I really enjoyed reading the manuscript, that is presenting an innovative methodology to classify drizzle presence in the clouds using a machine learning algorithm. The presented results are robust and potentially open new future studies and analysis. I find the paper well written and I only have some small minor comments regarding the exposition of some scientific concepts. Despite the completeness of the exposition, I could find a couple of open questions regarding the machine learning algorithm and the physical interpretation of the classification, that I would like to pose to the authors, for possibly improving the manuscript further. Please find below the detailed list of my comments.
Best regards

Technical comments:

line 55: what is autoconversion? At this line the term is introduced without any previous definition. To make the reading easier, you might first define what is the autoconversion process. I mean, it does not exist in nature as a process… it is model representation of the droplet growth… maybe some reflections on this could help the reader to understand better the problem you are tackling.

Line 57-58: cloud organizational structure… do you mean cloud spatial organization? And the boundary layer system… is it the boundary layer structure?

Line 70: Van Zanten , not VanZanten, there's a space missing, I guess.

Line 99: To maybe integrate on what you wrote: The threshold chosen on the skewness is based on the estimations of the skewness variability in non-drizzling conditions derived in Acquistapace et al., 2017 (https://doi.org/10.5194/amt-10-1783-2017). In that work, in section 3.1.1 we explained how to obtain that threshold: "Standard deviations of the skewness time series in the nondrizzling cloud using 2s integration time and spectral resolutions of 256, 512 and 1024 range between 0:389 and 0:369 with a mean value over the three cases of 0:379." Maybe this helps to make the threshold choice clearer.

Line 170: I don't understand how the sigma-t is derived. Which retrieval do you apply? and why you do not use a set of sigma t values instead of just one? I am asking because turbulence can strongly alter the spectra shape and it can be used to test your algorithm using values in the tails of your distribution.

Comment relative to figure 3.C:

From your classification, drizzle seems to occur anywhere in the cloud. Are you able to interpret this evidence in a physical way? How is this evidence for example correlated with turbulence in the cloud? Is there any statistical correlation between more turbulent bins and machine learning classified pixels? A possible approach to test this would be to derive EDR using Borque et al, 2016 approach for example.

General questions:

1) Why did you pick the combination of parameter Gamma=50 C=1, giving 0.95 for precision and 0.85 for recall? Many other combinations of the parameters give the same outcome or even better ones (for example, from the table, I could spot gamma = 1, C = 1000). How do the other choices affect the results? Can you justify more why you pick your choice? It would help to better understand the role of gamma and C to have a plot showing the variability of the black line with respect to the choices of gamma, and C, for maybe some sets.

2) how does the reflectivity profile of the clouds whose radar bins fall in the region between -30 and -20 dBz (that you highlight in the nice figure 2) look like? In our past work, we used the cloud adiabaticity (slope of the gradient of Ze in the profile) to identify clouds in which the embryonic drizzle onset was starting but no skewness signature was already evident. I would be interested to know if there's any correspondence. This point comes back to the physical interpretation of what's happening when drizzle develops, and in which conditions. The idea was that when drops grow with diffusion of water vapor they follow an adiabatic profile, and when other interactions take over, the profile ceases to be adiabatic.