We thank the reviewer for a detailed review. Please find below our responses to your comments. Your comments are in black and our responses are in red.

1. *One of the main findings of this paper concerns the importance of considering scattering effects in microwave retrievals. However, I have a few concerns with the ways these results are obtained. A first way to evaluate the scattering effect is through comparisons between the retrievals and their associated a priori values from a neural network algorithm that doesn't consider scattering. I am not very convinced with the impact on* Cf*, which seems to stay close to its a priori value, but a reduction of LWPt is indeed clearly observed. Is the optimal estimation framework used for this study based on a Levenberg-Marquardt scheme, i.e. is a departure from the a priori value actually showing a reduction of the cost function (rather than being possible iteration noise in a Gauss-Newton approach)? Please comment on this, and for future study I'd suggest using more quantitative metrics like the cost function, information content or degrees of freedom to reach such conclusions.*

Yes, the convergence is monitored through a reduction of the cost function and through a convergence criterion as explained in the 2017 paper (C2017) eq 4. Because the problem is fairly well defined the convergence is very quick. The aspects mentioned by the reviewer are very relevant and they are at the very heart of the problem. The main reason why they were not addressed in more details in this work is because they were analyzed in detail in C2017 and here we wanted to focus more on the application of the retrieval rather than the retrieval itself, and also not to repeat previous analysis. Nonetheless, given the importance of the topic we have expanded section 4 and included more references to the results from the 2017 paper.

As it can be seen in Fig. 6 and 8 and Table I of C2017 the C_f showed a small improvement with respect to the a priori. The degrees of freedom were also analyzed in Fig. 7 where it was found that the DOF of the system for C_f varied depending on the physical constraints on the system. In this work a few changes were made, in particular only 3 quantities are retrieved and the drizzle DSD is provided. The a priori information for C_f is also better constrained because is derived with the help of the active retrieval. Although it is true that the change in C_f is not large, it is also possible that the a priori information provided is in within the limits of what can be achieved with this technique. The a-posteriori uncertainty of this parameter shown in Fig. 7a (this work) does show a reduction.

We have now included in Fig. 7 c and d more details on the retrieval. Fig. 7c shows the third element of the averaging kernel matrix A(3,3) defined in Eq. 5 of C2017 in relation to the average drizzle diameter. Fig. 7d shows one example of convergence. In this case (as in the majority of the cases that were able to converge) convergence is achieved at the 3rd iteration. In Fig. 7d for example the C_f parameter is quickly adjusted from 0.73 to 0.59. On the right axis of Fig. 7d we now show the cost function is shown. A(3,3) represents the varying contribution of the measurements to Cf that depends partially on the amount of scattering that the model attributes to the scene. The discrete values are due to the truncation of the DFS values to the first decimal digit.

Another way the importance of scattering is quantified is by comparing the retrievals of the new technique to those of MWRRET2, a similar retrieval algorithm. Considering the importance of these results, more details of the similarities / differences between the retrieval algorithms should be given in section 2.1. But why not simply turn scattering off (forcing the single-scattering albedo to 0) in your current retrieval algorithm, instead of using a different retrieval algorithm? That would avoid being impacted by retrieval technique differences and be much more convincing.

We thank the reviewer for this comment. This is actually something that was debated during the writing of the manuscript. The rationale for showing the comparison with MWRRET was that, because users are going to utilize MWRRET, they may be interested in knowing how much scattering is affecting that retrieval. However, we also see the point of the reviewer here and agree that using the same radiative transfer code and methodology has its merits as it eliminates possible differences and biases due to the different retrievals. Therefore, we have rerun the retrievals for the open cell cases setting the drizzle to zero. The results are now shown in Fig. 8 c,d.

2. *The impact of shafts on retrievals is discussed, first in the algorithm description and then in the result discussions. But it is still not clear to me, especially in the discussions surrounding Tables 2 and 3, what part of the conclusions concern impacts from retrieval limitations or from actual microphysics differences during shafts. Please clarify the exact (expected) impact of shafts on retrievals, so that the readers can more clearly understand your results.*

Although the retrievals were performed on a time resolution of 1 minute, the results were analyzed statistically, in terms of shaft averages. This because instantaneous properties of drizzle shafts may be dominated by turbulent processes, however average properties are important to understand physical processes that affect the larger scales.
It is our opinion that the different characteristics between open cell and close cell systems evidenced in Tables 2-5 (2-3 in the previous version) are actual micro- and macro-physical differences and not artifacts of the retrievals. Tables 3 and 5 report results from the active part of the retrieval which is a fairly well-established technique. As for the passive retrievals (Tables 2 and 4) there are two main limitations that affect the results: the first limitation concerns the lack of sensitivity of the microwave to drop sizes smaller than ~ 100 μm. This limitation affects both open and closed call cases, however given that the frequency of occurrence of small drops is higher in closed cell systems it will probably lead to a larger underestimation of in-cloud DWP in these systems. The second limitation concerns the inability of the microwave to retrieve during the time of more intense precipitation. This will only affect the open cell cases and will result in an underestimation of the average shaft CWP and DWP. The quantification of the impact will likely require an LES model. We added these comments in section 4, lines 341-349.