Comment on amt-2021-123
Annika Lauber (Referee)

The paper “The University of Washington Ice-Liquid Discriminator (UWILD) improves single particle phase classifications of hydrometeors within Southern Ocean clouds using machine learning” by Atlas et al. introduces a random forest algorithm referred to as UWILD for the discrimination of ice and water particles recorded with a 2D-S. It is shown that UWILD outperforms two other existing classification algorithms and the strengths but also weaknesses of the algorithm are discussed in detail. Accurate phase discrimination is important to draw conclusions from measurements on research questions. I support publishing this paper in AMT and only have a few comments, which should be addressed before the final publication.

1) Independency of the test set

For training and testing the algorithm, the authors used liquid-only and ice-dominated periods because hand-labeling the data, which introduces a human bias, can be avoided like this. I generally think that this is a good approach. However, the authors should be aware that the training and test sets are not completely independent like this. The particles were all measured during similar conditions (temperature, RH, one-phase). This would be only a minor issue for liquid droplets as they should always look the same and only the size distribution can change, which was partly considered by dividing the test set into different size ranges. However, ice crystal habit strongly depends on the temperature and RH but also on the liquid water content as riming as an effect as well. The conditions during the ice-dominated period are therefore not completely comparable to other periods. Therefore, I suggest making an additional test by hand-labeling a random sample from all other periods (not used for training) and compare the three algorithms, being aware that a human bias is introduced like this. However, it would be a more fair comparison between the different algorithms.

2) Page 5, first line (something went wrong here with the line numbers, which is why I
Lawson et al. (2006) only mention a 10µm pixel resolution and not an optical resolution. Please verify if "optical resolution" is the correct term.

3) Page 6, line 138

How can the measured voltage from RICE be used to identify flight periods where the hydrometeors are most likely of the same phase? The authors later say that ice-dominated periods are defined as having no RICE response but do not explain why that is the case.

The word “from” is repeated two times after each other in the same line.

4) Page 8, line 213 "This variable would be more difficult to incorporate in an image-based deep learning model."

To my knowledge, it is not difficult and uncommon to add variables to image-based deep learning models.

5) Page 11, line 307: "Holroyd and UWILD classify too many medium-sized (0.1 mm < D_{eq} 3 mm) and large (D_{eq} >0.3 mm) particles as ice at warm temperatures..."

Holroyd and UWILD classify too many medium-sized particles compared to what?

6) Page 14, lines 399-405 and Figure 10:

In the text and the description of the figure, the solid lines are referred to as the bootstrapped median and the dashed lines as deterministic. However, if I understand the figure correctly, it should be the other way round. Furthermore, the text refers to a,b,c... but the different plots in the figure are not labeled with letters.

7) Page 17, Line 520-522 “Large droplets are necessary for Hallett-Mossop rime splintering and droplet freezing and estimating their concentrations from liquid classifications of 2D-S images can aid in understanding the ice-production process.”

I am not sure if one can refer to “the ice-production process“ as there are several ice-production processes and the processes described in this sentence are secondary-ice-production processes.