The authors modify an existing rain gauge quality control (QC) algorithm, to make it better able to handle data from non-professional gauges. The modified algorithm is used to quantify the accuracy of both professional and non-professional data. The added value of including the non-professional data in multi-source precipitation products is demonstrated, both for a longer period and for individual events.

There is a growing interest in using non-professional precipitation observations to complement professional observations from national or regional meteorological and hydrological services. A major challenge concerns the higher degree of uncertainties and errors in the non-professional data. Therefore, efforts to develop and share practically applicable QC algorithms are timely and must be encouraged. I find the approach presented overall sensible (although some further clarification is needed; see below). The presentation is clear and the calculations and analyses appear well performed, as far as I can judge, but some revision is needed in light of the following comments.

General:

Much of the results and conclusions are based on measurements from four single months, considered typical for their corresponding season. This is a quite heavy assumption, firstly because one month is very short in a rainfall/precipitation context – it may contain just a few single events – and secondly because a single month may differ substantially from seasonal climatology. I would suggest to include as much “seasonal data” as you have available for this analysis. Alternatively, but much less preferably, show that the selected months are sufficiently good “seasonal representatives” for supporting the conclusions made.

> We understand the reviewer's approach, which suggests to perform verification on data reflecting the climatology of precipitation. However, in this case this is, on the one hand, impossible due to the heterogeneity of the data on which such an analysis could be carried out and, on the other hand, the purpose of our analysis was different:
>
> 1. The radar network in Poland was completely replaced between 2021 and early 2023, and moreover two new radars were installed. Therefore, it is not possible to carry out verification on data from these years, and on the other hand, the radars operating earlier (installed in 2002-2004) were of a different type, with single polarisation and therefore quality control was less advanced. Since radars play a key role in the current version of RainGaugeQC, verifying its performance on data from old radars would give a distorted results of its performance.
> 2. During the same period most of the telemetric rain gauges were upgraded (from tipping-bucket to weighing ones) and their number significantly increased, so the rain gauge data also changed completely in terms of their number and quality, which drastically affects the effectiveness of the individual RainGaugeQC algorithms.
> 3. We initially thought of verifying on the whole four seasons (Dec-Feb, etc.), but found that more important are the characteristic types of precipitation in Poland: snow in winter and deep convection in summer. Our intention was to investigate how the developed algorithms perform with different types of precipitation, and so we chose the four months in which these types of precipitation occur most frequently. Currently, snowfall in Poland does not last all winter, it is most frequent in January (in January 2024: 12 days in Warsaw located in the centre of Poland), and the most intense storm precipitation occur in July. Stratiform rainfall is most common in October, and weak convective rainfall starts to appear in April.
>
> So we did not aim to carry out analyses that took into account the climatology of precipitation. This, of course, would also be interesting, but would require reasonably homogeneous rain gauge and also radar data series.

On lines 347-349 is written that (sufficiently dense) PWS data is not available at IMGW, and therefor tests have not been made on PWS data. I wonder whether the EUMETNET Sandbox with Netatmo data (Netatmo, 2021), which cover Poland, could be used for this purpose. If so, this would be a very interesting addition to the paper.

We downloaded Netatmo data, which for Poland has good coverage. However, there were several reasons why we did not use them:

- We only had access to data from 2020 but these data were not suitable because this is the period before radar replacement. The newer data are not available for free.
- The 2020 data are, at least for Poland, collected in an unclear way: often time steps between successive measurements are irregular (from several to several tens of minutes) without any information on accumulation time. This uncertainty would introduce too much error in the statistics.

We are still waiting for the network of PWS stations at IMGW, which is currently being set up, but we would have to wait next two years for the data (sufficient station numbers plus time to collect measurements).

Parts of the description of the updated algorithm are a bit hard to follow (see examples below and also comments from RC1), more explanation and justification needed.

Thanks to the comments of both Reviewers, we hope that we have sufficiently improved the transparency of the algorithm descriptions.

