

Anonymous referee#1: The manuscript is well written even if some clarifications are required. I think some additional parameters should be inserted in the figures. Please find below some specific comments.

-Line 19. One should not forget that the 3D structure of the flashes might be different in winter compared to the one in summer. **This is correctly pointed out by the referee and will be mentioned in the text as an additional potential reason: “An obvious decrease is observed in the percentage of outliers during May-Sept, compared to the other months of the year. This feature could be related to the fact that more sensor upgrades occur during winter or because precipitation of winter thunderstorms is more difficult to detect with the weather radars. In addition, the 3D structure of lightning flashes in winter compared to summer is somewhat different (López et al., 2017), which could increase the difficulty to locate those in winter accurately. [Full reference: López J.A., Pineda N., Montanya J., van der Velde O., Fabro F., Romero D.: Spatio-temporal dimension of lightning flashes based on three-dimensional Lightning Mapping Array., Atmospheric Research, 197, 255-264, 2017].**

-Line 38. Is there a missing word after “more”? **Correct, the word “important” was missing.**

-Lines 64-67. Why only these two regions? And not a larger domain covered by both EUCLID and the European radars? **The following will be added to the summary: “The latter two regions were chosen specifically for their difference in topography and because high spatial and temporal resolution radar data were readily available. However, a similar approach can be performed in the future on a larger spatial scale, based for instance on the radar composite imagery produced by the Eumetnet Operational Programme for the Exchange of Weather Radar Information (OPERA, Huuskonen et al. 2014) and related EUCLID domain” [Full reference: Huuskonen, A., E. Saltikoff, and I. Holleman, 2014: The Operational Weather Radar Network in Europe, Bull. Amer. Meteor. Soc., 95, 897–907, <https://doi.org/10.1175/BAMS-D-12-00216.1>]**

-Line 98. Please provide some physical and/or technological explanations on your statement that “timing only sensors often increase the number of outliers”. **The following will be added to the text: “Timing only sensors often increase the number of outliers if those are used in solutions determined by two or three sensors only.”**

-Line 108. What do you mean by “overshooting beams”? Please rephrase. **The text will include following explanation: “Moreover, since the height of the radar beam above ground increases with increasing distance from the radar, precipitation can be underestimated or even undetected at far range by overshooting when precipitation is produced below the height of the lowest radar beam.”**

-Lines 108-113. I suspect the precipitation product you have been using has been validated. It might be relevant to provide some references on such validation in your paper. **We agree completely with the referee. The radars used in this study are well calibrated. This can be done in different ways using the internal test signal, intercomparison of multiple radar observations in overlapping areas, ... Most of those tests are part of the operational work done at RMI and Austro Control and are not published. Anyway, the validation for Belgium of the product used in the manuscript is presented in Goudenhoofd and Delobbe (J. Hydromet., 2016) where it is referred to as the QPE1 product. The product used for Belgium is not a quantitative precipitation product. It is actually a reflectivity product, with a simple Z-R relation to produce rainfall rates. The product is used in this study to detect the presence of precipitation and not to estimate the intensity. For more information on the validation of the Austrian composite, the interested reader will be referred to Kaltenboeck and Steinheimer (2015).**

-Line 109. Is the Marshall-Palmer relationship valid whatever the precipitation regime? I suspect in your case you are more interested in low precipitation amount where potentially you might find the lightning outliers. So do you think that the radar product used here is sensitive enough to deliver a reliable and accurate product for your investigation? Why did you choose a radar-based precipitation product and not for example the reflectivity composite? Discharges are not only propagating where precipitation occurs (e.g. spider lightning). So I wonder if your choice to use the radar-based precipitation product does not lead to a larger uncertainty. Do you have any comment? **As written above, the radar products used in this work are actually reflectivity products, with a simple Z-R relation to produce rainfall rates. As a lower threshold a reflectivity of 12dBZ is used, which translates to a rainfall rate of 0.2mm/h using the Marshall-Palmer relation. Note that radars easily can measure much smaller values. However, to eliminate non-meteorological echoes the threshold is set at 12dBZ. Precipitation associated with thunderstorms generally produces much larger reflectivity values. We note that ‘bolts from the blue’ (which are rare events) will be classified as outliers by our method, but will be so as well using for instance satellite products. Moreover, satellite information will introduce uncertainty as well since cloud tops can shift by strong high altitude winds and layers of high clouds might overlay structures below.**

