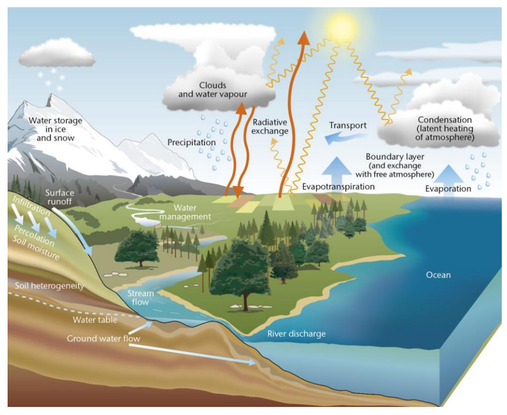
Improving Machine Learning Predictions of Thermodynamic Cloud Phase

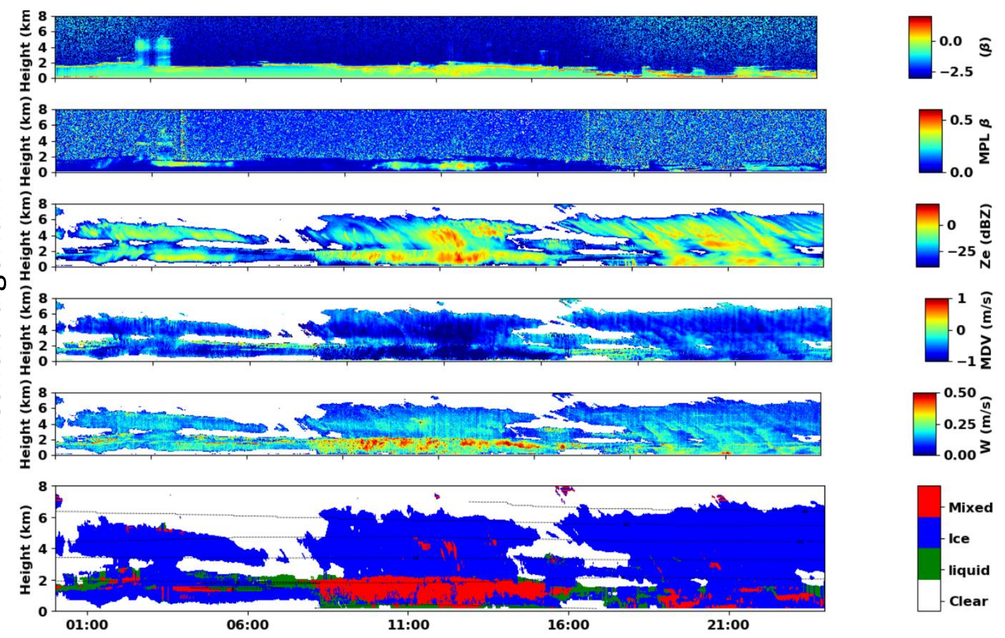
Harris, Carlandra*; Levin, Maxwell**; Zhang, Damao**; Goldberger, Lexie**
*Alabama State University
** Pacific Northwest National Laboratory

Introduction

Accurately predicting cloud phase is crucial for refining cloud models and understanding Earth's energy balance. The DOE's Atmospheric Radiation Measurement (ARM) facility produces a value added product (VAP) called 'thermocouldphase'. This VAP takes input from multiple instruments and using a math-based threshold algorithm outputs eight possible cloud phases: liquid, ice, mixed-phase, drizzle, liquid-drizzle, rain, snow, & unknown. The algorithm occasionally produces not-physically-possible errors. We correct algorithmic errors due to instrument or algorithm failures and apply a random forest model to evaluate if these corrections improve cloud phase prediction accuracy.

