Improving Satellite-Based Airglow Measurements Using Machine Learning Cloud Detection

Anh Phan, Pierre-Dominique Pautet, Dallin Tucker, Jiarong Zhang, Connor Waite, Ludger Scherliess, and Yucheng Zhao
Physics Department, Utah State University, Logan, UT

Background

The Atmospheric Waves Experiment (AWE) investigates how Earth's weather influences space weather through the study of atmospheric gravity waves.

The Advanced Mesospheric Temperature Mapper (AMTM) instrument on the International Space Station (ISS) captures airglow emissions originating at ~87 km altitude using narrow-band IR filters.

Cloud contamination reduces the accuracy of temperature measurements from these images.

Problem

Standard cloud maps detect all clouds, not just high-altitude clouds visible in the $Q_1(1)$ channel.

Manually labeling clouds is impractical because of the massive data volume (~1 image every 1.1 seconds).

Our Approach

Use machine learning (ML) to classify cloud vs. no-cloud in small image regions (segmentation).

Model uses temporally stacked inputs and a ResNet-50 backbone.

Workflow

This workflow outlines the full pipeline, from manual labeling to ML predictions and post-processing.

Labeling	Manually inspect satellite images to identify and mark cloud intervals
Creating Training Data	Generate training images from marked intervals, accurately representing cloud and no-cloud conditions
Training Model	Train the model using transfer learning with ResNet-50
Generating Predictions	Create Images for prediction and predict cloud presence in portions of each image
Post Processing	By projecting these portions' prediction, expand the prediction to other portions of each image
Viewing Results	Visualize the final processed data, highlighting cloud regions within each segmented image

Labeling, training, and prediction pipeline

Method Overview

Preprocessing

Each image is split into 45 rectangular boxes (15×3 grid).

Labels are manually assigned to central boxes (7,0), (7,1), and (7,2) in selected orbits.

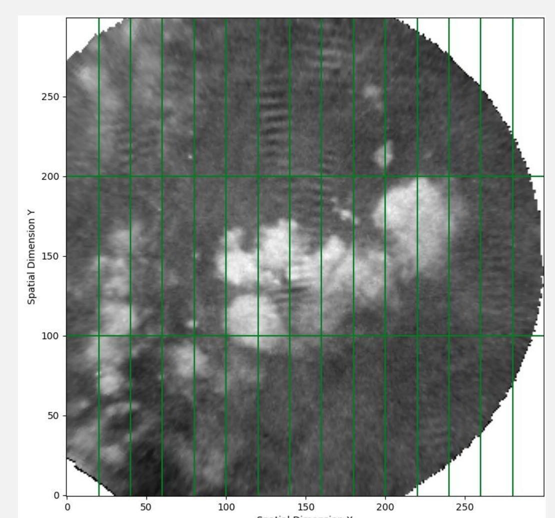
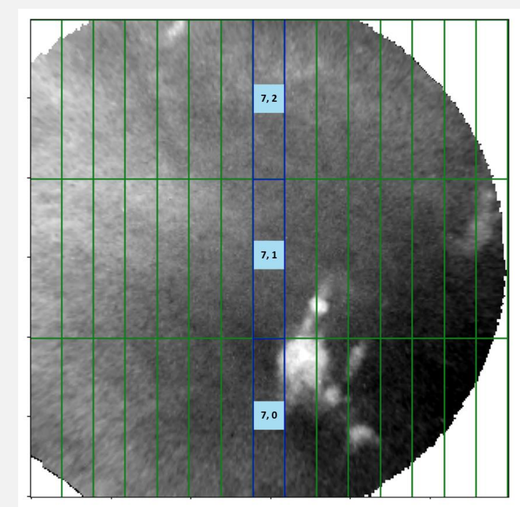
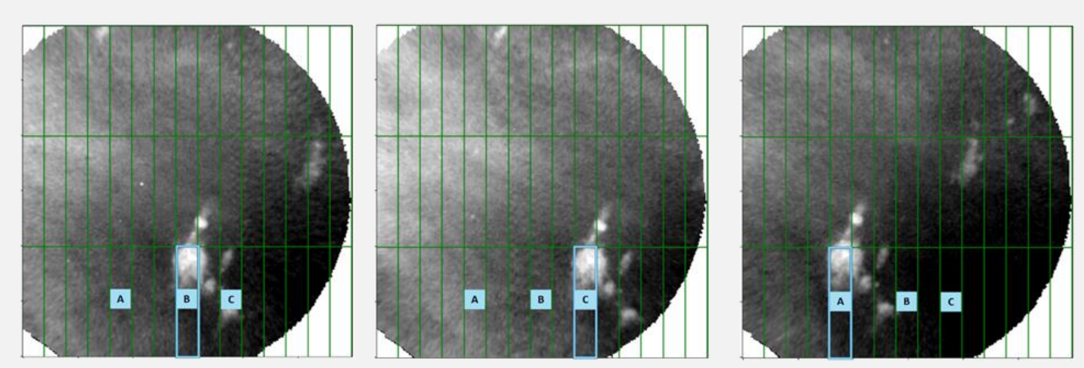


