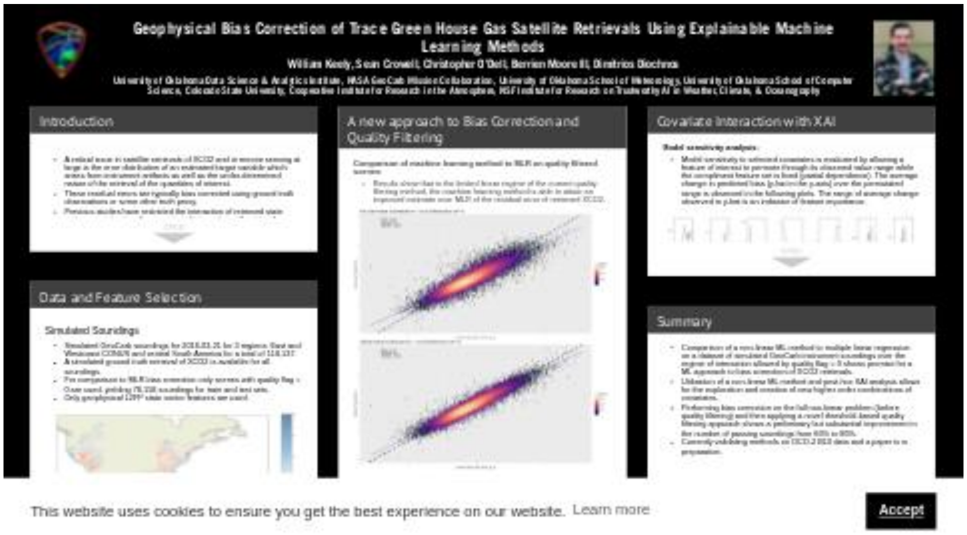
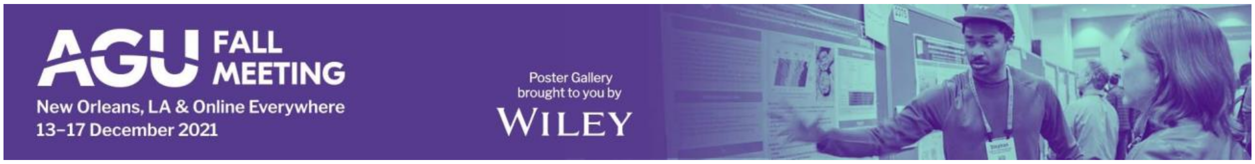
Geophysical Bias Correction of Trace Green House Gas Satellite Retrievals Using Explainable Machine Learning Methods



William Keely, Sean Crowell, Christopher O'Dell, Berrien Moore III, Dimitrios Diochnos

University of Oklahoma Data Science & Analytics Institute, NASA GeoCarb Mission Collaboration, University of Oklahoma School of Meteorology, University of Oklahoma School of Computer Science, Colorado State University, Cooperative Institute for Research in the Atmosphere, NSF Institute for Research on Trustworthy AI in Weather, Climate, & Oceanography

PRESENTED AT:



