https://doi.org/10.5194/amt-2022-202
Preprint. Discussion started: 18 July 2022

Atmospheric Measurement Techniques Discussions

Correcting 3D cloud effects in XCO2 retrievals from OCO-2

Steffen Mauceri[1], Steven Massie[2], Sebastian Schmidt[2]

[1]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA
[2]Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder, Colorado 80303, USA

Correspondence to: Steffen Mauceri (Steffen.Mauceri@jpl.nasa.gov)

Abstract. The Orbiting Carbon Observatory-2 makes space-based radiance measurements in the Oxygen A-band and the Weak and Strong carbon dioxide (CO_2) bands. Using a physics-based retrieval algorithm these measurements are inverted to column-averaged atmospheric CO2 dry-air mole fraction (X_{CO2}). However, the retrieved X_{CO2} are biased due to calibration issues and mismatches between the physics-based retrieval and nature. Using multiple linear regression, the biases are empirically mitigated. However, a recent analysis revealed remaining biases in the proximity of clouds caused by 3D cloud radiative effects (Massie et al., 2021) in the current processing version B10. Using an interpretable non-linear machine learning approach, we developed a bias correction model to address these 3D cloud biases. The model is able to reduce unphysical variability over land and ocean by 31% and 55%, respectively. Additionally, the 3D cloud bias corrected X_{CO2} show better agreement with independent ground-based observations from the Total Carbon Column Observation Network (TCCON). Overall, we find that OCO-2 underestimates X_{CO2} over land by -0.4 ppm in the tropics and northward of 45° N. The approach can be expanded to a more general bias correction and is generalizable to other greenhouse gas missions, such as GeoCarb, GOSAT-3 and CO2M.

1 Introduction

The Orbiting Carbon Observatory OCO-2 (Crisp et al., 2004; Eldering et al., 2017) makes space-based top-of-atmosphere radiance measurements in three spectral bands: Oxygen A band at 0.76 µm, the Weak CO2-band at 1.61 µm, and the Strong CO2 band at 2.06 µm. Using an optimal estimation retrieval (Rodgers, 2000) called ACOS (O'Dell et al., 2018), these measurements are converted to column-averaged atmospheric CO_2 dry-air mole fraction (X_{CO2}). ACOS employs a physics-based forward model that takes into consideration viewing and solar geometry and various atmospheric and surface parameters. Since OCO-2 generates on the order of 100,000 soundings per day, ACOS makes multiple approximations to speed up the retrieval algorithm. Most importantly, the retrieval makes the independent pixel approximation, where the radiance in a given sounding only depends on the properties (e.g. surface reflectance, aerosols, trace gas concentration) within the field of view of this sounding. Thus, the retrieval does not account for the horizontal exchange of photons as caused by nearby clouds, also referred to as 3D cloud effect. This leads to negative biases in retrieved X_{CO2} in the vicinity of clouds (Massie et al., 2021; Massie, Sebastian, Eldering, & Crisp, 2017; Merrelli, Bennartz, O'Dell, & Taylor, 2015).

https://doi.org/10.5194/amt-2022-202
Preprint. Discussion started: 18 July 2022

Atmospheric Measurement Techniques Discussions

To mitigate biases in retrieved X_{CO2}, a linear bias correction and threshold-based filtering is applied to the data. For the current version (B10) bias correction and filtering are based on co-retrieved elements from the state vector that are used to bring retrieved X_{CO2} into agreement with multiple truth sources (Kiel et al., 2019). These truth sources include a 'small areas analysis' which assumes that X_{CO2} is constant over small distances (<100 km) within the same orbit, comparisons to ground-based observations from the Total Carbon Column Observation Network (TCCON) (Wunch et al., 2010) and comparisons to a multi model mean of six models that assimilate in-situ data. Nevertheless, there are remaining negative 3D cloud biases present in the proximity of clouds with an average of -0.4 and -2.2 ppm for high quality (QF=0) and low quality (QF=1) data (Massie et al., 2021). To address these biases Massie et al., (2021) developed a linear bias correction and filtering approach using a set of features indicative of 3D cloud effects calculated from Moderate Resolution Imaging Spectroradiometer (MODIS) and OCO-2 files. However, biases in X_{CO2} caused by nearby clouds are highly non-linear and the 3D cloud effect features underrepresent the complexity of those effects and how they impact X_{CO2} biases. Consequently, the present study has two goals. The first goal is to explore if a non-linear bias correction can reduce 3D cloud biases further than a linear approach. While the developed cloud features (H3D, HC, CSNoiseRatio, Cloud Distance, discussed below) more directly capture 3D cloud effects, co-retrieved variables from the state vector might be more indicative of the resulting X_{CO2} biases. Thus, the second goal is to investigate if additional variables, co-retrieved with X_{CO2}, can be used to further reduce 3D cloud biases.

2 Data

We make use of the OCO-2 B10 lite files (https://disc.gsfc.nasa.gov/datasets/OCO2_L2_Lite_FP_10r/, last access: 05/2022) from September 2014 to July 2019. These files contain bias corrected X_{CO2} for ocean glint and land nadir observations that we wish to correct for remaining 3D cloud biases, a variety of parameters describing the retrieved atmospheric state vector, viewing and solar geometry, results from pre-processors, location and time, and a quality flag (QF) for each sounding. The QF is determined by a series of hand tuned thresholds for various variables derived from state vector elements that are indicative of retrieval biases in X_{CO2}. Similarly, the bias correction is performed with hand tuned linear fits to various state vector elements (Kiel et al., 2019). As a truth metric to determine the bias correction and QF, B10 utilizes the small areas analysis, comparisons to TCCON and a multi model mean.

In addition to the B10 lite files we utilize ground-based observations by TCCON from all 27 stations that are in close proximity in time (24 h) and space (2.5° in latitude, 5° in longitude) to OCO-2 observations (https://tccondata.org, last access: 05/2022). The ground-based observations are used for validation only. However, they can only provide comparisons for a limited number of locations, with relatively few ground-based sites in the Tropics and island locations.

Finally, we make use of four 3D cloud effect features: **H3D** (Liang, Di Girolamo, & Platnick, 2009; Massie et al., 2017) describes the normalized standard deviation of the radiance field, **HC** is calculated from differences in continuum radiances of an observation point and adjacent points in three rows (frames) of footprints, **CSNoiseRatio** is the ratio of the continuum radiance spatial standard deviation and noise level at the continuum radiance level, and **Cloud Distance** (Massie et al., 2021)

https://doi.org/10.5194/amt-2022-202
Preprint. Discussion started: 18 July 2022

Atmospheric Measurement Techniques Discussions

which is the distance of the nearest cloud as determined from MODIS imagery. The calculated 3D cloud features can be found for OCO-2 lite files from September 2014 to July 2019 at https://doi.org/10.5281/zenodo.4008764.

3 Methods

3.1 Small Areas and TCCON as Truth metric

As a pre-processing step we match the 3D cloud variables and B10 lite files by OCO-2 sounding id and TCCON by time and location. Afterwards, we remove soundings where no 3D cloud variables are available. To develop the bias correction model, we use the 'small areas analysis', which is based on the assumption that CO_2 is a well-mixed gas and assumed to be constant over spatial scales of less than ~100 km (there can be exceptions for strong CO_2 emitters such as mega cities). To exploit this constraint on X_{CO2} we split OCO-2 soundings from the same orbit into small areas using the k-means algorithm (Hartigan & Wong, 1979). This groups soundings where variations in X_{CO2} can be interpreted as non-physical variability, or retrieval biases. For each group we define the median B10 bias corrected X_{CO2} of this group as the true X_{CO2} and any differences to this median are treated as biases. Note, that this assumes that each small area contains a subset of soundings that are not affected by 3D cloud biases, which might not be accurate for some small areas dominated by clouds (e.g. in the tropics). Finally, we remove groups where soundings cover an area bigger than 100 km or there are fewer than 20 soundings. This results in approximately 10^6 land nadir soundings and $11 \cdot 10^6$ ocean glint soundings, with a small subset of the soundings having coincident TCCON measurements. TCCON can only provide comparison for a limited set of regions with most stations in the northern hemisphere and on land. This challenges the development of a bias correction approach based on X_{CO2} - TCCON differences that would be representative of areas far away from existing stations, such as Africa, South America and most of the ocean. Therefore, we use TCCON only as an independent truth metric for validation and not to develop the model itself.

