Overall

The presented manuscript develops a method of using REFIR-PAD spectroradiometer data to identify and track cloud properties over the Concordia station. Ancillary instrumentation is used to train a machine-learning algorithm to be applied to REFIR-PAD data. Three years of data are then used to track cloud properties and ultimately report on cloud statistics over Concordia. In principle, the goal of presenting cloud occurrence statistics seems both reasonable and achievable given the availability of data and methods presented. Assuming the trained data and classification scheme are correct, I see no reason to doubt the presented cloud statistic data. Furthermore, this data comes from a very data sparse part of the world and such information would be very beneficial to the community.

However, crucially, the data set used to train the algorithm must be above reproach. It is here that major concerns arise for me as a community member as I believe there are major deficiencies in the treatment of the training data. If the trained data or training method is to be doubted, the rest of the scientific value of this manuscript is degraded substantially. This is especially true given the above statement that the data come from a very sparsely sampled part of the world that would be potentially heavily relied upon to be correct and accurate.

It is my opinion that the presented manuscript contains some fairly fundamental deficiencies that need to be addressed before it should be considered by the editor for publication.

Specific Major Comments

1. While I completely recognize the paper presented does not focus on lidar, the authors seem to heavily rely on a lidar instrument, which is poorly described. It seems to me that the lidar system cannot remain as transparent as it is presented here because the reader does need to be able to evaluate the quality of the training data set. The main reference given is Palchetti et al. 2015, which is a BAMS article that seems to lack technical detail of the instrument. The Palchetti et al. 2015 paper further references a website for lidar data that seems to be defunct (at least I can't get to it on any of the computers I have tried). I am left wondering some very fundamental things about the construction of the lidar system that heavily influence its data quality. Some of the major ones (this is by no means an exhaustive list) are:
    a. Is the system coaxial or biaxial? This will affect the height range of detectable signal as well as the observed signal strength.
    b. What is the signal detection system and expected dynamic range? Does the system use photon counting or analog signals? Is the detector a photo-multiplier tube (PMT) or avalanche-photo diode (APD) or something else entirely? This affects both the height of the observable signal as well as the apparent oscillatory depolarization structure from Figure 2. For example the claim in the Palchetti et al. 2015 paper that the range is from 30-7000 m would require a

minimum signal dynamic range of 4.5 orders of magnitude (assuming a completely uniform scene). That is a tough ask even for systems that employ both analog and photon counting techniques, which introduce complexity in combining the two.

   c. What is the system field of view? This will directly affect depolarization measurements via multiple scattering.
   d. What is the laser system's divergence? Is it matched to the field of view? This combines with the above primarily relating in my mind to the possibility of observing multiple scattering.
   e. How is the system's depolarization sensitivity calibrated? Systematic effects such as internal depolarization, diattenuation, and retardance can affect all of this. For details here see for example: Biele et al. 2000, Alvarez et al. 2006, Hayman and Thayer 2009 or 2012, or Freudenthaler 2016.
   f. Do the authors use any sort of algorithm to make the backscattered signal threshold a quantitative and repeatable measure? Klett or Frenald inversions are 2 examples, which admittedly have a number of limiting assumptions required. However, it is my understanding that the authors are inspecting lidar data signal strength directly, which is neither quantitative nor repeatable. Furthermore, signal strength is complicated by alignment issues, long term degradation of optical components, atmospheric structure, and system dynamic range and design.

2. It appears that the authors are using a non-quantitative method to identify clouds using lidar data. They say on Line 213-215 that cloud identification is done by visual inspection of backscatter and depolarization profiles. If my understanding of this process is true, that is completely non-repeatable and lacks any metric whereby a reviewer or reader can either replicate or even compare results. If there is a more quantitative method to identifying clouds than what I have just described, it needs to be much more clearly stated. If this is the method, it should not really be considered quantitative at all, which undercuts the lidar data used as a standard to train the machine learning cloud identification code.
3. The authors seem to have created the following simple table to classify clouds via lidar data, which is then used to verify spectral classifications.

Low relative signal = Clear air	High Signal High Depol (d>0.15) = Ice
	High Signal Low Depol (d<0.15) = Liquid or mixed phase

Given the lack of overall description of the lidar instrument, it is not possible to evaluate if these value are reasonable. For example, the threshold between columns seems arbitrary. Furthermore, this classification scheme is very simplified (in comparison to for example Shupe 2007 or Nott and Duck 2011) and will miss a lot of instrument related effects such as:

   a. Multiple scattering induced increase in depolarization with range
   b. Long term calibration drifts of polarization parameters
   c. Basic error propagation, e.g. is a depolarization value of $d = 0.149 \pm 1$ clear air, liquid or just bad data?
   d. Complex cloud scenes masking multilayer clouds
   e. Long term signal degradation