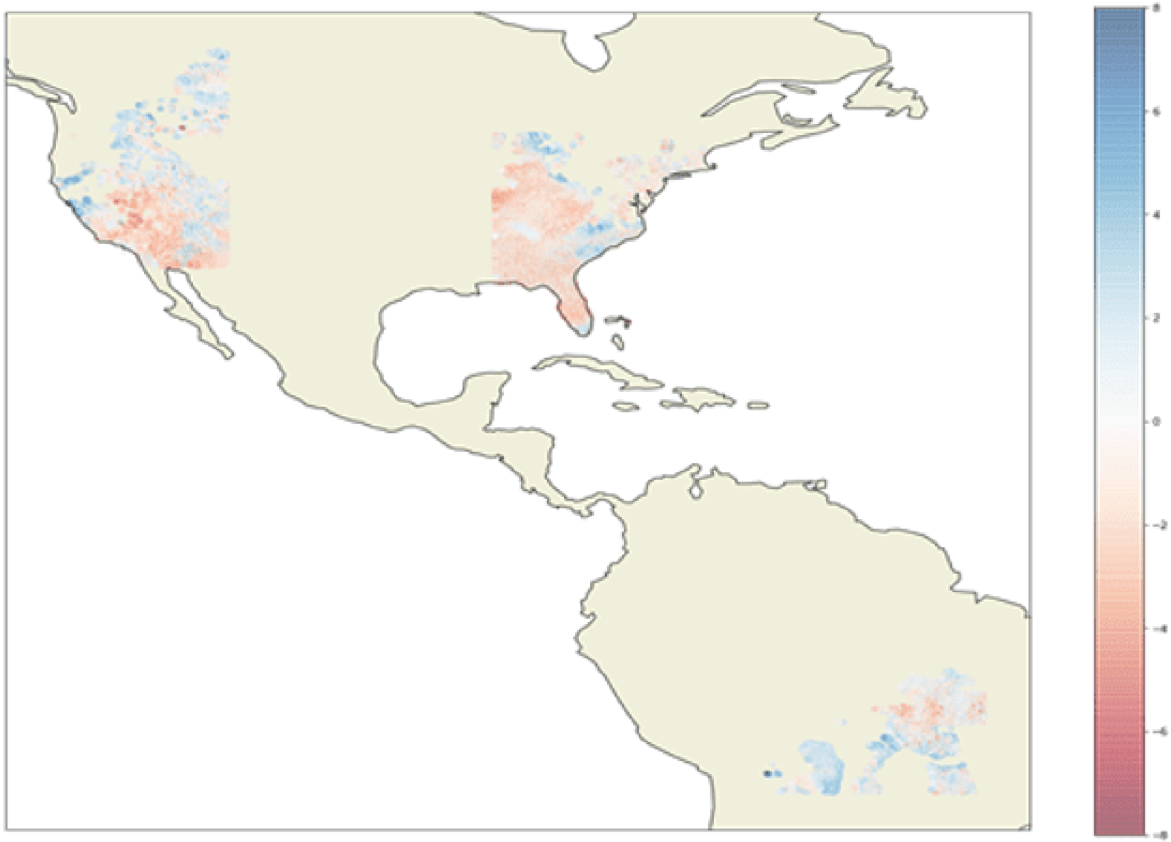
INTRODUCTION

- A critical issue in satellite retrievals of XCO2 and in remote sensing at large is the error distribution of an estimated target variable which arises from instrument artifacts as well as the under-determined nature of the retrieval of the quantities of interest.
- These residual errors are typically bias corrected using ground truth observations or some other truth proxy.
- Previous studies have restricted the interaction of retrieved state vector covariates to a linear regime of interaction with retrieval error and have employed a multiple linear regression (MLR) to model the error distribution (O'Dell et al. 2018).
- We employ an interpretable regularized gradient booster (XGBoost denoted in this presentation as XGB) that can capture potential non-linear interaction among the state vector covariates and is resistant to the effects of large outliers (Chen 2016). The machine learning technique is compared to an MLR model on a quality filtered simulated sounding dataset of a GeoCarb instrument.
- We also explore and propose a new method to quality filtering by opening the problem domain to the full non-linear problem and employing a novel bias thresholding approach.

DATA AND FEATURE SELECTION

Simulated Soundings

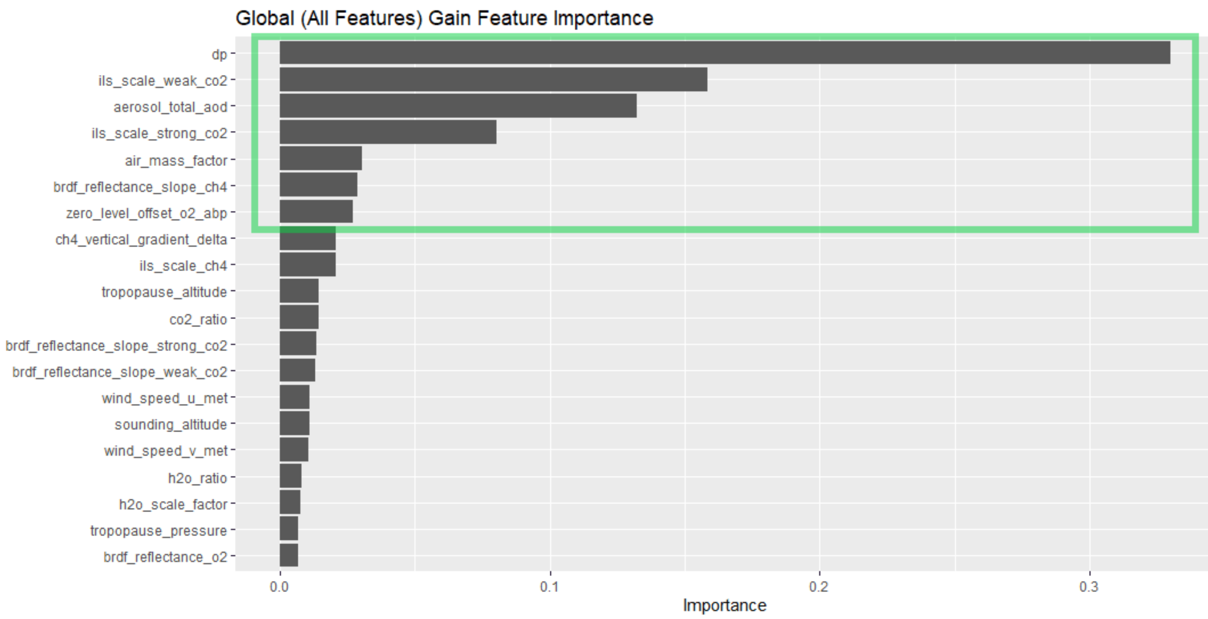
- Simulated GeoCarb soundings for 2016-03-21 for 3 regions: East and Westcoast CONUS and central South America for a total of 118,137.
- A simulated ground truth retrieval of XCO2 is available for all soundings.
- For comparison to MLR bias correction only scenes with quality flag = 0 are used, yielding 78,158 soundings for train and test sets.
- Only geophysical L2FP state vector features are used.



