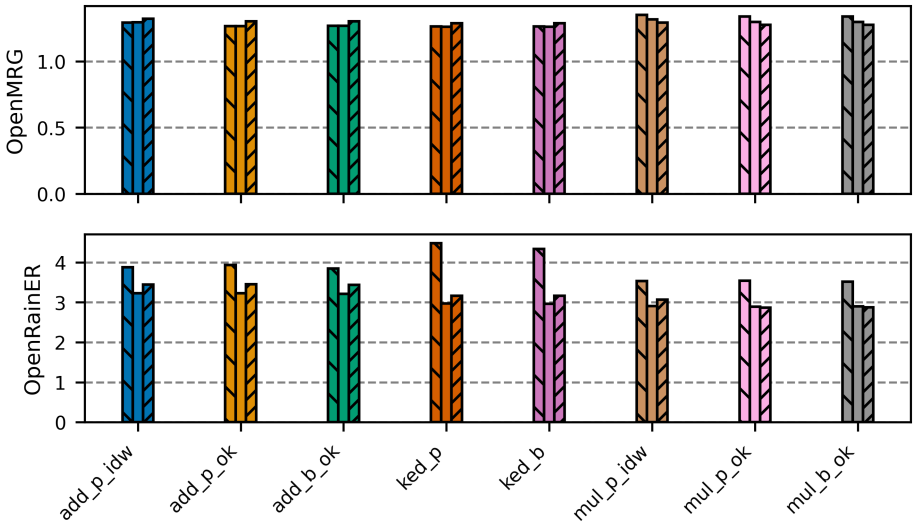


Figure 8. Root Mean Square Error (RMSE) ~~, Percent Bias (BIAS), Mean Absolute Error (MAE), and Pearson Correlation Coefficient (PCC)~~ for all merging methods under three range check configurations (no check, standard, and strict) for the OpenMRG ~~dataset~~(top) and OpenRainER (bottom) datasets.

~~Root Mean Square Error (RMSE), Percent Bias (BIAS), Mean Absolute Error (MAE), and Pearson Correlation Coefficient (PCC) for all merging methods under three range check configurations (no check, standard, and strict) for the OpenRainER dataset.~~

5 Discussion

The purpose of this study was to evaluate how CMLs can be used to improve radar QPEs and to compare the performance of different merging methods. This was done by assessing their overall performance, their behavior under varying rainfall intensities, and their accuracy across different distances to rain gauges. We also analyzed the impact of range checks on merging performance and explored specific scenarios through case studies.

5.1 Dataset characteristics and their influence on merging performance

The radar product in OpenMRG outperforms the OpenRainER for low and moderate rainfall. This could be due to the study area in OpenMRG having a relatively uniform terrain and moderate distance (78 km) to the closest weather radar (Andersson et al., 2022). Moreover, all rain gauges are located in the city center (Fig. 1), a localized area where the radar rainfall estimates appear to perform well. In contrast, OpenRainER covers a larger and more complex region with diverse terrain, contributing to higher errors. Despite this, OpenRainER's radar outperforms OpenMRG's radar during heavy rainfall. This could be due to the capping applied to reflectivity in the dataset (Covi et al., 2026), which helps prevent overestimation caused by hail but also limits the representation of extreme intensities.

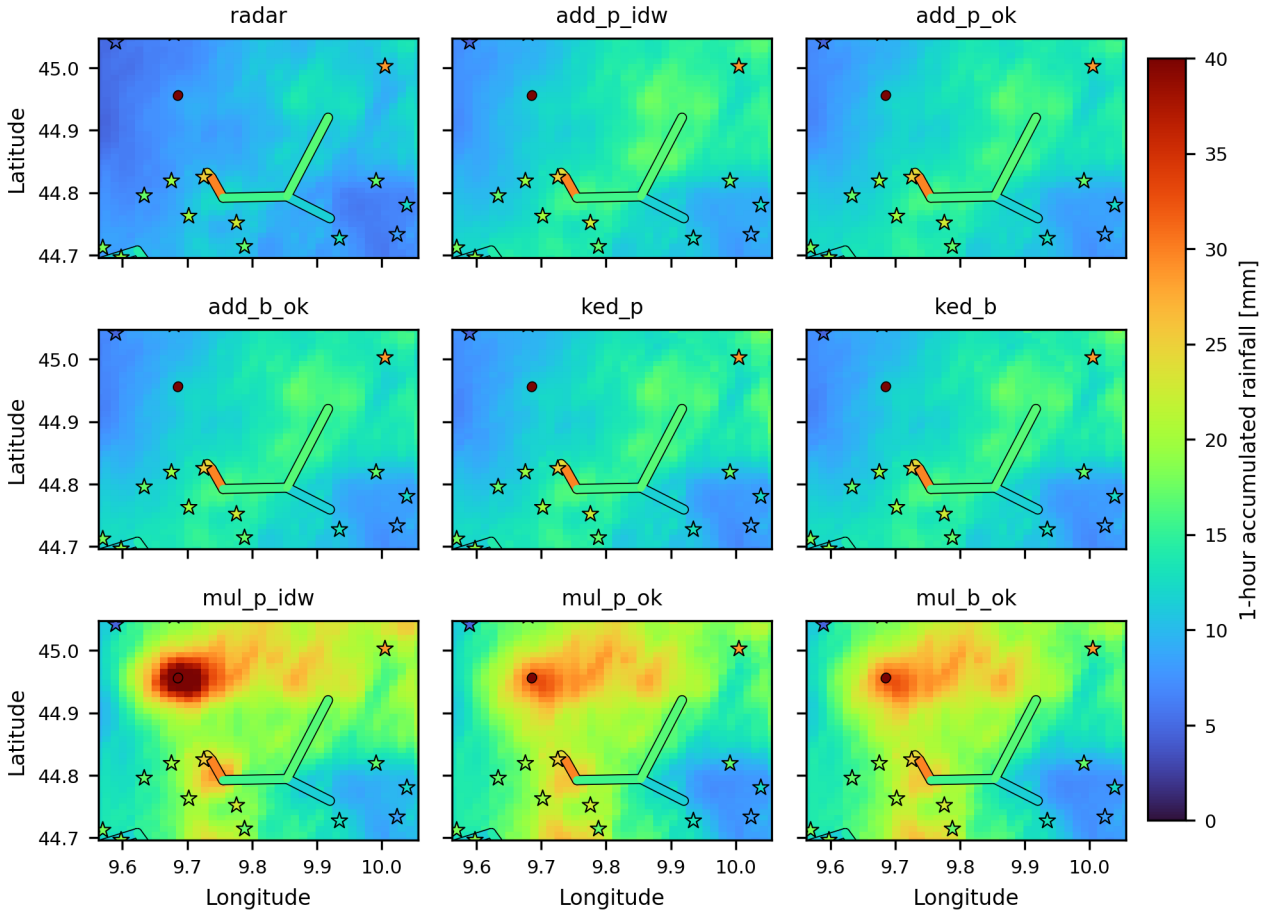


Figure 9. Spatial rainfall field from radar and merging methods for the OpenRainER dataset, with rain gauges (stars) and CMLs (lines) colored by rainfall amounts. Reference time: 2022-08-18T10:00.

