Atmos. Meas. Tech., 10, 2759–2771, 2017
https://doi.org/10.5194/amt-10-2759-2017

Atmospheric Measurement Techniques
Open Access
EGU

Evaluation and attribution of OCO-2 XCO_2 uncertainties

John R. Worden[1], Gary Doran[1], Susan Kulawik[2], Annmarie Eldering[1], David Crisp[1], Christian Frankenberg[3,1], Chris O'Dell[4], and Kevin Bowman[1]

[1]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA
[2]Bay Area Environmental Research Institute, Petaluma, CA, USA
[3]Geological and Planetary Sciences, California Institute for Technology, Pasadena, CA, USA
[4]Colorado State University, Fort Collins, CO, USA

Correspondence to: John R. Worden (john.worden@jpl.nasa.gov)

Received: 23 May 2016 – Discussion started: 21 July 2016
Revised: 23 June 2017 – Accepted: 29 June 2017 – Published: 31 July 2017

Abstract. Evaluating and attributing uncertainties in total column atmospheric CO_2 measurements (XCO_2) from the OCO-2 instrument is critical for testing hypotheses related to the underlying processes controlling XCO_2 and for developing quality flags needed to choose those measurements that are usable for carbon cycle science.

Here we test the reported uncertainties of version 7 OCO-2 XCO_2 measurements by examining variations of the XCO_2 measurements and their calculated uncertainties within small regions ($\sim 100\,km \times 10.5\,km$) in which natural CO_2 variability is expected to be small relative to variations imparted by noise or interferences. Over 39 000 of these "small neighborhoods" comprised of approximately 190 observations per neighborhood are used for this analysis. We find that a typical ocean measurement has a precision and accuracy of 0.35 and 0.24 ppm respectively for calculated precisions larger than ~ 0.25 ppm. These values are approximately consistent with the calculated errors of 0.33 and 0.14 ppm for the noise and interference error, assuming that the accuracy is bounded by the calculated interference error. The actual precision for ocean data becomes worse as the signal-to-noise increases or the calculated precision decreases below 0.25 ppm for reasons that are not well understood. A typical land measurement, both nadir and glint, is found to have a precision and accuracy of approximately 0.75 and 0.65 ppm respectively as compared to the calculated precision and accuracy of approximately 0.36 and 0.2 ppm. The differences in accuracy between ocean and land suggests that the accuracy of XCO_2 data is likely related to interferences such as aerosols or surface albedo as they vary less over ocean than land. The accuracy as derived here is also likely a lower bound as it does not account for possible systematic biases between the regions used in this analysis.

1 Introduction

Variations of total column CO_2 (XCO_2) resulting from photosynthesis and respiration in tropical forests (e.g., Parazoo et al., 2013), urban emissions (e.g., Kort et al., 2012) or tropical fires range from 2–5 ppm. Consequently, in order to use space-based measurements of XCO_2 to infer fluxes or properties of the processes controlling these variations, uncertainties in XCO_2 should ideally be much smaller than this variability (Miller et al., 2007). The Orbiting Carbon Observatory-2 (OCO-2) was launched in July 2014, to measure the atmospheric column averaged carbon dioxide (CO_2) dry air mole fraction, XCO_2 with the precision, accuracy, and coverage needed to quantify variations on regional scales at monthly intervals. These measurements are being used to investigate the underlying carbon cycle processes controlling atmospheric CO_2. The radiative transfer and XCO_2 estimation (or retrieval) algorithms (Boesch et al., 2006, 2011; Connor et al., 2008; O'Dell et al., 2012) were developed and tested using observed radiances from the Japanese TANSO GOSAT instrument (Kuze et al., 2009; Yoshida et al., 2011), which measured similar spectral regions as the OCO-2 mission. As discussed in Wunch et al. (2011), Crisp et al. (2012), and Mandrake et al. (2013), these algorithms also allowed extensive evaluation of quality flags and metrics needed to reject estimated XCO_2 values that were outside the expected range for XCO_2, likely because of poorly estimated values

for aerosols, clouds, surface albedo or surface pressure. In this paper we evaluate the calculated uncertainties due to noise and interferences in the OCO-2 data product (version 7). Our approach follows the methodology described in Boxe et al. (2010) and Kuai et al. (2013) in which variations of the observed trace gas over a small "area" are compared to the calculated errors.

2 Overview of OCO-2 data

The OCO-2 instrument measures radiances in the molecular oxygen (O_2) A-band (0.765 µm), the "weak" CO_2 band at 1.61 µm, and the "strong" CO_2 band at 2.06 µm. The OCO-2 instrument is an imaging spectrometer that collects eight samples, or "spatial footprints" across a narrow (0.8°) swath track. Observations are taken in three different modes: (1) "Nadir", where the spacecraft points the instrument's aperture at the ground directly downward along the orbit track; (2) "Glint", where the spacecraft points instrument's aperture near the "glint spot" where sunlight is specularly reflected by the surface, near the specular reflection point for sunlight; and (3) "Target", where the spacecraft points the instrument aperture at a stationary surface target, such as a validation site or city.

Nadir observations usually return useful measurements only over land. Glint observations return useful data over both land and ocean. Here, we discriminate land-glint and ocean-glint observations because they have different error statistics. We do not evaluate Target data in this analysis due to spurious statistics that are observed with the Target data.

As discussed in Boesch et al. (2006), Connor et al. (2008), and O'Dell et al. (2012 and references therein), total column estimates of XCO_2, are derived from OCO-2 observed radiances using a Bayesian optimal estimation approach that depends on CO_2, all the geophysical parameters or interferences that affect the radiances in these bands, and a priori statistics of the atmosphere and these interferences.

We use version 7 of the OCO-2 data, the first OCO-2 product distributed for general users. These data, like those described for GOSAT data in Wunch et al. (2011), are bias-corrected using a fit to retrieved aerosol optical depth and the retrieved vertical CO_2 gradient based on comparisons between OCO-2 and total column measurements from the ground-based Total Carbon Column Observing Network (TCCON) and regions where XCO_2 variations are expected to be small relative to the measurement uncertainties (Wunch et al., 2011, 2017). We find that use of the bias-corrected data greatly improves comparisons between expected variability within a neighborhood and the actual observed variability (see Appendix). Data quality is evaluated using a variety of metrics that depend on the estimated cloud, aerosol, and surface properties, convergence, and known statistics of the retrieved XCO_2 values (e.g., Mandrake et al., 2013). Data quality flags are given as "warn levels" with values ranging from 0 (best) to 19 (worst). Data with lower warn levels are more likely to represent the statistics of the observed CO_2 whereas data with higher warn levels are likely too strongly affected by interfering effects. The warn levels are primarily evaluated empirically; for these reasons we conservatively use only data with warn levels of 10 or smaller to ensure that the corresponding errors are likely well characterized: https://docserver.gesdisc.eosdis.nasa.gov/public/project/OCO/OCO2_DUG.V7.pdf.

We find empirically that use of data with warn levels less than 10 improves the comparison between the calculated uncertainties and observed variance as discussed in the Appendix.

3 Evaluation of uncertainties

We evaluate the uncertainties of the XCO_2 observations by examining the variations of XCO_2 within small neighborhoods of approximately 10.5 by 100 km in size. Within a neighborhood there are about 190 observations that are taken consecutively. After warn level filtering, this "small neighborhood" test set is composed of approximately 1.5 million land-nadir soundings, 1.0 million land-glint soundings, and 5.0 million ocean-glint soundings. We only select neighborhoods that contains at least 50 soundings that pass these criteria. There are approximately 39 000 small neighborhoods in total across the three modes. stretching from approximately 30° S to 30° N. The strict filtering used in this analysis (warn levels < 10), and the need for at least 50 measurements per bin limits this analysis to latitudes between 30° S and 30° N, primarily over drier, subtropical regions over land but no obvious preferential distribution over the ocean (not shown).

