Universität Stuttgart
Institute for Modelling Hydraulic and Environmental Systems (IWS)
Stochastic Simulation and Safety Research for Hydrosystems (LS3)

A benchmarking framework for PWS quality control algorithms

Damaris Zulkarnaen & Jochen Seidel

Introduction

In recent years personal weather stations (PWS) have outnumbered official stations from national met services. PWS present a valuable data source, providing precipitation data of high temporal resolution in a dense network.

As PWS rain gauges are not professionally placed and maintained, a thorough quality control (QC) prior to the application of PWS data is essential. In recent years several QC frameworks have been developed. Those are: first, the **pypwsqc** package (Chwala et al., 2026), which was developed in particular as QC for PWS networks and includes algorithms developed by de Vos et al. (2019) and Bárdossy et al. (2021); and second, **RainfallQC** applied as adaptable **rulebase**, which covers the GSDR-QC framework developed by Lewis et al. (2021).

To evaluate the effectiveness of the existing methods, we aim at developing a systematic benchmarking framework. The approach applies individual and sequential quality control (QC) processes to the PWS data. The QC´d PWS data will then be evaluated against data from a nearby reference station.

Data

For our tests we used three datasets. Those are located in: the Amsterdam metropolitan area in the Netherlands (dataset from de Vos 2019), the Lazio region in central Italy, and the area of Reutlingen in southwestern Germany.

Location	Data Period	N PWS	N Reference	Comment on Reference
Amsterdam	5.2016 – 6.2018	134	20	random pixels from radar
Lazio	1.2015 – 12.2022	339	230	HydroNet gauges
Reutlingen	1.2020 – 12.2025	141	21	12 Pluvio, 3 DWD, 6 Agrarm.

Benchmarking Framework

The framework currently includes functions that:

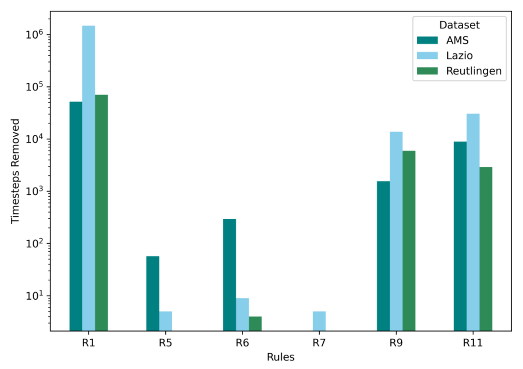
- Load and ensure data format (dims order, var names, x,y projection)
- Resample and apply flags („flex“ & „strict“ mode)
- Create polars dfs for RainfallQC and rulebase application from .nc
- Apply several different QC pipelines
- Run an evaluation for close PWS-Ref pairs

Pipelines tested

1. **pypwsqc**: faulty zeros, high influx, „flex“ resample, peak removal
2. **pypwsqc_ref**: pypwsqc + gamma bias correction + indicator correlation
3. **rulebase**: rules R1, R5, R6, R7, R9, R11
4. **pypwsqc_rulebase**: pypwsqc + rulebase

The pypwsqc functions were used with recommended default parameters. The rulebase was adapted to the following:

