Interactive comment on "Evaluation of GPM IMERG Early, Late, and Final rainfall estimates with WegenerNet gauge data in southeast Austria" *by* Sungmin O et al.

We appreciate your comments and suggestions, which helped to enhance the quality of this manuscript. We have discussed all issues indicated in the review report and provide a point-to-point response to the comments below.

Major comments:

#1 Gauge representativeness I think that 40 and 39 for given 0.1◦ × 0.1◦ (roughly 10 × 10 km2) grid cells are definitely good numbers of gauges for estimating areal average rainfall at a typical temporal scale of satellite rainfall products (e.g., three-hourly or monthly). However, since this study evaluates 30-min products for which random variability is much higher, I think that the authors should justify that the gauge representativeness error is not significant at the space and time scales used in this study. The authors could provide the structure of spatial correlation and variance reduction for the study area as shown in Villarini and Krajewski (2007). I also think that the use of 200 × 200 m2 gridded rainfall data at 5-min scale is not reasonable (but this does not significantly affect the results of this study because the gridded data are aggregated over 30-min and 10 × 10 km2 scales and then used). The 5-min gridded map contains so much variability (in terms of gauge representativeness) due to high space and time scales used as well as the tipping bucket rain gauge error itself with tip counts within 5-min (this will also decrease with longer time integration).

> ***=> We thank the reviewer for this comment and address as follows: Given the spatially uniform and high resolution configuration of WEGN, and additionally considering that the network region is characterized by low relief, we can ensure that WEGN provides a reliable ground reference, particularly in terms of spatial representativeness for IMERG validation.***
>
> ***Fig.1 below plots VRF (variance reduction factor) of WEGN 30-min data at Grid 15.85 (the pixel with 40 gauges) as a function of number of gauges randomly sub-sampled from the 40 gauges. Based on this result, we will add the following sentence in the <Introduction>: "The Variance Reduction Factor (VRF, Villarini et al. 2008), examined with half-hourly WEGN data of the 40-gauges grid box, is about 0.02 for 10 gauges (average of 40 random combinations) and little or no improvement is observed in the VRF beyond 10 gauges." (Page 3, Lines 1)***

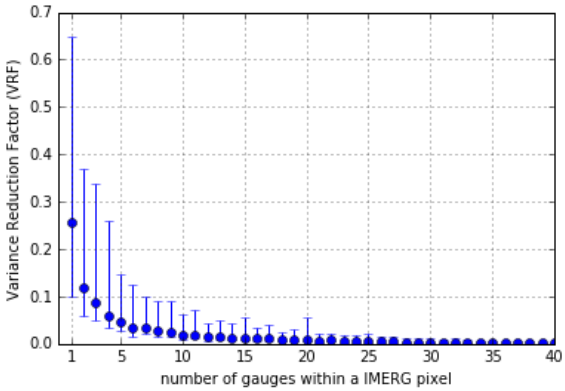


Fig. 1 VRF for different gauge densities, calculated with 40 random combinations for each gauge number. Filled circles represent the average of VRF for a combination. Vertical bars show the associated min-max range.

Fig.2 below shows normalized RMSE of WEGN mean areal rainfall (the average of all 40 gauges is assumed to be true) as a function of gauge number for 30-min and 5-min accumulation times, for Grid 15.85. The errors are calculated at a time step of maxima (left panels) and median (right panels) rainfall spatial standard deviation (when more than 20 gauges detected rainfall). Each boxplot is based on 3,000 possible combinations (with the exception of gauge number 1 and 39, based on 40 possible combinations; and gauge number 2 and 38, based on 780 possible combinations).

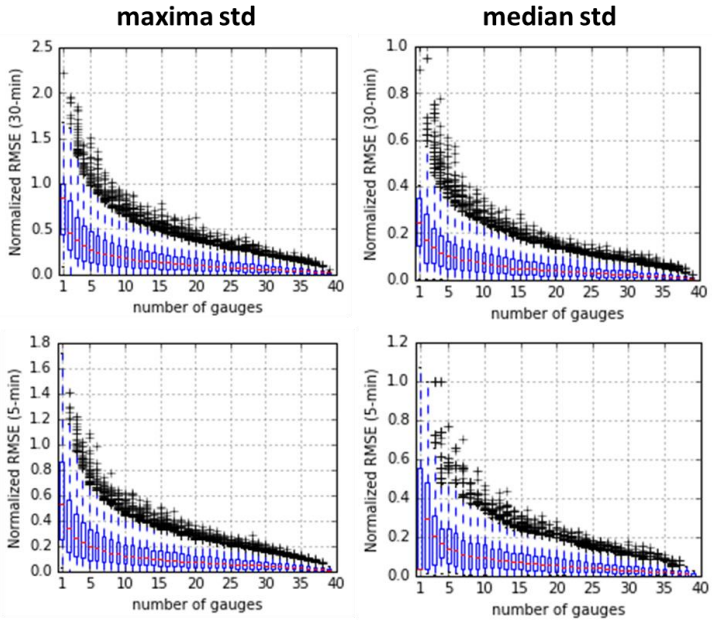


Fig. 2 Dependence of uncertainty of IMERG pixel-averaged rainfall estimation on the number of gauges. Note that the ranges of y-axis vary between plots.