As discussed in O'Dell et al. (2012), a CO_2 profile is simultaneously estimated with all other geophysical parameters that affect the observed radiance such as aerosols, albedo, and surface pressure. The "column-averaged dry air mole fraction" of CO_2 or XCO_2 is then calculated by applying the column operator (e.g., Connor et al., 2008; Worden et al., 2015) to the estimated CO_2 profile. As discussed in Rodgers (2000), Worden et al. (2004), Connor (2008), and Bowman et al. (2006), when this nonlinear retrieval converges to a solution, the estimated XCO_2 can be written as

$$\hat{X} = X_a + \boldsymbol{h}^T \mathbf{A}_{xx} (\boldsymbol{x} - \boldsymbol{x}_a) + \boldsymbol{h}^T \mathbf{A}_{xy} (\boldsymbol{y} - \boldsymbol{y}_a) + \boldsymbol{h}^T \mathbf{G} \boldsymbol{n} + \boldsymbol{h}^T \mathbf{G} \sum\nolimits_i \mathbf{K}_i \boldsymbol{\delta}_i, \quad (1)$$

where $\hat{X}$ is the estimated total column for CO_2, $\hat{X}_a$ is the a priori value used to help regularize the retrieval, and the vector $\boldsymbol{x}$ is the "true" CO_2 profile in units of volume mixing ratio (VMR), discretized onto the forward model atmospheric pressure grid used to calculate the transfer of radiation needed to model the observed radiance. The $\boldsymbol{x}_a$ is the a priori for the CO_2 profile. The vector "$\boldsymbol{y}$" contains all the other parameters that are simultaneously estimated with x

such as aerosol properties, surface albedo, surface pressure. The vector "$\boldsymbol{n}$" is the actual noise in the radiance. The quantities $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{n}$ are not known exactly, only their statistical properties can be estimated. The vector "$\boldsymbol{h}$" is the column operator which maps a profile on the pressure grid defined by "$\boldsymbol{x}$" into a dry air total column. The averaging kernel matrix $\mathbf{A}$ describes the sensitivity of the estimate to each retrieved parameter (Rodgers, 2000). In Eq. (1) the averaging kernel matrix is composed of two parts, $\mathbf{A}_{xx}$ and $\mathbf{A}_{xy}$, described by

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{xx} & \mathbf{A}_{xy} \\ \mathbf{A}_{yx} & \mathbf{A}_{yy} \end{bmatrix}. \quad (2)$$

For example $\mathbf{A}_{xx}$ describes the sensitivity (or $\frac{\partial \hat{x}}{\partial x}$) of the estimated CO_2 on each level, $\boldsymbol{x}$, to its true value, whereas $\mathbf{A}_{xy}$ describes the sensitivity of the estimated CO_2 on each level, $\boldsymbol{x}$, to all other simultaneously estimated parameters, e.g., aerosols. The matrix $\mathbf{G}$ is the gain matrix, which is the derivative of the estimated CO_2 on each level, $\boldsymbol{x}$, to the observed radiance, L (or $G = \frac{\partial \hat{x}}{\partial L}$). The matrix $\mathbf{K}$ is the Jacobian, or sensitivity of the observed radiance to a parameter (e.g., $\mathbf{K} = \frac{\partial L}{\partial x}$). The last term, δ, describes the error in all parameters that are not estimated for this retrieval, but are assumed constant, such as absorption coefficients or instrument functions (e.g., Connor et al., 2008). The mean CO_2 column is written as

$$\begin{aligned} \hat{X}_{\text{mean}} = X_a + \boldsymbol{h}^T \frac{1}{N} \sum\nolimits_{j=1}^{N} \mathbf{A}_j \left(\boldsymbol{x}_j - \boldsymbol{x}_a\right) \\ + \frac{1}{N} \boldsymbol{h}^T \sum_{j=1}^{N} \mathbf{A}_{xy}^{j} \left(\boldsymbol{y}_j - \boldsymbol{y}_a\right) \\ + \frac{1}{N} \sum_{j=1}^{N} \boldsymbol{h}^T \mathbf{G}_j \left(\boldsymbol{n}_j + \sum\nolimits_{i,j} \mathbf{K}_{i,j} \boldsymbol{\delta}_{i,j}\right), \end{aligned} \quad (3)$$

where N is the number of observations within the small neighborhood and for simplicity we assume the column operator h is constant across the domain.

For the next three sections, we test the following hypotheses regarding the observed distributions within the collection of "small neighborhoods" and their calculated uncertainties:

- H1: observed variability in small neighborhood is due to natural XCO_2 variability.
- H2: observed variability in small neighborhood is due to measurement noise.
- H3: observed variability is correlated.
- H4: observed variability within a small neighborhood is described by a slowly varying bias that is not explained by natural XCO_2 variability.

We look at the variability with respect to the neighborhood mean in two ways: (1) for small neighborhoods, the predicted errors for a neighborhood are averaged from the observations that comprise that neighborhood, making the statistics technically a sum of Gaussians, and (2) the variability with respect to the neighborhood mean, sorted by predicted error and aggregated over many neighborhoods – the statistics in this case should be Gaussian; however the locality of the analysis is somewhat reduced.

To evaluate whether measurement noise in the radiances is the primary factor driving variability within a small neighborhood we first assume that the terms $\mathbf{A}_{xy}(\boldsymbol{y}_j - \boldsymbol{y}_a)$ and systematic errors $\mathbf{K}_{i,j}\delta_{i,j}$ do not vary. Based upon these approximations, the difference between an observation and its mean is given by

$$\hat{X}_{\text{obs}} - \hat{X}_{\text{mean}} = \delta_{\text{obs}} = \delta_{\text{XCO}_2} + \mathbf{G}_{\text{obs}} \boldsymbol{n}_{\text{obs}} - \frac{1}{N} \sum_{j}^{N} \mathbf{G}_j \boldsymbol{n}_j, \quad (4)$$

where $\delta_{\text{XCO}_2} = h^T \mathbf{A}(x_{\text{obs}} - x_{\text{mean}})$ and is the difference between the individual "true" XCO_2 and the mean of the "true" XCO_2 values within the neighborhood. Assuming that the measurement noise is spatially uncorrelated, the variance within the small neighborhood (e.g., Bowman et al., 2006) is

$$\begin{aligned} \text{Var}\left|\left|\hat{X}_{\text{obs}} - \hat{X}_{\text{mean}}\right|\right| = \sigma_{\text{obs}}^2 = \sigma_{\text{XCO}_2}^2 \\ + \sigma_{\text{noise}}^2 + \frac{1}{N^2} \sum_{j=1}^{N} \sigma_j^2 - \frac{2}{N} \sigma_k^2, \end{aligned} \quad (5)$$

where $\sigma_{\text{noise}}^2 = \mathbf{G}_K \mathbf{S}_k \mathbf{G}_K^T$ is the measurement uncertainty due to noise. The σ_{XCO_2} is the variability of the true XCO_2 within the small neighborhood. The S_k is the spectral instrumental noise covariance and is calculated during calibration of the instrument. The individual σ_{noise} values are provided for each measurement in the OCO-2 product files. For large N, Eq. (5) is approximately equal to

$$\sigma_{\text{XCO}_2}^2 + \sigma_{\text{noise}}^2.$$

3.1 H1: observed variability is due to natural XCO_2 variability

In order to test whether natural variability, or $\sigma_{\text{XCO}_2}^2$, affects the observed variance of XCO_2 within each neighborhood we examine XCO_2 estimates from the NASA GMAO high-resolution free-running GEOS-5 CO_2 simulation available at https://gmao.gsfc.nasa.gov/global_mesoscale/7km-G5NR/data_access. We use model fields that correspond to each measurement within each neighborhood. There is a spatial mis-match because the model fields are at 7 km × 7 km resolution whereas the OCO-2 data are taken every 3 km; however, we discount the role of spatial resolution because our results do not fundamentally change when smoothing the data from 7 km × 7 km (lat/long) to 14 km × 7 km (lat/long). The dates of the model run (2006)

also do not match the dates of the OCO-2 data (2014–2015). However, since we are looking to quantify the approximate range of natural variability we would not expect inter-annual differences in XCO_2 from winds or fluxes to substantively alter our conclusions when looking at XCO_2 over a large swath of the Earth; as discussed next, comparisons with other data and model are consistent with this conclusion. From the GMAO model, we find that natural variability typically accounts for about 0.08 ppm of the variability within a ~ 100 km neighborhood over land and about 0.06 ppm within a typical ~ 100 km neighborhood over the ocean (see the Appendix). We subsequently assume that natural variability has negligible impact on our conclusions as we find it is on average much smaller than the observed variability.

