Anonymous Referee #1
Received and published: 17 April 2016

The aim of the paper is twofold: (i) present and assess a novel operational methodology to include lightning information in radar-gauge precipitation accumulations and (ii) analyze the impact of different integration time intervals in the radar-gauge correction method.

The topic of the paper is of interest for the readers of the journal and the manuscript is well written and concise. The idea of including lightning information in precipitation estimation for intense events is challenging and very interesting both from operational and research points of view.

Nevertheless, the methodology used for the assessment of the new method is not adequate to the purposes of the method and masks out any improvement provided by the method itself, that, as is currently presented, looks almost useless. For this reason I recommend that the study undergoes a major revision before publication. In the following my major concerns and a list of minor comments.

AUTHORS:
The authors want to thank the reviewer for the professional and thorough revision of this paper. The new updated article version is attached as Supplement (see PDF-file)

Major comments:

1. The phenomenon of lightning is usually associated to convection, that is generally characterized by relatively small spatial scales. Such meteorological events are known to be difficult in terms of quantitative precipitation estimation (QPE) because: (i) owing to their small spatial scales are difficult to be adequately sampled by gauges and (ii) radar system may experience important problems due to attenuation of the signal, hail contamination and other issues. Therefore, the use of the LDA potentially represents an important source of information for improving the QPE for such situations. Despite this, results presented in this work show no significant improvement when LDA is used together with the already implemented system (Radar + RandB). If I understood correctly, the information provided by LDA is equivalent to a radar profile of reflectivity corresponding to locations and times in which a lightning occurred. This information is local in terms of space and time (as shown in Fig 6), therefore the potential effects of the use of LDA cannot be detected when large scales (the whole Finland) and long periods (seasonal) are used for the assessment, as they would be masked out. The authors partially recognize this problem and focus on a shorter period (the short, 4-days period) but keep on analyzing the country-scale picture. Furthermore, the use of only 7 independent gauges strongly limits the potential of the study, because of the small scales in which lightning information is available. In fact Tab. 2 confirms this: absolutely no information is available for the short study period (the more interesting one). I would recommend to revise the analysis as follows: (1) limit the analyses, both in space and time, to rainfall events characterized by lightning strikes; (2) select independent gauges in meaningful location for each event.

AUTHORS ANSWER:
We fully agree to the concerns expressed by the reviewer in above comments. Though, at that time, setting up this system during 2014-2015, this was the best we could do due to many reasons (please see below comments). We did learn much during this study and will improve the methods in future developments, accordingly.
Answer to 1): The focus is to improve the operationally running precipitation accumulation analyses, which use the spatial- and time resolution of 3 km and 1 hour, respectively. Gauge

information is available as 1 hour accumulation from our real-time database. Therefore, the time resolution for analyzed accumulation is bound to be on hourly data. The verification within this article was performed during operational runs. Hence, rerunning longer periods would require resources not available due to all the extensive data input, which would have to be re-generate (including retrieval/extraction of data and format conversions). For the 4-days period (year 2014) we manually saved the input data, in order to rerun experiment where we exclude/include lighting from the data ingest and test different profile relationship generations.
Answer to 2): Since the verification was performed during operational runs, the independent stations had to be set beforehand (i.e. excluded from the assimilation). Rerunning long periods with different independent stations, manually set for each event, would require extensive resources (see above explanation). By running the operational system for whole summer, we intended to retrieve a large statistical sample for verification. Unfortunately, summer 2015 was a period with very small amount of lightning cases. In the introduction we have added a short explanation of these limitations.

2. How are Fig 4 and Fig 8 obtained? Are they based on the dependent gauges? Do they show 1h estimates (I assume so since the figures show "mm/h" for the accumulations)? Using 1h estimates for the comparison with the dependent gauges (that are used on 1h scale for the RandB process) will have the Rad_LDA_RandB (1h product) necessarily being the best.

AUTHORS ANSWER:
Note: After the revision of the paper, all figures and much of the text have been reorganized.
Figure 4 was obtained from verification against dependent gauges and in Fig. 8 the independent gauges are used. This is now corrected and clarified in the new figure captions.
Yes, they are both given as hourly accumulation values. We changed units to read "mm" and then, in text and figure captions, we mention that it is "hourly accumulation values".
Table 1 and 2 shows results for the dependent gauges, here one can see that Rad_LDA_RandB give same results as the Rad_RandB. The same result is achieved from the independent gauges (now also mentioned in the article text).

3. I suggest to choose one between r2 and Pareson's correlation coefficient since the two statistics provide the same information. Moreover, basing results on RMSE can be tricky because errors are not weighted.

AUTHORS ANSWER:
Yes, we agree. We have now removed the coefficient of determination (R^2) from the verification.
We would prefer to keep the RMSE, since it is widely used in literature and it is something that readers are used to interpret.

<u>Minor comments:</u>

1. The title should include more clearly the second objective of the study (impact of different integration time intervals in the radar-gauge correction method)

AUTHORS ANSWER:
The title is now changed: "*Improving the precipitation accumulation analysis using lightning*

***measurements and different integration periods*"**

2. lines 1-5: the sentence is difficult to read. Moreover the second objective of the study should be better stressed. What about: "Two main objectives are addressed: (i) the assimilation of lightning observations in radar and gauge measurements and (ii) the analysis of the impact of different integration time intervals in the radar-gauge correction method."

AUTHORS ANSWER:
We agree and have now changed the first paragraph to read:
"*The focus of this article is to improve the precipitation accumulation analysis, with special focus on the intense precipitation events. Two main objectives are addressed: (i) the assimilation of lightning observations together with radar and gauge measurements and (ii) the analysis of the impact of different integration periods in the radar-gauge correction method.*"

3. line 6: is the reference Gregow et al. (2013)?

AUTHORS ANSWER:
Yes, thank you. This is now corrected (year 2011 → 2013).

4. The state of art section (lines 28-39) is rather short and can be organized in a clearer way

AUTHORS ANSWER:
We have added text and references to the Introduction and organized it to become more clear. Added text include paragraphs:
"*The research of combining radar and surface observations, to perform corrections to precipitation accumulation, is well explored. Many have made developments in this field and much literature is available, for example Sideris et al. (2014), Schiemann et al. (2011) and Goudenhoofdt and Delobbe (2009). Recently, Jewell and Gaussiat (2015) compared performances of different merging schemas, and noted a large difference between convective and stratiform situations. In their study, the non-parametric kriging with external drift (KEDn) outperformed other methods in accumulation period of 60 minutes. Wang et al (2015) developed a sophisticated method for urban hydrology, which preserves the non-normal charactersitics of the precipitation field. They also noticed that common methods have a tendency to smooth out the important but spatially limited extremes of precipitation.*"
And:
"*Lightning is associated with convective precipitation, but in areas where a large portion of precipitation is stratiform, lightning data alone is not adequate for precipitation estimation. However, lightning has been used to complement and improve other datasets. Morales and Agnastou (2003) combined lightning with satellite-based measurements to distinguish between convective and stratiform precipitation area and achieved a remarkable 31% bias reduction, compared to satellite-only techniques. Lightning has also been assimilated to numerical weather prediction models to improve the initialization process of the model. This can be done by blending them with other remote sensing data to create heating profiles (e.g. estimating the latent heat release when precipitation is condensed). Papadopulos et al. (2005) used lightning data to identify convective areas and then modified the model humidity profiles, allowing the model to produce convection and release latent heat using its own convective parameterization scheme. They combined lightning with 6-hourly*

***gauge data, within a mesoscale model in the Mediterraiean area, and showed improvement in forecasts up to 12 hours lead time.*"**

5. line 47: " usually with a higher quality than radar" a reference can be helpful

AUTHORS ANSWER:
We have clarifed the sentence and added a reference in Sect. 2.1: "*Rain gauges provide point observations of the accumulation. They are usually considered more accurate than radar, as point values, and are frequently used to correct the radar field (Wilson and Brandes, 1979).*"

6. lines 54-55: "long" rather than "longer", "short" rather than "shorter"

AUTHORS ANSWER: This has now been changed.

