Machine Learning for Gravity Wave Identification

Wendy Carande[1], James Craft[1], Cora Randall[1], Jeff France[1], Lynn Harvey[1], Jim Russell[2], Scott Bailey[3], Justin Carstens[3]

[1]Laboratory for Atmospheric and Space Physics, [2]Hampton University, [3]Virginia Polytechnic Institute and State University

Why? To save time by automating gravity wave detection!

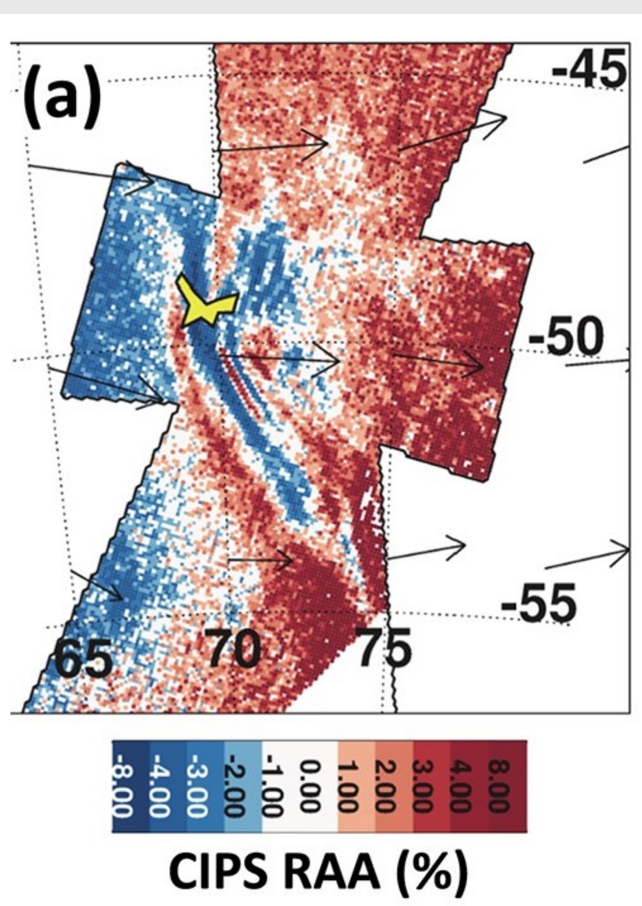
The Cloud Imaging and Particle Size (CIPS) instrument on board the Aeronomy of Ice in the Mesosphere (AIM) satellite was launched to measure polar mesospheric clouds (PMCs) in the altitude of 80 – 85 kilometers. Through development of new algorithms CIPS is now able to detect gravity waves (GWs) in the altitude of 50 - 55 kilometers in the new Rayleigh Albedo Anomaly (RAA) product. However, CIPS produces approximately 675 MB of data per day, which is a cumbersome amount of information for a human to process. We present an automated gravity wave detection algorithm using machine learning techniques and describe how this will be implemented in a production environment. The use of this automated gravity wave detection algorithm will save time and allow the researchers to focus on areas of interest rather than sifting through piles of data.

