Correcting 3D cloud effects in X_{CO2} 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 CO_2 dry-air mole fractions (X_{CO2}). However, the retrieved X_{CO2} are biased due to calibration issues and mismatches between the physics-based retrieval radiances and observed radiances~~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 develop~~ed~~ a bias correction model to address these 3D cloud biases. The model is able to reduce unphysical variability over land and ~~ocean~~ sea by ~~3~~1~~1~~0% and ~~5~~5~~3~~0%, 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 the published OCO-2 data record underestimates X_{CO2} over land by -0.~~4~~ 3 ppm in the tropics and -0.2 ppm northward of 45° N. The approach can be expanded to a more general bias correction and is generalizable to other greenhouse gas experiments~~missions~~, such as GeoCarb, GOSAT-3 and CO2M.

1 Introduction

The Orbiting Carbon Observatory OCO-2 (Eldering et al., 2017; Crisp et al., 2004) makes space-based top-of-atmosphere radiance measurements in three spectral bands: Oxygen A band at 0.76 µm, the Weak CO_2-band at 1.61 µm, and the Strong CO_2 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 fractions (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 pressure, surface reflectance, aerosols, trace gas concentration) within

the field of view of this sounding. This approximation exploits that for most clear sky observations there is no significant horizontal exchange of photons.

However, Nnearby clouds, however, can scatter a significant number of photons into the field of view of OCO-2 which enhances the observed radiance. This horizontal exchange of photons due to clouds, or 3D cloud effect,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 is not accounted for in the ACOS retrieval. Nevertheless, the forward model attempts to match the enhanced radiances which leads to errors in the converged state vector and most importantly,leads to negative biases in retrieved X_{CO2} in the vicinity of clouds (Massie et al., 2021; Massie et al., 2017; Merrelli et al., 2015; Emde et al., 2022; Kylling et al., 2022; Yu et al., 2021). Merrelli et al. (2015) applied the Spherical Harmonics Discrete Ordinate Method (SHDOM) 3D radiative transfer code (Evans, 1998) to perturb OCO-2 type spectra, and calculated OCO-2 retrievals without and with the 3D radiance perturbations. Retrieved X_{CO2} values were *lower* than clear sky retrievals by 0.3, 3, and 5-6 ppm for surfaces characterized by bare soil, vegetation, and snow-covered footprints, respectively. From an empirical perspective, Fig. 6 of Massie et al. (2021) demonstrates that retrievedocean X_{CO2} over sea generally decreases in value when the distance between observations and clouds becomes less than 5 km.

Nearby clouds can also cause radiance dimming due to cloud shadows. But only 25 % of observed radiances pass into the operational retrieval, s However, recent work indicates that cloud shadows play an insignificant role for retrievals from OCO-2 compared to radiance enhancements due to horizontal photon scattering (Massie et al., in prep).ince two cloud pre-processors (Taylor et al., 2016) exclude many observed radiances. Cloud brightening occurs on both sides of clouds since 40% of OCO-2 observations are within 4 km of clouds (Massie et al., 2021), and cloud brightening extends over a 5 to 10 km horizontal scale. A cloud shadow occurs only on one side of a cloud, with the shadow covering a limited angular portion of the side. Since the majority of OCO-2 observations are next to low-level clouds (think of an observation embedded in low-level Amazon cloud streets), the cloud shadows project only about aone km or so from the low-level clouds. -Using a year's data volume, Massie et al. (in prep.) discuss detailed calculations, based on an analysis of OCO-2 O2 A-band continuum radiances, that yield an estimate of cloud shadowing frequency to be on the order of 4%, compared to 96% for the observations influenced by cloud brightening.

To mitigate biases in retrieved "raw" X_{CO2}, a linear bias correction and threshold-based filtering is applied to the data, yielding "biased corrected" X_{CO2}–. For the current version (B10) bBias 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' 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 biases in retrieved X_{CO2}, that have been linked to 3D cloud effects, 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 latest public data record provided by OCO-2 (B10) lite files (https://disc.gsfc.nasa.gov/datasets/OCO2_L2_Lite_FP_10r/, last access: 05/2022) data from September 2014 to July 2019. These files contain bias corrected X_{CO2} for soundings over sea in glint mode (in which instrument tracks near the location where sunlight is directly reflected by the Earth's surface upwards towards OCO-2) ocean glint and soundings over land with a nadir viewing geometry. observations We that we wish to correct for remaining 3D cloud biases by utilizing s, a variety of parameters describing the retrieved atmospheric state vector, viewing and solar geometry, results from OCO-2 cloud screening 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}. High-quality data has a QF=0 and low-quality data has QF=1. Similarly, the operational 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 additionaddition, 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 variables indicative of 3D cloud effects (Massie et al., 2021) features: H3D, HC, CSNoiseRatio, and Cloud Distance. H3D (Liang et al., 2009; Massie et al., 2017) describes the normalized standard deviation of the MODIS radiance field, and is calculated based on off-line MODIS radiance data files (Cronk, 2018). The radiance standard deviation is calculated in a circle with aof radius of 10 km surrounding each OCO-2 data point. HC is calculated from differences in O2 A-band continuum radiances of an observation point and adjacent points in three rows (frames) of footprints. A frame has eight adjacent OCO-2 footprints, with each footprint on the order of 2 km in size., CSNoiseRatio is the ratio of the O2 A-band continuum radiance spatial standard deviation and noise level, calculated within a footprint (which has 20 "ColorSlice" sub-pixel elements). These three variables are indicative of 3D cloud effects since radiance gradients are present when clouds are

next to observation footprints (radiance enhancements become larger as cloud distance decreases). ~~and~~ Cloud Distance (Massie et al., 2021) is the distance of the nearest cloud to each observation point, as determined from t~~MODIS imagery. T~~he off-line radiance data files (Cronk, 2018), which contain 500 m MODIS radiances, geolocation and cloud mask data. C~~e~~alculated 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 soundings, ~~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 (though, 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 with a maximum size of 100 km. Each small area is generated by collecting soundings (ordered by observation~~s~~ time) until the distance between the first and last sounding exceeds the 100 km threshold. Afterwards, the collection process of the next small area is started~~using the k-means algorithm~~. ~~This groups soundings where variations in X_{CO2} can be interpreted as non-physical variability, or retrieval biases.~~ For each small area ~~group~~ we define the median retrieved ~~B10 bias corrected~~ X_{CO2} of this ~~group~~ small area 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). Additionally, this processing~~it~~ will interpret real X_{CO2} enhancements, for example from power plants, as positive biases. However, we postulate that these cases are rare and that a model that is robust to outliers can still learn a useful bias correction from these data. Next, we remove outliers with large X_{CO2} errors by applying a series of thresholds to the variables from the state vector. The variables and their thresholds are given in Table 1. Note~~,~~ that these filters remove only a small fraction of soundings (4%) and are not comparable to the operational quality flag processings ~~used by OCO-2~~. ~~Next~~Finally, we remove small areas ~~groups where soundings cover an area bigger than 100 km or there are~~with fewer than 20 soundings. This results in approximately 10^6 soundings over land ~~nadir soundings~~ and $11 \cdot 10^6$ soundings over the ocean~~sea~~~~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 N~~n~~orthern ~~hemisphere~~ Hemisphere and ~~on~~ over 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.

Table 1: Variables and their thresholds used to remove outliers

Variable	Description	Land	Sea
co2_ratio	Ratio of retrieved X_{CO2} in WCO_2 and SCO_2 bands	x < 1 or x > 1.04	x < 1 or x > 1.03
co2_grad_del	Change between the retrieved CO_2 profile and the a priori profile	x < -100 or x > 100	x < -50 or x > 100
deltaT	Retrieved offset to a priori temperature profile		x < 0
dpfrac	Retrieved X_{CO2} multiplied by difference in retrieved and a priori surface pressure (Kiel et al., 2019)	x > 7	
rms_rel_sco2	Root Mean Squared error of the L2 fit residuals for the SCO_2 band, relative to the continuum signal		x > 0.5
snr_sco2	Signal-to-noise ratio in SCO_2 band		x < 200

The distribution of nearest cloud distance, biases from the small area analysis and comparison to TCCON for land nadir and ~~ocean~~ sea 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 in the current OCO-2 product (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 soundings over sea ~~ocean glint~~. For soundings over land ~~nadir B10~~the small area bias and bias compared to ~~B10~~ TCCON are similar with a mean and standard deviation of 0.1 ± ~1 ppm. For QF=1 the distribution of biases ~~has~~ have a larger standard deviation for ~~B10~~ small area biases (land: 2.9 ppm, ~~ocean~~sea: 1.9 ppm) and compared to ~~B10~~ TCCON (land: 3.9 ppm, ~~ocean~~sea: 2.1 ppm), ~~is~~ are skewed, and contain~~s~~ negative biases that far exceed positive biases, as analysed with the small areas (land: -0.5 ppm, ~~ocean~~sea: -0.9 ppm) and compared to TCCON (land: -1.4 ppm, ~~ocean~~sea: -1.2 ppm). This long tail distribution of negative biases is indicative of 3D cloud effects (Massie et al., 2021) and should be mitigated with a successful 3D cloud bias correction.

