Review of "A dataset of microphysical cloud parameters, retrieved from Emission-FTIR spectra measured in Arctic summer 2017"

Overall, the manuscript is appropriate for ESDD. The dataset that is described is interesting and unique. I am not aware of many ground based datasets of arctic clouds over the ocean. The methodology is sound, using standard OE techniques, with supporting analysis with simulated data to characterize the expected range of retrieval errors. The data and software are all available via DOI.

There are issues with the methodology and the explanation that need improvement (described below). There are also a very large number of small technical and grammatic errors (listed at the end of the review). I strongly advise the authors to review the manuscript again (including a spell check!) as there are so many typos that I am certain I did not notice all of them. Because of the large number of issues I would say the manuscript needs major revision to be accepted.

My main concern with the manuscript is related to the comparisons to Cloudnet. First, the manuscript does not clearly describe the Cloudnet data. In the abstract, when referring to the Cloudnet data: "...liquid water path retrievals from microwave radiometer ..." (line 8) which suggests the LWP from Cloudnet is from solely the microwave data. On the other hand, the Cloudnet data is "combined cloud radar, lidar, and microwave radiometer" (line 5). Please clarify exactly how the Cloudnet works here - it is important to know how Cloudnet is retrieving the variables that are compared to the variables from the emission FTIR (IWP, LWP, particle size). Are the different variables simply retrieved from the individual remotely sensed measurements? How would that then work for particle size? Doesn't that require joint lidar/radar?

A main conclusion of the manuscript is that the Cloudnet LWP measurements are more accurate than the 20 g/m2 uncertainty that is quoted on the product. The supporting evidence is that the LWP retrievals from both methods (cloudnet and TCWret) show high correlation even when LWP < 20 g/m2. I am not sure this follows - it depends on how the original 20 g/m2 uncertainty estimate was derived for the cloudnet data. If this was assessed by comparison to independent "truth" estimate, then if the true errors in both
Cloudnet and TCWret are correlated, one would see this correlation even though the true error in both methods is still 20 g/m². For example, the 'parameter' uncertainties discussed in section 5.6 could drive correlated errors in both retrievals.

In section 5.5 (and 6.2), it would be much more informative to also show the posterior correlation matrix. An important point in the discussion section is the tradeoff between \( r_{\text{liquid}} \) and \( r_{\text{ice}} \). It would be very useful to know if the output of the OE algorithm shows this correlation (e.g. the \( r_{\text{liquid}} - r_{\text{ice}} \) correlation term should be negative and have a large magnitude).

The parameter error discussion in section 5.6 has some unclear aspects. At line 257 "Each of these modifications is applied individually, creating three new datasets". If the modifications are made as described (e.g., add +1K to each cloud's temperature), it seems like this would only tend to create a mean bias in the retrieval, not increase the uncertainty. If it was done in this way, then these parameter errors would seem to be significantly underestimating the actual parameter error magnitude.

At line 280-283 at the end of the section, I believe the authors are attempting to combine the various error estimates into one final combined error, but I cannot follow the explanation. Where does Delta \( T = 2K \), Delta \( q = 17.5\% \) come from? How should the reader interpret these Deltas from the blackbody emissivity and temperature versus the radiance error att line 256? What are these final "deltas" supposed to represent? If these are supposed to be the combined parameter and calibration errors, these are much larger than the range of the OE errors as reported in Table 3.

Section 5.7 was confusing at first because I think the explanation at line 286 is wrong. Table (5) is not "standard deviations of \( r_{\text{ice}} \)" but rather the standard deviation of the differences in the retrieved \( r_{\text{ice}} \) between two variations of the retrieval that assumed different ice crystal habits. This section would benefit from improved explanation. It is still unclear to me what these results imply about the retrieval product.

Section 6.3 introduces a cutoff value in PWV (1 cm) which is used to categorize the data. This is based on Cox 2016 ESDD, but the Cox et al manuscript does not address this issue at all. And more importantly, Cox 2016 does not address the water vapor transmission relevant to the specific spectral ranges used in TCWret. A plot or table should be added with the total atmosphere transmission through the selected microwindows at the cutoff value of PWV (1 cm), and I would even add the limit values observed during the campaign (by eye, in Figure 8, this is roughly 0.7 - 1.65 cm).

Section 6.4, Line 335: The ice crystal habit selection needs more explanation. If the habit was randomly chosen for \( r > 30 \), wouldn't that imply all the habits except droxtal should have a roughly equal percentage of the total retrievals? Was there some other criteria used for selecting the habit (which does not appear in the manuscript?) Also, if the habit is changing between retrievals, then how is this captured in the output product? I do not see any way this was tracked in the output netCDF file.