Soundings with True Bias ($XCO2_{raw}$- $XCO2_{true}$) & QF = 0

Feature Selection

- Initial XGB models constructed with large feature set for each region and quality flag = 0. Feature Importance is then calculated by percentage *information gain*.
- Feature importance varies by region, this is also seen in the correlation between covariates and bias for each region.
- A final subset of features that contribute a high percentage of information gain and correlation to bias are selected for a final *global* model. Final XGB and MLR models use the same subset of features.

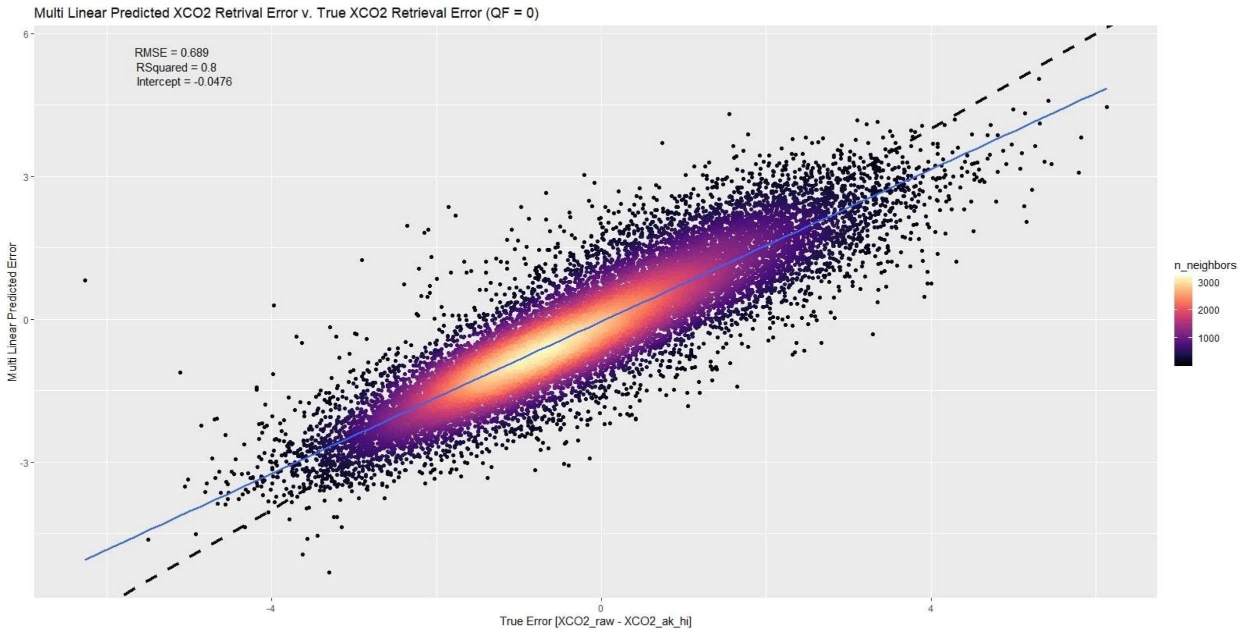


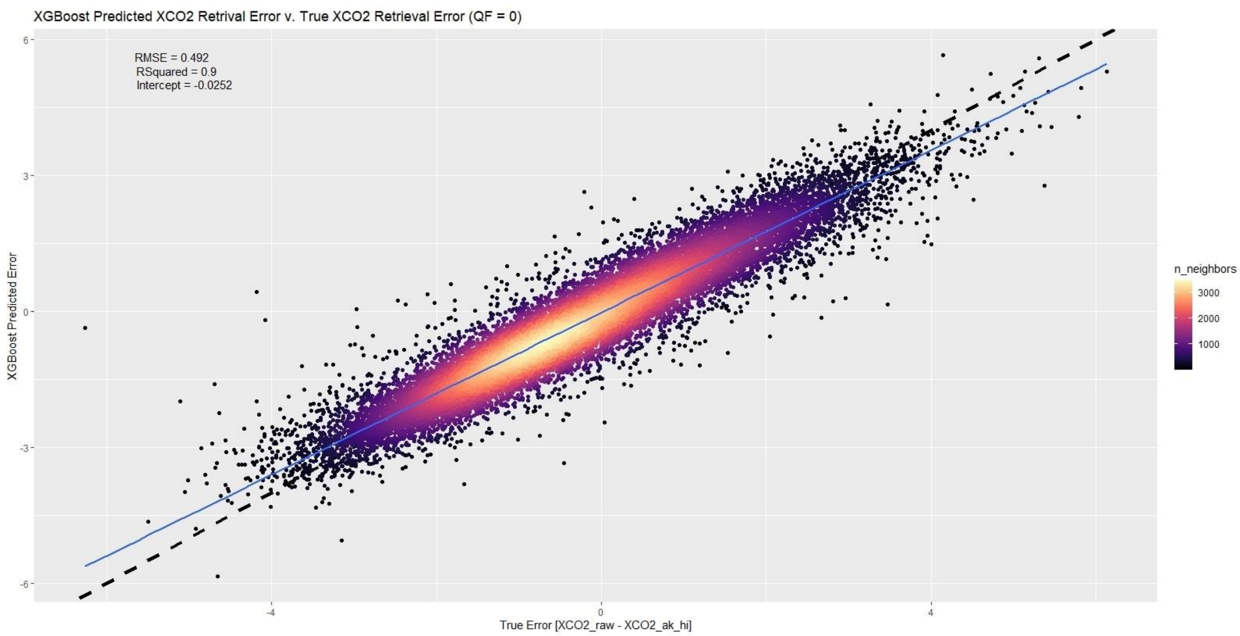
- Final selection includes: dP (surface pressure retrieved - prior surface pressure, however may use dP frac as described in Kiel et al. 2019) , *ils scale CO2 weak, ils scale CO2 strong, total AOD, air mass factor, BRDF slope CH4, and zero level offset O2.*

A NEW APPROACH TO BIAS CORRECTION AND QUALITY FILTERING

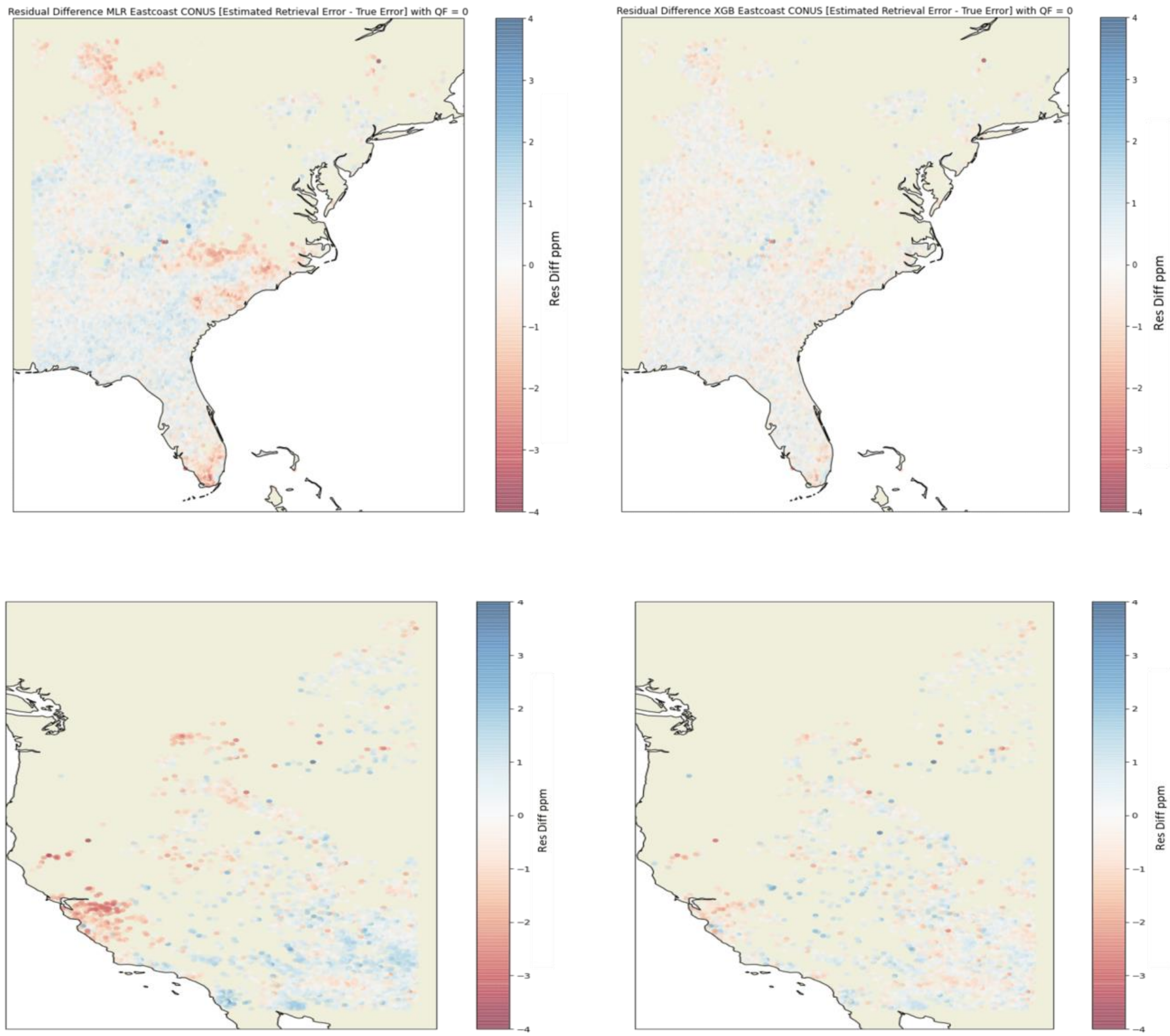
Comparison of machine learning method to MLR on quality filtered scenes:

