Answers to review by Hidde Leijnse (reviewer's comments in black, answers in blue)

In this paper rainfall estimates from a weather radar and an underground gravimeter are compared. A straightforward method of estimating rainfall from the gravimeter is presented. Gravity measurements are first corrected for tidal effects and atmospheric loading, and are then converted to rain accumulations by applying a moving average, and a linear relation between gravity and accumulated water. Results that are presented show that the gravimeter indeed has a strong precipitation signal. It is also shown that the gravimeter data can help in the case of hail, where radars typically overestimate precipitation rates. The paper is well-written and very interesting for readers of HESS. As far as I know, this is the first time that rainfall estimates from an underground gravimeter are reported, and hence the paper is very novel. The paper could benefit from adding known information about uncertainties in gravimeter data in order to facilitate comparison of the two. I have a few further suggestions for minor changes to the paper, after which I think the paper is ready for publication. Specific remarks are given below.

The authors would like to thank Hidde Leijnse for the very positive comments and the very relevant remarks and suggestions.

Uncertainties in gravimeter data are already described in the paper but we will add more explicit information on these uncertainties in a revised version of our paper.

Specific remarks

1. In Figs 2, 4, and 6-9, please add error margins to the gravimeter data so that it is immediately apparent what the expected uncertainty of these measurements are

For the time series (Figs 2 and 8) we propose to use a rolling standard deviation to plot a 2-sigma margin on the data. For the 5-min gravity changes in Fig. 4 and in event gravity changes in fig 6-7-9 (scatter plots) we propose to take the average standard deviation before and after the event. This requires some re-processing. We need also to check whether adding uncertainty information in the scatter plots will not reduce the readability of the figures.

2. On p.6, line 23, it is stated that the 5-minute gravity change is averaged over 5-minute intervals in order to remove high-frequency gravity fluctuations due to other sources. It would be very interesting to know how this 5-minute time scale is related to the time scales of hydrological processes that would affect the gravity measurements through the redistribution of water. Please add a statement on the typical timescales of these processes. This can then be related to the 5-minute scale of averaging, but also to the typical time scales of individual rainfall events.

We chose to average gravity data at 5-min scale in order to match radar data. As shown in Fig. 5, most rainfall events are very short (less than 15 minutes). Besides, when evaluating the change in gravity produced by such events, it is reasonable to average the gravity over short time intervals as well. Averaging the gravity, for example, over one hour and evaluating the gravity change by taking the 1h-hour average before the event and after the event would incorporate in the gravity changes various effects which are not directly related to precipitation, e.g. pressure changes, ground-water processes, …

Concerning the timescales, we have seen that the intense rainfall events analyzed here occur over timescales of minutes. As recommended by the reviewer, this should be compared with timescales of other processes affecting the redistribution of water. Evapotranspiration occurs typically at diurnal timescales (2-3 mm/day during sunny summer days, Van Camp et al., GRL 2016) and infiltration at timescales of hours. Characteristic timescales of runoff are more difficult to quantify and depend on the status of the soil saturation. We think that runoff is the predominant process affecting gravity at timescales close to those of precipitation. However, we find in most cases that gravity does not substantially and rapidly increase after the rainfall events, which suggests that runoff is slower than precipitation process.

3. In Fig.2, it is clear that there are high-frequency fluctuations present in the gravimeter data. I would recommend discussing these fluctuations, and potential ways to remove them. For example, would it be possible to average gravity differences over longer time intervals (say, 15 or 30 minutes) to remove most of these fluctuations? It would be interesting to see the effect of different averaging time scales on this apparent noise. Please consider adding a sensitivity analysis to the scale of averaging.

We use data with a one-minute sampling rate : it means that periods longer than 120 s were filtered out. Then, looking at the power spectral densities of SGs, the noise reaches a flat, lower level at periods ranging 50-500 s. Hence, averaging on 300 s (or 5 minutes) allows benefiting from the period at which the SG presents the best performances.