Image split into boxes for individual classification



Manual labeling for boxes (7,0), (7,1), (7,2)

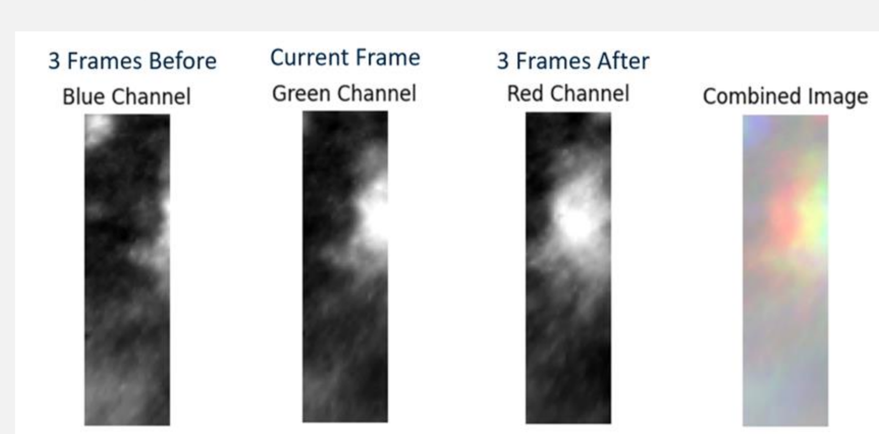
Clouds drift with ISS motion.
If Box B has clouds at time t, then Box A at t+19 and Box C at t–12 are also labeled.



Motion-based Augmentation: 3x3 boxes per frame

Model Input

RGB input created from 3 consecutive frames (t–3, t, t+3) per box.



Model input constructed from 3 time steps

Model Architecture

ResNet-50 with frozen layers extracts features.

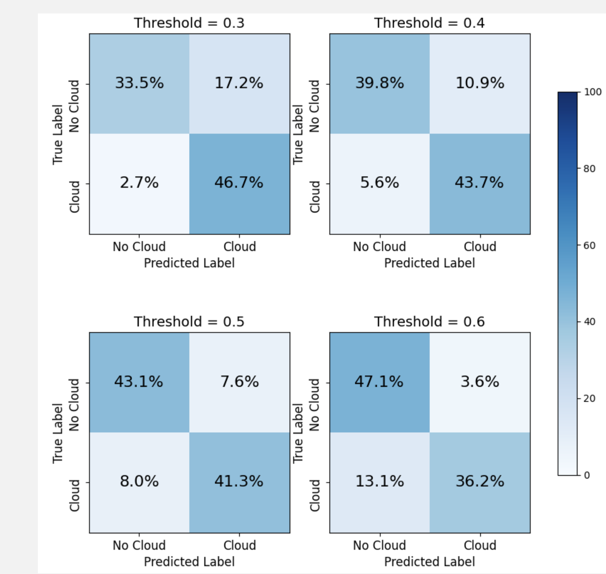
Two dense layers (512, 256) with ReLU and dropout (0.5).
Final sigmoid layer outputs cloud probability.

Lightweight, accurate, and easy to retrain on new data.

Confusion Matrices

At the default threshold of **0.5**, the model achieves **84.4% accuracy** (true positives + true negatives).

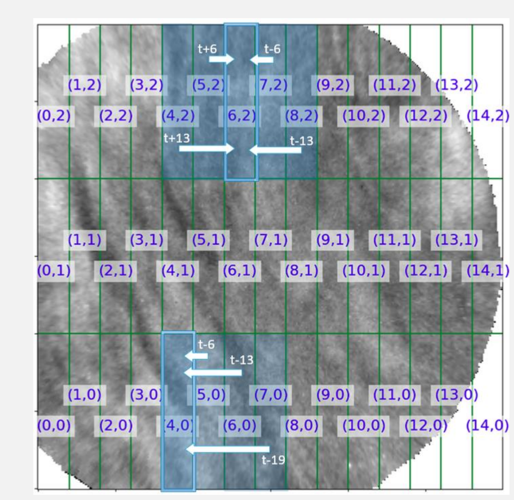
Lowering the threshold increases **recall** (better cloud detection) but also raises **false positives**.