The density of the CML network differs significantly between the two datasets, strongly influencing merging performance. OpenMRG benefits from a dense CML network (1 CML per 7.7 km^2, with even higher density near the majority of rain gauges in central Gothenburg), providing strong observational constraints that reduce variability in the merged fields and minimize differences between merging methods. For comparison, Nielsen et al. (2024) reported a similar observational density in Denmark (1 PWS per 2.1 km^2). These densities are notably higher than in many other studies. For instance, Goudenhoofdt and Delobbe (2009) used a rain gauge network in Belgium with a density of 1 gauge per 135 km^2, Delrieu et al. (2014) used a dense network in the Cévennes-Vivarais region of France with 1 gauge per 128 km^2, and Jewell and Gaussiat (2015) used a sparser network in England and Wales with 1 gauge per 204 km^2. The density in OpenRainER (1 CML per 137 km^2) is comparable to these earlier studies. However, unlike rain gauge networks, which are typically designed for even spatial

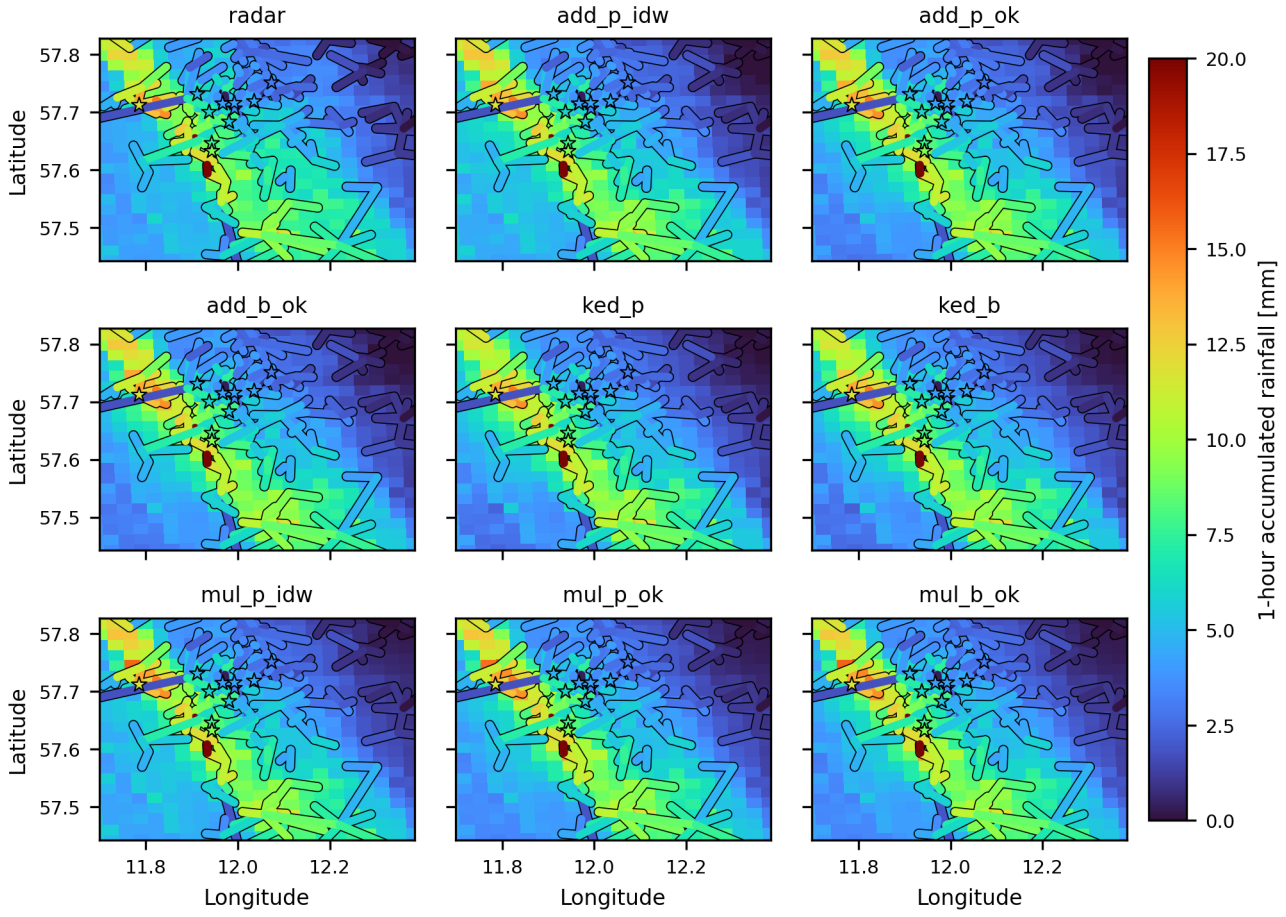


Figure 10. Spatial rainfall field from radar and merging methods for the OpenMRG dataset, with rain gauges (stars) and CMLs (lines) colored by rainfall amounts. Reference time: 2015-08-25T08:00

coverage, CMLs are often concentrated along roads and urban areas. This uneven distribution can result in lower densities in some regions, posing challenges for merging methods.

The sparser observation network in OpenRainER increases reliance on radar, which can lead to greater uncertainty in the merged fields. Mismatches between radar and CML observations can arise due to spatial or temporal discrepancies, particularly in regions with complex terrain or areas where wind shifts rainfall patterns. These mismatches can lead to adjustments that do not reflect the actual rainfall distribution and sometimes cause large corrections and errors. For example, in the case study for OpenRainER (Fig. 9), we observed that sparse CML coverage in the north resulted in overcorrections, where a single CML observation drove adjustments over a large area. Conversely, in the south, strict range checks excluded valid CML observations.

5.2 Performance of merging methods

In OpenMRG, all merging methods improved the radar field~~with relatively~~ , with additive and KED showing similar performance (e.g., ~~MAE reduced by about 18%), suggesting robustness in data-dense settings. By~~ 20% MAE reduction). Multiplicative methods exhibited slightly higher errors and a negative bias, likely because dense networks provide strong constraints that favor smoother methods like KED over the sharp gradients of multiplicative merging. In contrast, Nielsen et al. (2024) ~~used a similarly dense rain gauge network and reported larger differences between merging methods . One~~ reported larger performance gaps between methods in a similarly dense network. A possible explanation is that ~~the mean field bias adjustment (MFB) and inverse distance weighting bias adjustment (IDWB) applied in their study do not include~~ their simpler adjustments (MFB and IDWB) lack the local neighborhood filtering ~~mechanism introduced in their moving median bias adjustment~~ inherent in their more advanced methods (MMB) ~~. In our case, all approaches rely on local neighborhoods, which naturally reduce the influence of outliersand provide stability~~ and in all our approaches. Our sensitivity analysis (Fig. B1) shows that performance is robust across various neighborhood sizes, including when using all observations, suggesting that spatial weighting in combination with range checks provides such inherent stability in dense networks that restricting the neighborhood size offers no additional protection against far-away outliers. This suggest that the presence of spatial weighting and range checks provides sufficient stability against outliers in dense networks~~. We~~ , rendering performance largely independent of neighborhood size. While we did not implement ~~a median-based interpolator, nor did we explore approaches that include all observations without neighborhood constraints, so we cannot directly assess the added value of median filtering. This makes it difficult to disentangle whether the performance gains reported for MMB arise primarily from the use of local neighborhoods or from the median filtering itself~~MMB, our results suggest it would be worth investigating whether median-based filtering offers additional gains beyond spatial weighting alone.

In contrast, OpenRainER exhibits more variation between merging methods. The improvements observed in OpenRainER are consistent with findings from other studies, with similar network densities. For example, Goudenhoofdt and Delobbe (2009) found that KED and kriging with radar-based error correction reduced radar MAE by about 40% for daily rainfall in Belgium. Similarly, Jewell and Gaussiat (2015) reported significant reductions in RMSE and MAE for hourly rainfall at different rainfall intensities for KED in England and Wales, with improvements of up to 30-40% in MAE and 25-40% in RMSE. ~~Overeem et al. (2024b) found that PWS reduced radar MAE by 20% in a European wide study.~~ In OpenRainER, reductions in MAE were ~~20~~21% for additive, ~~32~~41% for multiplicative, and 38% for KED, aligning well with these earlier findings. Overeem et al. (2024b) found that PWS reduced radar MAE by 20% in a European wide study. However, the reductions in RMSE were more modest~~in OpenRainER, with an improvement of about 8~~: 10% for additive, ~~13~~17% for KED and ~~no improvement~~ 19% for multiplicative. This could be due to the less evenly distributed CML network and the complex terrain, which amplify radar-CML mismatches and create greater uncertainty in merged fields. It could also be due to signal fluctuations causing CML rainfall outliers, or the path averaging effect of the CML dampening rainfall extremes.