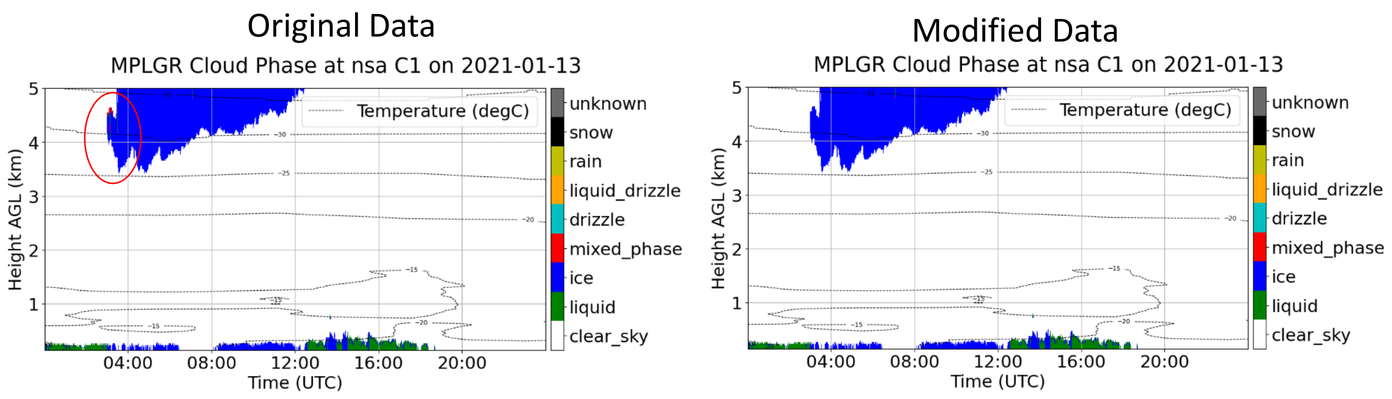


Above: Diagram of earth's energy budget. Clouds can have a warming or cooling effect on the planet
Right: raw data from instruments(LIDAR, RADAR, microwave radiometer, temperature from a radiosonte)



Hand Labeling Data

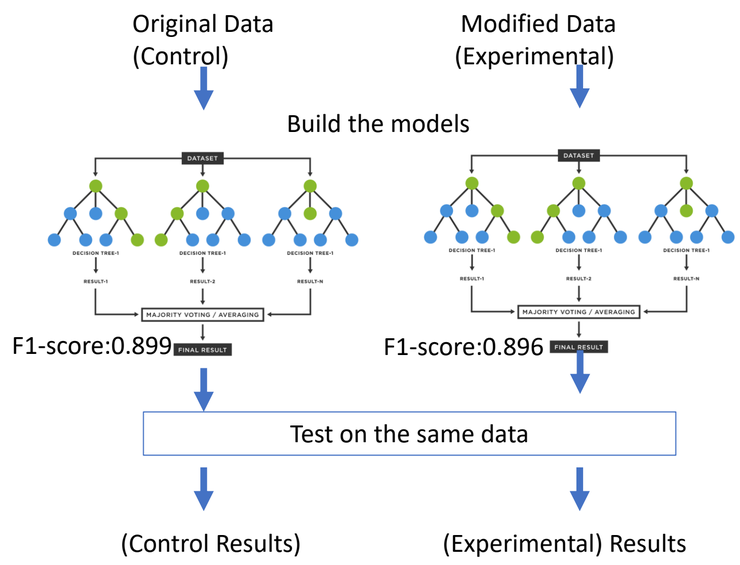
Hand labeling is the process of manually correcting data. We manually corrected two months worth of 'thermocloudphase' data taken from the North Slope of Alaska ARM measurement site. Some data could portray abnormalities, but most of those are derived from the algorithm.



Identifying errors in the algorithm and correcting them improved the thermocloudphase data
Left: highlighted is what correction needed to be made using hand labeling
Right: corrected data

