University of Stuttgart

SMHI

LUND UNIVERSITY | LTH FACULTY OF ENGINEERING

KIT-Campus Alpin
IMK-IFU: Atmospheric Environmental Research

OpenSense

Recent Developments in the Quality Control of Personal Weather Station Data

Jochen Seidel (1), Louise Petersson Wårdh (2,3), Nicholas Illich (1) and Christian Chwala (4)

Introduction

The use of personal weather stations (PWS) has gained much attention in recent years, as they clearly outnumber professional rain gauges. However, the data quality of such sensors is typically low and thus their information cannot be used without thorough quality control (QC).

Various QC algorithms for PWS rainfall data have been developed and published within the EU COST Action 20136 "Opportunistic Precipitation Sensing Network" (OpenSense) in the past years and are available on OpenSense's GitHub repository (El Hachem et al. 2024). These have been updated and are now available as an installable Python package (`pypwsqc`).

New functions for these QC filters include :

(1) an improved indicator correlation filter (Bárdossy et al., 2021) which now provides a skill score for the accepted PWS to assess quality of the indicator correlation with neighbouring references

(2) an algorithm to correct rainfall peaks in PWS data which may be caused by connection interruptions between the rain gauge and the base station

(3) a Python implementation of the QC algorithms originally developed in R by de Vos et al. (2019).

References

El Hachem, A., Seidel, J., O'Hara, T., Villalobos Herrera, R., Overeem, A., Uijlenhoet, R., Bárdossy, A., and de Vos, L.W (2024), Technical note: A guide to using three open-source quality control algorithms for rainfall data from personal weather stations, Hydrol. Earth Syst. Sci., 28, 4715–4731.

Bárdossy, A., Seidel, J., and El Hachem, A. (2021), The use of personal weather station observations to improve precipitation estimation and interpolation. Hydrol. Earth Syst. Sci., 25, 583–601.

de Vos, L.W., Leijnse, H.,Overeem, A., and Uijlenhoet, R. (2019), Quality control for crowdsourced personal weather stations to enable operational rainfall monitoring. Geophysical Research Letters, 46, 8820–8829.

(1) Updated Indicator Correlation Filter

The Indicator Correlation Filter, initially developed by Bárdossy et al. (2021), now includes a skill score that uses a normalised rank sum weighting method for range bins. Higher weights are assigned to bins with shorter ranges. Additionally, a plotting feature for the acceptance threshold (indicated by the black line in the figures below) has been added, enabling visual inspection of the results. This acceptance threshold can be adjusted by the user. Figures 1.1 to 1.3 illustrate examples of various Indicator Correlation Skill Scores (ICSS) of accepted PWS.

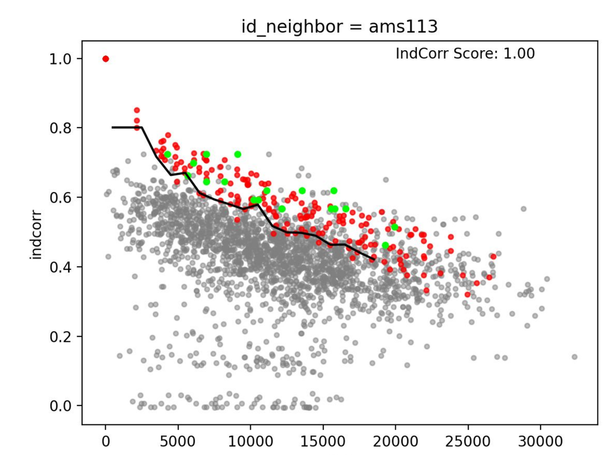


Fig 1.1: Example of a PWS with a perfect skill score as the indicator correlation in all ranges bins is at least as good as the reference

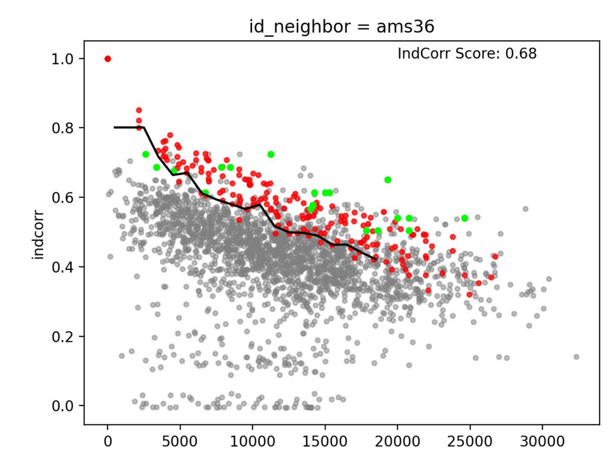


Fig 1.2: Example of a PWS with a mediocre skill score. The indicator correlation for the shorter range bins is lower the the reference