The intensity-based analysis suggests that ~~KED performs~~ multiplicative and KED methods perform better in terms of RMSE for low and moderate rainfall~~in both OpenRainER and OpenMRG, but performs worse relative to other methods for heavy~~ ,

while radar and additive perform best for heavy and extreme rainfall, particularly in OpenRainER. This performance gap could be explained by KED's inherent tendency to produce smoother rainfall fields , (as observed in Fig. 2. In KED, the interpolated rainfall field combines a deterministic component (drift), estimated by linearly relating the radar field to CML observations, with a stochastic component that adjusts the drift near observation locations. If the relationship between radar and observations is non-linear, enforcing a linear regression in KED can introduce bias, causing both over- and underestimation of rainfall values. Additionally, the path-averaging nature of CMLs inherently smooths rainfallestimates, further dampening extremes.Together, these factors can produce a smoothed drift surface that misrepresents extreme rainfall, particularly in areas far from observations. Overall, KED works best with close to medium-range, high-quality). This smoothing effect is likely due to the KED drift condition (Eq. 7). By forcing the weighted sum of path-averaged radar values to match a single pixel value at the target location, the method effectively smears localized peaks across the length of the contributing CMLs. As link length increases, the observations and for low to moderate rainfall intensities, but may underperform in data-sparse regions or represent a broader spatial mean, eventually losing the high-frequency variability required to reconstruct local peaks. In this limit, the drift term becomes spatially uniform and fails to provide the local 'guidance' necessary for KED to resolve intense, localized rainfall. This explains why KED excels at capturing moderate rainfall patterns but consistently underperforms during extreme events where localized intensity peaks are dampened. Further, the distance-based study (Fig. 7) reveals that KED and multiplicative methods follow a similar bias trend, with both exhibiting a negative bias that increases in magnitude as distance to the CML decreases. Despite this, KED maintains a superior PCC across all distances, outperforming all other methods for medium ranges. Consequently, while KED excels when utilizing high-quality, medium-range observations during low-to-moderate events, it tends to underperform during extreme rainfalldue to its smoothing and linear assumptions. This smoothing effect appears to be a general property of the method rather than a dataset-specific artifact, as it remains consistent across both OpenMRG and OpenRainER datasets. Furthermore, the neighborhood size sensitivity analysis (Fig. B1) confirms that this smoothing is intrinsic to the interpolation itself and not primarily driven by neighborhood configuration, as results remain relatively stable across different configurations.

In contrast, additive methods outperform KED in terms of RMSE for heavy rainfall, and in terms of PCC for regions 15 km away from rain gauges. This may be and multiplicative merging during heavy and extreme rainfall. This is likely because additive methods apply corrections directly to adjust the radar field without introducing as much smoothing, which helps retain more of the radar's variability or better preserves extreme rainfall values in areas with sparse observational data. In OpenRainER, KED outperforms the other methods when rain gauges are close to CMLs. However, in OpenMRG, where all rain gauges are within 2 km of a CML, KED performs slightly worse, especially for heavy rainfall.This discrepancy could be due to that KED benefits from the radar external drift in locations in medium range from rain gauges.In contrast, in CML-dense areas, using local differences, which preserves the underlying spatial variability and high-intensity peaks more effectively than scaling or re-interpolation. While KED and multiplicative methods also aim to improve the radar field, they are more prone to underestimating extremes. As shown in the additive approach might preserve more of spatial variability of rainfall, where KED provide more smoothed estimates. Overall, additive methods are most effective for capturing extremes and maintaining detail where observations are sparse or at greater distances from gauges. intensity-based analysis (Fig. 6), both exhibit a significant

negative bias during high-intensity periods in OpenRainER, directly resulting in higher RMSE. For KED, this could stem from its inherent smoothing effect (Fig. 2), while for multiplicative merging, it could be driven by sensitive ratio-based range checks that frequently remove high-intensity observation pairs (Fig. 8).

~~Multiplicative merging shows mixed performance. In dense networks like OpenMRG~~Multiplicative merging achieves the best overall RMSE and MAE in OpenRainER, and the smallest absolute BIAS (-2%). This is noteworthy because a key source of radar error — variability in the Z-R relationship — is inherently multiplicative (Villarini and Krajewski, 2010), which makes multiplicative adjustment theoretically well-motivated. The large bias reduction in OpenRainER drives the RMSE and MAE improvement, since these metrics are sensitive to systematic offsets. However, multiplicative merging has slightly lower PCC than KED and additive methods in both datasets (OpenRainER: 0.74 vs 0.76 for KED; OpenMRG: 0.69 vs 0.70), and in OpenRainER multiplicative PCC does not improve over radar (0.74 vs 0.75), in contrast to additive and KED which both exceed radar PCC. Multiplicative methods also perform poorly for heavy and extreme rainfall, and remain sensitive to single erroneous CML observations in sparse networks, ~~it performs comparably to other methods for heavy rainfall and can even yield lower bias, likely due to the stabilizing effect of dense CML coverage. However, in sparser networkssuch as OpenRainER, multiplicative methods are highly sensitive to outliers, as single extreme CML values can dominate the adjustment and lead to overcorrections. While multiplicative merging sometimes achieves the lowest MAE, especially in localized, data-rich settings where sharp rainfall features are supported by nearby observations , it is generally less reliable for RMSE and PCC, and caution is warranted in areas with sparse or unevenly distributed data~~as illustrated in the case study (Fig. 9), where a single unchecked CML produced a large overcorrection. Range checks reduce this sensitivity, but at a cost — they can also filter valid large ratios associated with heavy rainfall events. The sensitivity to erroneous observations in sparse areas is further reflected in the distance analysis: multiplicative PCC falls below that of radar at larger distances from CMLs.

In OpenMRG, the improvement in radar PCC after merging is substantial (from 0.55 to 0.70 with KED)~~. This aligns with findings from other studies; ~~, aligning with earlier findings — for example, Delrieu et al. (2014) reported radar PCC values ~~ranging from of~~ 0.4~~to 0.9, ~~0.9 improving to 0.8~~to 0.9 ~~0.9 with KED for hourly rainfall in France. In contrast, ~~the improvement in PCC is more modest in OpenRainER (from~~ OpenRainER shows only a modest PCC improvement (0.85 to 0.86), ~~possibly~~ reflecting the already high correlation between radar and reference data~~or the limits imposed by a sparser and less uniformly distributed CML network. However, when considering MAE , the ~~. The MAE pattern is reversed: OpenMRG shows a smaller ~~improvement (18% with KED), while OpenRainER exhibits a much larger MAE improvement (38% with KED). The larger reduction in MAE in OpenRainER indicates that merging was particularly effective at correcting magnitude errors and biases, even though the overall rainfall pattern(as measured by PCC) was already well captured by the radar. These differences highlight that merging scores are strongly dataset-dependent, reflecting the combined influence of terraincomplexity~~relative improvement while OpenRainER improves substantially, indicating that merging in OpenRainER primarily corrects magnitude errors rather than spatial pattern. These dataset-dependencies — driven by terrain, rainfall type, sensor distribution, and radar quality ~~. Radar-based rainfall estimates are subject to uncertainties such as beam blockage, attenuation, and calibration errors, ~~— make cross-study comparisons of merging scores challenging. Moreover, validation gauges themselves carry biases such as wind undercatch (Nešpor and Sevruk, 1999), further complicating evaluation.

5.3 Variogram estimation and its effect on merging performance

The variogram study (Fig. 3) reveals systematic differences between observation types and datasets. Variograms estimated from CML observations generally exhibit higher and more temporally variable nugget and range values than those estimated from rain gauge observations. In OpenMRG, the range estimated from CML observations is consistently larger than that from rain gauges, a pattern that also holds when comparing radar evaluated at CML midpoints versus radar evaluated at gauge locations. This points to a spatial extent effect rather than path-averaging. As Western and Blöschl (1999) show, the apparent correlation length is biased low when the sampling domain is small relative to the true correlation length. In OpenMRG, this effect is pronounced because rain gauges are confined to the city center while CMLs span the wider network. In contrast, for OpenRainER, CML- and ~~may also be influenced by wind-driven advection, which could misalign precipitation patterns and reduce correlation between radar and gauge measurements. While lower-quality radar products often benefit most from merging, very poor radar data may yield unreliable results. Consequently, comparisons across studies are challenging, as even similarmerging methods can produce different gains depending on the observational network and radar quality. Moreover, the rain gauges used for validation are themselves subject to biases such as wind undercatch Nešpor and Sevruk (1999), further complicating the evaluation~~gauge-estimated ranges are similar, suggesting that path-averaging does not substantially inflate the apparent correlation range for this dataset. The pronounced temporal variability in fitted parameters in both datasets highlights the difficulty of obtaining stable variogram estimates from limited networks.