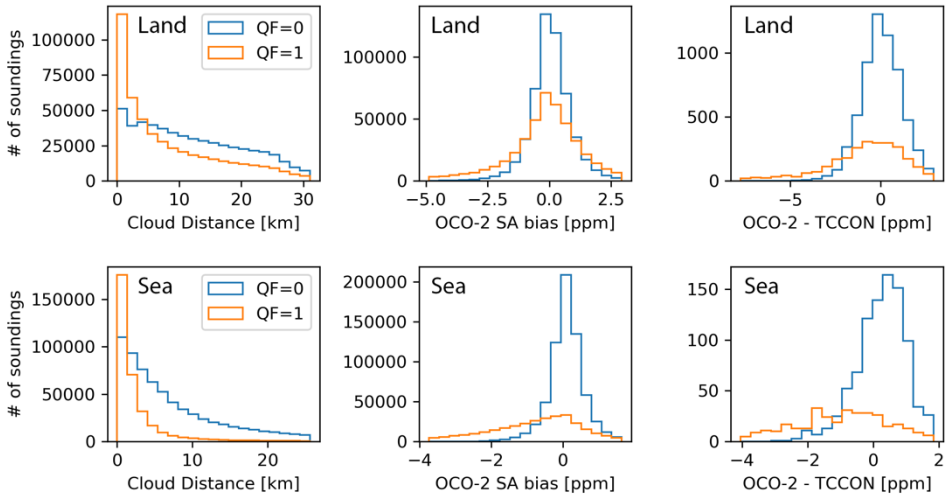

Figure 1: Histogram of data used in this study for nearest cloud distance (left), small area (SA) biases (middle), and biases compared to TCCON (right) for soundings over land ~~nadir~~ (top) and ~~ocean glint~~sea (bottom) ~~soundings~~. Higher quality data (QF=0) ~~data~~ is shown in blue, lower quality data (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 over sea ~~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 the 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~~ over the sea~~glint~~ ~~for~~with QF=0 and QF=1 ~~data~~, respectively, and $17 \cdot 10^4$ and $14 \cdot 10^4$ ~~for~~ soundings over land ~~for~~with ~~nadir~~ QF=0 and QF=1 ~~data~~, 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). Random Forests are easy to interpret and robust to outliers. 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 model performance on the validation set. The model inputs are a set of selected features from the OCO-2 retrieved state vector ~~Lite files~~ (e.g. co2_grad_del) and the model output is the remaining ~~bias in the B10 bias corrected~~ X_{CO2} bias derived from the small areas analysis.

Since the operational OCO-2 ~~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 (Hoerl and Kennard, 1970a, b). Thus, using the training set we seek to find the weights, $\boldsymbol{w}$, that minimize the following equation:

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

where $\boldsymbol{y}$ is the standardized (mean removed and divided by standard deviation) X_{CO2} bias, $\boldsymbol{X}$ are the standardized features, $\|\cdot\|_2$ is the Euclidean norm, 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 state 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~~ Two 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 are amplified at large solar zenith angles and for ~~so does a~~ brighter surface~~s albedo~~ (Okata et al., 2017). Finally, we removed highly correlated variables. This results in a set of 23 features for soundings over land ~~nadir~~, and 24 features for soundings over ~~ocean glint~~sea ~~soundings~~, that may be used to correct for 3D cloud biases in retrieved X_{CO2} (more information about each variable can be found o~~f~~n pages 29 to 40 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~~ sea ~~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 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 four~~three~~ most important variables for land and ~~ocean glint~~sea soundings~~and four most important variables for land nadir~~, as identified by the feature elimination. These variables explain most of the variance and ~~mostly~~ partially overlap for land and ~~ocean~~sea~~, a further indication that those variables have a robust relationship to 3D cloud biases~~. For land ~~nadir and ocean glint soundings~~ the four~~three~~ most important variables are ~~x~~**dp_abp~~co2_strong_idp~~** (retrieved surface pressure minus surface pressure from the GEOS-5 FP-IT model~~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~~), **h2o_ratio** (ratio of retrieved H_2O column from the WCO_2 band to that from the SCO_2 band), **co2_grad_del** (a measure of the difference in the retrieved and prior CO_2 vertical gradient~~change between the retrieved CO_2 profile and the a priori profile from the surface minus that at level 13~~), and ~~**dp_**~~**aod_water** (retrieved extinction optical depth of cloud water at 755 nm)~~difference of retrieved surface pressure and a priori surface pressure obtained from GMAO GEOS5-FP-IT model)~~. For sea the four most important variables are dp_abp, co2_grad_del, **albedo_sco2** (retrieved Lambertian albedo in the SCO_2 band), and **albedo_wco2** (retrieved Lambertian albedo in the WCO_2 band). ~~The fourth most important variable for land nadir is **albedo_wco2** (surface albedo in the weak CO_2 band).~~ Note that~~,~~ the final set of features does not include any of the 3D cloud metrics used in the bias correction discussed in~~by~~ Massie et al. (2021). Additionally, solar and viewing geometry were removed in the iterative process. However, the process~~it~~ includes the surface albedo in the strong and weak CO_2 band and dp_abp ~~albedo_wco2~~ which ha~~ve~~s a direct physical connection to 3D cloud effects. As discussed below in relation to Fig. 2b, dp_abp and nearest cloud distance are empirically correlated. ~~-~~Additionally, increased values in aod_water and deviations from unity for h2o_ratio are indicative of cloud contamination (Jet Propulsion Laboratory, 2018). ~~-~~This indicates that elements of the operational retrieval state vector (co2_grad_del, dp_abp, h2o_ratio, albedo_wco2, albedo_sco2) ~~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, using elements from the current retrieval state vector to correct 3D cloud biases ~~for OCO-2 the three features are available in the OCO-2 Lite files which~~ simplifies the bias correction ~~fo~~in~~r~~ ~~their inclusion in~~ future operational products. It also is ~~allows to apply the approach~~ more generally applicable to other missions that might not have available coincident cloud field measurements, ~~available~~ that can be applied to derive ~~the distance to the~~ nearest cloud distances, such as OCO-3 (Eldering et al., 2019). On the other hand, it reduces the interpretability of the developed model and does not allow to directly link 3D cloud biases to 3D cloud metrics. The OCO-2 and 3D cloud variables and their meaning are summarized in Table 2.

Note that it is not possible to clearly separate biases due to 3D cloud effects and other mismatches between the forward model of the retrieval algorithm and the observed radiances. For example, differences in modelled and real aerosol optical properties (Chen et al., 2022) or uncertainties in absorption profiles of various trace gases (Payne et al., 2020) likely are important. Additionally, uncertainties in the instrument calibration can cause systematic biases as well. Thus, some of the features might also correct for non-3D cloud effects. However, we tried to mitigate the effect of non-3D cloud biases by only adding features to the feature selection process that show some dependence to nearest cloud distance (see Figure 2) or have a direct physical relationship to 3D cloud biases. Additionally, our bias correction is applied to data that has already been corrected with the operational OCO-2 bias correction (our processing utilizes bias corrected X_{CO2}). Thus, biases independent to 3D cloud effects should be minimized.