- Results show that in the limited linear regime of the current quality filtering method, the machine learning method is able to attain an improved estimate over MLR of the residual error of retrieved XCO2.



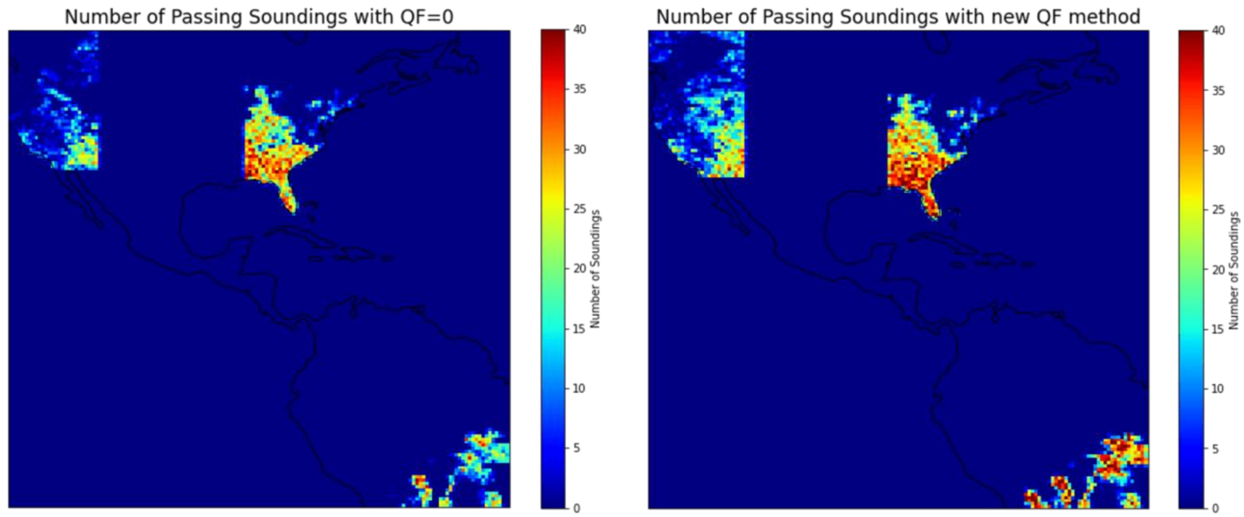


- Plotting of post BC residual of the two "global" models by each region illustrate an improved distribution of XGB estimated bias over MLR estimated bias.



