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.

Thanks for the interest in our work. We have used four years of data (2012-2015).

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.

We expect that the community members ask for clarifications and provision of details when their goal is a better understanding of the obtained results. The present research is the result of a huge effort by a community of scientists who work in this field with professionality from several years. Thus, we are fully available to respond to requests going in this direction and not affected by any preconceptual judgment.

Specific Major Comment

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.

The lidar system will be better described in the next version of the paper and new references will be added.
However, the description of the same lidar system was presented in many other articles (Di Natale et al., 2020; Maestri et al., 2019; Rizzi et al. 2016; Tomasi et al., 2015), and also in 3 previous works published in ACP (Ricaud et al., 2020, 2017; Chen et al., 2017). In all these publications the description of the lidar was considered satisfactory by the community. We are surprised that the reader has completely missed all the relative recent articles even the ones that are more focused on the lidar data.
Concerning the website indicated in Palchetti et al. (2015), the reader is right since the provided link was incorrectly typed.

The website is always active at the following link:
http://lidarmax.altervista.org/englidar/Antarctic%20LIDAR.php
(accessed on 17 May 2021).

Some of the major ones (this is by no means an exhaustive list) are:

Here below there is a long list of technical details. We agree with the reader that some of these details could be included in the next version of the article. Nevertheless, we would like to keep the focus on the identification and classification process of spectral radiances and the discussion of the results rather than the lidar.

a. Is the system coaxial or biaxial? This will affect the height range of detectable signal as well as the observed signal strength.

The system is biaxial, with 10 cm off-axis.
Please note that the Dome C site is located at almost 3300 m a.s.l and clouds are usually found between surface and 2 km above surface level.

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.

- LIDAR Detection: Hamamatsu PMT, analog mode. Automatic avoidance of signal saturation from g.l. upward through laser power modulation
- Signal averaging over 1000 laser shots
- True: signal dynamic range of 4.5 orders of magnitude is required assuming a completely uniform scene. But this is not the case: the molecular atmosphere at 7000m is not detected, just cirrus clouds are (with a scattering ratio of at least 10). 7000 m is the upper limit for detection: Concordia cirrus are well below 4000 m altitude. Moreover, in case of low clouds laser power is automatically reduced, so that a signal compression is always obtained. The figure of 4.5 magnitudes thus doesn't apply to this instrument.

c. What is the system field of view? This will directly affect depolarization measurements via multiple scattering.

FOV is approx. 2 mrad full angle.

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.

The nominal laser aperture is 1 mrad full angle.

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.

The calibration is obtained by inserting a lambda/2 plate at the laser exit (in order to have e roughly 50% power on both polarizations). In absence of clouds, a measurement of the ratio between the two (p,s) output signals (averaged over a 1000 m window) is obtained. The two pmts are exchanged of place (keeping everything else unchanged in the acquisition chain) and a second ratio (same window) is obtained. The geometric mean of the two ratios provides a measure of calibration ratio, insensitive with respect to changes in the atmosphere and laser power.

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.

No algorithm was used for this type of study: Concordia LIDAR is used as a range-finder for cloud base, top, vertical extension, time evolution and water phase (liquid/solid) from depolarization. Background/offset subtraction only is applied. No quantitative LIDAR data about backscatter, extinction, and else are in fact given in the paper. The ratio of offset-corrected signals is extremely reliable in providing our simple information. Methods like the one by Klett are unapplicable as automatic procedures, as the reader correctly suggests, and quite unreliable in complex atmospheres.

Clouds are very complicated objects. The parameters that define a cloud span over a large range of values. The lidar quicklooks are available to the reader (please visit the website). The list of the times of lidar measurements used in the study to define the training sets can be provided as additional material. For all the considered cases the atmospheric structure associated with the presence or the absence of a cloud are neatly identifiable. The selection of the training set elements was performed accurately choosing the observations and avoiding the most complicated cases.

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.

The reader is partially repeating the same question posed at point 1.f. See reply above.

Please note that we identified clearly three different atmospheric conditions through the lidar measurements: the clear sky, the ice cloud and a category called mixed phase.