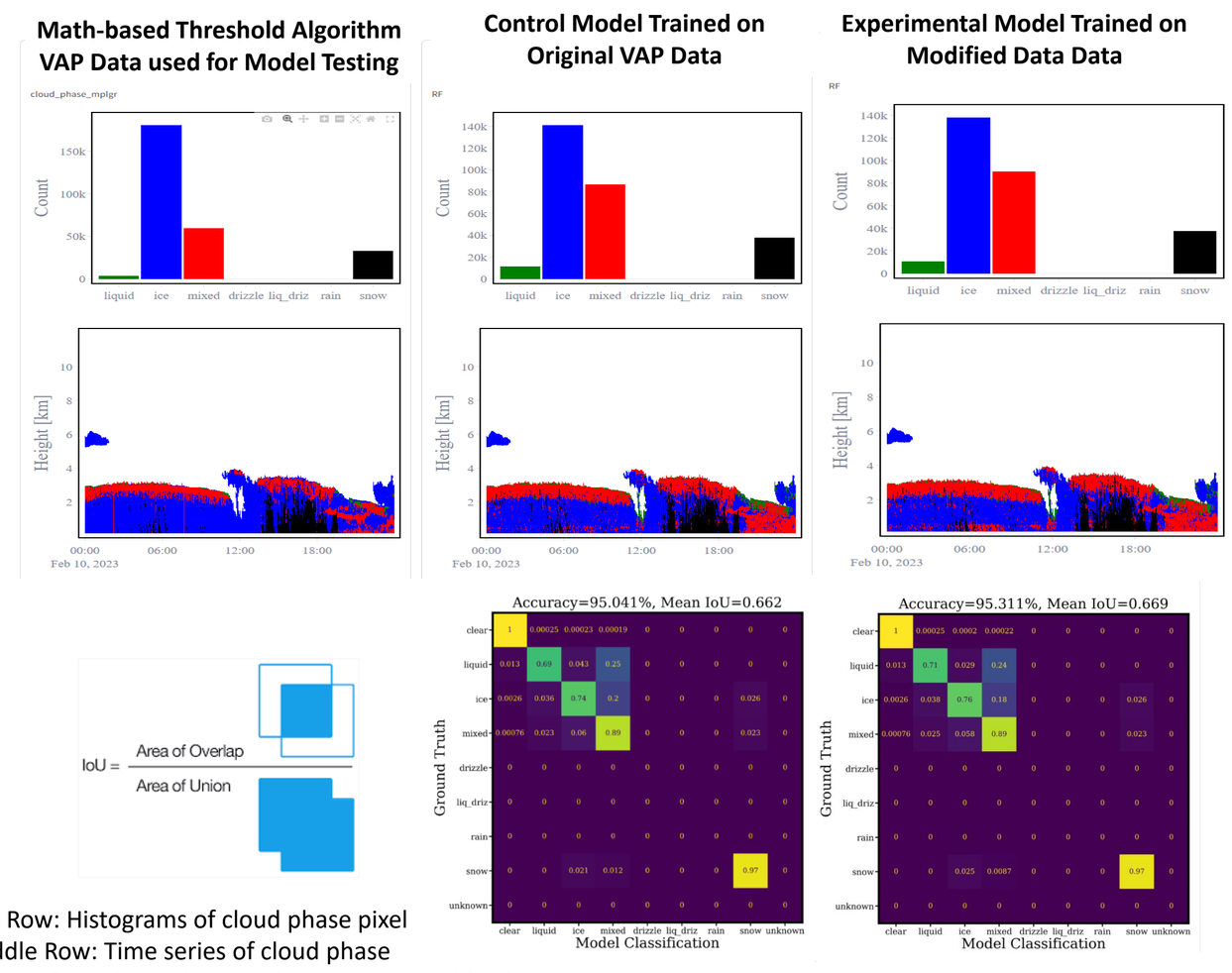
Experimental Design

We trained 2 random forest models one with original data and one with hand labeled data, then tested each model on 8 days worth of data. During model building we used 40,000 datapoints for training, 10,000 for validation, but was not enough datapoints for the rarer cloud phases like drizzle. For future work we would hand label more data to use when building the model.



Results

- Model accuracy and mean iou reported for each model for our 8 test days
- Experimental model preformed better
- The mean IOU shows how well the machine preformed when producing data(number>.5 were considered successful
- Last plot is a confusion matrix that shows how well the model did in predicting each cloud phase
- Larger values along the diagonal show good comparison



Top Row: Histograms of cloud phase pixel
Middle Row: Time series of cloud phase
Bottom Row: Confusion matrix of model classification
Bottom Left: Visual representation of the intersection over union(iou)
Left Column: Original ARM data used in testing
Middle Column: control model random forest results
Right Column: experimental model random forest results
confusion matrix = Table that visualizes the performance of a classification algorithm in machine learning between the ground truth and the model classified results

Model Performance on Testing Data

Date	Experimental Model Accuracy	Experimental Model Mean IOU	Control Model Accuracy	Control Model Mean IOU
February 10th, 2023	95.041%	0.662	95.311%	0.669
February 11th, 2023	97.433%	0.659	97.353%	0.654
February 12th, 2023	99.868%	0.465	99.853%	0.456
February 13th, 2023	99.311%	0.516	99.308%	0.508
February 14th, 2023	99.039%	0.639	99.032%	0.639
February 15th , 2023	98.820%	0.770	98.893%	0.625
February 16th , 2023	97.933%	0.582	97.915%	0.580
February 17th ,2023	97.376%	0.628	97.489%	0.632

Model accuracy and mean iou reported for each model for our 8 test days.

Conclusion

- The model is only as good as the data is trained on
- Contribute to advancements in understanding Earth's energy budget through more accurate cloud phase predictions.

Contact Information

- charris@myasu.alasu.edu

For additional information, contact:
charris0250@myasu.alasu.edu
000-000-0000
charris0250@myasu.alasu.edu