We have added a paragraph on line 283:
„First, the 10-day radar precipitation total $\sum_{10\ days}(R)$ is checked. If this is too low, then the correlation coefficient is not calculated, as it may not be reliable in such a case. In addition, it is checked whether the rain gauge rainfall $\sum_{10\ days}(G)$ differs significantly from the radar data (formulae 1 and 2) and depending on this, the quality index of G is reduced."

We changed the sentence in lines 283-284:
"If the both accumulations are below the assumed threshold values, then the quality index of the rain gauge data is not reduced and the check is stopped:"

We also changed the sentence in lines 286-288:
"If the amount of radar precipitation for the long series is below the assumed threshold and the amount of rain gauge precipitation is above the corresponding threshold, indicating large differences between the two accumulations, then the check is also stopped and the quality index of the rain gauge data is reduced by 0.05:"

We have removed the sentence in lines 290-291.

The English is overall understandable and not a big problem for me, but there are examples of curious expressions that could be improved by a "native check", I think (some examples below).

We have improved the English in the article. We hope that the text is now noticeably better.

Specific:

- 68: I think the P in PWS usually stands for Personal (although I do have seen also Private).

We have corrected to "personal"

- 76 and others: Change np. to e.g.

We have corrected.

- 79: Example of sub-optimal English (in my opinion): “relatively very large”.

  We have corrected.

- 124: et al. is missing.

  We have added.

- Table 1: What type is the DLP gauge? Weighing?

  We have added.

- 171: This expression is rather for the Introduction.

  In the Introduction there is already information about the dependence of measurement uncertainty on device type. We have therefore left it at line 171.

- Figs. 2 and 3: I suggest combine into one.

  The figures with the distribution of manual rain gauges and radars are in different Sections (2.1 and 2.2), so combining them would be unclear to readers. We will leave the decision to the editor of the journal.

- 204-207: Some more details here are needed to understand the following applications.

  We have added in line 207:

  “This adjustment is carried out from gauge-radar ratios determined at rain gauge locations, spatially interpolated over the entire domain.”

- 238: Is this done at each time step? Do you mean the sensor with highest quality?

  Yes, for each step - we have added this information in line 238:
  “This QI metric…” → „At each time-step this QI metric…”

  Yes, with the highest QI. In the case of two sensors, “higher” could also be used in our opinion, but to avoid confusion we have changed to “the highest”.

- 246: “proved unsuitable”, in what way?

  The sentence in lines 246-248:
  “The QC algorithms in the previous version of RainGaugeQC proved unsuitable for non-professional data, as they are often subject to greater uncertainty than from professional rain gauges, and besides, these gauges are generally not dual-sensor.”

  we have shortened into:
  “The QC algorithms in the previous version of RainGaugeQC turned out to be inadequte for non-professional data, as these gauges are generally not dual-sensor.”,
  to highlight the key reason.

- 255: Which time step (or interval)?

  There is only a very general description of the ideas for the various algorithms here, while their detailed descriptions can be found in later subsections of this chapter. We have added the sentence in line 253:

"Here is a brief overview of the changes made to the RainGaugeQC algorithms, whereas detailed information can be found in Sections 3.3 to 3.5."

- 268: English: "degree of outlying"

  We have corrected:

  "to determine the degree of outlying for individual data"
  → "to determine the degree to which individual data is an outlier"

- 276: How was 5 and 10 days selected? They are not that different, can you justify that they are short and long enough?

  In general, the idea is to test the correlation on a data series that is as short as possible, but still representative. The 5-day sequences optimally take into account the current correlation of the data. But we have found that, for example, during periods of low rainfall, the correlation calculated on short-time series can be random and then a 10-day sequence can give a more reliable correlation.

  We have removed the word "long" from line 272.

  We changed a sentence in lines 275-276:

  "Two time series aggregated from 10-min accumulations: "long" and "short" comprising 10 and 5 days, respectively, are analysed.
  → "Two time series aggregated from 10-min accumulations: "short" and "long" comprising 5 and 10 days, respectively, are analysed in order to test correlations on time series that are as short as possible and, on the other hand, sufficiently representative."