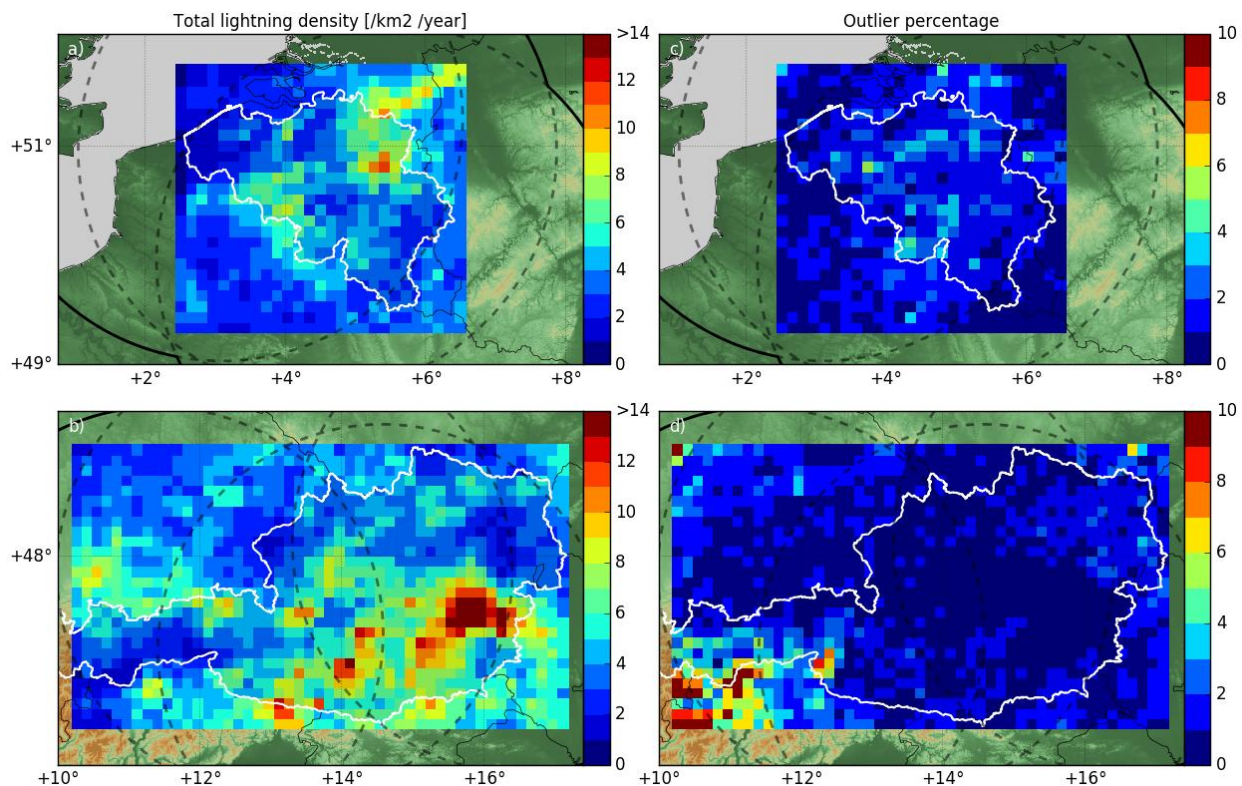
-Lines 116-127. How is the advection taken into account as a 5-min precipitation product is generated? And how do you take into account the advection in your lightning data? At which altitude does the radar-based precipitation product correspond? **No advection correction is performed in Belgium and Austria. However, this is not an issue since different (2, 5, and 10 km) spatial tolerances are used for the distance between precipitation and lightning location. The height of the radar-based precipitation product corresponds to 1500 m above sea level for Belgium. The product used in Austria**

is a vertically maximum surface projection, thus taking any signal into account at any height. This guarantees convective cell detection in complex terrain such as the Alps.

