Machine Learning in Analyzing Atmospheric Gravity Waves

Anastasia Brown
Utah State University Physics Department and Center for Atmospheric and Space Sciences
Kenneth Zia
Utah State University Physics Department and Center for Atmospheric and Space Sciences
P.-Dominique Pautet
Utah State University Physics Department and Center for Atmospheric and Space Sciences
Yucheng Zhao
Utah State University Physics Department and Center for Atmospheric and Space Sciences
Max Haehnel
Utah State University Physics Department and Center for Atmospheric and Space Sciences
Michel J. Taylor
Utah State University Physics Department and Center for Atmospheric and Space Sciences

UtahState University

Acknowledgments: This research project is supported by NSF grant 2029318.

Introduction and Background

Objective: **Improve Halley LightGBM Model** to better identify images of clear skies.

- The ANtarctic Gravity Wave Instrument Network (ANGWIN) is an international collaboration aimed at investigating the upper atmosphere dynamics over a continent-size region, using a network of all-sky imagers (Fig. 1).
- ANGWIN network began collecting All-Sky-Imager (ASI) data in 2012 (Table. 1). The ASI data is then sifted through to find windows of “clean” data (no clouds, aurora, or moonlight, Fig.2).
- Once found the “clean” image windows are processed and power spectrum graphs are made to identify wave activity (Fig. 3 and 4).
- In the middle of winter, each station collecting ASI data can easily produce well over 8,000 images a night.
- To streamline the sorting process, a machine learning algorithm is used to identify “clean” (marked as 0) images from “obscured” (marked as 1) images (Zia 2022).
  - Based on a machine learning algorithm that quickly sort through large Themis Aurora data sets reporting 96% accuracy (Clausen et al., 2018).
  - The algorithm we used: Light Gradient Boosted Machine (LightGBM)

Already, ASI data from 2 ANGWIN stations are sorted using LightGBM models. However, the model created to sort the third station, Halley, needs improvement:

The machine claims an accuracy of 99.2% but when validated, 62.3% of the “clean” windows identified were misidentified.

Airglow Images

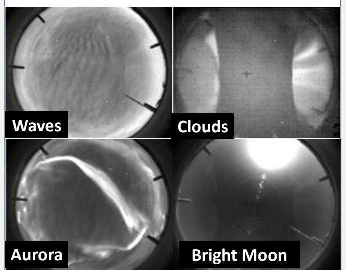

Fig. 2: Obstructions to the sky, such as clouds, aurora, bright moon, and fog will yield faulty power spectrum graphs.

Methods (Fig.6)

- Use current LightGBM model to identify “clean” windows.
- Validate the findings of the current model and identify strengths as well as areas that could be improved.
- Add frames depicting sky conditions the current model can not identify well to the training set to strengthen the next model.
- Evaluate training set and remove frames that confuse the computer.
- Increase the size of the training set and experiment with the ratio of clear and obscured images.
- Generated the next model with the new and improved training set.
- Compare each model’s ability to correctly label data from Halley, July 2012 (Fig. 4 and 5).
- Repeat until the machine learning algorithm effectively reduces the number of days that need to be processed.

Processed Airglow Images

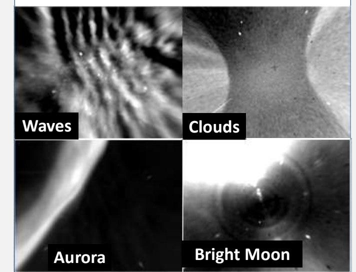

Fig 3: Airglow images are processed to remove noise and the fish-eye affect of the lenses.

USU ANGWIN Airglow Data

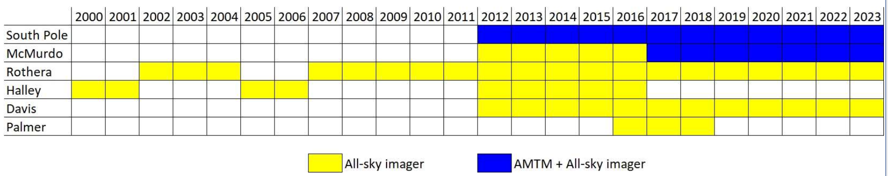

Table. 1: LightGBM is highly accurate and lightweight enough to use on a regular laptop (Ke et al., 2017). As a result, LightGBM models have been developed for four stations: Davis, McMurdo, Halley, and Rothera. Both the Davis and McMurdo models are fully functional and used to sort all sky data. The training set for the Halley model is under revision, and the Rothera model will be tested soon.