At longer periods, the noise becomes red (increasing power with increasing period). Hence, for rainfalls of 30-60 minutes, concurrently to the rain-induced drop in gravity, there is another signal induced by the red noise. However, this red noise, at those periods, contains essentially the hydrological (e.g., Van Camp et al., JGR 2010) and atmospheric effects. The hydrological signal is investigated here, while we assess the effect of the atmospheric effects by discussing the pressure signal before and after the event. Therefore, we think that presenting a sensitivity study to the scale of averaging is beyond the scope of the present study.

4. On p.10, lines 5-6, it is concluded that a rainfall signal can be detected when radar reflectivity exceeds 40 dBZ. However, this conclusion is based on comparison of the 5-minute signals. I think that it is very well possible to detect rainfall signals even if reflectivity values are lower than this if the rainfall is averaged over sufficiently long time periods. This is for example demonstrated in Fig.2, between 1:00 and 6:00 UTC, where the radar only exceeds 20 dBZ most of the time, and 30 dBZ on just a few occasions. Yet the total accumulation by the gravimeter nicely follows that of the radar. So I think that this conclusion is too hard on the method that is presented. Please rephrase this conclusion to reflect this.

This is a very relevant remark. We found indeed several moderate rainfall episodes where the radar reflectivity does not exceed 40 dBZ but where a very good correspondence is found between the cumulated rainfall derived from the radar and from the gravimeter along the day. However, there are also many cases where we don't find any correspondence when only moderate rainfall is observed. In the latter cases, it seems that hydrogeological processes dominate over direct precipitation effect. It is worth mentioning that gravity fluctuations due to hydrogeological processes are likely to be dependent on the seasons and status of the soil moisture saturation (Van Camp et al., JGR 2006), which has to be done. This a very interesting research topic which requires additional research. It is already mentioned in the last paragraph of the conclusion but we will elaborate a bit more on this in a revised version.

Coming back to the reviewer's comment, we agree that it is not justified to conclude that no gravity signal is observed when the radar reflectivity is lower than 40 dBZ. This will be adapted in the revised paper.

5. On p.10, lines 14-15, it is concluded that a 48-dBZ hail cap works better than a 55-dBZ hail cap. This 48-dBZ hail cap corresponds to 36 mm h−1 (for M-P) or 34 mm h−1 (for RADOLAN). These thresholds seem rather low, and could result in missing significant rainfall. My guess is that this optimal threshold is partially a result of compensation for error sources other than hail. Please comment about this in the paper.

It is well known that hail produces rainfall overestimation when the M-P relation is used but, even for rain, overestimation can occur. This is due, for example, to radar electronic miscalibration, ZR relation that is not adapted to real drop size distribution, or partial evaporation of rainfall below the height of the measurement (VPR effect). In page 9 line 18-19 we mention that the 48-bBZ can be influenced by other sources uncertainties. As underlined by the reviewer, hail cannot be blamed as the only responsible for radar overestimations. We will elaborate more on this in the revised version, taking also into account the comment given by the second reviewer.

6. In order to make the paper more concise, I suggest removing Figs 6 and 9. The points in Fig.6 are already given in Fig.7, and there is only a minor difference between Figs 7 and 9. Furthermore, results from Fig.9 are also summarized in Table 1.

There is indeed some redundant information in figures 6 and 7. Concerning Fig. 9 the aim is to show the best agreement that we get between radar- and gravimeter-derived rainfall amounts. It is a simple figure that can be easily extracted from the paper and used by the scientific community to summarize in one slide the use of gravity measurements to derive rainfall amounts. Fig. 6 can be used for that purpose as well. Therefore, we would propose to remove Fig. 6 or Fig. 9 but not both.

Minor remarks

1. On p.1, the title does not include the fact that the gravimeter is underground, but this is an essential element of the paper (it probably would no work so well if the gravimeter was at the surface). Please modify the title to reflect this. Suggestion:

replace “superconducting” by “underground”

We propose to use “Exploring the use of underground gravity monitoring to evaluate radar estimates of heavy rainfall” as a new title.

