# Identifying cloud phase from SOCRATES data using a random forest classification model

Jeremy Lu, Ian Hsiao , Hans Mohrmann, Rachel Atlas, Joe Finlon, and Rob Wood
University of Washington


## Summary

The **S**outhern **O**cean **C**louds, **R**adiation, **A**erosol **T**ransport **E**xperimental **S**tudy (SOCRATES) collected in-situ data on the clouds in the Southern Ocean (SO) region, in order to gain a better understanding of their impact on atmospheric and ocean processes and to inform climate model development. SO mixed-phase clouds contain supercooled liquid water and ice particles simultaneously and existing algorithms struggle to identify the phase of SO cloud particles, so we took a machine learning approach. Using Two-Dimensional Stereo (2D-S) optical array probe data and 15 parameters computed from that data, we built a random forest classification model that classifies cloud particle phase with over 90% accuracy for large particles.


## Motivation

Existing phase discrimination methods are unreliable, having a tendency to misclassify large liquid particles. This is problematic because larger particles are responsible for a majority of SO cloud mass. Our aim is to make a more accurate algorithm to classify cloud particle phase from particle images.

## Data

**2DS optical array probe** (Figure 1)

- Imaging using diode lasers; Nominal size range: 25–1280 µm; 10 µm resolution
- Collected data on 15 flights over SO region (Figure 2)

