Reply to comments of Reviewer # 1:

We want to thank Reviewer 1 for carefully reading our manuscript and the helpful comments. We are addressing the raised comments in a point-by-point way below:

0) Mark up on the manuscript made while reading it is contained in the attachment. Perhaps some of it will be of value to the authors.

Reply: The notes were certainly of use and we have followed many of the handwritten hints. Thank you for making the extra effort.

1) The essence of the proposed technique is stated in two places within the manuscript: the last full sentence on the bottom of Page 9 which reads "Features prominent enough in time and height to be still visible at all after averaging and smoothing are most probably physical" and the first full sentence at the top of Page 7 which reads "Doppler spectra peaks in low-turbulent liquid cloud droplet layers are very narrow and thus suited to determine the minimum width of a peak considered as physically meaningful." The authors use human analysis of Doppler spectra to identify thresholds in smoothness, peak powers, and peak width to identify physically meaningful peaks. There is some arbitrariness in this approach based on training data. So why not change the approach a bit: use Doppler spectra with no significant returns to identify smoothness, peak power, and peak width thresholds that eliminate all peaks because they are all noise? Any peaks that survive when applied to other observations must then come from hydrometeors. This was the idea that came to mind when reading the two parts of the paper above. It would then come down to a characterization of a radar, much like what the authors hope to do in future studies. Is it a worthwhile approach?

Reply:
The reviewer criticises the arbitrariness of human-created data to train the algorithm, and suggests to change the technique, not using peaks marked by users in Doppler spectra but spectra with no significant peaks at all as input data, to establish thresholds eliminating all noise. These are several good points that deserve a more detailed discussion.

1. We agree that there is arbitrariness in the approach using peaks detected by humans as training to the algorithm. But even though being by definition subjective, the human brain is perfectly tuned to pattern recognition and a role model to many machine learning applications. We tried to counter possible problems caused by human subjectivity by having the training data created by different experienced radar scientists.

2. The suggested alternative approach is definitely an interesting idea, however would not change our technique by 'a bit' but quite drastically. The phrase "Features prominent enough in time and height to be still visible at all after averaging and smoothing are most probably physical" was possibly misworded. The PEAKO algorithm does not primarily aim at determining whether there is physically meaningful signal in a spectrum or not, but at separating several meaningful signals within one spectrum. If this application can be trained using spectra with no significant returns at all is questionable. Usually, the problem is not to detect the main (primary) peak but to distinguish merged peaks. We changed the phrasing of the last sentence on page 9 to "Maxima prominent enough in time and height to be still visible after averaging and smoothing are most probably physical".

3. Eliminating noise is an aspect which is taken care of in part already by applying the Hildebrand & Sekhon criterion to the spectra. No peaks with magnitude below the noise threshold determined by their approach can be detected by PEAKO.

2) The current approach only identifies underlying hydrometeors when they produce separated peaks. The current approach does not work in identifying hydrometeors when the peaks they produce merge together to produce single peaks with or without shoulders. This needs to be pointed out because there are lots of hydrometeors out there that do not produce separate peaks.

Reply: True. The algorithm can only detect a hydrometeor population if it constitutes a maximum in the Doppler spectrum. We added the following sentence to the conclusions, which hopefully stresses this aspect more: "The described approach only identifies underlying hydrometeor populations if the particle types differ sufficiently in their terminal fall velocities to produce individual Doppler spectrum peaks."

3) Perhaps most importantly, the method itself needs to be perfectly described so that it is reproducible. Some might find reproducing the method difficult based on the current description, especially of how the spectra are initially smoothed. Manuscript lines 1-9 on Page 6 are hard to understand in this regard. If the spectra have a temporal resolution of 2 s, then it would take 8 consecutive spectra to cover a 16-s time window. Counting the current spectrum itself, it would take 4 spectra before and 3 after (or 3 before and 4 after) to cover the 16-s time window. Perhaps more likely, 4 spectra before and after the current one were used? If so, these would total 16 s and, together with the current one in the average, a total of 18 s worth of spectra would be averaged. Is this correct? Either way, make perfectly clear what was done. This is a small detail but an important one.
The description of the span in the first two sentences of Page 6 was also hard to understand. First, the phrase "to be considered for spectral smoothing" does not mean that the spectra were actually smoothed. But perhaps the smoothing routine was applied to all spectra in chunks dictated by the span? And the statement "Spectral smoothing is performed using local regression using weighted linear least squares and a 2nd degree polynomial model (loess)." is a bit ambiguous too. How are the local linear and 2nd degree polynomial fits related to each other and to the span? Perhaps a section that illustrates how all of this smoothing works on input spectra to generate an output spectrum would take care of these ambiguities in describing the smoothing. Perhaps it could go something like this: "First, the raw spectrum at the current time and height is replaced by an average spectrum obtained by averaging 27 spectra, 9 in time and 3 in height centered on the current one. Then the averaged spectrum is further smoothed over chunks of spectral bins determined by the span. [Keep going to describe how the span, local linear fits, and 2nd degree polynomial fits work together.]"

Finally, what does "loess" mean? In most dictionaries it is defined as a loamy sediment so not sure what it means in this context.

Reply: You are perfectly right about the time window: We used 4 spectra before and after the current one, totaling 18 s averaging time.
We also agree that the section describing averaging and smoothing of the Doppler spectra is not very well readable. According to your suggestions, we changed the section describing the algorithm. Now it reads:
"As a first step, the raw spectrum at the current time and height is replaced by an average spectrum obtained by averaging 27 spectra, 9 in time and 3 in height centered on the current one. For the given KAZR time-height resolution of 2 s (time) and 30 m (range), this translates to averaging of 18 s in the temporal and 90 m in the spatial dimension. With hydrometeor populations usually appearing in

distinct layers, which are persistent over a certain period of time, more neighbors in time than height are used for averaging. 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. The averaged spectrum is then further smoothed using local polynomial regression. The smoothing method applied, locally estimated scatterplot smoothing (loess), performs weighted linear least-square fitting on consecutive sub-sets of adjacent data points with a 2nd degree polynomial model. The span for smoothing is the fraction of the total number of data points (here: Doppler bins) of one Doppler velocity spectrum to be used for each local fit. loess smoothing was chosen empirically after testing different methods because it showed the best ability to capture peaks while filtering out noise. The span is varied in a range between 3.5 % and 13 %, regularly spaced with a distance of 0.5 %."

4) The red lines in all of the Doppler spectra figures represent the "maximum noise". But the definition of the maximum noise is never presented. Is it the maximum value of the power in a raw (unaveraged) spectrum among those Doppler spectral bins identified as noise by the Hildebrand and Sekhon technique? This is what it seems to be based on the pictures. This needs to be clear in the manuscript.

Reply: We added "Concerning the highest peak of the Doppler spectrum, the prominence is the power difference between the peak maximum and the mean of the spectral noise determined by Hildebrand and Sekhon (1974)." and "The red horizontal line marks the maximum value of the power in the raw (blue) spectrum among those Doppler spectral bins identified as noise by Hildebrand and Sekhon (1974). The black horizontal line is drawn at the mean power of the Doppler bins containing only noise." to the caption of Figure 2.

5) Comment 2) above raises another point. What are your criteria (or perhaps criterion) for eliminating spectral peaks that are too low in power? Are peaks below the maximum noise obtained from the raw spectrum allowed? What about peaks below the maximum noise obtained from the averaged spectrum? This needs to be clear.

Reply: No peaks below the maximum noise obtained from the raw spectrum can be detected by PEAKO. The noise obtained from the averaged spectrum is not considered at all for peak identification. To make this more clear to the reader, we added the following description:
"In the next step, local maxima are identified in the averaged and smoothed spectrum. Only peaks with powers above the raw spectrum's maximum noise are considered. Finally, peaks with prominences below the prominence threshold and widths smaller than the minimum peak width are excluded."

6) The current method seems to have two separate parts: smoothing and peak identification. The caption to Figure 7 indicates that the peak prominence and peak width thresholds were applied to the incoherently averaged and span-smoothed spectra whereas MicroARSCL and perhaps the other algorithms were not. It would be interesting to assess the importance of smoothing for all of the algorithms involved. To this end applying all four algorithms to the raw spectra and then to the incoherently averaged and span-smoothed spectra is of interest. At least all of the algorithms, and especially MicroARSCL, should be applied to the smoothed spectra and then compared. Doing so would help to differentiate the impacts of smoothing versus feature identification.

Reply: Indeed, but here the focus is on comparing existing algorithms with the new PEAKO algorithm, not on adjusting existing ones (e.g. to the same smoothing as is performed in PEAKO). Differences, especially to MicroARSCL, highlight the importance of smoothing for peak identification. In order to assess the importance of smoothing (as opposed to feature identification), we added an Appendix (D)

giving results of a sensitivity study applying e.g. smoothing or not; varying the temporal and spatial averaging and using different smoothing methods:

"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."
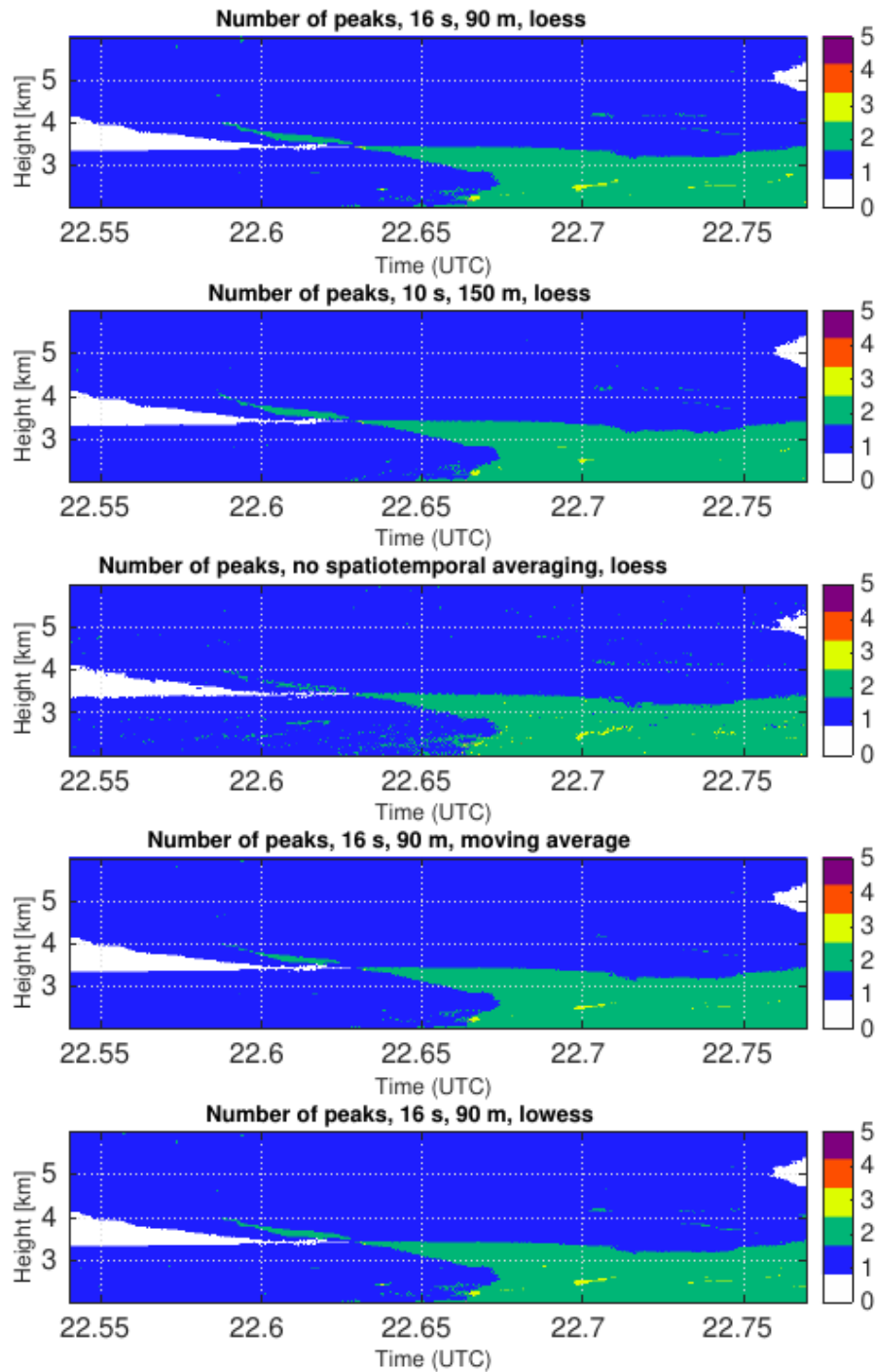
Figure D1. Number of Doppler spectrum peaks detected by PEAKO in five different configurations for the selected case study on 2014-02-21 from 22.54 to 22.77 UTC in 2 to 6 km altitude. Top to bottom: Number of peaks detected by PEAKO in the default configuration (16 s temporal and 90 m spatial averaging 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

7) The word "well" appears on line 13 of the abstract. Replace this subjective statement with something quantitative.

Reply: We agree that "well" is subjective and should be replaced. This issue was also raised by Reviewer #2. According to their suggestion, we rephrased this part of the abstract as follows: "The new algorithm consistently identifies Doppler spectra peaks and outperforms other algorithms by reducing noise and increasing temporal and height consistency in detected features."

8) Not sure what AMT guidelines are, but using past tense to describe events that happened in the past is perhaps preferable to using present tense in such descriptions. Same goes for describing what one did to pursue the study.

Reply: Agreed. We decided to stick to the past tense consistently throughout the manuscript when describing what was done in the data analysis.

9) Page 4, Lines 22-24: These lines describe levels of training and testing. This information is not carried through to Table 1. Table 1 and Lines 22-24 need to be strongly coupled in terms of wording as this would make clearer what data were used for which purpose.

Reply: We changed the table so that now three data sets are distinguished, a 1st training data set, a 2nd training data set and a testing data set, in accordance with the bullet points in Section 3.1 (algorithm description).

10) Page 7, Line 8: Figure 3 does not contain any purely red circles, but rather red dots surrounded by blue circles.

Reply: We rephrased the description of the figure in the text accordingly.

15) Starting around Figure 9, the figures are referenced out of order. Not sure what AMT guidelines are, but referencing figures in order makes them easy to find.

Reply: We moved the first mentioning of Figure 12 to a later paragraph so that figures are referenced in the correct order.

16) Figures 8 and 9 should be identically formatted in every way to make their comparison as easy as is possible.

Reply: We adjusted the grid and y-axis ticks of Figure 9 to match the format of Figure 8

17) A Doppler spectra peak identification procedure built for wind profilers might be applicable to cloud-radar Doppler spectra. It is based on fuzzy logic and received a fair amount of development effort:
Cornman et al. (1998) A Fuzzy Logic Method for Improved Moment Estimation from Doppler Spectra. Journal of Atmospheric and Oceanic Technology, 15, 1287-1305.
This approach has the attribute that thresholds on any one variable do not have to be fixed.

Reply: We thank the reviewer for pointing us to this interesting study. It is certainly true that different approaches for radar peak identification – including fuzzy logic – can be used to study cloud radar Doppler spectra. We acknowledge and cite this reference in the Introduction.

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 <u>minimum</u> 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."
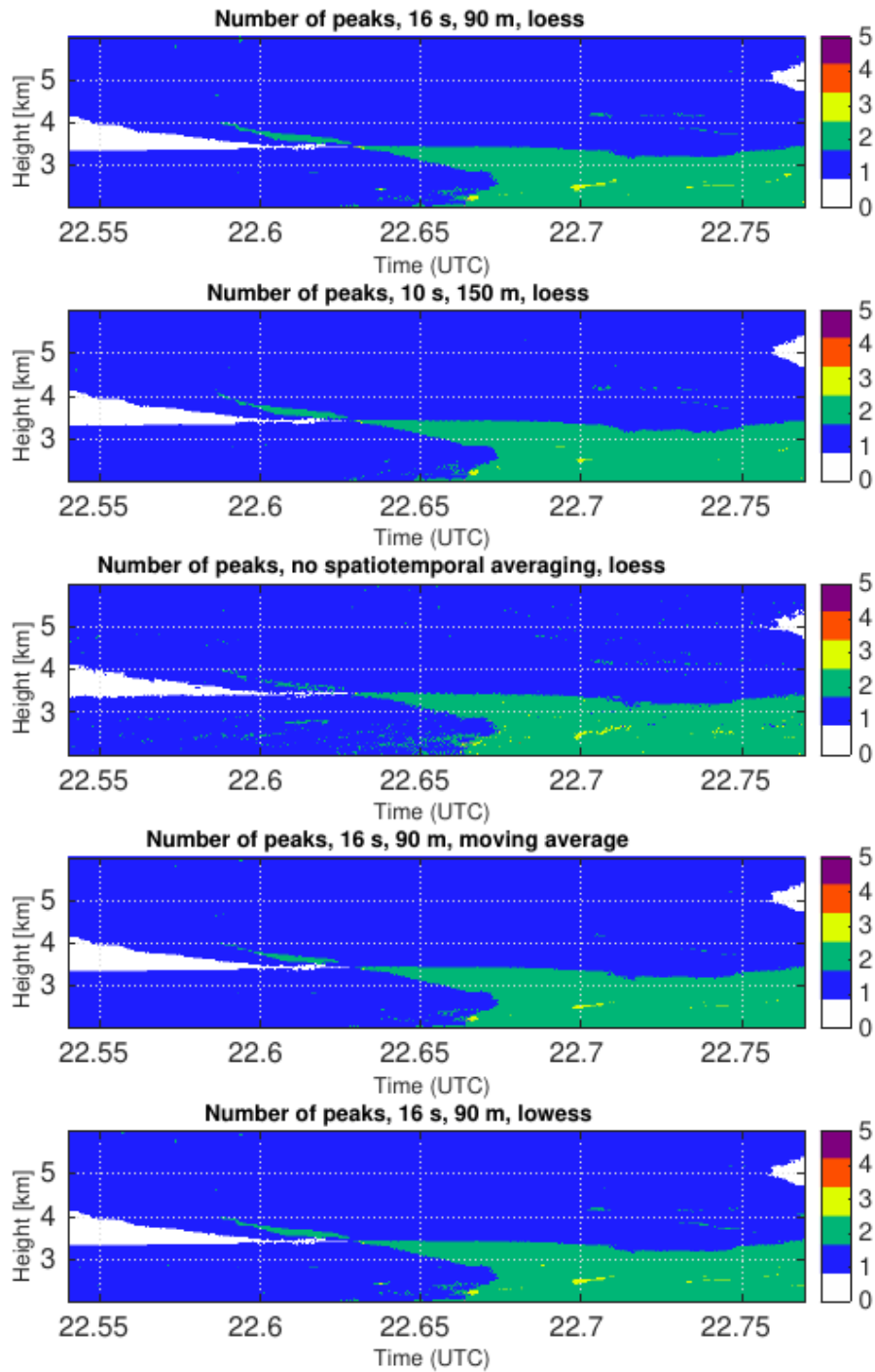
Figure D1. Number of Doppler spectrum peaks detected by PEAKO in five different configurations for the selected case study on 2014-02-21 from 22.54 to 22.77 UTC in 2 to 6 km altitude. Top to bottom: Number of peaks detected by PEAKO in the default configuration (16 s temporal and 90 m spatial averaging 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

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."