-Lines 130-143. Same questions as for lines 116-127. How is the advection taken into account? At which altitude does the radar-based precipitation product correspond? Is the precipitation product comparable in terms of accuracy for both domains of interest? **See previous comment + the spatial resolution of the radar product is 500m and 1000m in Belgium and Austria, respectively.**

-Line 146. How is the radar-based precipitation distributed in those two domains? Are they geographically uniformly distributed? The precipitation is not homogeneously distributed over those both areas. **In both regions, precipitation is not homogeneously distributed. More info can be found in Goudenhoofdt and Delobbe (2016) for Belgium and Kaltenboeck and Steinheimer (2015).**

-In Figure 4, you are giving the spatial distribution of the % of outliers. I would have added with iso-contours (in white) the actual lightning distribution from where you computed the %. **Very good idea. However, plotting iso-contours on top of the spatial distribution of percentage of outliers turns out to make the figure hard to read. We therefore opt to change the figure as follows, with on the left panels to total lightning density [km^2/year] and on the right the spatial distribution of the percentage of outliers. For Belgium, densities vary between 0.8 and 11 strokes $\text{km}^{-2}\text{yr}^{-1}$ with a median value of 3.4 strokes $\text{km}^{-2}\text{yr}^{-1}$ at 10km x 10km resolution. Overall the densities in Austria are somewhat higher compared to Belgium resulting in a median value of 4.4 strokes $\text{km}^{-2}\text{yr}^{-1}$. The highest total lightning densities are found towards the southeast of Austria with a maximum of 22 strokes $\text{km}^{-2}\text{yr}^{-1}$.**



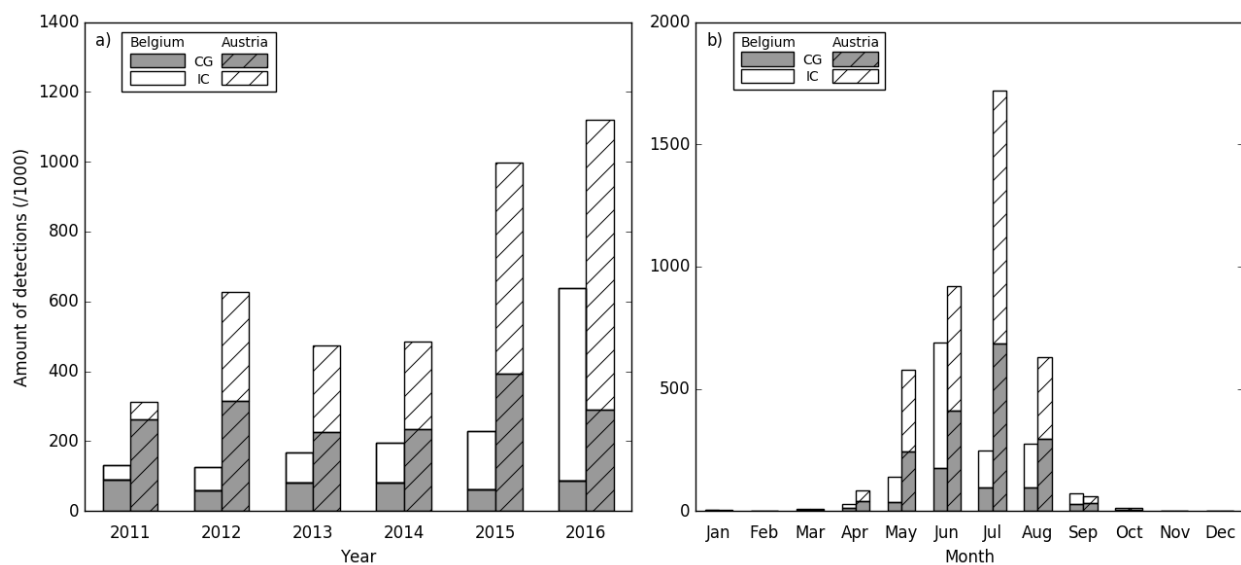
-Lines 148-149. Again how did you take into account the advection of the precipitation and of the lightning activity? The radar product is provided per 5-min period. How did you select the lightning data within that 5 min period and how does it fit with the way the radar product is built? **We have taken the 5min lightning data corresponding to the start and end time of the radar scan. The text will be slightly adapted to make this more clear: "Subsequently, CG strokes and IC pulses with timestamps within the start and end time of the radar scan are superimposed on the corresponding 5-min radar precipitation fields."**