**Figure 1**
Image of 2DS Instrument

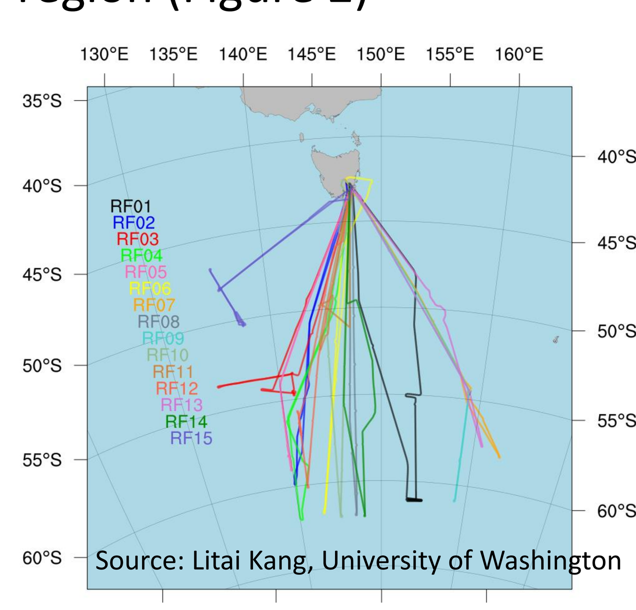


**Figure 2**
SOCRATES Flight Tracks

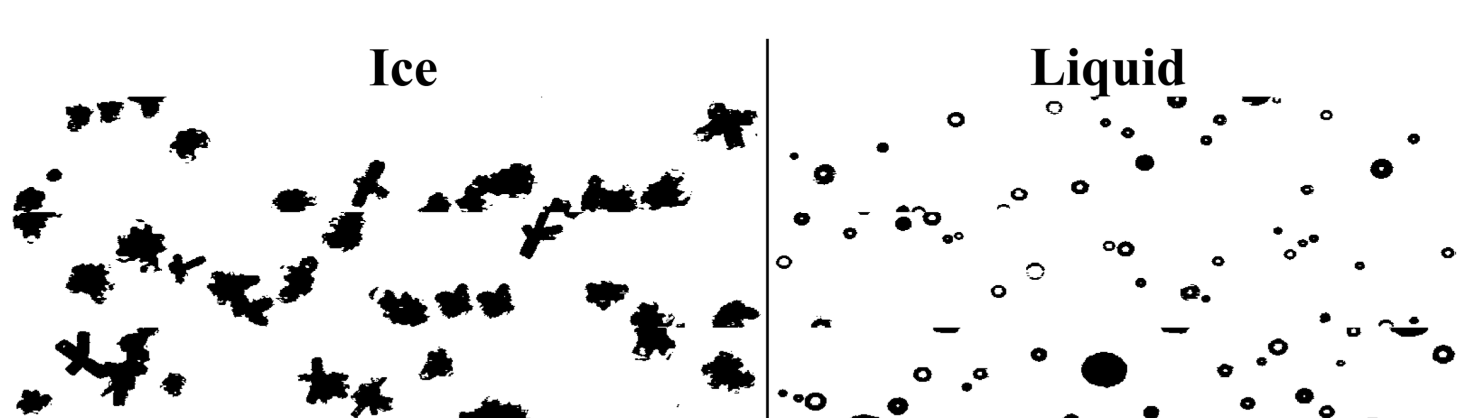


**Figure 3**
Shadow images of Ice and Liquid data collected during SOCRATES

## Model Architecture

- Used segments from three flights (RF01, RF03, RF04) for ice, and from one flight (RF05) for liquid training/validation/test data
- Total of 52,000 ice and liquid particles each and a 60/20/20 training/validation/test split
- Keep ice and liquid ratios in three size classes (25-100 pixels, 100-700 pixels, 700+ pixels) same as total data distribution in the flights