The merging methods utilized here demonstrate considerable stability across different variogram parameters, likely because the adjustment or residual fields vary more gradually than raw rainfall. While specific range and sill values have a limited practical impact, the inclusion of a non-zero nugget is essential for maintaining this performance. A zero nugget forces exact interpolation at CML locations, which can lead to the amplification of local errors and artifacts—especially for multiplicative methods in sparse networks where local uncertainties are high. This is corroborated by Haberlandt (2007), who found that a simple linear variogram with zero nugget resulted in the largest errors (highest RMSE) among all tested models. This poor performance likely stems from the model's inability to account for both local measurement noise and the finite climatological range of rainfall events. Although we did not explicitly test a linear variogram with a zero nugget, it is possible that the lack of a nugget—rather than the linear model itself—was a primary contributor to the high errors observed in that study. In contrast, their more effective models utilized spherical structures and standardized nuggets between 0.1 and 0.4. Our results confirm that while a reference variogram serves as a robust default, a non-zero nugget is critical for reliable geostatistical merging.

5.4 Comparison of interpolation methods

For the merging methods (additive, multiplicative, and KED), different interpolation approaches were used. These approaches include IDW, point kriging, and block kriging (Table 1). Note that, while KED is fundamentally an interpolation method, it is treated here as a kriging (point and block) variant for clarity in the following discussion. In ~~OpenMRG, where the dataset has a high density of observations, interpolation methods play a smaller role , possibly due to that data abundance provides strong constraints.~~

~~In OpenRainER, where observation density is lower, differences between interpolation methods are somewhat more visible, though still modest. For additive merging and KED, point and block kriging perform similarly, which can be explained by~~ the ~~large correlation lengths in~~ high-density OpenMRG dataset, the ~~rainfall field and in the radar–CML offsets. These long correlation structures reduce the influence of line-averaging. Under block kriging, CML observations are integrated assuming second-order stationarity, so that as link length increases, their contribution converges toward the mean rainfall along the path. Consequently, even if a long CML measurement deviates from the true variability, its interpolated value tends toward the mean, reflecting the limited information available about the underlying distribution. This explains why line-averaging effects play only a minor role in overall performance.~~ choice of interpolation method plays a negligible role due to strong spatial constraints. In the lower-density OpenRainER dataset, differences are more discernible, though performance gains remain modest.

~~Generally, we would expect block kriging to perform best, followed by point kriging and lastly IDW. For multiplicative mergingthis expectation was confirmed: IDW performed the worst, point kriging was better,~~ Theoretically, block kriging should outperform point kriging by accounting for the line-integral nature of CMLs. However, both methods yield nearly identical results in this study. In the sparse OpenRainER dataset, this parity is likely because the massive 107% radar bias makes the path-mean magnitude the primary driver of performance, rendering geometry a secondary refinement. Conversely, in the dense OpenMRG network, the high density of observations in combination with a relatively high nugget value constrains the interpolated fields so effectively that the choice between point and block representations has little practical impact. Furthermore, the large correlation lengths (approximately 30 km) and high nugget values (approximately 0.3 mm/h) identified in the variogram study minimize the impact of line-averaging across both datasets, allowing a midpoint point-approximation to effectively represent the local rainfall state. Nevertheless, block kriging may offer a latent advantage through its uncertainty estimates near CML observations, which could be used to blend merged and radar-only fields (Lussana et al., 2025; Delrieu et al., 2014), though this was not demonstrated in the present study.

The advantages of kriging become more apparent in multiplicative merging. Here, IDW performed worst, while point and block kriging ~~achieved the best results in terms of RMSE. One reason is that multiplicative merging tends to produce~~ showed similar performance. Distance analysis reveals this gap is most pronounced near CMLs. Because multiplicative merging produces adjustment fields with ~~stronger gradients, making the nugget effect of krigingand the line-integral properties of block kriging more important. Another explanation is that block kriging dampens large values, as reflected in consistently lower bias values, which helps mitigate the overestimation often associated with multiplicative merging. For additive merging, point kriging, block kriging~~sharper gradients, kriging's spatial weighting and nugget effect likely provide essential stability that IDW lacks. Theoretically, IDW is less suited for heterogeneous networks like OpenRainER because it weights nearby observations based only on distance, whereas kriging's 'de-clustering' property assigns more weight to isolated observations while accounting for redundancy in clusters. However, since IDW performs on par with kriging for additive methods, it appears that the weighting of isolated observations is not a decisive advantage, at least for additive merging. For the more variable multiplicative ratio fields, ~~and IDW yielded comparable results in terms of RMSE, suggesting that the choice of interpolation method had limited impact in this context. Similarly, for KED, block kriging and point kriging performed on pair. This may be due to the properties of the hourly rainfall fieldsin this study, such as intermittency and anisotropy, or the high density of~~

observations, which can reduce differences between interpolation approaches. These effects could be better represented using more advanced rainfall simulators (Green et al., 2024; Leblois and Creutin, 2013; Blettner et al., 2022; Roksvåg et al., 2020), and by allowing rainfall statistics to vary as a function of linear and nonlinear relationships between the observations and external variables such as radar (KED, where the relationship is linear), rainfall across time steps (Sideris et al., 2014) or terrain (Ingebrigtsen et al., 2015)the performance gap is much wider; yet, it remains difficult to disentangle whether kriging's advantage stems from this spatial configuration or primarily from the nugget effect, as both likely contribute to the stability required to handle high variability.

Although the impact of block kriging varies across methods, it offers an advantage by improving the representation of kriging uncertainty. This uncertainty can be used to combine the different merged fields, enabling a smoother transition from adjusted fields near observations to radar-only fields in areas farther from CMLs (Lussana et al., 2025; Delrieu et al., 2014) Block kriging's lack of improvement upon ordinary kriging may stem from the fact that kriging estimates the conditional expectation. Consequently, point-scale variability remains elusive regardless of path-averaging corrections. Capturing more complex, intermittent, or anisotropic structures may require stochastic spatial reconstructions (Green et al., 2024; Leblois and Creutin, 2013) or non-linear covariate integration (Sideris et al., 2014; Ingebrigtsen et al., 2015).

5.5 What we learned from working with the data and further work

Adjusted fields performed best close to the CMLs, indicating that radar–CML bias varies spatially. This finding is consistent with earlier work (Goudenhoofdt and Delobbe, 2009), which showed that spatially varying adjustment methods outperform a single mean correction factor applied across the entire field. Such variability in radar bias can arise from differences in radar beam height, precipitation type, or terrain. Both radar and CMLs tend to overestimate precipitation in wet snow conditions, while dry snow is not detected by CMLs (Øydvin et al., 2025b). In OpenRainER, the impact of wet snow is partly reduced by capping radar values. However, this cap complicates merging, as it introduces nonlinearity in the relationship between radar and ground observations. Further works should explore how external data, such as atmospheric models or terrain data (which are often incorporated into model data Lussana et al. (2021)) could be combined to make merging CML more robust.

