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?

The Cloudnet retrieval is described in Illingworth et al. (2007). A complete description of the cloudnet retrieval was not planned, so we have only provided a summary of the measurement equipment used.
The retrieval for this measurement campaign is described in Griesche et al. (2020) and we will give a brief description of the used quantities here and refer to this publication in the following.
The LWP is indeed determined from the measurements of the Humidity and Temperature Profiler(HATPRO). The frequency bands from 22.24 GHz to 31.4 GHz and from 51.0 GHz to 58.0 GHZ are used. A statistical retrieval is applied, which was set up using radiosonde data from Ny-Alesund. This retrieval corresponds to the retrieval described in Löhnert and Crewell (2003).
With the calculated LWP and radar reflectivity, r_liq is determined according to Frisch et al. (2002) for data points classified as liquid water.
We calculated the IWP ourselves via vertical integration of the IWC, which is determined by an empirical formula from temperature and radar reflectivity according to Hogan (2006).
R_ice is proportional to the ratio of IWC and the visible extinction coefficient and is thus also calculated using temperature and radar reflectivity.
IWC and R_ice is only calculated for data points classified as ice or supercooled water.
In principle, the LWC is also part of the cloudnet retrieval, which is determined on the basis of the LWP. However, since only the integrated quantities can be determined with the FTIR, the LWC is neglected here.
Before the individual retrievals, however, a classification of the data is performed (liquid, ice, aerosol, ...), which is made possible by the combined use of radar and lidar (Illingworth et al. 2007).

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.

The data on which the retrievals are based - with the exception of the atmospheric profiles - are different. TCWret also does not use Cloudnet's retrieval results as a priori information, so an influence on the results is not to be expected here either. However, TCWret and Cloudnet use the same radiosonde measurements of temperature and humidity for their retrievals, so the possibility of correlated errors cannot be completely ruled out. However, since Cloudnet and TCWret use the same atmospheric profiles, inaccuracies in them should not matter, so we believe the parameter errors are also negligible. This leaves only the retrieval errors.
The 20 gm-2 is the RMSE of the statistical retrieval, while the uncertainty of TCWret is that from the covariance matrix of the optimal estimation.

Additionally, all readings are averaged over a 2 minute period for better comparability. Again, I see the possibility that this reduces Cloudnet's true uncertainty. However, similar results show when results are averaged over one minute.

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_liquid and r_ice. It would be very useful to know if the output of the OE algorithm shows this correlation (e.g. the r_liquid - r_ice correlation term should be negative and have a large magnitude).

For clarification: We understand the posterior correlation matrix to be the matrix that contains the mutual correlations, for example Corr(tau_liq, tau_ice).
We found that the Averaging Kernel Matrix (AVK) already contains all the necessary information about mutual dependencies and dependencies on the A prior. In addition, the AVK provides information about the degrees of freedom (tr(AVK) = degrees of freedom). Nevertheless, we will provide the Posterior Correlation Matrix in Section 6.2, which will be used in the interpretation of the results. In Section 5.5, which will be moved to the Appendix, we will continue to consider only the AVK.
The correlation coefficient Corr(r_liq, r_ice), however, is not negative and also has no larger magnitude than the other correlations. The posterior correlation matrix of the mixed-phase clouds of the measurement campaign is shown:

	t_liq	t_ice	r_liq	r_ice
t_liq	1.000000	0.503039	-0.072331	-0.409233
t_ice	0.503039	1.000000	0.016336	-0.230436
r_liq	-0.072331	0.016336	1.000000	0.130602
r_ice	-0.409233	-0.230436	0.130602	1.000000

Here, just as in the AVK from equation 22, one can see that there is a greater correlation between t_liq and t_ice, which is in agreement with the difficult phase determination. In addition, there is a high correlation between r_ice and tau_liq, which also suggests that these quantities cannot be determined independently or that in the clouds the optical thickness for liquid water is indeed correlated with the ice crystal radius

The correlation between r_liq and r_ice is low. We suspect that the expectation that the r_liquid - r_ice correlation term should be negative and have a large magnitude is based on the explanation in Section 6. However, since bar{r} also contains f_ice and thus also the optical depths, the correlation is not as direct as stated here. We will revise the explanation of the mean biases in this respect. We also have removed bar{r} from the manuscript as it does not add any further value to the interpretation.