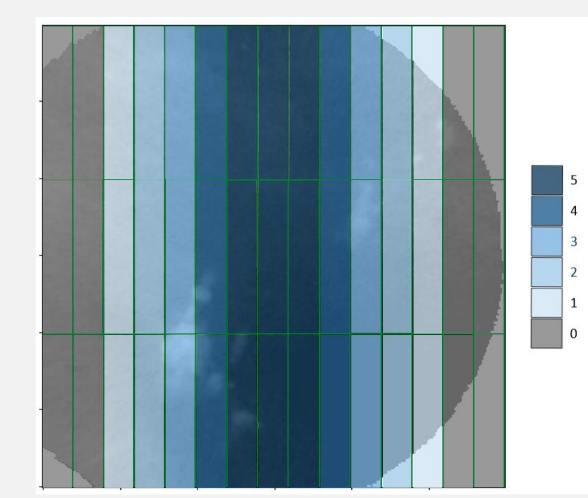
Confusion matrices for classification thresholds of .3, .4, .5, and .6.

Post Processing

Predictions from 15 boxes are averaged and propagated to cover 33 boxes.



Propagation and averaging strategy to refine spatial cloud classification.

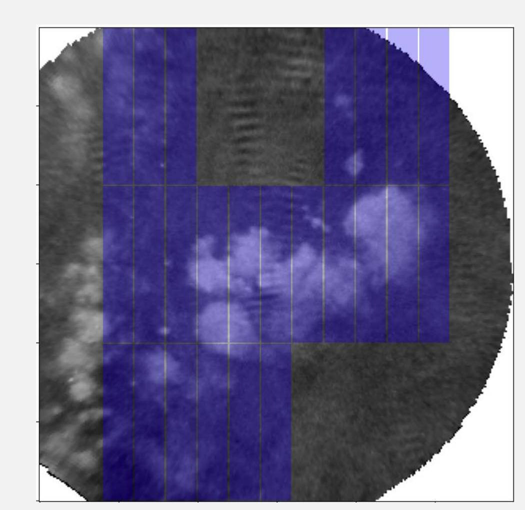


Number of boxes used for averaging. Central columns use up to 5 neighbors; side columns are excluded due to limited context.

Visualization

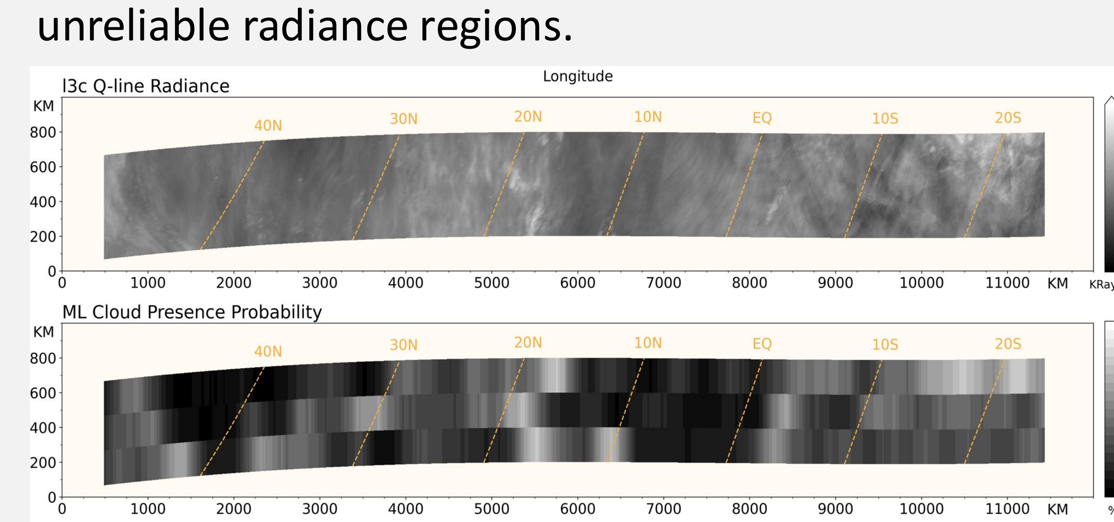
To assess model performance, we visualize cloud predictions for individual frames.

Boxes classified as **cloud** are highlighted in **blue**.



Segmented frame with cloud regions highlighted in blue.

Final cloud maps are written to NetCDF and used to flag unreliable radiance regions.



Orbit 82: Radiance swath (top) and machine learning cloud probability map (bottom). High-probability areas align with clouds. with cloud regions highlighted in blue.