Range checks are essential, particularly for OpenRainER. Without them, merged fields often perform worse than radar alone due to the presence of outliers. The reason OpenMRG is less sensitive to range checks could be the high density CML network in combination with a few rain gauges in the city center, where outlier data are quickly canceled. Another explanation could be that the rainfall intensities are lower in the OpenMRG dataset, causing fewer large outliers. Standard range checksare optimal for additive and KED methods, while stricter checksimprove multiplicative methods. Stricter range checks reduce outliers but may exclude valid observations , as seen in OpenRainERThe heightened sensitivity of the OpenRainER dataset also reflects its specific network configuration; operating within a narrow frequency band (24.6–25.6 GHz), link sensitivity varies significantly with length, complicating the application of uniform range checks. The coarser signal quantization in OpenRainER (1 dB) may also contribute to differences in data quality, which could also explain the grater sensitivity to range checks. Further work should explore different quality control methods and The stringency of these checks also varies significantly by method: multiplicative range checks removed 37% of observations in OpenMRG and 53% in OpenRainER, whereas additive checks

removed less than 2%. This disparity could be driven by the inherent sensitivity of ratio-based checks to near-zero radar values, where even minor CML signals produce large ratios. Future efforts should investigate dynamic range checks ~~that can detect outliers for different rainfall intensities~~ capable of adapting to varying rainfall intensities and local network geometries.

The ~~CML processing workflow in this studydepends on the classification of wet periods , which is used to estimate the baseline attenuation of the CMLs. We used weather radar data to identify wet periods, as this simplifies the processing workflow and ensures that CML data is available whenever radar data is available. Radar-based wet/dry classification has also been shown to outperform CML-based methods (Overeem et al., 2011). A potential drawback of using radar data with high temporal resolution is that it may introduce short, intermittent wet periods in the CML time series, which could lead to an elevated baseline attenuation estimate (Øydvin et al., 2024)~~uncertainty of CML-derived rainfall estimates is driven by several processing steps, including wet-dry classification, baseline estimation, and wet antenna attenuation (WAA) correction (Chwala and Kunstmann, 2019). In this study, identifying wet periods via radar simplifies the workflow, but errors in the timing of wet-period boundaries or the introduction of intermittent radar artifacts can lead to elevated baseline estimates and biased rainfall retrievals (Øydvin et al., 2024). Because these processing strategies are applied across the entire network, they can introduce systematic biases rather than just random local noise. These errors propagate directly into the merged rainfall field. Since geostatistical merging assumes that ground-based observations represent the truth, a network-wide bias in CML data will systematically shift the radar estimates toward that bias. Consequently, a merging method that provides a more correct bias correction may paradoxically yield a result that is more wrong than a less sensitive method; by more effectively forcing the radar to match a biased CML reference, the more effective method shifts the final estimate further from the actual rainfall. While dense networks can average out random local artifacts, they remain vulnerable to these systemic processing errors, which can degrade performance regardless of observational density. Further work is needed to evaluate ~~the impact of~~ how radar-based ~~wet/dry detection on the merged rainfall fields~~ processing affects these systematic biases and to explore ~~how it can be effectively combined with CML-based wet/dry detection~~more robust methods for integrating CML data without propagating these inherent uncertainties.

While this study focused on neighborhood-based merging methods, many other combinations remain to be explored. Neighborhoods could be extended to include both short- and long-range components, as in Overeem et al. (2023), or alternative approaches such as the median-based adjustment proposed by Nielsen et al. (2024). Exploring these options would help clarify how different neighborhood definitions and filtering strategies influence merging performance across diverse datasets.

6 Conclusions

In this work, we developed `mergeplg`, an open-source Python package for merging radar and CML rainfall data, ~~and used it~~ to assess the impact of ~~this merging~~ diverse merging strategies across varying network densities. The package and accompanying notebooks are released to promote openness, reproducibility, and usability (Øydvin et al., 2025a; Covi et al., 2025). ~~Researchers using `mergeplg` are encouraged to cite this work.~~

Our analysis demonstrates that merging radar and CML data consistently improves rainfall estimates, especially when the CML network is dense and close to rain gauges, but the degree of improvement depends on terrain, rainfall type, and radar quality. In OpenMRG, the improvement in PCC after merging is substantial (from 0.55 to 0.70), while the reduction in MAE is more modest (~~18~~20%). In contrast, OpenRainER's sparser network saw only a slight increase in PCC (from 0.85 to 0.86) but a much larger reduction in MAE (~~38%~~ up to 41% for multiplicative merging), suggesting that while the overall rainfall pattern was already well captured by the radar, merging primarily served to correct magnitude errors and biases.

These results highlight that merging effectiveness is strongly dataset-dependent. ~~KED performs best near observations and for moderate rainfall~~While KED and additive methods perform comparably in data-dense settings, additive methods are more effective for capturing extremes~~and in data-scarce regions, and~~ . In data-sparse settings, multiplicative merging achieves the lowest ~~error in localized, data-rich settings but is more sensitive~~ overall RMSE and MAE, outperforming other methods in bias correction despite its sensitivity to outliers. While KED provides the next best performance in terms of RMSE and MAE, it demonstrates greater spatial stability than multiplicative merging, maintaining higher PCC at larger distances from CML observations. Overall, the choice of merging method should match the observational network density and environmental conditions to maximize performance.

~~Although differences between interpolation methods were generally modest~~Regarding geostatistical parameters, all methods proved robust to variogram choice provided a non-zero nugget was used. While kriging approaches outperformed IDW in multiplicative merging, block kriging showed ~~showed an advantage in multiplicative merging. It could also better represent kriging uncertainty, allowing for more correct transitions between adjusted fields near observations and radar-only fields in areas farther from CMLs~~no consistent advantage over point kriging. This suggests that path-averaging corrections alone do not overcome the smoothing inherent in conditional expectation estimates. Finally, we conclude that robust range checks are essential for stability, particularly for multiplicative methods where ratio-based checks are sensitive to near-zero radar values.

Overall, this work demonstrates the ~~potential~~ clear value of combining radar and CML data for improved rainfall estimation and offers guidance on selecting merging methods for both dense and sparse networks, supporting more accurate hydrometeorological applications.

Code and data availability. `mergeplg` is available at https://github.com/OpenSenseAction/mergeplg (Øydvin et al., 2025a). The intercomparison notebooks used for this study are available at https://github.com/OpenSenseAction/radar_adjustment_intercomparison (Covi et al., 2025). The OpenRainER dataset is available from Covi and Roversi (2025), and the OpenMRG dataset is available from Andersson et al. (2022).

Appendix A: Block Covariance from Point-Support Structure

We refer to spatially averaged observations, such as CML path-integrated measurements, as block data. Although our focus is on path-averaged observations, the methodology applies to blocks of arbitrary shape and orientation. These block averages are assumed to arise from an underlying second-order stationary spatial process $Z(x)$. Under second-order stationarity, the semi-variogram γ and covariance function C are related by:

$$\gamma(h \mid \theta) = C(0 \mid \theta) - C(h \mid \theta), \tag{A1}$$

where $C(h \mid \theta)$ is the covariance function and $C(0 \mid \theta)$ is the sill, equal to the sum of spatial variance and the variance at $h = 0$. The parameter vector θ defines the covariance model. Assuming the observation errors are not correlated, the diagonal entries of the covariance matrix can be modified by adding the estimated error for each observation (Cecinati et al., 2018), and the covariance matrix becomes:

$$C_{ij} = \begin{cases} c + c_i, & \text{if } i = j \\ C(v_i - v_j) = c - \gamma & \text{if } i \neq j \end{cases} \tag{A2}$$

where c is the spatial variance of the point process and c_i is the individual uncertainty of each observation v_i and v_j.

