

Authors reply to interactive referee comments C1382:

We want to thank the reviewer for an encouraging introduction and thorough questions/comments. For detailed answers, see below.

Replies to Dr. Hidde Leijnse (referee #1, RC C1382)

Q1. It should be made very clear what the novel contribution of this paper is, given the papers on this topic that the authors cite. One of the novel aspects could be that this study is carried out at high latitudes, but then the difference between the summertime climate of Finland should be compared to that of e.g. Belgium (Goudenhoofdt and Delobbe, 2009).

A1. In this paper, the new combined Regression- and Barnes- (RandB) method is considered to be the main novel contribution. This will be made more clear in the revised version of this article (e.g. on page 2455, line 23-28 and on page 2456, lines 3-5). The new results can be used as a reference by those who use LAPS analyses in their applications (e.g. forest fire index) or apply LAPS in other environments, even though the method itself does not depend on LAPS.

As a secondary novel aspect, the location of our studies (high latitudes) can be accentuated, as suggested by the reviewer. We will therefore add a short description of Finland's climatology in the introduction. Preliminary suggestion:

“According to the classic Köppen classification, the climate of southern coastal Finland belongs to class Dfb, and the rest of the country to Dfc, i.e., a cool and moist continental/subarctic climate of cold and snowy winters and precipitation throughout the year. Summer is warm, not hot, and in the north it is also short (Jylhä et al., 2010). The only mountains are in northern Finland but do not exceed 1350 meters, while Finland is embraced by two Gulfs of the Baltic Sea (Gulf of Finland, Bothnian Bay) from two sides.”

Jylhä, K., Tuomenvirta, H., Ruosteenoja, K., Niemi-Hugaerts, H., Keisu, K., and Karhu, J. A.: Observed and projected future shifts of climatic zones in Europe and their use to visualize climate change information, Weather Climate Society, 2, 148–167, 2010.

Q2. One of the conclusions of the Goudenhoofdt and Delobbe (2009) paper was that geostatistical methods consistently outperformed other methods (for daily accumulations). What is the reason for not testing these methods in this study? It would have made sense to me

if this was at least discussed, and I think this should be done in a revised version of the manuscript.

A2. It has been found that there are many optional merging methods in the literature, which improves the results more or less, and some methods are more beneficial for operational usage (as discussed in Sect. 1, page 2455, lines 15-23). One of the results from Goudenhoofdt and Delobbe (2009) was that the geostatistical methods are more sensitive to the network density of gauges and that the more simple methods were less affected by the shortage of surface precipitation stations. In Finland we have a low density network of surface gauges and this is therefore one reason to first consider simpler methods, such as those used in this article.

In our study, the two methods Regression and Barnes have been selected because they are suitable for operational usage and improve the final results. Moreover, these two methods were available and used in other specific parts of the LAPS analysis (e.g. wind- and temperature processes) and it was therefore reasonable to further develop LAPS to make use of these methods for the radar-gauge assimilation. This is mentioned in the article page 2455, line 28 - page 2456, line 2.

Comparisons with more sophisticated methods, such as geostatistical methods, will be considered in upcoming future work and articles, but is unfortunately out of scope in this paper.

We suggest to add the following statement in the revised article, before the last paragraph in the Introduction, (preliminary suggestion of text):

“Geostatistical methods have shown good results in other studies for daily accumulation sums (e.g. Goudenhoofdt and Delobbe (2009)). However, they are sensitive to networks density, and the density of stations measuring hourly precipitation in Finland is very low. Therefore, in this paper we concentrate in further development of methods already used in LAPS.”

Q3. It is unclear to me where LAPS adds anything to the paper. In the description of LAPS (Section 2.1) it is stated that “the LAPS suite implemented at FMI is able to process several types of in-situ and remotely sensed observations” (p.2457, lines 12-13). However, it is not clear from this section how these data are processed. If I understand correctly from Section 3.1, the only thing LAPS does is convert radar reflectivities to rainfall rates and subsequently computes hourly accumulations. If this is the case, I don’t think it is necessary to mention LAPS (simply stating the employed Z R -relations and that accumulations are computed by summing the 5-

minute data should suffice). This would simplify the paper and improve its readability.

