A non-linear data driven approach to bias correction of XCO2 for OCO-2 NASA ACOS version 10

William Keely et al.

The points from the reviewer are shown in black, with our response in blue, and changes or additions to the manuscript in red.

Response to Reviewer #1

Specific Comments:

The authors correctly state that features measuring the deviation of retrieved quantities from the prior (surface pressure and vertical CO2 profile) have already been used in the operational bias correction. However, with a much more complex non-linear machine learning method, it actually needs to be checked again that there is no attenuation of actual CO2 signals when using these (or other) features in the bias correction. In section 5.3, two examples are used to demonstrate that the machine learning corrected product captures enhancements not present in the training data. This is a good thing. However, one objective of the non-linear correction was to largely reproduce the linear model for QF=0 (which is the currently accepted community standard) and Figure 6 additionally shows that the results associated to XGBoost_QFNew and B10_QF are very similar in terms of sounding throughput for both analysed regions (South Korea and Ohio, US). It is therefore reasonable to assume that most of the soundings in Figure 10 are of type QF=0 and that these examples are therefore unlikely to reflect the main innovations of QFNew quality filtering. The probably more critical part of the proposed non-linear bias correction, namely the increase of sounding throughput beyond QF=0 and the behaviour for related soundings with QF=1 in terms of potential over-correction, is thus (most likely) not explicitly investigated. This would be a worthwhile addition. Furthermore, Figure 9 shows that the specific features measuring the deviation of retrieved quantities from the prior are also important for QF=1. Thus, the following questions arise, whose answers would significantly further increase the conclusiveness of Section 5.3:

Are most of the soundings shown in Figure 10 actually of type QF=0?
How many soundings (absolute and relative) in Figure 10 are of type QF=1?
Can you introduce a graphical distinction of soundings with QF=0 and QF=1 in Figure 10?
Is it possible to find a plume, which is not present in the training data and largely consists of QF=1 data?

This is an excellent point, and we agree with the need for further assessment in Section 5.3. While the US plume is indeed primarily QF=0 data, the South Korea plume is primarily QF=1 (~65%) and we believe is suitable for the analysis. QFNew passes ~25% of QF=1 data for Taean, S. Korea much of which is within the plume feature itself. We have added an additional plot to Figure 10:

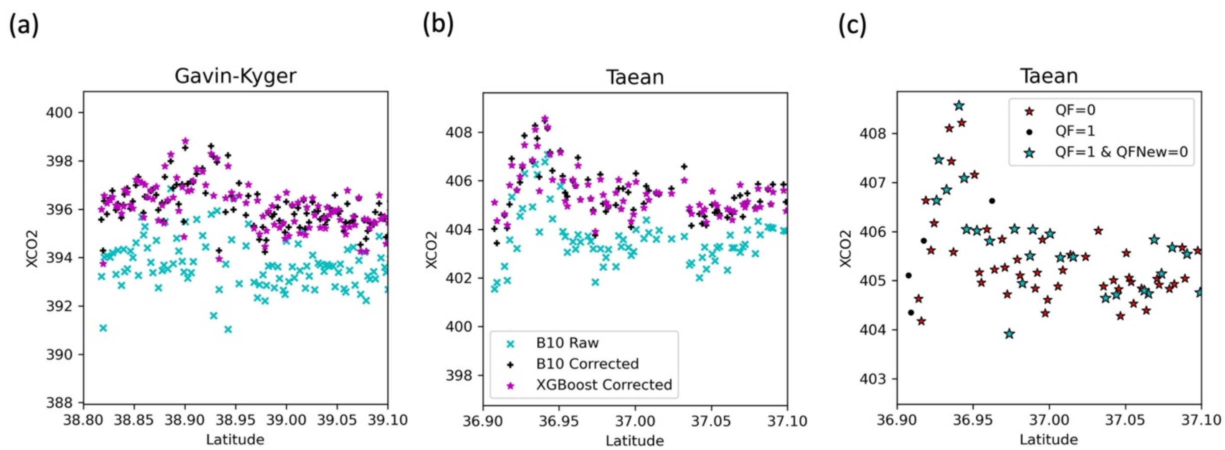

Figure 10. Two CO_2 plumes captured downwind from power plants (Nassar et al. 2021). An ocean glint and land nadir plume at Taean, South Korea, [lat 36.91°, lon 126.23°] on 2015-04-17 is shown in (a). A land nadir plume near the J. M. Gavin and Kyger Creek power plants in Ohio, USA, [lat 38.93°, lon -82.12°] on 2015-07-30. Regions with the example plumes are not present in the training dataset and consist of QF = 0 + 1 data. Plot (c) shows the increase in XGBoost corrected data for QFNew=0 that would be filtered by the B10 QF.