Next, as shown in (Journel and Huijbregts, 1978, p. 54), the expected squared difference between two block averages Z_{v_i} and Z_{v_j} can be expressed in terms of the underlying point-support covariance function:

$$\mathbb{E}[(Z_{v_i} - Z_{v_j})^2] = \bar{C}(v_{\underline{i}\underline{j}}, v_j) + \bar{C}(v_i, v_i) - 2\bar{C}(v_{\underline{j}\underline{i}}, v_j), \tag{A3}$$

where $\bar{C}(v_i, v_i)$ and $\bar{C}(v_j, v_j)$ are the average point covariances within blocks v_i and v_j, and $\bar{C}(v_i, v_j)$ is the average point covariance between them. Furthermore, the block-averaged process inherits second-order stationarity, allowing us to write a semi-variogram identity for block data:

$$\mathbb{E}[(Z_{v_i} - Z_{v_j})^2] = 2\left[C_b(0) - C_b(v_i, v_j)\right], \tag{A4}$$

$C_b(v_i, v_j)$ is the covariance function of the block averaged data and $C_b(0)$ is the sill, equal to the variance of the block averaged data. Combining Equation A3 and Equation A4, we obtain:

$$2\left[C_b(0) - C_b(v_i, v_j)\right] = \bar{C}(v_i, v_i) + \bar{C}(v_j, v_j) - 2\bar{C}(v_i, v_j), \tag{A5}$$

which leads to the block covariance expression:

$$C_b(v_i, v_j) = C_b(0) + \bar{C}(v_i, v_j) - \frac{1}{2}\left[\bar{C}(v_i, v_i) + \bar{C}(v_j, v_j)\right]. \tag{A6}$$

A common approach is to work with the semi-variogram when evaluating the correlation between observations (Murphy et al., 2025). We expand this approach to block kriging. Inserting Equation A6 into the ordinary kriging system (Equation 8) we get

$$\sum_{j=1}^{N} \lambda_j \{C_b(0) + \bar{C}(v_i, v_j) - \frac{1}{2}[\bar{C}(v_i, v_i) + \bar{C}(v_j, v_j)]\} + \mu = C_b(0) + \bar{C}(v_i, p) - \frac{1}{2}[\bar{C}(v_i, v_i) + \bar{C}(p, p)] \tag{A7}$$

Where v_i denotes block i and v_j denotes another block j. Following Equation A4 averaged covariance can be expressed in terms of averaged semi-variance, the sill cancels between the block-block and within-block covariances and we get

$$\sum_{j=1}^{N} \lambda_j \{C_b(0) - \bar{\gamma}(v_i, v_j) + \frac{1}{2}[\bar{\gamma}(v_i, v_i) + \bar{\gamma}(v_j, v_j)]\} + \mu = C_b(0) - \bar{\gamma}(v_i, p) + \frac{1}{2}[\bar{\gamma}(v_i, v_i) + \bar{\gamma}(p, p)], \tag{A8}$$

where p is the point we want to predict to. Following Equation A2, $C_b(0)$ on the r.h.s is not evaluated at $h = 0$, and is thus just the spatial variance of the data c added to all entries in the vector. $C_b(0)$ on the l.h.s contributes c to all cross-terms of the matrix and $c + c_i$ to the diagonal. Since $c\sum_{j=1}^{N} \lambda_j = c$ under $\sum_{j=1}^{N} \lambda_j = 1$, the constant c cancels on both sides, and we get the following system of equations:

$$\left[\begin{bmatrix} c_1 & 0 & \dots & 0 & 0 \\ 0 & c_2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & c_n & 0 \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix} - \begin{bmatrix} \bar{\gamma}(v_1, v_1) & \bar{\gamma}(v_1, v_2) & \dots & \bar{\gamma}(v_1, v_n) & 1 \\ \bar{\gamma}(v_2, v_1) & \bar{\gamma}(v_2, v_2) & \dots & \bar{\gamma}(v_2, v_n) & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \bar{\gamma}(v_n, v_1) & \bar{\gamma}(v_n, v_2) & \dots & \bar{\gamma}(v_n, v_n) & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} (\bar{\gamma}(v_1, v_1) + \bar{\gamma}(v_1, v_1)) & \dots & (\bar{\gamma}(v_1, v_1) + \bar{\gamma}(v_n, v_n)) & 0 & \\ (\bar{\gamma}(v_2, v_2) + \bar{\gamma}(v_1, v_1)) & \dots & (\bar{\gamma}(v_2, v_2) + \bar{\gamma}(v_n, v_n)) & 0 & \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ (\bar{\gamma}(v_n, v_n) + \bar{\gamma}(v_1, v_1)) & \dots & (\bar{\gamma}(v_n, v_n) + \bar{\gamma}(v_n, v_n)) & 0 & \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix} \right] \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \mu \end{bmatrix} = - \begin{bmatrix} \bar{\gamma}(v_1, p) \\ \bar{\gamma}(v_2, p) \\ \vdots \\ \bar{\gamma}(v_n, p) \\ 1 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} \bar{\gamma}(v_1, v_1) + \bar{\gamma}(p, p) \\ \bar{\gamma}(v_2, v_2) + \bar{\gamma}(p, p) \\ \vdots \\ \bar{\gamma}(v_n, v_n) + \bar{\gamma}(p, p) \\ 1 \end{bmatrix} \tag{A9}$$

Where $c_1, ...c_n$ is the uncertainty of each observation as defined in Equation A2. The within-block ($\bar{\gamma}(v_i, v_i)$) and between-block ($\bar{\gamma}(v_i, v_j)$) block-level variances are estimated by averaging the semi-variogram across all point pairs within and between blocks

$$\bar{\gamma}(v_i, v_j) = \frac{1}{P^2} \sum_{s=1}^{P} \sum_{s'=1}^{P} \gamma(\boldsymbol{u_s}, \boldsymbol{u_{s'}}), \tag{A10}$$

where P is the number of points used to discretize the two blocks v_i and v_j and the vectors $\boldsymbol{u_s}$ and $\boldsymbol{u_{s'}}$ represents the coordinates used to discretize the blocks. Note that in order to make the within-block and block-block variance cancel, the same number of discretization steps is needed in each block, regardless of block size.

Appendix B: Effect of neighborhood size

The Root Mean Square Error (RMSE) for all merging methods is plotted in Fig. B1. We observe that KED and multiplicative methods experience a small improvement when increasing the neighborhood size from 12 to 70, but there is little further

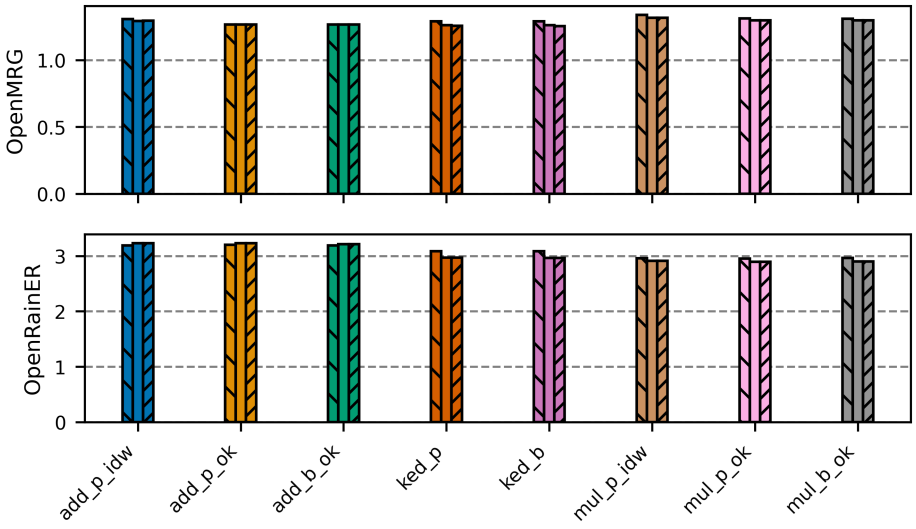


Figure B1. Root Mean Square Error (RMSE) for all merging methods under three neighborhood size configurations (12, 70, and 500 (all)) for the OpenMRG and OpenRainER dataset.

improvement when increasing the neighborhood to include all observations. For additive methods, their performance is more consistent across different neighborhood sizes.

Author contributions. Conceptualization: EØ, EC, MG, CC. Data curation: EØ, EC, MG, CC. Methodology: EØ, EC, CC. Software: EØ, EC, CC. Writing – Original Draft: EØ, EC, MG, CC. Writing – Review & Editing: EØ, EC, MG, CC.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Acknowledgements. This publication is based upon work from the COST Action "Opportunistic Precipitation Sensing Network" (OPENSENSE, ref. CA20136), supported by COST (European Cooperation in Science and Technology). This work was also supported by the following projects: CLIMAXPO (LIFE21-IPC-IT-LIFE CLIMAX PO-101069928) and MODMET agreement (years 2024/25) between Italian National Civil Department and Arpae-SIMC.