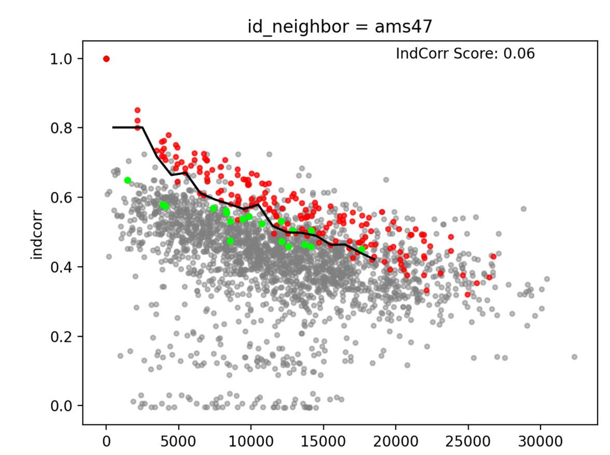


Fig 1.3: Example of a PWS with low skill score. Here, the indicator correlation is below the reference except for the distant range bins which have a low weight resulting in a low skill score

Indicator Correlation Skill Score

$$ICSS = \frac{\sum_{i=1}^{n} b_i \cdot \alpha_i}{\sum_{i=1}^{n} j_i \cdot \alpha_i}$$

$$b_i = \begin{cases} 1 & \text{indicator correlation above threshhold} \\ 0 & \text{else} \end{cases}$$

$$j_i = \begin{cases} 1 & \text{if } b_i \in \mathbb{R} \\ 0 & \text{else} \end{cases}$$

$$\alpha_i = \frac{m+1-i}{\sum_{k=i}^{m} k}, i = 1, 2, \ldots, m$$

$ICSS$ = Indicator Correlation Skill Score
b_i = range bin
α_i = weight for range bin i
m = number of range bins

(2) Peak Removal Filter*

Interruptions between the Netatmo rain gauge and the base station can result in artificial rainfall peaks, as the data collected during the disconnection is sent at the next timestamp once the connection is restored (Fig. 2.1). Although the total rainfall amount remains accurate, the timing of the rainfall data is incorrect. The filter identifies peaks that are preceded by NaN values (Fig. 2.2) and disaggregates them using data from nearby stations (Fig. 2.3).

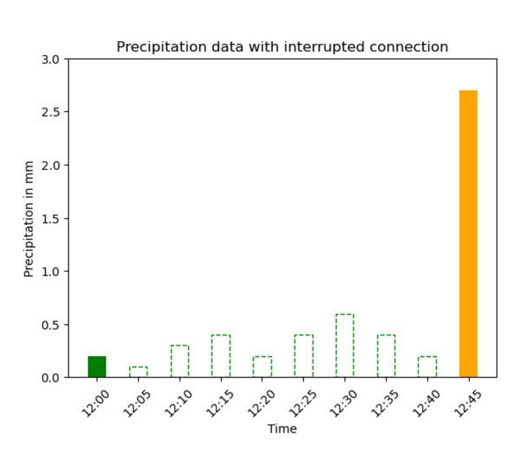


Fig 2.1: The dashed green bins show the true rainfall during an interrupted connection. These are NaN in the PWS data. The orange bin shows the data as it is transmitted after the connection is reestablished.

