

Reviewer:

This manuscript investigates the 2019 Brazilian fire season under non-extreme climate conditions using a two-step CTE-LW/SW inversion framework constrained by TROPOMI and MOPITT XCO observations. The topic is scientifically relevant and timely. The inversion design is carefully structured, and several sensitivity experiments are conducted to assess robustness. The study provides valuable insights into regional fire CO emissions and bottom-up inventory discrepancies. The manuscript is generally well organized and supported by data. However, several methodological aspects require further clarification and quantitative support before publication.

We thank the reviewer for their kind words and for taking their time to evaluate the manuscript. We address the suggested clarifications and corrections below (in blue text).

Major Comments*1. Identifiability of the Two-Step Inversion Framework*

The manuscript applies LW to optimize background CO and SW to optimize fire emissions. However: How does the LW step avoid absorbing part of the fire signal?

Thank you for this important question. As with any atmospheric CO inversion, we cannot absolutely guarantee that fire signals are not partially absorbed by other source terms; however, our framework is explicitly designed to minimise this aliasing through:

- *pre-selecting observations for assimilation:* The LW/SW is designed to decompose the inverse problem in two distinct spatio-temporal scales. In the long-window, we assimilate only background flask observations, which are rarely, if ever, near fresh fire plumes, and therefore not dominated by (Brazilian) fire-influenced CO signals. We acknowledge that background stations like ASC may pick up some fire signals every now and then from Africa; however, from the South American fire emission perspective, these would be part of the background signal.
- *LW state vector design:* The coarse and smooth spatial and temporal structure of the LW state poorly represents the sharp, episodic character of fire emission enhancements (and errors therein). Consequently, even if a residual fire signal were present in the background observations, the LW state vector has limited capacity to absorb it, and any aliasing would be confined to broad continental-to-latitudinal patterns averaged over seasonal timescales.

As a consequence of these choices, we largely rely on the prior fire emissions to capture inter-annual variability (IAV) of the fire signal. Even if we would alias some of the large/slow scale (continental, multi-month) variability in the fires, that is not our main concern, because our primary focus remains resolving the high-frequency temporal and fine-scale spatial variations of fire emissions for tropical South America's 2019 dry season in the SW step. Furthermore, for the case study presented we find that the background CO field in the domain is not strongly biased, so we can be confident that aliasing for this period/region is weak. We recognise that this assumption needs careful re-examination as we extend the framework (e.g. simultaneously cover multiple regions or apply it to different years) and consider this an important point of attention as we further develop/apply the framework.

Does LW optimization alter OH or NMVOC production patterns regionally?

The long-window inversion is given freedom to adjust the latitudinal structure of secondary CO production fields (including contributions from NMVOC oxidation), any scales more local than these are not explicitly scaled. Because OH concentrations directly govern the secondary CO production, any regional bias in OH will manifest as a bias in secondary CO production and will, to the extent that the observations inform on this, be aliased into the LW corrections of those fields. In practice, this means that the LW step can implicitly compensate for north-south gradients in secondary CO production and OH biases, even though OH itself is not an explicit state variable.

Could the SW step compensate for residual background bias?

We consider this risk to be limited, because compensation for residual background bias in the SW step would require that: (1) the bias is detectable in the set of super-observations assimilated in the short window; (2) those observations then must be within no more than ~6 days downwind of a fire source region; and (3) the residual background bias co-varies spatio-temporally with the fire emission signal. When there is no such covariance in background/fire residual, the SW state vector has no pathway through which to absorb this.

The manuscript mentions a remaining ~20 ppb positive bias in the free troposphere but argues it does not dominate SW results. This claim requires quantitative sensitivity tests. Please clarify information flow and provide supporting diagnostics.

Thank you for bringing up this point. Our statement was based on the finding that the SW ensemble spread in free-tropospheric observations was negligible. Upon reflection, we recognise that we did not address the possibility of a free-tropospheric bias projecting onto the fire emission estimates through our X_{CO} residuals ($y - H(x)$) in the SW assimilation step.

We conducted a sensitivity experiment (explained below) to quantify the potential impact of a free-tropospheric bias which demonstrates that this bias would have minimal impact on our reported fire emission estimates. Under a conservative scenario, we estimate the resulting change in fire emissions to be $\sim 10\%$ or ~ 4.7 TgCO. This falls well within the ± 15 TgCO structural uncertainty we report and would not alter our conclusions. We have clarified this information flow and its implication for the conclusions in Appendix A3.