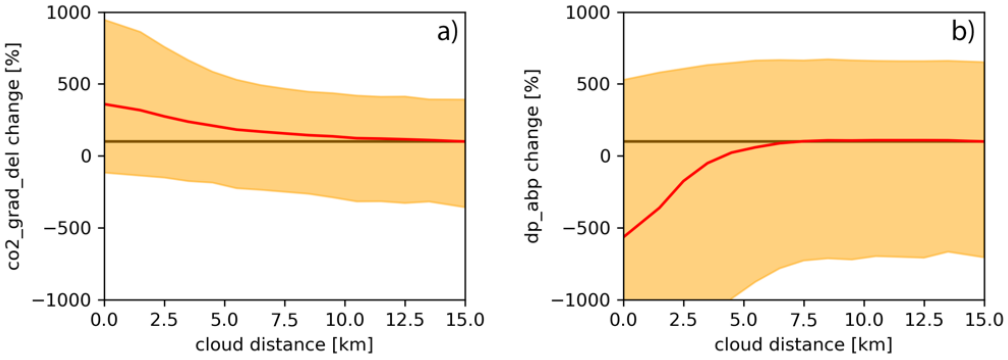

Figure 2: Change of variability and mean in percent of potential ~~model~~ features with respect to nearest cloud distance. Change in mean is shown in red; change of the 5th and 95th percentile is shown in yellow; no change (baseline) is shown with a brown straight line. Change is calculated with respect 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 be) dp_abp ~~(the mean of 'dp change' for a cloud distance of 0 km is -1300%)~~. Please refer to the text for a description of the two features.

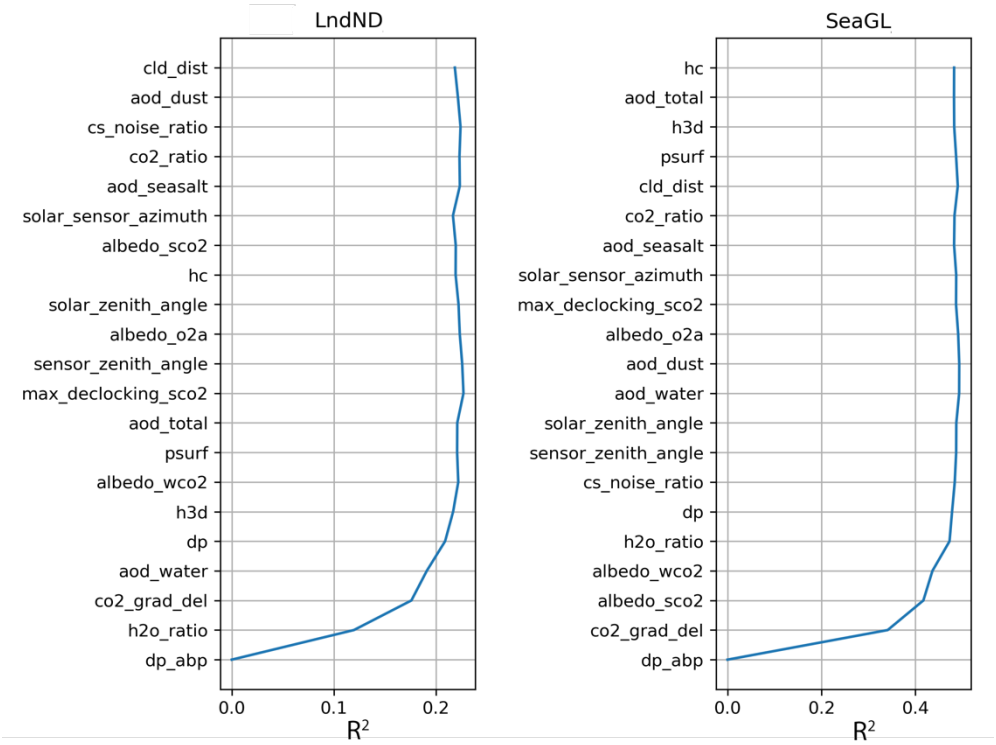

Figure 3: Feature ordering by importance as determined by recursive feature elimination. Features were removed from top to bottom with the most important features on the bottom. The model performance for removing a given feature is indicated with R^2 calculated on the validation set. Please refer to the pages 29 to 40 in (Jet Propulsion Laboratory, 2018) for a description of the individual features.

Table 2: Summary of OCO-2 state vector variables and 3D cloud variables

Variables	Description
dp_abp	Retrieved surface pressure minus surface pressure from forecast model
h2o_ratio	Ratio of retrieved H_2O column from WCO_2 band to SCO_2 band
co2_grad_del	Change between retrieved CO_2 profile and a priori profile
aod_water	Retrieved extinction optical depth of cloud water
albedo_sco2	Retrieved surface albedo in SCO_2 band
albedo_wco2	Retrieved surface albedo in WCO_2 band
H3D	Normalized standard deviation of the radiance field

HC	Differences in continuum radiances of an observation to adjacent observations
CSNoiseRatio	Ratio of continuum radiance spatial standard deviation and noise level
Cloud Distance	Distance to the nearest cloud

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~~ OCO-2 (as determined by the small areas analysis) with biases after our correction is applied (OCO-2 corr.~~B10 RF~~) for QF=0 and QF=1 soundings. For land ~~nadir~~ soundings X_{CO2} biases are reduced from a Root Mean Square Error (RMSE) ~~standard deviation~~ of 1.7~~8~~ ppm to 1.~~4~~6 ppm (see Figure 4c) ~~with the biggest correction applied to soundings that have biases less than -1.5 ppm~~. For sea~~ocean glint~~ soundings the bias correction has a significantly bigger impact and reduces biases from 1.2~~3~~ ppm to 0.8~~7~~ ppm (see Figure 4d). Over the ~~ocean~~ sea the bias correction mostly corrects negative biases less than -0.8 ppm (see Figure 4b).

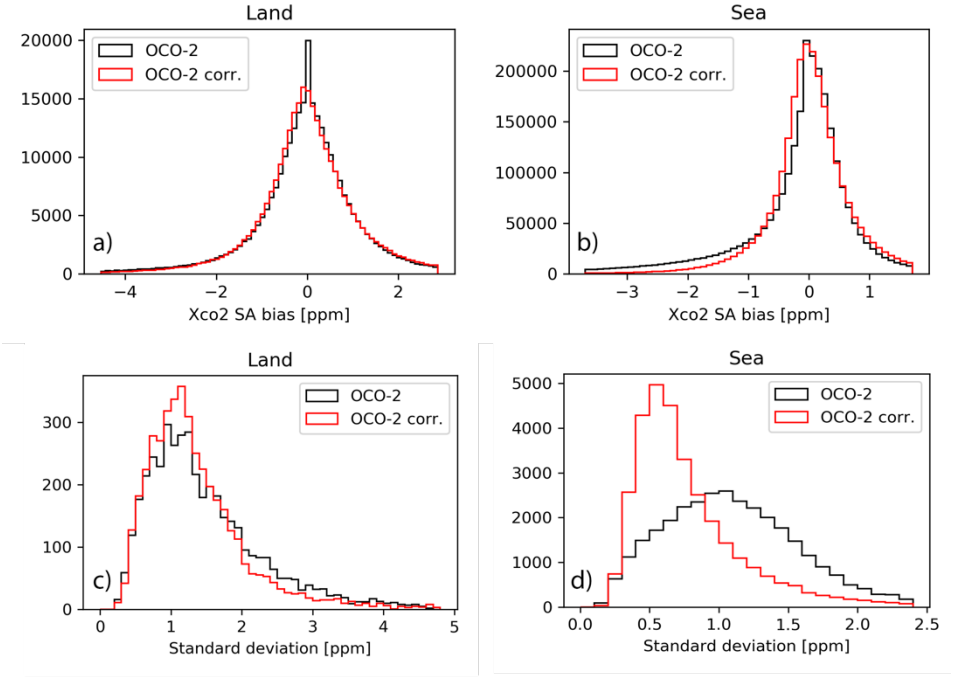

Figure 4: Reduction in non-physical variability in X_{CO2} for ~~B10~~ OCO-2 and the proposed bias correction approach (OCO-2 corr.~~B10 RF~~) for land ~~nadir~~ (left) and ~~ocean glint~~sea (right) for QF=0 and QF=1 data from 2018 to 2019. (Top) distribution of biases from individual soundings; (bottom) distribution of standard deviation for individual small areas.

Table 3~~1~~ shows the ~~Root Mean Square Error (RMSE)~~ RMSE by quality flag. For QF=0 and QF=1 the biases in X_{CO2} corrected with our model are less than the~~for~~ testing set operational ~~B10~~ OCO-2 biases~~on the testing set~~. However, for QF=0 improvements by our correction (~~B10 RF~~OCO-2 corr.) compared to ~~B10~~ OCO-2 are small (<~~-~~10%). The~~s~~ QF=0 ~~e~~data have significantly fewer soundings with clouds in close proximity (see Figure 1) which explains in part the smaller difference. Additionally, the quality flags are determined so that the operational linear bias correction of OCO-2 works well i.e., X_{CO2} biases have a mostly linear relationship to elements of the state vector where QF=0. ~~-~~For QF=1 the difference is more significant, reducing the RMSE from 2.~~76~~27 ppm to ~~1.91~~2.02 ppm over land and from 1.~~35~~86 ppm to ~~0.74~~1.22 ppm over ~~ocean~~sea.