A3. There is a reference to Albers et al. (1996) in Sect. 2.1 (page 2456, line 16) where the different processes and usage of in-situ and remotely sensed observations are explained in more details. It is correct that LAPS does the conversion between radar reflectivity and rain rates. Though, within LAPS ingest process it reads in the radar volume scans and remaps them on the LAPS Cartesian grid, then combining the different radar stations into a 3D radar mosaic and finally, calculates the rain rates from the low level scans in the accumulation process. Following recent developments, LAPS also now performs the merging/assimilation with rain gauges, with a choice of assimilation methods (discussed in this paper). All transformations and calculations therefore take place on the same platform. LAPS, which is a free of charge analysis tool, is used worldwide and we believe that these results are important, both for LAPS users (existing and potentially new user of this tool) and those who consider these assimilation methods. Therefore, we want to keep LAPS within the article.

We will better clarify the usage of LAPS in the text (Sect. 2.1, page 2457, line 16 and onward), along the following lines:

“Within LAPS ingest process the radar volume scans are read in as NetCDF format files, thereafter the data is remapped to LAPS internal Cartesian grid and finally the 3D mosaic process combines the different radar stations (Albers et al., 1996). The rain rates are calculated from the lowest levels of the LAPS 3D radar mosaic data via the standard Z-R equation formula, which is then used for precipitation accumulation calculations, either as radar only accumulation, see Sect. 3.1, or merged with gauge observations, see Sect. 3.2, 3.3 and 3.4.

Q4. Information on how surface precipitation is computed from volume radar data is scattered over several sections (Sections 2.1, 2.2, and 3.1). This is confusing, because there seem to be some contradictions. For example, it is still not clear to me whether a VPR correction is used or not. On p.2458, lines 15-18 the FMI VPR correction scheme is discussed, but on lines 20-21 of the same page HESSD it is stated that volume data are used in this study. This implies that no VPR correction is applied to these data, as the result of such a correction would not be volume data but surface reflectivity. In Section 3.1, in the description of the LAPS_radar product, no mention is made of a VPR correction. I suggest rewriting and merging Sections 2.1, 2.2, and 3.1. The resulting section should include a clear description of how hourly rainfall accumulations are computed from 5-minute volume radar data. This should also include information on

how rainfall estimated from different radars are combined. Rewriting this section will probably be easiest if LAPS is removed from the paper (see my previous comment).

A4. We consider it important to keep LAPS in the context and we will therefore clarify, rewrite and organize the involved sections along the following suggestion:

In Sect. 2.1 we will explain better how LAPS reads in the 3D volume radar data, in order to make the calculation of reflectivity to rain rate conversion in Sect. 3.1. See also answer to question 3.

In Sect. 2.2 we will reformulate the text accordingly:

"The uncertainty factors affecting radar reflectivity are the electronic mis-calibration, calibration differences between radars, beam blocking, attenuation due to both precipitation (Battan, 1973) and wet radome (Germann, 1999). At mid-latitudes, the main source of uncertainty of radar-based rainfall estimates is the vertical profile of reflectivity, which causes a range-dependent error (Zawadski, 1984). At large distances, the radar probes the upper parts of the cloud, where reflectivity is weaker. In FMI's general radar processing chain, this is compensated with the VPR- (Vertical Profile of Reflectivity) correction, which also compensates for overestimation in a melting layer when appropriate (Koistinen et al., 2003). The LAPS system used in this study processes 3D volumes so no separate correction is needed. Clutter is removed with Doppler-filtering, and any residual clutter with a post-processing procedure based on fuzzy logics (Peura, 2002).

Comparing radars and gauges, an additional challenge arises from the different sampling sizes of the instruments. Radar measurement volume can be several kilometers wide and thick (one degree beam is ca. 5 kilometres wide at 250 kilometres), while the measurement area of a gauge is 400 cm² (weighing gauges) or 100 cm³ (optical instruments).

In this study, the radar data were used as volume measurements, repeated every 5 minute and consisting of 5 elevation angles, typically between 0.4 and 45°. Details of the FMI radar network and processing routines are described in Saltikoff et al. (2010)."