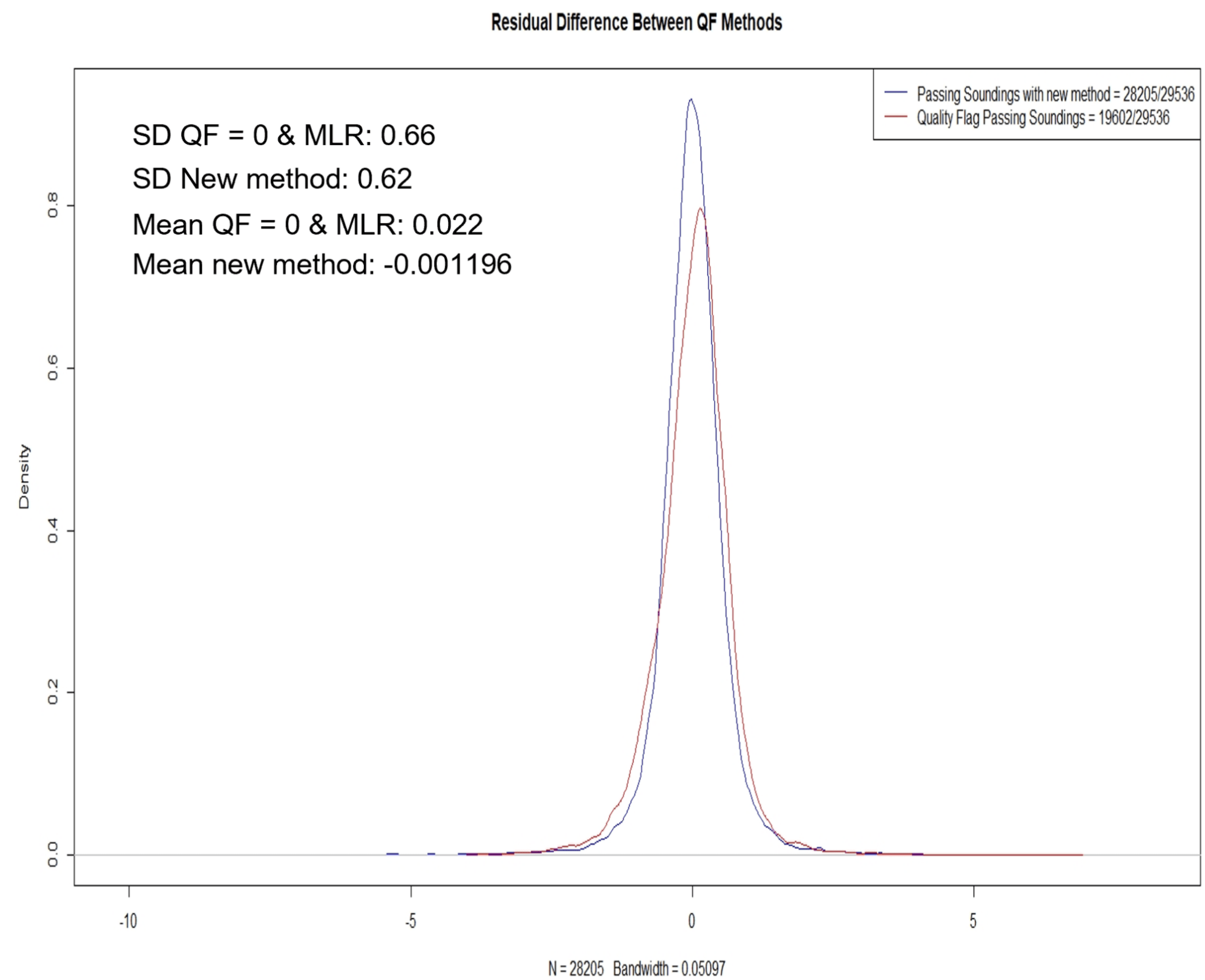
A novel Estimated Bias Threshold based approach to quality filtering:

- We now train a gradient booster with no quality filtering applied. A similar approach to feature selection is done but a larger subset of features is chosen for the final model for bias estimation in the non-linear regime.
- To apply our quality filtering technique, we simply select a desired absolute XGB estimated bias threshold and exclude soundings that whose estimated bias who fall outside of the selected range.
- Preliminary results are promising. Our first threshold selected for the XGB QF is the maximum absolute value of estimated bias by MLR & QF = 0. This improves the number of passing soundings from 66% to 95% of soundings as passing.



- In the distributions of the residual difference of estimated bias we can observe that several large outliers pass through the quality filtering of our new method (10ppm). However, a tighter mean about zero and standard deviation are maintained for the XGB bias correction over the MLR &

QF method.