The distribution of nearest cloud distance, biases from the small area analysis and comparison to TCCON for land nadir and ocean glint observations with QF=0 and QF=1 are shown in Figure 1. The plots show that the majority of OCO-2 soundings are taken within close proximity of clouds and that many of those soundings are filtered out (QF=1). This is especially problematic for areas such as the tropics that are dominated by clouds and, as a result, have few valid soundings. The small area and TCCON biases for QF=0 data are roughly normally distributed with a mean and standard deviation of 0.1 ± 0.5 ppm for B10 small area biases and 0.2 ± 0.8 ppm compared to TCCON for ocean glint. For land nadir B10 small area bias and B10 - TCCON are similar with a mean and standard deviation of 0.1 ± ~1 ppm. For QF=1 the distribution of biases has a larger standard deviation for B10 small area biases (land: 2.9 ppm, ocean: 1.9 ppm) and B10 – TCCON (land: 3.9 ppm, ocean: 2.1 ppm), is skewed, and contains negative biases that far exceed positive biases, as analysed with the small areas (land: -0.5 ppm, ocean: -0.9 ppm) and compared to TCCON (land: -1.4 ppm, ocean: -1.2 ppm). This long tail distribution of negative biases is indicative of 3D cloud effects and should be mitigated with a successful 3D cloud bias correction.

https://doi.org/10.5194/amt-2022-202
Preprint. Discussion started: 18 July 2022

Atmospheric Measurement Techniques Discussions

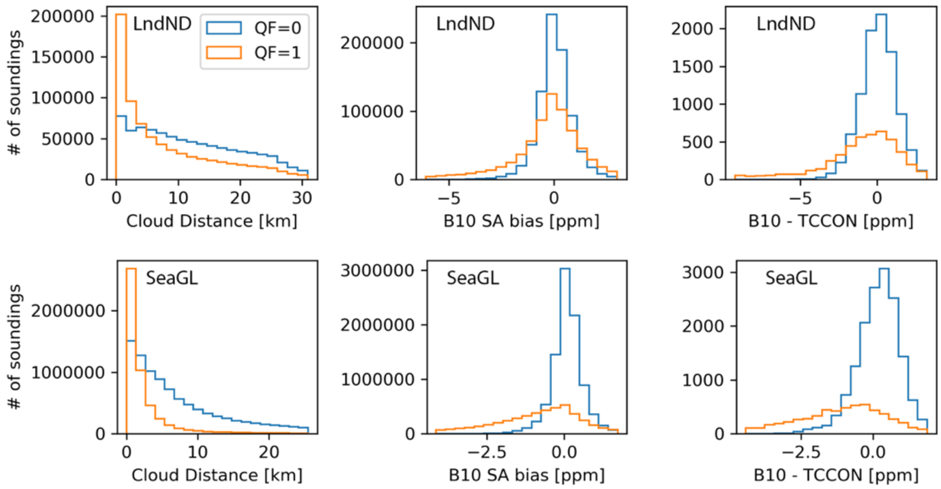

Figure 1: Histogram of data used in this study for nearest cloud distance (left), small area biases (middle), and biases compared to TCCON (right) for land nadir (top) and ocean glint (bottom) soundings. QF=0 data is shown in blue, QF=1 data in orange.

3.2 Train-, Validation-, Test-split

To fit, or *train*, the bias correction model we used soundings from September 2014 to the end of July 2017, totalling roughly $8 \cdot 10^6$ and $7 \cdot 10^5$ soundings for ocean glint and land nadir, respectively. To find the best model parameters and evaluate what features minimize biases the furthest we use a separate validation set containing soundings from the beginning of August 2017 to end of July 2018. Finally, to test how the trained model performs on new data we use a separate testing set of soundings from the beginning of August 2018 to the end of July 2019. The validation and testing set have $2 \cdot 10^6$ and $1.6 \cdot 10^6$ soundings for ocean glint QF=0 and QF=1, respectively, and $17 \cdot 10^4$ and $14 \cdot 10^4$ for land nadir QF=0 and QF=1, respectively.

3.3 Bias Correction Model

We train two types of models for the bias correction, non-linear models (Random Forest) and linear models (Ridge Regression) to provide a baseline comparison. A Random Forest is an ensemble of classifying decision trees and outputs the mean of those trees (Breiman, 2001). Each tree is trained in a supervised manner with a random subset (50%) of the available training data, also referred to as boot strapping. Using the training data, each tree iteratively splits the data using the feature that can minimize the mean squared error of the predictions the furthest, until it reaches a maximum user-provided number of splits, or *depth*. For our land model we used a depth of 8 and for our ocean model a depth of 15. The larger model size for the ocean is mostly due to there being more training data available over the ocean than over land which allows to fit a larger model that still generalizes to new data. Each random forest was composed of 100 individual trees. These parameters were chosen to maximize

https://doi.org/10.5194/amt-2022-202
Preprint. Discussion started: 18 July 2022

Atmospheric Measurement Techniques Discussions

model performance on the validation set. The model inputs are a set of selected features form the OCO-2 Lite files (e.g. co2_grad_del) and the model output is the remaining bias in the B10 bias corrected X_{CO2} derived from the small areas analysis.

Since the B10 bias correction uses a linear approach, we also perform a baseline comparison to a linear model. We choose multi-variate linear regression with a small Tikhonov regularization term (the regularization helps if some of the inputs are correlated, which is the case for most real-world applications), also referred to as ridge regression. Thus, using the training set we seek to find the weights, $\boldsymbol{w}$, that minimize the following equation:

$$\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{w}\|_2^2+\alpha\|\boldsymbol{w}\|_2^2 \tag{1}$$

where $\boldsymbol{y}$ is the standardized (mean removed and divided by standard deviation) X_{CO2} bias, $\boldsymbol{X}$ are the standardized features, and α controls the strength of the Tikhonov regularization. For our application we found $\alpha=10^{-5}$ to maximize performance on the validation set.

3.4 Feature Selection

First, we identified retrieved variables in the Lite files that show a strong dependence (change in mean or variability) to nearest cloud distance, indicating that they might be good candidates to correct for 3D cloud effects. Three examples are shown in Figure 2. In addition to the list of identified features we added solar and viewing geometries for land and ocean and surface albedo for land. Those variables have a direct physical impact on 3D cloud effects, for example, the sun being closer to the horizon amplifies 3D cloud effects so does a brighter surface albedo (Okata et al., 2017). Finally, we removed highly correlated variables. This results in a set of 23 features for land nadir, and 24 features for ocean glint soundings, that may be used to correct for 3D cloud biases in retrieved X_{CO2} (more information about each variable can be found in (Jet Propulsion Laboratory, 2018)). Next, we used recursive feature elimination to identify what subset of features can reduce biases the furthest. Reducing the number of features makes the model more robust to new data, or avoids *overfitting,* and aids interpretability.

For the recursive feature elimination, we removed one feature at a time, trained a small random forest model with 32 trees each on a random selection of $5 \cdot 10^5$ soundings with QF=0 and QF=1 from the training set. Afterwards we calculated the model performance on the full validation set. As the performance metrics we used the correlation coefficient (R^2) between modelled bias and existing bias as indicated by the small-areas calculations. The feature, that has been removed from the highest performing model is then permanently removed and the process is repeated until only one feature is left. The iterative process was performed separately for land nadir and ocean glint soundings. The order of the feature elimination and resulting R^2 is shown in Figure 3. The least important variables are shown at the top and were removed first. A low importance can either result from a variable varying independently of remaining biases in X_{CO2} or the variable could be correlated with another

https://doi.org/10.5194/amt-2022-202
Preprint. Discussion started: 18 July 2022

variable (e.g. dp and dp_abp) or set of variables that provide similar information, making one of them obsolete. The most important variables are shown on the bottom.