Development and validation of a supervised machine learning radar Doppler spectra peak finding algorithm

Heike Kalesse[1,2], Teresa Vogl[1,2], Cosmin Paduraru[3], and Edward Luke[4]

[1]Leibniz Institute for Tropospheric Research, Leipzig, Germany
[2]Institute for Meteorology, Universität Leipzig, Leipzig, Germany
[3]Department of Mining and Materials Engineering, McGill University, Montreal, Canada
[4]Environmental and Climate Sciences Department, Brookhaven National Laboratory, Upton, New York

Correspondence: Heike Kalesse (heike.kalesse@uni-leipzig.de)

Abstract. In many types of clouds, multiple hydrometeor populations can be present at the same time and height. Studying the evolution of these different hydrometeors in a time-height perspective can give valuable information on cloud particle composition and microphysical growth processes. However, as a prerequisite, the number of different hydrometeor types in a certain cloud volume needs to be quantified. This can be accomplished using cloud radar Doppler velocity spectra from profiling cloud radars if the different hydrometeor types have sufficiently different terminal fall velocities to produce individual Doppler spectrum peaks. Here we present a newly developed supervised machine learning radar Doppler spectra peak finding algorithm (named PEAKO). In this approach, three adjustable parameters (spectrum smoothing span, prominence threshold, and minimum peak width at half-height) are varied to obtain the set of parameters which yields the best agreement of user-classified and machine-marked peaks. The algorithm was developed for Ka-band ARM Zenith-pointing Radar (KAZR) observations obtained in thick snow fall systems during the Atmospheric Radiation Measurement Program's (ARM) mobile facility AMF2 deployment at Hyytiälä, Finland, during the Biogenic Aerosols – Effects on Clouds and Climate (BAECC) field campaign. The performance of PEAKO is evaluated by comparing its results to existing Doppler peak finding algorithms. The new algorithm ~~is found to perform well. Its advantage is that the detected features are less noisy and more consistent in time and height than the peak finding results of other algorithms~~ consistently identifies Doppler spectra peaks and outperforms other algorithms by reducing noise and increasing temporal and height consistency in detected features. In the future, the PEAKO algorithm will be adapted to other cloud radars and other types of clouds consisting of multiple hydrometeors in the same cloud volume.

1 Introduction

Determining cloud composition in terms of hydrometeor populations is a non-trivial task in thick cold precipitating clouds below 0°C. In these clouds, supercooled liquid water droplets and solid ice crystals of a variety of shapes and sizes can coexist at temperatures between -40°C and 0°C. Mixed-phase clouds and thick cold precipitating cloud systems play an important role in the Earth's climate, due to their strong influence on the radiative budget (Tan et al., 2016). Global climate models (GCM) still have problems in representing mixed-phase clouds, and especially the supercooled liquid fraction (SLF) accurately (Komurcu et al., 2014).

This motivates the need for highly time- and range-resolved observations of the occurrence of different hydrometeor populations and of cloud phase in the vertical column. The first step towards ~~this characterization of~~ characterizing hydrometeor types is ~~the determination of~~ determining the number of different populations within a certain cloud volume. Profiling cloud Doppler radars are well suited for this task for two reasons:

(i) They are able to penetrate the complete atmospheric column (except for strongly precipitating deep convective clouds), i.e., also beyond the range where lidar is fully attenuated, and

(ii) they can be used as a stand-alone means of inferring the number of different hydrometeor populations and in certain circumstances even cloud phase, because different ice particle populations (and sometimes liquid cloud droplets) and ice particles, which are present simultaneously within a radar sampling volume, are characterized by different terminal fall velocities due to their different particle size distributions and densities (Shupe et al., 2004; Verlinde et al., 2013; Kalesse et al., 2016; Radenz et al., 2019)

.

Each of these different particle size distributions thus generates a peak in the radar Doppler velocity spectrum (Kollias et al., 2016). However, sub-volume turbulence broadens the cloud Doppler spectra peaks and thus smears/smoothes the microphysical signature. Using narrow-beam width antennas and optimizing observational strategies with short dwell time and high vertical resolution reduces turbulence-induced spectrum broadening (Kollias et al., 2016). However, the observed Doppler spectrum is always a convolution of microphysical and dynamical effects.

In order to infer microphysical properties from the radar Doppler spectrum, the peaks have to be separated. ~~Since~~ Because spectra can be noisy and peaks can be merged, this is a non-trivial task, which has already been approached in multiple ways in the past for different cloud types: Shupe et al. (2004) were able to separate observed Doppler velocity spectra into a liquid and an ice spectral mode for a 30-minute long altostratus case study. They empirically defined criteria, which were applied by an algorithm to distinguish multiple peaks in the radar Doppler spectra.

The Microscale Active Remote Sensing of Clouds (MicroARSCL) data product (Kollias et al., 2007; Luke et al., 2008) is generated by a post-processing routine applied to Doppler spectra recorded by the U.S. Department of Energy (DOE) ~~'s~~ Atmospheric Radiation Measurement (ARM) Program millimeter wavelength cloud radars. It uses the morphology of the Doppler spectrum to determine shape parameters like skewness and kurtosis for both the primary peak (highest reflectivity) and, if applicable, an additional noise-separated secondary peak (of lower reflectivity). The peak power densities and modal velocities of up to two local maxima (sub-peaks) located within the primary peak are also included. The MicroARSCL product has~~e.g.~~, for example, been used by Riihimaki et al. (2016) and Oue et al. (2018). The former used it to infer hydrometeor phase in a tropical deep convective system, the latter to study hydrometeor populations in deep precipitating systems in the Arctic. Oue et al. (2018) found multimodal Doppler spectra in ~~the~~ a dendritic/planar growth layer as well as in mixed-phase layers. They also highlighted the added value of joint analysis of Doppler spectra and polarimetric variables from scanning cloud radar observations for snow microphysical studies.

~~Verlinde et al. (2013) were able to separate four~~ 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 ~~in an Arctic cloud case study, in combination~~ combined with high spectral resolution lidar (HSRL) and ~~in-situ observations .~~ in situ observations (Verlinde et al., 2013). Analyses of the Biogenic Aerosols – Effects on Clouds and Climate (BAECC) field campaign have

~~Kalesse et al. (2016) presented a case study of the BAECC field campaign, in which frontal snow was~~ also distinguished up to three noise-floor separated peaks in the recorded Doppler spectra for frontal snow falling through a supercooled ~~liquid~~ water layer (SWL) ~~, which lead to riming of the snow flakes. They could distinguish up to three noise-floor separated peaks in the recorded radar Doppler spectra, which~~ that produced rimed snowflakes (Kalesse et al., 2016). These respective peaks were then used to track microphysical processes along slanted fall streaks~~. However, it is stressed that this~~, although this documented case was special due to the separation of ~~the~~ peaks by the noise-floor ~~. Usually,~~ (merged peaks are

usually observed, motivating the need ~~for the development of~~ to develop robust cloud radar Doppler spectrum peak separation techniques.~~-~~

~~Williams et al. (2018)).~~ 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

5 ~~separation criteria for KAZR observations of liquid-only and mixed-phase clouds at Oliktok Point, Alaska.~~
the main separation criteria (Williams et al., 2018).

All these efforts, using ~~in part considerably~~ somewhat differing approaches, show that there is a need to correctly separate multiple merged peaks in Doppler spectra to aid microphysical understanding of mixed phase cloud processes as well as to improve hydrometeor classification techniques. ~~At the same time,~~ In the past, algorithms mimicing the feature detection

10 skill of human experts in analyzing Doppler spectra have been shown to achieve robust results (Cornman et al., 1998), while recent studies highlight the role of machine learning as a tool for hydrometeor classification based on remote sensing data ~~e.g. , Besic et al. (2016); Praz et al. (2017)~~(e.g. Besic et al., 2016; Praz et al., 2017). This study describes a new algorithm that adopts machine learning tools to classify Doppler spectra peaks in complex mixed phase cloud scenarios.

2 Data set description

15 The Biogenic Aerosols-Effects on Clouds and Climate (BAECC; Petäjä et al., 2016) campaign took place at the Station for Measuring Ecosystem-Atmosphere Relations II (SMEARII) in Hyytiälä, Finland (61°51′N, 24°17′E, 150 m above sea level). The ARM Program deployed their Second ARM Mobile Facility (AMF2) to Hyytiälä from February to September 2014. Within this time frame, a snowfall experiment (BAECC SNEX) took place as a collaborative effort between DOE ARM, University of Helsinki, the Finnish Meteorological Institute (FMI), the National Aeronautics and Space Administration (NASA),

20 and Colorado State University. An intensive operation period (IOP) from 1 February to 30 April 2014 was aimed at measuring snowfall microphysics using a comprehensive suite of remote sensing instruments, complemented by surface-based precipitation observations.

The AMF is constituted of several ground-based remote sensing instruments, including among other things a 35 GHz Ka-band ARM Zenith-pointing Radar (KAZR), as well as a W-, Ka-, and X-band Scanning ARM Cloud Radar (Kollias et al., 2014), a

25 High Spectral Resolution Lidar (HSRL), and a micropulse lidar (MPL). Supplementing these measurements, radiosondes were launched four times daily. This study will focus on the Doppler spectra recorded by the KAZR~~as main data basis~~, and will utilize other observations (e.g., ground-based ~~in-situ~~in situ, HSRL - if applicable) for comparison and validation purposes. The KAZR was operated with a temporal resolution of 2 s, a vertical range gate spacing of 30 m, and a Doppler velocity spectrum resolution (bin width) of 2.37 cm s^{-1}.

30 ## 3 Methodology

In the following section, the supervised Doppler spectra peak detection algorithm developed in this work is introduced. This description is followed by an introduction of the other Doppler spectra peak-finding algorithms which are compared to the new algorithm.

3.1 PEAKO algorithm description

5 In this study, a supervised Doppler spectra peak finding algorithm (in the following text referred to as PEAKO) was developed, which ~~is~~ was trained via hand-marked Doppler peaks as input. The learning process ~~is~~ was split into two phases, the training phase and the test phase, respectively. For that purpose, three data sets, each containing example input and the corresponding desired output, ~~are~~ were created:

- a first training data set, used to obtain an initial model;

10 - a second training data set, about half as large as the first training data set, used to tune the model;

- a testing data set, which has approximately the same size as the second training data set ~~and is~~, used for model evaluation.

With the help of a graphical Matlab interface, in which the currently to-be-marked Doppler spectrum and its surrounding neighbors (in time and height) are displayed in logarithmic space (see Fig. 1), pronounced Doppler spectrum peaks ~~are~~ were

15 hand-marked by an experienced user. Even though this approach is subjective, criteria such as peak width, dynamic range, i.e. the height above noise floor, skewness of the spectrum, and consistency of the feature (peak) in time-height ~~are~~ were taken into account. The locations of these hand-marked peaks (in mean Doppler velocity (V_D) units [m s^{-1}]), as well as their corresponding signal powers (in dBZ)~~are~~, were then saved as data matrices.

20 The training and test data sets were chosen as individual non-overlapping time-height areas rather than randomly splitting all hand-marked spectra into training and test categories. For the training phase, data recorded on 21 February ~~21~~ 2014, 22-23 UTC were utilized, a time period ~~,~~ which was studied in greater detail before (Kalesse et al., 2016; Mason et al., 2018). Data from 16 February 2014, 0-1 UTC, which were in part investigated in a case study presented in Kneifel et al. (2015), ~~are~~ were used in the training phase as well. The third case selected for the training data set, 21 February 2014, 23-24 UTC, was in part

25 analyzed by Kneifel et al. (2015) as well. The test set is comprised of two 1-hour-cases, which were recorded on 2 February 2014, 16-17 UTC, and on 7 February 2014, 23-24 UTC, respectively. In the case of 2 February 2014, 16-17 UTC, the study area was set within the lowest liquid layer where independent ~~lidar (HSRL) measurements can be~~ HSRL measurements were used to check the performance of the PEAKO algorithm for liquid-peak detection. Unfortunately, the HSRL was fully attenuated by near-surface liquid-layers during the other case studies. The chosen period on 7 February 2014 overlaps with another case

30 investigated by Kneifel et al. (2015). Table 1 gives a summary of which measurement periods were used for which of the data sets.
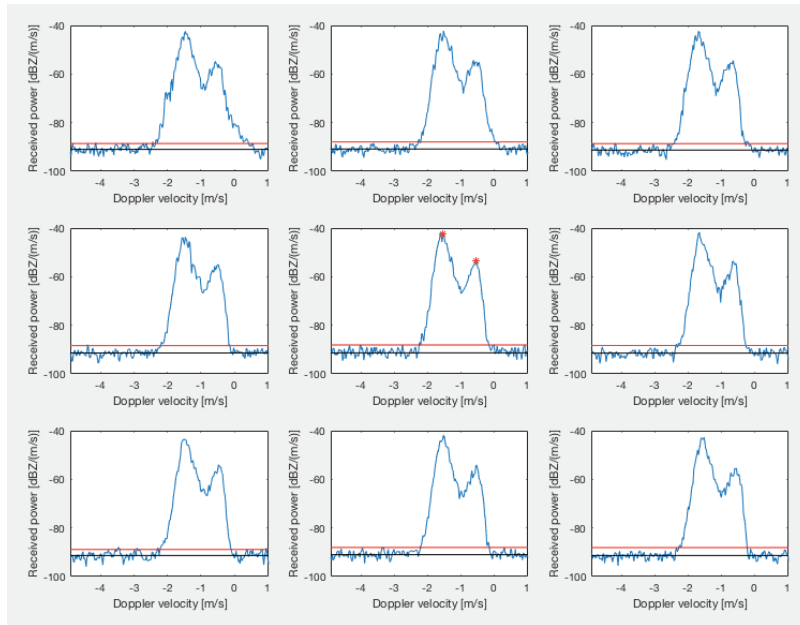
Figure 1. Example of graphical user interface for peak-marking by hand. For the Doppler spectrum in the center ~~panel~~ panel, two peaks (red stars) were marked by the user (HK). The surrounding ~~spectra~~ panels display the spatially and temporally neighboring spectra. Data: ~~random spectra of~~ KAZR spectra observed at TMP on February 16, 2014, 0.03-0.05 UTC between ~~1-1.2~~ 1.0-1.2 km height. The red line marks the maximum noise floor determined according to Hildebrand and Sekhon (1974), and the black line the mean ~~spectral~~ of the noise~~, determined according to Hildebrand and Sekhon (1974)~~.

Table 1. Overview of the measurement periods used in ~~training and test~~ each of the three data sets (first training data set, second training data set, testing data set) containing hand-marked peaks. Published studies of the selected periods, to which results can be compared, are noted as well.

	~~training~~ 1st training data set	~~test~~ 2nd training data set	testing data set
2014-02-02			X
~~16 - 17~~ 16-17 UTC			(comparison to HSRL)
2014-02-07			X
23-24 UTC			Kneifel et al. (2015)
2014-02-16	X	X	
0-1 UTC	Kneifel et al. (2015)	Kneifel et al. (2015)	
2014-02-21	X		
22-23 UTC	Kalesse et al. (2016)		
2014-02-21		X	
23-24 UTC		Kneifel et al. (2015)	

The PEAKO algorithm includes a set of three subsequently described adjustable parameters (smoothing span, prominence threshold, minimum peak width at half-height), which are varied to obtain the set of parameters which yields the best agreement of hand-marked and machine-marked peaks. The search for the best parameter combination is done via a search through a finite set of values ~~for~~ in the three-dimensional search space.

