Authors' Response to the comments on the manuscript by RC1:

Graupel and increased turbulence observed near small-scale intermittent lightning discharges at the top of intense thunderstorms

R van Loon, J D Assink, O Scholten, B H Hare, H Leijnse, A J van Delden

Preprint Egusphere-2025-6253

#. Reviewer's comment. AR: Authors response. CH: Changes to manuscript with new text and removed text

Note: Line numbers correspond to the document created with latexdiff, not the manuscript from first submission.

We thank the reviewer for their constructive comments. The manuscript has undergone thorough revisions. Many suggestions have been incorporated and have greatly improved the quality of the manuscript. Responses to the comments are listed below.

1. ***Beyond normalization constants and minimum cluster size, key DBSCAN settings should be explicitly reported (preferably in a parameter table).***

AR: We agree that the constant could be noted more clearly. Therefore we have added a table. Note that we only use the "n" constant in the DBSCAN algorithm. We manually normalize the spatial and temporal dimensions of the data.

CH: Added table 1.

2. ***The sensitivity tests must be accompanied and show whether sparkle counts, spatial patterns, and radar-based conclusions remain stable.***

AR: The parameters for the sparkles classification (two-stage clustering approach) are also based on physical insight and logic. For example, to eliminate imaging errors, we only need a n=2 within small distance (e.g. d=200 m, τ=5 ms). We also used our knowledge of lightning structures to judge how well the parameters were for the first stage (finding large lightning structures). As such, we are not sure how well a sensitivity test is useful.

On the other hand, your comment did encourage us to do a sensitivity analysis on the parameters. Since the parameter space is large, we only varied one parameter at a time from the baseline (table 1 in revised manuscript). The results can be found in the Fig. 1 and Fig. 2 below.

As expected, changing the variables does change the number of sparkles, and therefore the radar data near sparkles, quite dramatically. We also observe that the conclusion (near sparkles the Zh and Wrad values are significantly higher than near other VHF sources) remains intact. P-values (not shown) from the KS statistic, remain well below 0.01 for both Zh and Wrad distributions.

We are of the opinion that the figures are not of much added value to the manuscript. Therefore, we will add it as a supplementary material. In the revised text, we do refer to the sensitivity test.

CH: Sect. 3.3, line 284-: "Sensitivity tests … sparkle classification."
Added supplementary material S2.

3. ***Please provide at least a limited validation using manual labeling for a subset of LOFAR images or a weak ground truth based on Scholten et al. (2023), reporting approximate false-positive/false-negative rates or overlap metrics.***

AR: When imaging lightning, there is always a chance that a particular VHF source is mislocated and should be considered noise. It is very hard to be quantitative on this issue as a ground-truth check cannot be made. From imaging other lightning structures, we know that the amount of noise (for the used quality cuts on the data as presented in Scholten et al. (2023) is minor probably order 1 %. Since a sparkle is labelled as such when there are a few sources within a short distance, the chance is really negligible that a single sparkle is wrongly classified as such.

Furthermore, the observation of sparkles in LOFAR data (Scholten et al., 2023) has only been qualitative so far. Therefore, we do not have a ground truth for sparkle observations in the LOFAR data.

CH: No changes.

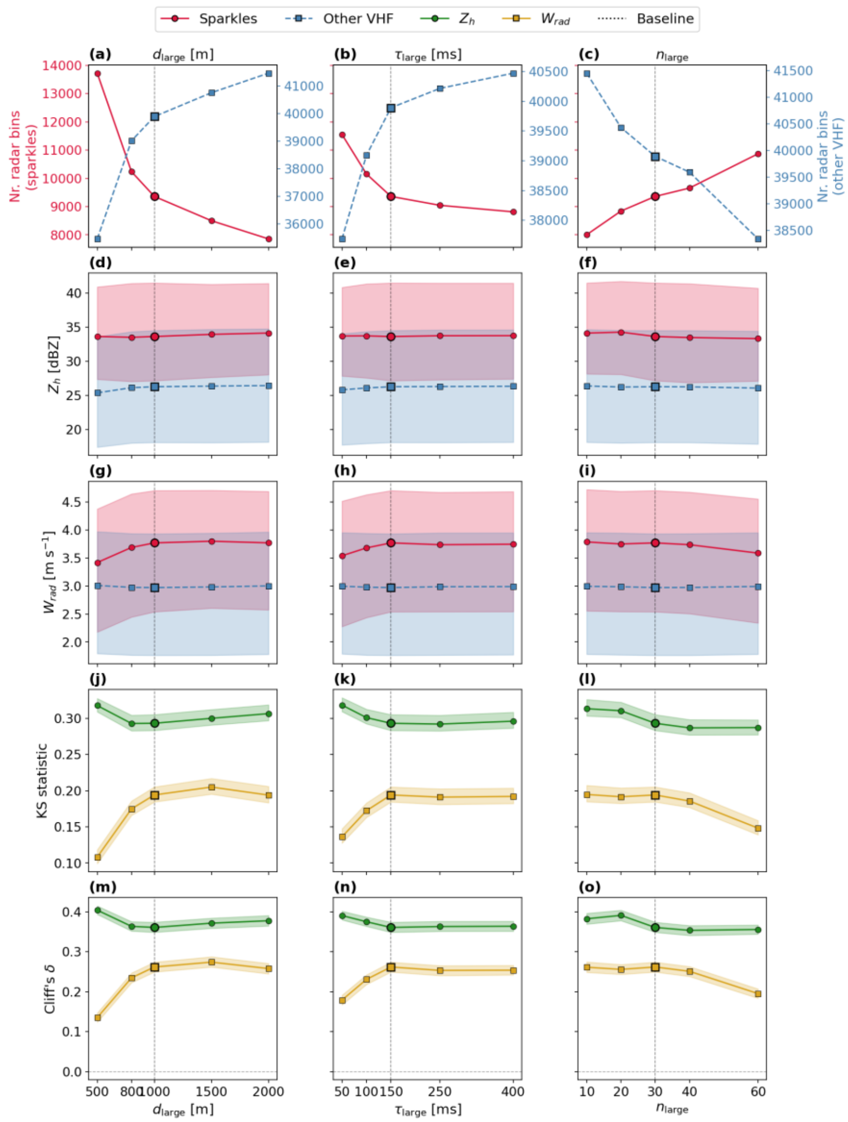

Fig. 1: Sensitivity of the radar data, comparing sparkles to other VHF sources, to the parameters of the large-scale clustering of the sparkle classification algorithm. a-c) Number of radar data point near sparkles and other VHF sources (on different vertical scales. d-f) Mean Zh values and interquartile range (shading). g-i) Mean Wrad values and interquartile range (shading). j-l) KS-statistic and 95% confidence interval (shading) from a bootstrap method. m-o) Cliff's δ and 95% confidence interval (shading). The spatial normalization factor, temporal normalization factor and the minimum number of sources, for the clustering of large lighting structures (d_{large}, τ_{large}, and n_{large} respectively), are varied from the baseline (values in table 1 of revised manuscript) in each of the columns.

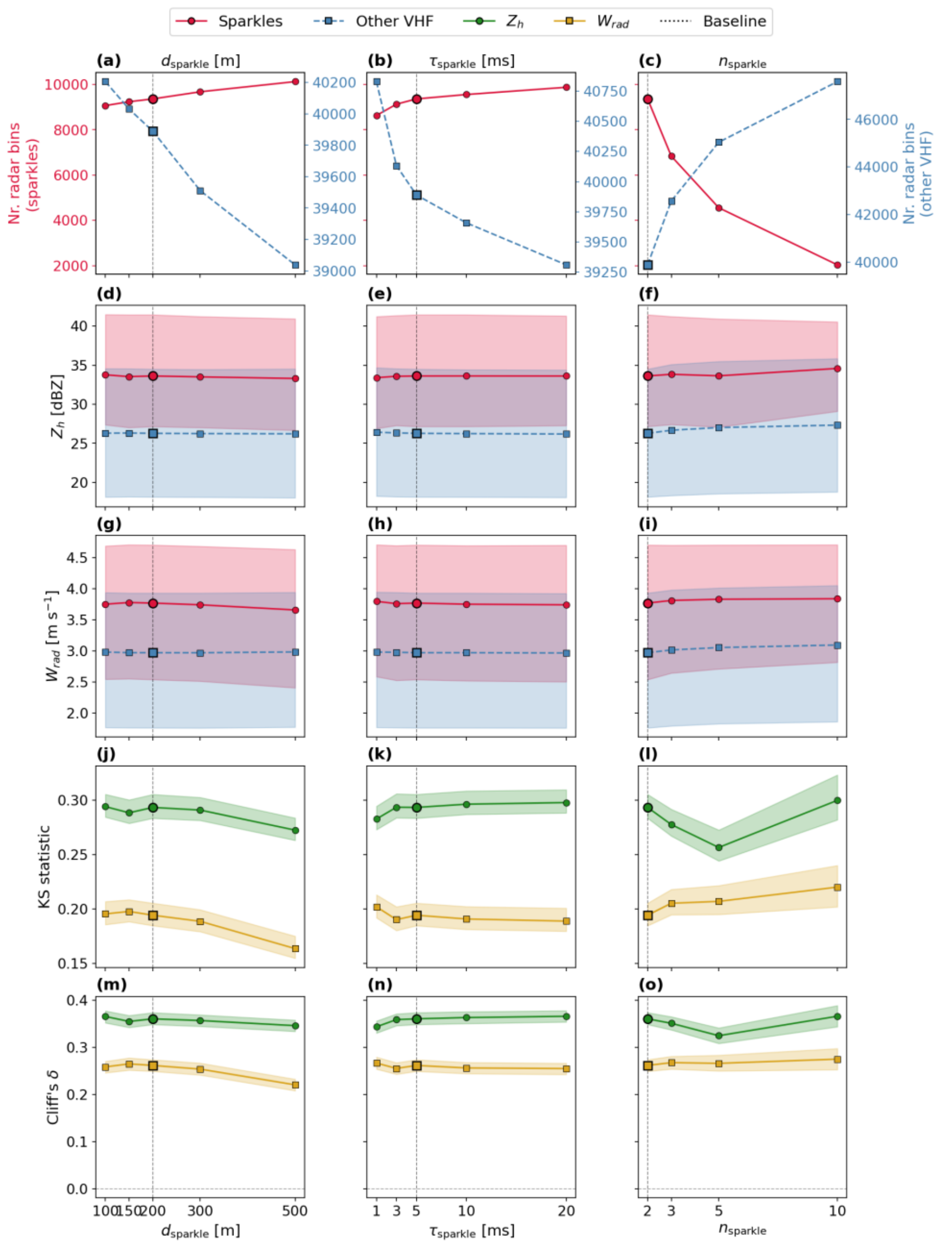

Figure 2: Similar to figure 1, but varying the spatial normalization factor, temporal normalization factor and the minimum number of sources, for the clustering of sparkles (d_{sparkle}, τ_{sparkle}, and n_{sparkle} respectively).

4. ***The method likely excludes sparkle-like activity linked to positive leader breakdown (needles). This scope limitation should be stated more explicitly in the abstract and methods to prevent over-generalization of "sparkles".***

AR: We do mention this in the discussion (line …: “Our two-stage … positive-leader breakdown.”)
However, we could indeed be clearer about the scope of our “sparkles” and we have added a some text to the abstract, introduction, and method section.

CH: Abstract, line …: “distinguish sparkles, defined as small and isolated lighting structures, from other lightning activity.”

Sect. 1, line 48 - 50: “At first glance, … on sparkles.”

Sect 2.2, line 120-121: “By design, … as sparkles.”

5. ***ERA5 may deviate from storm-scale winds, and within-sweep timing differences exist (Appendix B1). Please quantify the uncertainty of advection by comparing alignment metrics before/after correction, and report typical horizontal uncertainties.***

AR: Thank you for bringing this to our attention. We have added an alternative method for advection. Namely, tracking cell motion, based on a 2D, 1500 m altitude, radar composite and the Lukas-Kanade method from the pySTEPS package. The differences between the ERA5 and the pySTEPS method are visualized in Fig. 3 below.

We do see a major difference in the absolute advection distance up to 2 km. On average, ERA5 advects the data about 330 m less than the movement according to cell tracking. What is interesting, when comparing the statistical results with the pySTEPS method, there are no notable differences. Therefore, we decide to stick with the ERA5 advection. We have added a paragraph to the appendix about the advection error, and we briefly mention it in the Sect. 2.4.

We are aware that it would be preferred to derive the horizontal advection, with a 3D radar composite. However, we think that the effort of computing a 3D gridded radar product is not worthed the benefit.

Also note that in our advection scheme, we do actually take into account the time differences withing a single sweep. Each radar data point has a timestamp (radar points along the same azimuth and elevation angle are instantaneous) and is advected according to this timestamp.

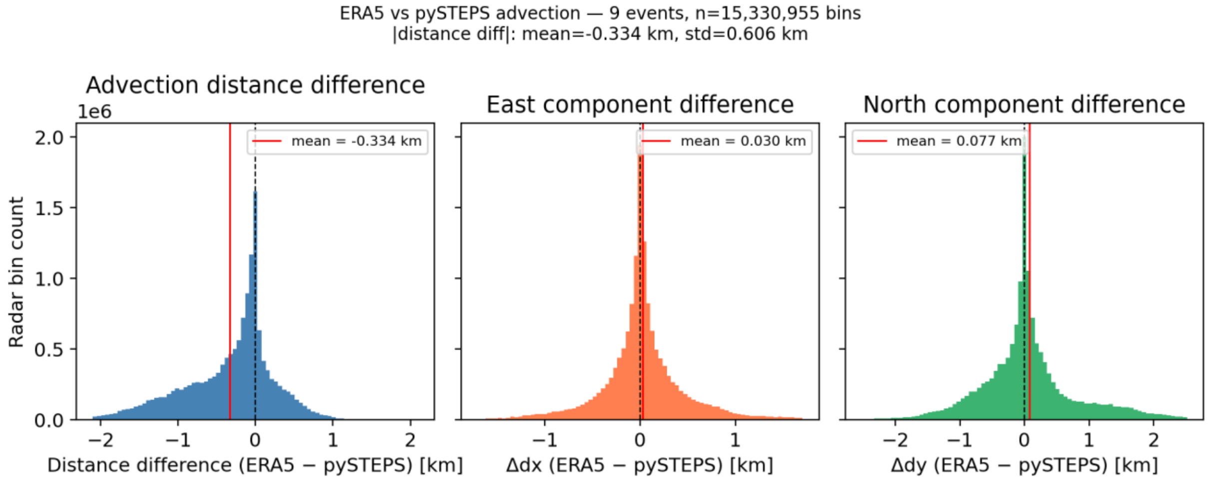

Figure 3: Histograms of differences in advection distance between ERA5 wind-based advection and advection based on pySTEPS' Lukas–Kanade convective cell motion tracking. (a) Absolute difference in the advection distance. (b) Difference in the longitudinal component. (c) Difference in the latitudinal component.

CH: Sect. 2.4. line 171: "We acknowledge ... (see Appendix B)."

Appendix B, line 555-563: "Convective systems ... our analysis."

6. ***Radius(2 km) is large relative to sparkle scales; include sensitivity tests (1/1.5/3 km) to show the Zh/Wrad/HMC contrasts are robust.***

AR: 2 km is indeed large, but this helps to mitigate errors regarding the advection of radar data, classification of sparkles and LOFAR imaging. We have done some extra checks on the sensitivity regarding the radius. This results in Fig. 2 below.

As expected, the number of datapoints is reduced when the r=500 m.

The most important observation is that the main conclusions are relatively invariant for a 0.5<r<4 km. The distributions between sparkles and other VHF sources are always significantly different. The p-value of the Kolmogorov-Smirnov (KS) statistics stay well below 0.01. The Cliff's δ value also does not change notably.

We are of the opinion that the figure is not of much added value to the manuscript. Therefore, we will add it as a supplementary material. We do refer in the text to the sensitivity test.

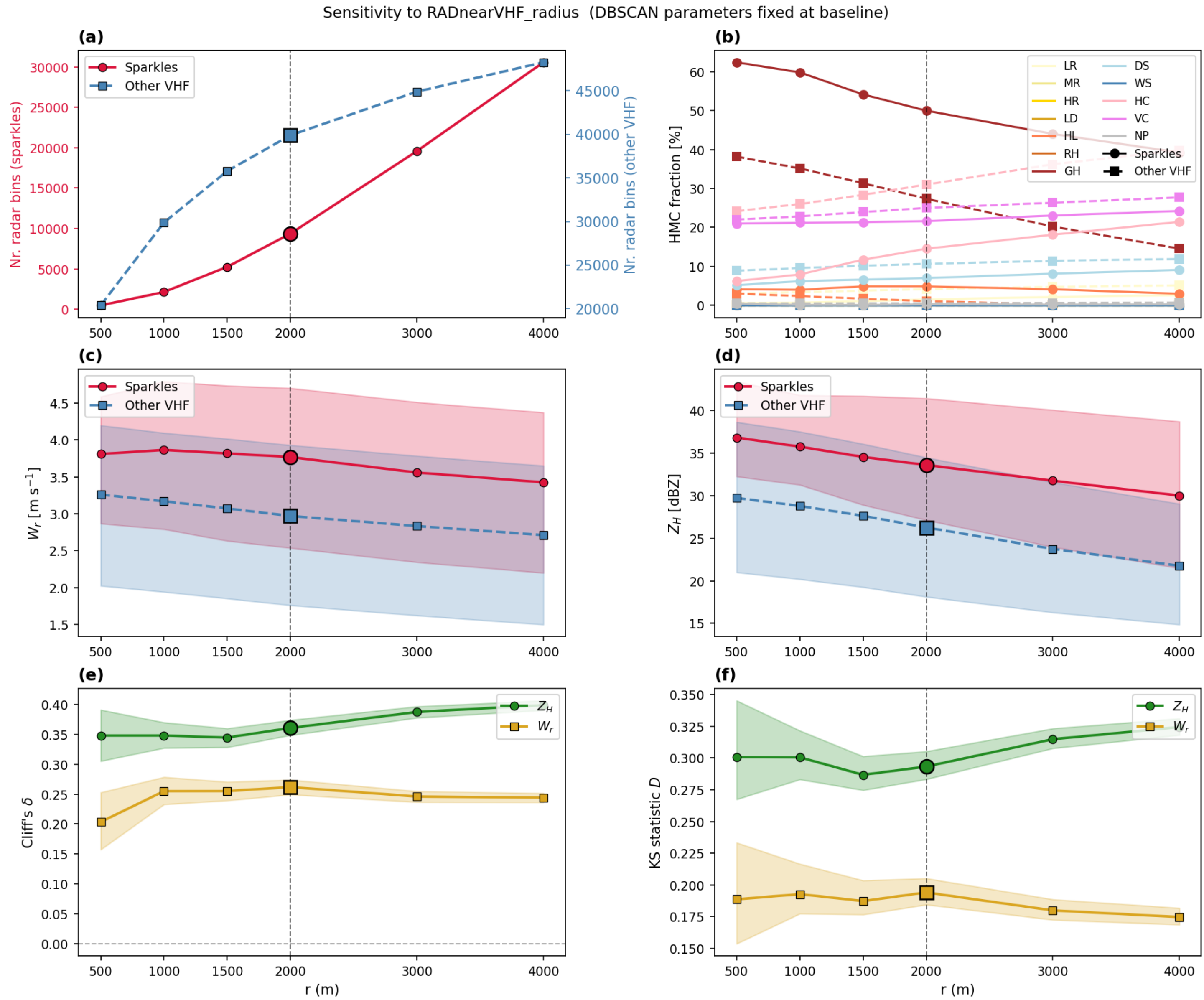

Fig. 4: Sensitivity test of the radius **r** to collect radar data near VHF sources. (a) Number of radar points near sparkles (red circles, solid line) and near other VHF sources (blue squares, dashed line). (b) Fraction of the data in different HMC (hydrometeor classification) categories. (c) Mean Wrad values and with shading indicating the interquartile range. (d) Mean Zh values and interquartile ranges. (e) Cliff's Delta between the distribution of sparkles versus the distribution of other VHF sources. Green for Zh distributions, and yellow for Wrad distributions. Shading shows the 95% confidence intervals of Cliff's Delta (CD) based on a bootstrapping method. (f) Kolmogorov-Smirnov (KS) test statistic between sparkles and other VHF

distributions. Shading indicates the 95% confidence interval from a bootstrapping method.

CH: Sect. 2.4, line 175: "In a … different results."

7. ***Even with an 8-km threshold, the detailed height distributions may differ; provide height histograms and stratified comparisons or explicitly control for height in a statistical model.***

AR: This is a very good point. We have made some additional checks. The results are shown in Fig. 5 below. There is a clear negative correlation between Zh and the altitude. For the Wr, there is negative correlation. For Zh, the mean values are significantly higher for all altitudes. Below 12 km, the Wrad values are also significantly higher, but above 12 km this is not the case.

We are of the opinion that the figure is not of much added value to the manuscript. Therefore, we will add it as a supplementary material. In the revised text, we do refer to this test.

CH: Sect. 3.3, line 286-288: "Comparing the … 12 km altitude."

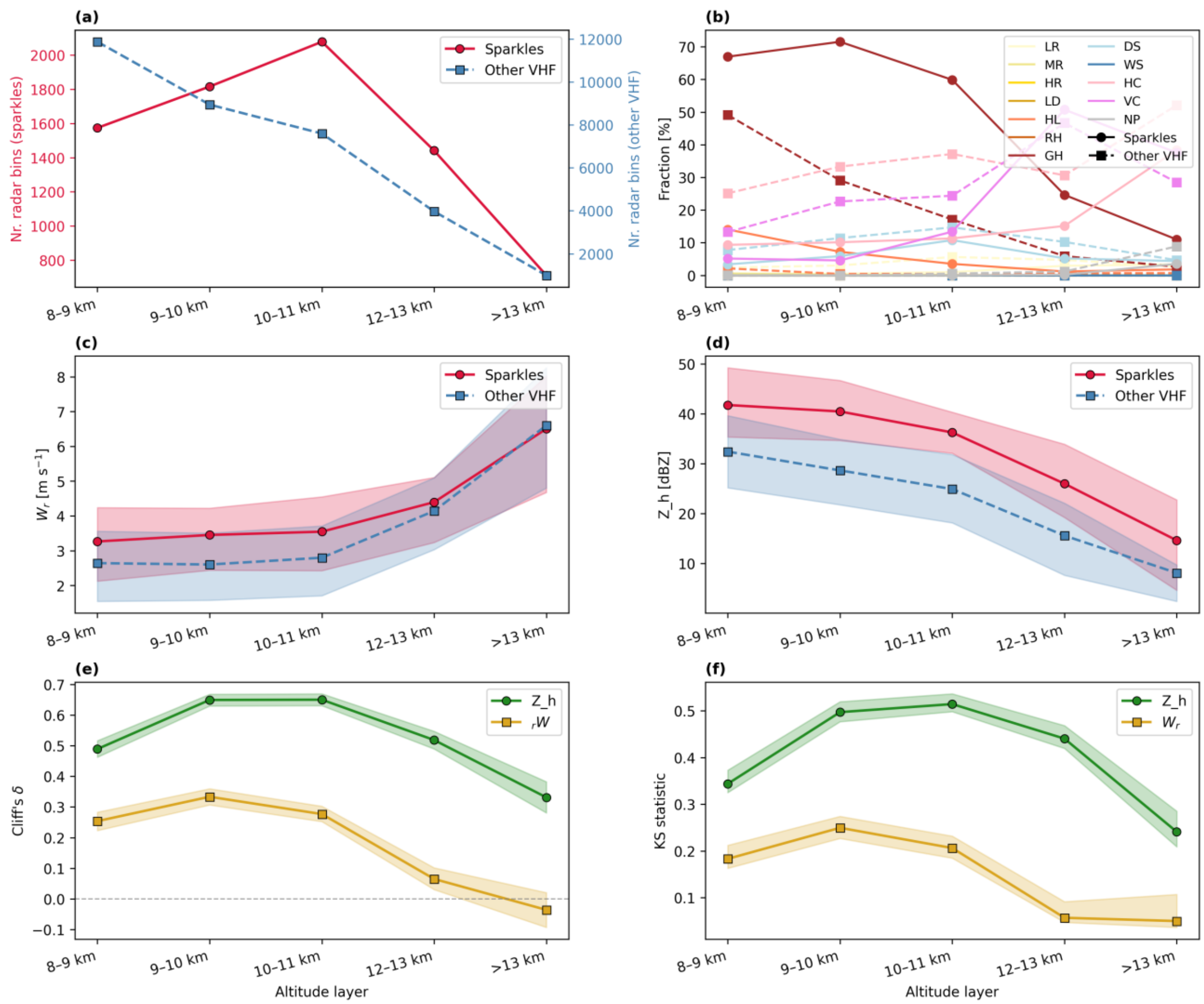

Figure 5: Comparing sparkles data to data near other VHF sources, in different altitude layers. (a) Number of radar points near sparkles (red circles, solid line) and near other VHF sources (blue squares, dashed line). (b) Fraction of the data in different HMC (hydrometeor classification) categories. (c) Mean Wrad values and with shading indicating the interquartile range. (d) Mean Zh values and interquartile ranges. (e) Cliff's Delta between the distribution of sparkles versus the distribution of other VHF sources. Green for Zh distributions, and yellow for Wrad distributions. Shading shows the 95% confidence intervals of Cliff's Delta (CD) based on a bootstrapping method.

8. ***Figures 6–7 are primarily descriptive. Add at least one significance test/effect size (KS, Mann-Whitney U, Cliff's delta, etc.), while noting limitations due to correlation.***

AR: In line with your suggestion, we have computed the KS test statistic and computed the Cliff's delta effect size. This has been added to the manuscript.

CH: Sect. 2.5, line 176-186: "We compare ... the difference."

Sect 3.3, line 281-288: "The distribution ... VHF sources."

Sect. 3.3, line 283-285: "The distributions ... very sources."

9. ***The interpretation of positive Vrad patches as signatures of strong updrafts plus rotation (Figs. 10–11) may be correct, but needs clearer geometry and quantification.***

AR: Indeed we do not quantify the updraft strength or rotation of the updraft. The main goal of this section is a qualitative analysis (hence the name of the section). In this case we want to qualitatively show how updrafts, rotation, and the associated turbulence or shear (indicated by the high values of W_r) are related to the locations of sparkles. To do quantitative analyses requires more extensive analyses that are outside the scope of this paper.

However, considering your comment, we have rephrased the paragraph somewhat.

CH: Sect. 3.4, line 330-332: "We interpret ... hydrometeors. The intensity ..."

Sect. 3.4, line 330-334: "... white ellipse. This ellipse ... of (1) and (2). "

10. ***Add a schematic conceptual model showing the hypothesized geometry of updrafts, graupel/hail charge regions, screening layer, turbulence, and sparkles.***

AR: In this study, we compare our results to previous hypotheses. We also hypothesize on the possibility of turbulence enhance non-inductive charging. We do not lean towards a single hypothesis. We also do not have an extensive analysis of the updraft geometry or spatial distribution of graupel/hail.

We don't think that a schematic will be necessary to bring the main message of our paper across.

CH: None

11. ***The manuscript repeatedly uses "lighting" where "lightning" is intended (including the abstract). Please perform a systematic language/typo pass (also fix minor misspellings such as "systesm", etc.).***

AR: Our apologies for the errors. We have performed additional language checks.

CH: Multiple places: "lightning"

12. ***For Fig. 6, consider reporting medians and IQRs (or 90% ranges) in addition to means. For Fig. 7 difference plots, mask or flag low-count bins as low confidence.***

AR: In Fig. 6, we have added additional symbols to show the median, and IQRs. In Fig. 7, we have added a contour to mark the "significant data" with more than 10 counts per cell.

CH: Fig. 7: Added interquartile range and 90% quantile range.

Fig. 7 caption: "The solid … range respectively."

13. ***Add an objective echo-top/overshooting-top metric (radar or satellite) to directly test the proposed relationship.***

AR: Unfortunately, we have no independent measure for the cloud top or cloud overshooting. The radar data is coarse to reliably estimate the cloud top height. Perhaps this will be possible for future studies.

In our manuscript, we do not directly link overshooting tops to sparkles.

CH: None