Please elaborate in the manuscript on the difference QFNew\B10_QF of the two sets QFNew and B10_QF (set of elements of QFNew not in B10_QF) in the context of potentially correcting out actual CO2 signals.

We have updated both 5.3 and 5.4 in the manuscript.

5.3 Preservation of CO_2 enhancements

We assess the risk of the proposed bias correction to correct out and remove plume features in the data. Several features heavily utilized by the XGBoost models and in operational correction such as the CO_2 gradient delta, and surface pressure terms (e.g., dpfrac, dp_o2a), are differences between the ACOS retrieved state, and the prior. Therefore, there is potentially a risk for the bias correction to use the delta terms to over correct the retrieved XCO_2 to the truth. We compare XGBoost corrected XCO_2 for two known plumes first identified in Nassar et al. 2021. The two example plumes are shown in Figure 10 (a) and (b): an ocean glint and land nadir plume in Taean, South Korea, and a land nadir plume observed over two co-located power plants in Ohio, US. We compare the uncorrected XCO_2 retrieval (B10 Raw), the operationally corrected XCO_2 (B10 Corrected) and the machine learning corrected XCO_2 (XGBoost Corrected) and note that the machine learning corrected product captures enhancements not present in the training data. These results are also consistent with the findings in Mauceri et al. 2023 which include similar delta terms. This is further illustrated with the Taean plume which consists of ~35% QF = 0 soundings and ~65 QF = 1 soundings. QFNew = 0 improves the passing rate to ~ 60% as shown in Figure 10 (c). The red stars show data that is passed by QF = 0 (and by construction QFNew = 0) and the blue stars show data that would be removed by QF = 1 but is passed by QFNew = 0, indicating where the increase of available data for the plume feature. Of particular interest is the increase of data within the feature around 36.95° which includes maximum observed enhancement value.

5.4 Potential for further improving data throughput

Figure 11 further illustrates how the shape of the filtering or decision surface can affect data throughput. Soundings are binned by two state vector features: h2o_ratio and dpfrac. Figure 11b, and Figure 11d show the improvement in reduction of mean ΔXCO_2 and in the error divided by the posterior uncertainty, from the non-linear correction. The QF filters for each feature are indicated by the black dashed lines and the interior of the intersection of these filters indicates the region of state space that is labelled as QF = 0 (Note: the additional filters of the QF further reduce the data that is passed in this region). Significant portions of the distribution, where the non-linear method can accurately correct, lay outside of this filtered region and are labelled QF = 1. A data driven filter can be constructed using similar interpretable machine learning techniques and produce a unified correction/filtering product. Furthermore, moving away from the binary quality flag to a ternary ("very good", "good", "bad") will likely provide an improved data product for end users. Data driven methods for quality filtering have already proven to be useful in the northern high latitudes (Mendonca et al. 2021) and a genetic algorithm was previously used to derive the Warn Levels which complement the operational quality flag found in early OCO-2 data versions (Mandrake et al. 2015). An important task for such future work will be to ensure that the machine learning method learns a physically consistent filter that can increase data throughput while still limiting variance of error and ΔXCO_2. We also acknowledge that while the Taean plume shown in Figure 10 illustrates an empirical example of the ability of a non-linear correction to improve throughput of good quality data, further evaluation of the intersection (QF = 1 & QFNew = 0) will be required before bringing such a method to operation.