As a first step, the raw spectrum at the current time and height is replaced by an average spectrum obtained by averaging 27 spectra, 9 in time and 3 in height centered on the current one. For the given KAZR time-height resolution of 2 s (time) and 30 m (range), this translates to averaging of 18 s in the temporal and 90 m in the spatial dimension. With hydrometeor populations usually appearing in distinct layers, which are persistent over a certain period of time, more neighbors in time than height are used for averaging. 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. The averaged spectrum is then further smoothed using local polynomial regression. The smoothing method applied, locally estimated scatterplot smoothing (loess), performs weighted linear least-square fitting on consecutive sub-sets of adjacent data points with a 2nd degree polynomial model. The span for smoothing ~~defines~~ is the fraction of the total number of data points (here: Doppler bins) of one Doppler velocity spectrum to be ~~considered for spectral smoothing. Spectral smoothing is performed using local regression using weighted linear least squares and a 2nd degree polynomial model (loess). This smoothing method~~ used for each local fit. loess smoothing was chosen empirically after testing different methods ~~since~~ because it showed the ~~most promising results~~best ability to capture peaks while filtering out noise. The span is varied in a range between 3.5% and 13%, regularly spaced with a distance of 0.5%.~~Spectral smoothing is performed on an average of 16 s temporal and 90 m spatial dimension, i.e. for the given KAZR time-height resolution of 2 s (time) and 30 m (range), the mean of 4 neighbors in time and 2 neighbors in height-dimension was made before the machine-based peak finding was applied to smooth out some spurious features in individual Doppler spectra. With hydrometeor populations usually appearing in distinct layers which are persistent over a certain period of time, more neighbors in time than height are used for averaging.~~

In the next step, local maxima are identified in the averaged and smoothed spectrum. Only peaks with powers above the raw spectrum's maximum noise are considered. Finally, peaks with prominences below the prominence threshold and widths smaller than the minimum peak width are excluded. The prominence threshold is a measure of how much a peak stands out relative to the other peaks in the considered Doppler spectrum. The prominence of a peak is the power difference (dynamic range) of the peak's maximum and the signal's minimum between the considered peak and the nearest higher peak. Concerning the highest peak of the Doppler spectrum, the prominence is the power difference between the peak maximum and the ~~noise floor~~mean of the spectral noise determined by Hildebrand and Sekhon (1974). This parameter is varied between 0 and 2 dBZ m^{-1} s^{-1} in the training phase of PEAKO. ~~Fig.~~Figure 2 illustrates the definition of the peak prominence: A spectrum with three merged peaks is shown and their prominences are drawn as red vertical lines. In the case of the rightmost peak ($V_D \approx -0.5$ m s^{-1}), the prominence is defined as the power difference between the peak's maximum and the minimum between this peak and the nearest higher peak. This minimum is located approximately at the leftmost peak's right edge (marked with a solid black vertical line at around -1.1 m s^{-1}). For the peak with the lowest power at $V_D \approx -0.8$ m s^{-1}, the prominence of 0.25 dBZ m s^{-1} is barely visible because it is defined as only the distance between this peak's maximum and the minimum
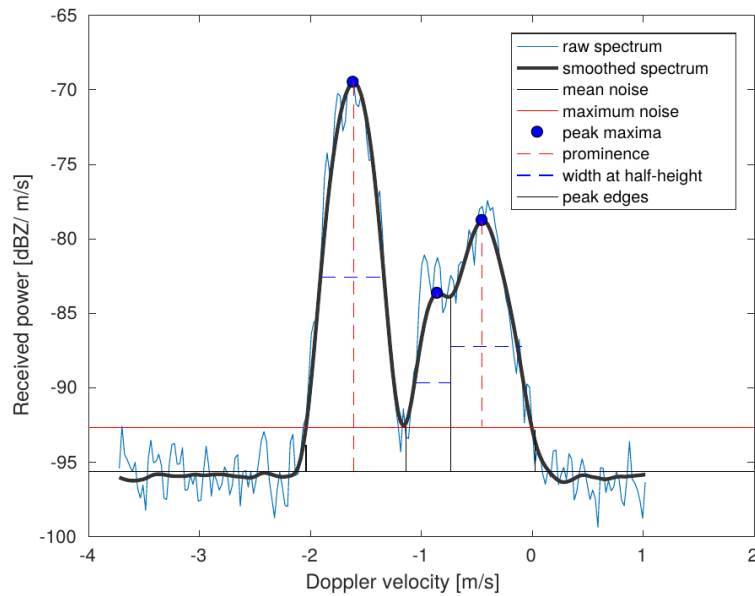
Figure 2. Example spectrum (blue line) with multiple merged peaks, recorded on 21 February ~~21~~ 2014, 22.7 UTC ~~in~~ at 2.44 km altitude. The smoothed spectrum (average of neighbor spectra in time and height domain, smoothed using a span of 8.5%) is shown as well (bold black line). For each of the three peaks marked in this spectrum (blue dots), the prominences (red dashed lines) and widths at half height (blue dashed lines), as well as the edges (vertical black lines) of the peaks are marked. The red horizontal line marks the maximum value of the power in the raw (blue) spectrum among those Doppler spectral bins identified as noise by Hildebrand and Sekhon (1974). The black horizontal line is drawn at the mean power of the Doppler bins containing only noise.

to the closest higher peak, which is the peak with the lowest absolute V_D.

The third adjustable parameter is the minimum peak width at half-height. The range of values in which it is varied (4.2 to 8.4 Doppler velocity bins, spaced with a distance of 1.05) was determined from a low-turbulence cloud region only consisting of liquid droplets (21 February 2014, 22.53 - 22.59 UTC 2.9-3.1 km). Doppler spectra peaks in low-turbulent liquid cloud droplet

5 layers are very narrow and thus suited to determine the minimum width of a peak considered as physically meaningful. At the given KAZR resolution, these peaks were found to be between 4.2 and 8.4 V_D bins wide corresponding to about 10 - 20 cm s^{-1}.

To determine the optimal parameter combination, a similarity measure is defined, based on the maximum overlapping area of detected peaks as illustrated in Fig 3: For a certain set of smoothing span, minimum peak width and prominence threshold,

10 the algorithm will detect certain peaks in a Doppler spectrum (shown as red ~~circles in Fig.~~ dots surrounded by blue circles in Figure 3). For these peaks, as well as for the peaks marked by a user in the same Doppler spectrum (red stars in Fig. 3), the edges (marked with vertical lines) are determined, which are defined as follows: The edge is either the first Doppler bin ~~,~~ where the spectrum power is smaller than the maximum noise floor ~~,~~ or, in case of a merged peak, the minimum (saddle point)

8
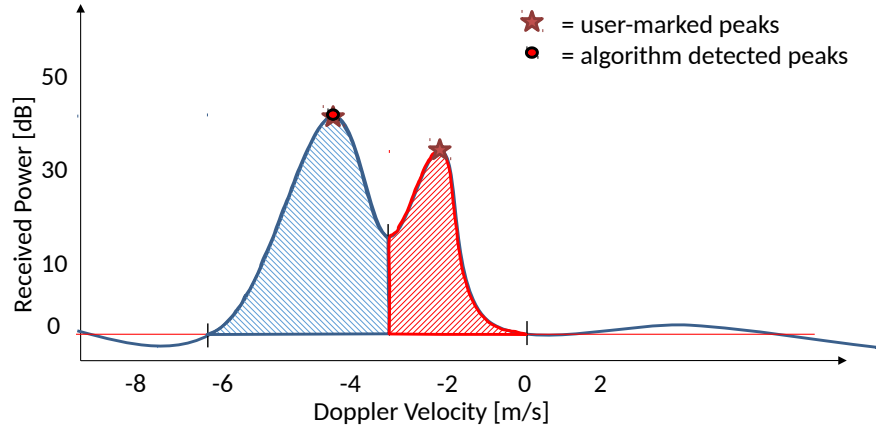
Figure 3. Schematic to visualize how the similarity measure to compare user-marked and algorithm-found peaks in Doppler spectra: Areas of matching peaks are summed up (blue hatched area), and the areas of mismatched peaks (red hatched) are subtracted.

between the merged peaks. In the next step, the overlapping area for each pair of hand-marked and machine-found peaks is identified: In case of multiple peaks in one spectrum, the sum of all overlapping areas is determined. Non-overlapping regions caused either by a mismatch in the number of hand-marked and machine-found peaks or by a different location (in x-direction, i.e., in Doppler velocity) of the pair of peaks, are penalized ~~in the similarity area measure~~ by subtracting the non-overlapping area (marked red hatched in Fig. 3) of the mismatched peaks _from the similarity area measure_. The optimum parameter combination is the triplet of span, prominence threshold, and minimum peak width ~~,~~ for which the similarity has its maximum value.

3.2 Description of other radar Doppler spectra peak-finding algorithms

Three other radar Doppler velocity spectrum peak finding algorithms were compared to PEAKO and will be briefly explained in the following.

The algorithm described in Shupe et al. (2004) (from now on referred to as "Shupe_04") uses peak finding criteria optimized to one relatively short (30 min) mixed-phase cloud case study period. In short, the power of the primary (strongest) peak must be at least 4 standard deviations of the noise above the mean spectral noise level. In addition, one or more secondary peaks are identified as maxima being at least 2.5 standard deviations of the noise greater than the mean spectral noise level. Both primary and secondary peaks must have a width of at least 0.448 m s^{-1}. Merged peaks are identified as separate spectral modes if the saddle point between the two maxima is lower than 65% of the lowest of the two peaks from the noise level.

The MicroARSCL data product (Kollias et al., 2007; Luke et al., 2008) decomposes noise floor subtracted and 3-bin-boxcar-filtered radar Doppler spectra into a primary peak (defined as the peak containing the bin with maximum spectral power

density), dominating the total reflectivity and containing up to two sub-peaks, and a possible secondary peak. Fixed thresholds, i.e. minimum primary or secondary peak width (Pw_{min}), minimum sub-peak height (Ph_{min}), minimum sub-peak separation (Ps_{min}), and minimum primary-secondary peak edge separation (Pn_{min}), are applied to extract a suite of variables from the Doppler spectra. Pw_{min} is set to 5 Doppler velocity bins ($0.12\,\mathrm{m\,s^{-1}}$), Ps_{min} to 3 Doppler velocity bins ($0.07\,\mathrm{m\,s^{-1}}$), Pn_{min}

5   is 1 Doppler velocity bin ($0.0237\,\mathrm{m\,s^{-1}}$), and Ph_{min} is $1\,\mathrm{dBZ\,m^{-1}\,s^{-1}}$. For the comparative study, a third technique, being a polynomial fitting algorithm as described in Kollias et al. (1999) and Kollias et al. (2003) was also applied to the radar Doppler spectra which are analyzed in this study. This routine first extracts the parts of the spectrum above the maximum noise floor (~~signal-to-noise~~ the noise threshold determined by Hildebrand and Sekhon (1974)) and extends the edges of the found peaks down to the mean noise floor. In a next step, each continuous sample of data above the noise floor is identified as a

10   sub-spectrum. Sub-spectra that are classified as being too narrow (with velocity ranges of the peak smaller than $0.2\,\mathrm{m\,s^{-1}}$) are excluded. For each of the remaining sub-spectra, polynomial fitting of 12th order is applied. The first and second derivatives are taken to identify minima and maxima. Peaks are defined as sequences of minimum-maximum-minimum. Peaks having a velocity range smaller than $0.2\,\mathrm{m\,s^{-1}}$ are ignored, as well as peaks with an amplitude smaller than $2\,\mathrm{dB}$, with amplitude being defined as the difference in reflectivity between consecutive minimum and maximum.

4   Results and Discussion

The following ~~chapter~~ section will be structured as follows: In ~~section~~ subsection 4.1, the best parameter values obtained during the training phase of the PEAKO algorithm are summarized. The peaks detected by one of these best parameter combinations are compared to peaks found by the Shupe_04, ~~the MicroARSCLproduct and the~~ MicroARSCL, and Polyfit12 algorithms in a case study. It should be noted that PEAKO was trained with a subset of data from the same distribution as this firstly

20   presented case study. This means that PEAKO has somewhat of an advantage over the other three algorithms when comparing on the training data set. Section 4.2 summarizes the testing phase and presents a comparative independent study case, in which PEAKO-found peaks are again compared to peaks detected by the three other algorithms and validated against HSRL retrievals of liquid water droplets. More case studies are presented in the Appendices.

4.1   Training phase of the PEAKO algorithm

25   The training phase was split into two steps: Initially, peaks marked manually in 1340 Doppler spectra were used for training the PEAKO algorithm and obtaining an initial model via a coarse parameter search. This initial training resulted in six equally well performing combinations of span, prominence threshold and width, which all yield the same value of the similarity measure. A more fine-resolved search for the three parameters was then performed, using 775 Doppler spectra with user-marked peaks. This second, refined training again resulted in several combinations of minimum peak prominence, minimum peak width,

30   and smoothing span yielding the same similarity. Table 2 gives an overview of the possible ranges found for the three PEAKO parameters after the initial training and the finer-resolved parameter search. The span for loess smoothing became slightly larger (increased by 0.5 - 1% in absolute terms) and converged to one single possible value (8.5%). The minimum peak prominence

Table 2. Ranges of the parameters yielding the highest similarity measure after the initial and the fine-tuned training using the first and second training data set, respectively.

	optimal parameter range after initial training	optimal parameter range after fine-tuned training
span for smoothing	7.5 - 8%	8.5%
peak prominence threshold	$0.15\,\mathrm{dBZm^{-1}s^{-1}}$	$0.1\,\mathrm{dBZm^{-1}s^{-1}}$
minimum peak width	4.2 - 6.3 V_D bins	4 - 6.25 V_D bins
	$(9.95 - 14.93\,\mathrm{cm\,s^{-1}})$	$(9.48 - 14.81\,\mathrm{cm\,s^{-1}})$

decreased by one third, i.e. from $0.15\,\mathrm{dBZm^{-1}s^{-1}}$ to $0.1\,\mathrm{dBZm^{-1}s^{-1}}$. This prominence threshold is very low compared to values used by other peak-finding techniques. However, in other approaches, spectra are usually not smoothed and neighbor-averaged. ~~Features~~ Maxima prominent enough in time and height to be still visible ~~at all~~ after averaging and smoothing are most probably physical, justifying the low prominence threshold. The possible values for the minimum peak width did not change

5   significantly between the initial and the more refined model and ranges between 0.09 to $0.15\,\mathrm{m\,s^{-1}}$ (i.e., V_D range, from 4 to $6.25\,V_D$ bins for the given KAZR Doppler spectra resolution). Doppler spectrum peaks detected by PEAKO configured in one of these combinations (span = 8.5%; prominence treshold =$0.1\,\mathrm{dBZm^{-1}s^{-1}}$; minimum peak width =$4\,\mathrm{m\,s^{-1}}$) are compared to peaks found by other methods for a study case on 21 February 2014, 22.54 to 22.77 UTC, at 2 to 6 km height. This parameter set containing the smallest possible minimum peak width was chosen because it is most stringentand thus best suited for the

10 detection of narrow supercooled liquid water peaks. The selected period was discussed in detail in Kalesse et al. (2016).

~~Fig.~~Figure 4 shows the first three radar moments, i.e. the radar reflectivity factor Ze, the mean Doppler velocity (MDV) and the Doppler spectrum width σ for the first selected case study, which is set from 21 February 2014, 22.54 UTC to 22.77 UTC in 2 to 6 km height. This time period is characterized by the passage of a warm occlusion in Hyytiälä, Finland, shown by the continuously lowering frontal snow cloud base characterized by high Ze. A midlevel mixed-phase cloud was present before

15 the front approached. It can be identified by its supercooled liquid water (SLW) layer near cloud top between approximately 2.9 and 3.2 km altitude. Before the snow cloud moves in (22.54 to 22.69 UTC), new ice is formed from this SLW layer and growing in size while sedimenting, leading to a slight increase in Ze, MDV (absolute value) and σ with decreasing altitude. As the frontal cloud moves in and snow begins to fall through the SLW layer, riming takes place along slanted fall streaks. A more in-depth analysis of the synoptic situation and the observed microphysical growth processes is given in Kalesse et al.