For our bias-correction model we decided to use the three most important variables for ocean glint and four most important variables for land nadir, as identified by the feature elimination. These variables explain most of the variance and mostly overlap for land and ocean, a further indication that those variables have a robust relationship to 3D cloud biases. For land nadir and ocean glint soundings the three most important variables are **xco2_strong_idp** (X_{CO2} retrieved from the strong CO_2 band with the IMAP-DOAS pre-processing algorithm, normalized by subtracting the mean of each small area), **co2_grad_del** (change between the retrieved CO_2 profile and the a priori profile from the surface minus that at level 13), and **dp** (difference of retrieved surface pressure and a priori surface pressure obtained from GMAO GEOS5-FP-IT model). The fourth most important variable for land nadir is **albedo_wco2** (surface albedo in the weak CO_2 band). Note, the final set of features does not include any of the 3D cloud metrics used in the bias correction by Massie et al. (2021). Additionally, solar and viewing geometry were removed in the iterative process. However, it includes albedo_wco2 which has a direct physical connection to 3D cloud effects. This indicates that elements of the state vector (co2_grad_del, dp, albedo_wco2) and results from the pre-processing algorithms (xco2_strong_idp) are more directly correlated with remaining biases in X_{CO2} (due to 3D cloud and other effects) than features that directly measure 3D cloud effects which perturb the radiation field (H3D, HC, CSNoiseRatio). From an operational standpoint for OCO-2 the three features are available in the OCO-2 Lite files which simplifies their inclusion in future operational products.

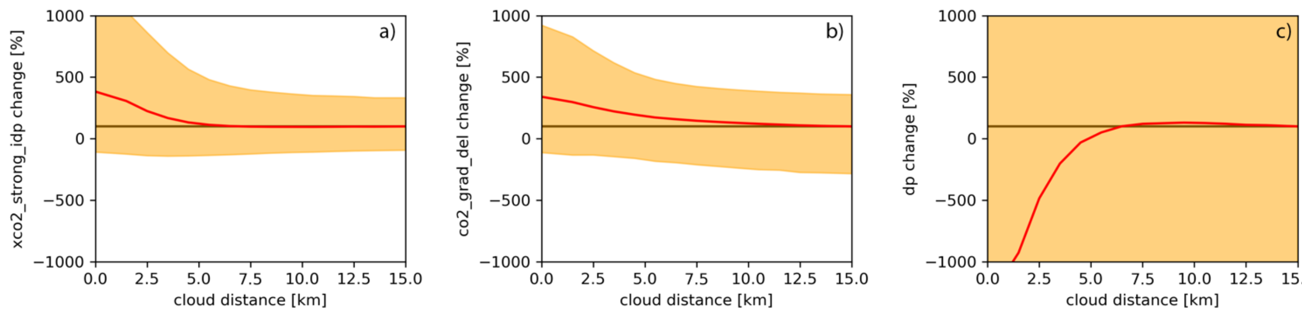

Figure 2: Change of variability and mean in percent of model features with respect to nearest cloud distance. Change in mean is shown in red; change of the 5th and 95th percentile is shown in yellow; no change (baseline) is shown with a brown straight line. Change is calculated with respect to feature mean for observations with a nearest cloud distance of 14 km to 15 km. a) shows change for xco2_strong_idp, b) co2_grad_del, and c) dp (the mean of 'dp change' for a cloud distance of 0 km is -1300%).

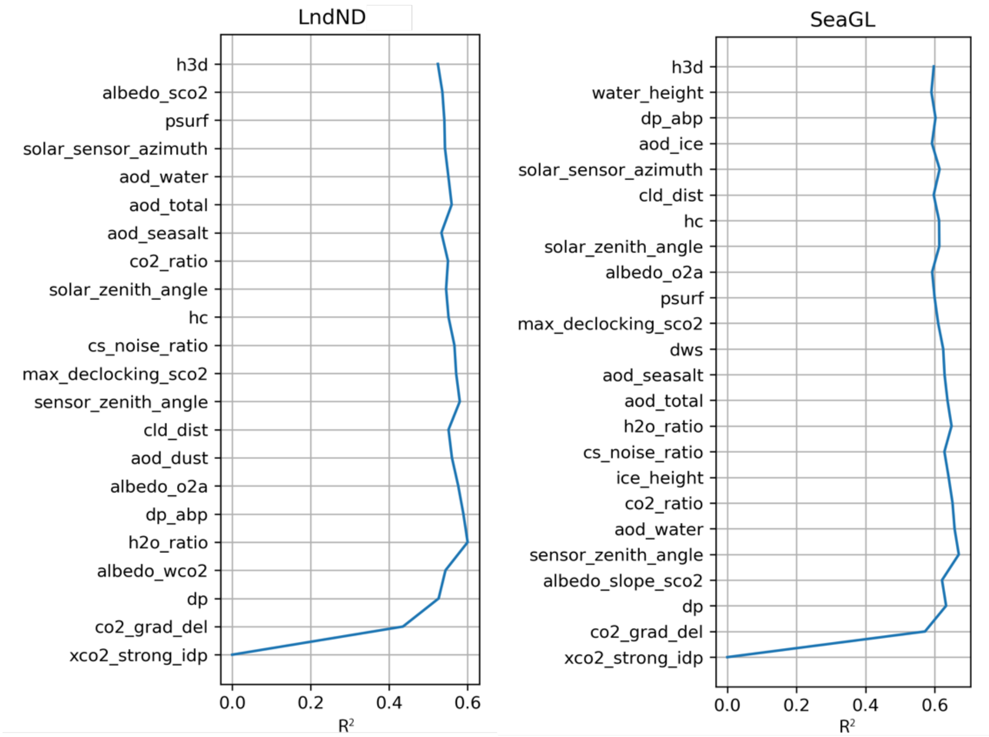

Figure 3: Feature ordering by importance as determined by recursive feature elimination. Features were removed from top to bottom with the most important features on the bottom. The model performance for removing a given feature is indicated with R^2 calculated on the validation set.

4 Results

4.1 Reduction in X_{CO2} biases

After the random forest was trained using the training set (09/2014 – 07/2017) we evaluated the model performance on the testing set (08/2018 – 07/2019). Figure 4 compares remaining X_{CO2} biases in B10 with biases after our correction is applied (B10-RF) for QF=0 and QF=1 soundings. For land nadir soundings X_{CO2} biases are reduced from a standard deviation of 1.8 ppm to 1.4 ppm (see Figure 4c) with the biggest correction applied to soundings that have biases less than -1.5 ppm. For ocean glint soundings the bias correction has a significantly bigger impact and reduces biases from 1.3 ppm to 0.7 ppm (see Figure 4d). Over the ocean the bias correction mostly corrects negative biases less than -0.8 ppm (see Figure 4b).