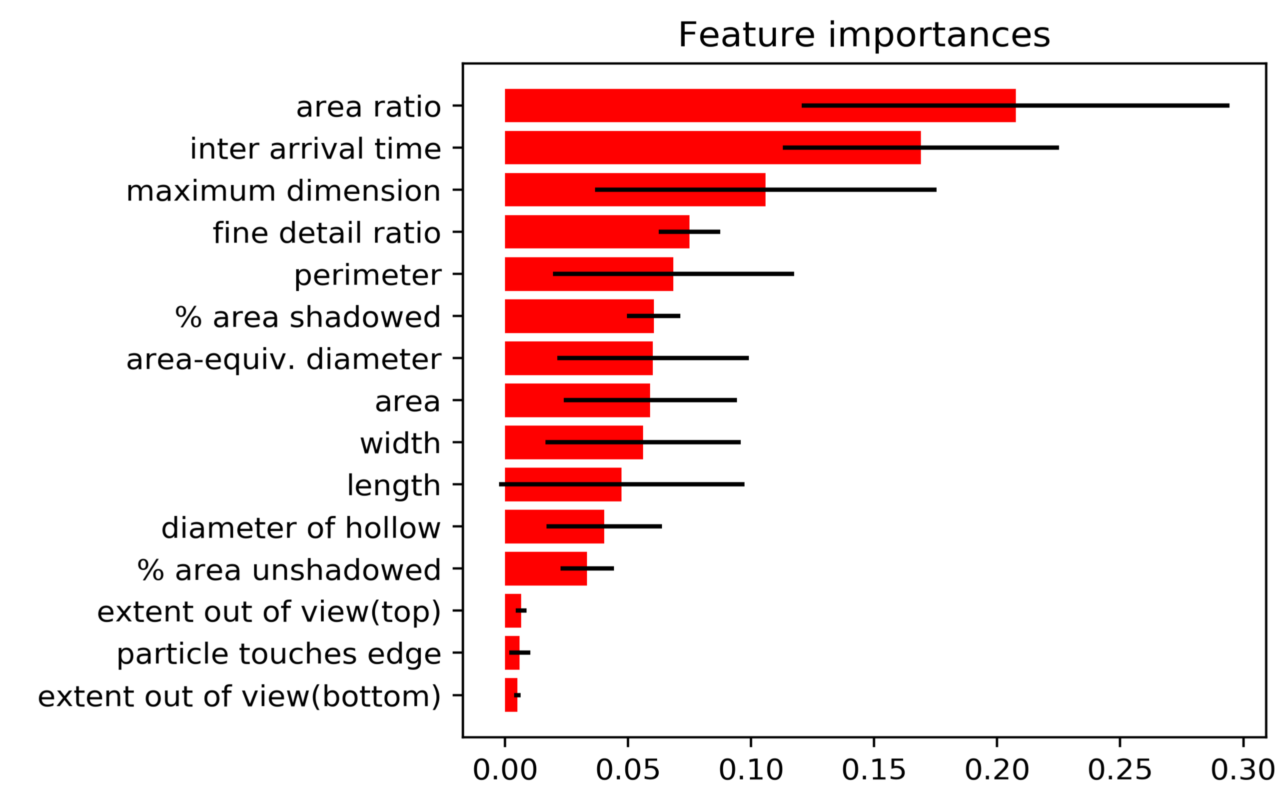


**Figure 4**
Feature Importance graph shows which parameters have biggest impact for random forest decisions

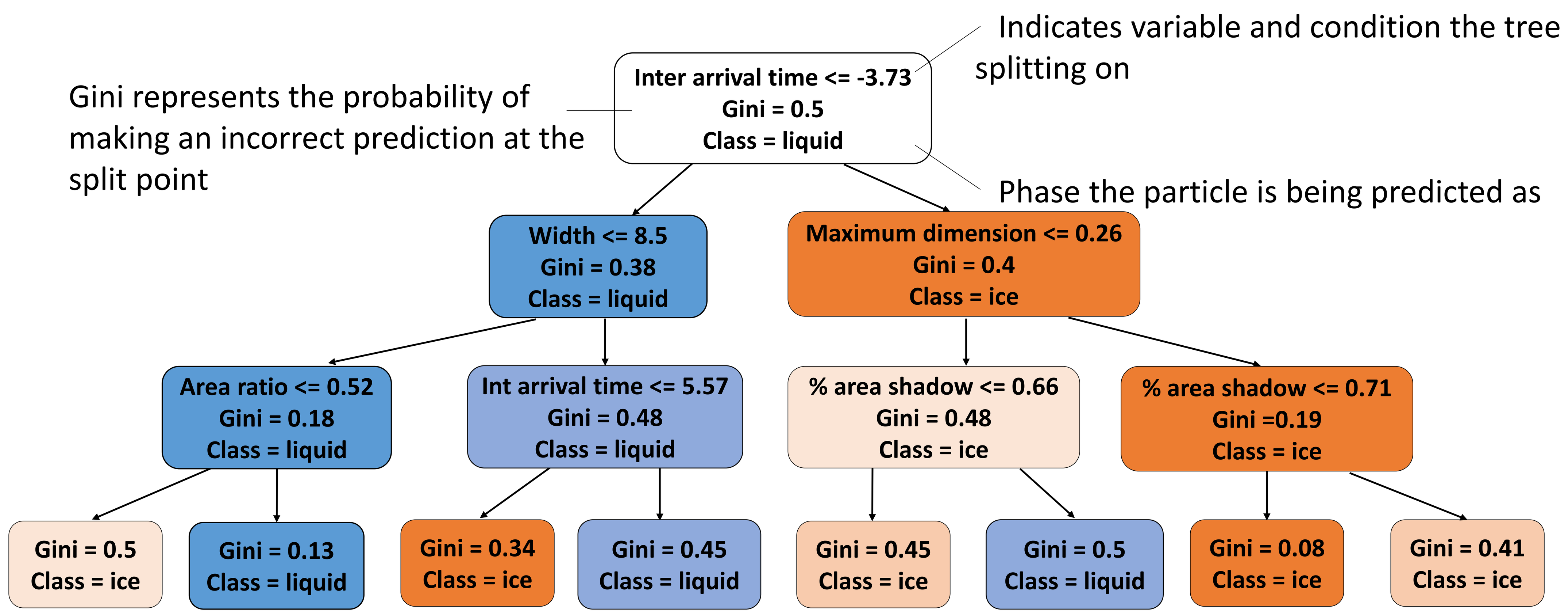


**Figure 5**
Example of single decision tree classifier with depth 3