Technical Corrections:

L28: Bovensmann
Corrected.
L43: under-constrained
Corrected
L65: provides
Corrected
L128-131: Please make it two sentences
Corrected
L191: train a set
Corrected
L261: There was already a Section 4.1 before. Please correct the section numbering
Corrected
L383: becomes
Corrected

Response to Reviewer #2

Minor

Figure 5/Table 7: It would be ideal to use only the 2018 evaluation data here. I understand the point the authors make about needing more data to plot (and thus switching to the average of three models for three years), but this is inconsistent with the rest of the paper. This is in theory fine for Figure 5 because it is explained, but it is not okay for Table 7 where the claim is made that the RMSE of the B10 correction is maintained. This is because Table 7 needs to be compared to Table 4 to see the RMSE has been maintained and Table 4 only evaluates on 2018 data. I suggest the authors use just 2018 data to be consistent with the rest of the paper.

Figure 5, Figure 6, and Table 7 now only show results from 2018.

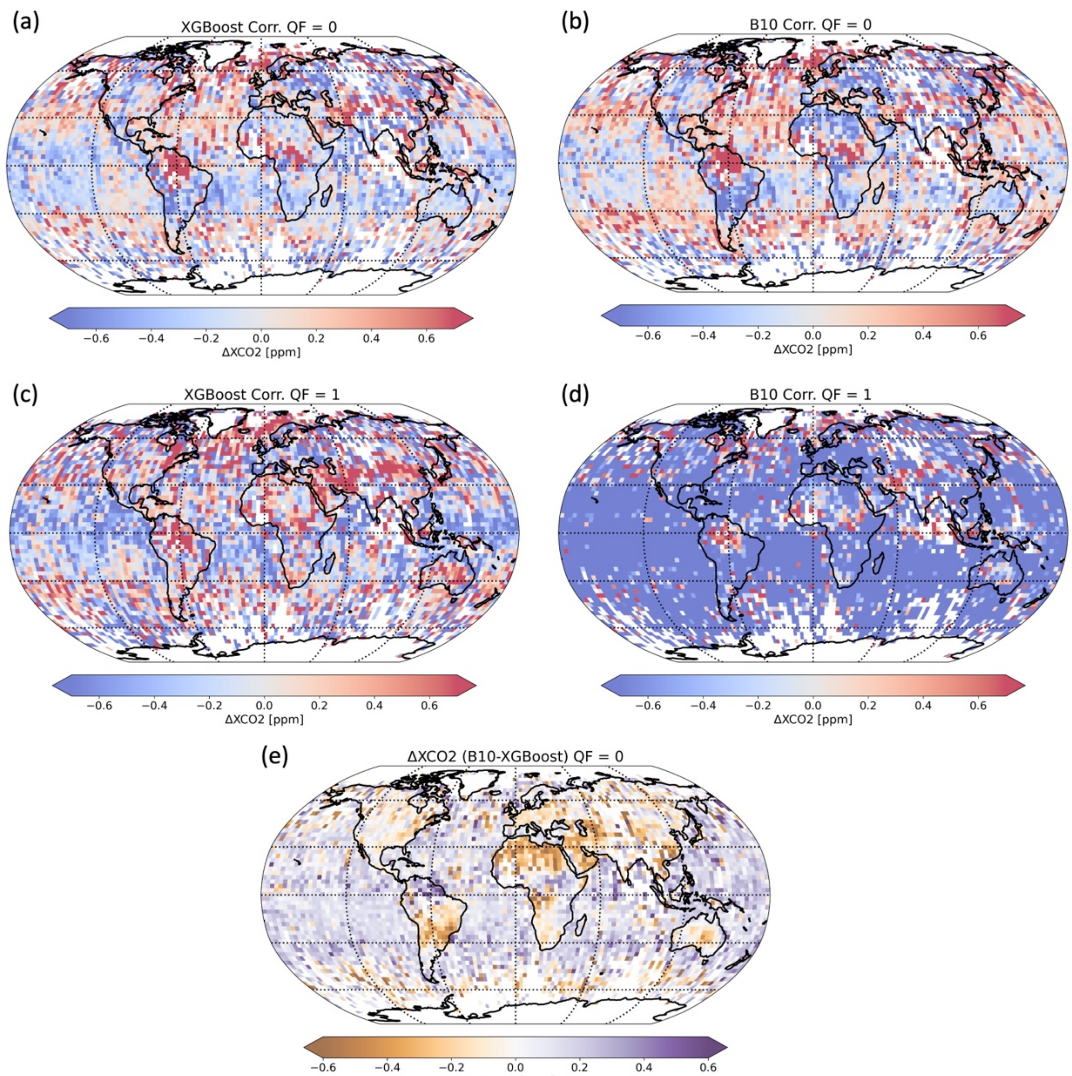

Figure 5. Remaining XCO_2 biases (ΔXCO_2) after correction for 2018 and model mean proxy, binned to a 3ºx3º resolution. ΔXCO_2 after the XGBoost correction for QF=0 is shown in (a), ΔXCO_2 after the B10 correction for QF=0 is shown in (b), ΔXCO_2

after the XGBoost correction for QF=1 is shown in (c), ΔXCO_2 after the B10 correction for QF=1 is shown in (d), and difference (B10 – XGB) for QF=0 is shown in (e).

Table 7. RMSE for combined XGBoost correction, B10 QF percent data throughput, and QFNew percent data throughput, by surface/mode, for 2018.

Surface (Mode)	XGBoost RMSE	B10 % Passing	QFNew % Passing
Land (Nadir+Glint)	1.07 ppm	59%	69%