The mixed phase category is better described in the next version of the paper.
Mixed phase clouds are characterized by a layer with small values of the depolarization ratio at cloud base (less than 15%), characteristics of liquid spheres. The depolarization ratio increases towards values typical of ice crystals near the cloud top. An increase is, in part, intrinsically related with liquid water layers, where multiple scattering determines a depolarization that gradually increases with the depth of penetration, in the lidar backscatter. For this reason, in some conditions,

the phase of the upper part of the cloud cannot be unambiguously defined based on the analysis of the depolarization ratio profile only. Nevertheless, the presence of liquid phase at bottom is unequivocally identified and the cloud is categorized as mixed-phase. Moreover, a common situation is the presence of falling ice from mixed-phased cloud layers, as shown in the mid panel of Figure 2 between 18 and 20 UTC. Typically, the quantity of the precipitating ice crystals is very small and the CIC algorithm is able to capture the radiometric signal from the upper liquid water layer as it will be shown in the case reported in Figure 7. For a classification point of view, the identification of liquid particles in the layer is the key information which makes the observed cloud to pertain to a specific category (mixed-phase in our nomenclature) that is different from 'pure' ice clouds.
The two categories (ice and mixed phase clouds) show peculiar radiometric features in the REFIR-PAD spectra which are captured by the CIC classificator. For this reason, we believe that multiple scattering above the liquid layer does not affect the classification results.

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 ± 1 clear air, liquid or just bad data?
d. Complex cloud scenes masking multilayer clouds
e. Long term signal degradation

Many authors use the threshold found by Intrieri et al. (2002) as a reference which uses a depolarization ratio of 0.11 to discriminate between ice and liquid water particles. We used the value (0.15) indicated by Sassen (1991). The lidar system is now better described and the reader can evaluate if the threshold is reasonable.

Note that the schemes proposed by Shupe (2007) exploits co-located measurements from lidar, radar, microwave radiometer, and temperature vertical profiles to classify clouds, which are not available in our case. In the experimental conditions encountered at Dome C clouds are very thin and composed of small particles that makes the proposed methodology totally ineffective.

Moreover, note that the use of the depolarization ratio to define the cloud phase has been used in many papers, such as the cited Nott and Duck, (2011).

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.

The depolarization ratio used is the simple signal depolarization (cross polarized total signal/ parallel polarized total signal *100), after background subtraction.

As the scattering ratio R of most Concordia cirrus exceeds 10, this means that a typical signal depolarization of 30% reported in the paper could in fact correspond to a cloud depolarization (cross polarized cirrus signal/ parallel polarized cirrus signal *100) of 33%,

Furthermore, the possible ambiguity between liquid phase clouds and oriented ice plates is avoided at Dome C by operating the lidar 4° off-zenith (Ricaud et al., 2020).

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.

Major comment?
We provided a general definition. We think that the reference book is appropriate, but we are available to add a new reference.

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.

Respectfully, we disagree with the reader's comment that, in principle, can be applied to every classification procedure based on automatic learning.
Moreover, the reader is assuming that the cases included in the test set are perfectly mirroring what is included in the training set.
The lidar data is used to define 3 distinct classes as clarified above. These classes correspond to 3 typical lidar backscatter and depolarization ratio observations in Antarctica. Moreover, they also correspond to specific radiometric signals in the FIR and MIR that are neatly recognized by the CIC algorithm in most of the cases (but not all).
See also replies to other community reader.

Differently to other methodologies, the classification is easily repeatable. We can provide the list of the times of the lidar and refir-pad measurements contained in the training set as additional material to the article. Note that CIC simply ingests the training set spectra and the dataset spectra and performs the classification without any other tuning. We are not aware of others methodology which work so straightforward. In the past we have used neural network and support vector machine methodologies.

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.

We have improved the description of the mixed-phase class and included a discussion of the multiple scattering effect that goes in the direction of the reader's comment. See also replies to reviewer 1 and 2.

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 ± 1). Second, if you are performing cloud identification (regardless of phase by visual inspection), optically thin clouds are very likely to be missed.

Yes, the sentence should be re-phrased. The IR radiance signal in presence of cloud approaches the clear sky radiance signal as the cloud optical depth becomes thinner.
The sensitivity of CIC to thin cirrus clouds has been tested in previous studies such as Maestri et al. (2019) and Magurno et al. (2020).

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?

We think that we made clear in the text that mixed-phase clouds are not considered in the cold season (i.e. from April to October included).
The low occurrence of mixed-phase clouds is in accordance with results from Listowski et al. (2019), being near to zero in MAM, JJA, and SON.