In Sect. 3.1 (page 2460, lines 24-26) we will clarify that LAPS has read in the radar reflectivity via ingest processes and that it is converted within LAPS

routines to precipitation intensity via the Z-R equation (Marshall and Palmer, 1948).

We believe that this will clarify how the radar information is read into and used in LAPS, and more clearly explain how the 1 hour rainfall accumulation is computed. In this way we can keep the structure where Sects. 2.1, 2.2 and 3.1 are included as they are.

Q5. In Section 2.2 the error sources in radar rainfall estimates are discussed, but no mention is made here of the errors related to variations in drop size distributions. This is discussed later, but should be included here. This should be solved by rewriting several sections (see my previous comment).

A5. This will be rewritten and clarified, see also answers and suggestions to question 4.

Q6. On p.2458, line 11, “calibration differences between radars” can be removed in the listing of sources of error in radar rainfall estimation because this is the result of the already mentioned “electronic mis-calibration” (p.2458, line 10) of one or more radars.

A6. We agree and this will be removed/changed.

Q7. On p.2459, line 8, the uncertainty of the weighing gauge is stated to be 0.2 mm. Is this including or excluding the sources of error discussed on lines 4-6 of the same page?

A7. Excluding, we will clarify this by using the word “random”. Re-formatting the text at page 2459, lines 4-6:

“Weighting gauges are subject to different sources of random errors such as mechanical malfunction, wind-drift (Hanna, 1995) and icing, which all affect the accuracy of measurements.”

Q8. On p.2459, lines 12-19, the FTA gauges are discussed. Can the authors give an indication of the uncertainty in these measurements? And how does this compare to that of the weighing gauges?

A8. The FTA has not published accuracy estimates and Vaisala instrument manual does not give any other information than stated in this article (e.g. precipitation detection sensitivity).

K. Wang has studied the performance of the PWD22 sensor used in FTA stations at the Egbert observatory in Canada, and the VRG weighting gauge, currently used at many FMI stations, comparing them to Geonor weighting gauges in double fence. His study shows that the PWD22 has larger negative mean error (underestimation) and more than four times larger standard error than the VRG. ("Performance of Several Present Weather Sensors as Precipitation Gauges "WMO TECO 2012).

Link to Wang article:

https://www.wmo.int/pages/prog/www/IMOP/publications/IOM-109_TECO-2012/Session1/P1_30_Wong_Performance_Wx_Sensors_Precip_Gauges.pdf

The error source we referred to are more related to the siting of the road stations than instrument errors. These are especially hard to quantify. All we can say is that it is obvious that the road stations sites have not been selected for best meteorological quality or representativeness.

For the comparison of weighting gauges and road stations we have the following plot (see Fig. 1) of Radar/gauge ratios ("+" is road station "O" is Weighting gauge). The coarse overestimation by road gauges is rather visible in the figure below, especially for the larger rainfall amounts. Fig. 1 is included here only for this discussion, it will not be include in revised version of the article, since it is based on different datasets than used in the article.

For the next version of the manuscript, we will attempt to illustrate this in other ways and add a summary of the discussion above.

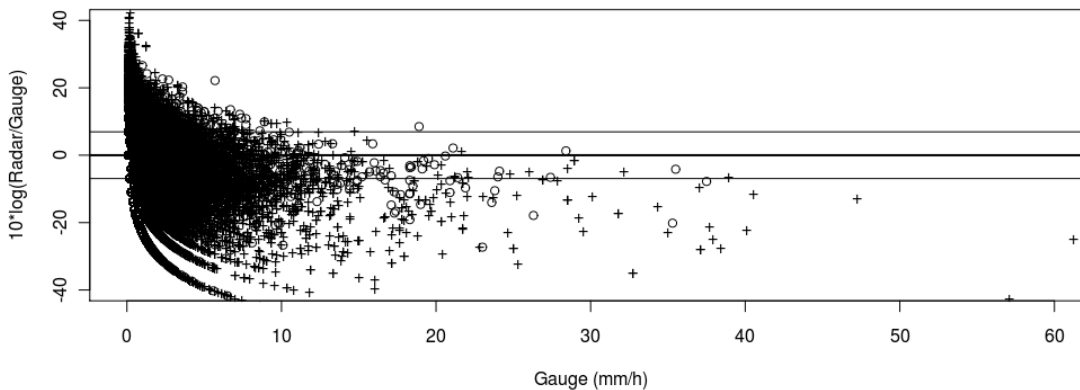


Fig. 1. Logarithmic radar-gauge relationship as function of precipitation amount as measured by the gauge. Below the 0-line radar < gauge, above the line radar > gauge. Weighing gauges in “o”, optical instruments in “+”.