https://doi.org/10.5194/amt-2022-202
Preprint. Discussion started: 18 July 2022

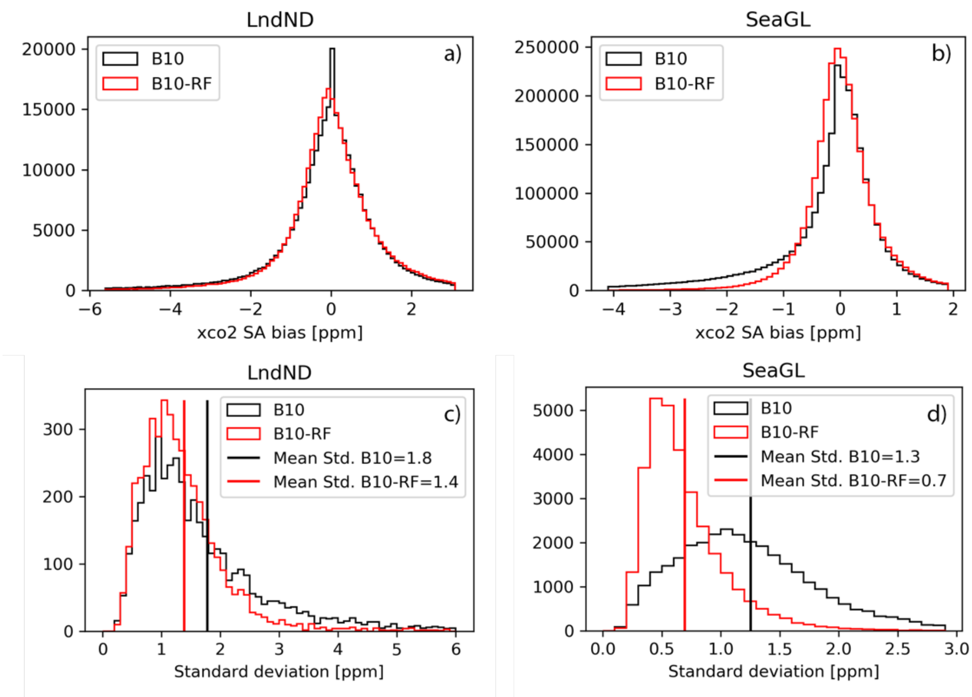

Figure 4: Reduction in non-physical variability in XCO_2 for B10 and the proposed bias correction approach (B10-RF) for land nadir (left) and ocean glint (right) for QF=0 and QF=1 data from 2018 to 2019. (Top) distribution of biases from individual soundings; (bottom) distribution of standard deviation for individual small areas.

Table 1 shows the Root Mean Square Error (RMSE) by quality flag. For QF=0 and QF=1 the biases in X_{CO2} corrected with our model are less than for B10 on the testing set. However, for QF=0 improvements by our correction (B10-RF) compared to B10 are small (~10%). These data have significantly fewer soundings with clouds in close proximity (see Figure 1) which explains in part the smaller difference. For QF=1 the difference is more significant, reducing the RMSE from 2.76 ppm to 1.91 ppm over land and from 1.35 ppm to 0.74 ppm over ocean.

Table 1: RMSE of X_{CO2} as determined by small areas analysis for the testing set (08/2018 – 07/2019). The RMSE is shown for the operational OCO-2 product (B10), the proposed random forest approach (B10-RF), a linear bias correction using the same three features than RF (B10-Ridge), and a random forest using dedicated cloud metrics (B10-Cloud).

	LndND XCO_2 [ppm]				SeaGL XCO_2 [ppm]			
	B10	B10-RF	B10-Ridge	B10-Cloud	B10	B10-RF	B10-Ridge	B10-Cloud
QF=0	0.83	**0.75**	0.77	0.82	0.52	**0.44**	0.47	0.49
QF=1	2.76	**1.91**	2.19	2.62	2.12	**1.18**	1.44	1.83
QF=0 + 1	2.08	**1.43**	1.67	1.99	1.35	**0.74**	0.89	1.19

https://doi.org/10.5194/amt-2022-202
Preprint. Discussion started: 18 July 2022

Atmospheric Measurement Techniques
Discussions

To more directly link the bias correction to 3D cloud effects we show biases with respect to nearest cloud distance in Figure 5. X_{CO2} from B10 shows a clear negative mean bias and increased variance for a nearest cloud distance of less than 3 km and 4 km over land and ocean, respectively. After applying our bias correction the mean bias in the proximity of clouds is close to zero. Thus, the bias correction effectively mitigates biases due to 3D cloud effects.

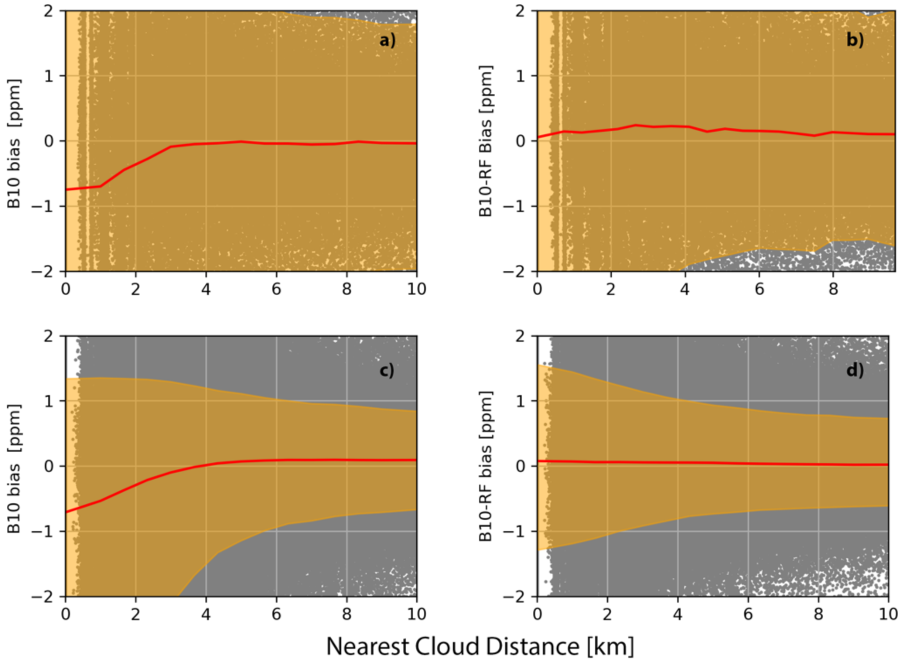

Figure 5: X_{CO2} bias vs cloud distance for land nadir B10 (a) land nadir B10 corrected (b) ocean glint B10 (c), and ocean glint B10 corrected (d) for QF=0 and QF=1 data from 2018 to 2019. The 5th and 95th percentiles are indicated with the yellow shaded area; the mean is shown with a red line and individual comparisons with grey dots.

4.2 Linear vs Non-linear bias correction

Building on the work by Massie et al., (2021) one of the guiding research questions was whether a non-linear approach based on interpretable machine learning techniques would improve upon a linear 3D cloud bias correction. To probe this question, we compare the performance of the non-linear random forest model to linear ridge regression (see Equation 1). To train the linear model we used the same features, training and testing set than for the random forest. The RMSE for the linear model (B10-Ridge) and non-linear model (B10-RF) is shown in Table 1. For QF=0 land nadir and ocean glint observations the linear and non-linear model have similar performance with the non-linear model allowing for a slightly lower RMSE. For QF=1 the non-linear random forest reduces remaining biases further than the linear ridge regression from 2.19 ppm to 1.91 ppm over land and from 1.44 ppm to 1.18 ppm over the ocean.

https://doi.org/10.5194/amt-2022-202
Preprint. Discussion started: 18 July 2022

Atmospheric Measurement Techniques Discussions

4.3 Comparison to Using Dedicated Cloud Variables

A second question that we wanted to answer was whether additional variables from the B10 lite files could improve the 3D cloud bias correction. As shown in Figure 3, the four cloud variables (h3d, hc, cs_noise_ratio, cld_dist) were removed during the recursive feature elimination step, indicating that other variables from the lite files are more directly correlated with remaining B10 X_{CO2} biases. To better understand how much of the model performance stems from the new set of features we performed a set of experiments. For the first experiment we trained a random forest using only the four cloud variables in addition to surface albedo, solar zenith angle, sensor zenith angle, and the difference between solar and sensor azimuth. The results are shown in Table 1 (B10-Cloud). As expected, using the cloud variables with the non-linear random forest model performs worse than using the random forest with the features identified using the recursive feature elimination. However, it also performs worse than using the linear model for QF=0 and QF=1 for land nadir and ocean glint observations. One caveat of this experiment is that our bias correction approach, aimed at 3D cloud biases, might also make corrections for biases stemming from other effects (e.g. aerosols) that are independent to clouds and, thus, cannot be explained with cloud variables. Unfortunately, clearly separating various sources of bias is not possible.