Table 3~~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 (OCO-2~~B10~~), the proposed ~~random forest~~bias correction approach (~~B10 RF~~OCO-2 corr.~~),~~), a linear bias correction using the same ~~three~~ features than the proposed approach ~~RF~~ (~~B10 Ridge~~OCO-2 lin. corr.), and a random forest using dedicated cloud metrics (OCO-2 cloud corr.)~~, and a random forest using dedicated cloud metrics (B10 Cloud)~~. The data is separated by high quality data (QF=0), low quality data (QF=1), and all data (QF=0 + 1).

	~~Lnd~~ND ~~-~~Land X_{CO2} [ppm]				Sea~~GL~~ X_{CO2} [ppm]			
	~~B10~~OCO-2	OCO-2 corr.~~B10 RF~~	OCO-2 lin. corr.~~B10 Ridge~~	OCO-2 cloud corr.	OCO-2~~B10~~	OCO-2 corr.~~B10 RF~~	OCO-2 lin. corr.~~B10 Ridge~~	OCO-2 cloud corr.
QF=0	0.83	**0.~~75~~81**	0.~~77~~82	0.82	0.5~~1~~2	**0.4~~6~~4**	0.~~47~~49	0.46
QF=1	2.~~76~~27	**~~1.91~~2. 02**	2.~~19~~15	2.19	~~2.12~~1.8 6	**1.~~18~~22**	1.~~44~~45	1.34
QF=0 + 1	~~2.08~~1. 73	**~~1.43~~1. 55**	1.~~67~~64	1.69	1.~~35~~19	**0.~~74~~83**	0.~~89~~97	0.88

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~~ OCO-2 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~~sea, 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.

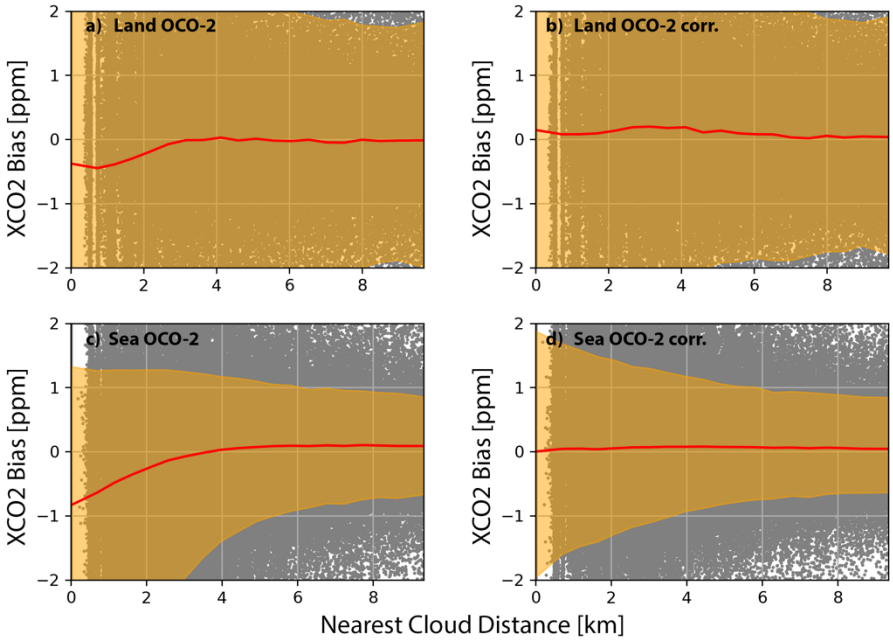

Figure 5: X_{CO2} bias vs cloud distance for OCO-2 soundings over land ~~nadir B10~~ (a), soundings over land ~~nadir B10~~ corrected by our proposed method (b), ~~ocean glint B10~~OCO-2 soundings over sea (c), and soundings over sea that are ~~ocean glint B10~~ corrected (d) for QF=0 and QF=1 data from 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 sets used~~than~~ for the random forest model development. The RMSE for the linear model (OCO-2 lin. corr.~~B10 Ridge~~) and non-linear model (OCO-2 corr.~~B10 RF~~) are~~is~~ shown in Table 3~~1~~. For QF=0 land ~~nadir~~ and ~~ocean glint~~sea 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~~15 ppm to ~~1.91~~2.02 ppm over land and from ~~1.44~~0.97 ppm to ~~1.18~~0.83 ppm over the ~~ocean~~sea.

4.3 Comparison to Using Dedicated Cloud Variables

A second question that we wanted to answer was whether additional variables from the OCO-2 retrieved state vector~~B10 lite files~~ could improve the 3D cloud bias correction. As shown in Figure 3, the four cloud variables (H3D, HC, CSNoiseRatio,

Nearest Cloud Distance)~~(h3d, hc, cs_noise_ratio, cld_dist)~~ were removed during the recursive feature elimination step, indicating that other variables ~~from the lite files~~form the state vector 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 3~~1~~ (OCO-2 cloud corr.~~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~~dp_abp, co2_grad_del, ~~dp~~ h2o_ratio, and aod_water for land ~~and ocean~~ and dp_abp, co2_grad_del, albedo_wco2, and albedo_sco2 for ~~land~~sea) 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 X_{CO2} 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: the features ~~a)~~ determined by the recursive feature elimination~~xco2_strong_idp, co2_grad_del, dp, albedo_wco2 (for land only)~~ in addition to~~and~~ a) 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 (see Table 1 and 2)~~ (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 4~~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~~sea QF=1 data the best set of features is c) which reduces the RMSE from 1.~~18~~22 ppm to 1.~~13~~17 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 is~~was~~ shown in Figure 5 and is presented in the next section with an independent comparison to TCCON.

Table 42: 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 bias correctionrandom forest approach (B10-RFOCO-2 corr.) and using the same approach but with additional features. In addition to the variables determined by the recursive feature elimination xco2_strong_idp, co2_grad_del, dp a) contains nearest cloud distance, b) CSNoiseRatio, c) nearest cloud distance, CSNoiseRatio, HC, H3D, and d) deltaT.

	LanddND OCO-2 corr.B10-RF X_{CO2} [ppm]					SeaGL X_{CO2} [ppm] B10-RFOCO-2 corr.				
		a)	b)	c)	d)		a)	b)	c)	d)
QF=0	**0.810.75**	0.7581	0.7581	0.7580	0.7581	**0.460.44**	0.4445	0.4445	0.4344	0.4445
QF=1	**2.021.91**	1.902.01	1.902.01	1.962.01	1.922.02	**1.221.18**	1.1620	1.1620	1.1317	1.1720
QF=0 + 1	**1.551.43**	1.4354	1.431.54	1.4854	1.5544	**0.830.74**	0.7381	0.7380	0.7178	0.7381

4.4 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 Nnorthern hemisphereHemisphere. 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 over for land nadir and 1305 (QF=0: 942, QF=1: 363) matches over seafor ocean glint observations. Note that, our bias correction model was trained without taking TCCON observations into consideration while B10-OCO-2 takes OCO-2 – TCCON biases explicitly into consideration for its linear bias correction, filtering, and to calculate global offsets. Thus, comparisons between B10-OCO-2 and TCCON are not independent.

Table 53 shows the mean and standard deviation of differences between B10-OCO-2 and TCCON and after we apply our bias correction (OCO-2 corr. - B10-RF - TCCON) for QF=0 and QF=1. For Over land and sea the nadir the bias correction reduces the standard deviation between OCO-2 and TCCON for QF=0 and QF=1. biases mostly for QF=1 data while there is practically no difference for QF=0. For observations over sea ocean glint observations the bias corrected X_{CO2} exhibits a systematic positive offset compared to TCCON- of about 0.8 ppmbut a reduced variability. The systematic offset could be addressed by recalculating the scaling factor used for ocean glint retrievals over sea in B10OCO-2. However, there are only few TCCON stations that can provide comparisons for ocean glintthose data and these stations are not equally distributed over the ocean.

Table 5~~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~~ OCO-2 - TCCON) and the proposed random forest approach by (OCO-2 corr.~~RF~~ - TCCON).