3.2 H2: observed variability in small neighborhood is due to measurement noise

We next compare observed variability across all the neighborhoods to the calculated uncertainties using two approaches. In the first approach we gather all observations that have approximately the same calculated measurement uncertainty due to noise, σ^2_{noise}, (to within 0.01 ppm) as provided in the OCO-2 product files and compare to the actual variability of these observations. The steps for this comparison are as follows:

1. Calculate the δ_{obs} or difference between an observation and its mean within a small neighborhood as shown in Eq. (4).

2. Collect all of the δ_{obs} values from all neighborhoods used in this analysis whose corresponding σ_{noise} values (measurement uncertainty) are the same to within 0.01 ppm and bin them as a function of σ_{noise}. There are typically about 1000 observations per σ_{noise} bin.

3. Compare the standard deviation of the collection of δ_{obs} values within each bin to the expected standard deviation due to noise or, σ_{noise}. Based on Eq. (5) we should expect to get a linear, one-to-one relationship if the dominant parameter affecting the variability within a small neighborhood is noise.

The results of these comparisons for land-nadir, land-glint, and ocean-glint observations are shown in the upper left panels of Figs. 1, 2, and 3 respectively. These results show the calculated measurement uncertainty due to noise has skill – i.e., there is a linear relationship between calculated and actual error. However, over land the observed random variability is approximately 0.4 ppm larger than the variability expected from noise. We discount synoptic variations in XCO_2 as discussed in the previous section. Other sources of variability could be due to the strong nonlinearities in the retrieval (e.g., Kulawik et al., 2008) or local variability between the true and a priori in the interferences, or non-retrieved parameters. Over the ocean there appears to be an even stronger one-to-one relationship between the calculated uncertainty and the actual uncertainty except for calculated uncertainties less than approximately 0.25 ppm which show a strong inverse relationship. We find that these observations (not shown) tend to occur in the tropics in cloudy regions and that the observations tend to have very high signal-to-noise ratios.

We next test whether the calculated measurement noise is a useful value for predicting the expected distribution of observations within a neighborhood. Because each δ_{obs} is drawn from a distribution with a different variance, we treat the sample of each set of observations, $[\delta_1, \delta_2, \ldots \delta_N]$, as being drawn from an uncorrelated distribution with individual variances σ^2_{obs}. Accordingly, the variance of this sample should be the average of the individual variances σ^2_{obs}:

$$\text{Var}|[\hat{X}_{\text{obs}} - \hat{X}_{\text{mean}}]| = ||[\delta_1, \delta_2, \ldots \delta_N]|| = \frac{1}{N}\sum_j^N \boldsymbol{\sigma}_j^2. \quad (6)$$

The top right panel of Fig. 1 shows a comparison of the observed variance of the XCO_2 distributions (using the left side of Eq. 6) within each neighborhood (black circles) versus the expected variance in XCO_2 using the right side of Eq. (6). Each black symbol represents a single neighborhood. In contrast to the top left panel of Fig. 1, this result suggests that the measurement error has no skill in predicting the observed variance of XCO_2 within a neighborhood.

We next test whether the observed variance, versus that due to measurement noise or sampling, explains the upper right panel of Figs. 1, 2, and 3. To perform this test, we perform the following steps:

1. Within each neighborhood, replace the calculated measurement error with the "actual" measurement error as shown by the solid red line in the upper left panel of Figs. 1, 2, and 3, for each observation.

2. Create a simulated distribution of observations based on this new uncertainty.

3. Randomly sample (or take) one of these observations $\rightarrow$ label this the "modeled" observation.

4. Repeat steps 1–3 for all observations in the neighborhood.

5. Calculate the variance of this "modeled" set of observations for each neighborhood.

The red dots in Figs. 1b, 2b, and 3b show the modeled distributions using the steps discussed above. The modeled distribution is more consistent with the mean of the observed distribution relative to the one-to-one line. However, it is clear from this simulation that errors due to random noise and sampling do not explain the observed variance for each neighborhood although the distribution of variances for the ocean show much better agreement relative to the land distributions.

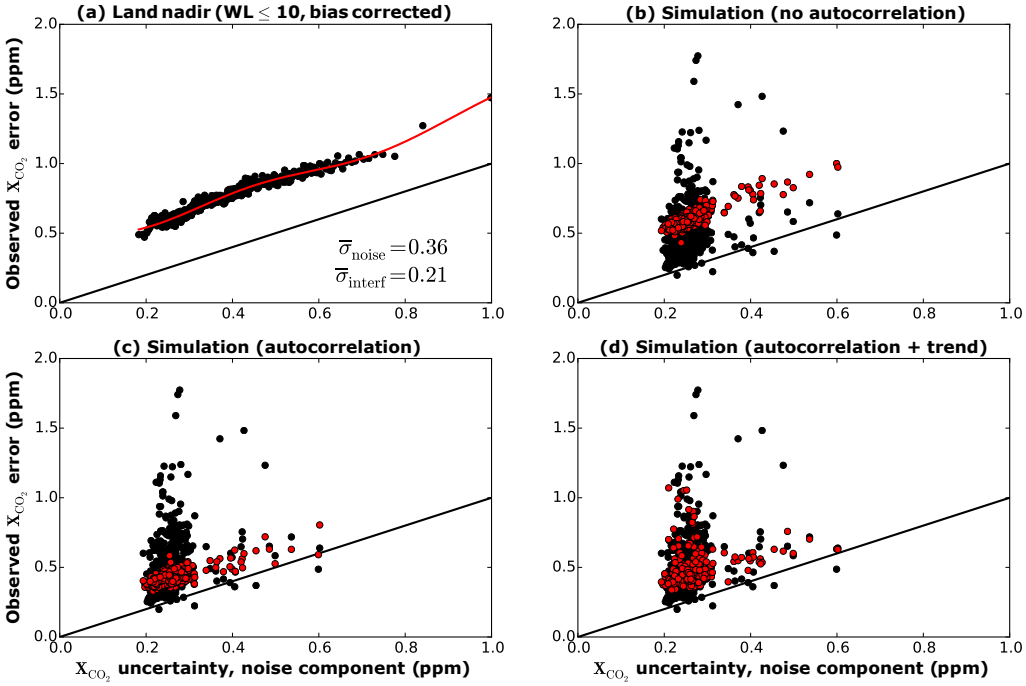

Figure 1. The calculated, observed, and modeled uncertainties for land-nadir observations. The black circles are the observed distributions and red circles are modeled distributions. **(a)** Comparison between expected and actual error when binning all the data by their calculated uncertainty; the solid black line is the one-to-one line. **(b)** Comparison between calculated and random error for each neighborhood (black) versus model (red) when using observed random error from Fig. 1a. **(c)** Same as Fig. 1b but now adding a correlation between adjacent data points. **(d)** Same as Fig. 1b but now accounting for distribution of observed gradients across the neighborhoods used for analysis.