For the other experiment we combine the 3D cloud variables with the variables determined by the recursive feature elimination (xco2_strong_idp, co2_grad_del, dp for land and ocean and albedo_wco2 for land) and compare the results to using only the features from the recursive feature elimination. If adding the 3D cloud variables would significantly reduce biases in XCO_2 further it would indicate that the set of identified features is mostly correcting for biases unrelated to 3D cloud effects. In total we compare the model performance of four sets of features: a) xco2_strong_idp, co2_grad_del, dp, albedo_wco2 (for land only) and nearest cloud distance, b) xco2_strong_idp, co2_grad_del, dp, albedo_wco2 (for land only) and CSNoiseRatio, c) xco2_strong_idp, co2_grad_del, dp, albedo_wco2 (for land only), nearest cloud distance, CSNoiseRatio, HC, H3D, and d) xco2_strong_idp, co2_grad_del, dp, albedo_wco2 (for land only), and deltaT (retrieved offset to a priori temperature profile). The last set of features serves as a control experiment where we quantify the effect of adding a random variable that is unrelated to 3D cloud effects to the set of chosen features. The results are shown in Table 2. For QF=0 there are practically no differences for the four test cases compared to our chosen set of features. For land nadir QF=1 soundings the feature sets a) and b) lead to a similar RMSE than our chosen set of features. For the feature sets c) and d) the RMSE is slightly larger. For ocean glint QF=1 the best set of features is c) which reduces the RMSE from 1.18 ppm to 1.13 ppm. Overall, the addition of 3D cloud variables (a, b, c) allows the models to lower the RMSE further compared to our proposed model, however, the improvements are only marginal. This indicates that the set of chosen features in our bias correction model accounts for the majority of 3D cloud biases in X_{CO2}. Further evidence for this was shown in Figure 5 and is presented in the next section with an independent comparison to TCCON.

https://doi.org/10.5194/amt-2022-202
Preprint. Discussion started: 18 July 2022

Atmospheric Measurement Techniques Discussions

Table 2: RMSE of X_{CO2} as determined by small areas analysis for the testing set (08/2018 – 07/2019). The RMSE is shown for the proposed random forest approach (B10-RF) and using the same approach but with additional features. In addition to xco2_strong_idp, co2_grad_del, dp a) contains nearest cloud distance, b) CSNoiseRatio, c) nearest cloud distance, CSNoiseRatio, HC, H3D, and d) detaT.

	LndND XCO_2 [ppm]					SeaGL XCO_2 [ppm]				
	B10-RF	a)	b)	c)	d)	B10-RF	a)	b)	c)	d)
QF=0	**0.75**	0.75	0.75	0.75	0.75	**0.44**	0.44	0.44	0.43	0.44
QF=1	**1.91**	1.90	1.90	1.96	1.92	**1.18**	1.16	1.16	1.13	1.17
QF=0 + 1	**1.43**	1.43	1.43	1.48	1.44	**0.74**	0.73	0.73	0.71	0.73

4.5 Comparison to TCCON

We further compare bias corrected X_{CO2} to TCCON. TCCON observations have low uncertainties and are used to validate OCO-2 retrieved X_{CO2}. However, they can only provide point measurements and are non-uniformly distributed, with most TCCON sites over land and in the northern hemisphere. For our comparison we consider coinciding observations of OCO-2 and TCCON for the period of the testing set (08/2018 - 07/2019). This results in 1768 (QF=0: 1397, QF=1: 371) matches for land nadir and 1305 (QF=0: 942, QF=1: 363) matches for ocean glint observations. Note, our bias correction model was trained without taking TCCON observations into consideration while B10 takes OCO-2 – TCCON biases explicitly into consideration for its linear bias correction, filtering, and to calculate global offsets. Thus, comparisons between B10 and TCCON are not independent.

Table 3 shows the mean and standard deviation of differences between B10 and TCCON and after we apply our bias correction (B10-RF - TCCON) for QF=0 and QF=1. For land nadir the bias correction reduces biases mostly for QF=1 data while there is practically no difference for QF=0. For ocean glint observations the bias corrected X_{CO2} exhibits a systematic positive offset compared to TCCON but a reduced variability. The systematic offset could be addressed by recalculating the scaling factor used for ocean glint retrievals in B10. However, there are only few TCCON stations that can provide comparisons for ocean glint data and these stations are not equally distributed over the ocean.

https://doi.org/10.5194/amt-2022-202
Preprint. Discussion started: 18 July 2022

Table 3: Mean and standard deviation of bias in X_{CO2} compared to TCCON observations for the testing set (08/2018 – 07/2019). The comparison for the operational OCO-2 product is indicated by (B10 - TCCON) and the proposed random forest approach by (RF - TCCON).

	LndND XCO_2 [ppm]		SeaGL XCO_2 [ppm]	
	B10 - TCCON	B10-RF - TCCON	B10 - TCCON	B10-RF - TCCON
QF=0	-0.19 ± 1.11	**-0.2 ± 1.00**	0.87 ± 0.75	**0.79 ± 0.70**
QF=1	-0.42 ± 2.27	**0.19 ± 1.63**	0.08 ± 2.08	**1.20 ± 1.72**
QF=0 + 1	-0.34 ± 1.69	**-0.18 ± 1.24**	0.65 ± 1.32	**0.88 ± 0.99**

To better understand how the bias correction addresses 3D cloud biases as compared to TCCON, Figure 6 shows X_{CO2} biases vs nearest cloud distance. For land nadir and ocean glint there exist negative biases in B10 in the proximity of clouds (Figure 6a and 6c). Interestingly, there is a positive bias for B10 ocean glint data when no clouds are close to OCO-2 soundings (> 4 km) that likely stems from B10 incorporating a multi model mean in its bias correction in addition to TCCON. After applying our bias correction, X_{CO2} biases in the proximity of clouds (< 4 km) have been mitigated for land nadir (Figure 6b). For ocean glint, the bias correction pushed X_{CO2} up by roughly 0.5 ppm in the proximity of clouds, resulting in a uniform positive bias of roughly 1 ppm independent of cloud distance (Figure 6d). Thus, the bias correction removed the dependency of X_{CO2} biases on nearest cloud distance but did not address the overall offset present in B10.

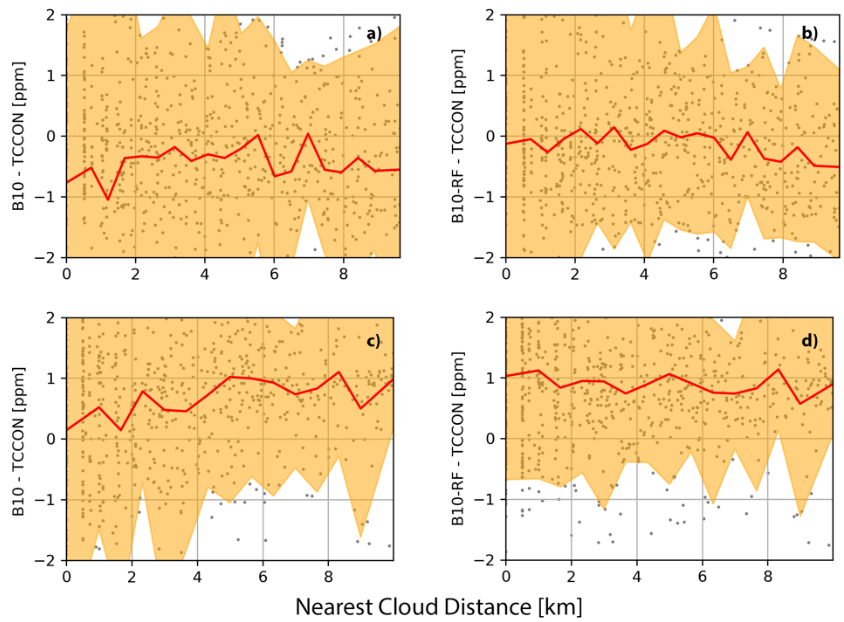

Figure 6: X_{CO2} bias vs cloud distance for land nadir B10 (a) B10 corrected (b) ocean glint B10 (c), and B10 corrected (d) for QF=0 and QF=1 data from 2018 to 2019. The 5[th] and 95[th] percentiles are indicated with the yellow shaded area, the mean is shown with a red line and individual comparisons with grey dots.