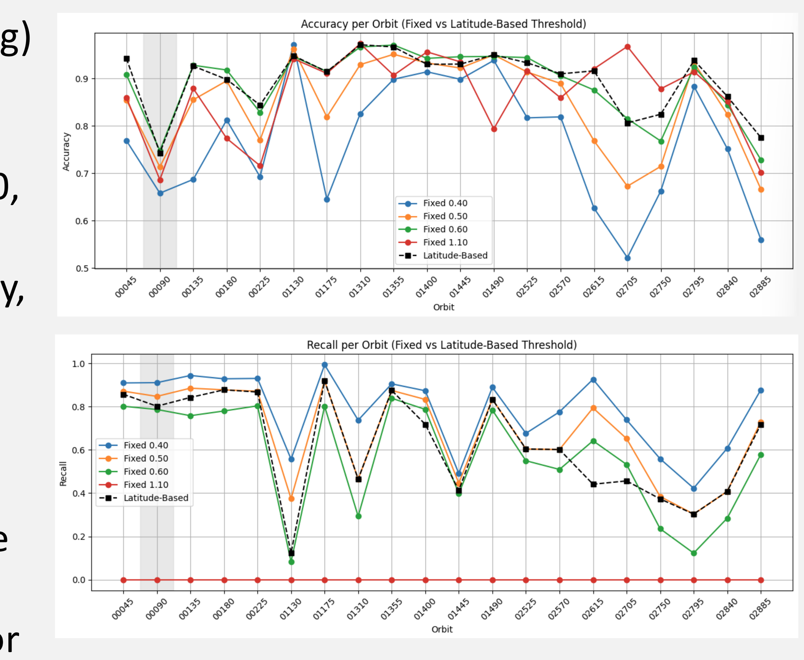
Model Evaluation

Tested on evaluation orbits every 45 steps.

Orbit 90 (in training) had high recall.

Orbits 45, 135, 180, 225 (unseen) performed similarly, showing generalization.

Later orbits had more variation due to changes in latitude, lighting, or cloud types.



Accuracy (top) and recall (bottom) across evaluation orbits. Later orbits show more variation.

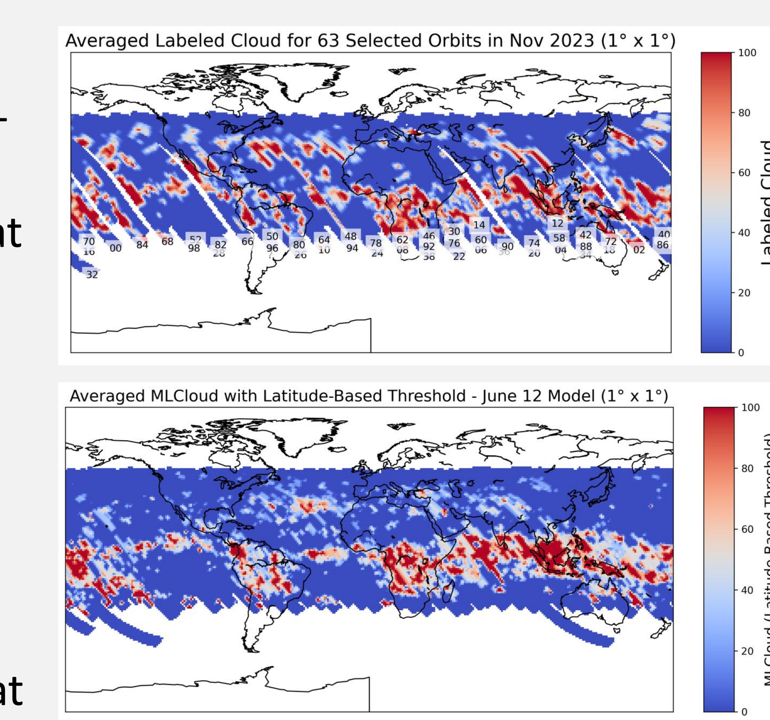
To improve generalization, we added orbits from **each month** into the training set, increasing exposure to diverse conditions.

Cloud Coverage Map

Compare expert labels vs. ML predictions.

The model matches human-labeled patterns well, especially at latitudes between −20° and 30°.

Latitude-based thresholding reduces cloud overestimation at higher latitudes.



Cloud coverage maps from expert labels (top) and ML predictions using a latitude-based threshold (bottom).

Conclusion

Our ML framework helps with cloud identification in AWE data, enhances temperature correction without discarding full images.

Scalable, interpretable, and integrable into the existing pipeline.

Future work: improve mid-latitude performance and expand to detect solar panels and ground lights.

Acknowledgement:
The material is based upon work supported by the National Aeronautics and Space Administration under Contract Number 80GSFC18C0007. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of NASA.