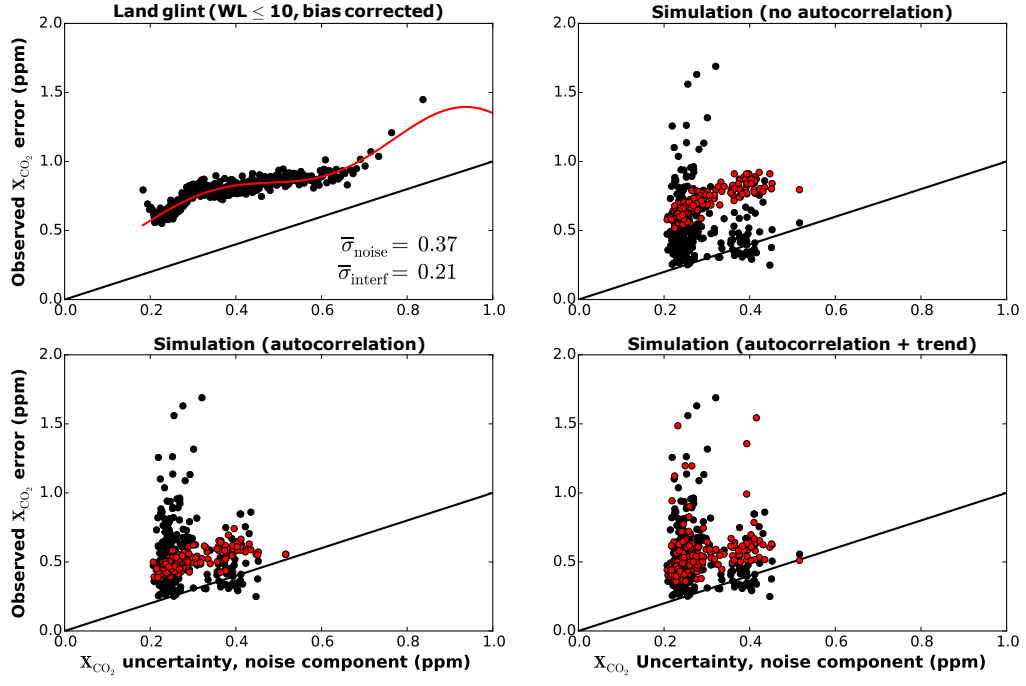

Figure 2. Observed and modeled distributions for land-glint data.

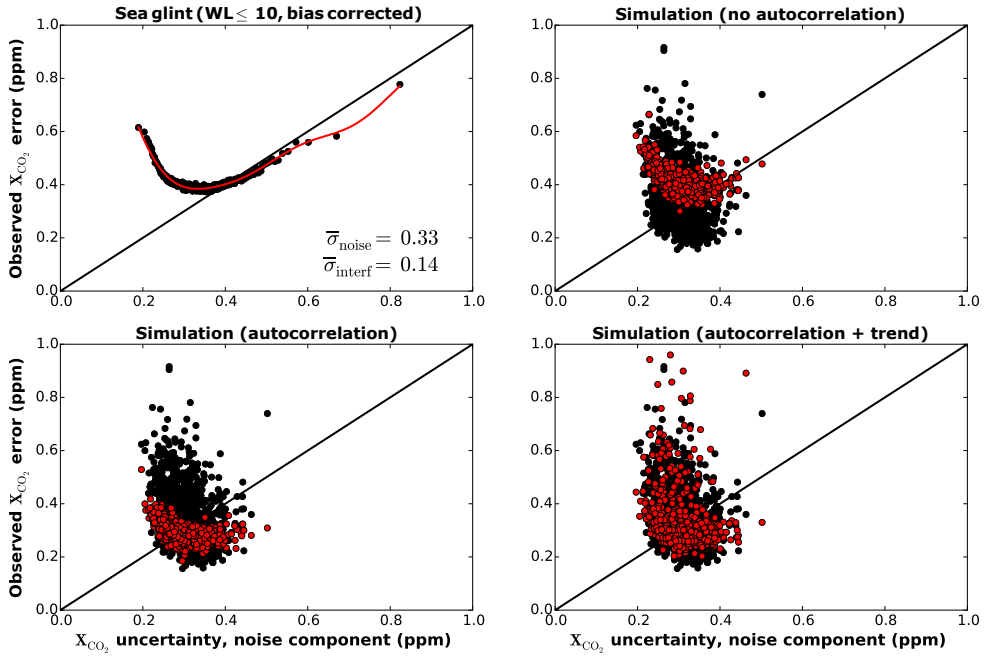

Figure 3. Observed and modeled distributions for sea-glint data.

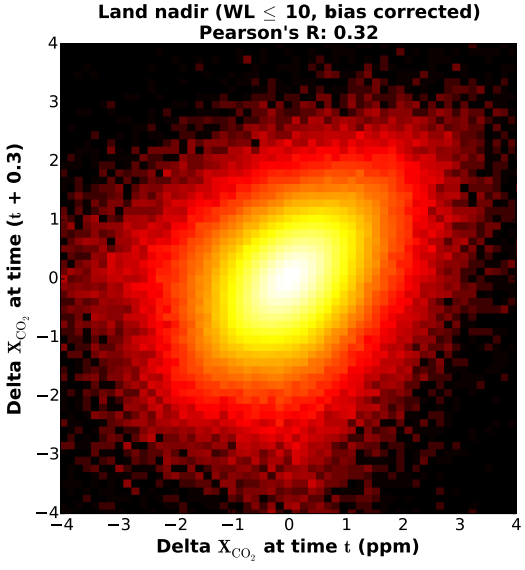

Figure 4. Distribution of XCO_2 values between time steps for the set of observations from each "small neighborhood" used in this analysis.

3.3 H3: uncertainties are correlated

We next test whether observed correlations in the data could explain the distributions of the data within a neighborhood. Figure 4 shows the joint distribution of the XCO_2 anomaly and a 0.3 s lagged anomaly in a neighborhood (consistent with observation sampling). If the data were uncorrelated then the joint distribution should be circular; the asymmetric distribution therefore implies that the errors, as empirically described by the differences, are correlated. Figure 5a and b show that autocorrelation is observed both in time for measurements made of the order of 1 s of each other, and with respect to the spatially adjacent "footprints", the eight simultaneous measurements made by the OCO-2 instrument at each time. The range of correlations for the different observation types, land nadir, land glint, and ocean glint are 0.45, 0.43, and 0.28 as a function of footprint and 0.31, 0.34, and 0.24 as a function of time.

In order to test whether these observed correlations could explain the distributions shown in Figs. 1, 2, and 3, we conservatively use a correlation coefficient of 0.7 for all observations (an extreme case). We then use the following procedure, building on the steps described in the previous section.

1. Within each neighborhood replace the calculated measurement error with the "actual" measurement error as shown in the upper left panels of Figs. 1, 2, and 3 for an observation.

2. Starting with the first observation (in time) within a neighborhood for Footprint #1, sample a value for the observation from the distribution of "actual" measurement errors. Label this the "modeled" observation.

3. For all subsequent observations in time for Footprint #1, sample each "modeled" observation from a distribution that is correlated with the modeled observation at the previous time step and has a variance corresponding to the "actual" measurement error.