Specific Minor Comments
1. Line 4: Probably mean 2015
Corrected

2. Line 67-68: LiDAR is first used in line 46 and should probably be defined there.
No. In that case the word LIDAR is used in the expression of the acronym for CALIOP.

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.
We adopt simply "lidar", as suggested.

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.
The colors (red for clear sky, blue for ice clouds, and green for mixed-phase clouds) are kept the same in all figures to facilitate the analysis for the readers.

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.

Yes, the reader is right, but we don’t have a solution for this at the moment.

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.

References

Chen, X.; Virkkula, A.; Kerminen, V.-M.; Manninen, H. E.; Busetto, M.; Lanconelli, C.; Lupi, A.; Vitale, V.; Del Guasta, M.; Grigioni, P.; et al. Features in Air Ions Measured by an Air Ion Spectrometer (AIS) at Dome C. Atmos. Chem. Phys., 2017, 17 (22), 13783–13800. https://doi.org/10.5194/acp-17-13783-2017.

Di Natale, G.; Bianchini, G.; Del Guasta, M.; Ridolfi, M.; Maestri, T.; Cossich, W.; Magurno, D.; Palchetti, L. Characterization of the Far Infrared Properties and Radiative Forcing of Antarctic Ice and Water Clouds Exploiting the Spectrometer-LiDAR Synergy. Remote Sensing, 2020, 12 (21), 3574. https://doi.org/10.3390/rs12213574.

Intrieri, J. M. An Annual Cycle of Arctic Cloud Characteristics Observed by Radar and Lidar at SHEBA. J. Geophys. Res., 2002, 107 (C10). https://doi.org/10.1029/2000jc000423.

Listowski, C.; Delanoë, J.; Kirchgaessner, A.; Lachlan-Cope, T.; King, J. Antarctic Clouds, Supercooled Liquid Water and Mixed Phase, Investigated with DARDAR: Geographical and Seasonal Variations. Atmos. Chem. Phys., 2019, 19 (10), 6771–6808. https://doi.org/10.5194/acp-19-6771-2019.

Maestri, T.; Arosio, C.; Rizzi, R.; Palchetti, L.; Bianchini, G.; Del Guasta, M. Antarctic Ice Cloud Identification and Properties Using Downwelling Spectral Radiance From 100 to 1,400 Cm −1. J. Geophys. Res. Atmos., 2019, 124 (8), 4761–4781. https://doi.org/10.1029/2018jd029205.

Maestri, T.; Cossich, W.; Sbrolli, I. Cloud Identification and Classification from High Spectral Resolution Data in the Far Infrared and Mid-Infrared. Atmos. Meas. Tech., 2019, 12 (7), 3521–3540. https://doi.org/10.5194/amt-12-3521-2019.

Magurno, D.; Cossich, W.; Maestri, T.; Bantges, R.; Brindley, H.; Fox, S.; Harlow, C.; Murray, J.; Pickering, J.; Warwick, L.; et al. Cirrus Cloud Identification from Airborne Far-Infrared and Mid-Infrared Spectra. Remote Sensing, 2020, 12 (13), 2097. https://doi.org/10.3390/rs12132097.

Nott, G. J.; Duck, T. J. Lidar Studies of the Polar Troposphere. Met. Apps, 2011, 18 (3), 383–405. https://doi.org/10.1002/met.289.

Ricaud, P.; Bazile, E.; del Guasta, M.; Lanconelli, C.; Grigioni, P.; Mahjoub, A. Genesis of Diamond Dust, Ice Fog and Thick Cloud Episodes Observed and Modelled above Dome C, Antarctica. Atmos. Chem. Phys., 2017, 17 (8), 5221–5237. https://doi.org/10.5194/acp-17-5221-2017.

Ricaud, P.; Del Guasta, M.; Bazile, E.; Azouz, N.; Lupi, A.; Durand, P.; Attié, J.-L.; Veron, D.; Guidard, V.; Grigioni, P. Supercooled Liquid Water Cloud Observed, Analysed, and Modelled at the Top of the Planetary Boundary Layer above Dome C, Antarctica. Atmos. Chem. Phys., 2020, 20 (7), 4167–4191. https://doi.org/10.5194/acp-20-4167-2020.

Sassen, K. The Polarization Lidar Technique for Cloud Research: A Review and Current Assessment. Bull. Amer. Meteor. Soc., 1991, 72 (12), 1848–1866. https://doi.org/10.1175/1520-0477(1991)072<1848:tpltfc>2.0.co;2.