Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2019-48-RC1, 2019


Atmospheric Measurement Techniques Discussions

Open Access



# *Interactive comment on* "Development and validation of a supervised machine learning radar Doppler spectra peak finding algorithm" *by* Heike Kalesse et al.

**Anonymous Referee #1**

Received and published: 18 May 2019

The paper presents a new method for partitioning cloud-radar Doppler spectra into physically meaningful peaks. There have been many methods adopted over the years for identification of such peaks but these other methods were usually just a step towards an end and not the end itself. The new proposed method is at the heart of this paper. There are interesting ideas in this paper so it is worthy of publication in AMT. Although no show stoppers showed up in reading the paper, there are places where the manuscript can be improved and these are listed in the comments below.

Comments:

0) Mark up on the manuscript made while reading it is contained in the attachment.



Printer-friendly version

Discussion paper

Perhaps some of it will be of value to the authors.

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?

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.

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





Printer-friendly version

Discussion paper

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.

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.

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.

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.

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

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.

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.

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

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.

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

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.

Please also note the supplement to this comment:
https://www.atmos-meas-tech-discuss.net/amt-2019-48/amt-2019-48-RC1-supplement.pdf

---

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2019-48, 2019.