General comments:

The manuscript introduces a multi-level data quality check for rainfall data measured by citizens and quantifies the improvement in time and space if the data is used for interpolation additionally to a conservative network. This kind of rainfall data is analysed for Germany for the first time to my knowledge and shows a high potential for further applications. Hence, this topic is of broad interest for the hydrological scientific community and suitable for a publication in HESS. The paper is well-written, with a clear and intuitive structure. I have no major comments, but a number of issues that should be solved before the manuscript should be published.

Specific comments:

L26-28 The short periods of available radar data should be mentioned in this context as well.

L77-79 It would be helpful if the authors are more concise regarding the number os PWS stations finally used in the study. To enable a transfer of the applied methods the authors should provide some information, which minimum time series length was chosen for the secondary time series and how was it chosen?

L85 From the first paragraph in Section 3 it sounds as only the two data quality filters will be explained. I suggest to provide a brief overview of all subsections at the beginning of Section 3 and an explanation, how they are interacting.

L102 Maybe the authors should explain briefly why they consider Y as a random field.

L120-123 The chosen criterion sounds reasonable. I'm wondering if an exclusion for too high correlations has to be applied as well. Later in Fig. 5 indicator correlations of 1 are shown for interstation distances of 10 km, which is way higher than from the primary network. Maybe the authors can report if an upper limit is required or not when working with the data as a result from their data analysis.

Also, I'm struggling with the final decision if a secondary time series remains in the potential useful data set or not. As far as I understand it a time series is "flagged as suspicious" if it does not meet the criterion in Eq. 2. That means the time series will be sorted out. Since the procedure is repeated for several α and Δt, I imagine the highest exclusion rate will be found for high values of α. Is a flagging for only one of the analysed values of α enough for an exclusion of that time series? Which values of α have been applied and what was the exclusion rate?

L130-133 The authors switch in their explanations from correct amount order in time to percentiles that can be applied to distribution functions for the correct estimation of precipitation amounts. This requires the assumption that the unknown bias mentioned before does not change over time and between events. This should be communicated to the reader as well.

L159-160 Does this approach introduce an upper limit for the point of interest, resulting from the maximum rainfall amount measured at the surrounding primary stations? Or are theoretical distribution functions applied and the information is missing (or I missed it)?

L228 The first filtering? I would have assumed that this extra-ordinary event-based deviation would have been eliminated by the second filtering.

Fig. 3 & Table 3 From Fig. 3 it is obvious that the minimum resolution is 0.01mm for the Pluvio, while it is 0.1mm for the PWS. This makes a comparison of p0 without its consideration biased. Was the different measurement resolution taken into account for the values of p0 in Table 1? Otherwise I would recommend to either neglect values <0.1mm or to sum rainfall amounts up to a minimum of 0.1mm. The pluvio will gain more dry time steps by doing so. It maybe has a neglectable effect for hourly time steps, but for the original temporal resolution of 5min it will be critical. Hence, it should be at least communicated to the reader.

Fig.3 I recommend to add x-y-lines to illustrate the perfect match since in the left figure it is not the diagonal.

Fig. 5 Indicator correlations with values below the minimum resulting from the primary network for similar distances are included in the right figure. From my understanding these were removed by (2)? Also, for the decision of keeping secondary stations or not indicator correlations for unknown distances resulting from the primary network have to be estimated. In general, this is done by fitting regression lines to the observations? Was it done similar in this study? If so, it could be useful for the reader to provide the type of regression line and it parameters. If not, how were values judged for unknown distances?
L289 "…With increasing…as the role of the bias increases". Is the bias the only reason therefore? I guess the much higher spatial correlation for longer time steps also gives less possibility for improvements, so a frontal event with 12h duration covers some of the stations from the primary network, while this is not the case for hourly time steps (it is mentioned later, L298).
L335-336 Do the authors mean "event" here instead of "data"? Otherwise I'm wondering to not find the values for the RMSE in Table 5.

Technical corrections:
L45 Bracket before the authors.
L62 If the authors refer to the Black Forest, it should be indicated for the reader who is not familiar with the German topography in Fig. 1 where the Black Forest is.
L67 Maybe at least the location of Reutlingen can be shown in Fig. 1
L75 Can the authors provide a reference for the 5°C threshold or how was it chosen?
L80-84 Again, it would be helpful for the reader if mentioned locations are shown in Fig. 1. Otherwise, the locations can be left out.
L94 „This relationship" Which relationship?
L96 „it" is missing
L98 „be be"
L99 „Δt durations" -> „durations Δt"
L100 „Δt" was introduced before.
L105 Maybe „exceedance probability" instead of „probability".
Eq. 1 „Δt" is missing as index on the left side of the equation.
L185 There is no subsection 3.4.2 .
L186 missing comma
Table 1 The criteria/abbreviations should be introduced in the table caption.
L213 we -> were
L218 on -> one
L233 over -> overestimate
Table 3, 4 and 5: Line 2 is redundant with line 2 from Table 2.
L284 Something is missing in this sentence, please double-check.
L349 based -> is based
Author contributions: sduy -> study