7. line 62: more information about how "poor data quality" stations are identified is needed

AUTHORS ANSWER:
Clarified in Sect. 2.1 by following sentence: "*If measurements consistently indicate poor data quality, either manually identified from station error-logs or by inspecting the data, those stations are blacklisted within the LAPS process and do not contribute to the precipitation accumulation analysis.*"

8. line 70: Lat-Lon information are not shown in the figure

AUTHORS ANSWER:
We have now rephrased the sentence to: "*As Finland has no high mountains, the horizon of all the radars is near zero elevation with no major beam blockage, and, in general, the radar coverage is very good except in the most northern part of the country.*"

9. lines 70-72: something is missing in the sentence

AUTHORS ANSWER:
Changed to read: "*During year 2014 and 2015 the utilization rate was > 99%.*"

10. line 108: I couldn't find the work by Pessi and Albers, 2014

AUTHORS ANSWER:
We have now updated the reference with a web-link of the presentation: https://ams.confex.com/ams/94Annual/webprogram/Paper238715.html

11. lines 120-124: this is not useful for the purposes of the paper

AUTHORS ANSWER:
We would like to keep these sentences about existing "default profiles". Because, we believe it is relevant to point out that for experimental/operational usage, anywhere in the world, there is a direct possibilty to use and test the LDA method without collecting new, own statistical relationships.

12. line 121 and 127: I couldn't find the work by Pessi, 2013

AUTHORS ANSWER:
We have now updated the reference with a web-link of the presentation: https://ams.confex.com/ams/93Annual/webprogram/Paper215562.html

13. line 184: why 0.3? more details are needed

AUTHORS ANSWER:
The threshold value for the hourly surface gauge measurements, retrieved from FMI real-time database, is 0.254 mm/h (i.e. everything below 0.254 mm/h is just 0).
The sentence changed and moved to Sect. 3.4: "*In this study we apply a filter to the verification datasets, where hourly accumulation data less than 0.3 mm are discarded (due to the lowest threshold value of surface gauge measurements from FMI real-time database).*"

14. Fig 7: the colors of the regression lines are not explained in the caption

AUTHORS ANSWER:
We have clarifeid and added the following to figure caption "*The corresponding regression lines are represented with same color as the markers, for each method.*"

H. Leijnse (Referee)
hidde.leijnse@knmi.nl

Received and published: 17 April 2016

This paper describes an assessment of the quality of quantitative rainfall estimates using a combination radar, cloud-to-ground lightning, and rain gauge data. The effect of adding GC lightning data to radar data on rainfall accumulations is investigated, both before and after gauge adjustment. In these analyses, several methods of estimating relations between lightning activity and rain intensity are utilized. Furthermore, the effect of the length of the accumulation interval used for gauge adjustment is also studied. The paper is interesting, and its topic relevant. It is not entirely clear to me what the main goal of this paper is. I think that the paper could benefit from a clearer description of what its main goals are, how the analyses that are presented contribute to these goals, and coming back to these goals in the Discussions and Conclusions section. When reading the paper for the first time I was sometimes confused because new analyses are proposed in the Results section and some of the methods described in the Methods section were not entirely clear to me. Hence, the paper could benefit from some restructuring, where all methods that are used are presented clearly in the Methods section. I think that the paper needs major revisions in order for it to be suitable for publication. More specific remarks are given below.

AUTHORS:
The authors want to thank the reviewer for the professional and thorough revision of this paper. The paper has undergone a significant reorganization and has now a better structure. Please see the new updated article version, attached as Supplement.

Specific comments

1. Section 1, Given the fact that there are not very many lightning strikes in Finland, how much would you expect that adding this information would influence the final rainfall estimates? I think that this should be thoroughly discussed in the introduction of the paper.

AUTHORS ANSWER:
Yes, it is correct that due to low lighting frequency in Finland the quantitative effect during a year is small. But the goal is to improve the quality of the few (but important) existing intense precipitation cases (i.e. causing flash floods etc). Here the LDA method have an impact, since the largest uncertainties took place during heavy rainfall (i.e. convective weather situations and lightning; Gregow et al., 2013). We have added text related to this in the Introduction section:
"Radar reflectivity can in some cases suffer from poor quality, resulting from electronic mis-calibration, beam blocking, clutter, attenuation and overhanging precipitation (Saltikoff et al., 2010). In some cases the radar can even be missing, due to upgrading or technical problems. Thunderstorms add probability of many of these problems in form of interruptions in electicity and telecommunications, and attenuation due to intervening heavy precipitation. In general, combining radar and rain gauge data is very difficult in the vicinity of heavy, local rain cells (Einfalt et al., 2005)."
Also, the intention is to enlarge the analysis area to whole Scandinavia. For this reason the LDA will have a larger contribution to the precipitaion accumulation analysis, since there are gaps in radar coverage for this area and the retrieval of data is not always stable (i.e. radars can be missing more frequently from neighbouring countries). This is mentioned in the Introduction

with following text:
"Our situation is different from the above mentioned experiments because lightning activity is usually low in Finland, compared to warmer climates (Mäkelä et al., 2011). Also, our analysis area already has a good radar coverage and relatively evenly distributed network of 1 hour gauge measurements. However, if we want to enlarge the analysis area, we will soon go to either sea areas or neighbouring countries where availability of radar data and frequent gauge measurements is low. Our principal goal is to have as good analysis as possible, which is different from having a best analysis to start a model."

2. Section 3.3, It's not entirely clear how radar and lightning data are merged to come up with a final rainfall estimate. Did I understand correctly that the number of recorded GC lightning strikes within a LAPS pixel (3x3 km) and in a 5-minute interval are counted. These counts are then related to a vertical reflectivity profile, and subsequently the maximum of the radar reflectivity and the 'lightning reflectivity' are taken at a given height. The rainfall estimate is then based on the lowest data point. In practice, this means that the rainfall estimate is based on the maximum of the lowest-level radar reflectivity and the lowest-level 'lightning reflectivity'. If this is indeed the case, the description of the method to estimate rain rates could be simplified and clarified. If not, I recommend clarifying this section.

AUTHORS ANSWER:
Sections 3.2 and 3.3 are now merged. The text is reorganized and we have clarified the process better.

3. l.137-149, Given the fact that lightning only occurs in convective situations, it would make sense to me if a Z-R relation specifically derived for convective rain is used wherever lightning is observed. This would be a simple addition to the LDA that could improve results even further.

AUTHORS ANSWER:
Thank you, this is a very good suggestion.
This article presents the first results in an on-going process of developing the LAPS-LDA system at FMI. We have thought of different ways to improve the system, learned much during this study and the plan is to implement new routines, in future versions of LAPS-LDA system. The suggestion by reviewer is clearly one that should be considered then.

4. l.174-177, the rationale behind the regression part of the RandB method is that radar rainfall estimates often suffer from large-scale multiplicative biases, and that using regression on radar and gauge data can correct for this error. When adding lightning data to radar data, the errors are likely to be very different, and this could have a large effect on the final rainfall estimates. Something similar can be said for the Barnes-part of the RandB method, where the influence of a gauge correction is in general relatively large compared to the area affected by lightning. I therefore strongly suggest to add a discussion of this in the paper.

AUTHORS ANSWER:
We agree that this is a simplification of mixing different scale-processes. In the Introduction we included following text:

"Lightning is associated with convective precipitation, but in areas where a large portion of precipitation is stratiform, lightning data alone is not adequate for precipitation estimation. However, lightning has been used to complement and improve other datasets. Morales and Agnastou (2003) combined lightning with satellite-based measurements to distinguish between convective and stratiform precipitation area and achieved a remarkable 31% bias reduction, compared to satellite-only techniques. Lightning has also been assimilated to numerical weather prediction models to improve the initialization process of the model. This can be done by blending them with other remote sensing data to create heating profiles (e.g. estimating the latent heat release when precipitation is condensed). Papadopulos et al. (2005) used lightning data to identify convective areas and then modified the model humidity profiles, allowing the model to produce convection and release latent heat using its own convective parameterization scheme. They combined lightning with 6-hourly gauge data, within a mesoscale model in the Mediterraiean area, and showed improvement in forecasts up to 12 hours lead time."