Q9. On p.2461, lines 3-4, two Z R -relations are given (one for rain and one for snow). How do you determine which one of these to use?

A9. In order to avoid confusion, we will change the sentence in page 2461, lines 3-4 to be:

“In FMI’s implementation of LAPS we used: $A=315$ and $b=1.5$ for liquid precipitation, which is relevant in this study carried out during summer period.”

(Note to the reviewer: The determination between rain and snow phase is done within LAPS, from the analyzed temperature profile performed at every grid point.)

Q10. In Section 3.2, the linear regression analysis method is discussed. Can the authors briefly describe how this method relates to methods that have been presented in the literature?

A10. Similar linear regression methods, e.g. “multiple linear regression” (several explanatory variables) compared to our used “simple linear regression” (one explanatory variable), have been used in other environments when merging radar and gauge observations for precipitation accumulation calculations (Sokol, 2003a and Sokol, 2003b). We will study this literature (and others if found) and add a short paragraph with reference, within the next version of the article.

Sokol, Z.: Utilization of regression models for rainfall estimates using radar-derived rainfall data and rain gauge data, J. Hydrol., 278, 144-152, 2003a.

<http://www.sciencedirect.com/science/article/pii/S0022169403001392>

Sokol, Z.: The use of radar and gauge measurements to estimate areal precipitation for several Czech river basins, Stud. Geophys. Geod., 47, 587-604, 2003b.

<http://link.springer.com/content/pdf/10.1023/A:1024715702575.pdf>

Q11. On p.2461, lines 20-28 and on p.2463, lines 1-20, it is discussed how some data are discarded because of extremely low or high gauge/radar ratios. Why are these criteria different for the different methods? And what effect does this have on the results?

A11. Finland is 1500 kilometres long and we often experience stratiform rain and drizzle in one part of the country, and at the same time rain showers in another part. In the linear regression analysis (see page 2462, lines 4-7) a systematic bias between gauge and radar values is estimated, considering whole domain (whole Finland). Here we don't want to include too high gauge/radar ratios, because those might reflect very local precipitation patterns, taking place in only part of the domain, and we don't want this to affect to the more general/systematic correction we are looking for with this method. Therefore we are using tighter thresholds within the Regression method. In the Barnes method we allow more freedom (e.g. a more aggressive approach) when doing the corrections, because this is taking place on local scale. Here are also tendencies (e.g. large gauge/radar ratios) weighted into the calculation (see page 2463, lines 11-20). The criteria, used in this article, are based on our sensitivity studies for different meteorological cases (convective, frontal etc.) in Finland.

Q12. On p.2462, Eq.(4), it is not clear to me how the variable Y (the “corrected radar estimate”) is computed. Please clarify.

A12. The article explains on page 2462 (lines 3-6) how Eq. 4 is being produced, e.g. the outcome are values for k and c. Then, as a next step, Eq. 4 is again used to produce the corrected radar value by inserting the first-guess radar value as X and the equation is solvable since k and c are known. This is

performed at every LAPS grid point where there exists a first-guess radar value.

This will be clarified in revised article with suggestion of text (page 2462, lines 3-6):

“Once these criteria are enforced, the remaining data form a dataset of representative gauge-radar pairs from which a linear regression can be established, calculated with the least-square method, which minimized the errors between the measurement pairs. The outcome are values for k and c in the linear regression formula:”

And changing the text on page 2462, lines 8-10:

“The next step is to calculate the new corrected radar estimate using Eq. (4). In Eq. 4, Y is the corrected radar estimate, X is the first-guess accumulation from radar and the regression coefficients; k (the slope) and c (the interception point with the y -axis) are derived from the regression analysis.”

Q13. On p.2464, Eq.(6), what is the value of W_b that was used here?

