Uncertainty quantification for a machine learning bias correction of OCO-2 column averaged CO2

William Keely[1], Steffen Mauceri[1], Otto Lamminää[1], Amy Braverman[1], Jonathan Hobbs[1], Joaquim Teixeira[1], and Vivienne Payne[1]

[1]Affiliation not available

December 10, 2023

Uncertainty quantification for a machine learning bias correction of OCO-2 column averaged CO_2

William Keely[1,2], Steffen Mauceri[1], Otto Lamminpää[1],
Amy Braverman[1], Jonathan Hobbs[1], Joaquim Teixeira[1], Vivienne Payne[1]
[1] *Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA*
[2] *Data Science and Analytics Institute, University of Oklahoma, Norman, OK, USA*

Jet Propulsion Laboratory
California Institute of Technology

I. Introduction

- Space based remote sensing of CO_2 is important for understanding the climate cycle.
- Flux inversion models depend on well bias corrected measurements of column averaged CO_2 observations.
- A nonlinear Machine Learning bias correction based on truth proxy validation improves correction over the operational linear approach used by OCO-2[1][2] .
- Uncertainty quantification (UQ) of machine learning is important for building trust in 'black box' models, e.g., when the model is extrapolating outside of the training data.

II. Objectives

1. Quantify the uncertainty of machine learning bias correction point estimate for the OCO-2 mission[1].
2. Explore the roles of the UQ product in quality filtering and explainability.

III. Methods & Data

TCCON truth proxy dataset:

- Total Carbon Column Observing Network (TCCON) is a set of ground-based instruments with large global distribution that produce precise total column measurements.[3]
- ΔXCO_2 or bias is the difference in column averaged CO_2 between OCO-2 and TCCON: $\Delta XCO_2 = XCO_{2,OCO-2} - XCO_{2,TCCON}$

Machine Learning Bias Correction:

- Compare three machine learning methods that provide a UQ of their estimates: Quantile Regression Forest (QRF), Gaussian Process Regressor (GPR) and a Bayesian linear regression as comparison to the operational model.
- Models are fit using a selection of co-retrieved atmospheric state features to estimate ΔXCO_2.
- All models use a temporal training/test split: data from 2015-2020 is used to train models and 2021 is holdout for evaluation.

Quantile Regression Forest:

Quantile Regression Forest[3] (QRF) is an extension of Random Forest that provides estimates for conditional quantiles prediction intervals of the response variable. Given a percentile α and a new input x, the percentile is calculated by taking the empirical distribution of the labels Y in the leaf nodes where x falls. The CDF of the empirical distribution is $F_{Y|X=x}(y) = \frac{1}{|\{Y_{leaf}(x)\}|}\sum_{y_i \epsilon \{Y_{leaf}(x)\}} \mathbb{I}(y_i \leq y)$ and the α percentile is estimated by finding the value of y for which the CDF reaches and is calculated as $Q_\alpha(x) = \frac{1}{M}\sum_{m=1}^{M} q_\alpha^{(m)}(x)$ where $q_\alpha^{(m)}(x)$ is the percentile estimate from m^{th} tree in the ensemble.

Gaussian Process Regression:

Gaussian Process Regression (GPR) is a Bayesian machine learning method that utilizes a prior mean and covariance functions to parameterize a Gaussian predictive distribution of a function $f(x)$ s.t. $f(x) \sim GP(m(x), k(x, x'))$. We use a Radial Basis Function kernel as our choice of covariance function defined as $k(x, x') = \sigma^2 \exp(-\frac{||x-x'||^2}{2\lambda^2})$ where λ is the length scale parameter. The predictive distribution is attained for a new x by the point estimate $\mu(x) = k(x, X)[k(X, X) + \sigma_n^2 I]^{-1} y$, and estimate of uncertainty is given as $\sigma^2(x) = k(x, x) - k(x, X)[k(X, X) + \sigma_n^2 I]^{-1} k(X, x)$.

Abstention Based Filtering:

A robust UQ can be used to indicate when we might distrust the estimate from the 'black box' machine learning model e.g., when the model extrapolates outside the domain of the training dataset. When uncertainty is high or above a selected threshold, the model "abstains" from making a prediction and thus servers as a form of quality filtering.