And we added text about this in the discussions section:

"In the RandB-method the Regression is used to correct for large-scale multiplicative biases between radar and gauge data. In this article we introduce lightning into the RandB-method, as an additional data source. However, lightning errors are likely to be different from those of radar and gauges and this could have an effect on the methodology used here. In future developments, after collecting longer time series to quantify the nature of uncertainty of lightning-based precipitation estimates, we intend to improve the analysis in this direction."

5. l.176-177, What does it mean that Rad_LDA_Accum is the reference?

AUTHORS ANSWER:

This sentence (now in Sect. 4.2) is changed to: *"Note that Rad_LDA_Accum (e.g. a method not using RandB, as an reference) is included when comparing the results of different integration periods."*

6. Section 4, Why are the graphs where rainfall intensities are compared plotted on log-log scales? If the aim is to study the performance of quantitative precipitation estimation algorithms for high intensities (as is stated in the paper), it would make most sense to me if these graphs were plotted using linear axes.

AUTHORS ANSWER:

Yes, the intention is to increase the readability of high precipitation values but without disturbing the overal readability. Plotting the values on linear axes will decrease the readability of the low-middle values. The log-log scales was the best way we could produce these plots (according to us), after testing different plotting techniques (see below). Therefore we suggest to keep Figs. 5 and 7 with log-scales.

As an example we plot Fig. 5 in log-scale vs linear-scale (please see the attached Fig. 1).

And it is the same with Fig. 4 and 8, the visualization of data is more clear with log-scales. Here we show Fig. 4a with log-scale vs linear-scale (please see the attached Fig. 2).

7. l.187-188, what exactly is meant by "the averaged (i.e. 50%-percentile) Rad-Lig reflectivity profiles from the LDA-method."? How were these profiles determined, and based on what data? I think this should be discussed in the Methods section.

AUTHORS ANSWER:
It is now moved into Methods, Sect. 3.2 (merged and changed with other sections). The related text now reads:
"For this study over Finland, climatological Rad-Lig reflectivity relationship profiles were estimated using NORDLIS-LLS lightning information and operational radar volume data from Finland area, during summer 2014. A total of approximately 220'000 lightning strokes were used for this calibration. The FMI-LAPS LDA is using 5 minutes interval of lightning- and radar data, within a LAPS grid-box of resolution 3*3 km. The collected strokes are divided into binned categories using an exponential division (i.e. $2^n...2^{n+1}$), following the same method used in Pessi (2013). This result in 6 different lightning categories (e.g. with 1, 2-3, 4-7, 8-15, 16-31 and 32-63 strokes) for the NORDLIS-LLS dataset. For each of these 6 categories, the average radar reflectivity profile is calculated and gives the Average Rad-Lig profiles (Fig. 3a), which is the baseline method. We extend this method to also calculate the 3'rd Quartile (i.e. 75%-percentile) and a Variable Quartile Rad-Lig profiles. The Variable Quartile method uses a range between 50%-percentile (for the lower dBZ values) up to the 95%-percentile (for the highest dBZ values)."

In this answer (to reviewer) we also provide a plot which visualize the process. For each category we collect the relevant radar reflectivity profiles. From these selections of profiles, the average is calculted and further used as the LDA-lightning profile (please see the attached Fig. 3).

8. l.192-200, I suggest to remove the R2 statistic, because it is simply the correlation coefficient squared (see Eqs (6) and (7)) and it hence doesn't add any information relative to CORR.

AUTHORS ANSWER:
We have now removed the R2 statistics from the text and tables.

9. l.202-207, The panels of Fig. 4 with LDA added (i.e. panels b and d) do not really add any information, as they are extremely similar to panels a and c, respectively. I therefore suggest making a remark in the text about this, and removing either panels a and c, or panels b and d.

AUTHORS ANSWER:
We have added this into Discussion:
"*The accumulation products generated from RandB-method are corrected using gauge information. This process is influencing the final accumulation results much more than the contribution from the LDA-method (seen in Fig. 4 results from dependent dataset, where a, c and b, d panels, respectively, are almost identical). The same result was seen for the independent dataset.*"
We suggest to keep Fig. 4 as it is. Removing either a,c- or b,d-panels, only mentioning this in the text, would most probably result in contradicting comments by other reviewers (i.e. that this should be shown with figures).

10. l.208-212, I would strongly suggest using different gauges for the independent measurements to test whether using LDA improves rain estimates, because this is what I understand the main objective of this paper to be.

AUTHORS ANSWER:

The verification in this article was performed during operational LAPS runs (i.e. products are used within end-users applications). Seven independent stations were pre-selected (from different parts of Finland). Because of this we could not set more stations aside, without risking the quality of the end product. Re-running longer periods with different independent stations, manually set for each event and re-generate the extensive input datasets (retrieval/extraction of data, format conversions etc), would require resources not available.
By running the operational system for whole summer, we intended to retrieve a large statistical sample for verification. Unfortunately, summer 2015 was a period with very small amount of lightning cases. This restriction is now mentioned and explained in the Introduction:
"The work reported here has been performed using the operational Local Analysis and Prediction System (LAPS), which is used in the wether service of Finnish Meteorological Institute (FMI). Testing new approaches in an operational system has its limitations in e.g. excluding independent reference stations. Also the possibilities to rerun cases with different settings have been limited. The benefit of the approach is that we can be sure that we only use data which is operationally available."

11. l.209-210, The use of a 25-day subset is introduced here. I suggest introducing this earlier in the paper (the Methods section). And if this subset is used, what is the added value of using the 4-day subset? I think the clarity of the paper would improve if either the 4-day or the 25-day subset is used.

AUTHORS ANSWER:
The lightning information is local in terms of time (e.g. also in space). Therefore, the potential effects of the use of LDA is not detected when long periods (seasonal) are used for the assessment, as they are masked out. We are trying to show this by using different verification periods (summer-, 25- and down to the 4-days periods). The 4-days subset (for which we have saved all the extensive input data) also fills another purpose, namely being able to rerun and test different developments (such as the verification of average-, 3'rd- and Variable Quartile Rad-Lig profiles).
The paper has undergone many changes and is reorganized. We now introduce the 25-day subset in Methods, Sect. 3.4, together with the other periods, as follows:
"The verification periods consists of one long period ranging from 1 April to 1 September, 2015 (i.e. to avoid the winter season and snow precipitation). This dataset includes many precipitating cases without lightning and therefore, the effective impact by lightning is diluted (e.g. no influence by the LDA-method). Therefore, a subset of 25 days with frequent lightning (e.g. > 100 CG strokes/day) were selected from summer 2015. Additionally, in order to perform several autonomous experiments with the FMI-LAPS LDA system, a dataset consisting of four days with heavy rain and strong convection were used: 03, 23, 24 and 30 of July 2014 (hereafter 4-days period). These were the 4 days with highest lightning intensity (e.g. > 100 strokes/day) in Finland, during year 2014."

12. l.225-238, It's unclear to me how the new profiles are exactly generated. I strongly suggest to include a good description of this in the Methods section (preferably in Section 3.2).

AUTHORS ANSWER:
Please, also see reply to comment 7 here above.
We have now moved and merged sections. The description of Rad-Lig relationship profiles is now better explained in Sect. 3.2.

13. l.240-245, Why not test sub-hourly scales?

AUTHORS ANSWER:
The gauge information is available as 1 hour accumulation, from our FMI real-time database, and this is used in our operational runs. Therefore, the time resolution for analyzed accumulation is bound to be on hourly data.

