“Random Error and Averaging Strategies for XCO2 Observations from Spaceborne IDPA Lidar Onboard DQ-1 Satellite”

By ZengCheng Fan et al., submitted to *Atmospheric Measurement Techniques*
egusphere-2026-3759

Referee Report

Summary

This study uses a full year of DQ-1/ACDL L2B data with the primary goal to propose averaging strategies of single-shot XCO_2. This is because single-shot XCO_2 is extremely noisy (with single-shot errors like 5-30 ppm, depending primarily on surface reflectance). They do this by comparing theoretical random errors to Allan variances, where the latter include systematic errors. They come up with different averaging strategies for different surface types, with longer averaging windows needed for lower-signal surfaces like water,snow,and ice, and shorter windows needed for brighter surfaces like deserts.

The works is valuable and the questions posed by the authors are useful to be answered. However, I recommend rejection for several severe structural problems which, when addressed, will likely change the manuscript into something pretty different.

Severe Structural Issues

1. The theoretical single-shot random error appears to omit signal photon (shot) noise, and is never tested against the observations.

The propagation framework (Eqs. 6–8, after Ehret et al., 2008) is correct. The problem is the SNR estimator at L151–154. The authors assume the noise can be taken from the standard deviation of the signal without any return. By construction, this completely ignores photon (shot) noise, or any other noise sources related to a measured signal (such as speckle). Basically they include background noise only. This may underestimate the actual random noise, perhaps severely.

Based on back-of-the-envelope calculations using parameters given in the instrument paper Zhu et al. (2021, Table 2), I calculate that SNR may be overstated by up to a factor of four over bright surfaces, with σ_XCO_2 understated correspondingly. Though it might be smaller, but I cannot see any way to tell from the present manuscript. The authors need to either justify the assumption that photon noise is small enough to ignore (and I find this hard to believe this would be the case), or include it by estimating it. The former could be

accomplished by plotting the measured standard deviation of single-shot XCO_2 retrievals against Eq. (8) over homogeneous scenes like deserts, where the shot noise would be largest. I'm no instrument expert, but I would think the photon noise might be estimated from instrument parameters using Zhu et al. (2021, equations 6-7). It's a quite important instrument parameter, and I have no doubt the authors can either estimate it themselves or work with the instrument engineers directly to do so.

The consequences are potentially significant. If photon noise is significant over bright surfaces, σ_random rises there while changing little over dark ones, basically invalidating all the follow-on results presented in the current manuscript.

2. The Allan variance analysis is not well posed, so the conclusions drawn from it are unsupported.

Eq. (10) requires that the random errors within a window be approximately equal. The only stated selection criterion for the continuous segments (Table 1) is length: N > 148 × 7 for land and N > 296 × 5 for ocean, roughly 350 km and 500 km at the ~337.5 m along-track spacing. No constraint is placed on surface type, reflectance or topography. For segments this long, the authors are no doubt mixing surface types, meaning that the random errors will not be close to equal, so the theoretical curve plotted against the Allan deviation in Fig. 7 becomes an invalid comparison.

The excess of Allan deviation over theory therefore has at least three unseparated causes: within-segment heterogeneity of σ, the possible missing photon noise term of item 1, and genuine along-track XCO_2 variability. Section 5.2 attributes it to correlated error plus real variability, having excluded geophysical bias by an IWF-weighting test run on those same segments — a regime where geophysical bias is small by construction, since it scales with Cov(XCO_2, IWF). That conclusion is not established.

3. Stratification is by land cover type rather than by surface reflectivity.

The noise has a simple and largely parameter-free dependence on surface reflectance, which I derive below because it's not obvious a priori. Writing the received power at either wavelength in the usual hard-target form,

$$P_\lambda = \frac{E_\lambda}{t_{\mathrm{eff}}} \cdot \frac{A_{\mathrm{tel}}\,\eta}{R^2} \cdot \rho^* \cdot T_\lambda^2,$$

where A_{tel} is the receiver collecting area and η its optical throughput, R the range, ρ^* the target reflectance per steradian, T_λ^2 the two-way atmospheric transmission, and t_{eff} the

effective pulse width, broadened by within-footprint relief and canopy structure. It is convenient to group the two surface-dependent factors into a single effective reflectance,

$$\rho_{\mathrm{eff}} \equiv \rho^{*}\, T^{2},$$

where T is part of the transmittance independent of CO_2 absorption (ie, common to both wavelengths, for instance due to aerosols). The online and offline wavelengths are nearly identical, so everything except the CO_2 absorption is common to both, and thus we get

