Greenland Surface Mass Loss

Data Science Institute
COLUMBIA UNIVERSITY

Kailande Cassamajor, Ke Li, Meggie Wen, Mingyue Xu, Yuezhu Xu, Jiawen Zhou

Data Science Capstone Project
with Dr. Marco Tedesco,
Lamont-Doherty Earth Observatory

Introduction

Greenland is the world's largest island country located between the Arctic and Atlantic Ocean. Due to global warming, it has been melting unusually fast which can cause rising sea level and have serious environmental impact around the world. Our study focused on one of the result parameter related to surface mass loss problem, bare-ice albedo, which quantifies the energy needed for ice melting. We want to discover and examine the drivers of bare-ice albedo by conducting exploratory analysis and implementing machine learning algorithms.

Figure 1. Greenland Ice Sheet

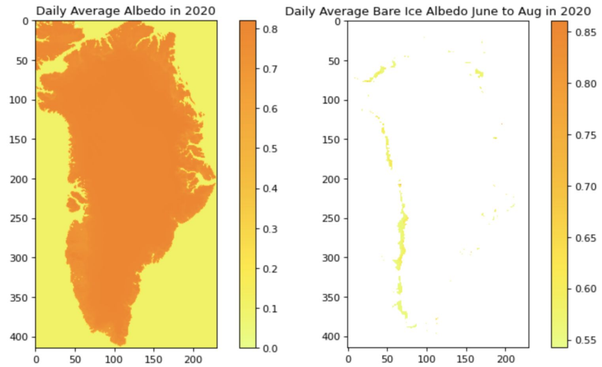

Figure 2. Daily Average Albedo in 2020

Data & Methods

Data: We only extracted bare ice pixels from June 1st to Aug 31st every year from 2000 to 2021. We selected 15 MAR (Modèle Atmosphérique Régional) variables to predict the MODIS (Moderate Resolution Imaging Spectro-radiometer) albedo. MAR variables include shortwave downward radiation (SWD), longwave net radiation (LWN), surface temperature (ST2), meltwater production (ME), snowfall (SF), rainfall (RF), sensible heat flux (SHF), latent heat flux (LHF), surface atmospheric pressure (SP), lower cloud cover fraction (CD), middle cloud cover fraction (CM), air temperature (TT), specific humidity (QQ), wind speed in x-direction (UU), and wind speed in y-direction (VV).

Data Pipeline: MAR, MODIS → reshape, sequence, log transform, standardize → model

Methods: Elastic Net, XGBoost, LightGBM, Random Forest, and LSTM.

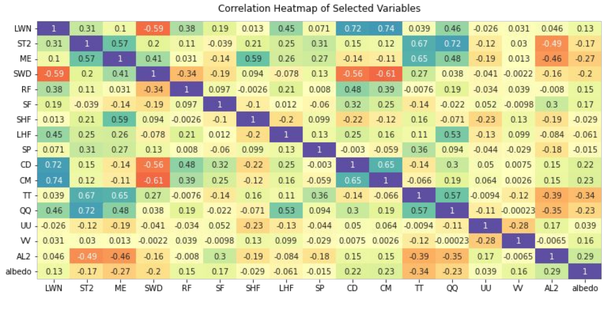

Figure 3. Correlation Heatmap

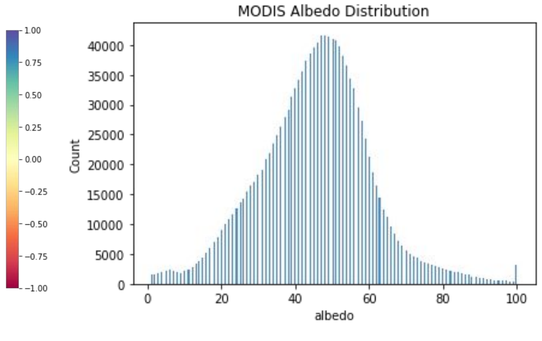

Figure 4. Distribution of Target Albedo

Results

The metric we used to evaluate model performance of this regression analysis is R^2 score. By its statistical definition, a low R^2 score means a low proportion of the variance of the target that is explained by the independent variables in the regression model.

In our results of comparing R^2 score on test dataset, Random Forest is the best model.

We examined feature importance to find possible factors contributing to changing albedo values. Figure 5 shows quantified feature importance of random forest model. Surface atmospheric pressure (SP) has the highest value and meltwater production (ME) has the lowest.

Model	R^2 score
Elastic Net	**0.21**
XGBoost with 10% preprocessed data	**0.34**
LightGBM	**0.28**
Random Forest with original dataset	**0.43**
Random Forest with 10% preprocessed data	**0.36**
LSTM	**0.30**

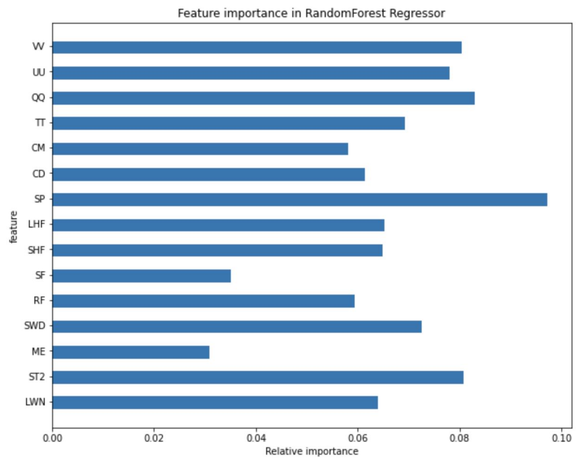

Figure 5. Feature Importance in Random Forest

Conclusions

After implementing appropriate models according to our exploratory results, we still have a relatively low R^2 score. The models share a problem of overfitting and underfitting. Further work may include reevaluating the performance metric, and improving understanding of the complex statistical relationships amongst variables. Noise in the data also affect the performance. Current computational limitations and model complexity can be improved in the future to raise model performance and by including more relevant MAR variables in the experiment would produce more interpretations for the topic.

Acknowledgments

We would like to thank our mentor Dr. Marco Tedesco and his PhD student Raf Antwerpen for their mentorship, expertise, and support throughout the progression of the project.

References

R. M. Antwerpen et al. "Assessing bare-ice albedo simulated by MAR over the Greenland ice sheet (2000–2021) and implications for meltwater production estimates". In: The Cryosphere 16.10 (2022), pp. 4185–4199. url: https://tc.copernicus.org/ articles/16/4185/2022/.

Jeremie Mouginot et al. "Forty-six years of Greenland Ice Sheet mass balance from 1972 to 2018". In: Proceedings of the National Academy of Sciences 116.19 (2019), pp. 9239–9244. url: https://www.pnas.org/doi/abs/10.1073/pnas.1904242116.