https://doi.org/10.5194/amt-2022-202
Preprint. Discussion started: 18 July 2022

5 Discussion

5.1 Model Interpretation

To better understand how the model utilizes the input features to calculate the bias correction we calculated the modelled bias (RF Bias) with respect to the three input features (xco2_strong_idp, co2_grad_del, dp) (see Figure 7). Overall, the 3D cloud bias correction depends similarly on the individual features for land and ocean observations. Additionally, the bias-feature relationship is non-linear for most features. This explains the lower model performance of the linear model we compared to in Section 4.2. **xco2_strong_idp** is positively correlated with modelled biases, thus, both the operational X_{CO2} retrieval as well as the IDP preprocessor X_{CO2} retrieval are both biased by 3D cloud effects. For ocean glint observations this relationship is roughly linear for negative biases in xco2_strong_idp below -2 ppm. Overall, the IDP preprocessor seems to be more strongly biased by 3D cloud effects than the operational retrieval, for example, a bias of -10 ppm in the IDP preprocessor (calculated by subtracting the mean of each small area) relates to a bias of -1 ppm and -3 ppm in the operational retrieval over land and ocean, respectively. **co2_grad_del** shows mostly a positive correlation for negative co2_grad del (surface CO_2 is underestimated compared to CO_2 higher up in the atmosphere) and a negative correlation for positive co2_grad_del. This indicates that 3D cloud effects challenge the accurate retrieval of the X_{CO2} profile. **dp** shows a positive correlation with X_{CO2} biases when the operationally retrieved surface pressure is underestimated. Overestimating the surface pressure shows no correlation with biases in X_{CO2}. Finally, **albedo_wco2** shows no clear dependence on X_{CO2} biases. Note, that our bias correction is applied in addition to the bias correction performed in B10 that implicitly removes some correlations of 3D cloud biases with co2_grad_del and dPfrac (highly correlated with dp).

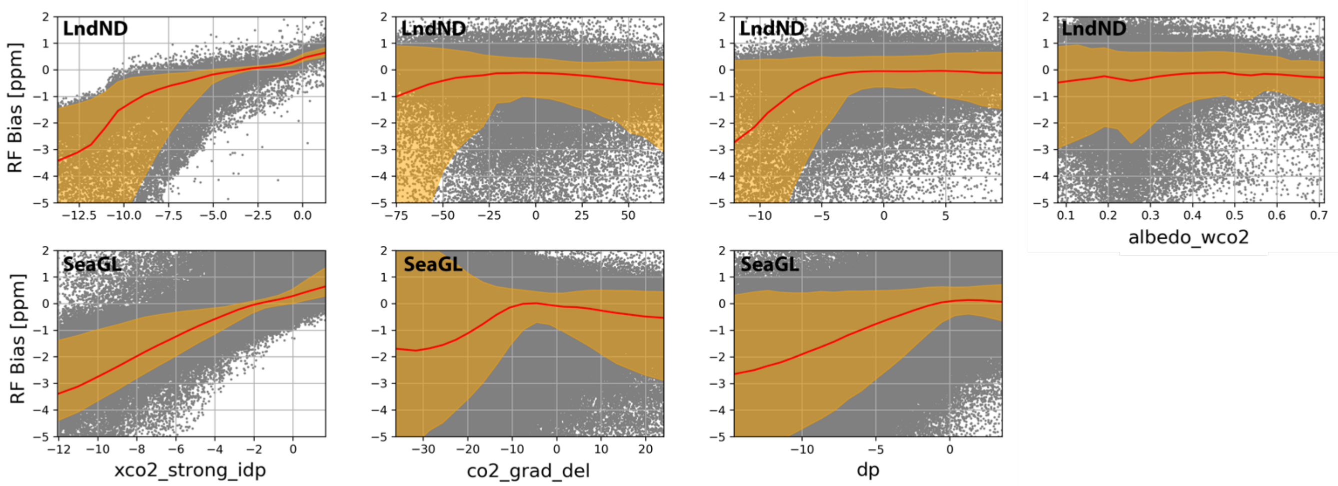

Figure 7: Bias correction by the proposed model (RF Bias) with respect to its three input features (xco2_strong_idp, co2_grad_del, dp) for land nadir (top) and ocean glint (bottom) observations for QF=0 and QF=1 data from 2018 to 2019. The 5th and 95th percentile are indicated with the yellow shaded area, the mean is shown with a red line and individual comparisons with grey dots. Note, the scale of the x-axis for each plot is different.

https://doi.org/10.5194/amt-2022-202
Preprint. Discussion started: 18 July 2022

Atmospheric Measurement Techniques
Discussions

A further look at the relative importance of the model features shows roughly the same ordering for land nadir and ocean glint observations with xco2_strong_idp being the most important feature followed by dp, co2_grad_del and, finally, albedo_wco2 for land nadir (see Figure 8). The feature importance was calculated as the normalized total reduction of mean square error brought by an individual feature. I.e., if we were to omit xco2_strong_idp from our model as a feature the bias correction would be less effective than if we were to omit co2_grad_del.

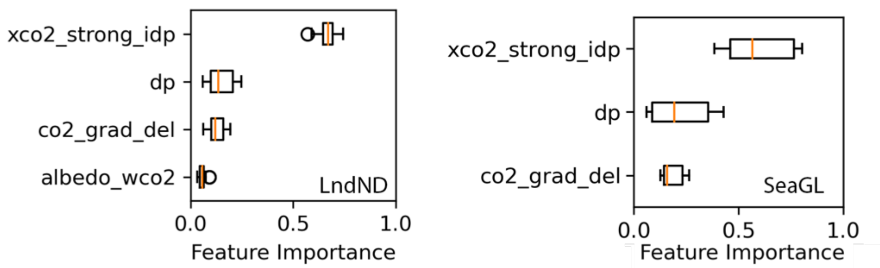

Figure 8: Feature importance for the bias correction model. Feature importance is shown for land nadir (left) and ocean glint (right) observations. Model was trained using the training set with QF=0 and QF=1 data.