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.

The aim of this section was to carry out an error propagation analogous to equation 16, but this time taking into account variables that enter the retrieval as parameters. These are temperature of the cloud, humidity and the temperature of the blackbody, which influences the calibration and thus the size of the radiance itself. According to equation 16, to determine the uncertainty with respect to one of these quantities, I need the variance or standard deviation, as well as the partial derivative with respect to the quantity. The former are the specified instrument errors, while the latter is unknown, since the retrieval by temperature. Humidity and radiance does not vary. Thus, the partial derivatives are determined separately in this case:

- We apply the described perturbation (e.g. add +1K to cloud temperature) to the test cases and calculate the cloud parameters under this perturbation.
- These results are compared with the test case retrievals without disturbance. I calculate the differences.
- The sigma (Table 4) is then the partial derivative by temperature, humidity and radiance normalised to unit size (1K, 1%rh and 1 RU). Since I want to calculate unperturbed partial derivatives, I can only apply one perturbation per data set. There is an error here because in the paper it is the standard deviation, although it must be the mean values.
- To calculate the total parameter error, Equation 18 then multiplies the unit errors by the partial derivatives.

Since we used these modifications to calculate the partial derivatives, which are weighted with the real instrument errors, this procedure should yield more than a mean bias.

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.

The Delta (Delta T, Delta q and Delta L) consist of

- The specified instrument errors of the radiosondes and the blackbody.
- What we have called here the interpolation error. This interpolation error takes into account the errors introduced into the temperature and humidity by linear interpolation between two radiosonde data setsThe interpolation error follows from the comparison between the linear interpolation between two radiosonde measurements and the ERA5 atmosphere at the position of the measurements. We query the ERA5 atmosphere for each hour. Then we calculate the atmospheric profiles from the radiosondes once per hour by linear interpolation. From this we calculate the difference, average over one day and calculate the standard deviation.

We added the two types of error and plotted them in Figure 4. Delta T = 2K and Delta q = 17.5% are estimates of the mean errors.
The magnitude of the errors is too high. This is because we are using the partial derivatives incorrectly as described above. Using the partial derivatives described above, the parameter errors are:
Tau_liq: 0.42 tau_ice: 0.28 r_liq: 3.30 um r_ice: 13.08 um lwp: 2.75 gm-2 iwp: 5.64 gm-2

We reformulated the entire section, improved the description and divided it into several subsections.

Section 5.7 was confusing at first becase I think the explanation at line 286 is wrong. Table (5) is not "standard deviations of r_ice", but rather the standard deviation of the differences in the retrieved r_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.

We have removed the entire section. The retrieval was performed again using each ice crystal shape for all spectra. Since neither r_ice of TCWret and Cloudnet correlate independently of ice crystal shape, we see no added value in this section for the interpretation of the results.

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.

In the data, the ice shape is stored as a number. We will add a table with the corresponding meanings in the appendix..

Since the shapes of the ice crystals are not known, the retrievals were carried out for all ice crystal shapes. However, this procedure leads to up to 8 results per measurement, so a selection was made. The accepted result was determined as follows:

- If r_ice for plates, bullet rosettes and solid columns or for droxtals are less than 30 um the result using ice crystals as droxtals is accepted.
- If the first condition is not fulfilled and r_ice for Droxtals are greater than 10 um, the result that uses plates, bullet rosettes or solid columns is accepted. To choose one of the datasets, a random number is drawn which selects plates in 35%, bullet rosettes in 15% and solid columns in 50%
- If none of the conditions above apply, the data for which the degrees of freedom of the outcome are highest is accepted.

The first condition ensures that all small ice particles are classified as droxtals, while the second ensures that all larger particles are classified as plates, solid columns or bullet rosettes. Stricter thresholds would more often result in only the last condition applying, which should be avoided as much as possible.
The ice crystal shapes shown in Figure 9 are determined from this procedure.

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).