	L~~ND~~and X_{CO2} [ppm]		Sea~~GL~~ X_{CO2} [ppm]	
	~~B10~~ OCO-2 - TCCON	~~B10 RF~~OCO-2 corr. - TCCON	OCO-2 - TCCON~~B10 - TCCON~~	OCO-2 corr. - TCCON~~B10 RF - TCCON~~
QF=0	-0.19 ± 1.11	**-0.26 ± 1.010**	0.87 ± 0.75	**0.~~79~~ 78 ± 0.~~70~~67**
QF=1	-0.~~42~~ 2 ± ~~2.27~~1.92	**0.~~19~~ 03 ± 1.8663**	0.~~08~~ 01 ± ~~2~~1.~~08~~81	**~~1.20~~0.87 ± 1.~~72~~44**
QF=0 + 1	-0.~~34~~ 31 ± 1.~~69~~59	**-0.~~18~~ 25 ± 1.~~24~~42**	0.65 ± 1.~~32~~19	**0.8~~1~~8 ± 0.~~99~~93**

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~~sea there exist negative biases in ~~B10~~ OCO-2 in the proximity of clouds (Figure 6a and 6c). Interestingly, there is a positive bias for OCO-2~~B10~~ ~~ocean glint~~sea data when no clouds are close to OCO-2 soundings (> 14 km) that likely stems from ~~B10~~ OCO-2 incorporating a multi model mean in its bias correction in addition to TCCON. After applying our bias correction, X_{CO2} biases over land show no dependence on nearest cloud distance anymore ~~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~~ 0.8 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~~OCO-2.

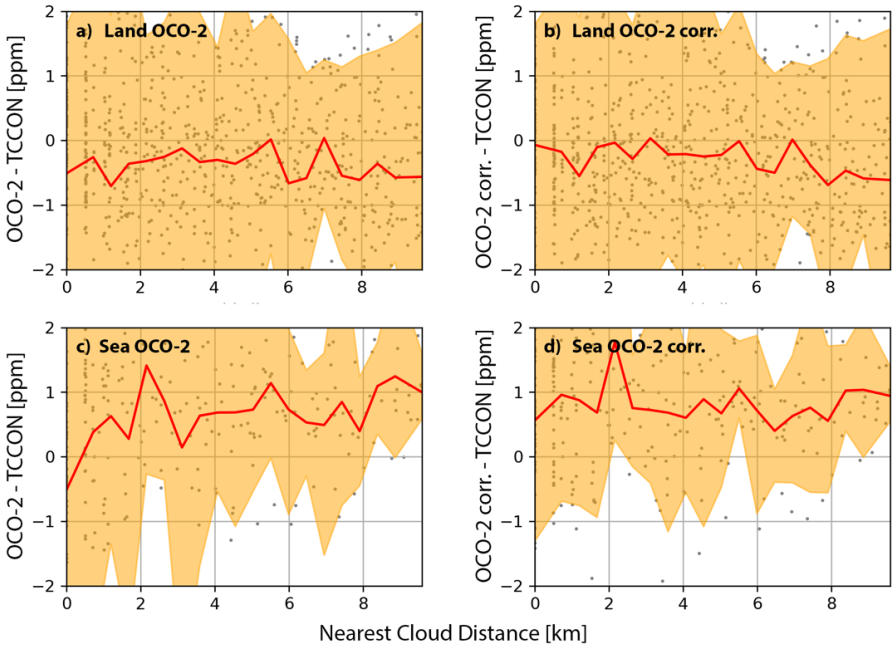

Figure 6: X_{CO2} bias vs cloud distance of OCO-2 ~~for~~ over land ~~nadir~~ ~~B10~~ (a), OCO-2 ~~B10~~ corrected (b), OCO-2 over sea~~ocean glint B10~~ (c), and OCO-2 corrected~~B10 corrected~~ (d), for QF=0 and QF=1 data from 2018 to 2019. The 5th and 95th percentiles are indicated with the yellow shaded ~~area,~~area, the mean is shown with a red line, and individual comparisons with grey dots.

5 Discussion

5.1 Model Interpretation

To better understand how the model utilizes the input features to calculate the bias correction we ~~calculated~~ show the modelled biases ~~(RF Bias)~~ with respect to the individual ~~three input~~ features ~~(xco2_strong_idp, co2_grad_del, dp)~~ in ~~(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 over the complete state space but linear over part of the state space. This explains the lower model performance of the linear model we compared to in Section 4.2. over the complete state space (QF=0+1) and the only marginal improvement compared to QF=0 data. Differences between retrieved surface pressure and surface pressure from a forecast model (~~xdp_abp~~)~~co2_strong_idp~~ show a positive correlation with X_{CO2} biases. When the operationally retrieved surface pressure is underestimated, X_{CO2} is underestimated as well. The ratio of the retrieved H_2O column from the WCO_2 band to that from the SCO_2 band (~~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.~~**h2o_ratio),** for soundings over land (Figure 7b), is independent to X_{CO2} biases for ratios of less than one and has a strong negative correlation for ratios above one. A ratio of 1.05 corresponds on average to an X_{CO2} bias of -1 ppm. The difference between the retrieved CO_2 profile and the a priori profile (~~-~~**co2_grad_del**) shows mostly a positive correlation for negative ~~co2_grad_del~~values (surface CO_2 is underestimated compared to CO_2 higher up in the atmosphere) and a negative correlation for positive ~~co2_grad_del~~values. This indicates that 3D cloud effects challenge the accurate retrieval of the X_{CO2} profile. The sensitivity of X_{CO2} biases to changes in co2_grad_del is approximately twice as strong over sea than land (see Figure 7c and 7f). This feature cannot be exclusively linked to 3D cloud effects since it is one of the most important features for the operational bias correction of OCO-2. The retrieved extinction optical depth of cloud water (~~**dp_**~~**aod_water)** shows a mostly negative linear correlation with a X_{CO2} bias of -1 ppm for an extinction optical depth of 0.1. ~~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, the surface albedo in the weak and strong CO_2 band (**albedo_wco2**~~,~~ **albedo_sco2)** have mostly no dependence to X_{CO2} biases for most of their range but show some positive and negative correlations with biases for brighter and darker surfaces, respectively. ~~shows no clear dependence on X_{CO2} biases.~~ Note~~,~~ that our bias correction is applied in addition to the bias correction that has already been performed in the operational OCO-2 retrieval. While the operational OCO-2 bias correction does not explicitly account for 3D cloud biases it might ~~B10 that~~ implicitly ~~remove~~mitigate such~~s some~~ biases ~~correlations of 3D cloud biases~~ with its linear bias correction (since the operational bias correction variable dP is correlated to nearest cloud distance, see the red line in Fig. 2b). ~~co2_grad_del and dPfrac (highly correlated with dp).~~

To understand why some variables of the OCO-2 retrieved state vector are correlated with 3D cloud biases it is important to remember that the operational retrieval, based on optimal estimation, tries to match the observed radiances with a forward radiative transfer model. However, while the observed radiances can be perturbed by 3D cloud effects, the forward model tries to match those radiances with an independent pixel approximation that does not physically include 3D cloud effects. In particular the 3D cloud effect enhances, or brightens, the radiances as compared to no clouds being present. To compensate for this brightening the forward model decreases the retrieved surface pressure (reduction in dp_abp), increases the optical depth of cloud water (aod_water) and increases the surface albedo in the WCO_2 band. These relationships are shown empirically in Figure 7. As shown in Fig. 2 of Massie et al. (2021), the spectral signature of the 3D cloud effect (the optical depth structure of the radiative perturbation of the 3D effect) differs from the spectral signatures of perturbations in surface pressure, surface reflectivity, aerosol, and X_{CO2}. Fig. 2 illustrates that a decrease in surface pressure and X_{CO2}, and an increase in surface reflectance will increase the observed radiance. In order to provide for extra radiance enhancement in the cloud brightened observed radiance, a variety of state variable adjustments (and their unique spectral contributions) are utilized by the retrieval to bring forward model radiances in agreement with the observed radiances. The relationship of 3D cloud biases to surface pressure differences and surface albedo are likely due to a combination of physically-based 3D cloud radiative effects and- operational retrieval algorithmic considerations.