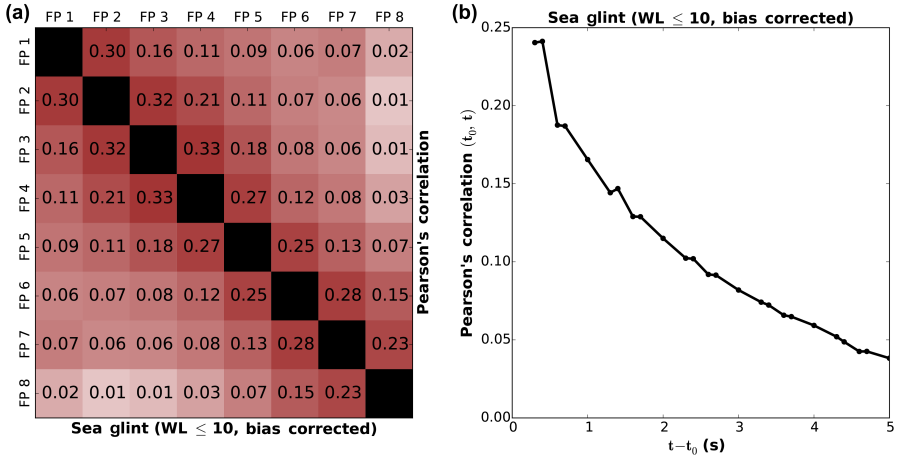

Figure 5. (a) Correlation of differences across pixels between observed minus mean within a neighborhood. **(b)** Correlation between observations for a single pixel.

4. For observations in Footprints #2–8, sample each modeled observation from a distribution correlated with the modeled observation at the same time step in the previous (adjacent) footprint, again with a variance corresponding to the "actual" error.

5. Calculate variance of this "modeled" set of observations, for each neighborhood.

As can be seen in the lower left panels of Figs. 1, 2, and 3, adding correlations to the data makes the comparison worse because the modeled distributions become much narrower relative to the modeled distributions in the upper right panels of these figures. Our conservative choice of a 0.7 correlation between observations at adjacent times and footprints illustrates this effect clearly. We therefore conclude that while correlations are empirically observed in the data, they cannot completely explain the observed distributions within the small neighborhoods.

3.4 H4: observed variability within a small neighborhood is described by a slowly varying bias that is not explained by natural XCO_2 variability

We next examine whether "non-random" uncertainties could explain the observed distributions in the upper right panels of Figs. 1, 2, and 3. For example, as shown in Eq. (1), the jointly retrieved parameters $(\boldsymbol{y} - \boldsymbol{y}_a)$ might remain constant across a neighborhood but the averaging kernel associated with this term, which is given by $\mathbf{A}_{xy} = \frac{\partial x}{\partial L}\frac{\partial L}{\partial y} = \mathbf{G}\mathbf{K}_y$, can vary across a neighborhood as the pointing angle varies. The effect of non-retrieved parameters such as instrument effects or spectroscopy on the estimate can vary for the same reason.

Figure 6 shows the variation of XCO_2 across one of the ocean neighborhoods for all eight OCO-2 footprints (denoted by "FP"). The right panel shows the observed distribution in black relative to the mean XCO_2 of the neighborhood. For reference, the red dashed line in the right panel indicates the expected distribution if only the calculated random noise explained the variability. The slope shown in Fig. 6 represents an extreme case but demonstrates that observations can pass the set of quality flags but still show this unlikely behavior over the ocean. Figure 7a shows the distribution of all slopes across all land-nadir neighborhoods used in this study and different fits (Gaussian, Lorentz, Laplace) to the distribution. The Laplace distribution provides the best overall fit so we use its functional form as a simple, convenient description of the shape of the sharply peaked slope distribution. More complex models such as Gaussian mixtures might also describe the shape of this distribution of slopes as drawn from several distinct "populations" of neighborhoods, but we leave such an analysis to future work. For comparison, Fig. 7b shows the expected distribution using the GMAO model XCO_2 fields described earlier. As with the OCO-2 data shown in Fig. 7a, the histogram in Fig. 7b describes the distribution of XCO_2 gradients across 100 km neighborhoods spatially corresponding to the OCO-2 data. The expected distribution of natural variability of XCO_2 across the 100 km neighborhoods is much smaller than observed (as with the conclusions about natural variability discussed for H1); we therefore do not expect that the natural carbon cycle can explain these observed variations.

For land-nadir, land-glint, and ocean-glint data the variance of the slopes is given by 1.28 ppm 100 km^{-1}, 1.12 ppm 100 km^{-1}, and 0.48/100 km respectively. As shown in the Appendix, these values are much larger than the gradients expected from natural variability, such as the latitudinal gradient in XCO_2. Consequently, we expect these gradients to be related to interferences in the XCO_2 data.

To test whether these slowly varying changes explain the distribution of XCO_2 within small neighborhoods we follow

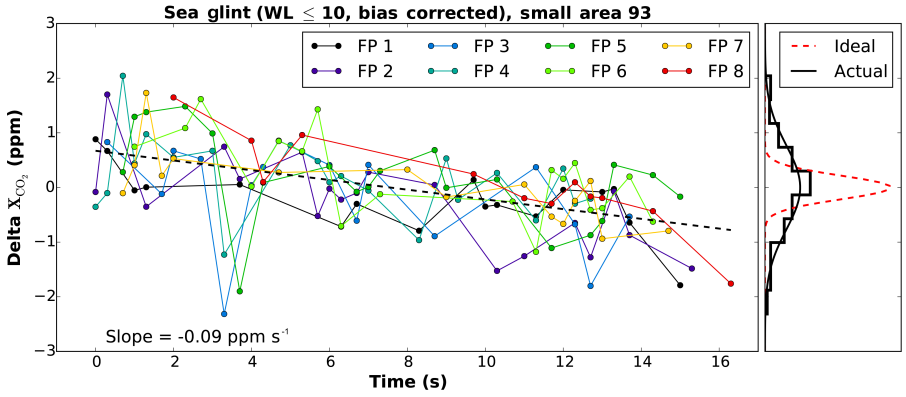

Figure 6. The difference between XCO_2 and the mean value (or delta XCO_2) for one of the small neighborhoods (or areas) used in this analysis. The left panel shows the differences for each footprint (FP), representative of one of the OCO-2 observations. The right panel shows the observed distribution (actual) and one calculated if the distributions were representative of the calculated random error.

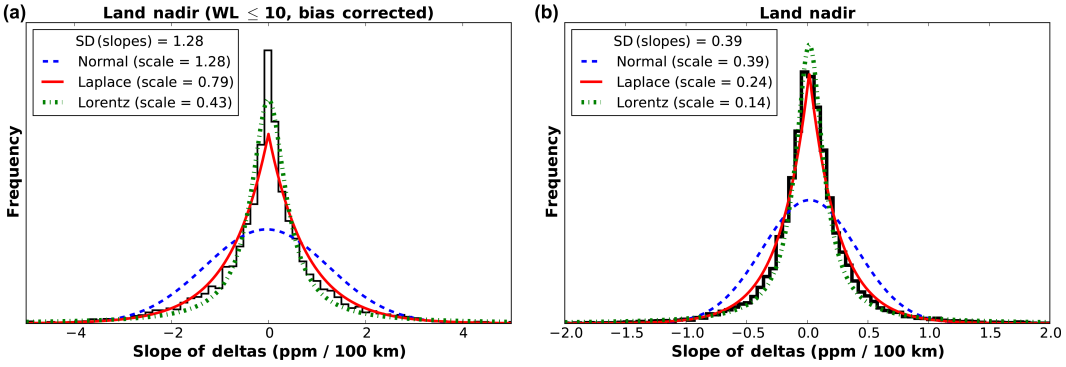

Figure 7. (a) The distributions of slopes of the observed XCO_2 gradients across all the small neighborhoods corresponding to land-nadir observations. **(b)** The expected distribution of slopes in XCO_2 based on the GMAO high-resolution model.

the same steps described in Sect. 3.2 and 3.3 but now add another:

1. Within each neighborhood replace the calculated measurement error with the "actual" measurement error as shown in the upper left panels of Figs. 1, 2, and 3 for an observation.

2. Starting with the first observation (in time) within a neighborhood for Footprint #1, sample a value for the observation from the distribution of "actual" measurement errors. Label this the "modeled" observation.

3. For all subsequent observations in time for Footprint #1, sample each "modeled" observation from a distribution that is correlated with the modeled observation at the previous time step and has a variance corresponding to the "actual" measurement error.