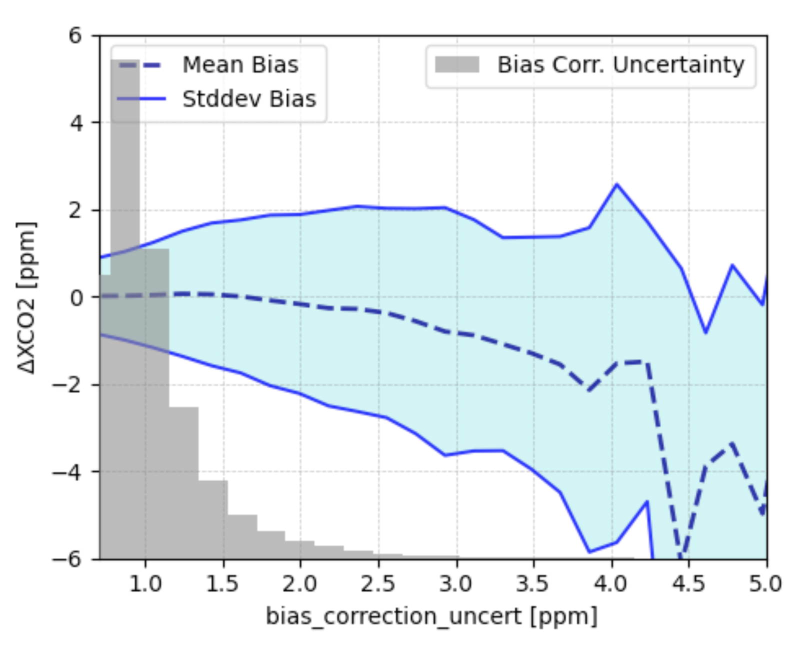


Fig. 1 | Abstention Filtering
The remaining bias (ΔXCO2) after bias correction is plotted on the y-axis and the uncertainty from the machine learning bias correction is plotted on the x-axis. The dashed-blue line indicates the mean ΔXCO2 and fill-between shows the SD of ΔXCO2. Given a threshold selection on the uncertainty of the bias correction [e.g., bias_correction_uncert < 1.5] the remaining variance of error can be reduced at the cost of some data throughput.

IV. Results & Conclusions

Model Comparison: We evaluate the robustness of the UQ estimates from each model by evaluating the coverage of the prediction intervals on out-of-distribution (OOD) data. Models are trained on quality filtered data that is of good-quality and evaluate on poor-quality data. The Estimated Calibration Error (ECE) is then plotted as shown in Figure 2. QRF has the best performance on OOD samples for the selected quantiles, followed by GPR. The prediction intervals from the Bayesian linear regression (GLM) are often overconfident i.e., do not contain the ground truth.

Spatial Distribution: The magnitude of uncertainty from the bias correction is consistently high over the tropics and Sub-Saharan Africa. For Northern Hemisphere summer (JJA) and fall (SON) there is a significant increase in uncertainty in extreme Northern Latitudes particularly for ocean glint scenes, as shown in Figure 3. Often the uncertainty is high in regions where TCCON co-locations are sparse or nonexistent, thus indicating when the machine learning bias correction is extrapolating outside of the training data.

Future Work: The machine learning bias correction and its uncertainty will be included in a future OCO-2 update along with a machine learning and abstention-based quality filter.