Minor remarks

1. l.16, replace "such as;" by "such as" (remove semicolon)

AUTHORS ANSWER:
This is done.

2. l.17, replace "eceonomically" by "economically"

AUTHORS ANSWER:
This is done.

3. l.39, replace "leass" by "less"

AUTHORS ANSWER:
This is done.

4. l.41, what is meant by "a timely accurate manner"?

AUTHORS ANSWER:
This is changed to "...timely manner (i.e. near real-time).".

5. l.133, replace "resulting from;" by "resulting from" (remove semicolon)

AUTHORS ANSWER:
This is done.

6. l.133-134, consider including clutter as an important source of error

AUTHORS ANSWER:
Clutter has been added to the sentence.

7. l.144-145, do you mean to say here that convective rain is important for flooding events? I so, I suggest changing "such situations" to "convective events". The first time I read this sentence I

interpreted “such situations” to be the drizzle that is mentioned in the previous sentence.

AUTHORS ANSWER:
Yes, this is what we meant and it is now changed to “convective events”.

8. l.187, the 50th percentile is not the average, but the median, and it is either the 50th percentile or the 50% quantile. So I suggest replacing “averaged (i.e. 50%- percentile)” by “median (i.e. 50% quantile).”

AUTHORS ANSWER:
Correct, well spotted. We have remove the “(i.e. 50%-percentile)” here and in other places in the text, where this occur.

9. l.192, I suggest calling STDEV “relative standard deviation” or “logarithmic standard deviation” to make clear that it is different from a regular standard deviation.

AUTHORS ANSWER:
We have now changed this to read “the logarithmic standard deviation”.

Anonymous Referee #3

Received and published: 17 April 2016

This paper addresses the question of how to the use of lightning data can help improve rainfall accumulation estimation. The topic is relevant and paper on this issue are welcomed. However the manuscript exhibit severe flaws, and cannot be published in its current form. The modification needed require in-depth modification.

AUTHORS:
The authors want to thank the reviewer for the professional and thorough revision of this paper. The new updated article version is attached as Supplement (see PDF-file).

General comments:

Overall the paper is quite difficult to read, many processing techniques are tested on various data sets, presented not at the same time. Maybe a scheme summarizing the techniques would help. I think the paper should also be organized better. A solution could be to present "data", then methods with a subsection on the various products from the radar, gauge and lightning, and a subsection on how the comparison is performed.

AUTHORS ANSWER:
The paper has undergone a significant reorganization, which has now improved the readability accordingly. The methods have now been bundled from different sections, the observations have a better structure and the result section is more concise.

There seems to be a contradiction between the abstract and the content. abstract l.8-9 : "the performed... usefullness of..." and l.201 "The overall result... neutral to positive impact..." and same comment on the dependent sub-set (l.211-212).

AUTHORS ANSWER:
We have changed the second paragraph in abstract to be more consistent with the result section, which has been rewritten:
"Lightning data does improve the analysis when no radar is available, and even with radar, lightnings have a neutral to positive impact on the results."

The conclusion seems to be that basically when radar data is available lightning data is rather useless. This is already a result that should be stressed (it is already mentioned). That said I have the feeling that the paper could be more interesting by shifting its scope to how to estimate rainfall (locally) from lightning data when no radar data is available (this would correspond to developing more in depth what is done with figure 7). After this analysis, you could practically test the interest by artificially removing some radar data.

AUTHORS ANSWER:
We now stress the importance of lightning data for situations of no radar information in the abstract, the result- and discussion sections.
The results in this article was performed during operational LAPS runs, i.e. with all input data

availabe at that time. Hence, rerunning longer periods would require resources not available, due to all the extensive data input needed (i.e. regenerate the data input and format conversions). For the 4-days period we manually saved the input data in order to rerun experiment, where we exclude/include lighting from the data ingest and, additionally, test different profile relationship generations. This is now explained in the introduction:
"The work reported here has been performed using the operational Local Analysis and Prediction System (LAPS), which is used in the wether service of Finnish Meteorological Institute (FMI). Testing new approaches in an operational system has its limitations in e.g. excluding independent reference stations. Also the possibilities to rerun cases with different settings have been limited. The benefit of the approach is that we can be sure that we only use data which is operationally available."

Detailed comments:

The title "radar -,gauge-" formula is not very clear.

AUTHORS ANSWER:
The title is changed to: *"Improving the precipitation accumulation analysis using lightning measurements and different integration periods"*

1) Introduction
- It should be extended to include a state of art section on the actual topic of the paper, i.e. lightning measurement assimilation for rainfall or more generally in meteorology. (ex among many others: Papadopoulos et al. 2005; Morales et al 2003)

AUTHORS ANSWER:
We have now included text in the Introduction section, where work on this topic is elaborated (including references) as follows:
"Lightning is associated with convective precipitation, but in areas where a large portion of precipitation is stratiform, lightning data alone is not adequate for precipitation estimation. However, lightning has been used to complement and improve other datasets. Morales and Agnastou (2003) combined lightning with satellite-based measurements to distinguish between convective and stratiform precipitation area and achieved a remarkable 31% bias reduction, compared to satellite-only techniques. Lightning has also been assimilated to numerical weather prediction models to improve the initialization process of the model. This can be done by blending them with other remote sensing data to create heating profiles (e.g. estimating the latent heat release when precipitation is condensed). Papadopulos et al. (2005) used lightning data to identify convective areas and then modified the model humidity profiles, allowing the model to produce convection and release latent heat using its own convective parameterization scheme. They combined lightning with 6-hourly gauge data, within a mesoscale model in the Mediterraiean area, and showed improvement in forecasts up to 12 hours lead time.
Our situation is different from the above mentioned experiments because lightning activity is usually low in Finland, compared to warmer climates (Mäkelä et al., 2011). Also, our analysis area already has a good radar coverage and relatively evenly distributed network of 1 hour gauge measurements. However, if we want to enlarge the analysis area, we will soon go to either sea areas or neighbouring countries where availability of radar data and frequent gauge measurements is low. Our principal goal is to have as good analysis as possible, which is different from having a best analysis to start a

model."

2) Observations and instrumentation
- p2 l.49 : "LAPS" is used but was not defined before in the manuscript

AUTHORS ANSWER:
LAPS is now defined properly.

- p3 l.76: "as as proxy", one "as" should be removed

AUTHORS ANSWER:
This is corrected.

- section 2.2 : how mosaicking is done ? Some more detail on the radar processing should be provided. How dual polarisation is used? The differences in terms of sampling area between rain gauges and radar are almost not mentioned (l.81). This discussion should at least be expanded because it can have an influence on the standard scores used after. See for example reference such as Jaffrain and Berne 2012; Gires et al. 2014; on this issue

AUTHORS ANSWER:
The text related to LAPS processes is reorganized (now in Sect. 3.1), including a reference to (Albers et al., 1996): *"The Finnish radar volume scans are read into LAPS as NetCDF format files, thereafter the data is remapped to LAPS internal Cartesian grid and the mosaic process combines data of the different radar stations (Albers et al., 1996)."*

A sentence related to the use of dual polarisation radar is included (Sect. 2.2): *"At the moment, the quantitative precipitation estimation based on dual-polarization is not used operationally in FMI, but the polarimetric properties contribute to the improved clutter cancellation (i.e. removal of non-meteorological echoes, especially sea clutter, birds and insects)."*

The differences in sampling size (both spacial and temporal) has now been inlcuded in the introduction section, with text and references as follows:
"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 approximately 5 kilometres wide at 250 kilometres), while the measurement area of a gauge is 400 cm^2 (weighing gauges) or 100 cm^3 (optical instruments). Part of the disparateness of radar and gauge measurements is due to variability of the raindrop size distribution within area of a single radar pixel. Jaffrain and Berne (2012) have observed variability up to 15% of the rainrate in a 1x1 km pixel, with timesteps of 1 minute. Gires et al (2014) have shown that the scale difference has an effect in verification measures (such as normalized bias, e.g. RMSE) but it decreases with growing accumulation time (e.g. from 5 to 60 minutes). In our study, the 60 minutes accumulation period is smoothing some of the differences."