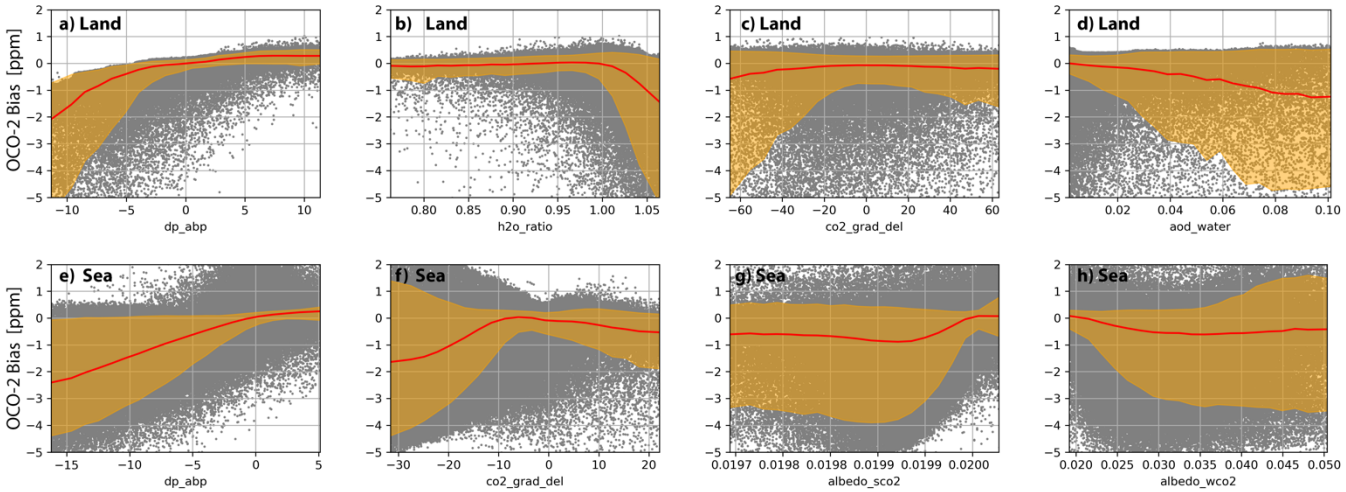

Figure 7: Bias identified by correction by the proposed model (~~RF Bias~~OCO-2 Bias) with respect to its ~~three input~~ features: ~~(~~dp_abp (a), h2o_ratio (b), ~~xco2_strong_idp,~~ co2_grad_del (c), aod_water (d~~d~~p) over land ~~for land nadir~~ (top) and dp_abp (e), co2_grad_del (f), albedo_sco2 (g), albedo_wco2 (h) over sea~~ocean glint~~ (bottom) ~~observations~~ for QF=0 and QF=1 data from 2018 to 2019. The 5th and 95th percentile ~~are~~ are indicated with the yellow shaded area, the mean is shown with a red line and individual comparisons with grey dots. ~~Note,~~ T~~t~~he scale of the x-axis for each plot is different. Please refer to Section 3.4 for a description of the individual features.

A further look at the relative importance of the model features shows dp_abp being the most important feature for land and sea observations. Over land dp_abp is followed by h2o_ratio, co2_grad_del, and aod_water. Over sea dp_abp is followed by co2_grad_del, albedo_sco2, albedo_wco2~~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~~dp_abp from our model as a feature the bias correction would be less effective than if we were to omit co2_grad_del.

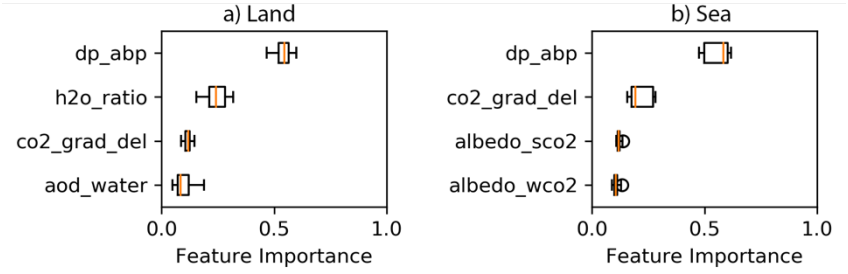

Figure 8: Feature importance for the bias correction model. Feature importance is shown for land ~~nadir~~ (left) and ~~ocean glint~~sea (right) observations. Model was trained using the training set with QF=0 and QF=1 data. Please refer to Section 3.4 for a description of the individual features.

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 averaged results over 2° by 2° cells (see Figure 9). I.e., int~~o~~ applying the proposed bias correction, ~~one would subtract~~ the results shown in Figure 9 are subtracted from ~~B10~~ OCO-2 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} from OCO-2 is underestimated ~~in B10~~) are present north of 45° in America, Europe and Asia, averaging -0.~~36~~23 ppm. ~~and~~ A~~a~~round the tropics within ±10° of the equator, average biases are near~~averaging~~ -0.~~43~~30 ppm. Positive biases are most dominant over the deserts of northern Africa and Saudi Arabia. Over ~~the ocean~~sea 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 (correlation coefficient between nearest cloud distance and OCO-2 bias is R=0.3). ~~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.~~

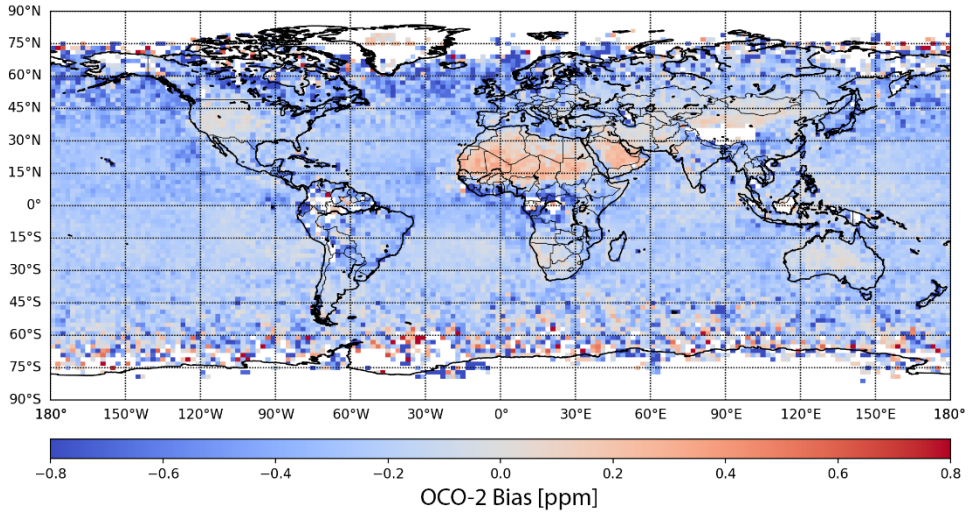

Figure 9: Biases in X_{CO2} identified by our model. Biases are averaged over 2° by 2° for all soundings (2014 to 2019, QF=0 and QF=1). Negative biases are shown in blue, positive biases in red, no data in white.

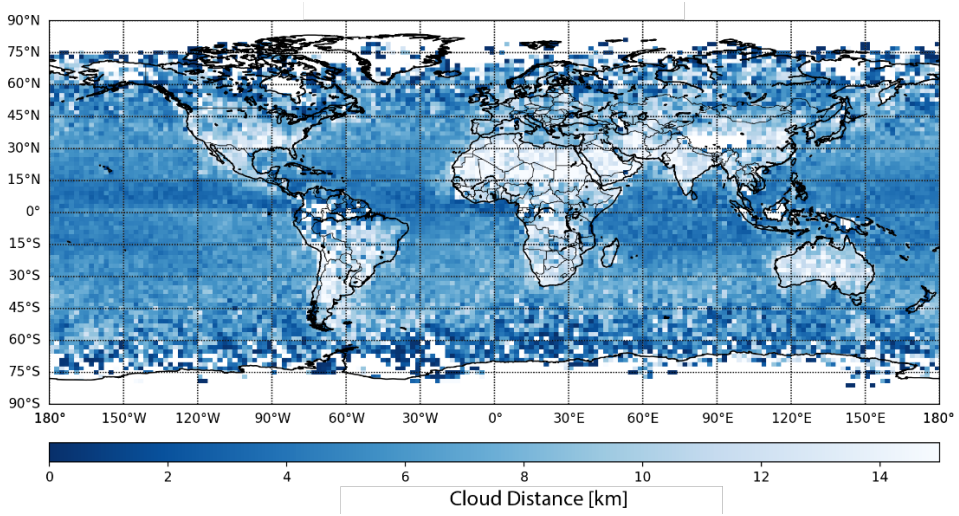

Figure 10: Nearest cloud distance derived from MODIS. Nearest cloud distances are averaged over 2° by 2° for all matched soundings (2014 to 2019, QF=0 and QF=1). Darker blues indicate closer clouds, no data is shown in white.