5.2 Regional Biases

To further understand regional impacts of our bias correction we calculate biases, as identified by our model (RF Bias), for soundings from 2014 to 2019 and average results over 2° by 2° cells (see Figure 9). I.e., to apply the proposed bias correction, one would subtract the results shown in Figure 9 from B10 X_{CO2}. Since using soundings only from the testing set leads to many areas with no data, we used all available data (2014 - 2019) for this visualization. Over land negative biases (i.e., X_{CO2} is underestimated in B10) are present north of 45° in America, Europe and Asia, averaging -0.36 ppm and around the tropics within ±10°, averaging -0.43 ppm. Over the ocean biases are more equally distributed than over land and of lower magnitude, except closer to the poles where OCO-2 retrievals have generally higher uncertainties. When comparing the regional biases to a map of nearest cloud distance (see Figure 10) there is a high degree of overlap between negative biases over land and areas dominated by clouds. Over the ocean there is less agreement between the two. Most notable our model identified a positive bias in X_{CO2} for the tropics between 60°E and 170°E where we would expect negative biases due to 3D cloud effects. This indicates that over the ocean our bias correction, aimed at 3D cloud effects, is also correcting for other biases.

https://doi.org/10.5194/amt-2022-202
Preprint. Discussion started: 18 July 2022

Atmospheric Measurement Techniques
Discussions

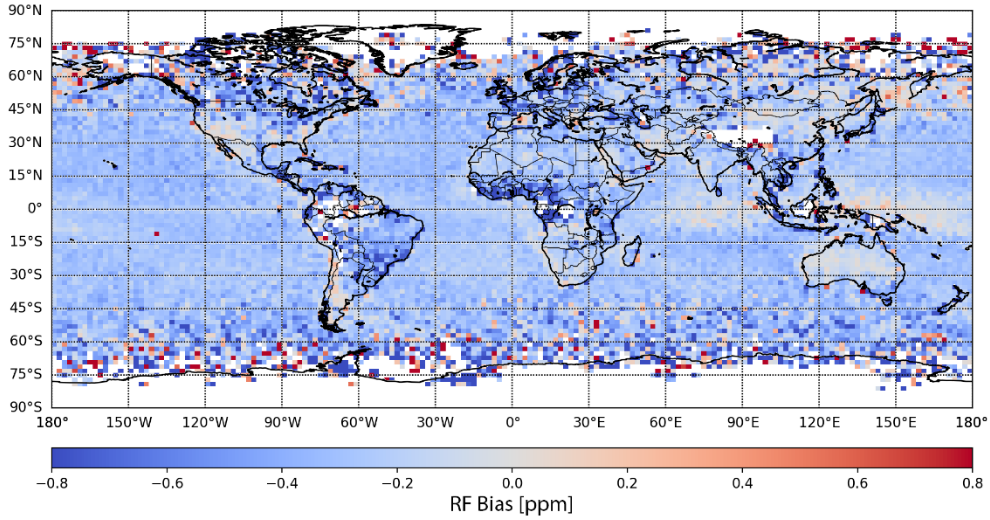

Figure 9: Biases identified by our model. Biases are averaged over 2° by 2° for all soundings (2014 to 2019, QF=0 and QF=1). Negative biases are shown in blue, positive biases in red, no data in white.

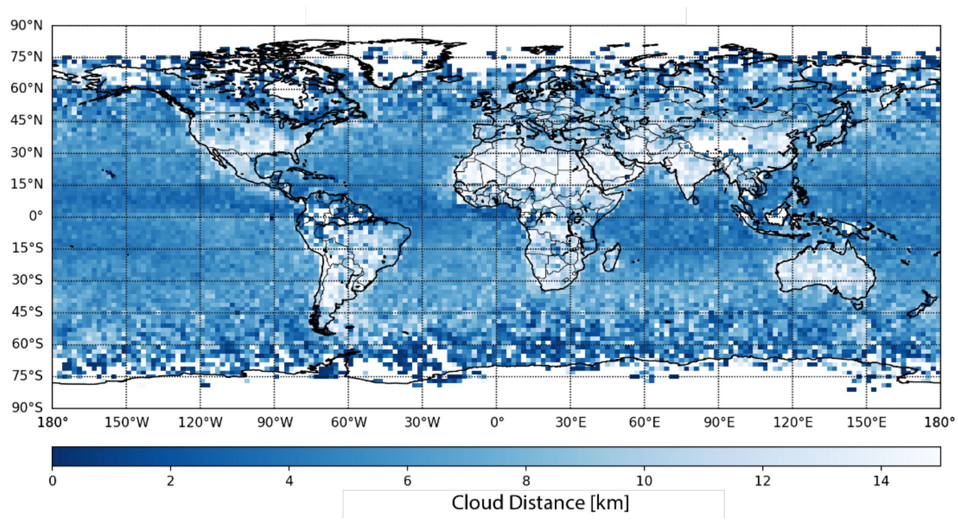

Figure 10: Nearest cloud distance derived from MODIS. Nearest cloud distances are averaged over 2° by 2° for all matched soundings (2014 to 2019, QF=0 and QF=1). Darker blues indicate closer clouds, no data is shown in white.

5 Future Work and Conclusion

5.1 Future Work

The developed bias correction approach was aimed at mitigating 3D cloud biases in B10, but could readily be expanded to a more general bias correction. Future research will need to show in how far the approach used in this research (determining the bias correction solely from small area biases) will work for correcting raw X_{CO2}. For such a correction a two-step approach might be necessary that combines a global (comparison to TCCON) and local (small area analysis) bias correction approach. However, developing such an approach would be challenged by the sparse coverage of TCCON stations.

The bias correction used for B10 is aimed at QF=0 data. This is highlighted by the significant reduction in X_{CO2} biases our correction was able to achieve on QF=1 data while improvements on QF=0 data where moderate. Filtering out low quality data is a simple approach to improve the overall quality of the OCO-2 X_{CO2} retrieval. However, it leaves certain areas with too few samples, most notably the tropics (due to clouds), higher latitudes (due to shallow solar zenith angles) and around Brazil, Bolivia, Paraguay (due to the South Atlantic Anomaly). Improving the bias correction of future OCO-2 versions that allow for less restrictive filtering would benefit applications that rely on those data.

Finally, one could expand the approach taken here, developing one model for land nadir and one for ocean glint data, to having multiple models for land and ocean to better capture the diverse causes for biases in XCO2 across Earth, for example, different types of aerosols dominate different areas and might lead to specific biases in different regions or seasons. Such a location based bias correction could also be expanded to a location-based filtering approach that would, for example, allow less restrictive filtering at higher latitudes (Jacobs et al., 2020; Mendonca et al., 2021) to have more of those soundings pass the filter and be available for scientific inquiry. A key challenge of such an approach will be validation due to the limited number of available TCCON stations.

5.2 Conclusion

We identified four variables (xco2_strong_idp, co2_grad_del, dp, albedo_wco2) that allow to correct for 3D cloud biases in B10 X_{CO2} from OCO-2. All variables are bi-products of the operational retrieval used by OCO-2 which simplifies their inclusion for bias correction in future versions of the operational product. Using the identified variables, we were able to reduce the remaining 3D cloud biases further than using dedicated cloud variables. The proposed non-linear bias correction is based on a random forest approach and able to reduce the RMSE from 2.08 ppm to 1.43 ppm over land and 1.35 ppm to 0.74 ppm over the ocean for QF=0 and QF=1 data on an independent testing set. We demonstrated a systematic approach to correct for biases in OE retrievals. Namely, (1) find a physical variable that is well understood and correlated with the cause of the bias (in our case cloud distance). (2) Identify elements from the retrieved state vector and other retrieval products that show a dependence to the variable from step (1) in addition to other variables that have a physical connection to the bias. (3) Use recursive feature elimination to identify which subset of the elements identified in (2) should be used for the bias correction. (4) Use a simple explainable machine learning model to map the features identified in (3) to the biases and correct for them.

https://doi.org/10.5194/amt-2022-202
Preprint. Discussion started: 18 July 2022

Author contribution

SMAU, SMAS, and SS conceptualized the research goals. SMAU and SMAS prepared the various datasets. SMAU developed the approach, implemented the experiments and visualized the results. SMAU prepared the manuscript with contributions from all co-authors.

The authors declare that they have no conflict of interest.

Acknowledgement

We acknowledge support by NASA grant 80NSSC21K1063 "Mitigation of 3D cloud radiative effects in OCO-2 and OCO-3 X_{CO2} retrievals". We appreciate the TCCON teams, who measure and provide ground-based X_{CO2} validation to the carbon cycle research community (Dubey, Henderson, et al., 2014; Dubey, Lindenmaier, et al., 2014; L. Iraci et al., 2014; L. T. Iraci et al., 2016; Toon & Wunch, 2014; Wennberg, Roehl, et al., 2016; Paul O. Wennberg et al., 2014; P. O. Wennberg, Debra Wunch, C. Roehl, et al., 2014; Wennberg, Wunch, et al., 2016; P. O. Wennberg, Debra Wunch, Y. Yavin, et al., 2014; Wunch et al., 2017; Wunch et al., 2015).