It is clearly shown that the spatial uncertainty of mean areal rainfall substantially decreases with increasing gauge numbers (little change from ~10 gauges, as expected from the VRF), and areal rainfall estimates tend to converge toward the 'true' value. As the reviewer pointed out, 5-min data contain higher rainfall variability but, nevertheless, the figure shows that we can still obtain reliable mean areal rainfall from WEGN for the study domain, with a significant decrease of spatial representation errors.

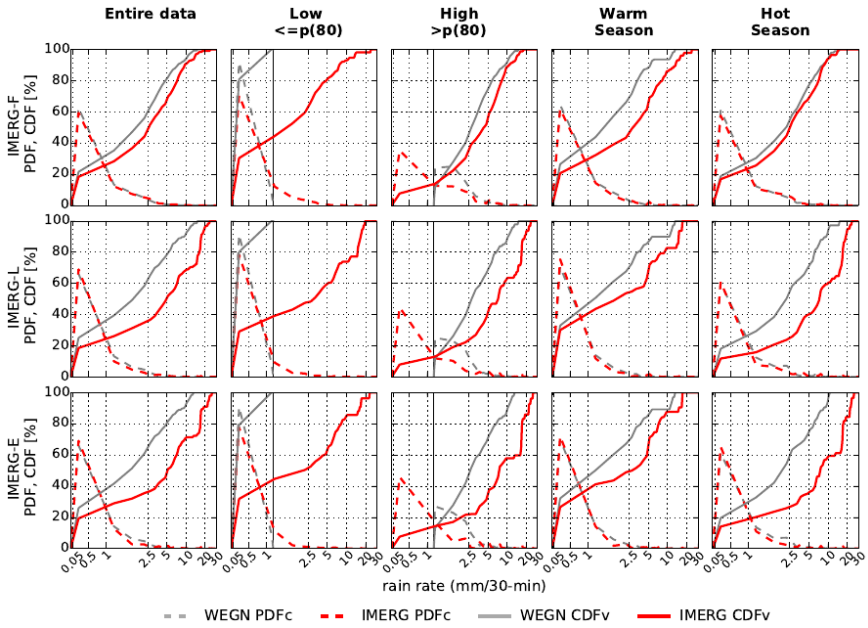
#2 Gauge data independence Please clarify that the rain gauge data used in this study are from an independent network. The authors state that "the WEGN is not a member network of the GPCC network" (Page 4 Line 32). However, there is another statement (Page 7 Line 33 – Page 8 Line 2) that a better performance may be attributed to an Austrian national station that are associated with the GPCC product. These are confusing.

=> We apologize for the confusion. There are two different networks in the study domain: 1) 40+39 WEGN stations and 2) an Austrian national weather station from the network of the national meteorological service (ZAMG). This ZAMG station is a member of the GPCC network. We did not include the station for our study since its data directly affect the performance of IMERG Final run.

We will rewrite as "… attributed to an Austrian national station, <u>which is not part of the WEGN gauges, but located</u> within the WEGN area (in Grid cell 15.85), of which measurements are integrated into the GPCC …" (Page 8 Line 1), so that readers can understand it clearly.

#3 Figure 5 (the second and third columns) I may miss something, but the WEGN PDFs and CDFs exist for the entire rain rate regions although they are presented for low (1.2mm) rain amounts each. For example, shouldn't the PDFs and CDFs start from 1.2 mm in the third column (high rain rate)? Why do the PDFs and CDFs exist for R<1.2mm?

=> Your understanding is correct. WEGN PDFs and CDFs have values only from/till 1.2 mm, but current plots contain lines extending beyond stated range because, for example, WEGN PDFc in the high rain rate cases (third column) starts from 0% at 0.05 mm (not at 1.2 mm) so we get to see values even in < 1.2 mm ranges. We now re-plotted the second and third columns so the starting/end points of WEGN PDFs and CDFs will be indicated by a vertical line and the WENG PDFc/CDFv end/begin at these lines. Thank you for the question.



#4 Time shift in Figures 7 and 8 It is hard to say that the observed patterns in Figures 7 and 8 show a time shift. I think that we can say there is a time shift only when the rainfall durations are the same between reference (WEGN) and IMERG products and starting times are different. It seems to me like that the observed patterns are just errors, probably by morphing and other reasons. In Figure 8, the shapes and peak times are all different and it is hard to find any consistent or systematic tendency.