485

5.3 Effect of bias correction on true CO_2 Enhancements

As discussed in section 3.1 we use the small areas analysis as a truth proxy to develop our model. This assumes that CO_2 is well mixed and constant over short spatial scales (<100 km). However, this assumption is violated for strong CO_2 emitters such as power plants. Even though these strong emitters are rare in the data and likely don't influence the bias correction
490 model, there is a risk that the model would "correct", i.e. remove real local CO_2 enhancements. To confirm that real CO_2 enhancements are still present after the proposed bias correction, we compare OCO-2 retrieved and corrected X_{CO2} from three OCO-2 overpasses over large coal power plants (see Figure 11), that have been used in a previous study (Nassar et al., 2017). The CO_2 enhancements of the retrieved and corrected X_{CO2} for the three overpasses (the singular spikes in X_{CO2} in the middle of the graphs) agrees closely and demonstrates that the *bias correction does not erroneously remove true CO_2 enhancements*
495 *from the OCO-2 data record*.

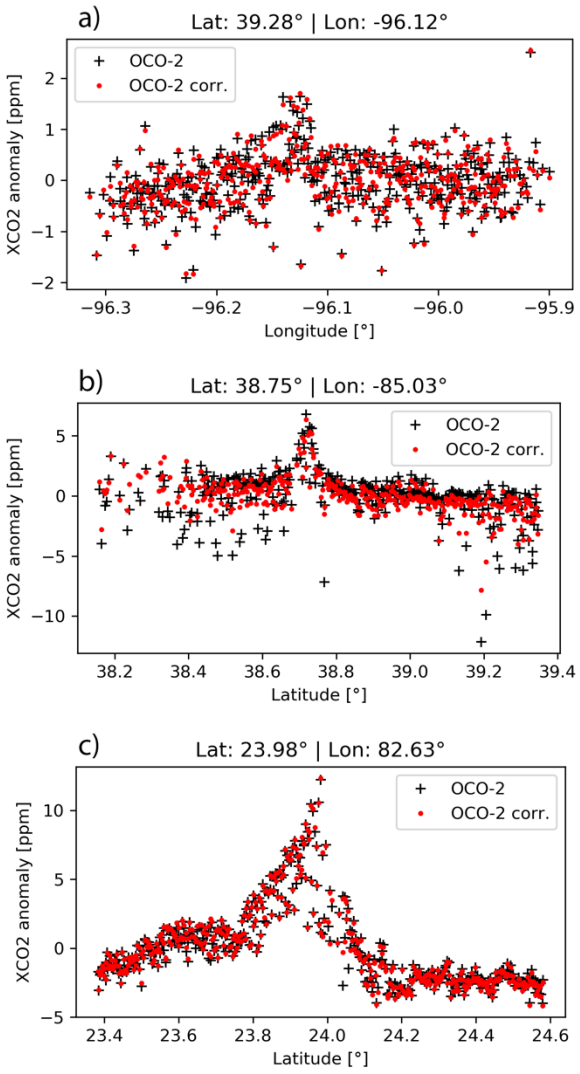

Figure 11: X_{CO2} anomalies for OCO-2 and bias corrected OCO-2 retrievals in the proximity of coal power plants. Power plant a) Westar at Lat: 39.28° Lon: -96.12° on 12/04/2015, b) Ghent at Lat: 38.75° Lon: -85.03° on 08/13/2015, c) Sasan at Lat: 23.98° Lon: -82.63° on 10/23/2014. Anomaly is calculated by subtracting the mean.

500

5 Future Work and Conclusion

5.1 Future Work

The developed bias correction approach is~~was~~ aimed at mitigating 3D cloud biases in ~~B10~~OCO-2, 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 previously uncorrected ~~raw~~ X_{CO2}~~.~~ (the "raw" XCO2 from the operational retrieval). 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 operational bias correction used for ~~B10~~ OCO-2 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 in~~on~~ QF=0 data where ~~moderate~~small. 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~~ large 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~~ sea ~~glint~~ data, to having multiple models for land and ~~ocean~~ sea to better capture the diverse causes for biases in X_{CO2} 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 (Mendonca et al., 2021; Jacobs et al., 2020) 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 from the state vector for OCO-2 retrievals over land (dp_abp, h2o_ratio, co2_grad_del, aod_water) and sea (dp_abp, co2_grad_del, albedo_sco2, albedo_wco2) ~~xco2_strong_idp, co2_grad_del, dp, albedo_wco2)~~ that are used in a machine learning model that ~~was~~is able to remove~~allow to correct~~ ~~for~~ 3D cloud biases in ~~B10~~ operational bias-corrected OCO-2 retrieved X_{CO2} ~~from OCO-2~~. We demonstrate that this machine learning model does not erroneously remove true CO_2 power plant enhancements from the OCO-2 data record. 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~~1.~~08~~ 73 ppm to 1.~~43~~ 55 ppm over land and 1.~~35~~ 19 ppm to 0.~~74~~ 83 ppm over ~~the ocean~~sea for QF=0 and QF=1 data on an independent testing set. We demonstrated a systematic approach to correct for biases in optimal estimation~~OE~~ retrievals. Namely, (1) find a physical variable that is well understood and correlated with the cause of the bias (in our case 'nearest 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.

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 et al., 2014a; Dubey et al., 2014b; Iraci et al., 2014; Iraci et al., 2016; Toon and Wunch, 2014; Wennberg et al., 2016a; Wennberg et al., 2014c; Wennberg et al., 2014a; Wennberg et al., 2016b; Wennberg et al., 2014b; 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.: Random forests, Machine learning, 45, 5-32, 2001.

Chen, S., Natraj, V., Zeng, Z.-C., and Yung, Y. L.: Machine learning-based aerosol characterization using OCO-2 O2 A-band observations, Journal of Quantitative Spectroscopy and Radiative Transfer, 279, 108049, 2022.

Crisp, D., Atlas, R. M., Breon, F. M., Brown, L. R., Burrows, J. P., Ciais, P., Connor, B. J., Doney, S. C., Fung, I. Y., Jacob, D. J., Miller, C. E., O'Brien, D., Pawson, S., Randerson, J. T., Rayner, P., Salawitch, R. J., Sander, S. P., Sen, B., Stephens, G. L., Tans, P. P., Toon, G. C., Wennberg, P. O., Wofsy, S. C., Yung, Y. L., Kuang, Z., Chudasama, B., Sprague, G., Weiss, B., Pollock, R., Kenyon, D., and Schroll, S.: The Orbiting Carbon Observatory (OCO) mission, Advances in Space Research, 34, 700-709, https://doi.org/10.1016/j.asr.2003.08.062, 2004.

Cronk, H.: OCO-2/MODIS Collocation Products User Guide, Version 3, June 2018 (3) [dataset], 2018.

Dubey, M., Henderson, B., Green, D., Butterfield, Z., Keppel-Aleks, G., Allen, N., Blavier, J. F., Roehl, C., Wunch, D., and Lindenmaier, R.: TCCON data from Manaus (BR), Release GGG2014R0, 10.14291/tccon.ggg2014.manaus01.R0/1149274, 2014a.

Dubey, M., Lindenmaier, R., Henderson, B., Green, D., Allen, N., Roehl, C., Blavier, J. F., Butterfield, Z., Love, S., Hamelmann, J., and Wunch, D.: TCCON data from Four Corners (US), Release GGG2014R0, 10.14291/tccon.ggg2014.fourcorners01.R0/1149272, 2014b.

Eldering, A., Taylor, T. E., O'Dell, C. W., and Pavlick, R.: The OCO-3 mission: measurement objectives and expected performance based on 1 year of simulated data, Atmos. Meas. Tech., 12, 2341-2370, https://doi.org/10.5194/amt-12-2341-2019, 2019.

Eldering, A., O'Dell, C. W., Wennberg, P. O., Crisp, D., Gunson, M. R., Viatte, C., Avis, C., Braverman, A., Castano, R., Chang, A., Chapsky, L., Cheng, C., Connor, B., Dang, L., Doran, G., Fisher, B., Frankenberg, C., Fu, D., Granat, R., Hobbs, J., Lee, R. A. M., Mandrake, L., McDuffie, J., Miller, C. E., Myers, V., Natraj, V., O'Brien, D., Osterman, G. B., Oyafuso, F., Payne, V. H., Pollock, H. R., Polonsky, I., Roehl, C. M., Rosenberg, R., Schwandner, F., Smyth, M., Tang, V., Taylor, T. E., To, C., Wunch, D., and Yoshimizu, J.: The Orbiting Carbon Observatory-2: first 18 months of science data products, Atmos. Meas. Tech., 10, 549-563, https://doi.org/10.5194/amt-10-549-2017, 2017.