4. For observations in Footprints #2–8, sample each modeled observation from a distribution correlated with the modeled observation at the same time step in the previous (adjacent) footprint, again with a variance corresponding to the "actual" error.

5. Adjust each modeled observation with a linear function where the slope of the linear function is randomly chosen from the fitted Laplace distribution to the slopes (e.g., the Laplace function shown in Fig. 7).

6. Calculate variance of this "modeled" set of observations, for each neighborhood.

Figures 1, 2, and 3 (lower right panels) show the best overall agreement between modeled distributions of XCO_2 relative to the mean and the expected distributions based on observations, demonstrating that a slowly varying bias is needed to best explain the observed distributions within a grid of approximately 100 km × 10 km.

Each typical observation has a random error related to noise and a systematic error that is in principle bounded by the calculated interference error (e.g., Boxe et al., 2010) and is approximately 0.2 ppm. Within a typical grid box an OCO-2 observed measurement over land is within 1.28/2, or ∼ 0.65 ppm of the mean XCO_2 value. For these reasons, we

expect that a typical observation over land has a systematic error of at least 0.65 ppm, about 2 to 3 times larger than the calculated interference error.

In contrast, the observed distributions of slopes for the ocean data is 0.48 ppm 100 km^{-1}, or a mean error of 0.24 ppm. This error is 70 % larger than the mean calculated interference error of 0.14 ppm. Because the distribution of ocean data within "bins" (Fig. 3, upper left panel) is also well described by the calculated random error, we conclude that the ocean-glint data are reasonably well characterized by their calculated uncertainties for this size of a grid box, except for calculated noise (or precision) uncertainties that are less than $\sim$ 0.25 ppm.

We find no relationship between the distribution of slopes for a neighborhood and the corresponding mean of the calculated interference error, suggesting that the calculated interference error does not explain the observed slope within a neighborhood, in contrast to the measurement error. However, there is a correlation between the slope and the estimated magnitude of interferences, such as aerosol optical depth, surface albedo, and surface pressure. For example, the correlation between the slopes of land-glint data with the mean uncertainty in the interferences is 0.06 whereas the correlation between the observed slopes in XCO_2 and similarly calculated observed slopes in aerosol optical depth is 0.37. This correlation suggests that the observed slow variations in XCO_2 across a neighborhood could be related to how interferences affect the XCO_2 estimate as OCO-2 takes observations across a neighborhood.

4 Summary

We compare XCO_2 variability from OCO-2 observed within small neighborhoods of $\sim$ 100 km $\times$ 10.5 km to evaluate the precision and accuracy of the XCO_2 data. Our analysis shows that the calculated precision of the OCO-2 data has skill as there is a linear relationship between the measurement noise and the random variation of the OCO-2 data. We find that the precision and accuracy of a typical ocean measurement is approximately 0.35 and 0.2 ppm respectively, consistent with the calculated errors (assuming that the accuracy is bounded by the calculated interference error and does not include smoothing error). The precision and accuracy of a typical land measurement (both nadir and glint) is approximately 0.75 and 0.65 ppm. These values can be compared to the calculated measurement and interference errors of approximately 0.36 and 0.2 ppm. Differences are likely due to nonlinearities in the retrieval or random components of interference error which are likely poorly characterized (Connor et al., 2016). The accuracy is estimated from observed gradients in XCO_2 of approximately 1.28 ppm 100 km^{-1} across the small neighborhoods used in this analysis.

This 0.65 ppm estimate for the accuracy of the land data is likely a lower bound because it is based on observed gradients across a region with the bias removed.

For example Wunch et al. (2017) shows that the root-mean-square difference between the land nadir and glint data is 1.36 ppm, which is twice the value that we obtain for the accuracy. However, both ours and that of Wunch et al. (2017) suggest a relationship between these larger than expected uncertainties in the OCO-2 data and interferences due to surface properties or aerosols.

This analysis sheds further light on the sources of uncertainty of the observed XCO_2 data. For example, the XCO_2 gradient variability in the small neighborhoods over the ocean as compared to the land suggests that the largest uncertainty in OCO-2 XCO_2 data is related to surface properties such as surface pressure or albedo because we expect larger variations of these geophysical parameters over land. The observed gradients could also be related to the variation in solar zenith angle as OCO-2 data take observations because the effect is manifested as a slowly varying quantity in addition to increased random variability. The observed distribution of these XCO_2 gradients over the whole globe, which has a Laplace distribution, is also a potential clue as any bottom-up or future analysis that attempts to model the XCO_2 uncertainties should also replicate this distribution. A future study in which the empirically calculated uncertainties presented here are tested using the more refined theoretical uncertainties discussed in Connor et al. (2016), as well as the TCCON data, will hopefully reveal and characterize the likely sources of 70 these uncertainties.

Data availability. The data used in this paper are publicly accessible at the following web page: https://disc.sci.gsfc.nasa.gov/datasets/OCO2_L2_Lite_FP_V7r/summary?keywords=OCO-2.

Appendix A: Ancillary comparisons of GMAO high-resolution XCO_2 data

This appendix provides supporting analysis for the results discussed in the main text by comparing the distribution of synoptic variability of XCO_2 as derived by the GMAO high-resolution 7 km × 7 km XCO_2 fields with the variability expected by the mean random error for all data with a bound of the variability expected by the interference error.

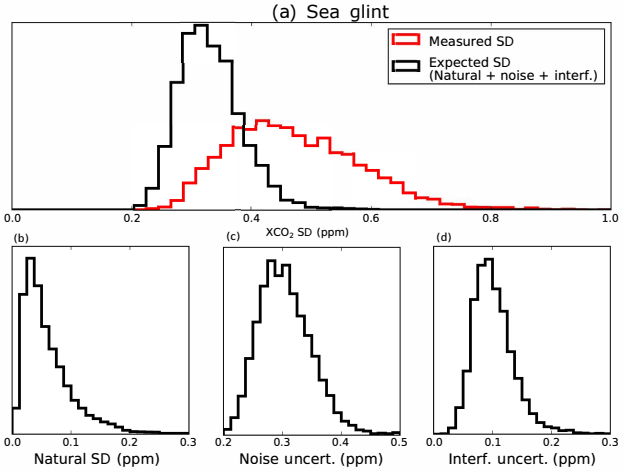

Figure A1. Distribution of all data used in this analysis for sea glint with no data quality flags used (all warn levels are used). **(a)** The expected **(b, c, d)** in black and the actual, bias removed, variability (red) line. The lower panels show the expected variability from natural sources (Appendix A), noise, and interferences.

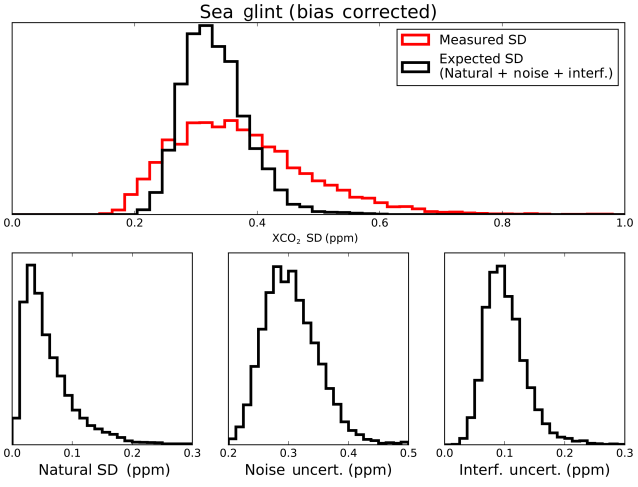

Figure A2. Same as Fig. A1 but using only data with "warn levels" < 10.