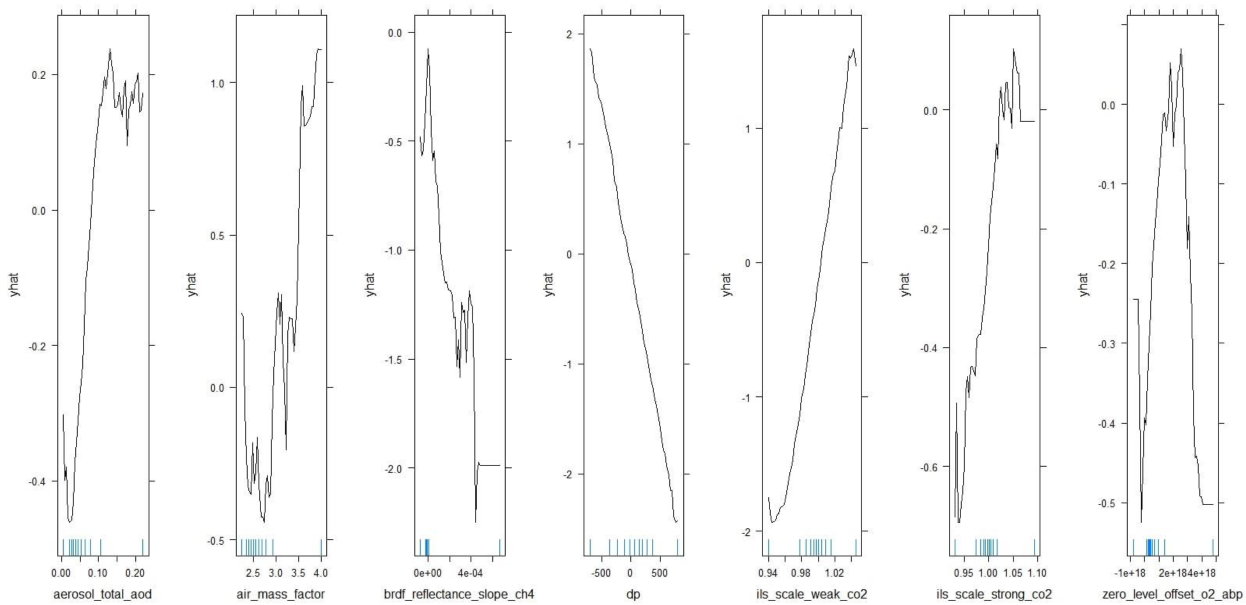


- By tightening the bound we can remove the outliers observed in the XGB thresholded method, as well as improve the overall bias correction. Number of passing soundings is still improved over MLR & QF at 86% soundings marked as passing. Note the results are preliminary on a small, simulated dataset. Currently validating methods on OCO-2 B10 data and paper is in preparation.

COVARIATE INTERACTION WITH XAI

Model sensitivity analysis:

- Model sensitivity to selected covariates is evaluated by allowing a feature of interest to permute through its observed value range while the compliment feature set is fixed (partial dependence). The average change in predicted bias (y-hat in the y-axis) over the permutated range is observed in the following plots. The range of average change observed in y-hat is an indicator of feature importance.

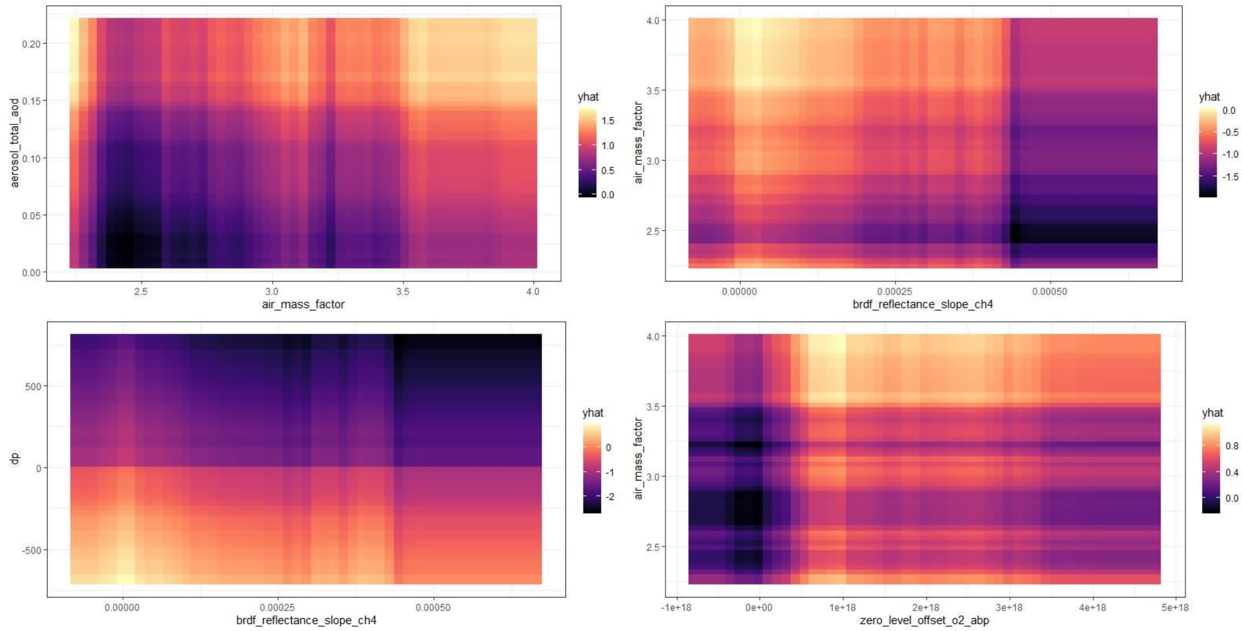


Friedman's H-Statistic and high order interactions among covariates:

- Higher order interactions among selected covariates can be explored using the H-Statistic (Friedman & Popescu 2008) which indicates the strength of an interaction by the amount of variance explained - with a value of 0 indicating no interaction.