$$P_{\mathrm{off}} = C\,\rho_{\mathrm{eff}}, \qquad P_{\mathrm{on}} = C\,\rho_{\mathrm{eff}}\, e^{-2\delta},$$

with δ the differential absorption optical depth due to CO_2, and $C = E\,A_{\mathrm{tel}}\,\eta/(t_{\mathrm{eff}}R^{2})$ is a nearly-constant instrument factor (the range varies by under 2 % from orbit). A stated in item (1) above, detected-signal noise is a signal-independent floor plus a photon (shot) term,

$$\sigma_P^{2} = \sigma_0^{2} + kP, \qquad \mathrm{SNR}(P) = \frac{P}{\sqrt{\sigma_0^{2} + kP}},$$

where σ_0 is "background" or "dark" noise, and k is an instrument constant controlling the photon noise. Substituting into the manuscript's own Eqs. (7)–(8), we get

$$\sigma_{X\mathrm{CO_2}} = \frac{1}{2\,\mathrm{IWF}}\sqrt{\frac{\sigma_0^{2} + kP_{\mathrm{on}}}{P_{\mathrm{on}}^{2}} + \frac{\sigma_0^{2} + kP_{\mathrm{off}}}{P_{\mathrm{off}}^{2}}} = \frac{1}{2\,\mathrm{IWF}}\sqrt{\frac{\sigma_0^{2}(1+e^{4\delta})}{C^{2}\rho_{\mathrm{eff}}^{2}} + \frac{k(1+e^{2\delta})}{C\,\rho_{\mathrm{eff}}}}$$

The factors in δ are not free parameters. Rearranging the manuscript's own Eq. (4), $\delta = X_{\mathrm{CO_2}} \cdot \mathrm{IWF}$, and thus is nearly controlled by IWF alone — that is, dominated by surface pressure, with a weaker dependence on the temperature profile and $X_{\mathrm{CO_2}}$. In any event, the expression above for $\sigma_{X\mathrm{CO_2}}$ is consequently a *two-parameter model of the entire global single-shot random-error field*, in the two observables ρ_{eff} and IWF, with only the instrument combinations σ_0^{2}/C^{2} and k/C to be fitted. That is the natural object of comparison, rather than seventeen land-cover classes regrouped into four. The two limits are

$$\sigma_{X\mathrm{CO_2}} \propto \frac{1}{\rho_{\mathrm{eff}}\,\mathrm{IWF}} \quad \text{(floor-limited)}, \qquad \sigma_{X\mathrm{CO_2}} \propto \frac{1}{\sqrt{\rho_{\mathrm{eff}}}\,\mathrm{IWF}} \quad \text{(signal-limited)},$$

with the transition at

$$\rho_{\mathrm{eff}}^{\mathrm{knee}} \simeq \frac{\sigma_0^{2}}{k\,C}\cdot\frac{1+e^{4\delta}}{1+e^{2\delta}}.$$

For a typical sea-level column with $X_{\mathrm{CO_2}}$=400 ppm, $\delta \approx 0.78$, giving

$$\frac{1+e^{4\delta}}{1+e^{2\delta}} \approx 4.1,$$

falling to about 2.1 at 600 hPa. This leads to $\rho_{\mathrm{eff}}^{\mathrm{knee}} \approx 4.1\,\sigma_0^2/kC$ at sea level. Plotting $\sigma_{X\mathrm{CO}_2} \cdot \mathrm{IWF}$ against ρ_{eff} on logarithmic axes, after dividing out the known δ brackets, should ideally collapse all surfaces onto a single curve with asymptotic slopes of -1 and $-1/2$, and the position of the knee returns σ_0^2/kC — the instrument's on-orbit noise floor relative to its shot-noise coefficient. Scatter about that curve is the quantity that would justify a categorical scheme in the first place.

A single plot of σ_XCO_2·IWF against ρ_{eff} would therefore test whether IGBP class carries information beyond reflectance, would reveal whether a shot-noise knee is present (bearing directly on item 1), and would yield the instrument's on-orbit noise-equivalent power as a by-product. Instead, the paper offers seventeen surface classes regrouped into four, with post-hoc narrative explanations — needleleaf versus broadleaf forest, for example — that are never tested against the obvious physical alternative, that both orderings are essentially reflectance effects. Such an analysis might therefore drastically alter the primary conclusions of the paper.

4. The chosen AVX scheme is likely not the best choice!

This is a critical choice and may really change the paper's results. From equations (1) and (3) from the manuscript, single-shot XCO_2 is derived from the measurements as

$$XCO_2 = \frac{DOAD}{IWF} = \frac{1}{2\,IWF} \ln \frac{P_{off}\,E_{on}}{P_{on}\,E_{off}}$$

where the measured return powers are the noisiest quantities, and generally follow normal distributions. When one averages XCO_2 over many shots, there are three generally ways to do it, first described in the GHG lidar community in Tellier et al. (2018) for MERLIN measurements of methane.