> ***=> Thank you for pointing this out. The term "shift" is used to refer to the time disagreement between WEGN and IMERG; to be clearer, we will re-write the sentence (Page 8 Line13-14) as "… with a time <u>difference</u> of about 2 hours earlier in starting time. This <u>false alarm</u> suggests that…"***

Minor comments:

#1 Page 3 Line 1. ". . .. has been suggested to guarantee a monthly error of under 10%". This cannot be directly applied to this study because of the temporal scale difference (monthly vs. 30-min). The gauge representativeness is a function of space and time scales used (Seo and Krajewski 2010).

> ***=> We will add the VRF value obtained from WEGN 30-min data in the <Introduction> (Page 3 Line 1) as you suggested, please also refer to Major comment #1.***

#2 Page 4 Line 24. Please explain "high native time resolution."

> ***=> We will modify the sentence as "…, WEGN gridded data with 5-min native resolution are used." (Page 4 Line 24)***

#3 Page 4 Line 27. Please clarify the rainfall threshold used in this study. In the caption of Figure 2, there is a phrase "≥ 0.1 mm at single station". How big is the tip resolution against the threshold?

=> The caption means that we applied 0.1 mm threshold to count the number of WEGN gauges (the tip resolution is 0.1 mm) that detected rainfall. So this threshold is used only in Figure 2. On the other hand, 0.05 mm threshold is the one used to define rain/no-rain data in each grid box (average of multiple gauges) for the study. We will modify the Figure 2 caption to say "(gauges with >= 0.1mm rainfall are counted)" to make it clearer.

#4 Page 6 Line 33. Please clarify the difference between Figures 4 and 5. Is the threshold 0.05 mm applied to both Figures 4 and 5?

=> Yes, the same threshold is applied. Figure 4 shows all rain data of IMERG and WEGN (i.e. not only hits but also false alarms and misses), whereas Figure 5 shows only matched data (i.e., only hits; this is explained at Page 6 Line 33 - Page7 Line 1).

#5 Page 7 Line 5. Isn't it 0.05 instead of 0.5?

=> It is 0.5. We are comparing the POD score of IMERG-WEGN for all rain intensity ranges versus higher intensities (for WEGN > 0.5) to show that misses of IMERG are observed at relatively low rain intensities. To make it clear, we will re-write the sentence of Page 7 Lines 4-6: "Indeed, we recomputed the POD score using only values when WEGN is above 0.5 mm 30-min^{-1} and found that the PODs are 0.70, 0.79, and 0.75 for IMERG-E, IMERG-L, and IMERG-F, …". Please also refer to the next comment – Minor comment #6.

#6 Page 7. Line 6. Please explain "the entire WEGN data." Is it without applying the threshold?

=> It is with applying the threshold. Here "the entire WEGN data" means all available WEGN data; we will correct it as "… against WEGN for the entire range of rain intensities above the 0.05 mm 30-min^{-1} threshold" in order to make it clear.

#7 Page 7 Line 31. It would be useful if the correlation coefficient values obtained from other evaluation studies are provided and compared with those of this study.

=> It seems to us that it will be very difficult to directly compare our correlation coefficient values with those of other studies, because the agreement score between ground reference and IMERG can be very different depending on i) the nature of the ground reference (e.g., radar vs gauge, number of gauges, rainfall regime) and on ii) temporal and spatial scales used

for comparison. Moreover, given that the study aims to compare the performance between three different IMERG runs, we show the indices in Table 2 to explore differences between the IMERG runs rather than to give absolute scores of each run.

#8 Page 9 Line 2-4. What (more PMW estimates or rain gauge correction) has more contribution to the improvement? I think that this is very important point in the satellite product evaluation

=> Thank you for the suggestion. We will add "From these two case studies, it appears that the gauges provide a greater improvement to IMERG Final estimates" after Page 9 Line 4.

#9 Section 4.3. Please add some implication of the result found in Section 4.3 and discussions on how to use the products for hydrologic applications.

=> We are grateful for this comment. We will add "This analysis identifies possible sources of error that should be considered in the context of hydrological applications of IMERG data. For instance, biases (overestimation in this case) in IMERG rainfall estimates will inevitably propagate through hydrologic models, and consequently this would lead to greater errors in runoff. The magnitude of biases can be reduced when IMERG Final estimates are used. Time offset bias, however, remains relatively stable across all three IMERG runs, especially in the warm season. Therefore, comparison or adjustment of IMERG estimates using local ground reference (if available) in terms of biases not only in amount of rainfall but also in its timing should be considered as an approach to meet the required level of accuracy in rain data." in <Section 4.3> (page 9, Line 23)