3) Methods

- l94 : randB method mentioned but not defined after

AUTHORS ANSWER:
This has been corrected.

- section 3.1 : more details on LAPS are needed. Sentences such as "LAPS uses statistical methods to perform high-resolution analysis" are too general for a scientific paper. It is not clear what is the purpose of LAPS and what data is used out of it, and how it is related to the radar mosaic product of FMI.

AUTHORS ANSWER:
We have now added more information and references to the LAPS processes:
"*The FMI-LAPS use a pressure coordinate system including 44 vertical levels distributed with a higher resolution (e.g. 10 hPa) at lower altitudes and decreasing with height. The horizontal resolution is 3 kilometres and the temporal resolution is 1 hour. The domain used in this article covers the whole Finland and some parts of the neighbouring countries (Fig. 1b). LAPS highly relies on the existence of high-resolution observational network, in both space and time, and especially on remote sensing data. The FMI-LAPS is able to process several types of in-situ and remotely sensed observations (Koskinen et al., 2011), among which radar reflectivity, weighting gauges and road weather observations are used for calculating the precipitation accumulation. The Finnish radar volume scans are read into LAPS as NetCDF format files, thereafter the data is remapped to LAPS internal Cartesian grid and the mosaic process combines data of 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 formula (Marshall and Palmer, 1948), which is then used for precipitation accumulation calculations (see Sect. 3.2). Other information on observational usage, first-guess fields, the coordinate system etc. is described in Gregow et al. (2013).*"

- Section 3.2: Some indication on the number of lightning strokes used to calibrate the relation should be given. Figure 3.b : the vertical scale should be changed to improve the visibility of the relation. Could you confirm that the temporal resolution is indeed of 5 min

AUTHORS ANSWER:
We included (in Sect. 3.2): "*A total of approximately 220'000 lightning strokes were used for this calibration.*"
We have changed y-axis scale in Fig. 3b, to make improve the readability.
Yes, the radar and lightning data have 5 minutes resolution and the final accumulation product is hourly.

- Eq 3 : so the dual pol capacities are not used ?
AUTHORS ANSWER:
At the moment, the quantitative precipitation estimation based on dual-polarization is not used, but the polarimetric properties contribute to the improved clutter cancellation (see also answer above).

- l.151-152 : when merging radar and lighting data, why choosing the maximum ? When radar data is available, is not radar more reliable than one choice among the 6 different profiles for lightning ? Please justify this choice.

AUTHORS ANSWER:
In most cases the radar data are more reliable, compared to lightning-profiles. Though, in cases where there occur attenuation or the grid-point is at the far end of radar coverage, the lightning-profiles could be of better quality and hence, in these situtaions would have higher reflectivity values (therefore we choose the dataset with higher value). We included text explaining this:
"The logic behind this is that the radars are more likely to underestimate, than overestimate the precipitation (due to attenuation, beam blocking or the nearest radar missing from network; e.g. Battan, 1973 and Germann, 1999), especially in thunderstorm situations."

- Section 3.4 : why only one sub section? You have to say at least few words on Rand B method and Barnes analysis. It is very difficult to understand this section with so little explanations.

AUTHORS ANSWER:
We admit we were perhaps too careful not to repeat too much from the earlier paper in Sect. 3.4. We now added new text describing the RandB-method:
"The FMI-LAPS RandB-method corrects the precipitation accumulation estimates using radar and gauges datasets. The first step in this method is to make the radar-gauge correction using the Regression method. Data of hourly accumulation values are derived from the gauge-radar pairs within the LAPS grid (i.e. from same location and time), and from this a linear regression function can be established. The corrections from Regression method is applied to the whole radar accumulation field and thereafter used as input for the second step, the Barnes analysis. Within LAPS routines the Barnes interpolation converge the radar field towards gauge accumulation measurements at smaller areas (i.e. for gauge station surroundings). Several iterative correction steps are performed within the Barnes analysis, adjusting the final accumuation. The FMI-LAPS RandB-method is described in more details in Gregow et al. (2013)."

4) Results and verification

- p. 7 is methodology and should be moved in the corresponding section. Please confirm that in eq. 4-7, values are taken at the hourly resolution ? It might be interesting to test other time steps.

AUTHORS ANSWER:
The result section has been rewritten and parts related to methods have been moved accordingly.
Yes, we confirm that these values are hourly.
We agree, it would be interesting and important to use lower time-resolution. However, the surface gauge measurements, coming from FMI real-time database, are given as hourly values. Therefore we are restricted to make the corrections using the time resolution of 1 hour.

- l.187-188 : "the avg Radlig reflectivity profile. Please clarify ?

AUTHORS ANSWER:
This text is moved and rewritten in methods section (Sect. 3.2, now merged with Sect. 3.3).

- Figure 4 : please clarify what is plotted and how it was obtained. It is also almost not commented in the text.

AUTHORS ANSWER:
The figure caption has been updated with more information:
"***Figure 4. The FMI-LAPS precipitation accumulation (described in plots with density iso-lines of hourly accumulation values, in mm and log-scale) calculated using 4 different methods: a) Rad_Accum, b) Rad_LDA_Accum, c) Rad_RandB and in d) Rad_LDA_RandB. Results are from the dependent gauge dataset during summer 2015. Shown is also the best fit line (1:1).***"

- Figure 5 : again mention time steps used (1h ?). l.220-224 : the figure should be more commented. - l.228-230 : the quantiles mentioned are not clear.

AUTHORS ANSWER:
We added information about hourly accumulation into the caption of Fig. 5. The result section has been rewritten to become more readable and asserted.

- Figure 7 and comments : the change in quantile seems to improve rainfall estimates. Why was not it used in the first place ?

AUTHORS ANSWER:
Note: Figs. 6-8 have now changed order.
Why not use the Quartile method in first place? This is because we had a test-dataset, a 4-days case from year 2014 (with all the extensive input data), which was used to perform autonomous experiments/reruns. But, the idea of calculating/testing different Quartile profiles were not thought of until we saw the results from summer 2015. Then, we made reruns using data from 2014 to establish these findings (i.e. Variable Quartile approach give the best profiles).

- Figure 8 : please comment more the figure...

AUTHORS ANSWER:
We have added text to the figure caption.

References (reviewer):

Extending the Capabilities of High-Frequency Rainfall Estimation from Geostationary-Based Satellite Infrared via a Network of Long-Range Lightning Observations Carlos A. Morales and Emmanouil N. Anagnostou Journal of Hydrometeorology 2003 4:2, 141-159

Improving Convective Precipitation Forecasting through Assimilation of Regional Lightning Measurements in a Mesoscale Model Anastasios Papadopoulos, Themis G. Chronis, and Emmanouil N. Anagnostou Monthly Weather Review 2005 133:7, 1961-1977

Gires, A. et al., 2014. Influence of small scale rainfall variability on standard comparison tools between radar and rain gauge data. Atmospheric Research, 138(0): 125-138.

Jaffrain, J. and Berne, A., 2012. Influence of the Subgrid Variability of the Raindrop Size Distribution on Radar Rainfall Estimators. Journal of Applied Meteorology and Climatology, 51(4), 780-785.

S. Ochoa Rodriguez (Referee)
s.ochoa-rodriguez@imperial.ac.uk

Received and published: 18 April 2016

This paper explores the use of cloud-to-ground lightening data for improving radarbased (and raingauge-adjusted radar-based) precipitation estimation, with a focus on high intensity, convective storms. The study includes an evaluation of resulting rainfall estimates at different temporal accumulations.

The topic of the paper is interesting and the results are potentially useful. However, the paper has a number of major flaws which should be addressed before it can be published.

AUTHORS:
The authors want to thank the reviewer for the professional and thorough revision of this paper. The new updated article version is attached as Supplement (see PDF-file).