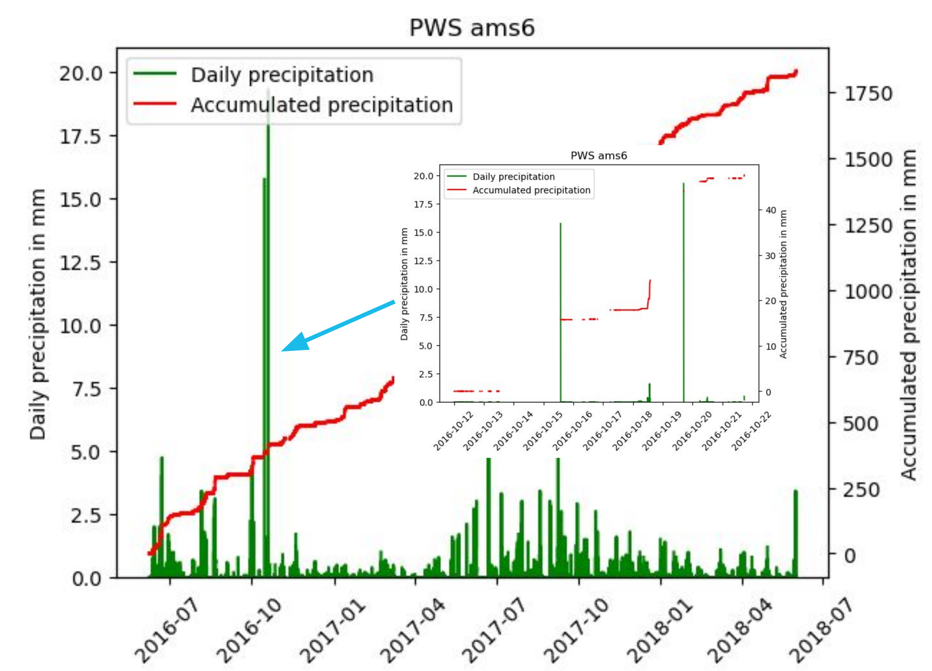


Fig 2.2: Example of a PWS with two peaks in October 2016 caused by connection interruptions.

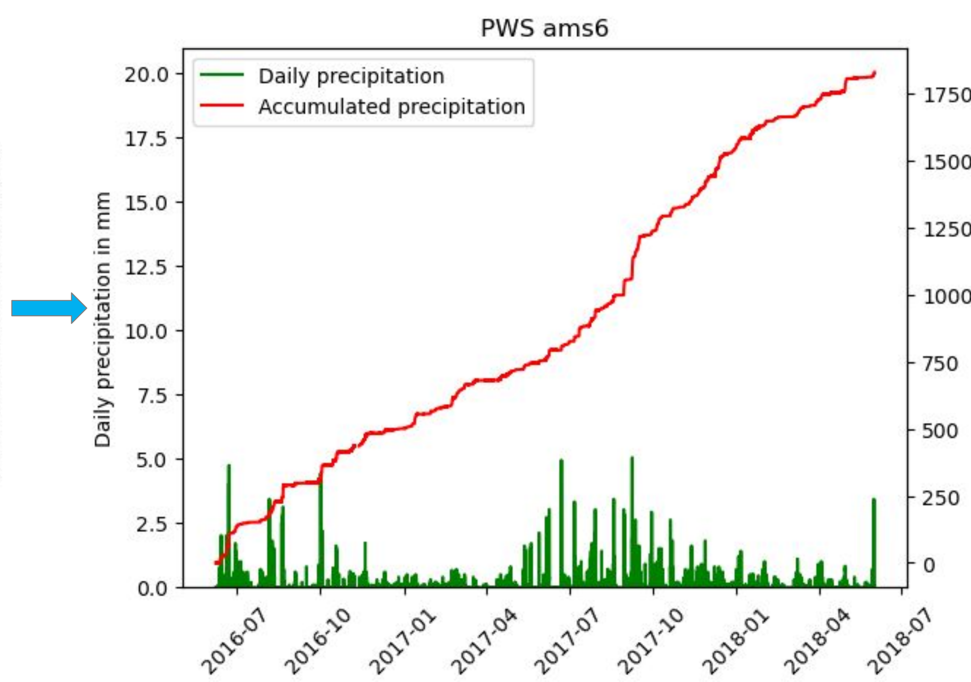


Fig 2.3: PWS rainfall time series after the application of the peak removal filter.

* not available on GitHub/implemented in `pypwsqc` yet

(3) Python implementation of faulty zeroes, high influxes and station outliers filter