20 (2016). The number of detected peaks for this case study is shown in ~~Fig.~~Figure 5. All algorithms show a similar general picture with an increasing number of spectral peaks as the snow front moves in and snow starts falling through the SLW layer of the midlevel mixed-phase cloud. A closer examination however reveals some differences between the methods: PEAKO and Shupe_04 algorithm have very similar results except for some areas in the snowfall region where PEAKO detects 3 peaks and Shupe_04 detects 2 peaks. MicroARSCL generally shows higher variability than the other algorithms, the small areas of

25 higher peak number often coincide with increased spectrum width in ~~Fig.~~Figure 4. Polynomial fitting shows 3 peaks in the area where snow falls through the top of the SLW layer and is otherwise very similar to PEAKO and Shupe_04. The areas where the
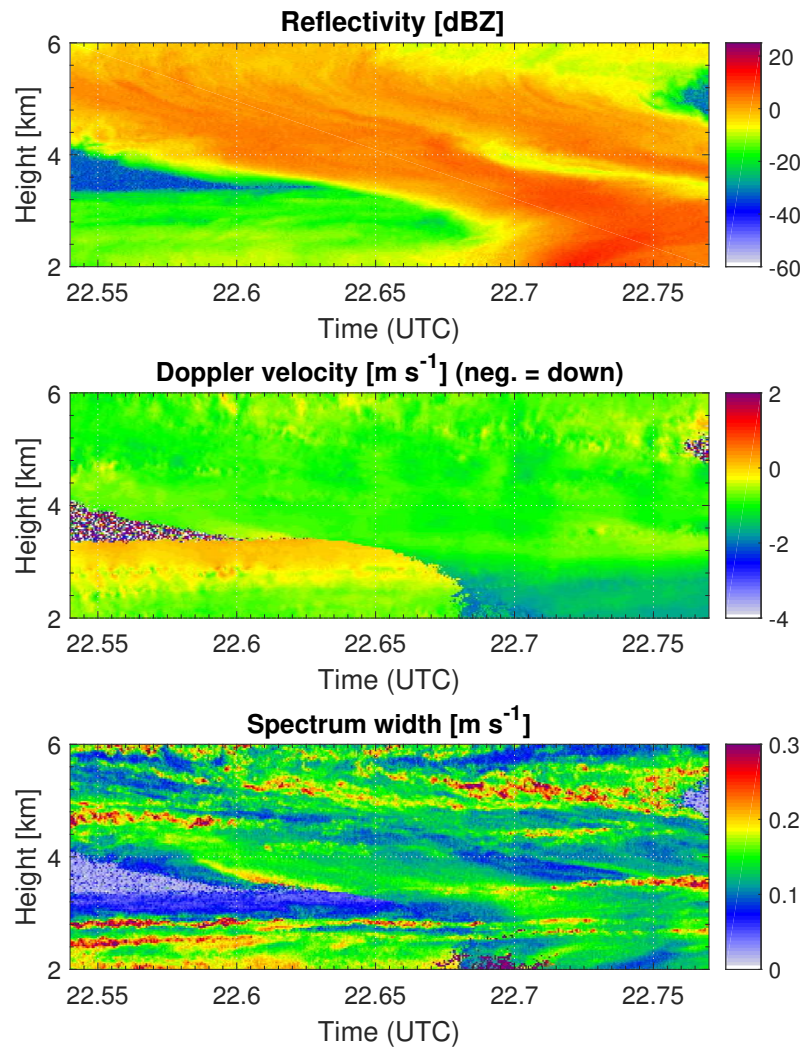
Figure 4. Case study period from 2014-02-21 22.54 UTC to 22.77 UTC in 2 to 6 km height. Top to bottom panels show the radar reflectivity factor Ze, the mean Doppler velocity (MDV), and the Doppler spectrum width σ of the main peak in the Doppler spectrum.
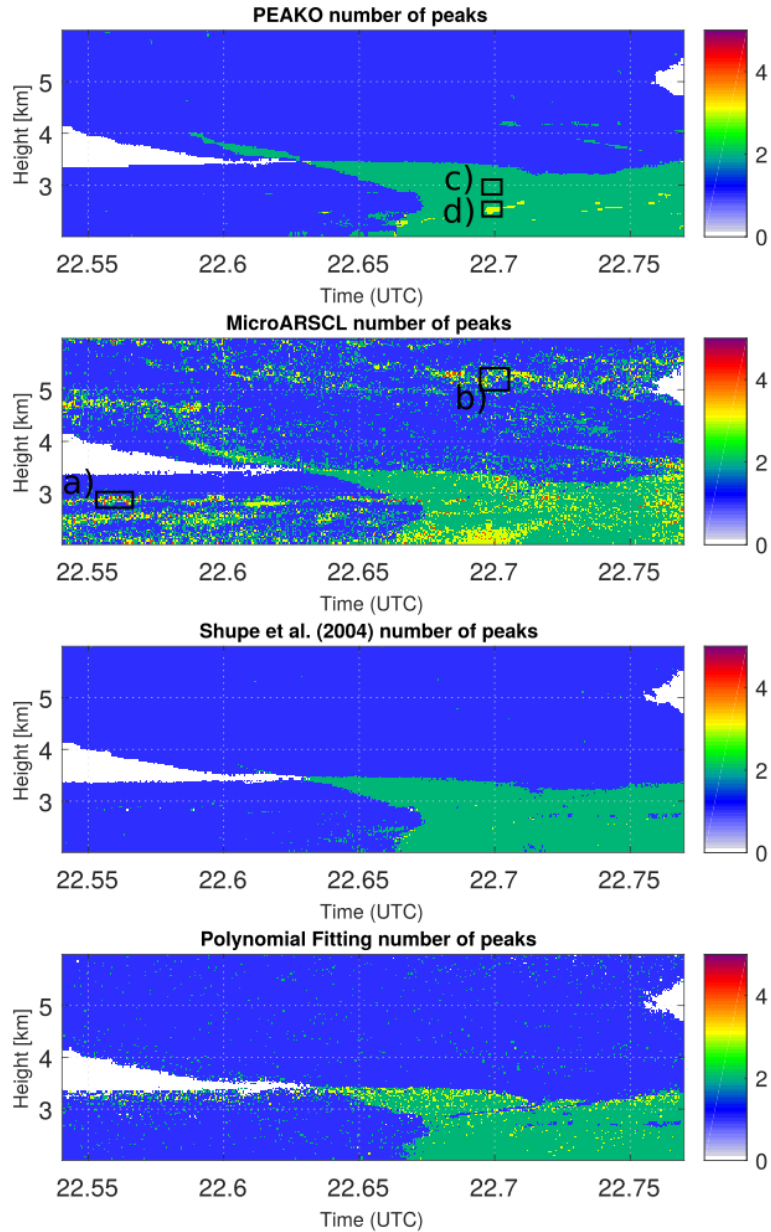
Figure 5. Number of Doppler spectrum peaks detected by different algorithms for the selected case study on 2014-02-21 from 22.54 to 22.77 UTC in 2 to 6 km altitude. Top to bottom: Number of peaks found by the PEAKO algorithm for one of the "best parameter" combinations obtained in the training phase of the algorithm; Number of peaks in MicroARSCL data product; Number of peaks detected using the criteria of Shupe_04; Number of peaks determined by Polyfit12
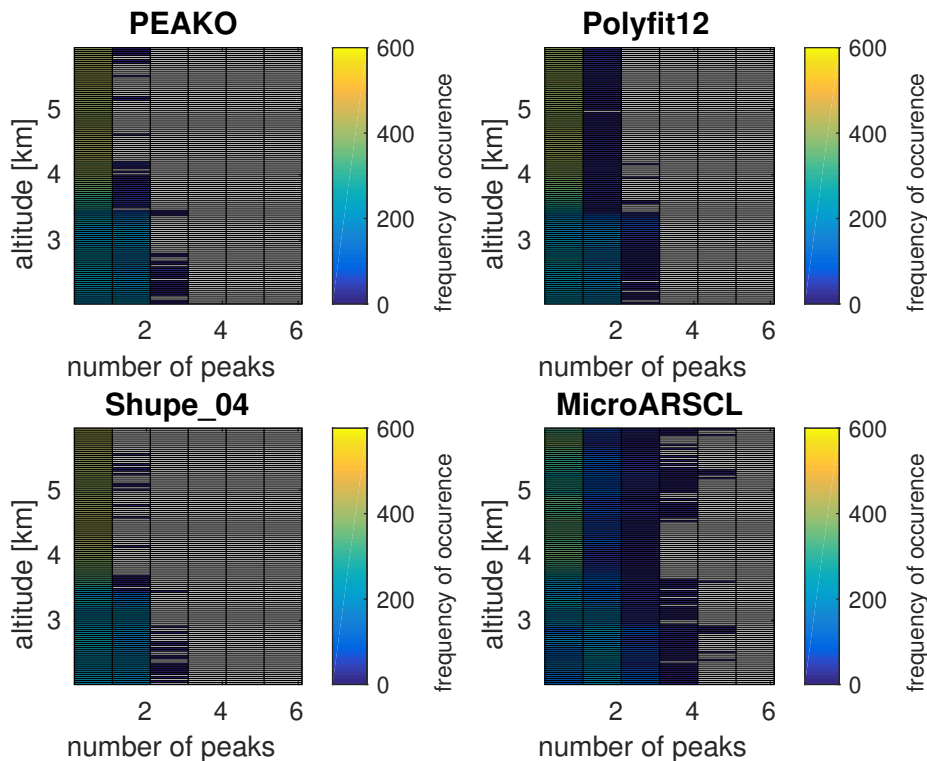
Figure 6. Contoured frequency by altitude diagram (CFAD) for the frequency of occurrence of number of detected peaks from different algorithms for the case study period on 2014-02-21 from 22.54 to 22.77 UTC in 2 to 6 km altitude.

different algorithms show discrepancies are now examined in more detail. For that purpose, contoured frequency by altitude diagrams (CFAD, Fig. 6) are created to compare the results of the algorithms in a different way. The CFAD shows the number of detected peaks (abscissa) at different heights (ordinate) as a colored frequency of occurrence for the total case study sample. For all four compared algorithms, it is most common that only one peak is detected. This is especially true for higher altitudes

5   within the snow front (4 km and above). It is visible that the PEAKO algorithm and the results obtained using the Shupe_04 approach agree to a large extent. In the polynomial fitting approach, two or three peaks are detected more often, especially in the layer just above 3 km altitude, where several spectra are classified to contain three peaks. The MicroARSCL data product contains even more Doppler peaks, often three or more, over the complete altitude range.

In ~~Fig.~~ Figure 7, four exemplary spectra from regions where the algorithms show discrepancies are shown along with the peaks

10   detected by each of the four algorithms. The spectrum in ~~Fig.~~ Figure 7a is recorded in 2.83 km height at 22.56 UTC, below the SLW and before the snow front moves in. As discussed in Kalesse et al. (2016), ice particles which are nucleated in the SLW layer of the midlevel cloud and growth through water vapor deposition lead to this Gaussian shaped monomodal Doppler peak. The spectrum is relatively broad (as can also be seen from ~~Fig.~~ Figure 4) and noisy, pointing to turbulence. MicroARSCL is

14
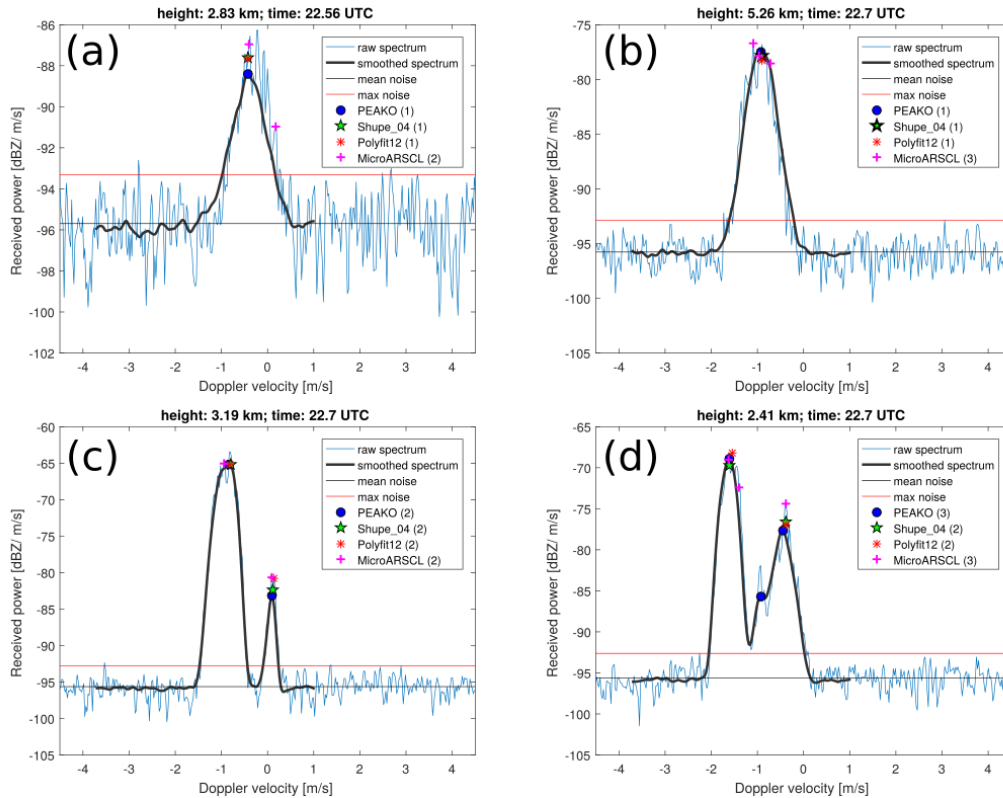
Figure 7. Four exemplary Doppler spectra picked from the case study on 21 February 2014, 22.54 to 22.77 UTC in 2 to 6 km altitude. The averaged and smoothed spectrum which is used as input to PEAKO is drawn in bold over the original spectrum. The peaks detected by the four algorithms are marked. The number of peaks found by each algorithm is noted in parenthesis in each figure legend. Please note the different y-scales.

sensitive to small-scale noise of the original spectrum which the other algorithms are not sensitive to and thus overestimates the number of spectral peaks. ~~Fig.~~ Figure 7b shows a spectrum from later on, at 22.7 UTC, in the upper part of the frontal snow cloud in 5.26 km height. In this time-height region, PEAKO, Shupe_04 and Polyfit12 detect the Gaussian-shaped snow peak, but MicroARSCL is again sensitive to small-scale fluctuations in the Doppler spectra and finds three peaks. The example in

5 ~~Fig.~~ Figure 7c is taken from 3.2 km altitude at 22.7 UTC where snow from the frontal cloud starts to fall through the SLW. All four algorithms are able to detect the very narrow noise-floor separated peak produced by the supercooled liquid droplets with V_D near $0\,\mathrm{m\,s^{-1}}$. ~~Fig.~~ Figure 7d shows a spectrum recorded at the same time but below the SLW layer in 2.41 km height, where freshly generated ice (V_D = -0.5 m s^{-1}), unrimed snow (V_D = -1 m s^{-1}), and rimed snow ((V_D = -1.7 m s^{-1})) are present. These hydrometeor populations do not have sufficient differences in terminal fall velocities and thus produce a spectrum with

10 merged peaks. For this example, Shupe_04 and Polyfit12 detect the two main maxima (freshly generated ice and rimed snow), whereas MicroARSCL and PEAKO find three peaks. However, closer examination of the example spectrum shows that the two

algorithms each detect an additional peak in different locations: MicroARSCL finds a sub-peak in the faster falling hydrometeor population (the Gaussian peak of the rimed snow) while PEAKO detects a sub-peak in the slower-falling mode which exhibits a strong skewness. This subpeak is most likely caused by snow from the upper layers which remained unrimed as discussed in Kalesse et al. (2016).

5   The results of the PEAKO comparison to the other three peak-finding approaches for the other two training data sets, i.e., for 16 February 2014, 0.67 to 0.92 UTC and for 21 February 2014, 23.01 to 23.25 UTC are shown in Appendix A and B. Time-height plots of peak number are shown in ~~Fig.~~ Figure A2 and B2; CFAD diagrams can be found in ~~Fig.~~ Figure A3 and B3. Both of these time periods were as well analyzed by Kneifel et al. (2015) in depth, so it was possible to compare the microphysical signatures reported in this study to the Doppler peaks detected by the four algorithms. For 16 February 2014,

10   0.67 to 0.92 UTC, Kneifel et al. (2015) reported high values of microwave radiometer derived liquid water path of 100 to 500 g m^{-2} and clear signatures of large aggregates in the dual-wavelength ratios of Ka-W-band below 2 km as well as in the reflectivity fall streak feature at 0.85 UTC which were also seen in the X-Ka dual-wavelength ratios around 0.85 UTC. The general structure of the layers of Doppler peak number detected by PEAKO, Shupe_04 and Polyfit12 again agree to a large extent, whereas MicroARSCL detects a higher number of peaks in both cases (Fig. A2 and B2). The fall streak feature

15   which exhibits particle size sorting was better detected by PEAKO and MicroARSCL than by the other two algorithms. The increase in number of peaks at 0.67 to 0.75 UTC can be explained by the presence of large needle aggregates of sometimes more compact and sometimes very open structure as explained in ~~(Kneifel et al., 2015)~~ Kneifel et al. (2015) which lead to the interesting multi-modal Doppler spectrum with up to four peaks as shown in ~~Fig.~~ Figure B4d). Ground-based ~~in-situ~~ in situ observations show that during 0.75 - 0.85 UTC aggregates and rimed particles with enhanced terminal velocities were present

20   and that the number of large aggregates was further found to decrease while number of increasingly rimed aggregates further increased until 1 UTC. Radio sounding observations on 15 February 2014 at 23.2 UTC show a thin layer at 0.8-0.9 km altitude which is subsaturated with respect to ice and liquid and which might explain the decrease to 1 found Doppler peak at this altitude. For 21 February 2014, 23.01 to 23.25 UTC, Kneifel et al. (2015) report the transition from a low concentration of strongly rimed particles (lump graupel) to aggregate snowfall with large snowfall rates and increasing size and number of

25   the aggregates. The fast transition of the snowfall from rimed particles to aggregates results in the bimodal Doppler spectra (with two found peaks) at 23 to 23.05 UTC and monomodal spectra afterwards. For this case study, PEAKO and Shupe_04 and Polyfit12 agree well with the situation described in Kneifel et al. (2015) while MicroARSCL overestimates the number of peaks especially in the turbulent boundary layer and near 4 km altitude.