General comments / major flaws:

- The structure of the article is rather unorganised and the description of data, methods and results are often unclear. The description of the way in which lightening and radar data are merged is very unclear (and bits and pieces of the methodology are spread throughout several sections of the paper), descriptions of the methods are included in the data and results sections, amongst other things (more details are provided below).

AUTHORS ANSWER:
Also other reviewers have commented on this and therefore the article has now been reorganized in several ways:
- Methodology is now better collected into Sect. 3 (moved text from Sect. 4, merged 3.2 with 3.3 etc).
- Lightning and radar merging is now better explained.
- Result section is now more asserted and concise.
- Plus many other changes (see below comments).
We believe this has improved the readablilty and objectives of the paper.

- The fact that only 7 independent rain gauges are used for evaluation renders the results statistically weak. This issue is so critical that in one of the cases (Table 2), there are simply no data available for evaluation from any of the independent gauges. Why out of 472 available rain gauges would you only select 7 for independent evaluation?
I am aware that you also present the results of performance statistics at nonindependent rain gauge locations; however, I truly believe that more interesting results could be achieved if either more independent rain gauges were used or if a crossvalidation approach were implemented.

AUTHORS ANSWER:
The results in this article were performed during operational LAPS runs, we could not set more stations aside, without risking the quality of the end-users products and applications. This is now mentioned in the introduction section:

"The work reported here has been performed using the operational Local Analysis and Prediction System (LAPS), which is used in the wether service of Finnish Meteorological Institute (FMI). Testing new approaches in an operational system has its limitations in e.g. excluding independent reference stations. Also the possibilities to rerun cases with different settings have been limited. The benefit of the approach is that we can be sure that we only use data which is operationally available."
Unfortunately the summer of 2015 had an unusually low frequency of lightning, which limited the statistics for this study.

- The results, as currently presented, are too vague and far from the objectives initially set in the abstract and introduction. As stated in the title of the paper, you aim at "improving precipitation accumulation analysis using radar, gauge and lightening measurements". Throughout the paper, the focus is mostly on the added value of lightening data, which does not really lead to significant improvements in radar-based QPEs, let alone gauge-based adjusted radar QPEs. One option would be to change the focus (and consequently the title) of the paper to "the added value of lightening measurements" for generation of QPEs. The added value could be more clearly assessed under different scenarios, including with and without radar data available, with / without rain gauge data. From the results you present, the real advantage of lightening data appears to be in cases when radar data are not available (which can occur either because there simply are no weather radars, due to malfunctioning of the radar or to issues such as beam blockage and/or attenuation); this is very valuable and should be made clearer (and, from my point of view, justifies changing the focus of the paper/ changing the way in which objectives and results are described).

AUTHORS ANSWER:
We now stress the importance of using lightning data (especially in case of no radar data availability), in the abstract, result and discsussion sections.
Also, the title is changed: *"Improving the precipitation accumulation analysis using lightning measurements and different integration periods"*

Detailed comments:

Introduction: Work by the co-authors of the paper is often cited and a proper review of relevant literature on the actual topic of the paper is lacking. The introduction should be extended to include:
- A review on the use of lightening data for QPE generation

AUTHORS ANSWER:
We have now included text in the introduction section, where work on this topic is elaborated (including references) as follows:
"Lightning is associated with convective precipitation, but in areas where a large portion of precipitation is stratiform, lightning data alone is not adequate for precipitation estimation. However, lightning has been used to complement and improve other datasets. Morales and Agnastou (2003) combined lightning with satellite-based measurements to distinguish between convective and stratiform precipitation area and achieved a remarkable 31% bias reduction, compared to satellite-only techniques. Lightning has also been assimilated to numerical weather prediction models to improve the initialization process of the model. This can be done by blending them with other remote sensing data to create heating profiles (e.g. estimating the latent heat release when precipitation is condensed). Papadopulos et al. (2005) used lightning data to identify convective areas and then

modified the model humidity profiles, allowing the model to produce convection and release latent heat using its own convective parameterization scheme. They combined lightning with 6-hourly gauge data, within a mesoscale model in the Mediterraiean area, and showed improvement in forecasts up to 12 hours lead time.
Our situation is different from the above mentioned experiments because lightning activity is usually low in Finland, compared to warmer climates (Mäkelä et al., 2011). Also, our analysis area already has a good radar coverage and relatively evenly distributed network of 1 hour gauge measurements. However, if we want to enlarge the analysis area, we will soon go to either sea areas or neighbouring countries where availability of radar data and frequent gauge measurements is low. Our principal goal is to have as good analysis as possible, which is different from having a best analysis to start a model."

- A more in depth and critical discussion of adjustment of radar QPEs. The only comment so far is that "the use of monthly adjustment factors leads to less than optimal results". Several studies have been carried out which have shown a clear advantage of adjusting radar QPEs based on rain gauge data and other sources of information, with corrections implemented at significantly shorter time scales (as compared to the monthly one that you mention). Also, since the focus of your study is mostly on convective storms, it may be worth discussing the performance of merging techniques for convective storms, which is still a problematic issue (see Jewell & Gaussiat, 2015; Wang et al., 2015) and which could be a case in which lightening information could be useful.

AUTHORS ANSWER:
Together with previous answer (see above) we have also added the following text into the introduction part:
"The research of combining radar and surface observations, to perform corrections to precipitation accumulation, is well explored. Many have made developments in this field and much literature is available, for example Sideris et al. (2014), Schiemann et al. (2011) and Goudenhoofdt and Delobbe (2009). Recently, Jewell and Gaussiat (2015) compared performances of different merging schemas, and noted a large difference between convective and stratiform situations. In their study, the non-parametric kriging with external drift (KEDn) outperformed other methods in accumulation period of 60 minutes. Wang et al (2015) developed a sophisticated method for urban hydrology, which preserves the non-normal charactersitics of the precipitation field. They also noticed that common methods have a tendency to smooth out the important but spatially limited extremes of precipitation."

- A review and discussion on the topic of the impact of temporal aggregation on precipitation products. The impact of the temporal scale at which adjustments are performed is a key part of this study. This issue has been discussed in a number of papers which should be reviewed and in the light of which the results of the present study should be analysed (e.g. you should discuss how the optimal temporal resolution that you found (1h) compares to that found in previous studies for the case of convective storms). See for example Berndt et al. 2014.

AUTHORS ANSWER:
This is addressed by including following section in the introduction:
"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 approximately 5 kilometres wide at 250 kilometres), while the measurement area of a gauge

is 400 cm^2 (weighing gauges) or 100 cm^3 (optical instruments). Part of the disparateness of radar and gauge measurements is due to variability of the raindrop size distribution within area of a single radar pixel. Jaffrain and Berne (2012) have observed variability up to 15% of the rainrate in a 1x1 km pixel, with timesteps of 1 minute. Gires et al (2014) have shown that the scale difference has an effect in verification measures (such as normalized bias, e.g. RMSE) but it decreases with growing accumulation time (e.g. from 5 to 60 minutes). In our study, the 60 minutes accumulation period is smoothing some of the differences."

- I would suggest to remove irrelevant information, such as L26-27 (projected annual precipitation in Northern EU) and keep the introduction focused on the topic of the paper.

AUTHORS ANSWER:
This sentence has now been removed.

Data Section:
- L47-50 include description of methods and should be removed from the data section. I would suggest that you simply say that three data sources are employed in this study and then go on to explain them (without starting to describe the LAPS, which should be done in the methods section).

AUTHORS ANSWER:
Text have been modified (text about rain gauges moved into Sect. 2.1) and now follows your suggestion (Sect. 2): "*Here we describe the three data sources employed in this study: rain gauge observations, radar and lightning observations.*"

- Section 2.1 – I would change title to "Ground rain gauges" or would at least include the word "rain gauge" in it.

AUTHORS ANSWER:
Title has been changed to *"Rain gauge observations"*

- Section 2.1, L57-58 is a repetition of L52; try merging these sentences.