Emde, C., Yu, H., Kylling, A., Van Roozendael, M., Stebel, K., Veihelmann, B., and Mayer, B.: Impact of 3D cloud structures on the atmospheric trace gas products from UV–Vis sounders–Part 1: Synthetic dataset for validation of trace gas retrieval algorithms, Atmospheric Measurement Techniques, 15, 1587-1608, 2022.

Evans, K. F.: The spherical harmonics discrete ordinate method for three-dimensional atmospheric radiative transfer, Journal of the Atmospheric Sciences, 55, 429-446, https://doi.org/10.1175/1520-0469(1998)055<0429:TSHDOM>2.0.CO;2, 1998.

Hoerl, A. E. and Kennard, R. W.: Ridge regression: applications to nonorthogonal problems, Technometrics, 12, 69-82, 1970a.

Hoerl, A. E. and Kennard, R. W.: Ridge regression: Biased estimation for nonorthogonal problems, Technometrics, 12, 55-67, 1970b.

Iraci, L., Podolske, J., Hillyard, P., Roehl, C., Wennberg, P. O., Blavier, J. F., Landeros, J., Allen, N., Wunch, D., Zavaleta, J., Quigley, E., Osterman, G. B., Barrow, E., and Barney, J.: TCCON data from Indianapolis (US), Release GGG2014R0, 10.14291/tccon.ggg2014.indianapolis01.R0/1149164, 2014.

Iraci, L. T., Podolske, J., Hillyard, P. W., Roehl, C., Wennberg, P. O., Blavier, J. F., Allen, N., Wunch, D., Osterman, G. B., and Albertson, R.: TCCON data from Edwards (US), Release GGG2014R1, 10.14291/tccon.ggg2014.edwards01.R1/1255068, 2016.

Jacobs, N., Simpson, W. R., Wunch, D., O'Dell, C. W., Osterman, G. B., Hase, F., Blumenstock, T., Tu, Q., Frey, M., Dubey, M. K., Parker, H. A., Kivi, R., and Heikkinen, P.: 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, 5033-5063, 10.5194/amt-13-5033-2020, 2020.

Jet Propulsion Laboratory: Orbiting Carbon Observatory–2 (OCO-2) Data Product User's Guide, Operational L1 and L2 Data Versions 8 and Lite File Version 9, 2018.

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

Kylling, A., Emde, C., Yu, H., van Roozendael, M., Stebel, K., Veihelmann, B., and Mayer, B.: Impact of 3D cloud structures on the atmospheric trace gas products from UV–Vis sounders–Part 3: Bias estimate using synthetic and observational data, Atmospheric Measurement Techniques, 15, 3481-3495, 2022.

Liang, L., Di Girolamo, L., and Platnick, S.: 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, https://doi.org/10.1029/2008GL037124, 2009.

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

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

Mendonca, J., Nassar, R., O'Dell, C. W., Kivi, R., Morino, I., Notholt, J., Petri, C., Strong, K., and Wunch, D.: 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, 7511-7524, 10.5194/amt-14-7511-2021, 2021.

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

Nassar, R., Hill, T. G., McLinden, C. A., Wunch, D., Jones, D. B. A., and Crisp, D.: Quantifying CO2 Emissions From Individual Power Plants From Space, Geophysical Research Letters, 44, 10,045-010,053, https://doi.org/10.1002/2017GL074702, 2017.

O'Dell, C., Eldering, A., Wennberg, P. O., Crisp, D., Gunson, M. R., Fisher, B., Frankenberg, C., Kiel, M., Lindqvist, H., Mandrake, L., Merrelli, A., Natraj, V., Nelson, R. R., Osterman, G. B., Payne, V. H., Taylor, T. E., Wunch, D., Drouin, B. J., Oyafuso, F., Chang, A., McDuffie, J., Smyth, M., Baker, D. F., Basu, S., Chevallier, F., Crowell, S. M. R., Feng, L., Palmer, P. I., Dubey, M., García, O. E., Griffith, D. W. T., Hase, F., Iraci, L. T., Kivi, R., Morino, I., Notholt, J., Ohyama, H., Petri, C., Roehl, C. M., Sha, M. K., Strong, K., Sussmann, R., Te, Y., Uchino, O., and Velazco, V. A.: Improved retrievals of carbon dioxide from Orbiting Carbon Observatory-2 with the version 8 ACOS algorithm, Atmos. Meas. Tech., 11, 6539-6576, https://doi.org/10.5194/amt-11-6539-2018, 2018.

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

Payne, V. H., Drouin, B. J., Oyafuso, F., Kuai, L., Fisher, B. M., Sung, K., Nemchick, D., Crawford, T. J., Smyth, M., and Crisp, D.: Absorption coefficient (ABSCO) tables for the Orbiting Carbon Observatories: version 5.1, Journal of Quantitative Spectroscopy and Radiative Transfer, 255, 107217, 2020.

Rodgers, C. D.: Inverse methods for atmospheric sounding: theory and practice, World scientific2000.

Taylor, T. E., O'Dell, C. W., Frankenberg, C., Partain, P. T., Cronk, H. Q., Savtchenko, A., Nelson, R. R., Rosenthal, E. J., Chang, A. Y., and Fisher, B.: Orbiting Carbon Observatory-2 (OCO-2) cloud screening algorithms: validation against collocated MODIS and CALIOP data, Atmospheric Measurement Techniques, 9, 973-989, https://doi.org/10.5194/amt-9-973-2016, 2016.

Toon, G. C. and Wunch, D.: A stand-alone a priori profile generation tool for GGG2014, 10.14291/tccon.ggg2014.priors.R0/1221661, 2014.

Wennberg, P. O., Roehl, C., Blavier, J. F., Wunch, D., Landeros, J., and Allen, N.: TCCON data from Jet Propulsion Laboratory (US), 2011, Release GGG2014R1, 10.14291/tccon.ggg2014.jpl02.R1/1330096, 2016a.

Wennberg, P. O., Wunch, D., Roehl, C., Blavier, J. F., Toon, G. C., and Allen, N.: TCCON data from Caltech (US), Release GGG2014R1, 10.14291/tccon.ggg2014.pasadena01.R1/1182415, 2014a.

Wennberg, P. O., Wunch, D., Yavin, Y., Toon, G. C., Blavier, J. F., Allen, N., and Keppel-Aleks, G.: TCCON data from Jet Propulsion Laboratory (US), 2007, Release GGG2014R0, 10.14291/tccon.ggg2014.jpl01.R0/1149163, 2014b.

Wennberg, P. O., Roehl, C., Wunch, D., Toon, G. C., Blavier, J. F., Washenfelder, R. a., Keppel-Aleks, G., Allen, N., and Ayers, J.: TCCON data from Park Falls (US), Release GGG2014R0, 10.14291/tccon.ggg2014.parkfalls01.R0/1149161, 2014c.

Wennberg, P. O., Wunch, D., Roehl, C., Blavier, J. F., Toon, G. C., Allen, N., Dowell, P., Teske, K., Martin, C., and Martin, J.: TCCON data from Lamont (US), Release GGG2014R1, 10.14291/tccon.ggg2014.lamont01.R1/1255070, 2016b.

Wunch, D., Toon, G. C., Sherlock, V., Deutscher, N. M., Liu, C., Feist, D. G., and Wennberg, P. O.: The Total Carbon Column Observing Network's GGG2014 Data Version, Pasadena, California, 43-43, 10.14291/tccon.ggg2014.documentation.R0/1221662, 2015.

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

Wunch, D., Mendonca, J., Colebatch, O., Allen, N., Blavier, J.-F. L., Roche, S., Hedelius, J. K., Neufeld, G., Springett, S., Worthy, D. E. J., Kessler, R., and Strong, K.: TCCON data from East Trout Lake (CA), Release GGG2014R1, 10.14291/tccon.ggg2014.easttroutlake01.R1, 2017.

Yu, H., Emde, C., Kylling, A., Veihelmann, B., Mayer, B., Stebel, K., and Van Roozendael, M.: Impact of 3D Cloud Structures on the Atmospheric Trace Gas Products from UV-VIS Sounders–Part II: impact on NO 2 retrieval and mitigation strategies, Atmospheric Measurement Techniques Discussions, 1-34, 2021.