Line 30: the authors quote an LWP uncertainty from a microwave retrieval in the literature; is this using the data from microwave radiometers at the same frequency as Cloudnet? I do not think the MWR frequency for the Cloudnet/OCEANET instruments is mentioned anywhere.

Line 60: I would suggest adding a couple more simple pieces of information to help understand the dataset: how many days of data were in each "cruise leg", and what was the approximate fraction of time the vessel was in cloudy conditions?
Minor technical errors, typos, short clarifications, etc:

Line 8: "a uncertainty" -> "an uncertainty"

Line 12: this is unclear: " ... dataset ... allows to perform ..." suggest " ... dataset ... allows researchers to perform calculations ...", is that the intended meaning?

Line 24: "places" -> "place"

Line 44: "where low absorption of gases occur" - this is false since the spectral range includes the CO2 absorption band; add a sentence here about the fact that the TCWret is using selected microwindows within that range.

Line 67: "The spectrometer was permanently rinsed with dry air." I have never heard the term "permanently rinsed" used in this context, so this is unclear. Can you explain this in more detail? Is the internal air continuously recirculated with desiccated ambient air, or was it purged with dry air and then sealed during the measurement campaign?

Line 85: what is the length of time for one complete calibration cycle? (specifically, how much time elapses between views of the blackbody at the same temperature?) And what is the duty cycle? (specifically, what fraction of the time is spent looking at the blackbodies versus the atmosphere)

Line 101: "Informations about the cloud ceiling were recorded..." -> "Information about the cloud ceiling was recorded..."

Line 124: was the CO2 concentration also the standard atmosphere value, or did you pick a more appropriate value for 2017?

Line 133: "Temperature depended" -> "Temperature dependent"

Line 138: Were the droplet size and ice crystal size distributions both gamma functions?

Line 159: standard notation for this variable uses "chi", not "xi", and "xi" was already used for the cost function, which is an entirely different quantity. (\( \chi = \chi \), \( \xi = \xi \)) The expression in (7) is incorrect, assuming this is supposed to be a standard reduced \( \chi^2 \) variable, it should be:
\[
\chi^2 = \text{Sum} \left( \frac{(y - F(x))^2}{\sigma^2} \right) / \text{DOF}
\]

Line 173: is the retrieval done in log-space, or linear space for \( \tau \)? (this line seems to contradict what is said just above).

Line 175: in standard notation, the extinction coefficient is beta, and the extinction cross section is sigma.

Line 188: By my reading of the Ceccherini and Ridolfi 2010 notation, the left term in parentheses in equation (4) is \( M_i \) inverse, not \( M_i \). Please double check.

Line 206: Suggest changing the section title to 'Retrieval performance on simulated data' a similar phrase, to make it clear this section is not using real measurements.

Line 207: "artifical" -> "artificial"

Line 215: "parametern" -> "parameters", "stndard" -> "standard"
Table 3: Can you quote the number of test cases used? Is ERR(OE) is the mean of the posterior uncertainty predicted by the OE algorithm?

Line 225: Here, the text states: \( f_{\text{ice}} = \tau_{\text{ice}} \times \tau_{\text{cw}} \), I think this should be \( f_{\text{ice}} = \tau_{\text{ice}} / \tau_{\text{cw}} \).

Line 241: This sentence is unclear, could it just be deleted? I am not sure what the authors intend here.

Line 249: More detailed is needed. Does this sentence imply that all TCWret retrievals (in particular, those performed on the real measurements from Polarstern) have scaling applied to the posterior errors as predicted by the OE?

Line 250: "Errors" -> "Errors"

Line 251: "humidty" -> "humidity"

Line 278: equations in text are missing closing parentheses.

Line 280: \( T_{BB} \) should be 100 C, not 100 K

Figure 5 caption: "retreived" -> "retrieved". Also, these histograms are not the counts, some normalization was done - are these PDFs (meaning they integrate to 1)?

Line 330 "intransparent" is not a word. I think what the authors intended to say is "Atmopsheric transmission in the far-infrared spectral region drops to zero for PWV > 1 cm." See earlier comment about this statement.

Line 353 "Withouth" -> "Without"

Figure 12: The units on the axes are wrong, I think this should be (um)?

Line 363 - 365: These sentences are unclear.

Line 367: The "very thin clouds" should be the LWP cutoff, not the PWV cutoff?

Line 385: "Jupyer" -> "Jupyter"

Line 400: I would reiterate that the utilized test cases are simulated or synthetic data, not real observations with some independent estimate.