4. What definition of depolarization are you using? There are several in the literature, well summarized by Flynn et al. 2007 or Hayman and Thayer 2009 or 2012. Depolarization ratio vs. the Mueller matrix element d (also called depolarization) can differ by factors approaching 2. This will directly impact your ice/liquid phase classification.
5. The reference to Liou and Yang 2016 as summation of depolarization lidar is not appropriate in my opinion. There are multiple papers dating to at least Schotland et al. 1971 that are more fundamentally related to lidar such as Sassen 1991 and more recent papers such as Gimmestad 2008 or Hayman and Thayer 2012 that are complete and well known.
6. On line 283, you specify that 98% of spectra are correctly classified. That really just says that your training and test data sets are self-consistent. Furthermore, it really just says that you are pushing your reference to the lidar system. If you take the above comments numbered 2 and 3, it makes it very difficult to analyze how accurate the classification really is. Furthermore, it is impossible to replicate in any meaningful way.
7. There are a number of physical interpretations given that seem both counterintuitive and relatively easy to link to poor control of lidar data. Some examples are:
   a. Multiple scattering: You say a number of times that liquid sits below ice layers. This is counterintuitive to all the results I have seen from Arctic studies (summarized nicely by Morrison et al. 2012 and references therein). However, this is really easy to explain given the presence of multiple scattering. Even in the presence of non-depolarizing scatterers, multiple scattering can cause monotonic increases in depolarization measurements with range.
   b. You mention on line 288-289 that optically thin cloud phase is problematic to define? Without error bounds on your depolarization measurements, you cannot define how accurately you are measuring clouds, which could easily affect physical interpretations (as in the above example of $d = 0.149 \pm 1$). Second, if you are performing cloud identification (regardless of phase by visual inspection), optically thin clouds are very likely to be missed.
   c. I am really puzzled by the results in Table 5 indicating almost no observations of mixed phase clouds for 9 months out of the year. I wonder if thick liquid clouds with high occurrences of multiple scattering are being misclassified?

Specific Minor Comments

1. Line 4: Probably mean 2015
2. Line 67-68: LiDAR is first used in line 46 and should probably be defined there.
3. The capitalization of LiDAR seems odd to me. It, much like the acronym radar, is in my experience most commonly used as a word. For example, Palchetti et al. 2015 simply uses “lidar”. I would suggest adopting this convention.
4. Color scheme of Figure 8. It is a minor point but using blue for ice instead of mixed phase or liquid is an odd choice to me.
5. I would also point out that Figures 2, 5, 7, 8, 10, 11 and 12 would be difficult to read for those who are red/green colorblind.

Suggested References:

1. Schotland, R.M., K. Sassen, and R.J. Stone, 1971: Observations by lidar of linear depolarization ratios by hydrometeors. J. Appl. Meteor., 10, 1011-1017.
2. Sassen, K. (1991). The Polarization Lidar Technique for Cloud Research: A Review and Current Assessment, Bulletin of the American Meteorological Society, 72(12), 1848-1866.
3. Jens Biele, Georg Beyerle, and Gerd Baumgarten, "Polarization lidar: Corrections of instrumental effects," Opt. Express 7, 427-435 (2000)
4. Alvarez, J. M., M. A. Vaughan, C. A. Hostetler, W. H. Hunt, and D. M. Winker. " Calibration Technique for Polarization-Sensitive Lidars". Journal of Atmospheric and Oceanic Technology 23.5 (2006): 683-699.
5. Connor J. Flynna, Albert Mendozaa, Yunhui Zhengb, and Savyasachee Mathurb, "Novel polarization-sensitive micropulse lidar measurement technique," Opt. Express 15, 2785-2790 (2007)
6. Shupe, M. D. (2007), A ground-based multisensor cloud phase classifier, Geophys. Res. Lett., 34, L22809, doi:10.1029/2007GL031008.
7. Gary G. Gimmestad, "Reexamination of depolarization in lidar measurements," Appl. Opt. 47, 3795-3802 (2008)
8. Matthew Hayman and Jeffrey P. Thayer, "Explicit description of polarization coupling in lidar applications," Opt. Lett. 34, 611-613 (2009)
9. Nott, G.J. and Duck, T.J. (2011), Lidar studies of the polar troposphere. Met. Apps, 18: 383-405. https://doi.org/10.1002/met.289
10. Matthew Hayman and Jeffrey P. Thayer, "General description of polarization in lidar using Stokes vectors and polar decomposition of Mueller matrices," J. Opt. Soc. Am. A 29, 400-409 (2012)
11. Morrison, H., de Boer, G., Feingold, G. et al. Resilience of persistent Arctic mixed-phase clouds. Nature Geosci 5, 11–17 (2012). https://doi.org/10.1038/ngeo1332
12. Freudenthaler, V.: About the effects of polarising optics on lidar signals and the Δ90 calibration, Atmos. Meas. Tech., 9, 4181–4255, https://doi.org/10.5194/amt-9-4181-2016, 2016.