Figures A1–A6 show comparisons of the expected distribution for all data used in this analysis (black line) with the actual distribution (red line). The components due to natural variability, noise, and interferences are shown on the bottom of each figure. Note that the x axes for the three bottom figures are different in order to illustrate the full range and shape of the distribution. Figures A1–A6 show that the natural variability on average is about a factor of 3 less than that expected from noise and interferences. Use of the warn levels greatly improves the comparison between the actual and expected distribution, especially for the ocean. The actual variability is consistent with the results shown in Figs. 1–3 with the variability primarily attributed to the observed gradients in OCO-2 XCO_2 data across the ∼ 100 km neighborhoods.

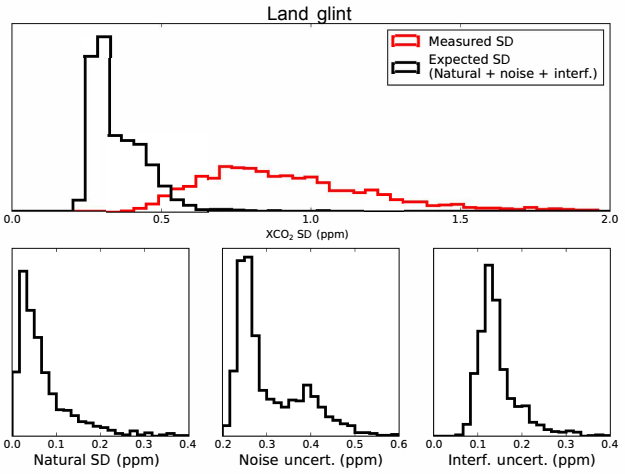

Figure A3. Same as Fig. A1 but for land glint.

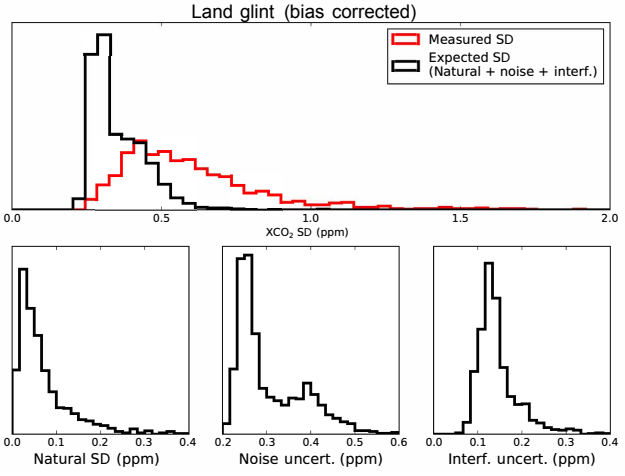

Figure A4. Same as Fig. A2 but for land glint.

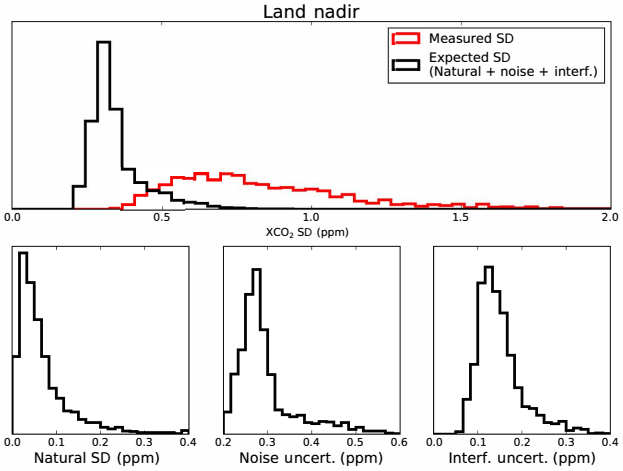

Figure A5. Same as Fig. A1 but for land-nadir scenes.

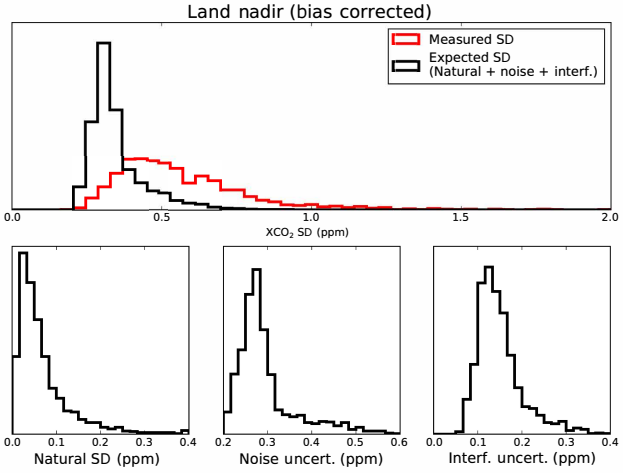

Figure A6. Same as Fig. A2 but for land-nadir scenes.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. Part of this research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. Funding for Susan Kulawik provided by NASA Roses NMO710771/NNN13D771T, "Assessing OCO-2 predicted sensitivity and errors". ftp://ftp.nccs.nasa.gov/Ganymed/7km/.

Edited by: Ilse Aben
Reviewed by: two anonymous referees

References

Bösch, H., Toon, G. C., Sen, B., Washenfelder, R. A., wennberg, P., buchwitz, M., de Beek, R., Burrows, J., Crisp, D., Christi, M., Connor, B., Natraj, V., and Yung, Y.: Space-based near-infrared CO_2 measurements: Testing the Orbiting Carbon Observatory retrieval algorithm and validation concept using SCIAMACHY observations over Park Falls, Wisconsin, J. Geophys. Res.-Atmos., 111, D23302, https://doi.org/10.1029/2006JD007080, 2006.

Boesch, H., Baker, D., Connor, B., Crisp, D., and Miller, C.: Global Characterization of CO2 Column Retrievals from Shortwave-Infrared Satellite Observations of the Orbiting Carbon Observatory-2 Mission, Remote Sensing, 3, 270–304, https://doi.org/10.3390/rs3020270, 2011.

Bowman, K. W., Rodgers, C. D., Kulawik, S. S., Worden, J., Sarkissian, E., Osterman, G., Steck, T., Lou, M., Eldering, A., and Shephard, M.: Tropospheric emission spectrometer: Retrieval method and error analysis, IEEE T. Geosci. Remote, 44, 1297–1307, 2006.

Boxe, C. S., Worden, J. R., Bowman, K. W., Kulawik, S. S., Neu, J. L., Ford, W. C., Osterman, G. B., Herman, R. L., Eldering, A., Tarasick, D. W., Thompson, A. M., Doughty, D. C., Hoffmann, M. R., and Oltmans, S. J.: Validation of northern latitude Tropospheric Emission Spectrometer stare ozone profiles with ARC-IONS sondes during ARCTAS: sensitivity, bias and error analysis, Atmos. Chem. Phys., 10, 9901–9914, https://doi.org/10.5194/acp-10-9901-2010, 2010.

Connor, B. J., Boesch, H., Toon, G., Sen, B., Miller, C., and Crisp, D.: Orbiting Carbon Observatory: Inverse method and prospective error analysis, J. Geophys. Res, 113, D05305, https://doi.org/10.1029/2006JD008336, 2008.

Connor, B., Bösch, H., McDuffie, J., Taylor, T., Fu, D., Frankenberg, C., O'Dell, C., Payne, V. H., Gunson, M., Pollock, R., Hobbs, J., Oyafuso, F., and Jiang, Y.: Quantification of uncertainties in OCO-2 measurements of XCO_2: simulations and linear error analysis, Atmos. Meas. Tech., 9, 5227–5238, https://doi.org/10.5194/amt-9-5227-2016, 2016.