## Model Accuracy and Evaluation

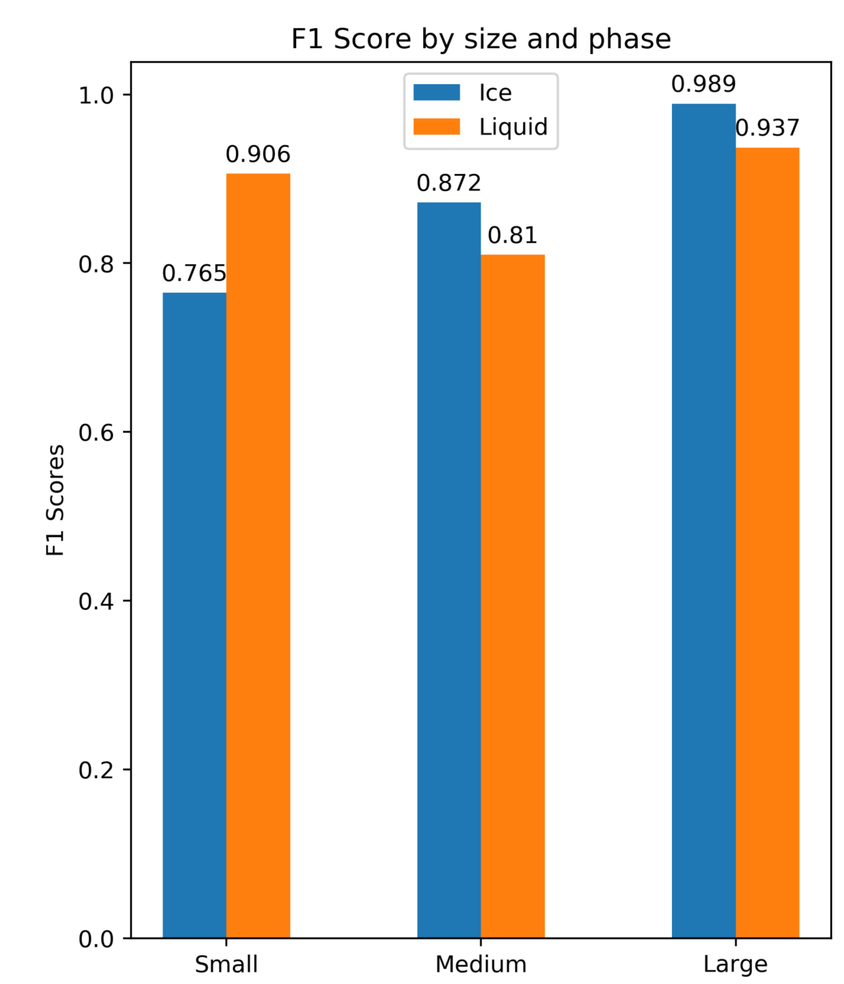


**Figure 6**
F1 scores show that the model does very well classifying large particles and is less accurate with small ice particles