-Lines 151-152. Again this suggests that precipitation at the ground is required to verify the lightning data. How are we sure that precipitation is always required where lightning flashes occur? Please provide some arguments to strengthen your methodology. I think a similar analysis using the reflectivity composite should be performed in order to identify cloud-free outliers to in-cloud non-precipitation outliers. At which stage of the lightning flash do the lightning outliers correspond? **We do make use of "complete" composites, i.e. if one of the radars did not participate in a 5-min composite then this**

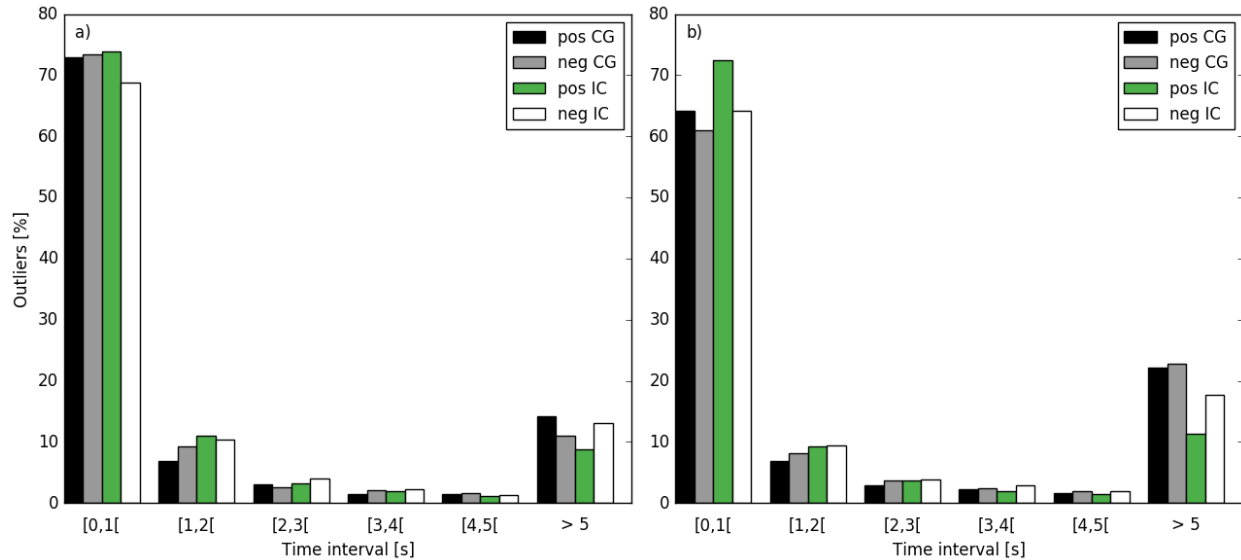
composite is not used at all. Our methodology makes use of radar data, and thus precipitation/reflectivity is required (but not at ground!, see answer related to your remark for Lines 116-127) to discriminate between the outliers and well-located lightning events. However, we admit that lightning produced by “dry” thunderstorms or bolts from the blue will be misclassified as outliers. Nevertheless, we believe that this particular phenomenon is extremely limited. We think that a methodology based on satellite cloudiness products would not allow a proper identification of outliers. Note that the majority of the lightning outliers are single-stroke flashes. The drawback of the used method will now be mentioned in the summary.