Data

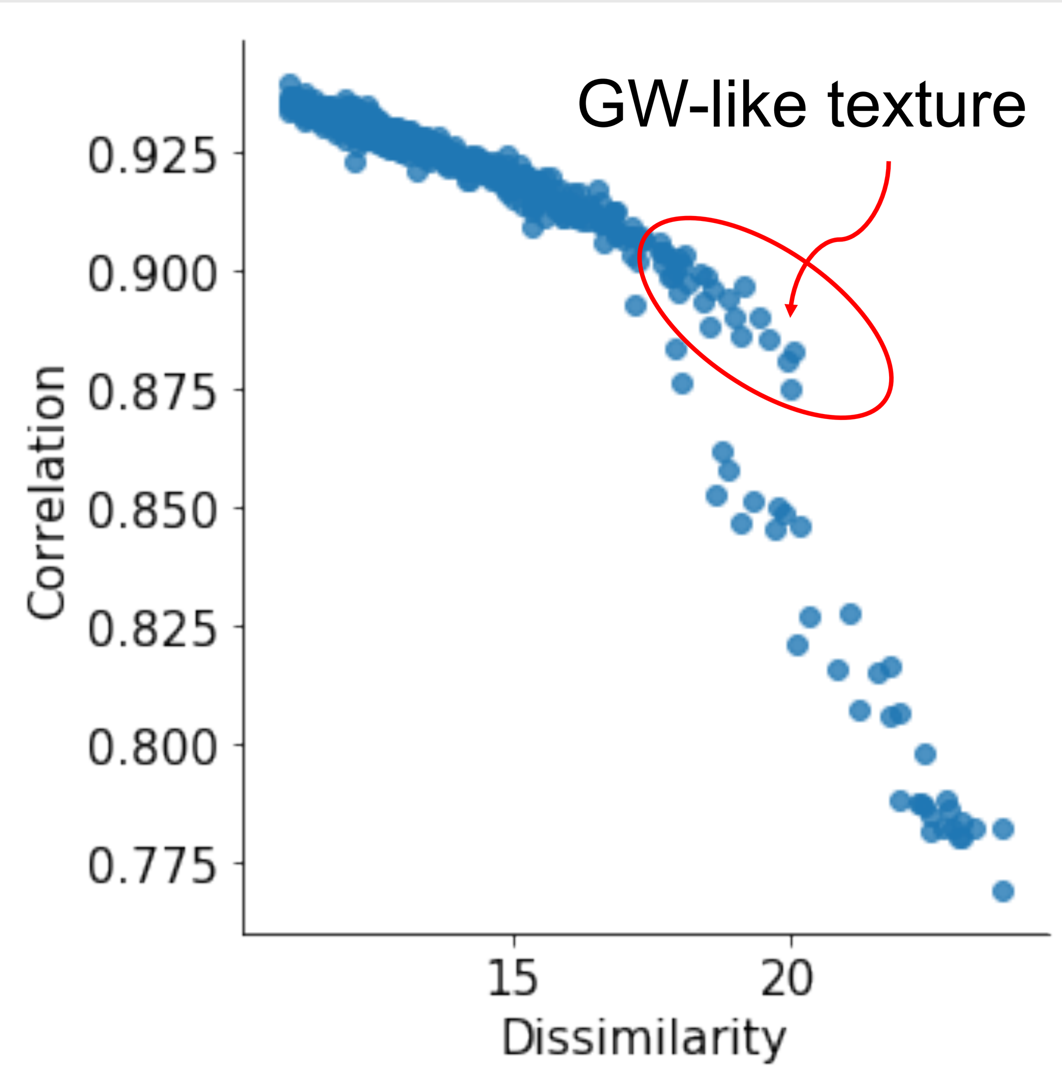
We use images created from level 2a CIPS RAA data products from 2016.

