

Review of egusphere-2025-6349

Title: Scalable radar-driven approach with compact gradient-boosting models for gap filling in high-resolution precipitation measurements

Authors: Peter Lünenschloß, Antje Claussnitzer, Thomas Schartner, Mirjam Brunner, Timo Houben, David Schäfer, and Jan Bumberger

Summary

This manuscript presents a novel and practical AI/machine-learning framework that uses radar-derived QPE as predictors for sub-hourly surface rain gauge imputation, with the primary goal of addressing missing data and temporal gaps in surface precipitation observations. The use of a compact gradient-boosting model is particularly appealing given its computational efficiency. Overall, the authors demonstrate encouraging statistical predictive skill when relating radar-derived QPE to the DWD rain gauge network, highlighting the promise of AI-based approaches for improving precipitation datasets. Although, there are several areas where the manuscript would benefit from additional clarification and strengthening. Further discussion of the radar data quality and associated uncertainties is needed, especially given the known limitations of QPE in convective and high-impact precipitation regimes. The robustness of the statistical analysis could also be improved through additional event-based evaluation and clearer justification of methodological choices. Furthermore, several sections of the manuscript would benefit from expanded explanation. My comments are largely minor in nature and are primarily intended to improve the clarity, rigor, and overall presentation. Pending satisfactory revision of these points, I recommend this manuscript for publication in EGUSphere.

Major comments

- 1) Section 2.2: Please provide additional detail regarding the 17 dual-polarization DWD weather radars, including radar band (S-, C-, or X-band), range and azimuthal gate spacing, effective beamwidth, maximum range, and the extent to which the radar beams clear the local horizon (e.g., lowest usable elevation angle, maximum elevation angle, and terrain blockage considerations). These characteristics can substantially influence QPE quality and uncertainty. Radar-derived QPE is often sensitive to factors such as beam geometry, attenuation, beam blockage, storm mode, bright band contamination, and sampling resolution, all of which may lead to over- or underestimation of rainfall. Given that the radar-based QPE retrievals are a major component of this manuscript, the current description does not provide enough detail about the radar systems themselves or the quality-control and validation methodologies used to ensure reliable precipitation estimates across different storm types and geographic regions. These are only slightly mentioned later in the manuscript and often glossed over (e.g., lines 280-290, 409-417). Elaborate how you all combat these issues with these specific radars.

- 2) Lines 367-368: Why were observations from 2024 specifically selected for the final model evaluation? Was this year chosen randomly (which it should be), or was there a particular rationale for its selection? Additionally, do the results shown in Table 2 remain similar if a different evaluation year is used? I am also somewhat concerned that the statistics presented in Table 2 may obscure model performance for the most societally impactful events, such as high-frequency and/or high-accumulation precipitation cases, because these events are averaged together with non-precipitation or weak-precipitation periods. Since extreme convective rainfall events are often of greatest socioeconomic importance, it would strengthen the manuscript to include a more targeted evaluation of AI/machine learning performance during these high-impact events. For example, it would be helpful to examine how the reported statistics change when only strongly convective or heavy-rainfall cases are considered. This concern also relates to Lines 427–430, where the authors note reduced robustness for such events. However, the current figures and tables do not clearly quantify or highlight this degradation in performance, making the issue easy to overlook.

Minor comments

- 1) A few times the authors describe the 10-minute precipitation dataset (or even data with a frequency of < 30 min; line 54) as “high-frequency” or “high-resolution” (e.g., lines 2, 13, 17, etc.). Some may not consider 10-min data to be truly high-frequency given that in-situ instruments and radar observations can provide data at 1-min, 10-Hz, or even finer temporal resolution. Consider rewording this terminology or clarifying the context in which the dataset is considered high-frequency and/or high-resolution. For example, the authors could emphasize that the description is relative to the temporal and spatial scales typically available in climatological or gridded precipitation datasets.
- 2) Line 3: Are data gaps truly only from “sensor malfunctions”? The authors do go into more detail in lines 27-31 about what causes data gaps, and not all are due to sensor malfunctions (i.e., outages, quality control, human error, etc.).
- 3) Line 57: The DWD acronym has not been introduced yet until line 60. Replace with “the German Weather Service (DWD) national gauge network” in line 57 and remove from line 60.
- 4) Line 79: Is RADOLAN an acronym?

- 5) Line 81: Is QPECommon an acronym? Lines 146-147 introduces the acronym. Please introduce here and just keep the acronym at lines 146-147.
- 6) Line 134: Go into detail about the precipitation accumulation instrument, such as type (e.g., tipping-bucket), response times, etc.
- 7) Fig. 2: What is being plotted here? Pearson correlation coefficient (r) or coefficient of determination (R^2)? Explain how the correlations were calculated and include these into the text of the manuscript, too.
- 8) Line 232: Is ANNs an acronym?
- 9) Lines 360-365: What is the total run process time (end-to-end) for all 1100 gauges? 15 minutes seems like a relatively long time to process just one station and one variable (accumulated precip). Please elaborate further on how this process is more cost effective than running on a high-performance computer.
- 10) Lines 409-412: Duplicate sentences.