4.2   Testing phase of the algorithm

30   Using the tuned parameter pairs obtained in the training phase, the PEAKO algorithm is again compared to the other three algorithms, as well as to data measured by an independent instrument, the HSRL. For this purpose, a case of a frontal passage associated with snow on 02 February 2014, 16 to 17 UTC was analyzed. During this time, a liquid-topped mixed-phase cloud with cloud top temperature of (T = -4°C) and cloud top height of 2.6 km was present (Fig. 9). A deeper precipitating cloud system with cloud top around 8 km (cloud top temperature of -40°C) was approaching the TMP site at about 16.27 UTC. The
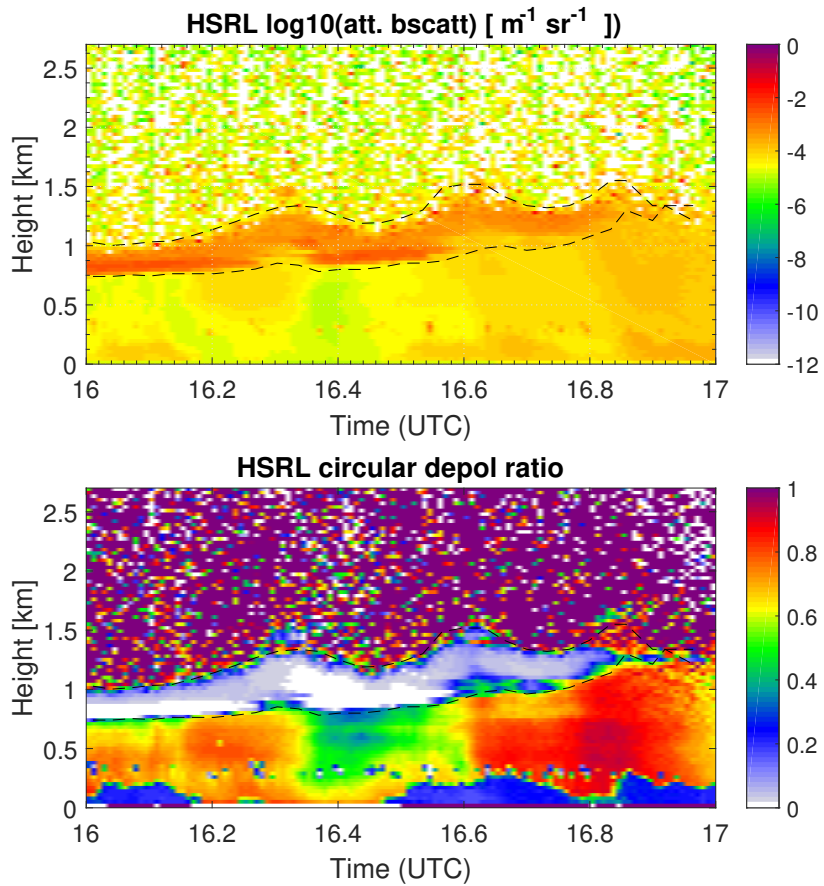
Figure 8. HSRL measurements from the case study period from 02 February 16 UTC to 17 UTC in 0 to 3 km height. The top panel shows the attenuated backscatter cross section and the lower panel the circular depolarisation ratio. The black dashed line marks the boundary of the supercooled liquid layer, indicated by high backscatter and low depolarisation ratio.

surface temperature was -5°C. During the first half of the hour-long case study, the HSRL detected an embedded layer of SLW in the mid-level cloud, characterized by high backscatter coefficient values and low depolarisation ratio values (Luke et al., 2010) in ~~Fig.~~ Figure 8. The SLW layer is located between 0.8 and 1 km height and is slightly lifted as the front is moving in. Its base and top are traced with dashed lines in ~~Fig.~~ Figure 8. After 16.8 UTC, the microwave radiometer derived LWP decreases

5   from around $300\,g\,m^2$ to approximately $60\,g\,m^2$ and the lidar does not detect the cloud base anymore due to the scarcity in small liquid droplets to which the lidar is sensitive to and due to the strong snow fall. The strong decrease in LWP is again pointing to riming~~, which is substantiated by a closer look at the Doppler spectra (Fig. 12)~~. Analysis of the ground-based ~~in-situ~~ in situ Particle Imaging Package (PIP) data shows a variety of different precipitating particles during that one hour time period (Annakaisa von Lerber, personal communication): Around around 16 UTC, oblate particles, possibly needles, and some

10   small needle aggregates are present. When Ze decreases around 16.3 - 16.55 UTC (Fig. 9), no large particles are present at all, just very small ones, maybe single pristine crystals (the resolution of PIP is not good enough to distinguish). At 16.55 - 16.8 UTC when Ze increases strongly and LWP decreases significantly, the PIP observes a clear change to round, dense, fast falling particles, indicative of small graupel. Finally, from 16.8 UTC onward, particle sizes at the ground increase, there are more (quite dense) aggregates, resulting in Ze of up to 10 dBZ.

15   ~~Fig.~~ Figure 9 shows the first three radar moments (Ze, MDV and σ) of the main peak for the selected case study. The supercooled liquid layer at the top of the mid-level cloud extends from about 2.1-2.4 km and is characterized by MDV of near $0\,m\,s^{-1}$. Snow fall rate is at first low and increases at about 16.6 UTC (surface meteorological observations, not shown). Pronounced fall streaks can be seen coinciding with large values of spectrum width, indicating the presence of several hydrometeor populations, producing Doppler spectra with broad merged peaks. ~~Fig.~~ Figure 10 reveals that the number of peaks detected

20   by the four algorithms differ significantly for this case study. Shupe_04 and Polyfit12 again agree to large extents although Shupe_04 does mostly only detect one peak at 2-2.5 km height before 16.3 UTC and after 16.5 UTC where all other algorithms mostly detect two peaks. MicroARSCL generally detects a larger number of peaks in the Doppler spectra, PEAKO is in this case more similar to MicroARSCL than to the other two algorithms and often detects three to four and sometimes even five peaks along certain fall streaks. However, due to the smoothing performed within PEAKO, the detected features are less noisy

25   and more consistent in time-and-height than for MicroARSCL.

~~Fig.~~ Figure 11 shows CFAD diagrams for each of the four algorithms for the case study on 02 February 2014, 16 to 17 UTC. These graphs confirm that Polyfit12 and Shupe_04 estimate the number of peaks more conservatively than PEAKO and MicroARSCL. It is obvious that MicroARSCL often detects three or four peaks in the lowermost radar height bins, which can probably be attributed to turbulence. The HSRL-detected SLW layer (varying between 0.7 to 1.5 km height) is most obvious

30   in the Shupe_04 and PEAKO CFAD plot, the number of spectra which are assigned two or three peaks is noticeably higher in this altitude range.

~~Fig.~~ A closer look at the Doppler spectra substantiates the occurence of riming during the selected study period: Figure 12 shows four exemplary Doppler spectra from the case study on 02 February 2014, 16 to 17 UTC. Only PEAKO is able to detect a narrow peak near $0\,m\,s^{-1}$ in all four example spectra. While these peaks can be attributed to SLW in ~~Fig.~~ Figure 12a,

35   ~~Fig.~~ Figure 12b, and ~~Fig.~~ Figure 12c, it is more likely that this small subpeak in ~~Fig.~~ Figure 12d is caused by small ice ice
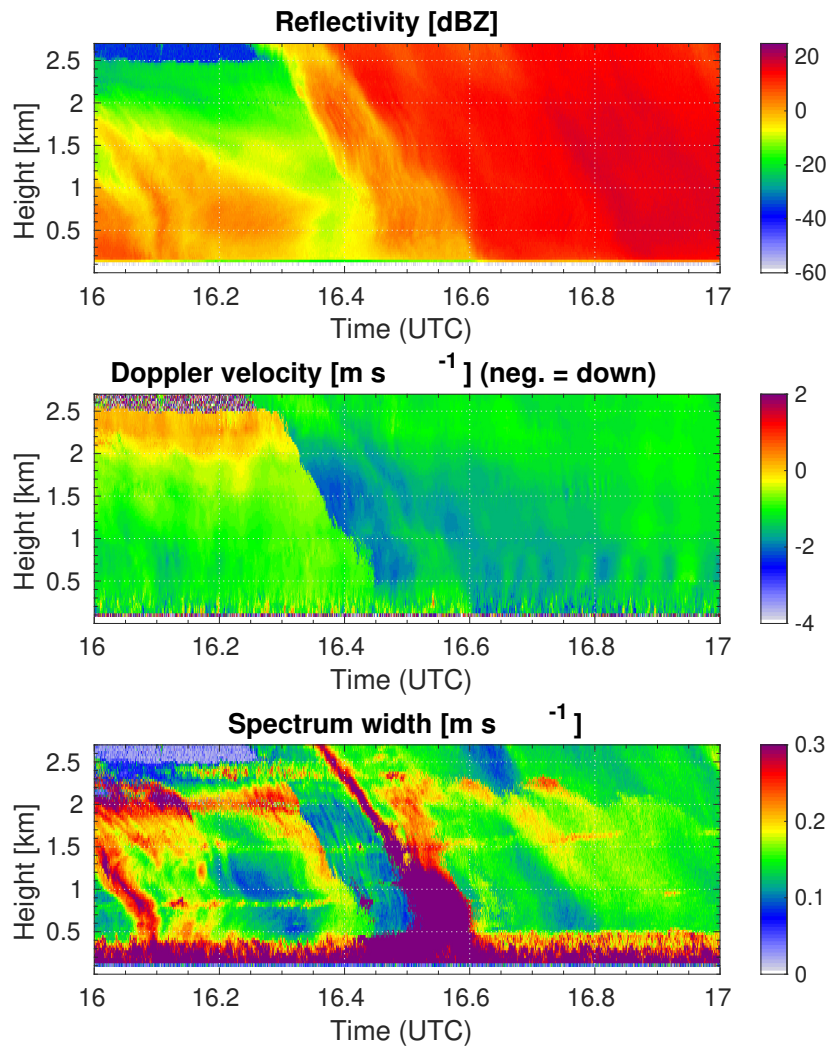
18

Figure 9. Like Fig. 4 but for 02 February 2014, 16.00 to 17.00 UTC in 0 to 2.7 km height.
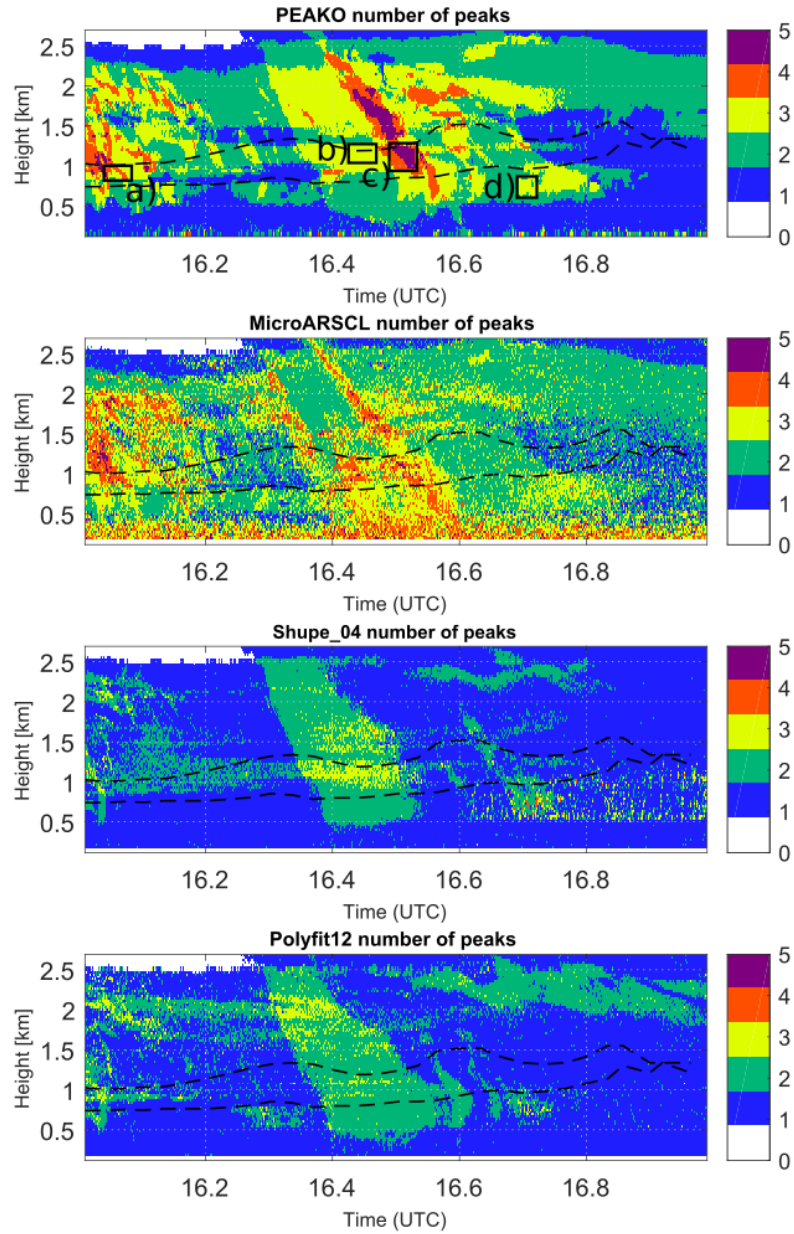
Figure 10. Like Fig. 5 but for 02 February 2014, 16 UTC to 17 UTC in 0 to 2.7 km height. Boxes mark the points in time-height where peaks detected in single spectra are analyzed in greater detail. The dashed lines mark the bottom and top of the SLW layer detected by the HSRL (Fig. 8)
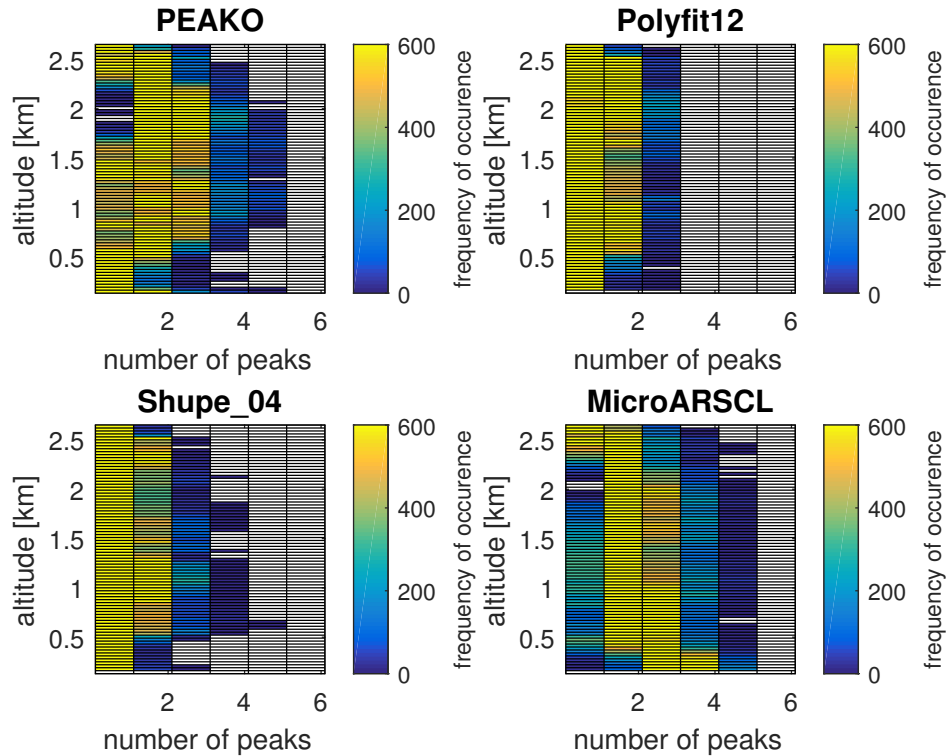
.

20

Figure 11. Like Fig. 6 but for the case study period on 02 February 2014, 16.00 to 17.00 UTC in 0 to 2.7 km altitude.

particles nucleated in the SLW layer situated slightly above because the HSRL does not detect liquid at 0.7 km altitude around 16.7 UTC. Besides the peak near $0\,\mathrm{m\,s^{-1}}$, all shown spectra are characterized by broad merged snow peaks pointing to snow particles of different size, shape, and density falling at different terminal velocities. In ~~Fig.~~Figure 12a, the three merged modes of snow, as well as the SLW peak are detected by PEAKO and MicroARSCL, while Shupe_04 and Polyfit12 both only detect one maximum. The spectrum shown in ~~Fig.~~Figure 12b is near the top of the SLW layer detected by the HSRL. The narrow liquid peak with fall velocity near $0\,\mathrm{m\,s^{-1}}$ is only detected by PEAKO and Shupe_04. Both algorithms find two more snow peaks with larger fall velocity. These two peaks are also detected by Polyfit12. MicroARSCL detects three peaks as well, however is not able to detect the liquid peak. ~~Fig.~~Figure 12c shows a spectrum which was chosen in an area where PEAKO finds five peaks. Again, one of them is a SLW peak within the SLW layer detected by the HSRL (Fig. 10). This peak is also detected by MicroARSCL and Shupe_04 but not Polyfit12. The four other peaks found by PEAKO are merged snow peaks with different fall velocities which hint to various degrees of riming that the other algorithms have difficulties to detect. In ~~Fig.~~Figure 12d, the number of peaks detected by the four algorithms differ significantly: The peak with highest reflectivity at around $-1.5\,\mathrm{m\,s^{-1}}$ fall velocity is found by all algorithms. PEAKO detects two sub-peaks, which are each ~~detected~~ identified by at least one other algorithm as well. However, none of the other methods finds both other ice sub-peaks.
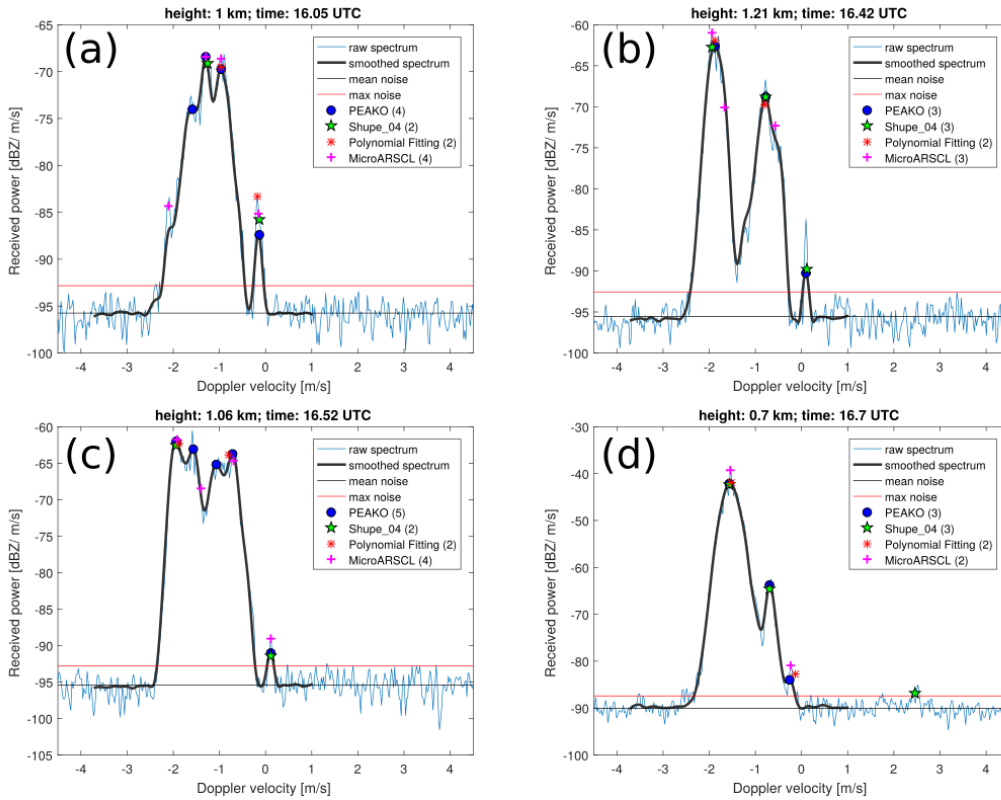
Figure 12. Four exemplary Doppler spectra picked from the case study on 02 February 2014, 16 to 17 UTC in 0 to 2.5 km altitude. The averaged and smoothed spectrum which is used as input to PEAKO is drawn in bold over the original spectrum. The peaks detected by the four algorithms are marked. The number of peaks found by each algorithm is noted in parenthesis in each figure legend. Mean and maximum noise floor are presented by black and red horizontal lines, respectively. Please note that the scale of the y axis is different in each plot.