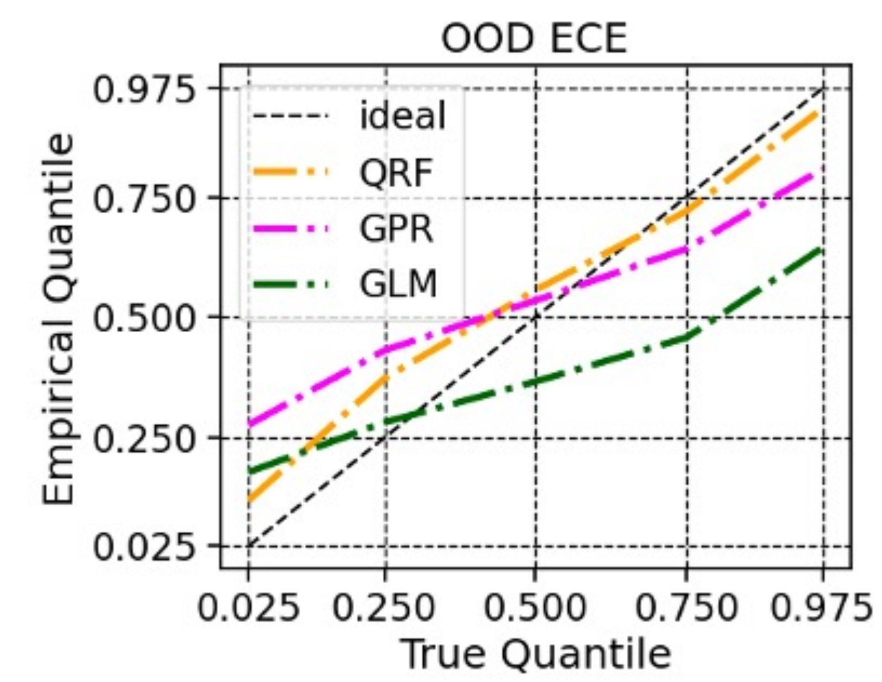
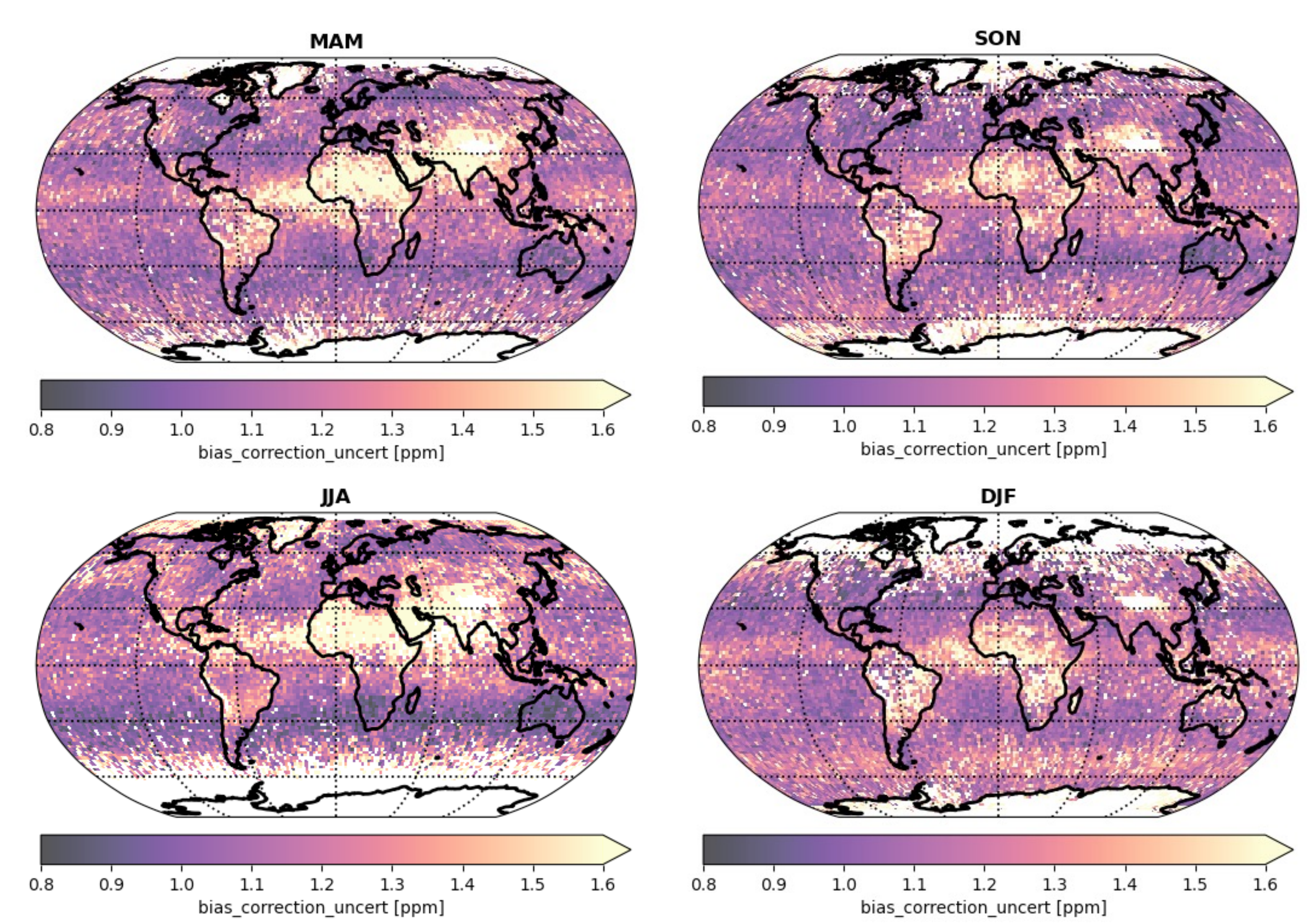


Fig. 2 | Estimated Calibration Error
Estimated Calibration Error (ECE) is calculated for each of the three models. The 1-sigma prediction intervals are z-score standardized and the percentage of ground truth observations that fall within each quantile are observed (empirical quantile) and compared to the expected quantile coverage (true quantile). A perfectly calibrated uncertainty should fall along the 1:1 line. The QRF show the best coverage on out-of-distribution (OOD) observations over the GPR and Bayesian linear regression (GLM).

Fig. 3 | Seasonal
Seasonal plots of uncertainty from the machine learning bias correction on B11.1 data for N. Hemisphere spring (MAM), summer (JJA), fall (SON) and winter (DJF), for 2021 using a 2 x 2 degree binning.



VI. Acknowledgements

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

References

[1] Crisp, D., Atlas, R. M., Breon, F. M., Brown, L. R., Burrows, J. P., Ciais, P., . . . Schroll, S. (2004). The Orbiting Carbon Observatory (OCO) mission. Advances in Space Research, 34(4),
[2] Keely, W. R., Mauceri, S., Crowell, S., and O'Dell, C. W.: A nonlinear data-driven approach to bias correction of XCO2 for NASA's OCO-2 ACOS version 10, Atmos. Meas. Tech., 16, 5725–5748, https://doi.org/10.5194/amt-16-5725-2023
[3] Meinshausen, N., & Ridgeway, G. (2006). Quantile regression forests. Journal of machine learning research, 7(6).

william.r.keely@ou.edu