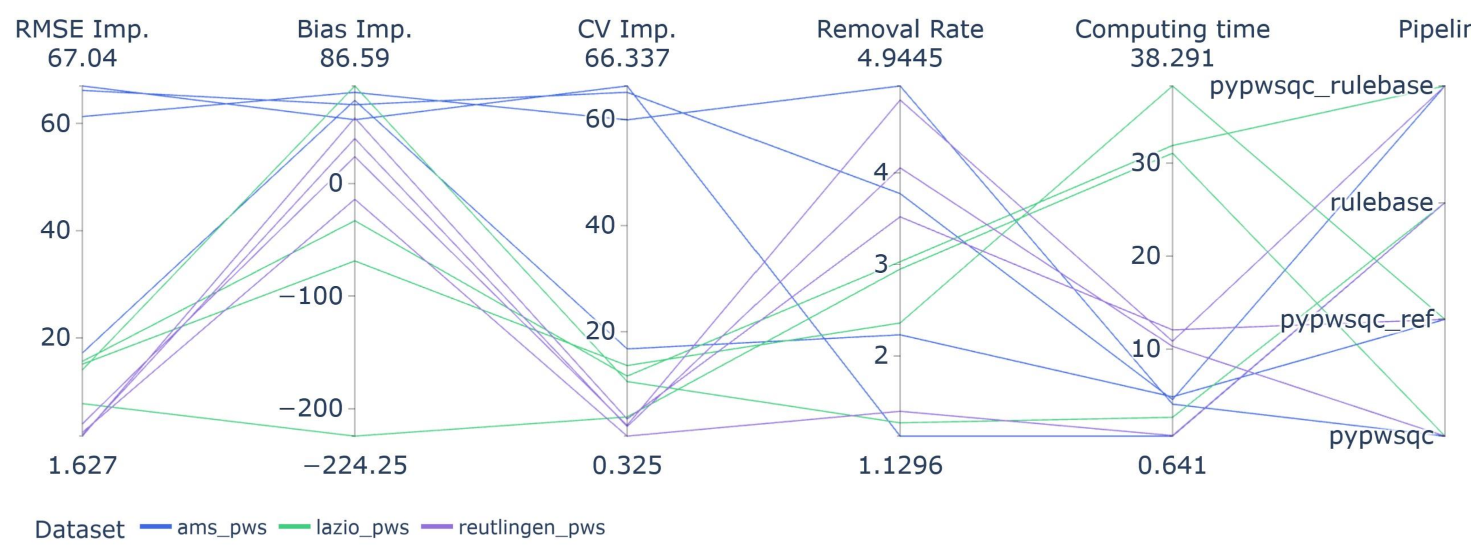
Rule	Description: Rule removes
R1	whole years if k-largest is zero (k=12)
R5	timesteps exceeding world record rainfall
R6	timesteps exceeding stations Rx1day threshold
R7	timesteps during wet spells if flag ==2
R9	timesteps if hourly dry spell flag & CDD for that station is exceeded
R11	month of data if sum is >= 2x wettest neighbour