In the Appendices A-C, three more case studies from the training and test phase are presented. Comparative results of PEAKO to the other peak finding algorithms are similar to the cases presented here. Appendix D contains a sensitivity study on the effect of different smoothing schemes and spatiotemporal averaging scales.

5 Conclusions and Outlook

5.1 Summary of findings and outlook

The presented study focuses on the description of a new supervised cloud radar Doppler velocity spectrum peak-finding algorithm (PEAKO). Its performance was compared to different existing Doppler spectrum peak-finding algorithms. It was found that the PEAKO algorithm generally agrees well with results from Shupe_04 and a polynomial fitting approach. PEAKO is however capable to detect narrower merged peaks with a smaller power contribution than Shupe_04. The polynomial fitting

approach has mostly similar results as Shupe_04 but is not very practical due to its long computation time. The MicroARSCL product was usually more sensitive to small perturbations in the radar Doppler spectrum and thus often detected a higher number of peaks than the other three algorithms and produces more "speckled" results. Some areas where peaks are overestimated by MicroARSCL are in high-turbulent regions with large spectrum width like the turbulent boundary layer while others seem

5 more random and not consistent in time and height. Consistency in time and to a lesser extent height is a good indicator of the performance of a peak-finding algorithm because hydrometeor populations and cloud microphysical processes generally occur in layers (unless in high turbulent regions). The number of found cloud radar Doppler velocity spectrum peaks within mixed-phase wintertime snow clouds in Finland ~~where~~ were validated with independent ground-based ~~in-situ~~ in situ observations described in Kneifel et al. (2015) and if available HSRL observations.

10 The described approach only identifies underlying hydrometeor populations if the particle types differ sufficiently in their terminal fall velocities to produce individual Doppler spectrum peaks. In upcoming projects, it is planned to test if the found best three-parameter pairs of PEAKO ~~can easily be applied~~ are applicable to other radar systems (like METEK-MIRA ~~35GHz~~ 35 GHz radars or RPG ~~94GHz~~ 94 GHz FMCW radars) or to which extent further refinement is needed for different radar sampling parameters. Additionally, the effect of stronger cloud dynamics will be evaluated.

15 Determining the *number* of different hydrometeor populations in the same radar volume based on morphological features of the radar Doppler spectrum as presented in this comparative study is the first step towards cloud particle classifications ~~which is however not the focus of this paper.~~ Having this easily-adjustable cloud radar Doppler spectrum peak detection algorithm available will facilitate carrying out microphysical process studies, involving applications such as peak tracking.

Appendix A: Case study from 16 February 2014, 0.67 - 0.92 UTC (training data set 2)

The results of the PEAKO comparison to the other three peak-finding approaches for the training data set of 16 February 2014, 0.67 to 0.92 UTC is shown in Appendix. This time period is also analyzed by Kneifel et al. (2015) in depth, so it was possible to compare the microphysical signatures reported in this study to the Doppler peaks detected by the four algorithms. For 16 February 2014, 0.67 to 0.92 UTC, Kneifel et al. (2015) reported high values of microwave radiometer derived liquid water path of 100 to 500 g m^{-2} and clear signatures of large aggregates in the dual-wavelength ratios of Ka-W-band below 2 km as well as in the reflectivity fall streak feature at 0.85 UTC which were also seen in the X-Ka dual-wavelength ratios around 0.85 UTC. The general structure of the layers of Doppler peak number detected by PEAKO, Shupe_04 and Polyfit12 again agree to a large extent, whereas MicroARSCL detects a higher number of peaks in both cases (Fig. A2 and B2). The fall streak feature which exhibits particle size sorting was better detected by PEAKO and MicroARSCL than by the other two algorithms. The increase in number of peaks at 0.67 to 0.75 UTC can be explained by the presence of large needle aggregates of sometimes more compact and sometimes very open structure as explained in (Kneifel et al., 2015) which lead to the interesting multi-modal Doppler spectrum with up to four peaks as shown in Fig. B4d). Ground-based ~~in-situ~~ in situ observations show that during 0.75 - 0.85 UTC aggregates and rimed particles with enhanced terminal velocities were present and that the number of large aggregates was further found to decrease while number of increasingly rimed aggregates further increased until 1 UTC. Radio sounding observations on 15 February 2014 at 23.2 UTC show a thin layer at 0.8-0.9 km altitude which is subsaturated with respect to ice and liquid and which might explain the decrease to 1 found Doppler peak at this altitude.

Appendix B: Case study from 21 February 2014, 23.01 - 23.10 UTC (training data set 3)

The training data set of 21 February 2014, 23.01 to 23.25 UTC which is also a case study of Kneifel et al. (2015) is described in Appendix B. For 23.01 to 23.25 UTC, Kneifel et al. (2015) report the transition from a low concentration of strongly rimed particles (lump graupel) to aggregate snowfall with large snowfall rates and increasing size and number of the aggregates. The fast transition of the snowfall from rimed particles to aggregates results in the bimodal Doppler spectra (with two found peaks) at 23 to 23.05 UTC and monomodal spectra afterwards. For this case study, PEAKO and Shupe_04 and Polyfit12 agree well with the situation described in Kneifel et al. (2015) while MicroARSCL overestimates the number of peaks especially in the turbulent boundary layer and near 4 km altitude.

Appendix C: Case study from 7 February 2014, 23.75 - 24 UTC (test data set 2)

The second test data set of February 7, 2014 23.75 to 24 UTC is characterized by dendritic ice particles (Kneifel et al., 2015) and a slanted fall streak feature extending from near 4 km to 1 km from 23.75 to 23.9 UTC (Fig. C1) with bimodal Doppler spectra (Fig. C2 and Fig. C4). Ground-based ~~in-situ~~ in situ observations report mostly small, open-structured aggregates (which are later replaced by more compact spheroidal habits) as well as a small number of spherical probably rimed particles.
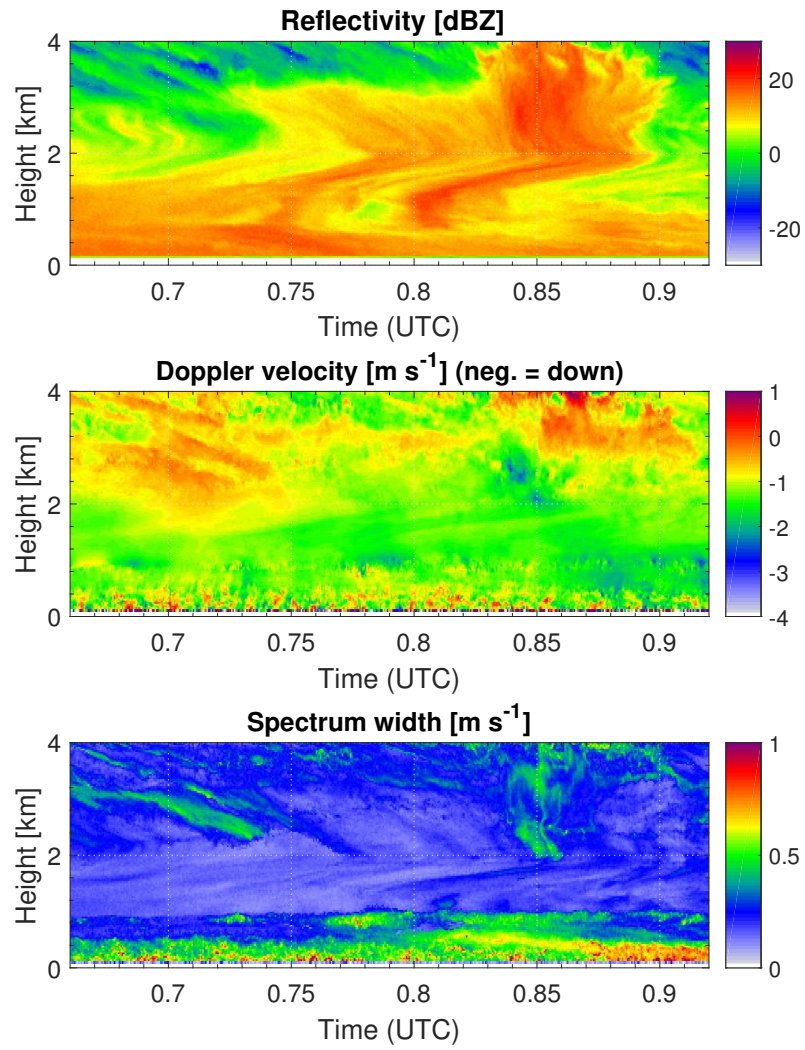
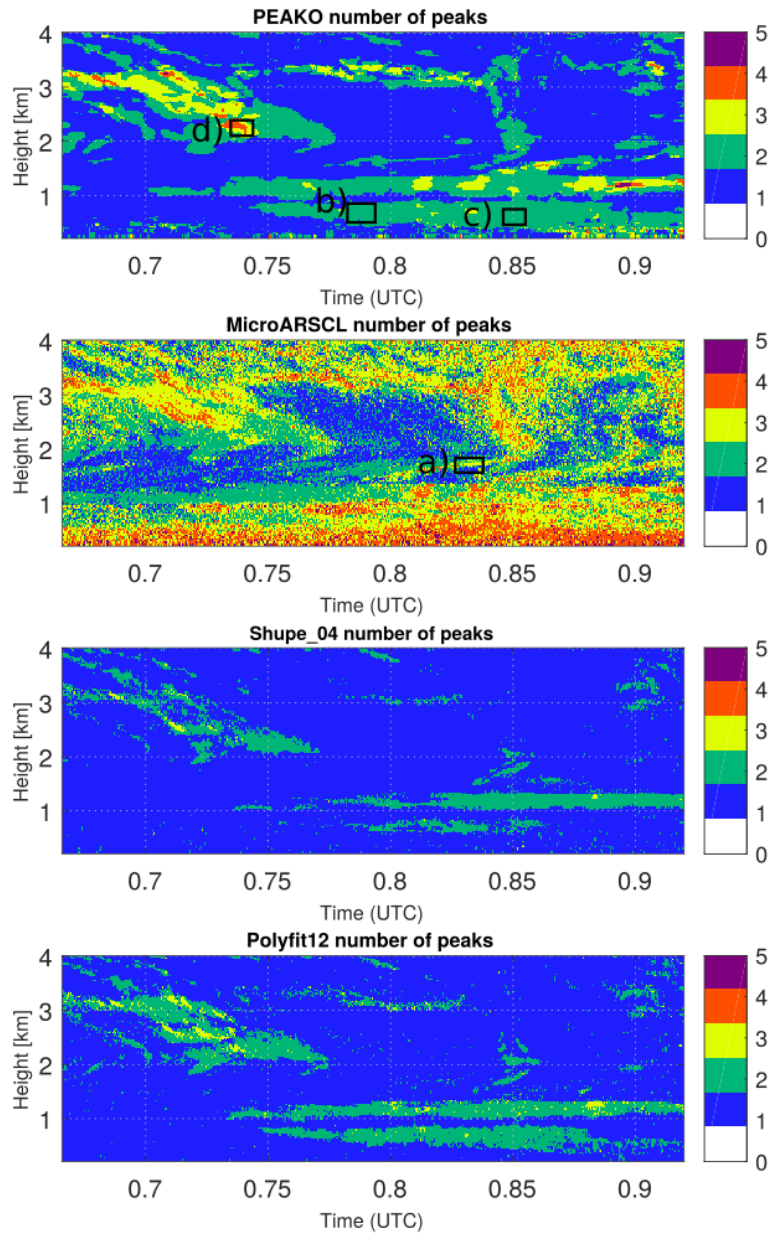**Figure A1.** Like Fig. 4 but for 16 February 2014, 0.67 - 0.92 UTC in 0 - 4 km height.

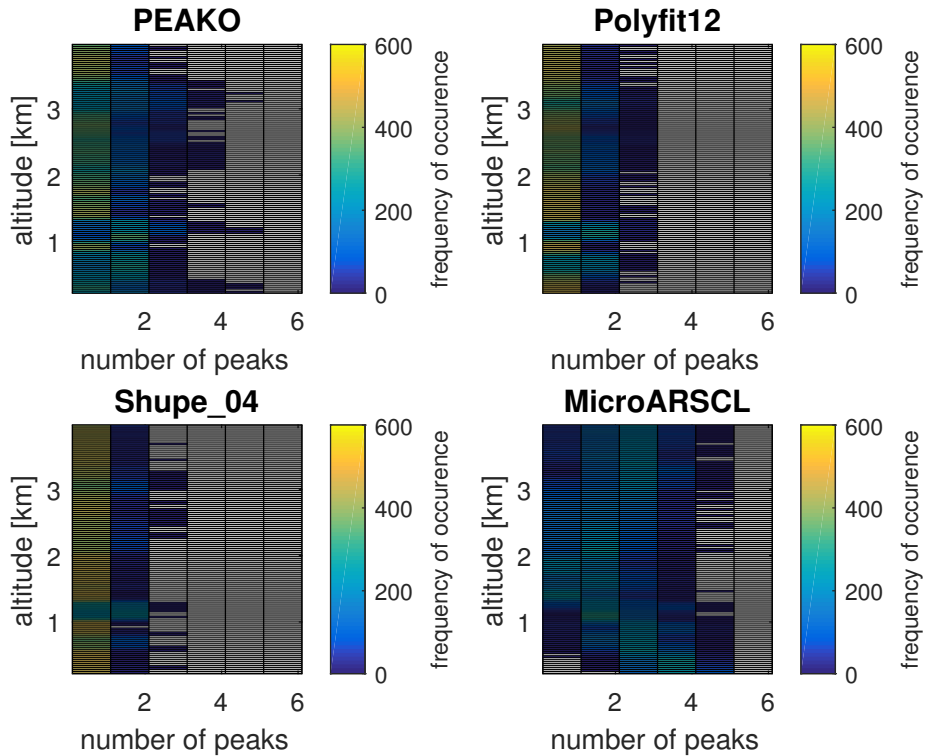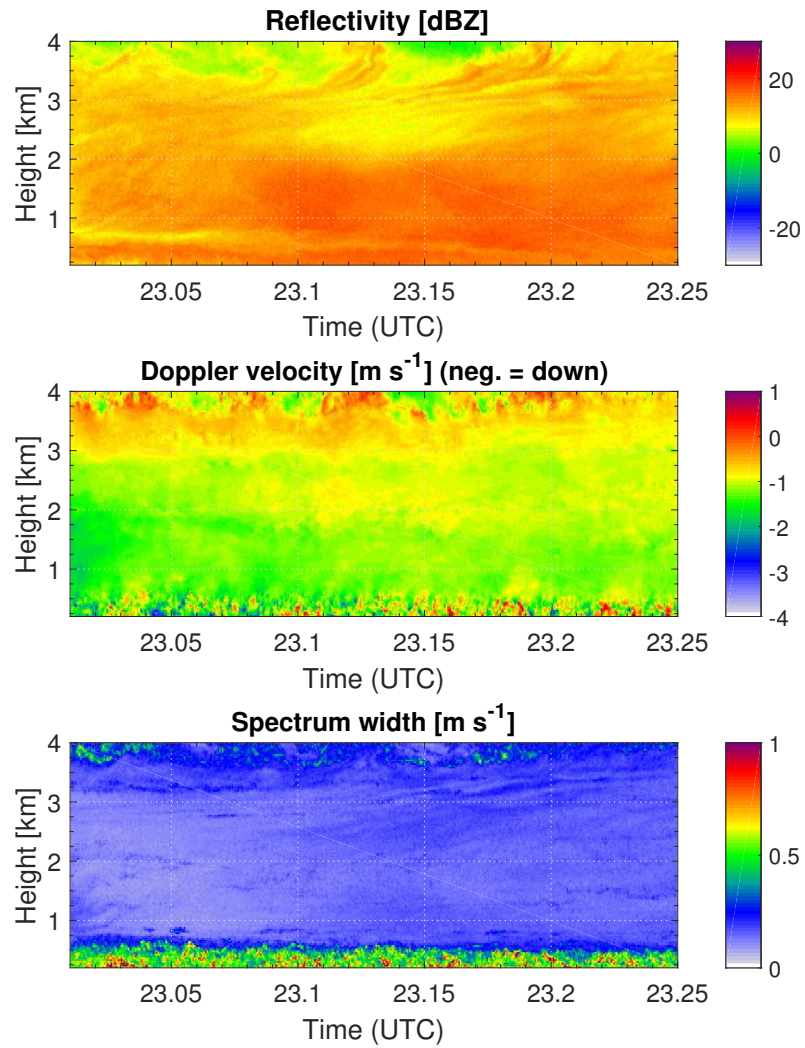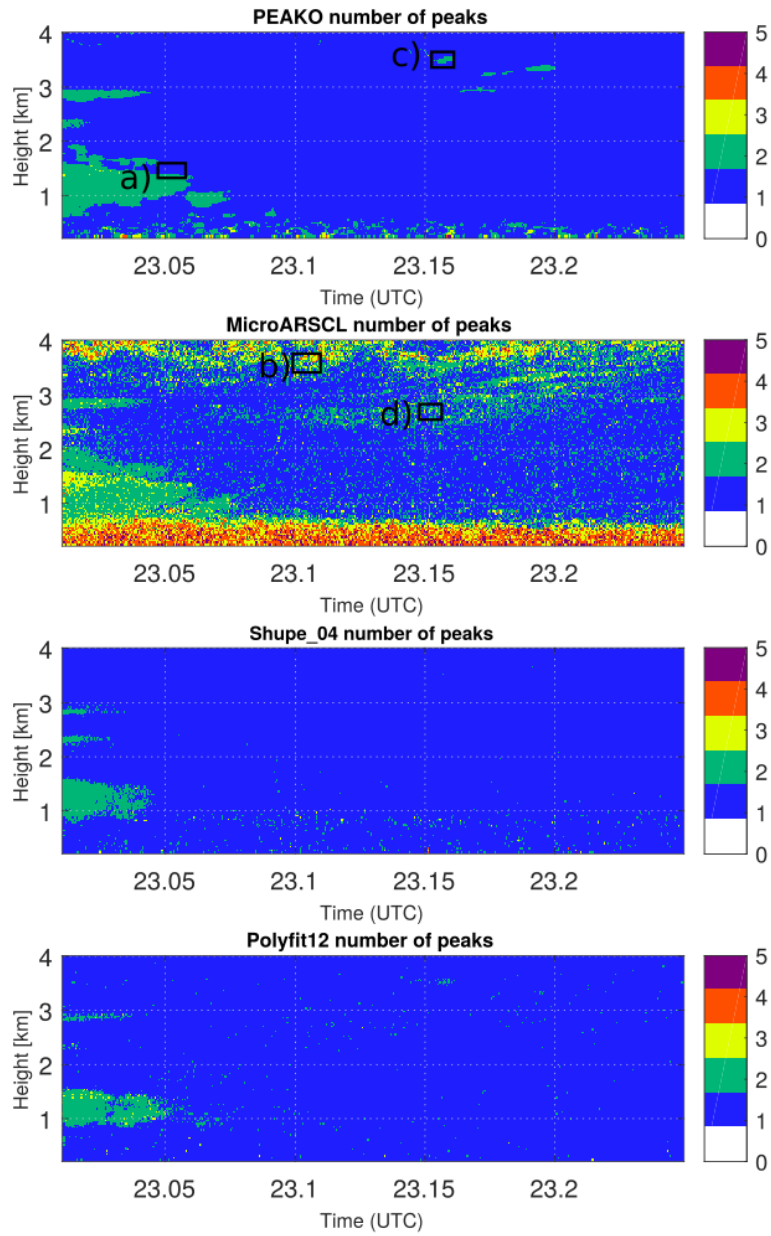**Figure A2.** Like Fig. 5 but for 16 February 2014, 0.67 - 0.92 UTC in 0 - 4 km height.