Cox et al. (2016, ESSD) is the paper that describes the testcases. The correct paper is Cox et al. (2015, Nature Communications.) We will correct the citation. The limit values are 0.67 cm and 1.62 cm. Since we only mention this cut-off value with reference to the literature, but then also interpret the data with PWV > 1cm, we don't think that a plot the transmission is necessary and rather refer to figure 3 of Cox et al. (2015), where the radiance for PWV = 0.36cm and PWV = 1cm are shown.

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.

The LWP of the Cloudnet retrieval is from a HATPRO microwave radiometer and uses two frequency bands (22.24 GHz to 31.4 GHz and 51.0 GHz to 58.0 GHz). The retrieval corresponds to that from Lohnert/Crewell of 2003, as mentioned in Griesche et al.(2020, AMT). With the used frequencies, the RMSE found in Lohnert/Crewell is 22.4 gm-2. We have added the frequencies and the RMSE.

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?

We will add such information in the section „Observations“ based on the number of successful retrievals.

Minor technical errors, typos, short clarifications, etc:

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

Line 12: this is unclear: " ... dataset ... allows to perform ..."
suggest " ... dataset ... allows researchers to perform calculations ...", is that the intended meaning?<
That is true. We rephrased the sentence

Line 24: "places" -> "place"
Done

Line 44: "where low absorption of gases occur" - this is false since the spectral range includes the CO_2 absorption band; add a sentence here about the fact that the TCWret is using selected microwindows within that range.
A sentence that mentions the microwindows is added

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?
Continuously dry air was brought into the spectrometer to avoid damage to the hygroscopic beam splitter. We have reformulated it.

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)
Each calibration measurement cycle took 10 minutes and each atmosphere measurement took cycle 15 minutes.

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

Line 124: was the CO_2 concentration also the standard atmosphere value, or did you pick a more appropriate value for 2017?
We used the value for the standard atmosphere (330ppm). This potential source of error is discussed in a section in the appendix. We performed retrievals using spectra from the 11th June 2017 with CO_2 concentrations of 330 ppm and 410 ppm. Mean biases and root-mean-square errors are
Tau_liq: -0.0 +- 0.1
Tau_ice: 0.0 +- 0.1
R_liq: 0.1 +- 1.1
R_ice: 0.0 +- 1.3

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

Line 138: Were the droplet size and ice crystal size distributions both gamma functions?
Correct, both size distributions are 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 = χ , xi = ξ)
We changed it from xi to chi

The expression in (7) is incorrect, assuming this is supposed to be a standard reduced chi^2 variable, it should be:
chi^2 = Sum((y - F(x))^2 / sigma^2) / DOF
We corrected the expression.

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

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

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.
That is correct. We have corrected our equation in the manuscript.

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

Line 207: "artifical" -> "artificial"
Done

Line 215: "parametern" -> "parameters", "stndard" -> "standard"
Done

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?
ERR(OE) is the mean uncertainty from the covariance matrix. There are 253 test cases included in this analysis. We will mention the number in the caption.

Line 225: Here, the text states: f_ice = tau_ice * tau_cw, I think this should be f_ice = tau_ice / tau_cw.
That is true. We have corrected the equation.

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

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?
Ja genau, das habe ich mit den Fehlern gemacht. Habe es umformuliert.

Line 250: "Erorrs" -> "Errors"
Done

Line 251: "humidty" -> "humidity"
Done

Line 278: equations in text are missing closing parentheses.
Done

Line 280: T_BB should be 100 C, not 100 K
Done

Figure 5 caption: "retreived" -> "retrieved". Also, these histograms are not the counts, some normalization was done - are these PDFs (meaning they integrate to 1)?
We have corrected the typo. Also the figure was exchanged. Now it represents the counts.

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.
That is what this sentence was meant to say. We have exchanged the term to opaque.
Line 353 "Withouth" -> "Without"
Done

Figure 12: The units on the axes are wrong, I think this should be (um)?
That is correct. The units in the figure have been corrected.

Line 363 - 365: These sentences are unclear.
We have rephrased this sentence in the manuscript.

Line 367: The "very thin clouds" should be the LWP cutoff, not the PWV cutoff?
That is correct. We have corrected the wording.

Line 385: "Jupyer" -> "Jupyter"
Done

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