Results and Conclusion

- Use of machine learning in ASI cleaning removes the bottle neck created by the large data set.
- A more accurate model results a lower cost of analysis and a faster turnout. After much fine tuning, the 2nd, 3rd, and 4th iteration of the LightGBM Halley Model all preformed better than the 1st .

The 4th iteration surpassed the others in avoiding obscured windows while still correctly clear skies (Fig. 7).

ANGWIN Research Stations Over Antarctica

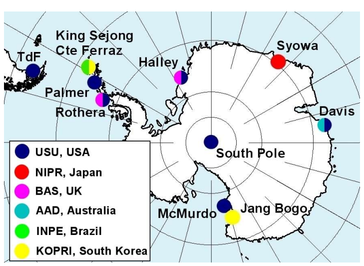

Fig. 1: International ANGWIN institutions and research sites.

Power Spectrums

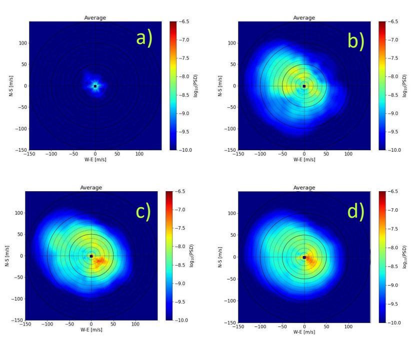

Fig. 4 Power spectrums from windows shown in figure 4. For comparison see a) Halley Model 1, Jul12-13 5:35 – 10:04 UT. b) Halley Model 2, Jul12-13 0:55 – 3:34 UT. c) Halley Model 3, Jul12-13 20:35 – 23:36 UT. d) Halley Model 4, Jul12-13 18:08 – 1:42

Improve Model

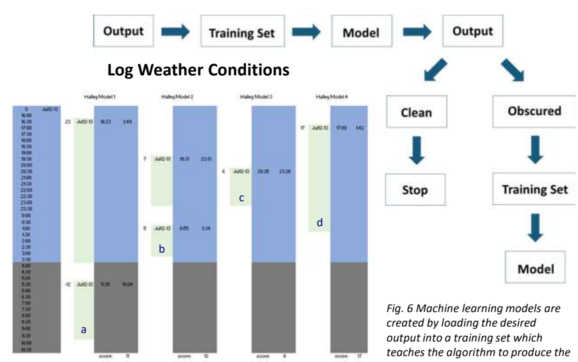

Fig. 5: The models’ ability to find clean windows was compared by grading each model’s capacity to find clean windows in July. Those points were then tallied to estimate the net number of “clear” hours each model identified.

Fig. 6 Machine learning models are created by loading the desired output into a training set which teaches the algorithm to produce the desired output. If the LightGBM models correctly identify “clean” data, the model is done. Else, if the models mislabel data, the faulty output is labeled as “obscured” and added to the training set to produce a new model.

Winner: **Halley Model 4**

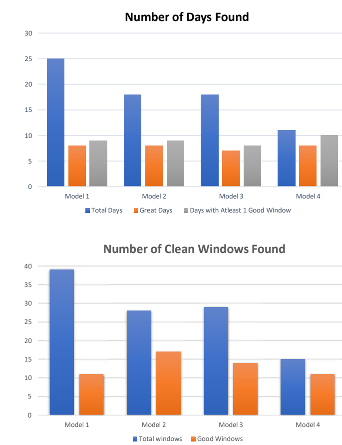

Fig. 7: Comparing the number of “clean” windows identified by each model to the number of clean windows each model correctly identified shows the validated accuracy of each model. Comparing the number of days that the models claimed to have “clean” windows to the number of days that truly had clean windows displays the amount of time and resources each model saves.

References:

Clausen, L. B. N., & Nickisch, H. (2018). Automatic classification of auroral images from the Oslo Auroral THEMIS (OATH) data set using machine learning. *Journal of Geophysical Research: Space Physics*, 123, 5640– 5647. https://doi.org/10.1029/2018JA025274

Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, Tie-Yan Liu. (2017) “LightGBM: A Highly Efficient Gradient Boosting Decision Tree“, *NeurIPS Proceedings.* *https://proceedings.neurips.cc/paper_files/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf*

Zia, Kenneth I., "Investigating Atmospheric Gravity Waves Using 3-Dimensional Spectral Analysis" (2022). All Graduate Theses and Dissertations. 8645. https://doi.org/10.26076/b507-e96e