Gravimeters installed underground are usually sensitive to a large area at the surface, which improves their rainfall detection. Conversely, gravimeters installed at ground level generally sample a much smaller portion of the top layer, but this varies from site to site. On the one hand, the buildings hosting gravimeters installed at ground level act as an umbrella, which prevents direct infiltration of rainwater below the gravimeter. It has been proved that such a building mask reduces significantly the signal recorded after rainfall events (Creutzfedldt et al., 2008). Reducing the size of the building, which hosts a gravimeter, as well as raising the gravimeter on a pillar tends to mitigate such building mask effects, as shown by Güntner et al. (2017). On the other hand, specific topography settings can also favor rainfall to be detected by surface gravimeters. This is for example the case of gravimeters installed at the top of a hill. However, underground gravimeters are in any case much more suitable to study rainfall processes because they are not concerned by building mask effects and much less affected by topography effects, while sampling larger areas of the surface layer.

References:

Creutzfeldt et al., 2008 https://library.seg.org/doi/10.1190/1.2992508

Guntner et al., 2017 https://www.hydrol-earth-syst-sci.net/21/3167/2017/

2. On p.2, line 12, consider replacing “puntual” by “point-scale”.

Will be done.

3. On p.2, line 23, consider mentioning that this is mostly the case for C- and X-band radars (not so much for S-band radars).

Will be done.

4. On p.2, line 25, consider referring to Fabry et al. (1994; https://doi.org/10.1016/0022-1694(94)90138-4)

Will be done.

5. On p.2, line 33, consider adding a remark that the radar sampling area is at least 1 million times as large as a gauge sampling area.

Very good suggestion. The comparison between sampling areas of the radar, the gravimeter, and a rain gauge is very instructive. Will be done.

6. On p.5, lines 3 and 4, there’s a typo in the units (should be “nm/s2/hPa”; this occurs twice: on lines 3 and 4). Consider using using notations for units such as “nm s−2 hPa−1” instead of using the “/” character throughout the paper to increase readability.

Will be done.

7. On p.5, lines 3-5, it is mentioned here that the values of the coefficients vary with time as well as the frequency of pressure fluctuations. In the conclusion that is drawn that there is a 15% uncertainty in the gravimeter data is based on the differences between the minimum and maximum values of these coefficients. It is hence implicitly assumed that the time variations of these values is much less than the variation with frequency (or cpd). Is this correct? If so, please add a statement that the time variation is much smaller than the variation with cpd.

We will add this statement.

8. On p.5, line 8, what are "tares"? I assume this refers to the "gap, steps, or spikes" from the previous sentence. Consider removing this word such that it reads "These are essentially...".

Tares refers indeed to all kinds of anomalies. It is frequently used in gravimetric jargon. We will remove it from this paper, clearly mentioning gaps, steps and spikes.

9. On p.5, line 17, it is unclear to me what "precision" means in this context. Is this the noise expressed in the power spectral density of the gravity signal? Or is it something else (such as mentioned in the abstract on p.1, line 15). Please briefly mention in the paper what is meant by the precision here.

We propose to replace :

"At this period, the SG at the Membach station is able to monitor with a precision of 20 $(nm/s^2)^2/Hz$, corresponding to 0.2 nm/s^2 at a period of 300 s (Van Camp et al., 2005)."

by

"The power spectral density of the SG at the Membach station is at the level of 20 $(nm/s^2)^2/Hz$, which corresponds to a precision of 0.2 nm/s^2 at a period of 300 s (Van Camp et al., 2005)."

and, in the abstract, to replace

"The precision of the gravimeter is a few nm/s^2;"

by

"The precision of the gravimeter is a few tenths of nm/s^2;"

10. On p.5, line 32, consider removing the word "large".

Will be done.

11. On p.8, line 23, the mean bias is defined as the ratio of the radar sum and the gravimeter sum. I generally interpret "bias" to mean the systematic error, that becomes negative in case of underestimation (and I think more readers would, too). I therefore recommend expressing the bias as the sum of differences divided by the sum of the reference (i.e., the radar). In practice, this means subtracting 1 from the original numbers. But, in my view, it does give more clarity.

Good suggestion.

12. On p.8, line 30, consider removing the word “falls”.

Will be done.

13. On p.10, line 32, consider removing the word “falls”.

Will be done.

14. On p.11, line 8, consider removing the word “fall”.

Will be done.