- Section 3.3: Generally, many numbers/thresholds here that are given without explanation or motivation. I assume they have been carefully set, but some clarification would be good.

  We have added a new paragraph after line 269:
  "All parameters of the algorithms described in sections 3.3 to 3.5 were selected empirically by comparing the calculated QI values with the expected ones derived from our assessment of the data reliability."

- 285: Why different limits for R and G?

  Regarding equation (1): if the precipitation is low and the correlation may not be reliable, then if the radar confirms this low precipitation, we do not calculate the correlation and decrease the QI. However, we allow some tolerance, hence the different thresholds for R and G.

  In formula (2): if rain gauge precipitation (G) is high and radar precipitation (R) is low, but this difference must be evident, then we decrease the QI but do not calculate the correlation as it would not be reliable. Hence, the threshold on G is significantly higher than on R.

  We have revised the text relating to formulae (1) and (2) to improve clarity. Details of these changes are in response to an earlier Reviewer's comment, which we have marked with the symbol (*).

- 295: Has c in this eq. been defined?

  Yes, "c" is defined in line 278 (we have corrected the missing italic in this symbol).

- 323: This eq. is one example of something that needs more clarification.

  We have changed the notation of equation (7) to make it clearer:

$$SF(\Sigma G, \Sigma R) = \begin{cases} \text{true} & 1.3 \cdot \min(\Sigma G, \Sigma R) + 7.0 > \max(\Sigma G, \Sigma R) \\ \text{false} & 1.3 \cdot \min(\Sigma G, \Sigma R) + 7.0 \leq max(\Sigma G, \Sigma R) \end{cases}$$

- 336: So, PWS is always biased? Is this a reasonable assumption?

  The reduction of the quality by at least 0.1 is due to their generally lower reliability. We have reason to assume that while professional stations meet WMO standards and are properly maintained and supervised, we are not so sure about private stations.

- 359: English: “additionally non-outlier”

  We have changed “and additionally non-outlier” into: “or to the class of correct data”.

- 373: How is this classification made? More explanation needed here.

  We have completed the sentence in line 373, before the colon:
  „(see: Ośródka t al., 2022)”

- 411: Insert “, respectively,” between “gauges” and “is”.

  We have corrected.

- Fig. 5: The font is a bit small.

  We have increased the size of all the fonts in this figure.

- 458: Any influence of snow?

  Probably yes. We have added this observation in line 460:

  “Probably the reason for the worse results for January is the occurrence of snowfall, which is more challenging for radars to detect.”

- 570-571: A bit strong statement, in my opinion, that these two cases show that non-professional are useful “in most cases”. Probably/hopefully they are, but the statement would require more cases to be evaluated.

  We changed in line 571:

  “can in most cases play a positive role in the estimation of the precipitation field.”
  → “is likely to play a positive role in the estimation of the precipitation field in many cases.”

- 597-599: Is this a conclusion from the present study? I do not really see this.

  The Reviewer is right. We have removed this point from the Conclusions.

- 602-604: This is clearly not a conclusion from this work. I suggest extend this to a final paragraph about future efforts, remaining issues, etc.

Thank you for this suggestion! We have moved point 5 of the conclusions into a separate paragraph and completed it:

"The development of the quality control system for telemetric rain gauge measurements will be continued. Plans include incorporating precipitation data from other non-professional networks to supplement the IMGW rain gauge network. This will increase the proportion of data with potentially lower reliability, which may require even more sophisticated algorithms for the quality control. Moreover, IMGW is in the process of establishing a network of personal rain gauges. Once this network is operational, it will be possible to test the quality control algorithms proposed in this paper using data from these rain gauges."

Reference:

Netatmo (2021): EUMETNET Sandbox: Netatmo observing network data v1. NERC EDS Centre for Environmental Data Analysis, 2025-03-15. https://catalogue.ceda.ac.uk/uuid/e8793d74a651426692faa100e3b2acd3/

We thank the Reviewer for this information. We used Netatmo data from 2020 in our initial work on RainGaugeQC, but we do not have access to this data from the period for which we were able to test the current version of RainGaugeQC.