Ocean (Glint)	0.72 ppm	60%	74%

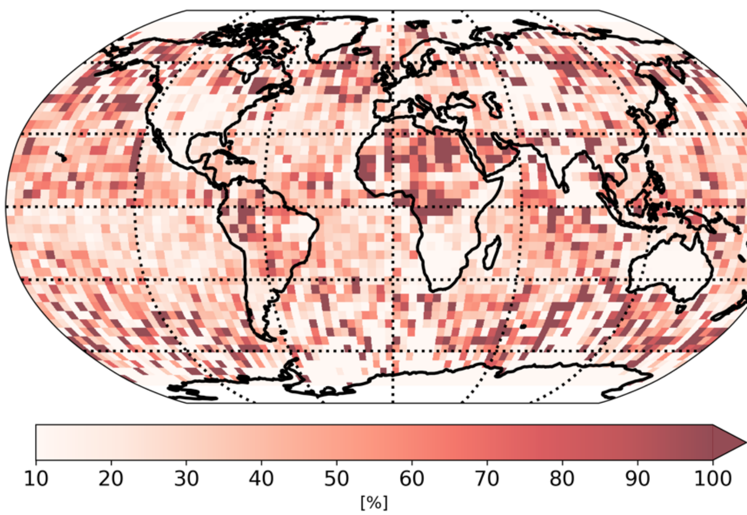

Figure 6. Relative increase in percent passing QFNew over B10 QF for 2018 aggregated by 4º×4º bins.

Technical

Line 31: Orbital -> Orbiting
Line 44: remove ")"
Line 63: rephrase sentence starting "A drawback of applying…"
Line 65: or is -> and are (?)
Line 86: 2022 -> 2023
Line 96: kernel -> averaging kernel
Line 106: is still -> it is still
Line 131: co-located TCCON -> co-located with TCCON
Line 191: run on sentence. Maybe it is meant to split between "nodes" and "when" (?)
Line 272: It is mentioned earlier that data is available until February 2019. Is January 2019 to February 2019 ignored? This should be specified.
Line 302: train as -> train a
Line 314: correction to -> correction
Line 352: for operational -> for the operational

Line 354: missing words before “and prior” (?)
Figure 2: make x-axis labels consistent with Table 3
Line 560: repeated section number (not corrected as noted in response to reviewers)
Lines 565-566: recalculate numbers in bold
Section 4.2 -> clarify that you are using the XGBoost (trained on QF = 0 + 1)
Line 602: 1.37 -> 1.38
Line 826: order of y vs. x is flipped
Line 951: add “et al.”
Line 1024: appears to be an incomplete sentence (maybe a former figure caption?)
Line 1027: sentence that begins “likely” is incomplete
Line 1045/1054: QF -> QFNew
Line 1046: spelling of parentheses
Line 1053: diamonds -> pluses
Table C1: define abbreviations in third column; co2_grad_del description is not consistent with Table 3 (“large unphysical”), define what the asterisk means.

Thank you for thoroughly reviewing and catching these errors. We have corrected for these and additional mistakes throughout the manuscript.