AUTHORS ANSWER:
The sentences have been merged and the whole section reorganized.

- Section 2.3: A bit more details about lightening sensors would be desirable. It would help the reader understand how is it that lightening measurements can be translated into vertical 'radar' rainfall profiles and so on. I am aware that this information can be found in other papers, but I think it would be helpful and interesting for the reader to find a brief description of the sensors here. Just in the same way that you provide a brief overview of radar QPE generation.

AUTHORS ANSWER:
We have now added more details about the lightning sensors:
"Lightning location sensors detect the electromagnetic (EM) signals emitted by lightning return

strokes, measure the signal azimuth and exact time (GPS). Sensors send these information to the central processing computer in real time which combines them, optimises the most probable strike point and outputs this information to the end user. More detailed information of LLS principles are described in Cummins et al. (1998)."
Also, the section related to translating radar-lightning into profiles has been rewritten and made more clear.

- While all but one of the Finnish radars are dual-pol, it appears that dual-pol parameters are not being used for QPE generation. The authors mention that a single Z-R relationship is used, which implies simplification and assumptions about variable dropsize distribution and the like. Such simplification could clearly be avoided were dual-pol parameters used. It should be made clear (in Section 2.2) that dual-pol parameters are not being used at all and the implications of this should be discussed (I reckon that the use of dual-pol parameters would lead to much larger improvements in the quality of radar QPEs than the use lightening information).

AUTHORS ANSWER:
Thank you, we agree.
Due to cold climate and relatively long distances between radars, a significant part of our radar measurements are made above melting layer, so the benefits of pure KDP-based QPE are not as obvious as in some other areas. We have tested the use of PhiDP-based attenuation correction methods, but some uncertainties remains. We have now made this clearer by adding following sentence into Sect. 2.2:
"At the moment, the quantitative precipitation estimation based on dual-polarization is not used operationally in FMI, but the polarimetric properties contribute to the improved clutter cancellation (i.e. removal of non-meteorological echoes, especially sea clutter, birds and insects)."

Methods Section: this section is rather unorganised and a thorough re-structuring would be desirable.

AUTHORS ANSWER:
We have made a thorough reorganization of the methods sections, following the suggestions by reviewer (see comments below). Note: Also the other reviewers comments have been implemented and the final structure has now become better.

Some specific comments/suggestions are the following:
- Section 3.1, L97: ": : : where a dense observational input, from several sources, are fitted to a coarser background model first-guess field". Please indicate which data sources are used.

AUTHORS ANSWER:
This section has now been enlarged and we inlcude a sentence and reference related to this:
"The FMI-LAPS is able to process several types of in-situ and remotely sensed observations (Koskinen et al., 2011), among which radar reflectivity, weighting gauges and road weather observations are used for calculating the precipitation accumulation."
We prefer not to mention all observational input, because most of them are not relevant for the accumulation process (in previous article, i.e. Gregow et al., 2013, this was pointed out by reviewers and we had to remove this information). But for your information, FMI-LAPS is using input from satellite, radar, lightning, synop, metar, local surface observational network, soundings, air-plane reports, lidars and background model.

- Section 3.1, L102: you indicate the spatial resolution of the FMI-LAPS output, but not the temporal resolution. From subsequent sections I gather that the temporal resolution is 5 min, but it should be clearly indicated in Section 3.1.

AUTHORS ANSWER:
FMI-LAPS temporal resolution is 1 hour and this has been added (together with more information):
"The FMI-LAPS use a pressure coordinate system including 44 vertical levels distributed with a higher resolution (e.g. 10 hPa) at lower altitudes and decreasing with height. The horizontal resolution is 3 kilometres and the temporal resolution is 1 hour."
Note: FMI-LAPS produce hourly analysis, i.e. the accumulation product is for 1 hour and calculated using 5 minutes segments of radar and lightning data. Correction with rain gauges are done using the hourly data (since rain gauge data is availabe as 1 hour accumulation from our FMI real-time database).

- Sections 3.2 and 3.3: these two sections present overlapping information and a clear and integrated description of the integration of lightening data with radar data is missing. I suggest merging these two sections and producing a new and clearer description of the merging method.

AUTHORS ANSWER:
The sections 3.2 and 3.3 have now been merged, according to suggestion, and the new section has become more clear and readable.

Results section:
- Why using coefficient of determination AND Pearson's correlation coefficient? Their only difference is that the correlation coefficient has a sign, so it may be useful to include both in cases where you expect the correlation coefficient to take negative values. Since this is not the case here, the two performance measures provide very similar information. I would suggest to use only one of the two.

AUTHORS ANSWER:
Yes, we agree. We have now removed the coefficient of determination (R^2) from the verification.

- A description of the log STDEV is missing and a justification for using a log-STDEV instead of the non-log STDEV should be included.

AUTHORS ANSWER:
An explenation to STDEV has now been included:
"STDEV quantifies the amount of variation (i.e. spread) of a dataset. A low STDEV indicates that the data points tend to be close to the mean value of the dataset. Here we use the logarithm of the quotients, in order to get the datasets closer to be normally distributed."

- As mentioned above, this section (results) is unorganised and includes a great deal of description of methods, which should be transferred to Section 3. Also, results should be presented in a more concise

and assertive manner (comments such as “neutral to positive impact” should be removed). As suggested above, a shift in the focus of the paper could make it easier to describe the results in a more assertive / critical manner.

AUTHORS ANSWER:
The results are now presented in a more clear and assertive way. Text related to methods have been moved accordingly, part of text is moved to discsussion and unnecessary subjective comments have been removed.

- A scale bar should be included in Figure 1.
AUTHORS ANSWER:
We have now included a scale bar in Fig. 1. (Scale bar in Fig. 2 was also corrected)

- Why using log scales in figures 4, 5, 7 and 8? I think a normal (linear) scale would be better. A linear scale is normally used in papers on this topic, so readers are used to it. I do not see any added value in using a log scale and I do think that it hinders interpretation of results.

AUTHORS ANSWER:
The intention is to increase the readability of high precipitation values but without disturbring the overal readability. Plotting the values on linear axes will decrease the overal readability, which has been seen after testing different plotting techniques. The log-log scales was the best way to produce the plots (according to us). Therefore we prefer to keep Fig. 5 and 7 with log-scales. As an example we here plot log-scale vs linear-scale (please see Fig. 1 attached below).

It is the same with Fig. 4 and 8, the visualization of the results is more clear and better with log-scales. Here we show Fig. 4a with log-scale vs linear-scale (please see Fig. 2 attached below).

- I suggest using mm to indicate rainfall accumulations (accompanied by the temporal aggregation scale), instead of using mm/h (which is the unit normally used for intensities).

AUTHORS ANSWER:
We have now changed the units to “mm” and explain that this is hourly accumulation values in the figure captions (Fig. 4, 5, 7 and 8).

Other comments:
- Why not work with sub-hourly temporal accumulations? The lifetime of convective cells is often < 1 h. Since the focus of the study is on convective storms and lightening and radar data are available at high temporal resolution (5 min), it would make sense to evaluate sub-hourly scales.

AUTHORS ANSWER:
The limiting factor for this is the gauge information, which is available as 1 hour accumulation from FMI real-time database. Therefore, the time resolution for analysis is bound to be as hourly accumulation data.

- Misuse of semicolons throughout the paper. The semicolon should be either removed or replaced by a colon. E.g.:
L16: “such as;” (remove semicolon)
L133: “resulting from;” (either remove or change to colon)
L218: “analysis time;” (change to colon) Many others! Please check.

AUTHORS ANSWER:
We have now made corrections to the above comments (plus several other places in the text).

REFERENCES (by reviewer):

Berndt, C., Rabiei, E. & Haberlandt, U. (2014). Geostatistical merging of rain gauge and radar data for high temporal resolutions and various station density scenarios. Journal of Hydrology, 508, 88-101.