Crisp, D., Fisher, B. M., O'Dell, C., Frankenberg, C., Basilio, R., Bösch, H., Brown, L. R., Castano, R., Connor, B., Deutscher, N. M., Eldering, A., Griffith, D., Gunson, M., Kuze, A., Mandrake, L., McDuffie, J., Messerschmidt, J., Miller, C. E., Morino, I., Natraj, V., Notholt, J., O'Brien, D. M., Oyafuso, F., Polonsky, I., Robinson, J., Salawitch, R., Sherlock, V., Smyth, M., Suto, H., Taylor, T. E., Thompson, D. R., Wennberg, P. O., Wunch, D., and Yung, Y. L.: The ACOS CO_2 retrieval algorithm – Part II: Global XCO_2 data characterization, Atmos. Meas. Tech., 5, 687–707, https://doi.org/10.5194/amt-5-687-2012, 2012.

Kort, E. A., Frankenberg, C., Miller, C. E., and Oda, T.: Space-based observations of megacity carbon dioxide, Geophys. Res. Lett, 39, L17806, https://doi.org/10.1029/2012GL052738, 2012.

Kuai, L., Worden, J., Kulawik, S., Bowman, K., Lee, M., Biraud, S. C., Abshire, J. B., Wofsy, S. C., Natraj, V., Frankenberg, C., Wunch, D., Connor, B., Miller, C., Roehl, C., Shia, R.-L., and Yung, Y.: Profiling tropospheric CO_2 using Aura TES and TCCON instruments, Atmos. Meas. Tech., 6, 63–79, https://doi.org/10.5194/amt-6-63-2013, 2013.

Kulawik, S. S., Bowman, K. W., Luo, M., Rodgers, C. D., and Jourdain, L.: Impact of nonlinearity on changing the a priori of trace gas profile estimates from the Tropospheric Emission Spectrometer (TES), Atmospheric, 8, 3081–3092, 2008.

Kuze, A., Suto, H., Nakajima, M. and Hamazaki, T.: Thermal and near infrared sensor for carbon observation Fourier-transform spectrometer on the Greenhouse Gases Observing Satellite for greenhouse gases, Appl. Opt., 48, 6716–6733, 2009.

Mandrake, L., Frankenberg, C., O'Dell, C. W., Osterman, G., Wennberg, P., and wunch, D.: Semi-autonomous sounding selection for OCO-2, Atmos. Meas. Tech., 6, 2851–2864, https://doi.org/10.5194/amt-6-2851-2013, 2013.

Miller, C. E., Crisp, D., DeCola, P. L., Olsen, S. C., Randerson, J. T., Michalak, A. M., Alkhaled, A., Rayner, P., Jacob, D. J., Suntharalingam, P., Jones, D. B. A., Denning, A. S., Nicholls, M. E., Doney, S. C., pawson, S., Boesch, H., Connor, B. J., Fung, I. Y., O'brien, D., Salawitch, R. J., Sander, S. P., Sen, B., Tans, P., Toon, G. C., Wennberg, P. O., Wofsy, S. C., Yung, Y. L., and Law, R. M.: Precision requirements for space-based XCO 2data, J. Geophys. Res, 112, D10314, https://doi.org/10.1029/2006JD007659, 2007.

O'Dell, C. W., Connor, B., Bösch, H., O'Brien, D., Frankenberg, C., Castano, R., Christi, M., Crisp, D., E ldering, A., Fisher, B., Gunson, M., McDuffie, J., Miller, C. E., Natraj, V., Oyafuso, F., Polonsky, I., Smyth, M., Taylor, T., Toon, G. C., Wennberg, P. O., and Wunch, D.: Corrigendum to "The ACOS CO_2 retrieval algorithm – Part 1: Description and validation against synthetic observations" published in Atmos. Meas. Tech., 5, 99–121, 2012, Atmos. Meas. Tech., 5, 193–193, https://doi.org/10.5194/amt-5-193-2012, 2012.

Parazoo, N. C., Bowman, K., Frankenberg, C., Lee, J.-E., Fisher, J. B., Worden, J., Jones, D. B. A., Berry, J., Collatz, G. J., Baker, I. T., Jung, M., Liu, J., Osterman, G., O'Dell, C., Sparks, A., Butz, A., Guerlet, S., Yoshida, Y., Chen, H., and Gerbig, C.: Interpreting seasonal changes in the carbon balance of southern Amazonia using measurements of XCO_2 and chlorophyll fluorescence from GOSAT, Geophys. Res. Lett, 40, 2829–2833, https://doi.org/10.1002/grl.50452, 2013.

Wofsy, S. C.: The HIPPO Science Team and Cooperating Modellers and Satellite Teams: HIAPER Pole-to-Pole Observations (HIPPO): fine-grained, global-scale measurements of climatically important atmospheric gases and aerosols, Philos. T. R. Soc. A, 369, 2073–2086, https://doi.org/10.1098/rsta.2010.0313, 2011.

Worden, J., Kulawik, S., Shepard, M., Clough, S., Worden, H., Bowman, K., and Goldman, A.: Predicted errors of tropospheric emission spectrometer nadir retrievals from

spectral window selection, J. Geophys. Res, 109, D09308, https://doi.org/10.1029/2004JD004522, 2004.

Worden, J. R., Turner, A. J., Bloom, A., Kulawik, S. S., Liu, J., Lee, M., weidner, R., Bowman, K., Frankenberg, C., Parker, R., and Payne, V. H.: Quantifying lower tropospheric methane concentrations using GOSAT near-IR and TES thermal IR measurements, Atmos. Meas. Tech., 8, 3433–3445, https://doi.org/10.5194/amt-8-3433-2015, 2015.

Wunch, D., Wennberg, P. O., Toon, G. C., Connor, B. J., Fisher, B., Osterman, G. B., Frankenberg, C., Mandrake, L., O'Dell, C., Ahonen, P., Biraud, S. C., Castano, R., Cressie, N., Crisp, D., Deutscher, N. M., Eldering, A., Fisher, M. L., Griffith, D. W. T., Gunson, M., Heikkinen, P., Keppel-Aleks, G., Kyrö, E., Lindenmaier, R., Macatangay, R., Mendonca, J., Messerschmidt, J., Miller, C. E., Morino, I., Notholt, J., Oyafuso, F. A., Rettinger, M., Robinson, J., Roehl, C. M., Salawitch, R. J., Sherlock, V., Strong, K., Sussmann, R., Tanaka, T., Thompson, D. R., Uchino, O., Warneke, T., and Wofsy, S. C.: A method for evaluating bias in global measurements of CO_2 total columns from space, Atmos. Chem. Phys., 11, 12317–12337, https://doi.org/10.5194/acp-11-12317-2011, 2011.

Wunch, D., Wennberg, P. O., Osterman, G., Fisher, B., Naylor, B., Roehl, C. M., O'Dell, C., Mandrake, L., Viatte, C., Kiel, M., Griffith, D. W. T., Deutscher, N. M., Velazco, V. A., Notholt, J., Warneke, T., Petri, C., De Maziere, M., Sha, M. K., Sussmann, R., Rettinger, M., Pollard, D., Robinson, J., Morino, I., Uchino, O., Hase, F., Blumenstock, T., Feist, D. G., Arnold, S. G., Strong, K., Mendonca, J., Kivi, R., Heikkinen, P., Iraci, L., Podolske, J., Hillyard, P. W., Kawakami, S., Dubey, M. K., Parker, H. A., Sepulveda, E., García, O. E., Te, Y., Jeseck, P., Gunson, M. R., Crisp, D., and Eldering, A.: Comparisons of the Orbiting Carbon Observatory-2 (OCO-2) X_{CO_2} measurements with TCCON, Atmos. Meas. Tech., 10, 2209–2238, https://doi.org/10.5194/amt-10-2209-2017, 2017.

Yoshida, Y., Ota, Y., Eguchi, N., Kikuchi, N., Nobuta, K., Tran, H., Morino, I., and Yokota, T.: Retrieval algorithm for CO_2 and CH_4 column abundances from short-wavelength infrared spectral observations by the Greenhouse gases observing satellite, Atmos. Meas. Tech., 4, 717–734, https://doi.org/10.5194/amt-4-717-2011, 2011.