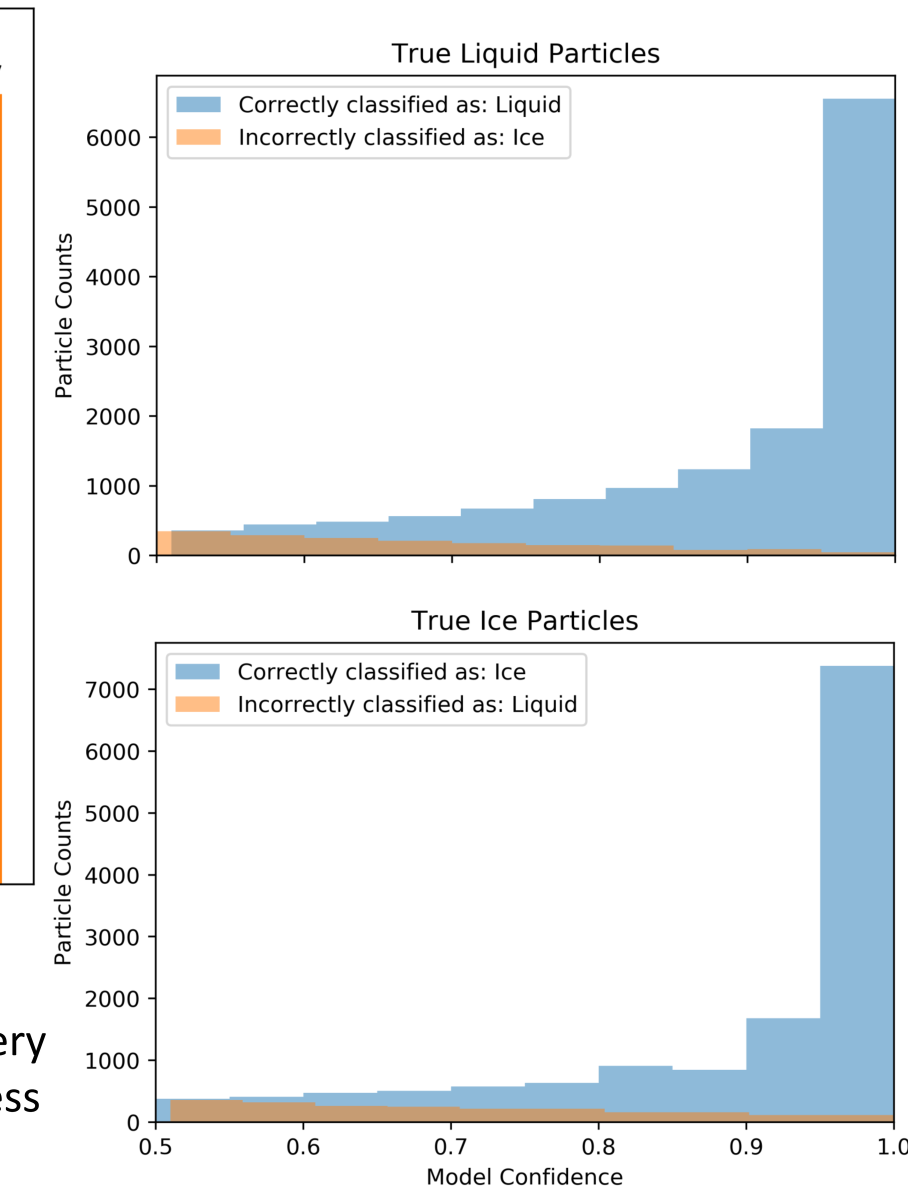


**Figure 7**
Model confidence is higher for correctly classified particles (blue) and lower for incorrectly classified ones (orange).