Jewell, S. A. & Gaussiat, N. (2015). An assessment of kriging-based rain-gauge-radar merging techniques. Quarterly Journal of the Royal Meteorological Society.

Wang, L.-P., Ochoa-Rodríguez, S., Onof, C. & Willems, P. (2015). Singularity-sensitive gauge-based radar rainfall adjustment methods for urban hydrological applications. Hydrology and Earth System Sciences, 19 (9), 4001-4021.

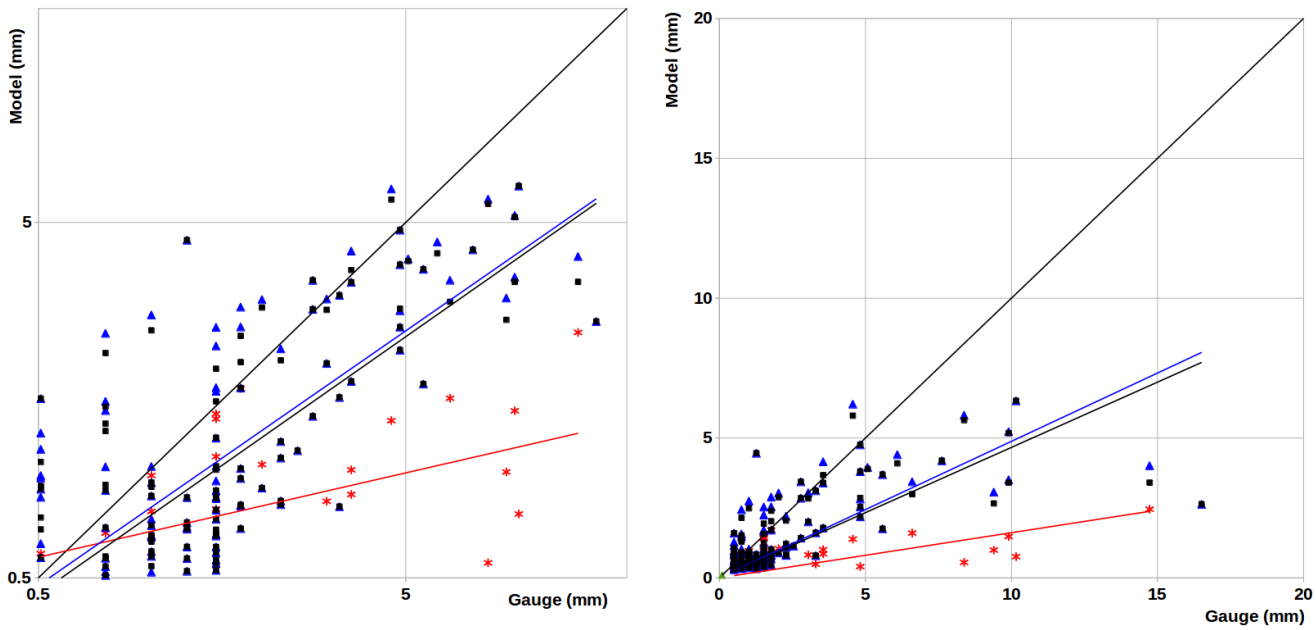
Model (mm)
5
0.5
0.5
5
Gauge (mm)
Model (mm)
20
15
10
5
0
0
5
10
15
20
Gauge (mm)

Figure 1.

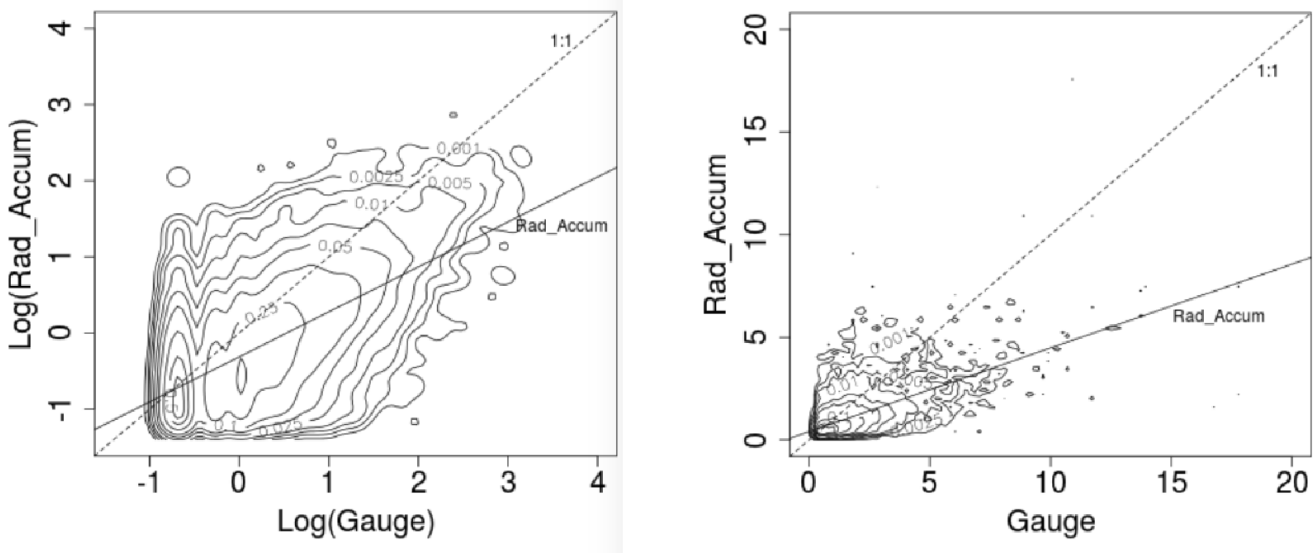
Log(Rad_Accum)
Log(Gauge)
1:1
Rad_Accum
0.001
0.0025
0.005
0.01
0.05
0.25
0.1
0.025
Rad_Accum
Gauge
1:1
Rad_Accum
0.001
0.01
0.005
0.1
0.002

Figure 2.

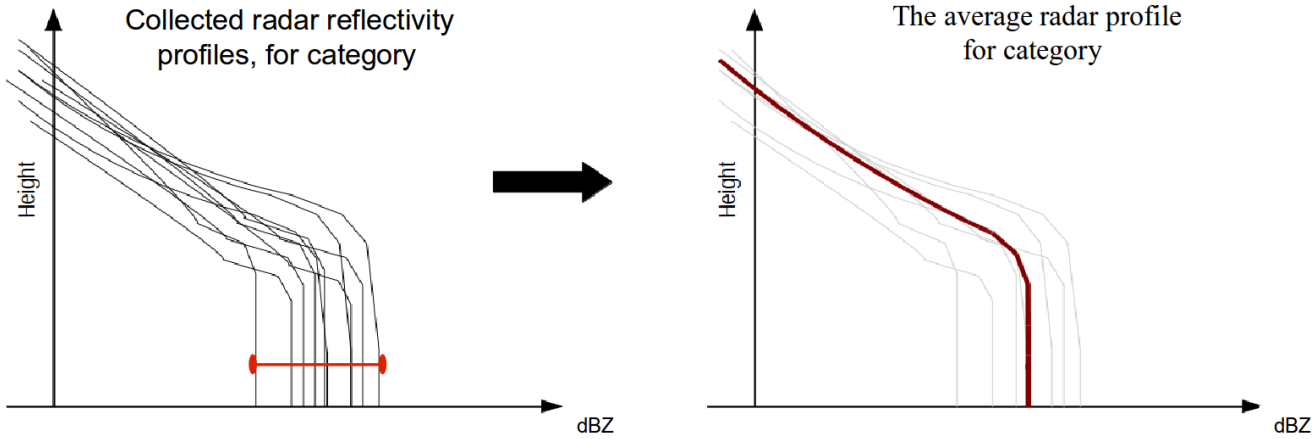
Collected radar reflectivity profiles, for category
Height
dBZ
The average radar profile for category
Height
dBZ

Figure 3.