The research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (80NM0018D0004).

References

Breiman, L. (2001). Random forests. *Machine learning, 45*(1), 5-32.

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), 700-709. doi:https://doi.org/10.1016/j.asr.2003.08.062

Dubey, M., Henderson, B., Green, D., Butterfield, Z., Keppel-Aleks, G., Allen, N., . . . Lindenmaier, R. (2014). TCCON data from Manaus (BR), Release GGG2014R0. In. Pasadena, California.

Dubey, M., Lindenmaier, R., Henderson, B., Green, D., Allen, N., Roehl, C., . . . Wunch, D. (2014). TCCON data from Four Corners (US), Release GGG2014R0. In. Pasadena, California.

Eldering, A., O'Dell, C. W., Wennberg, P. O., Crisp, D., Gunson, M. R., Viatte, C., . . . Yoshimizu, J. (2017). The Orbiting Carbon Observatory-2: first 18 months of science data products. *Atmos. Meas. Tech., 10*(2), 549-563. doi:https://doi.org/10.5194/amt-10-549-2017

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. *Journal of the royal statistical society. series c (applied statistics), 28*(1), 100-108.

Iraci, L., Podolske, J., Hillyard, P., Roehl, C., Wennberg, P. O., Blavier, J. F., . . . Barney, J. (2014). TCCON data from Indianapolis (US), Release GGG2014R0. In. Pasadena, California.

Iraci, L. T., Podolske, J., Hillyard, P. W., Roehl, C., Wennberg, P. O., Blavier, J. F., . . . Albertson, R. (2016). TCCON data from Edwards (US), Release GGG2014R1. In. Pasadena, California.

Jacobs, N., Simpson, W. R., Wunch, D., O'Dell, C. W., Osterman, G. B., Hase, F., . . . Heikkinen, P. (2020). Quality controls, bias, and seasonality of CO2 columns in the boreal forest with Orbiting Carbon Observatory-2, Total Carbon Column Observing Network, and EM27/SUN measurements. *Atmos. Meas. Tech., 13*(9), 5033-5063. doi:10.5194/amt-13-5033-2020

Jet Propulsion Laboratory. (2018). *Orbiting Carbon Observatory–2 (OCO-2) Data Product User's Guide, Operational L1 and L2 Data Versions 8 and Lite File Version 9*. Retrieved from https://docserver.gesdisc.eosdis.nasa.gov/public/project/OCO/OCO2_DUG.V9.pdf

Kiel, M., O'Dell, C. W., Fisher, B., Eldering, A., Nassar, R., MacDonald, C. G., & Wennberg, P. O. (2019). How bias correction goes wrong: measurement of XCO2 affected by erroneous surface pressure estimates. *Atmos. Meas. Tech., 12*(4), 2241-2259. doi:10.5194/amt-12-2241-2019

Liang, L., Di Girolamo, L., & Platnick, S. (2009). View-angle consistency in reflectance, optical thickness and spherical albedo of marine water-clouds over the northeastern Pacific through MISR-MODIS fusion. *Geophysical research letters, 36*(9). doi:https://doi.org/10.1029/2008GL037124

Massie, S., Cronk, H., Merrelli, A., O'Dell, C., Schmidt, K. S., Chen, H., & Baker, D. (2021). Analysis of 3D cloud effects in OCO-2 XCO2 retrievals. *Atmos. Meas. Tech., 14*(2), 1475-1499. doi:10.5194/amt-14-1475-2021

Massie, S., Sebastian, S., Eldering, A., & Crisp, D. (2017). Observational evidence of 3-D cloud effects in OCO-2 CO2 retrievals. *Journal of Geophysical Research: Atmospheres, 122*(13), 7064-7085. doi:https://doi.org/10.1002/2016JD026111

Mendonca, J., Nassar, R., O'Dell, C. W., Kivi, R., Morino, I., Notholt, J., . . . Wunch, D. (2021). Assessing the feasibility of using a neural network to filter Orbiting Carbon Observatory 2 (OCO-2) retrievals at northern high latitudes. *Atmos. Meas. Tech., 14*(12), 7511-7524. doi:10.5194/amt-14-7511-2021

Merrelli, A., Bennartz, R., O'Dell, C. W., & Taylor, T. E. (2015). Estimating bias in the OCO-2 retrieval algorithm caused by 3-D radiation scattering from unresolved boundary layer clouds. *Atmos. Meas. Tech., 8*(4), 1641-1656. doi:10.5194/amt-8-1641-2015

O'Dell, C., Eldering, A., Wennberg, P. O., Crisp, D., Gunson, M. R., Fisher, B., . . . Velazco, V. A. (2018). Improved retrievals of carbon dioxide from Orbiting Carbon Observatory-2 with the version 8 ACOS algorithm. *Atmos. Meas. Tech., 11*(12), 6539-6576. doi:https://doi.org/10.5194/amt-11-6539-2018

Okata, M., Nakajima, T., Suzuki, K., Inoue, T., Nakajima, T., & Okamoto, H. (2017). A study on radiative transfer effects in 3-D cloudy atmosphere using satellite data. *Journal of Geophysical Research: Atmospheres, 122*(1), 443-468. doi:https://doi.org/10.1002/2016JD025441

Rodgers, C. D. (2000). *Inverse methods for atmospheric sounding: theory and practice* (Vol. 2): World scientific.

Toon, G. C., & Wunch, D. (2014). A stand-alone a priori profile generation tool for GGG2014. In. Pasadena, California.

Wennberg, P. O., Roehl, C., Blavier, J. F., Wunch, D., Landeros, J., & Allen, N. (2016). TCCON data from Jet Propulsion Laboratory (US), 2011, Release GGG2014R1. In. Pasadena, California.

Wennberg, P. O., Roehl, C., Wunch, D., Toon, G. C., Blavier, J. F., Washenfelder, R. a., . . . Ayers, J. (2014). TCCON data from Park Falls (US), Release GGG2014R0. In. Pasadena, California.

Wennberg, P. O., Wunch, D., Roehl, C., Blavier, J. F., Toon, G. C., & Allen, N. (2014). TCCON data from Caltech (US), Release GGG2014R1. In. Pasadena, California.

Wennberg, P. O., Wunch, D., Roehl, C., Blavier, J. F., Toon, G. C., Allen, N., . . . Martin, J. (2016). TCCON data from Lamont (US), Release GGG2014R1. In. Pasadena, California.

Wennberg, P. O., Wunch, D., Yavin, Y., Toon, G. C., Blavier, J. F., Allen, N., & Keppel-Aleks, G. (2014). TCCON data from Jet Propulsion Laboratory (US), 2007, Release GGG2014R0. In. Pasadena, California.

Wunch, D., Mendonca, J., Colebatch, O., Allen, N., Blavier, J.-F. L., Roche, S., . . . Strong, K. (2017). TCCON data from East Trout Lake (CA), Release GGG2014R1. In. Pasadena, California.

Wunch, D., Toon, G. C., Sherlock, V., Deutscher, N. M., Liu, C., Feist, D. G., & Wennberg, P. O. (2015). *The Total Carbon Column Observing Network's GGG2014 Data Version*. Retrieved from Pasadena, California: http://dx.doi.org/10.14291/tccon.ggg2014.documentation.R0/1221662

Wunch, D., Toon, G. C., Wennberg, P. O., Wofsy, S. C., Stephens, B. B., Fischer, M. L., . . . Biraud, S. C. (2010). Calibration of the Total Carbon Column Observing Network using aircraft profile data. *Atmospheric Measurement Techniques, 3*(5), 1351-1362. doi:https://doi.org/10.5194/amt-3-1351-2010