-Line 153. I suspect you have projected the lightning data on the same temporal and spatial grid as the radar data, right? **This is Correct.**

-Figure 3. Could you please plot as well the number of CG strokes and IC pulses in order to see how your dataset spreads over the years. **The following figure will be included in the manuscript. It plots the total (CG + IC) amount of detections as a function of a) year and b) month in Belgium and Austria. With regard to the IC detections, one notices a sharp increase in 2016 in Belgium and from 2015 in Austria. This increase is not climatological in nature, but is attributed to the increased amount of LS700x sensors in EUCLID and its capability to detect IC pulses in the low-frequency domain. Looking at the distribution of the total monthly stroke count, it is found that peak in activity is observed in June and July for Belgium and Austria, respectively. For both regions about 95% of all the observed lightning activity occurs between May and September.**



-Lines 161-163. Am I correct when I say that you did not try to associate the outliers with “correct” events based on a time criteria for example? I wonder if a time criteria should not be added to your analysis in order to dissociate isolated events to outliers mis-located from a group of events. **Point well taken. The following figure, which will be added to the paper, plots the distribution of the time difference between the outliers and its closest ‘ok’ event (in time) for a) Belgium and b) Austria. We would argue that all the outliers lying within 1 second of an ‘ok’ event are simply badly located lightning events (~65%-70%), whereas those larger than 1 second (~30-35%) are outliers in time and space or so-called ghost outliers. We find that everything is quite independent of polarity and classification for Belgium and Austria.**



-Line 185. See comment on Line 146 for Figure 4. **Figure 4 has been changed: see above**

-Lines 194-196 Could not we say as well that the October-to-April flashes might have different vertical and horizontal structures that make them detected by EUCLID sensors with more difficulties? One should not only criticize the radar sensibility or the sensor upgrade. **True, we will add this as an additional potential reason (see above).**

-Lines 203- 204. I understood that the two geographical domains you have been studying should not really suffer of any long range issue. Am I right? **Correct, the domains are chosen specifically to reduce those kinds of effects to a minimum, if any.**

-Line 208. "which" instead of "wich". **OK, typo will be adapted in the text.**

-Lines 212-214. I agree with you on the climatology point of view, but then it depends on the application you want to do with your outliers data. **Thanks for your comment. However, we feel this is exactly what is written in L212-213.**

-Lines 217-219. This confirms the interest of considering a temporal criterion in your analysis in order to discriminate isolated (in time and space) outliers to isolated in space only outliers. What do you think? **Thanks for this idea (see fig. above)**

-Figure 6. Without the actual number of events considered for each year, it is difficult to identify how statistically representative is the dataset you are studied in Figure 6. Could you please add that information? **We have added an additional figure showing the number of events per month and per year for both domains (see fig. above)**

-Line 228. What the % of outliers with an absolute current above 20 kA? **"The percentage of CG outliers with absolute peak currents above 20 kA is a factor of about three lower compared to the total percentage of outliers found within [0-20] kA, and is a factor of 15 lower in case of IC. The larger drop in case of IC results from the lower amount of IC events with absolute peak currents larger than 20 kA compared to CG events."**

-Line 238. I would add "positive current outliers" instead of "positive outliers". **Thanks for the comment. We will adapt to the terminology of the negative event and write "positive IC and CG outliers".**

-Lines 228-238. I have a question about all those low current events. How confident are you in their detection and on their classification not only in terms of IC or CG but also in terms of polarity? **We do not present any data on polarity errors because by comparing LLS data with independent E-field measurement data we have never observed such errors since we started those measurements (see Schulz et al. 2016: The European lightning location system EUCLID: Performance analysis and validation). With respect to the IC/CG classification, very recently Kohlmann et al. looked into the classification accuracy (CA) of EUCLID and found that CA varies between 89% and 97% for years 2012 and 2015 in Austria. [reference: H. Kohlmann, W. Schulz, and S. Pedebay: Evaluation of EUCLID IC/CG classification performance based on ground-truth data; to be presented during the International Symposium on Lightning Protection SIPDA, October 2017]**