References

Aldana, M., Pulkkinen, S., von Lerber, A., Kumjian, M. R., and Moisseev, D.: Benchmarking KDP in rainfall: a quantitative assessment of estimation algorithms using C-band weather radar observations, Atmospheric Measurement Techniques, 18, 793–816, https://doi.org/10.5194/amt-18-793-2025, 2025.

Andersson, J. C. M., Olsson, J., van de Beek, R. C. Z. ., and Hansryd, J.: OpenMRG: Open data from Microwave links, Radar, and Gauges for rainfall quantification in Gothenburg, Sweden, Earth System Science Data, 14, 5411–5426, https://doi.org/10.5194/essd-14-5411-2022, 2022.

Blettner, N., Chwala, C., Haese, B., Hörning, S., and Kunstmann, H.: Combining Commercial Microwave Link and Rain Gauge Observations to Estimate Countrywide Precipitation: A Stochastic Reconstruction and Pattern Analysis Approach, Water Resources Research, 58, https://doi.org/10.1029/2022WR032563, 2022.

Blettner, N., Chwala, C., Wiegels, R., Lang, R., and Kunstmann, H.: Evaluating CML Rainfall Processing Methods on a Countrywide Scale in Zambia, Journal of Hydrometeorology, 26, 1603–1629, https://doi.org/10.1175/JHM-D-24-0157.1, 2025.

Cecinati, F., Moreno-Ródenas, A. M., Rico-Ramirez, M. A., claire Ten Veldhuis, M., and Langeveld, J. G.: Considering Rain Gauge Uncertainty Using Kriging for Uncertain Data, Atmosphere, 9, 446, https://doi.org/10.3390/atmos9110446, 2018.

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

Chwala, C., Graf, M., Øydvin, E., Walraven, B., and Seidel, J.: OpenSenseAction/poligrain, https://github.com/OpenSenseAction/poligrain, 2024.

Chwala, C., Graf, M., Polz, J., Blettner, N., DanSereb, Øydvin, E., keis f, overeem11, toufikshit, Kaleta, L., and yboose: pycomlink/pycomlink: v0.5.0, https://doi.org/https://doi.org/10.5281/zenodo.17304578, 2025a.

Chwala, C., Overeem, A., Øydvin, E., Petersson Wårdh, L., Seidel, J., Graf, M., Walraven, B., Covi, E., Habi, H. V., Fencl, M., de Vos, L., Giannetti, F., Green, A., O'Hara, T., Blettner, N., Keel, T., Schutz, G., El Hachem, A., Illich, N., Polz, J., Shit, T., Kaleta, L., Zulkarnaen, D., and Bareš, V.: Open-source tools for processing opportunistic rainfall sensor data: An overview of existing tools and the new OpenSense software packages poligrain, pypwsqc and mergeplg, EGUsphere, 2025, https://doi.org/10.5194/egusphere-2025-5438, 2025b.

Covi, E. and Roversi, G.: OpenRainER, https://doi.org/10.5281/zenodo.14731404, 2025.

Covi, E., Graf, M., Chwala, C., and Øydvin, E.: OpenSenseAction/radar_adjustment_intercomparison, https://github.com/OpenSenseAction/radar_adjustment_intercomparison, dOI will be added upon publication, 2025.

Covi, E., Roversi, G., and Nebuloni, R.: OpenRainER: an open-source dataset for studying the opportunistic sensing of rainfall in Emilia-Romagna, Italy, in preparation, 2026.

Cristiano, E., ten Veldhuis, M.-C., and van de Giesen, N.: Spatial and temporal variability of rainfall and their effects on hydrological response in urban areas – a review, Hydrology and Earth System Sciences, 21, 3859–3878, https://doi.org/10.5194/hess-21-3859-2017, 2017.

Delrieu, G., Wijbrans, A., Boudevillain, B., Faure, D., Bonnifait, L., and Kirstetter, P.-E.: Geostatistical radar–raingauge merging: A novel method for the quantification of rain estimation accuracy, Advances in Water Resources, 71, 110–124, https://doi.org/10.1016/j.advwatres.2014.06.005, 2014.

Djibo, M., Chwala, C., Graf, M., Polz, J., Kunstmann, H., and Zougmoré, F.: High-Resolution Rainfall Maps from Commercial Microwave Links for a Data-Scarce Region in West Africa, Journal of Hydrometeorology, 24, 1847–1861, https://doi.org/10.1175/JHM-D-23-0015.1, 2023.

Fencl, M., Nebuloni, R., C. M. Andersson, J., Bares, V., Blettner, N., Cazzaniga, G., Chwala, C., Colli, M., de Vos, L., El Hachem, A., Galdies, C., Giannetti, F., Graf, M., Jacoby, D., Victor Habi, H., Musil, P., Ostrometzky, J., Roversi, G., Sapienza, F., Seidel, J., Spackova, A., van de Beek, R., Walraven, B., Wilgan, K., and Zheng, X.: Data formats and standards for opportunistic rainfall sensors [version 2; peer review: 2 approved], Open Research Europe, 3, https://doi.org/10.12688/openreseurope.16068.2, 2024.

Goovaerts, P.: Kriging and Semivariogram Deconvolution in the Presence of Irregular Geographical Units, Mathematical Geosciences, 40, 101–128, https://doi.org/10.1007/s11004-007-9129-1, 2008.

Goudenhoofdt, E. and Delobbe, L.: Evaluation of radar-gauge merging methods for quantitative precipitation estimates, Hydrology and Earth System Sciences, 13, 195–203, https://doi.org/10.5194/hess-13-195-2009, 2009.

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

Graf, M., Hachem, A. E., Eisele, M., Seidel, J., Chwala, C., Kunstmann, H., and Bárdossy, A.: Rainfall estimates from opportunistic sensors in Germany across spatio-temporal scales, Journal of Hydrology: Regional Studies, 37, 100 883, https://doi.org/10.1016/j.ejrh.2021.100883, 2021.

Green, A. C., Kilsby, C., and Bárdossy, A.: A framework for space–time modelling of rainfall events for hydrological applications of weather radar, Journal of Hydrology, 630, 130 630, https://doi.org/10.1016/j.jhydrol.2024.130630, 2024.

Haberlandt, U.: Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event, Journal of Hydrology, 332, 144–157, https://doi.org/10.1016/j.jhydrol.2006.06.028, 2007.

Hänsler, A. and Weiler, M.: Enhancing the usability of weather radar data for the statistical analysis of extreme precipitation events, Hydrology and Earth System Sciences, 26, 5069–5084, https://doi.org/10.5194/hess-26-5069-2022, 2022.

Ingebrigtsen, R., Lindgren, F., Steinsland, I., and Martino, S.: Estimation of a non-stationary model for annual precipitation in southern Norway using replicates of the spatial field, Spatial Statistics, 14, 338–364, https://doi.org/10.1016/j.spasta.2015.07.003, 2015.

ITU: ITU-R P.838-3: Specific attenuation model for rain for use in prediction methods, International Telecommunication Union, Geneva, Switzerland, pp. 1–8, https://www.itu.int/dms_pubrec/itu-r/rec/p/R-REC-P.838-3-200503-I!!PDF-E.pdf, 2005.

Jewell, S. A. and Gaussiat, N.: An assessment of kriging-based rain-gauge–radar merging techniques, Quarterly Journal of the Royal Meteorological Society, 141, 2300–2313, https://doi.org/10.1002/qj.2522, 2015.

Journel, A. G. and Huijbregts, C. J.: Mining geostatistics, The Blackburn Press, 1978.

Leblois, E. and Creutin, J.-D.: Space-time simulation of intermittent rainfall with prescribed advection field: Adaptation of the turning band method, Water Resources Research, 49, 3375–3387, https://doi.org/10.1002/wrcr.20190, 2013.