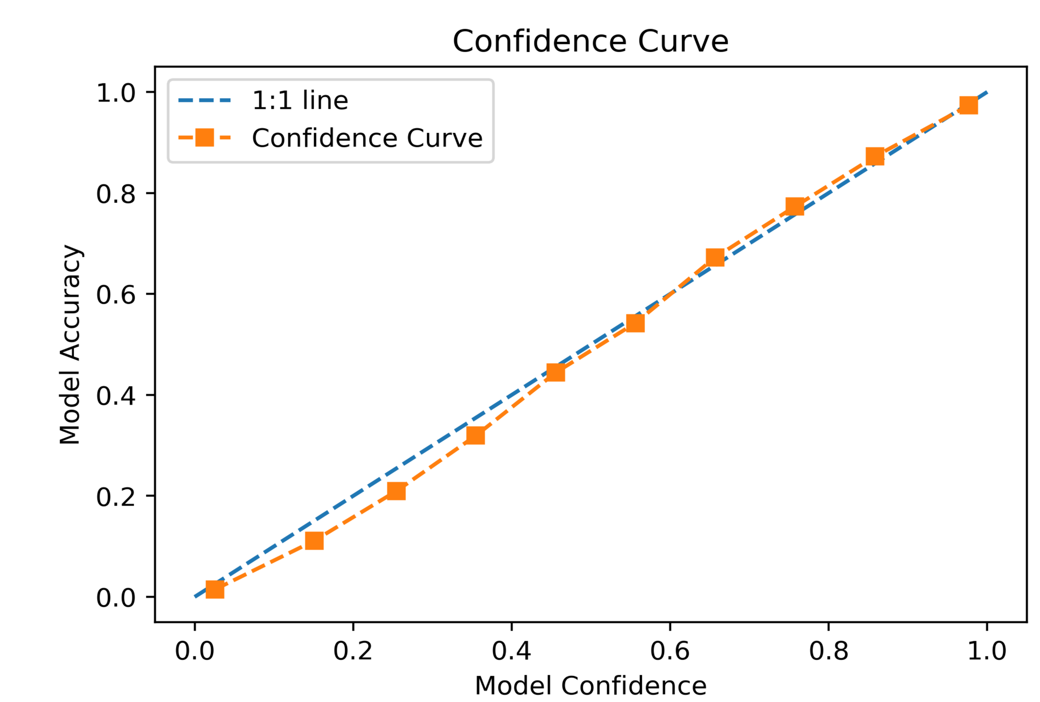


**Figure 8**
Model confidence and accuracy has a 1:1 relationship, which means we can use model confidence to estimate the uncertainty of its predictions

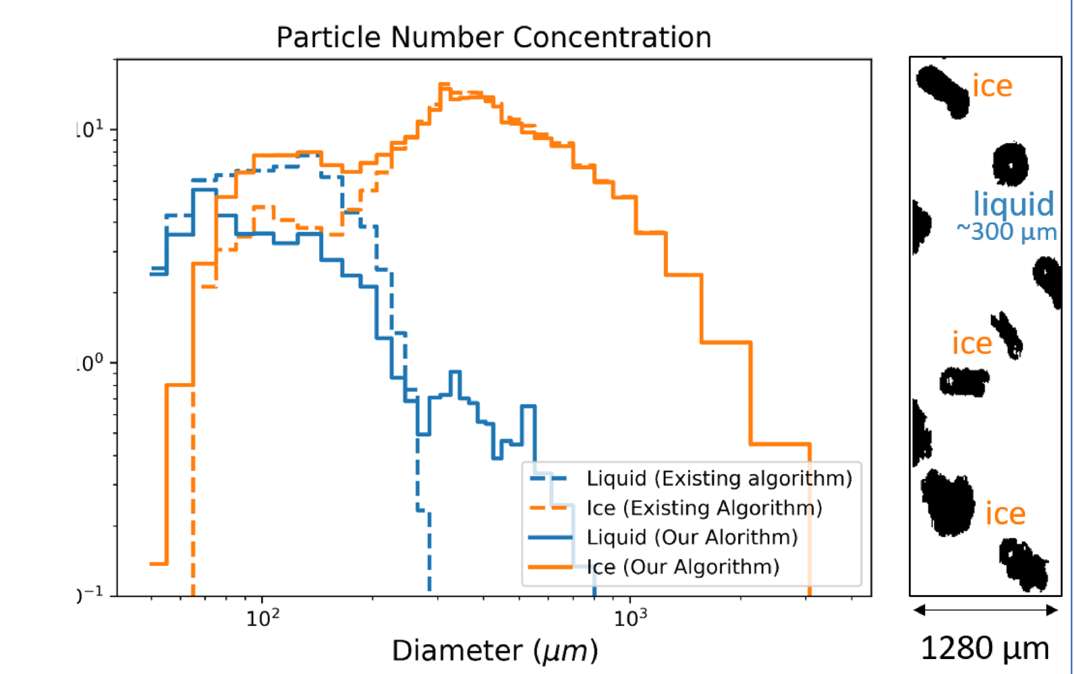


**Figure 9**
Particle size distributions of liquid and ice from a 5-minute period show that ML model predicts large liquid particles better than existing models

## Future Work

- Eliminate atypical particles in the dataset which may be reducing the accuracy of the model (such as splattered droplets)
- Tune model by trying it out with ranges of different hyperparameters
- Try different learning models (deep learning, gradient boosted tree, etc...)