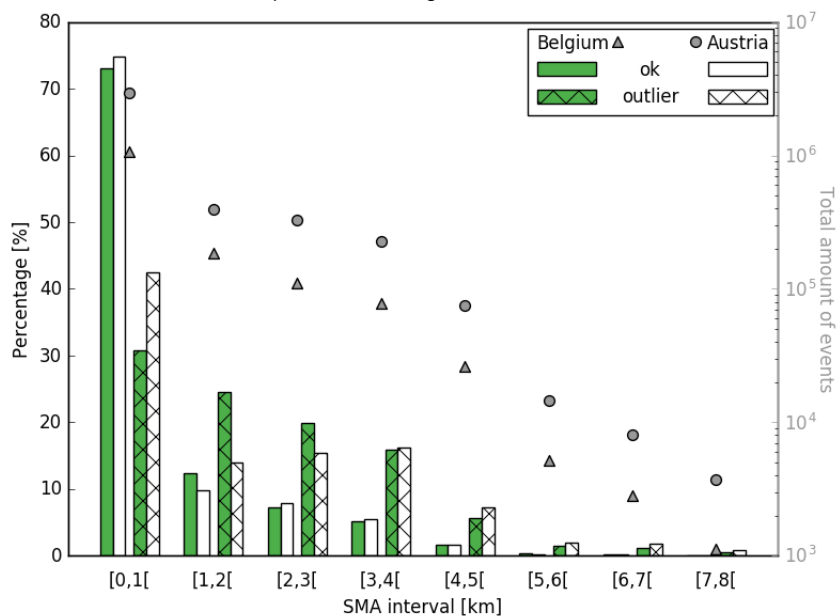
-Lines 238-253. How accurate are the event locations when two lightning sensors are only used? Do you usually keep flashes detected with only two EUCLID sensors? Have you plotted the same parameter but by range of current? I would like

to see how the number of lightning sensors influences the detection of the outliers according to their estimated current. You could plot it with 2D cumulative distribution. **It is clear that the more sensors participate in a solution, the more accurately detected the lightning event will be. The percentage of events detected by only two sensors is ~10% and ~30% in case of CG and IC, respectively, as can be deduced from Fig. 8. Thus, the majority of the events are detected by more than two sensors. Obviously, the higher the peak current of a lightning event, the higher the number of sensors reporting the event. It is therefore clear that low number of sensors reporting is for low peak currents.**

-Line 257. Do you see any difference between IC and CG outliers separately? **There is a slight difference, which was mentioned in the same paragraph as follows: "Although not shown in this plot, it is found that the average SMA for CG strokes is smaller by a factor of two compared to IC pulses. This is expected since more sensors participate in a solution for CG strokes compared to IC pulses as discussed in Fig. 9."**

-Lines 256-259. Do you have a way to get the number of events that were rejected by the central processor for the period of data you have studied? **Thanks for your comment. However, we don't see exactly what it will bring to our study.**

-Figure 9. The number of samples per SMA range would provide an idea on the statically representativeness of the dataset used here. Please add that parameter in Figure 9. **We have added this information in the following plot:**



-Figure 9 (continued). Similarly to what I suggested for Lines 238-253, have you looked on how the SMA is distributed according to the number of lightning sensors used to locate the different event categories? If yes, what are your main conclusions? If not, please take a look on that plot and provide some information in response to the present comment. **The SMA is indirectly related to the number of sensors reporting. Thus, small SMA values are related to a high number of sensors reporting. From the above plot, it means that ~ 30% and 40% of the outliers exhibit a SMA between [0-1[km for Belgium and Austria, respectively, and are therefore detected by a high number of sensors.**

-Lines 265-267. I do not understand your statement. Please explain it. **To avoid any misunderstanding/confusion and since no results are shown related to χ^2 anyway, we chose to simply remove L266-268 from the text.**

-Lines 280-281. I would insert the actual numbers, i.e. outliers and total number of samples. **The total number of "samples" (as a function of year and month) is now included in an extra plot (see above) and will be mentioned in the text related to the figure.**

%%% End of review