How? Step 1: Feature Engineering

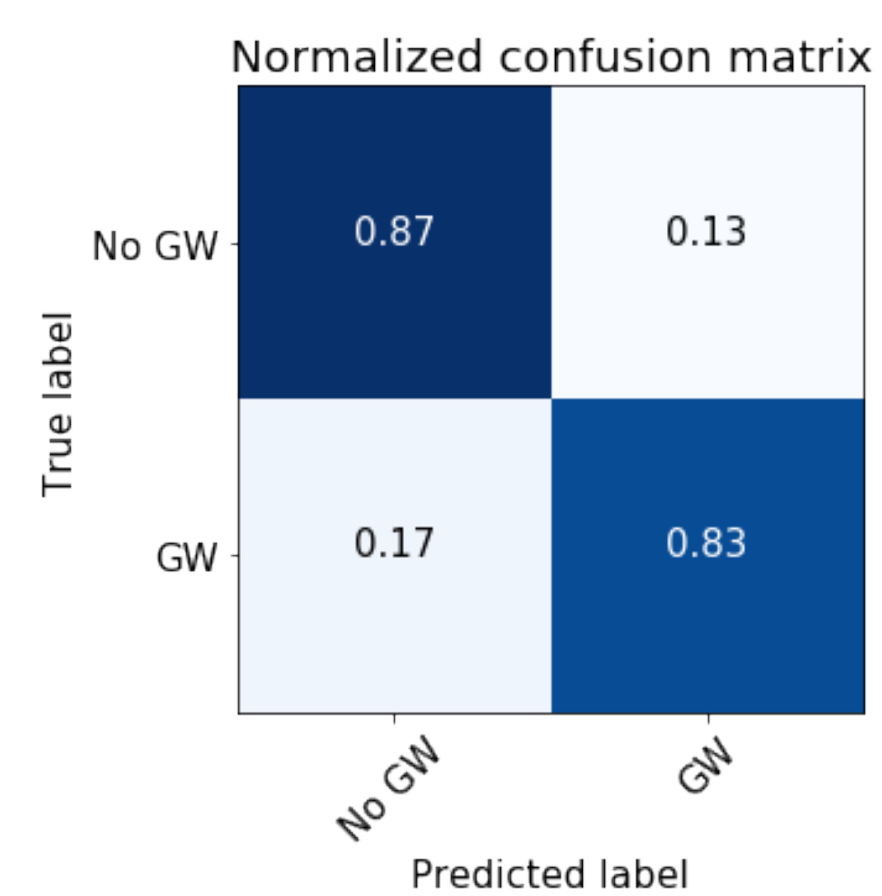
Gravity waves can be separated from "normal" RAA frames by texture analysis. Using the scikit-image package in Python, we calculate two features related to grey level co-occurrence matrices (GLCMs): **dissimilarity** and **correlation** (see plot below). This method also allows us to discern images with "bad pixels," which are filtered from the dataset.



Example of GWs in RAA data from Figure 2a of Randall et al. (2017).



Random Forest	
Training/Testing	70/30
# of Trees	100
Splitting Criterion	Gini impurity



How? Step 2: Random Forest

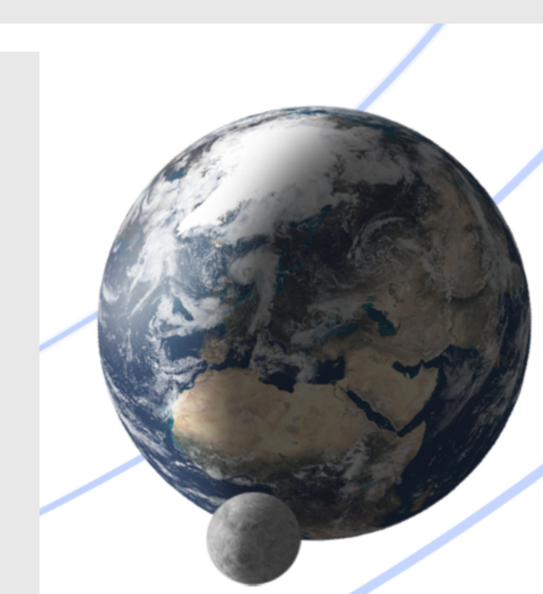
We use a Random Forest Classifier from scikit-learn to learn patterns in the training data. Gravity waves frames are less common than non-GW frames, making this an inherently imbalanced dataset. To correct for this imbalance, we use the Synthetic Minority Over-sampling Technique (SMOTE), which improves the identification performance of our classifier. The values we chose for training/testing split, number of trees in the forest, and splitting criterion are listed in the table to the left.

Does it work?

The confusion matrix to the left shows how the classifier performs on the test data. 83% of the time, we classify gravity waves correctly. This could be improved with additional features, better labeling, and more data.

Lessons learned…

- **Start simple:** The original concept for this work included the use of a neural network. However, the problem turned out to be much better suited to a simpler, more elegant solution.
- **Data preparation and exploration** account for 90% of the work.



Reference: Randall, C. E., et al. (2017), New AIM/CIPS global observations of gravity waves near 50–55 km, *Geophys. Res. Lett.*, 44, 7044– 7052.