Lussana, C., Nipen, T. N., Seierstad, I. A., and Elo, C. A.: Ensemble-based statistical interpolation with Gaussian anamorphosis for the spatial analysis of precipitation, Nonlinear Processes in Geophysics, 28, 61–91, https://doi.org/10.5194/npg-28-61-2021, 2021.

Lussana, C., Nipen, T. N., Ménétrier, B., and Seierstad, I. A.: Ensemble-based statistical interpolation of atmospheric variables near the surface, Quarterly Journal of the Royal Meteorological Society, https://doi.org/10.1002/qj.5046, 2025.

Marshall, J. S. and Palmer, W. M. K.: The distribution of raindrops with size, Journal of Meteorology, 5, 165–166, https://doi.org/10.1175/1520-0469(1948)005<0165:TDORWS>2.0.CO;2, 1948.

Messer, H. and Sendik, O.: A New Approach to Precipitation Monitoring: A critical survey of existing technologies and challenges, IEEE Signal Processing Magazine, 32, 110–122, https://doi.org/10.1109/MSP.2014.2309705, 2015.

Murphy, B., Yurchak, R., and Müller, S.: GeoStat-Framework/PyKrige: v1.7.3, https://doi.org/https://doi.org/10.5281/zenodo.17372225, 2025.

Nešpor, V. and Sevruk, B.: Estimation of Wind-Induced Error of Rainfall Gauge Measurements Using a Numerical Simulation, Journal of Atmospheric and Oceanic Technology, 16, 450–464, https://doi.org/10.1175/1520-0426(1999)016<0450:EOWIEO>2.0.CO;2, 1999.

Nielsen, J., van de Beek, C., Thorndahl, S., Olsson, J., Andersen, C., Andersson, J., Rasmussen, M., and Nielsen, J.: Merging weather radar data and opportunistic rainfall sensor data to enhance rainfall estimates, Atmospheric Research, 300, 107 228, https://doi.org/10.1016/j.atmosres.2024.107228, 2024.

O, S. and Foelsche, U.: Assessment of spatial uncertainty of heavy rainfall at catchment scale using a dense gauge network, Hydrology and Earth System Sciences, 23, 2863–2875, https://doi.org/10.5194/hess-23-2863-2019, 2019.

Olsson, J., Horváth-Varga, L., van de Beek, R., Graf, M., Overeem, A., Szaton, M., Bareš, V., Bezak, N., Chwala, C., Michele, C. D., Fencl, M., Seidel, J., and Todorović, A.: How Close Are Opportunistic Rainfall Observations to Providing Societal Benefit?, Journal of Hydrometeorology, 26, 1585–1602, https://doi.org/10.1175/JHM-D-25-0043.1, 2025.

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

Overeem, A., Leijnse, H., van Leth, T. C., Bogerd, L., Priebe, J., Tricarico, D., Droste, A., and Uijlenhoet, R.: Tropical rainfall monitoring with commercial microwave links in Sri Lanka, Environmental Research Letters, 16, 074 058, https://doi.org/10.1088/1748-9326/ac0fa6, 2021.

Overeem, A., van den Besselaar, E., van der Schrier, G., Meirink, J. F., van der Plas, E., and Leijnse, H.: EURADCLIM: the European climatological high-resolution gauge-adjusted radar precipitation dataset, Earth System Science Data, 15, 1441–1464, https://doi.org/10.5194/essd-15-1441-2023, 2023.

Overeem, A., Leijnse, H., de Vos, L., and Silver, M.: RAINLINK, https://doi.org/10.5281/zenodo.12211069, 2024a.

Overeem, A., Leijnse, H., van der Schrier, G., van den Besselaar, E., Garcia-Marti, I., and de Vos, L. W.: Merging with crowd-sourced rain gauge data improves pan-European radar precipitation estimates, Hydrology and Earth System Sciences, 28, 649–668, https://doi.org/10.5194/hess-28-649-2024, 2024b.

Overeem, A., Walraven, B., Leijnse, H., and Uijlenhoet, R.: Four-year commercial microwave link dataset for the Netherlands, https://doi.org/10.4121/be252844-b672-471e-8d69-27269a862ec1.v1, 2024c.

Pastorek, J., Fencl, M., Rieckermann, J., and Bares, V.: Precipitation Estimates From Commercial Microwave Links: Practical Approaches to Wet-Antenna Correction, IEEE Transactions on Geoscience and Remote Sensing, 60, 1–9, https://doi.org/10.1109/TGRS.2021.3110004, 2022.

Roksvåg, T., Steinsland, I., and Engeland, K.: Estimation of annual runoff by exploiting long-term spatial patterns and short records within a geostatistical framework, Hydrology and Earth System Sciences, 24, 4109–4133, https://doi.org/10.5194/hess-24-4109-2020, 2020.

Rombeek, N., Hrachowitz, M., Droste, A., and Uijlenhoet, R.: Evaluation of high-intensity rainfall observations from personal weather stations in the Netherlands, Hydrology and Earth System Sciences, 29, 4585–4606, https://doi.org/10.5194/hess-29-4585-2025, 2025.

Schleiss, M., Olsson, J., Berg, P., Niemi, T., Kokkonen, T., Thorndahl, S., Nielsen, R., Nielsen, J. E., Bozhinova, D., and Pulkkinen, S.: The accuracy of weather radar in heavy rain: a comparative study for Denmark, the Netherlands, Finland and Sweden, Hydrology and Earth System Sciences, 24, 3157–3188, https://doi.org/10.5194/hess-24-3157-2020, 2020.

Sideris, I. V., Gabella, M., Erdin, R., and Germann, U.: Real-time radar–rain-gauge merging using spatio-temporal co-kriging with external drift in the alpine terrain of Switzerland, Quarterly Journal of the Royal Meteorological Society, 140, 1097–1111, https://doi.org/10.1002/qj.2188, 2014.

Uijlenhoet, R., Overeem, A., and Leijnse, H.: Opportunistic remote sensing of rainfall using microwave links from cellular communication networks, WIREs Water, 5, https://doi.org/10.1002/wat2.1289, 2018.

van de Beek, C., Leijnse, H., Torfs, P., and Uijlenhoet, R.: Seasonal semi-variance of Dutch rainfall at hourly to daily scales, Advances in Water Resources, 45, 76–85, https://doi.org/10.1016/j.advwatres.2012.03.023, 2012.

Velasco-Forero, C. A., Sempere-Torres, D., Cassiraga, E. F., and Gómez-Hernández, J. J.: A non-parametric automatic blending methodology to estimate rainfall fields from rain gauge and radar data, Advances in Water Resources, 32, 986–1002, https://doi.org/10.1016/j.advwatres.2008.10.004, 2009.

Villarini, G. and Krajewski, W. F.: Review of the Different Sources of Uncertainty in Single Polarization Radar-Based Estimates of Rainfall, Surveys in Geophysics, 31, 107–129, https://doi.org/10.1007/s10712-009-9079-x, 2010.

Western, A. W. and Blöschl, G.: On the spatial scaling of soil moisture, Journal of Hydrology, 217, 203–224, https://doi.org/10.1016/S0022-1694(98)00232-7, 1999.

Øydvin, E., Graf, M., Chwala, C., Wolff, M. A., Kitterød, N.-O., and Nilsen, V.: Technical note: A simple feedforward artificial neural network for high-temporal-resolution rain event detection using signal attenuation from commercial microwave links, Hydrology and Earth System Sciences, 28, 5163–5171, https://doi.org/10.5194/hess-28-5163-2024, 2024.

Øydvin, E., Chwala, C., Covi, E., and Graf, M.: OpenSenseAction/mergeplg: v0.1.0, https://github.com/OpenSenseAction/mergeplg, DOI will be added upon publication, 2025a.

Øydvin, E., Gaban, R., Andersson, J., van de Beek, R. C. Z., Wolff, M. A., Kitterød, N.-O., Chwala, C., and Nilsen, V.: Combining commercial microwave links and weather radar for classification of dry snow and rainfall, Atmospheric Measurement Techniques, 18, 2279–2293, https://doi.org/10.5194/amt-18-2279-2025, 2025b.