Figure A3. Like Fig. 6 but for the case study period on 16 February 2014, 0.66 - 0.92 UTC in 0 - 4 km altitude.

Appendix D: **Sensitivity study on the effect of different smoothing schemes and spatiotemporal averaging scales**

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 the 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
5 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.

10 – 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
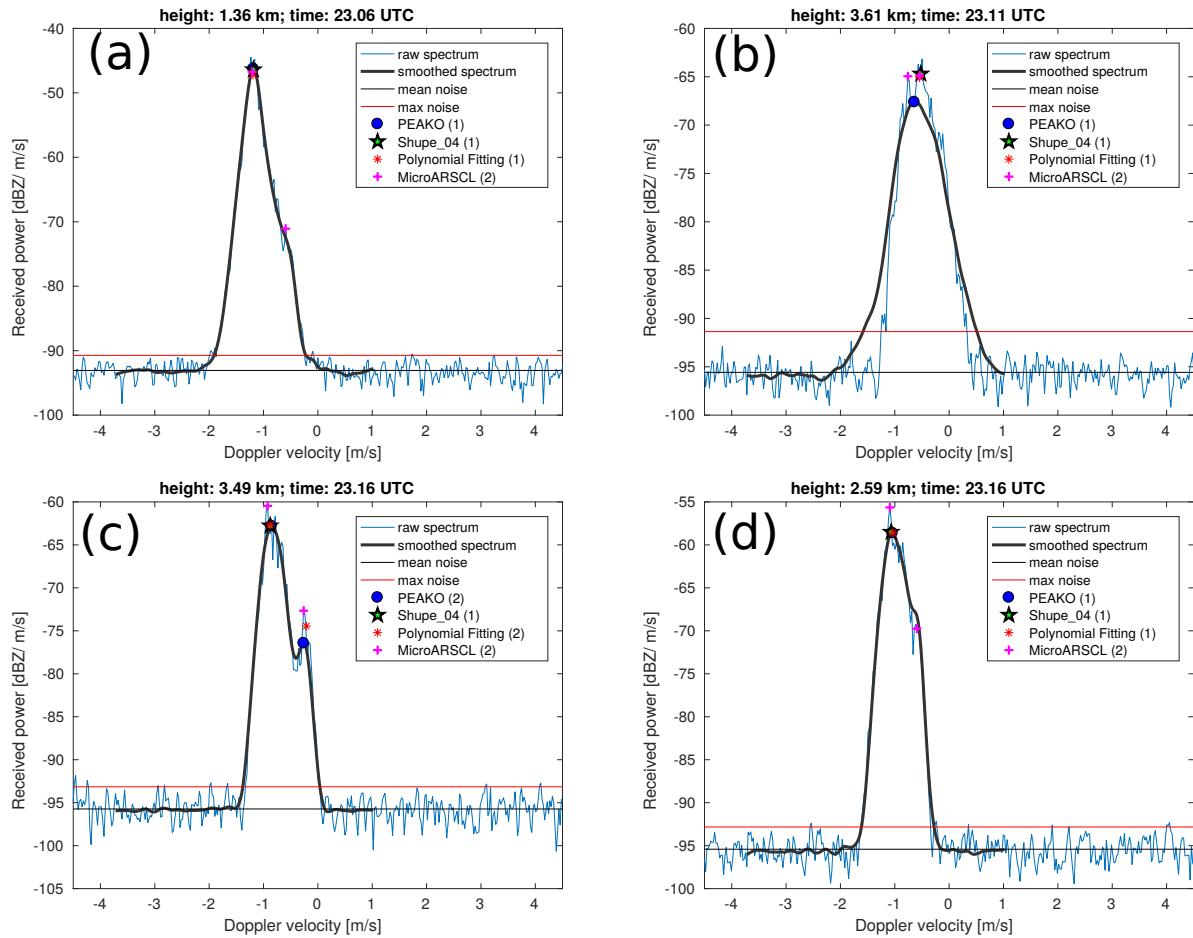
27

Figure A4. Four example spectra selected from the case study on 16 February 2014, 0.63 - 0.92 UTC. Please note that the y axis scale is different for each of the spectrum plots.

The optimized parameters obtained after training PEAKO in each of above listed configurations were applied to ~~the same case which was already~~ presented in Fig. 5. Figure D1 shows the ~~number of peaks found for the original PEAKO settings and for each of the four PEAKO settings.~~ The five ~~plots~~ 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

Reflectivity [dBZ]

## Doppler velocity [m s^{-1}] (neg. = down)

Spectrum width [m s^{-1}]

Figure B1. Like Fig. 4 but for 21 February 2014, 23.01 - 23.25 UTC in 0.2 - 4 km height.

Figure B2. Like Fig. 5 but for 21 February 2014, 23.01 - 23.25 UTC in 0.2 - 4 km height.
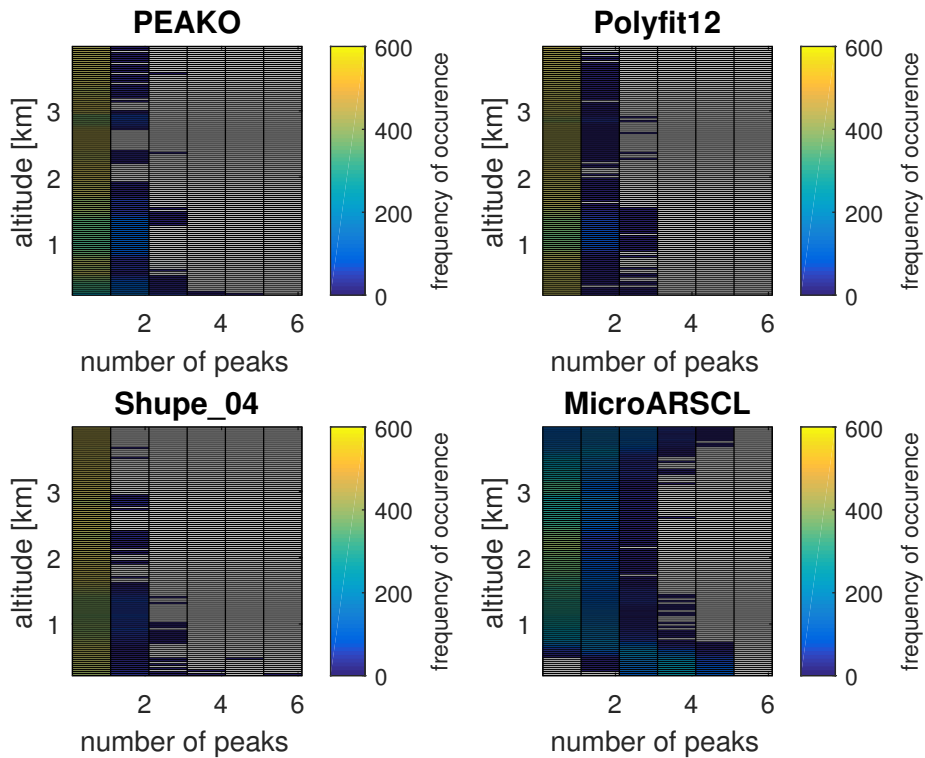
Figure B3. Like Fig. 6 but for the case study period on 21 February 2014, 23.01 - 23.25 UTC in 0.2 - 4 km altitude.

and lowess methods are able to reproduce the features detected by PEAKO in the default configuration only with some minor deviations.