A13. The weight W_b is set to be 0.02 in the LAPS code. We will add this clarification to W_b in the revised article.

Q14. On p.2464, lines 4-5, can the authors describe how the values of gamma are successively decreased?

A14. The radius of influence is decreasing by a factor of 2, for each pass/iteration. This information will be added to those lines, in the revised article.

Q15. On p.2464, line 6, the conditions for the iteration to terminate are given. How is the RMSE that is used as a criterion computed? And how often is the iteration terminated because the maximum number of iterations has been reached, and what are typically the values of RMSE in these cases?

A15. Information on the number of iterations is available from log-files but these files are not stored after the runs (due the amount of archiving space). The iteration is usually terminated by the RMSE threshold (e.g. RMSE=0.13), seen from an experimental run (5 days period with precipitation). The results showed that the iteration was terminated 100 % of the time due to that the RMSE value was reached and 0 % due to that the maximum iteration number was reached. The average number of iterations performed was 3, for this experimental run. Theoretically, if there are a lot of stations of varying densities, these are the cases which could enforce the calculations to use the maximum iteration count.

Q16. On p.2465, Eq.(8), the expression given in this equation is simply the mean error (or bias). I think that the sum should be over |Analysis - Gauge|

A16. Correct, the denominator will be changed in Eq. (8) to be the sum over |Analysis-Gauge|

Q17. On p.2465 line 23 - p.2466, line 13, the authors draw some conclusions from the analyses based on the comparison of the corrected data to the rain gauges that have been used to correct the radar data. I don't think these conclusions can be drawn based on these analyses. Instead the authors could consider using techniques such as cross-validation (i.e. removing a gauge from the dataset used for correction of the radar data to use it for verification)

A17. In the validation we have used both dependent (gauges used when correcting the radar data) and independent (gauges removed from dataset, i.e. not used when correcting the radar data) datasets. The independent gauges were used for "cross-validation". See text related to the independent dataset: page 2466, line 14 - page 2467, line 2.

Q18. On p.2466, lines 22-25, the reasons for the differences between the independent and dependent verification are discussed. It is argued that high accumulations have a large impact. Can the authors elaborate on why this is the case? Are there relatively more high-intensity values in the independent dataset?

A18. In the general case of a linear curve fit, values far away from the average tend to dominate. In the case of precipitation, which is not normally distributed, the average is near the lower end of the distribution and hence the larger values dominate linear the fit. We will explain this more in the corrected version article (perhaps even with distributions).

Q19. On p.2467, lines 1-2, for which method and at what intensity is this systematic overestimation observed?

A19. This statement was based on the dataset from RandB method in Fig. 5d. Looking at Fig. 5a-b one can see an increasing error of the absolute difference. While in Fig. 5c-d, this trend is not that pronounced and the absolute difference is on average about 5 mm/h, when considering higher accumulation values (> 7 mm/h).

We will improve the images (see answer to question 22 below) with better visual appearance of the data and evolve the discussion in order to make this statement more substantial. Preliminary new text at page 2466, line 28 - page 2467, line 2:

"The same trend as with dependent station data is observed: less dependence of the RandB method with increasing precipitation accumulation (>7 mm/h) tending an average absolute difference of approximately 5mm/h, which is lower than with LAPS_radar or Regression data."

Q20. On p.2468, lines 10-14, new information is given on the correction methods. This should be described in the sections describing the methods, not in the conclusions.

A20. This will be corrected. Move this information (page 2468, lines 9-14) to Sect. 3.2.

Q21. In Tables 1 and 2, the standard deviation of the R=G ratio is given. Because of the high skewness of this ratio (minimum 0, maximum infinity, and mean approximately 1), I think the standard deviation of the logarithm of this ratio would make more sense.

A21. We will change (or add) this information in the revised article.

Q22. In Figures 2-5, I suggest using shading or colors to indicate density of points in these graphs (bivariate histograms). This will aid interpretation of these graphs.

A22. We will improve the readability of these graphs by using another plotting tool, either by using shading or colors, as suggested.

Q23. In Figure 6, could the mean rainfall intensity values be added to the figure?

A23. The mean rainfall intensity will be added to Fig. 6.