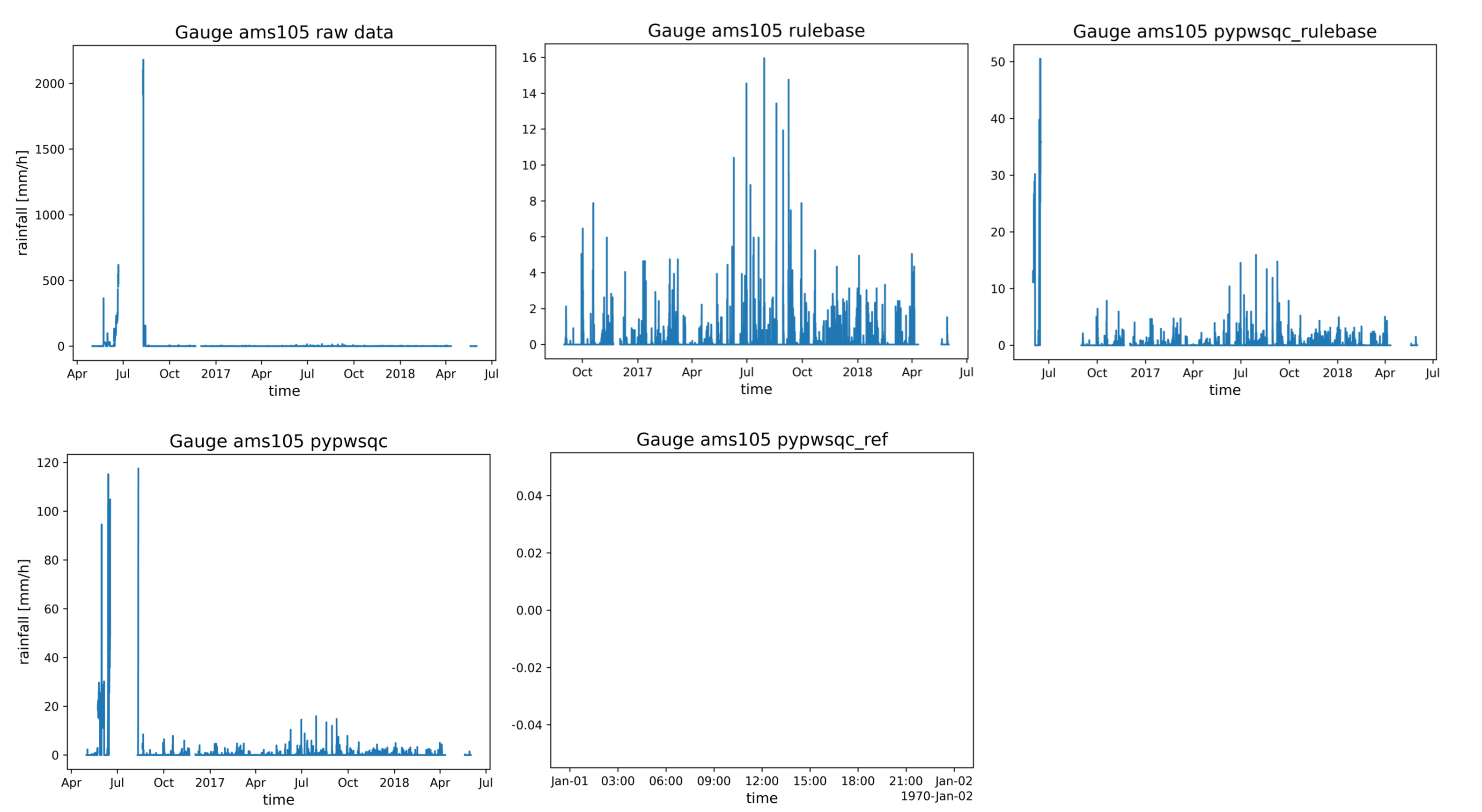


Evaluation Metrics

- Pearson r
- RMSE
- Bias
- CV (of residuals)
- Computing time
- Removal Rate
- Valid timesteps
- Empty stations
- Max rainfall
- Total rain sum
- Valid timesteps
- Improvement [%] / metric

Preliminary Results





In the ams105 example the RainfallQC rulebase yields the best results, while the pypwsqc_ref pipeline completely removed the stations data (not passing IC filter). Several pipelines that include pypwsqc methods show that data were removed, however not all of the high influxes. Pypwsqc methods have a long computing time for the large Lazio dataset. Plotting the percentage improvement of a metric can be misleading because the bias improvement of -224.25% corresponds to a bias of 0.0014 and 0.0046 before and after QC. Similar QC results for different QC pipelines of the same dataset can be observed.

Open Questions & Outlook

- Test on other regions, datasets, parameter sets
- Develop new pipelines and QC rules/ methods
- What are the effects of parameter
- Effect of evaluating different time resolutions (RainfallQC uses 15min, either 100% or 66% completeness from 5 min data?)
- Assumed high quality of the reference justified?
- Comparison by absolute metric values pre and post QC

Conclusions

So far no general pipeline that performs well among all metrics and datasets could be found. Further development of the tool and the QC methods is needed.

RainfallQC

References

de Vos, L. W., Leijnse, H., Overeem, A., and Uijlenhoet, R.: Quality control for crowdsourced personal weather stations to enable operational rainfall monitoring, Geophysical Research Letters, 46, 8820–8829, 2019. DOI:10.1029/2019GL083731
Bardossy, A., Seidel, J., El Hachem, A.:The use of personal weather station observations to improve precipitation estimation and interpolation, Hydrology and Earth System Sciences, 25, 583-601, 2021. https://doi.org/10.5194/hess-25-583-2021
Chwala, C. et al.: Open-source tools for processing opportunistic rainfall sensor data: An overview of existing tools and the new opensense software packages poligrain, pypwsqc and mergeplg. Submitted to Hydrology and Earth System Sciences, 2026.
Lewis, E., Pritchard, D., Villalobos-Herrera, R., Blenkinsop, S., McClean, F., Guerreiro, S., Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Rustemeier, E., Fowler, H. J.: Quality control of a global hourly rainfall dataset, Environmental Modelling & Software, 144, 2021. https://doi.org/10.1016/j.envsoft.2021.105169