AVX: simply average up the single-shot XCO2

AVD: Average up DOAD and IWF separately, and take their ratio.

AVS: Average P_{on}/E_{on} and P_{off}/E_{off} separately, then take the natural log of their ratio; then divide by 2 * <IWF>.

This exact scheme was subsequently analyzed for the DQ-1 CO_2 lidar in Cao et al. (2024), which concluded that over uniformly-bright surfaces, AVS is the best, but over variable surfaces (either in topography or albedo), AVX is the the best (AVD was about the same as AVX). Based on the Cao et al. findings, this manuscript chose to simply adopt AVX as it seemed the most general and it simplified things. Ie, average AFTER the logarithm, rather

than before. This is strangely in contrast to the pioneering work of Tellier et al. (2018), which found AVS to *always* be the best scheme (and often by a substantial amount).

What causes this discrepancy? It appears that Cao et al. have made a critical mistake in their implementation of AVS. That paper's AVS divides the averaged signals by a uniformly averaged ⟨IWF⟩ (their Eq. 8), where Tellier et al. (2018, Table 1) requires IWF to be weighted by the offline calibrated signal. If you carefully evaluate the equation, the AVS numerator is intrinsically offline-weighted, so the mismatch generates a bias proportional to $\mathrm{Cov}(P_{off}/E_{off}, \mathrm{IWF})$. Because P_{off} varies from both surface reflectivity as well as surface elevation variations, this covariance vanishes over flat terrain and grows with surface variability, exactly the pattern Cao et al. report at one of their study regions (Ya'an) and attribute to an irreducible "new geometric error."

Against this, Tellier et al. find corrected AVS compliant at every reflectivity and topography tested; the present authors' own airborne experiment (Fan et al., 2024, *Appl. Opt.* **63**, 2121, Table 3) found AVS lowest in both scatter and bias and states that averaging before the logarithm "performs better"; and even Cao et al.'s flat-site values give biases of 0.18, 0.18 and −0.01 ppm for AVX, AVD and AVS, respectively.

Therefore, we strongly suggest that the authors construct the AVS average XCO_2 correctly (by constructed the proper weighted average of IWF, as done in Tellier et al.) and evaluate it for themselves. I suspect it may be *substantially better* than AVX, and therefore could improve their findings, perhaps dramatically so.

Secondary Issues (that will remain after the above are addressed)

I raise a number of secondary issues below, which I think may apply to any revised manuscript.

a. Definition and selection of continuous segments. "Continuous" is never defined: no gap tolerance is stated, and it is unclear whether isolated missing shots break a segment. Table 1's filter criteria are length thresholds only. The threshold was explicitly relaxed for ocean to preserve sample size but not elsewhere, with no explanation of the inconsistency. Any reanalysis should state the gap tolerance and constrain segments on surface and topographic homogeneity.

b. The Allan curve is smoothed by an unspecified method, and the raw series is not shown. The slope criterion $|\mathrm{slope}| < 0.005$ is applied to the smoothed curve, so every recommended averaging number depends on an undocumented operation. The unsmoothed series appears in the Fig. 7 legend but is not evident in either panel.

c. Effect of the quality control on the reported statistics. Removing the additional QC changes the global mean single-shot random error from 18.46 to 23.51 ppm (L250–255) — a 27 % change from a filtering step, reported in a single sentence. The monthly 5th–95th percentile trim discards a tenth of the distribution before the mean is taken, which is a strong operation on a quantity whose tails are physically meaningful. Therefore, the sensitivity of all headline numbers to these choices should be discussed.

d. Unstated quality-control parameters. The authors need to state and justify their combined SNR threshold (used to remove cloud contamination). It may be fine what they do, but it's not clear.

e. Detector gain state is never addressed. Zhu et al. (2021) document two instrument gain channels with four levels each, with linearity guaranteed only within a channel's design range. Gain switching makes the noise law piecewise, and is most likely to occur at exactly the sharp reflectance transitions the boundary scheme targets. Whether gain transitions occur within the boundary cases should be reported.

f. No external validation anywhere. All "goodness metrics: are internal — propagated error and Allan deviation. If the authors plan to make a useful product out of their ultimate averaging scheme, they need to actually validate against reference XCO_2 values such as from TCCON.

k. Editorial Two different papers are both cited as "(Fan et al., 2024)" at L55 and L65 and both appear in the reference list. Please add a/b suffixes so it's clear which paper is being cited.