Reply to comments by Anonymous Referee #2.

Our response to the comments by Referee #2 is provided in bold font.

My overall evaluation of the article presented to me for review is high. The few comments I have written below are rather secondary and are more questions to the Authors, who may not necessarily agree with them.

The main strength of the paper in my opinion is that it addresses the latest developments in QPE and QC of rainfall data on European ground.

We thank the reviewer for providing a positive and constructive assessment of our manuscript.

Most important doubts (questions):

- Why were only OPERA and Netatmo data used, without rain gauges of even a few NMHSs? More rain gauges = their QC is more effective.

The reviewer is right that ideally all available rain gauge data would be combined. This could be beneficial for the QC of PWS rain gauge data, but also to obtain the best possible gauge-adjusted radar precipitation accumulations (i.e., for the merging itself). Concerning the latter, we already mention "Improving the climatological gauge-adjusted radar dataset EURADCLIM by merging both the PWS and the ECA&D rain gauge data at once is another avenue that can be explored.". Since the aim of this study is to show the potential of PWS gauge observations for improving radar data, we decided to use the NMHS gauges only for an independent evaluation. We added to the manuscript (L. 402-403): "Finally, NMHS rain gauge data could also be taken into account in the quality control of PWS rain gauge data.".

- Does the scheme shown include the use of some form of quality index (QI) or quality flag? E.g. in the case of 'flex filtering', when there are less than 5 other rain gauges in a neighbourhood, then one could keep the value in a given rain gauge but lower its QI. If we included the QI in these algorithms, then that rainfall height would enter into spatial interpolation and/or merging, but with lower weight.

Currently, the scheme assumes all gauge data to have the same quality, but, as already mentioned in the manuscript, "The merging algorithm employed in this study can use quality information about the gauge observations, giving lower weight to lower-quality gauges (in this case the PWS gauges).". We agree that incorporating quality information on PWS rain gauge data would be relevant, where its weighing would, for instance, depend on the outcome of the quality control and the type of rain gauge being used. This will require a thorough evaluation of the appropriate weights that should be used given the outcome of the quality control. Especially, when NMHS and PWS rain gauge data would be combined in the merging, assignment of

weighing factors will become even more relevant to avoid that the more accurate NMHS gauge data would be overwhelmed by the many more observations from the less accurate PWS gauges.

- P. 9, l. 178-180: What about the reverse case, where the radar does not see the weak precipitation found by the rain gauge(s), which happens at greater distances from the radar site as a result of a radar beam overshooting the precipitation? This happens when the rainfall is from low clouds, especially in colder periods.

It could indeed occur that the radar accumulation becomes lower than 0.25 mm in 1 hour due to overshooting of precipitation, implying that the radar-gauge pair is not used to compute an adjustment factor field. Note that in case precipitation is entirely missed in the radar image, the radar accumulation of zero cannot be increased anyway.

- P. 9-10, Sect. 3: How about providing the full formulas for the statistical metrics used? This always makes analysis easier.

We added the equations for relative bias, coefficient of variation, Pearson correlation coefficient and mean absolute error to Section 3.

- Table 1: Are these numbers in the top row of the table for 'no threshold'? It might be worth writing this in the relevant lines of the table.

Yes, the numbers in the top row are for "no threshold". We added "No threshold" to the lines where no threshold is applied.

- Table 1 and p. 11, 222-223: How to interpret this table? Do the small differences between the PWS data without and with QC mean that the PWS data are good on their own, or rather that QC is too ineffective (or too moderate)? The former possibility might be suggested by the large improvement in the "OPERA + Netatmo No QC" data relative to "OPERA".

It is probably a combination of the quality of the PWS data already being relatively good, and the merging algorithm acting as a kind of quality control. We added to the beginning of the Results Section (L. 228-230): "No QC implies that the quality control in Section 3.1 has been omitted, except that the 1 h PWS accumulations are used for merging if both the unadjusted radar value and the Netatmo gauge value are larger than 0.25 mm." (because it was not entirely clear from the manuscript what is meant with "No QC") & (L. 233-236) "The relatively good performance for No QC could be attributed to the quality of PWS gauge data and to the merging algorithm acting as kind of quality control. Only radar-gauge pairs are used in the merging for which radar and PWS observe more than 0.25 mm. Moreover, a spatial adjustment factor is computed by distance-weighted averaging of radar and PWS values, which can average out outliers.". Note that when we discuss the scatter density plots we do see a positive and

important impact of the quality control: "The scatter density plots also reveal how quality control has a positive impact: a group of large precipitation accumulations in case of lower gauge accumulations is removed.".

- P. 15, l. 249-265: For me, these are very valuable insights!

Thank you.

- Fig. 8: What are the benefits of using the PWS data in this figure? The spatial distribution of the precipitation field is very similar in the three maps shown, so is the scaling the main benefit? This is what the commentary in l. 275-289 suggests as well, so why not also present such statistics to show this impact on the distribution? E.g., the correlation coefficient...

We indeed do not show clear benefits of the quality of the PWS-based dataset. Given these results, presenting metrics will likely not provide new insights. There is, however, one clear benefit of the PWS-based dataset, that is already mentioned: "Moreover, the density of the PWS network is higher than that of the ECA\&D network (Fig. 8d) and the data could potentially be available in real-time, which is usually not the case for gauges in the ECA\&D dataset.".

- Fig. 9 is also very extremely interesting!

Thank you.

Good luck!