We would like to thank the reviewers for their suggestions, edits and questions which contributed to a hopefully improved revised manuscript. In the following, please find our point by point response to the reviewer' comments.

Anonymous Referee #2

Received and published: 4 May 2020

The authors report an interesting work of applying a deep learning model to 16 years of satellite data to create an observational classification of marine low cloud mesoscale morphology. The deep learning technique is quite novel in this area of remote sensing measurement and analysis. The science topic is also of interest to the atmospheric and climate science community. The paper is well written. I only have a few minor comments and questions for the authors to consider for improving the presentation quality of the paper.

Specific comments:
Line 18: Considering that AMT is an international journal, the authors might want to clarify on "NASA funded project" or remove it (which I don't think is critical to mention here)
Changed.

Line 21, Line 52, Line 77: Are these (128x128 or 256x256) the number of pixels? Is the pixel size 250 m? Please clarify in the main text. I wonder how the size of each scene has been determined. I imagine that a too big or too small size might cause some ambiguity in the classification of mesoscale cloud morphology. For example, some of the disorganized MCC scenes in Figure 7 look like evolving open-cell or closed-cell MCC. Have any sensitivity tests been performed to decide on the scene size for the training data?
Great point. We indeed spent months thinking about this question before deciding on 128x128 pixels. The pixel size is close to 1km. The main consideration is that if the size gets too large, e.g. 256x256, the chance of mixed types in a scene increases. On the other hand, if the size is too small, the lack of context renders classification by even humans hard because it can become quite ambiguous.

The example you raised for Figure 7 is important in showing that the scale really matters. The difference is more apparent at the native 128x128 scale. Looking at the scene when zooming out, some of the disorganized MCCs indeed can be confused with open-cell MCC.

Line 97-98: Except for the scenes got filtered out, does each scene have to belong to one of the six types when being analyzed for the frequency distribution? Please clarify.
Yes. We added a sentence for this point.

Line 105: Is the droplet size information used for disorganized MCC in the classification algorithm? This could be useful to remove ambiguity mentioned above.
We did not include the droplet size information. It would indeed provide extra information in many circumstances. However, including it would make the algorithm less general. We opt to not include it in this trade-off.

Line 141: how does rotating or flipping scenes help to increase the open-cell MCC sample size? That makes me wonder how the orientation of each scene affects the pattern recognition of the deep learning model here.
Rotating and flipping are standard operations to enhance the sample size as well as the robustness of the algorithm. A robust algorithm should be agnostic to orientation and vantage point. We increased the sample size of open-cell scenes to reduce the imbalance between cloud types.

Line 160-162: Was each scene in the training dataset labeled by at least two people? How if there is a disagreement?
Not every scene was repeatedly labelled by two experts, but there are hundreds of scenes that are labeled by at least two experts. When there is disagreement, an accompanying discussion can be found online. In these situations, we also examine the scenes closely to determine the true label.

Line 226-227: Please clarify on the "internal mechanisms". Are you referring to the self-organizing mechanisms (e.g., Feingold et al., 2010)? Feingold G, Koren I, Wang H, Xue H, Brewer WA. (2010): Precipitation-generated oscillations in open-cellular cloud fields. Nature 466:doi:10.1038/nature09314.
Here we are making broad separations between two camps, one advocating for large-scale forcing and the other for an internal mechanism. But yes, this paper would definitely count as supporting the internal mechanism hypothesis.

Line 330: More details are needed for the PDFs in Figure 4. How many scenes? What time periods and regions?
Good point. We added relevant information in the revision. We randomly selected 1000 scenes for each cloud type from 2006 data in the southeast Pacific region.

Line 78: No more than 10% of the scenes got filtered out? Please clarify.
Changed to clarify. We remove scenes with more than 10% land cover.

Line 117: classify -> classifying Line 146: remove the first "low"
Changed.

Line 311: units of LWP in Figure 2 are wrong.
Nice catch! Changed.