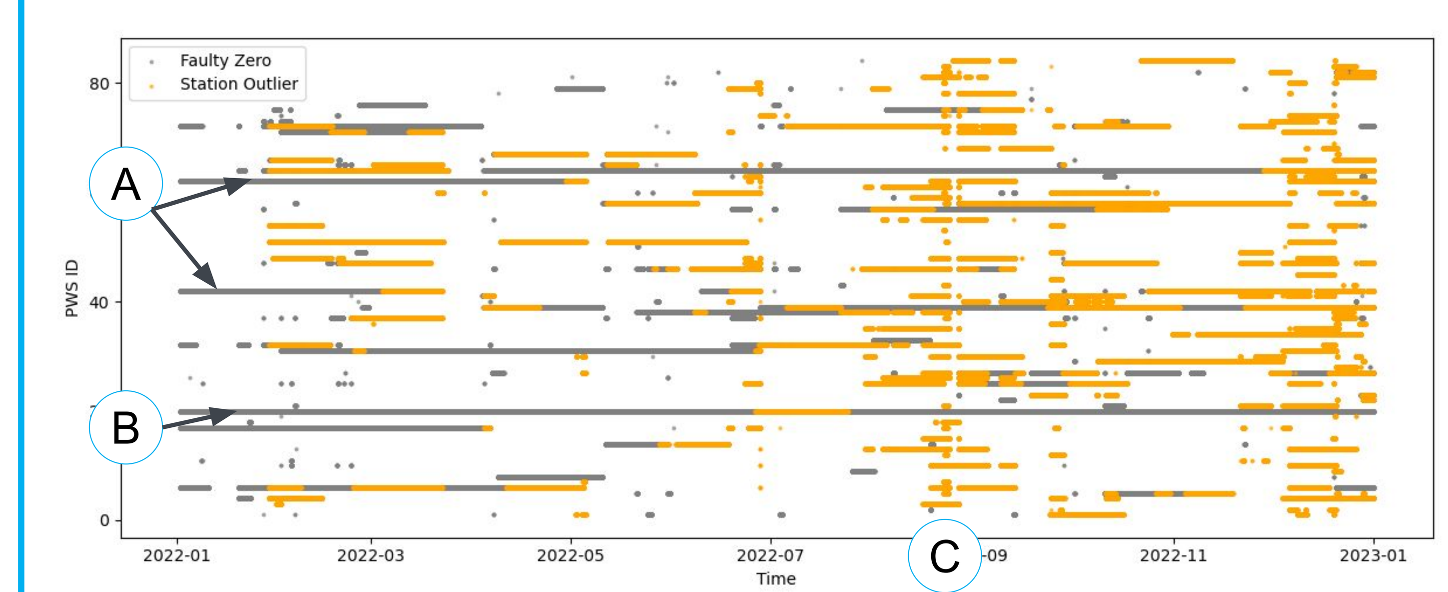
The QC protocol originally developed by de Vos et al. (2019) is now available in a Python package (`pypwsqc`). The algorithm assess the performance of individual PWS, per time step, based on the behaviour of neighbouring PWS within a user-specified range. The protocol consists of three filters:

Faulty Zeroes: Erroneous measurement of zero rainfall

High influxes: Unrealistically high rainfall rates

Station outliers: Low rolling correlation with neighbours over a user-specified period.

Example data set of 85 PWS (Fig 3.1) in southwestern Sweden, 2022:



A PWS placed indoors during winter and outdoors in the summer months

B PWS continuously measuring zero rainfall throughout 2022

C Convective storm in August triggered many station outliers, which highlights the problem with neighbour-checks for spatially variable rainfall

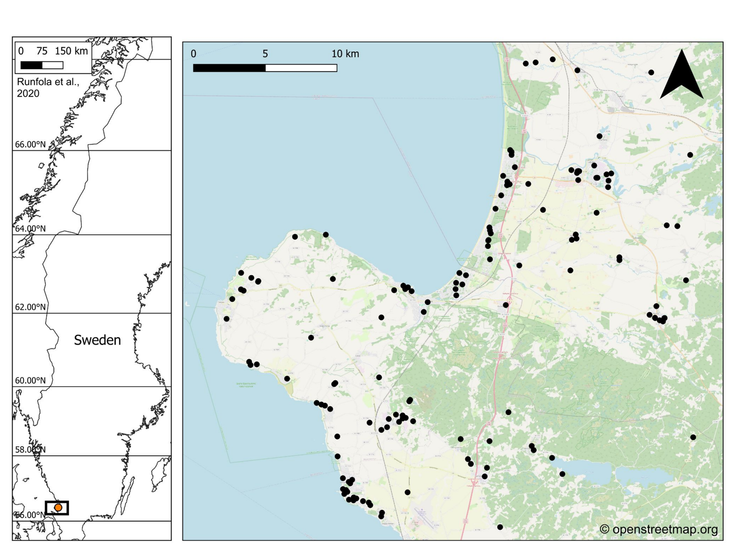


Fig 3.1: PWS in Bjärehalvön, Sweden, 2022.

(1) Institute for Modelling Hydraulic and Environmental Systems, University of Stuttgart, Germany (jochen.seidel@iws.uni-stuttgart.de)
(2) Swedish Meteorological and Hydrological Institute, Hydrological Research Unit, Norrköping, Sweden
(3) Division of Water Resources Engineering, Faculty of Engineering, Lund University, P.O. Box 118, 22100 Lund, Sweden
(4) Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology, Campus Alpin, Garmisch-Partenkirchen, Germany

EGU25-10988

More information about OpenSense: