Reply to comments of Reviewer # 2:

We want to thank Reviewer 2 for their comments, which certainly aided to improve our manuscript. We are addressing the raised comments in a point-by-point way below:

1. Abstract, lines 12-14: The authors state that "The new algorithm is found to perform well." "Well" is a very subjective description. A more quantitative descriptor, or at least less subjective language, is preferred. Possible alternative wording that combines two sentences:
The new algorithm consistently identifies Doppler spectra peaks and outperforms other algorithms by reducing noise and increasing temporal and height consistency in detected features.

Reply: We agree that a more quantitative wording is preferable to using "well". Thus, we decided to accept the proposed alternative phrasing and combine the two sentences.

2. Lines 23-25: Suggest changing nominalized language for stronger writing:
The first step towards characterizing hydrometeor types is determining the number of different populations within a certain cloud volume.

Reply: According to the suggestion, we adjusted the sentence using the gerund.

3. Introduction, Lines 21-31: This section of the introduction is very fragmented. The authors inject various Doppler spectra analysis studies in a somewhat non-coherent manner. Maybe it's just a spatial issue (i.e., the authors indent the various studies as stand-along paragraphs comprised of 1-3 sentences). One way to mitigate this issue to allow the science themes, rather than the referenced studies, to drive the content. I envision these lines recast in terms of scientific topic that would allow a more natural flow to the discussion. A suggestion:
Other studies have utilized Doppler spectra analyses to identify cloud microphysical composition and cloud processes operating in Arctic clouds. For instance, four Arctic cloud hydrometeor populations (background ice, cloud, drizzle and new ice) were successfully classified using continuity of spectral modes in time and height combined with high spectral resolution lidar (HSRL) and in-situ observations (Verlinde et al. 2013). BAECC {what does the BAECC acronym represent?!} field campaign analyses have also distinguished up to three noise-floor separated peaks in the recorded Doppler spectra for frontal snow falling through a supercooled water layer (SWL) that produced rimed snowflakes (Kalesse et al 2016). These respective peaks were then used to track microphysical processes along slanted fall streaks, although this documented case was special due to the separation of peaks by the noise-floor (merged peaks are usually observed, motivating the need to develop robust cloud radar Doppler spectrum peak separation techniques). Finally, KAZR observations of liquid-only and mixed-phase clouds at Oliktok Point, Alaska have been used to identify multiple Doppler peaks using the depth of the local minimum between the main peak and sub-peak as the main separation criteria (Williams et al. 2018).
All these efforts, using somewhat differing approaches, show that there is a need. . .
[continue with the rest of the content from the last introductory paragraph].

I also suggest adding a final sentence to the introduction that briefly introduces what the current study will accomplish. For example, "This study describes a new algorithm that adopts machine learning tools to classify Doppler spectra peaks in complex mixed phase cloud scenarios" – or something similar to this statement that properly whets the readers' appetites.

Reply: We agree that this section of the text is rather incoherent and gratefully accept the proposed alternative text. This greatly improves the readability of the manuscript. We added the abbreviation for the Biogenic Aerosols – Effects on Clouds and Climate (BAECC) field campaign, which is mentioned in the Abstract for the first time.

4. Figure 1: Are these truly random spectra chosen from 16 February 2014? Or are they neighboring spectra, where neighboring can be defined as either spatial (height) or temporal?

Reply: The caption of Fig. 1 was misworded. The algorithm picks spectra for the user to marks peaks randomly in a previously defined time-height chunk of Doppler spectra. Fig. 1 (center panel) shows one of these randomly picked spectra. The spectra displayed in Fig. 1 in the panels around the center plot are neighboring spectra (in time and height). We changed the label of Fig. 1 for clarity to "The surrounding spectra display the spatially and temporally neighboring spectra." and changed from "[…] Data: random spectra of KAZR observed at TMP [...]" to "[…] Data: KAZR spectra observed at TMP [..]".

5. Page 5, Line 2: 21 February 21 2014 -> 21 February 2014

Reply: We adjusted the text accordingly.

6. Page 6, Lines 3-4: How did the chosen smoothing method produce the most promising results? Is there any quantitative measure to optimally select the smoothing method (like line fitting parameters)?

Reply: We picked the smoothing method by visual inspection, testing different smoothing methods which are by default available in the Matlab Signal Processing Toolbox. The loess-smoothing yielded the best results at separating peaks in a quick sensitivity study. We re-phrased the text to "This smoothing method was chosen empirically after testing different methods since it showed the best ability to capture peaks while filtering out noise". Please also see the reply to the next comment for a more complete answer to the question.

7. Page 6, Lines 5-9: Was there a compelling reason to choose the 16 s temporal and 90 m spatial smoothing parameters? This question is probably related to the previous comment. The obvious answer is that spatiotemporal smoothing needs to capture the multi-modal peaks shown in Fig. 2 without completely smearing out the features. I guess I'm having a difficult time being convinced that one could empirically derive the best smoothing parameters and method based only on an "eye test" without further quantitative support.
Post-hoc comment: The appendix content nicely lends further support for how the algorithm works with the adopted spatiotemporal constraints. I was initially going to suggest appendix material that shows how the algorithm would perform with different smoothing methods and parameters - maybe include a final brief appendix section illustrating the sensitivity of one or two cases to different smoothing schemes or spatiotemporal averaging parameters?

Reply: Averaging is performed over more neighboring spectra in time than in space, because Doppler spectra features, e.g. liquid peaks, occur usually in layers and are more consistent in time than in height. E.g. Figure 9a in Bühl et al. (2016, https://doi.org/10.5194/acp-16-10609-2016) shows minimum liquid layer depth was on the order of 50-100 m equivalent to 2-3 range gates assuming 30 m vertical range gate spacing, which motivates our choice of 90 m. We added the following sentence to the methods description to make this clear to the reader: "Fig 9a) in Bühl et al. (2016) shows minimum liquid layer depths on the order of 50 to 100 m equivalent to 2-3 range gates assuming 30 m vertical range gate spacing, which motivated our choice of 90 m."

Following the recommendation, we added another appendix section (Appendix D) which assesses the influence of different smoothing schemes and spatiotemporal averaging on the performance of the algorithm:

"To assess the influence of different smoothing schemes and spatiotemporal averaging space on the algorithm's performance, a sensitivity study was performed. Two smoothing methods available in Matlab are the *moving average* and the locally weighted scatterplot smoothing (*lowess*) schemes. *Lowess* smoothing is very similar to *loess* smoothing with the difference that *lowess* utilizes a first-degree polynomial which is fit to the data subset defined by span.

We trained PEAKO in different configurations using the first training dataset (Table 1). The PEAKO configurations tested were the following:

- Averaging over five spectra in temporal and five spectra in spatial scale, which results in an averaging time scale of 10 s and an averaging height of 150 m. The average spectrum is smoothed using the *loess* method.
- Omitting time-height averaging altogether prior to smoothing the spectra using *loess* smoothing.
- Keeping the spatiotemporal averaging fixed at the default of 16 s and 90 m but using *moving average* smoothing instead of the *loess* method
- Keeping the spatiotemporal averaging scale fixed at the default and using *lowess* smoothing instead of *loess* smoothing

The optimized parameters obtained after training PEAKO in each of above listed configurations were applied to the case study presented in Fig. 5. Figure D1 shows the results.

The panels in Fig. D1 all display a similar pattern with respect to peak number. This is not surprising because the training process of PEAKO is the same for each of the methods, i.e. the three adjustable parameters are adjusted to obtain the best agreement with the human-created training data. A change in the spatiotemporal averaging scale towards more neighbors in height and less neighbors in time does not alter the result significantly. However, performing time-height averaging prior to smoothing at all is important as can be seen in the third panel in Fig. D1: If no spatiotemporal averaging is carried out before smoothing, the features detected by PEAKO become less coherent and more noisy. The two lower panels in Fig. D1 explore the effect of different smoothing schemes on the algorithm performance. Both *moving average* and *lowess* methods are able to reproduce the features detected by PEAKO in the default configuration only with some minor deviations."