Creation of new covariates:

- We select higher order non-linear combinations of covariates using Friedman's H-statistic. Exploration of these combinations can be done the with same post-hoc analysis as above (Gilpin et al. 2018). Two-dimensional partial dependence is plotted and combinations that contribute to high variance in predicted bias are selected as potential features for the XGB model.



- Models run with new features created from high order interactions of state vector covariates have improved predictive performance.

SUMMARY

- Comparison of a non-linear ML method to multiple linear regression on a dataset of simulated GeoCarb instrument soundings over the regime of interaction allowed by quality flag = 0 shows promise for a ML approach to bias correction of XCO2 retrievals.
- Utilization of a non-linear ML method and post-hoc XAI analysis allows for the exploration and creation of new higher order combinations of covariates.
- Performing bias correction on the full non-linear problem (before quality filtering) and then applying a novel threshold-based quality filtering approach shows a preliminary but substantial improvement in the number of passing soundings from 66% to 86%.
- Currently validating methods on OCO-2 B10 data and a paper is in preparation.

AUTHOR INFORMATION

William Keely is a Data Science PhD student at The University of Oklahoma and Graduate Research Assistant for the NASA GeoCarb Mission working under Dr. Sean Crowell.

email: william.r.keely@ou.edu

ABSTRACT

OCO-2, launched in 2014, uses reflected solar spectra and other retrieved geophysical variables to estimate (“retrieve”) the column averaged dry air mole fraction of CO2, termed XCO2.

A critical issue in satellite estimates of trace greenhouse gasses and remote sensing at large is the error distribution of an estimated target variable which arises from instrument artifacts as well as the under-determined nature of the retrieval of the quantities of interest. A large portion of the error is often incurred during inference from measurement of retrieved physical variables. These residual errors are typically corrected using ground truth observations of the target variable or some other truth proxy. Previous studies used multilinear regression to model the error distribution with a few covariates from the retrieved state vector, sometimes termed “features.” This presentation will cover the bias correction of XCO2 error attributed to retrieved covariates with a novel approach utilizing explainable Machine Learning methods (XAI) on simulated sounding retrievals from GeoCarb.

Utilization of non-linear models (Zhou, Grassotti 2020) or models that can capture non-linearity implicitly (Lorente et al. 2021) have been shown to improve on linear methods in operation. Our approach uses a gradient boosted decision tree ensemble method, XGBoost, that captures non-linear relations between input features and the target variable. XGBoost also incorporates regularization to prevent overfitting, while also remaining resilient to noise and large outliers – a feature missing from other ensemble DT methods. Decision Tree based models provide inherent feature importance that allows for high interpretability. We also approach post training analysis with model agnostic, *explainable* methods (XAI). XAI methods allow for rigorous insight into the causes of a model’s decision (Gilpin et al. 2018).

By applying these techniques, we will demonstrate our approach provides reduced residual errors relative to the operational method as well as yielding an uncertainty estimate in bias corrected XCO2, which is currently not treated separately from the posterior uncertainty estimate derived from the retrieval algorithm.

REFERENCES

Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). Association for Computing Machinery, New York, NY, USA, 785–794. DOI:https://doi.org/10.1145/2939672.2939785

Friedman, Jerome H., and Bogdan E. Popescu. "Predictive learning via rule ensembles." The Annals of Applied Statistics 2.3 (2008): 916-954.

Gilpin, Leilani H., et al. "Explaining explanations: An overview of interpretability of machine learning." 2018 IEEE 5th International Conference on data science and advanced analytics (DSAA). IEEE, 2018.

Kiel, Matthäus, et al. "How bias correction goes wrong: measurement of X CO 2 affected by erroneous surface pressure estimates." Atmospheric Measurement Techniques 12.4 (2019): 2241-2259.

Lorente, Alba, et al. "Methane retrieved from TROPOMI: improvement of the data product and validation of the first 2 years of measurements." Atmospheric Measurement Techniques 14.1 (2021): 665-684.

O'dell, Christopher W., et al. "Improved retrievals of carbon dioxide from Orbiting Carbon Observatory-2 with the version 8 ACOS algorithm." Atmospheric Measurement Techniques 11.12 (2018): 6539-6576.

Zhou, Yan, and Christopher Grassotti. “Development of a Machine Learning-Based Radiometric Bias Correction for NOAA’s Microwave Integrated Retrieval System (MiRS).” Remote Sensing 12.19 (2020): 3160. Crossref. Web.