5

31

Figure B4. Three example spectra selected from the case study on 21 February 2014, 23.01 - 23.25 UTC. Please note that the y axis scale is different for each of the spectrum plots.
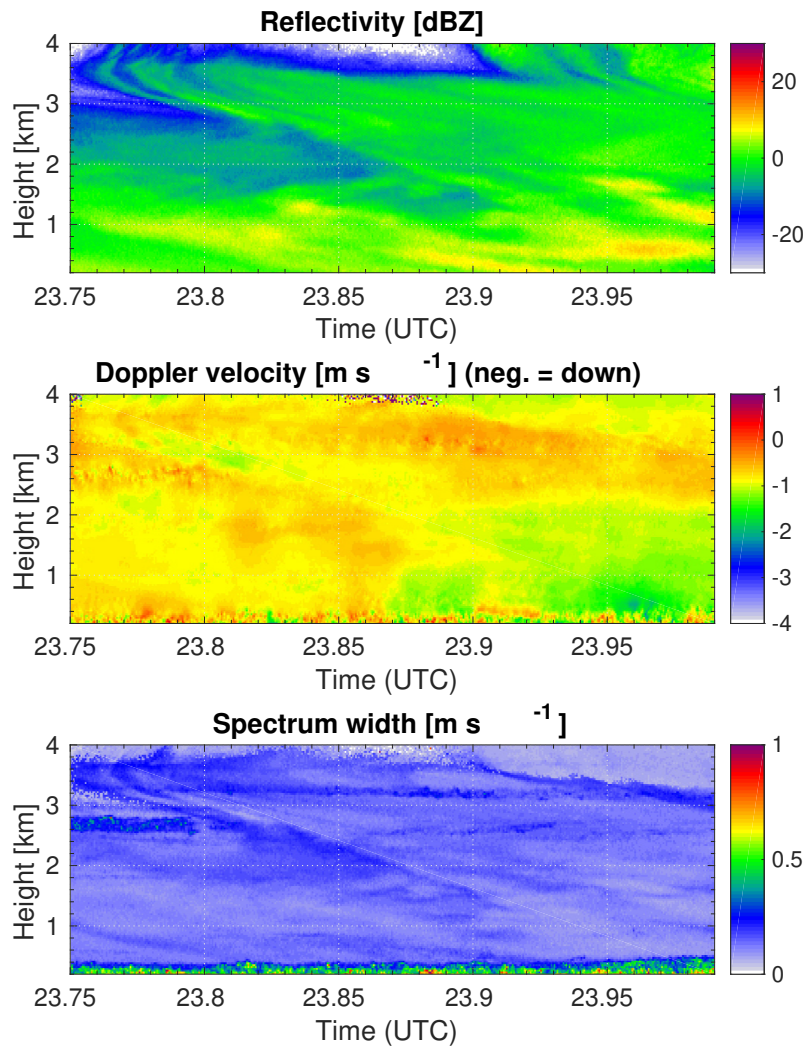
Figure C1. Like Fig. 4 but for 7 February 2014, 23.75 - 24 UTC in 0.2 - 4 km height.
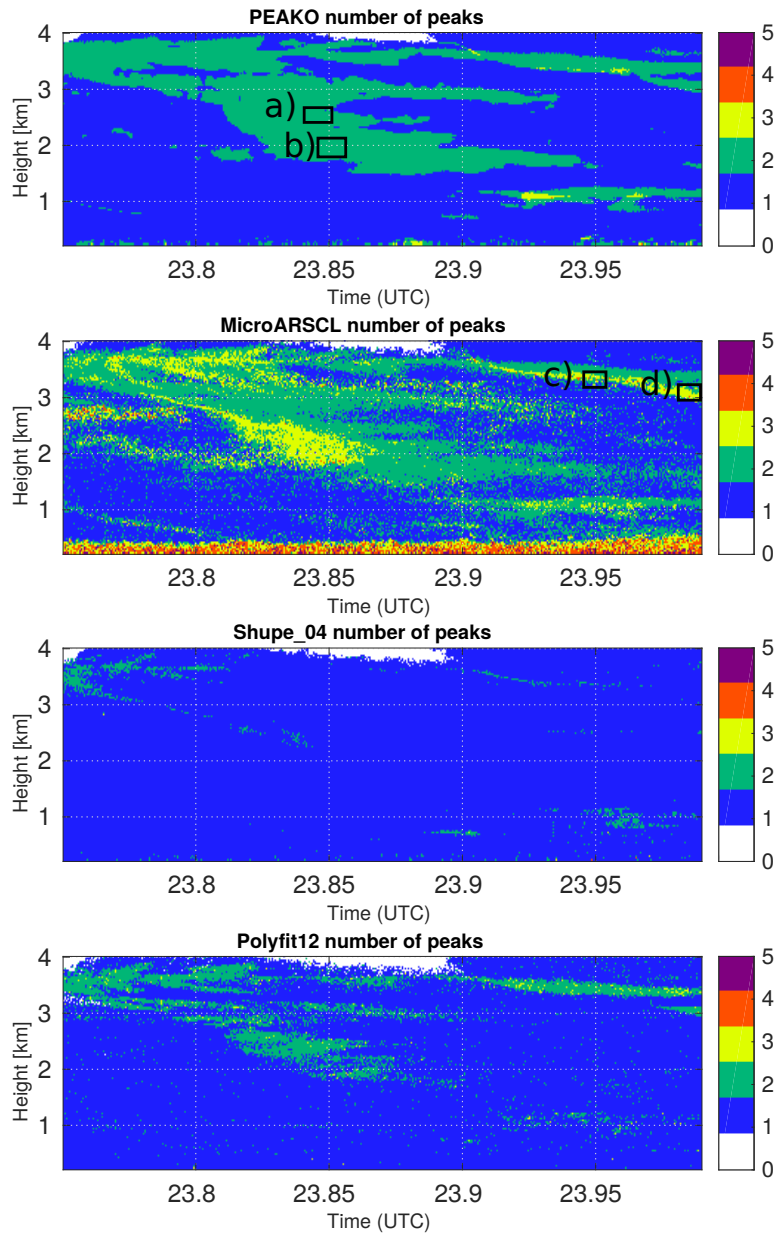
Figure C2. Like Fig. 5 but for 7 February 2014, 23.75 - 24 UTC in 0.2 - 4 km height.
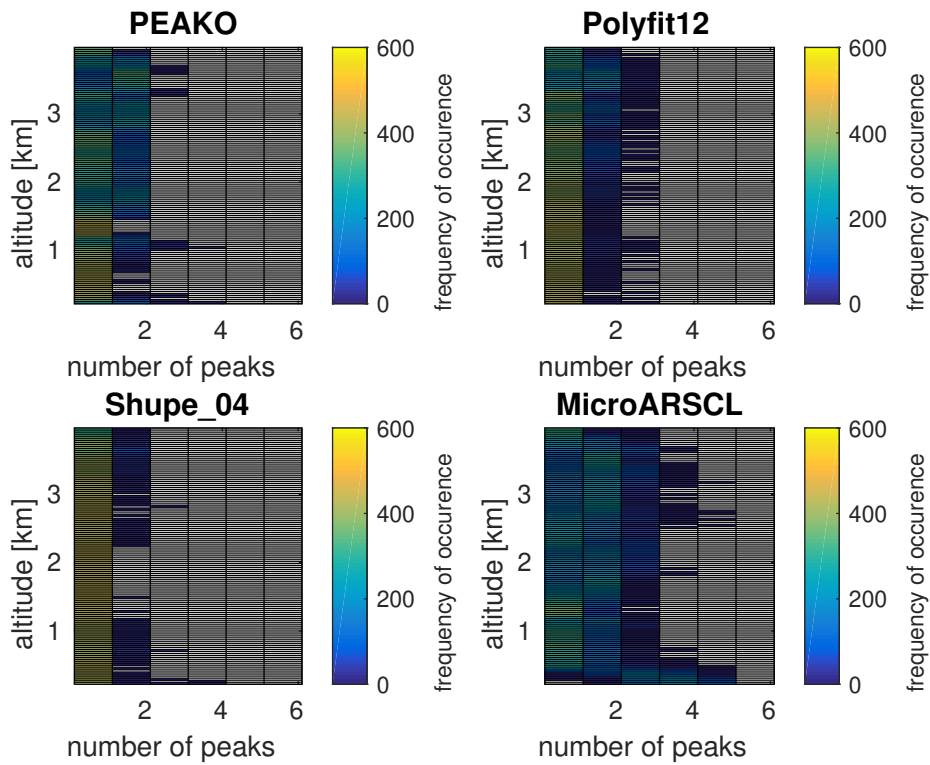
Figure C3. Like Fig. 6 but for the case study period on 7 February 2014, 23.75 - 24 UTC in 0.2 - 4 km altitude.
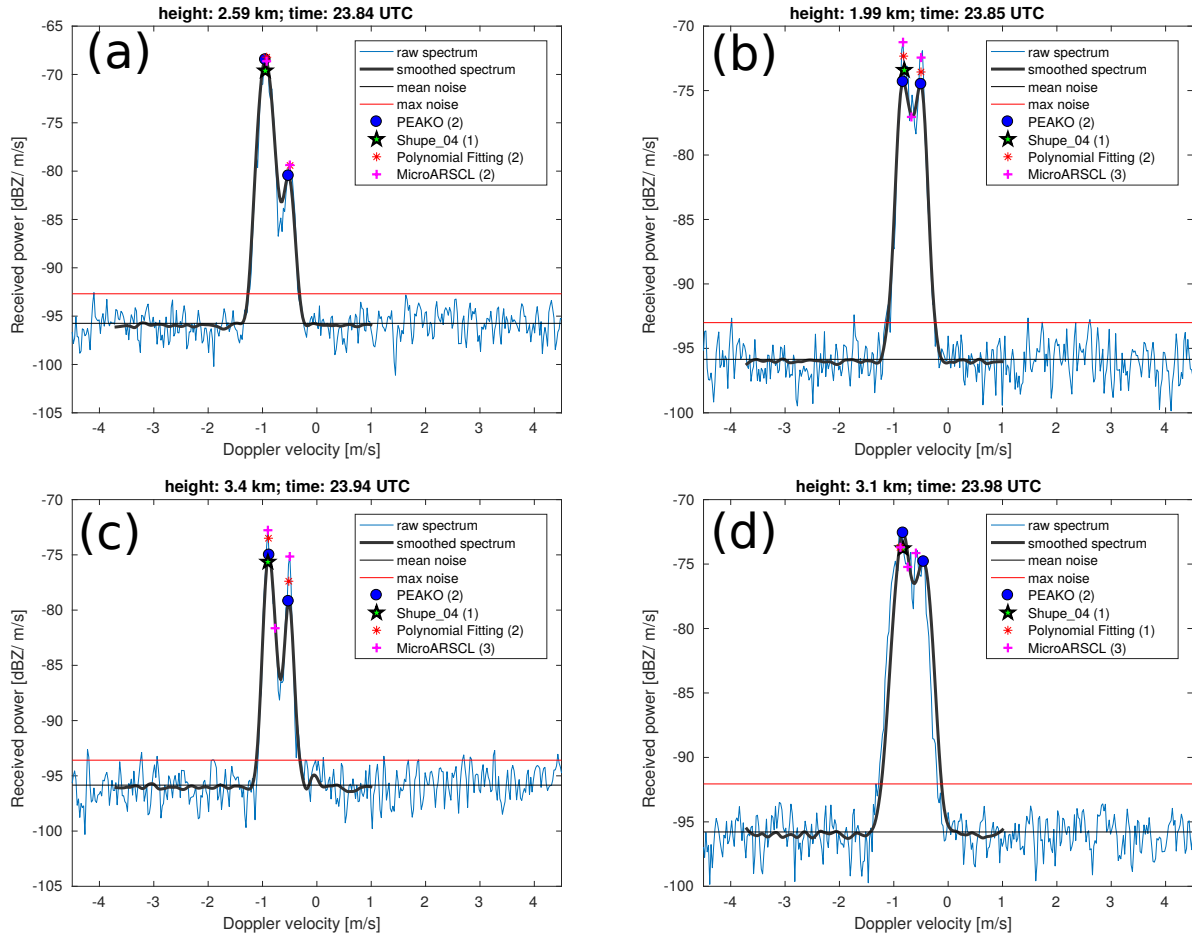
Figure C4. Four example spectra picked from the case study on 7 February 2014, 23.75 - 24 UTC. Please note that the y axis scale is different for each of the spectrum plots.
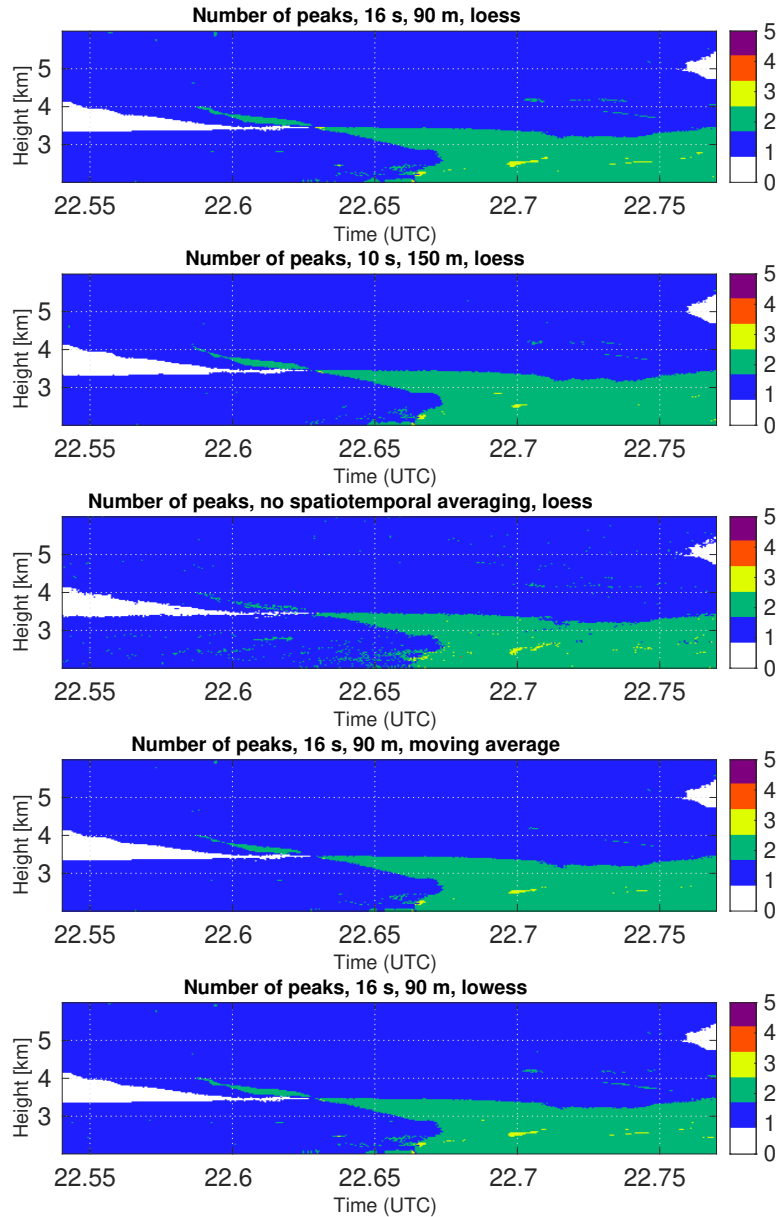
Figure D1. Number of Doppler spectrum peaks detected by PEAKO in five different configurations for the selected case study on 2014-02-21 from 22.54 to 22.77 UTC in 2 to 6 km altitude. Top to bottom: Number of peaks detected by PEAKO in the default configuration (16 s temporal and 90 m spatial averaging prior to loess smoothing), this plot is equivalent to the top panel in Fig. 5; number of peaks detected using 10 s 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

References

Besic, N., i Ventura, J. F., Grazioli, J., Gabella, M., Germann, U., and Berne, A.: Hydrometeor classification through statistical clustering of polarimetric radar measurements: a semi-supervised approach, Atmos. Meas. Tech., 9, 4425–4445, https://doi.org/10.5194/amt-9-4425-2016, 2016.

5 Bühl, J., Seifert, P., Myagkov, A., and Ansmann, A.: Measuring ice- and liquid-water properties in mixed-phase cloud layers at the Leipzig Cloudnet station, Atmospheric Chemistry and Physics, 16, 10 609–10 620, https://doi.org/10.5194/acp-16-10609-2016, 2016.

Cornman, L. B., Goodrich, R. K., Morse, C. S., and Ecklund, W. L.: A Fuzzy Logic Method for Improved Moment Estimation from Doppler Spectra, Journal of Atmospheric and Oceanic Technology, 15, 1287–1305, https://doi.org/10.1175/1520-0426(1998)015<1287:AFLMFI>2.0.CO;2, 1998.

10 Hildebrand, P. H. and Sekhon, R. S.: Objective Determination of the Noise Level in Doppler Spectra, J. Appl. Meteor., 13, 808–811, https://doi.org/10.1175/1520-0450(1974)013<0808:odotnl>2.0.co;2, 1974.

Kalesse, H., Szyrmer, W., Kneifel, S., Kollias, P., and Luke, E.: Fingerprints of a riming event on cloud radar Doppler spectra: observations and modeling, Atmos. Chem. Phys., 16, 2997–3012, https://doi.org/10.5194/acp-16-2997-2016, 2016.

Kneifel, S., von Lerber, A., Tiira, J., Moisseev, D., Kollias, P., and Leinonen, J.: Observed relations between snowfall microphysics and
15 triple-frequency radar measurements, J. Geophys. Res. Atmos., 120, 6034–6055, https://doi.org/10.1002/2015jd023156, 2015.

Kollias, P., Lhermitte, R., and Albrecht, B. A.: Vertical air motion and raindrop size distributions in convective systems using a 94 GHz radar, Geophys. Res. Lett., 26, 3109–3112, https://doi.org/10.1029/1999GL010838, 1999.

Kollias, P., Albrecht, B. A., and Marks Jr., F. D.: Cloud radar observations of vertical drafts and microphysics in convective rain, J. Geophys. Res. Atmos., 108, https://doi.org/10.1029/2001jd002033, 2003.

20 Kollias, P., Miller, M. A., Luke, E. P., Johnson, K. L., Clothiaux, E. E., Moran, K. P., Widener, K. B., and Albrecht, B. A.: The Atmospheric Radiation Measurement Program Cloud Profiling Radars: Second-Generation Sampling Strategies, Processing, and Cloud Data Products, J. Atmos. Ocean. Tech., 24, 1199–1214, https://doi.org/10.1175/jtech2033.1, 2007.

Kollias, P., Bharadwaj, N., Widener, K., Jo, I., and Johnson, K.: Scanning ARM Cloud Radars. Part I: Operational Sampling Strategies, J. Atmos. Ocean. Tech., 31, 569–582, https://doi.org/10.1175/jtech-d-13-00044.1, 2014.

25 Kollias, P., Clothiaux, E. E., Ackerman, T. P., Albrecht, B. A., Widener, K. B., Moran, K. P., Luke, E. P., Johnson, K. L., Bharadwaj, N., Mead, J. B., Miller, M. A., Verlinde, J., Marchand, R. T., and Mace, G. G.: Development and Applications of ARM Millimeter-Wavelength Cloud Radars, Meteor. Mon., 57, 17.1–17.19, https://doi.org/10.1175/amsmonographs-d-15-0037.1, 2016.

Komurcu, M., Storelvmo, T., Tan, I., Lohmann, U., Yun, Y., Penner, J. E., Wang, Y., Liu, X., and Takemura, T.: Intercomparison of the cloud water phase among global climate models, J. Geophys. Res. Atmos., 119, 3372–3400, https://doi.org/10.1002/2013jd021119, 2014.

30 Luke, E. P., Kollias, P., Johnson, K. L., and Clothiaux, E. E.: A Technique for the Automatic Detection of Insect Clutter in Cloud Radar Returns, J. Atmos. Ocean. Tech., 25, 1498–1513, https://doi.org/10.1175/2007JTECHA953.1, 2008.

Luke, E. P., Kollias, P., and Shupe, M. D.: Detection of supercooled liquid in mixed-phase clouds using radar Doppler spectra, J. Geophys. Res. Atmos., 115, https://doi.org/10.1029/2009JD012884, 2010.

Mason, S. L., Chiu, C. J., Hogan, R. J., Moisseev, D., and Kneifel, S.: Retrievals of Riming and Snow Density From Vertically Pointing
35 Doppler Radars, J. Geophys. Res. Atmos., https://doi.org/10.1029/2018jd028603, 2018.

Oue, M., Kollias, P., Ryzhkov, A., and Luke, E. P.: Toward Exploring the Synergy Between Cloud Radar Polarimetry and Doppler Spectral Analysis in Deep Cold Precipitating Systems in the Arctic, J. Geophys. Res. Atmos., 123, 2797–2815, https://doi.org/10.1002/2017jd027717, 2018.

Petäjä, T., O'Connor, E. J., Moisseev, D., Sinclair, V. A., Manninen, A. J., Väänänen, R., von Lerber, A., Thornton, J. A., Nicoll, K., Petersen, W., Chandrasekar, V., Smith, J. N., Winkler, P. M., Krüger, O., Hakola, H., Timonen, H., Brus, D., Laurila, T., Asmi, E., Riekkola, M.-L., Mona, L., Massoli, P., Engelmann, R., Komppula, M., Wang, J., Kuang, C., Bäck, J., Virtanen, A., Levula, J., Ritsche, M., and Hickmon, N.: BAECC: A Field Campaign to Elucidate the Impact of Biogenic Aerosols on Clouds and Climate, B. Am. Meteorol. Soc., 97, 1909–1928, https://doi.org/10.1175/bams-d-14-00199.1, 2016.

Praz, C., Roulet, Y.-A., and Berne, A.: Solid hydrometeor classification and riming degree estimation from pictures collected with a Multi-Angle Snowflake Camera, Atmos. Meas. Tech., 10, 1335–1357, https://doi.org/10.5194/amt-10-1335-2017, 2017.

Radenz, M., Bühl, J., Seifert, P., Griesche, H., and Engelmann, R.: peakTree: A framework for structure-preserving radar Doppler spectra analysis, Atmospheric Measurement Techniques Discussions, 2019, 1–18, https://doi.org/10.5194/amt-2019-76, 2019.

Riihimaki, L. D., Comstock, J. M., Anderson, K. K., Holmes, A., and Luke, E.: A path towards uncertainty assignment in an operational cloud-phase algorithm from ARM vertically pointing active sensors, Adv. Stat. Climatol. Meteorol. Oceanogr., 2, 49–62, https://doi.org/10.5194/ascmo-2-49-2016, 2016.

Shupe, M. D., Kollias, P., Matrosov, S. Y., and Schneider, T. L.: Deriving Mixed-Phase Cloud Properties from Doppler Radar Spectra, J. Atmos. Ocean. Tech., 21, 660–670, https://doi.org/10.1175/1520-0426(2004)021<0660:dmcpfd>2.0.co;2, 2004.

Tan, I., Storelvmo, T., and Zelinka, M. D.: Observational constraints on mixed-phase clouds imply higher climate sensitivity, Science, 352, 224–227, https://doi.org/10.1126/science.aad5300, 2016.

Verlinde, J., Rambukkange, M. P., Clothiaux, E. E., McFarquhar, G. M., and Elorant, a. E. W.: Arctic multilayered, mixed-phase cloud processes revealed in millimeter-wave cloud radar Doppler spectra, J. Geophys. Res. Atmos., 118, 13 199–13 213, https://doi.org/10.1002/2013JD020183, 2013.

Williams, C. R., Maahn, M., Hardin, J. C., and de Boer, G.: Clutter mitigation, multiple peaks, and high-order spectral moments in 35 GHz vertically pointing radar velocity spectra, Atmos. Meas. Tech., 11, 4963–4980, https://doi.org/10.5194/amt-11-4963-2018, 2018.