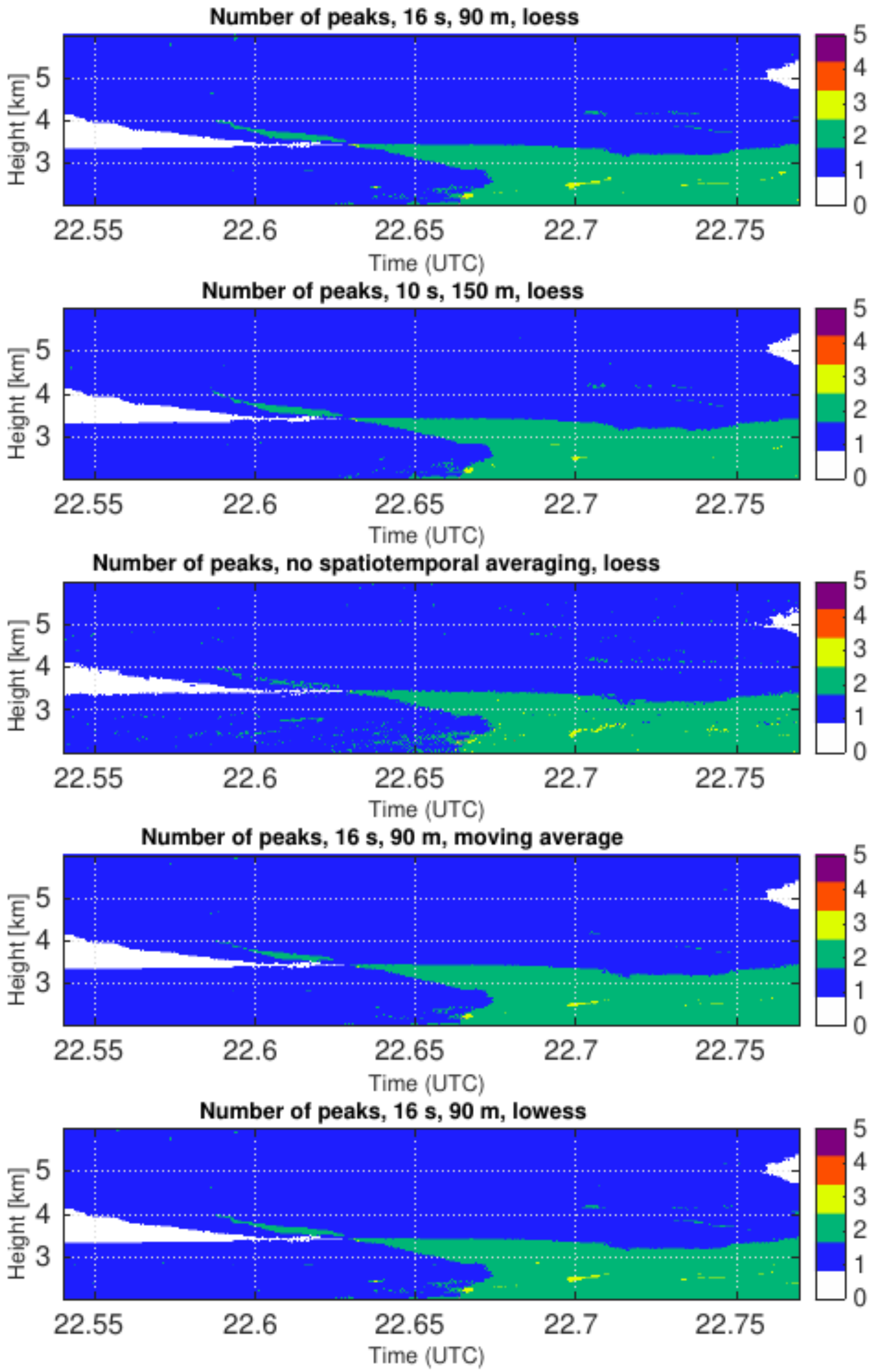

Figure D1. Number of Doppler spectrum peaks detected by PEAKO in five different configurations for the selected case study on 2014-02-21 from 22.54 to 22.77 UTC in 2 to 6 km altitude. Top to bottom: Number of peaks detected by PEAKO in the default configuration (16 s temporal and 90 m spatial averaging prior to loess smoothing), this plot is equivalent to the top panel in Fig. 5; number of peaks detected using 10 temporal and 150 m spatial averaging followed by loess smoothing; number of peaks detected without time-height averaging prior to loess smoothing; number of peaks detected using 16 s and 150 m time-height averaging followed by smoothing using the moving average method; number of peaks detected using 6 s and 150 m time-height averaging followed by lowess smoothing

8. Figure 8 caption: I recommend adding what the black dashed line indicates. It is obviously the SLW layer that is again repeated in a later figure, but it should probably be mentioned here, too.

Reply: We added the following sentence to the figure caption: “The black dashed line marks the boundary of the supercooled liquid layer, indicated by high backscatter and low depolarisation ratio.”