Description sensitivity experiment:

We applied a uniform perturbation of -20 ppb to the TM5 model profiles between ~ 600 and ~ 150 hPa. Although the Manaus profiles reach until ~ 5 km, we assumed in this experiment that the bias persists to the upper troposphere as conservative upper bound. The perturbation was applied to all assimilated X_{CO} within 10° of the Manaus profiles (2.6°S , 60.2°W), in total > 40 k samples. Recalculating the X_{CO} (using the column averaging kernels) for TROPOMI, resulted in a change of about 10% in X_{CO} ($(\frac{X_{CO}^{FT-20} - X_{CO}^{base}}{X_{CO}^{base}}) \times 100$) as demonstrated in Figure 1 by their kernel density estimates that we report per $\sqrt{R}$ quantile. We found comparable numbers for MOPITT.

Assuming linearity of the observation operator $H(x)$ with respect to the state vector x , this 10% ΔX_{CO} translates to a 10% change in the posterior fire emission estimates. Given that the observed free-tropospheric bias was predominantly positive, this suggests that our posterior fire emissions may be underestimated by $\sim 10\%$ if such bias would be present over the whole domain, up to the tropopause, for the entire period.

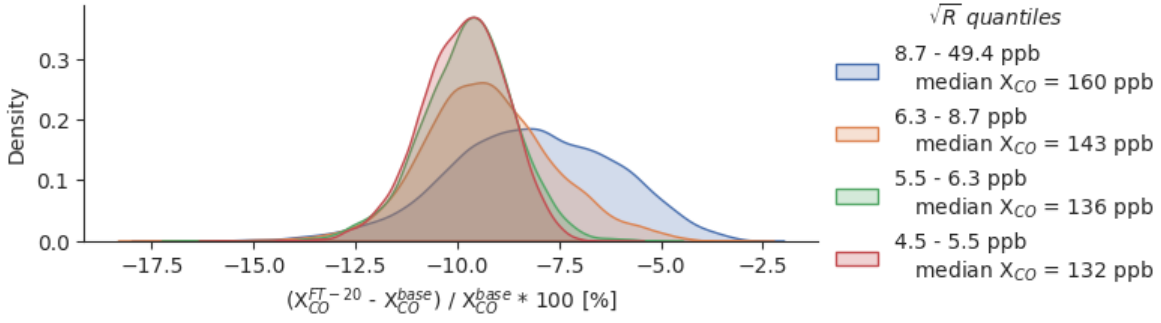


Figure 1: Kernel density estimates per $\sqrt{R}$ quantile with the change in X_{CO} (%) when subtracting 20 ppb for assimilated TROPOMI samples in the GFED5.1-TROPOMI inversion. The legend show the range of the $\sqrt{R}$ quantile and the median X_{CO} for that quantile (before subtracting the 20 ppb).

2. Justification of the $X_{CO} \geq 125$ ppb Filtering Threshold

The chosen threshold is used to isolate fire plumes, but: How sensitive are total emissions to this threshold? Could low-intensity or diffuse fires be excluded? Please provide emission estimates under different thresholds (e.g., 100, 125, 150 ppb). Also include statistics of excluded observations (< 125 ppb) to demonstrate that the filtering does not bias total fire emissions.

Thank you for the suggestion, we conducted an additional set of (GFED5.1 + TROPOMI) inversions with the 100 ppb and 150 ppb thresholds and updated Section 2.5.4 in the manuscript accordingly. Figure 2 demonstrates that our total emissions are not very sensitive to this threshold, even at the biome scale. The (max. - min.) range in posterior total Brazil fire emissions for the simulation period between the TROPOMI inversions with different thresholds (1.1 TgCO, or 2-3%, of the Brazil season total) is smaller when starting from different priors (1.7 TgCO). Between the different threshold runs, the Caatinga emissions are most affected in the 150 ppb threshold. This is logical as these columns usually are relatively clean and therefore get filtered out first, which limits the constraint on the Caatinga's fire emissions.

Please find the mean absolute errors of the excluded observations in Table A2 of the manuscript's Appendix A2 for 125 ppb threshold inversion and in Table 1 for the new runs. These indicate that the statistics for the match to non-assimilated obser-

vations does not degrade considerably after assimilation, consistent with minor changes in emissions.

Regarding low-intensity or diffuse fires, we agree these are likely under-constrained, as we discuss in Sections 4.2 and 4.3 of the manuscript. However, we hypothesise that the observation threshold is not the main limiting factor on constraining diffuse fire emissions, given that the inversion currently is limited in its capacity to scale these fires. This is because we apply multiplicative scaling factors with prior magnitude-dependent covariances, limiting the freedom to change emissions: zero/small prior emissions can only result in zero/small emission changes (original manuscript lines 478-484).

We updated Section 2.5.4 to clarify the impact of the threshold on the emissions.

Table 1: Mean Absolute Error over samples that were not selected for assimilation.

Experiment	n	MAE	
		Prior	Posterior
100 ppb	319730	5	6
125 ppb	497780	6	7
150 ppb	556250	7	7

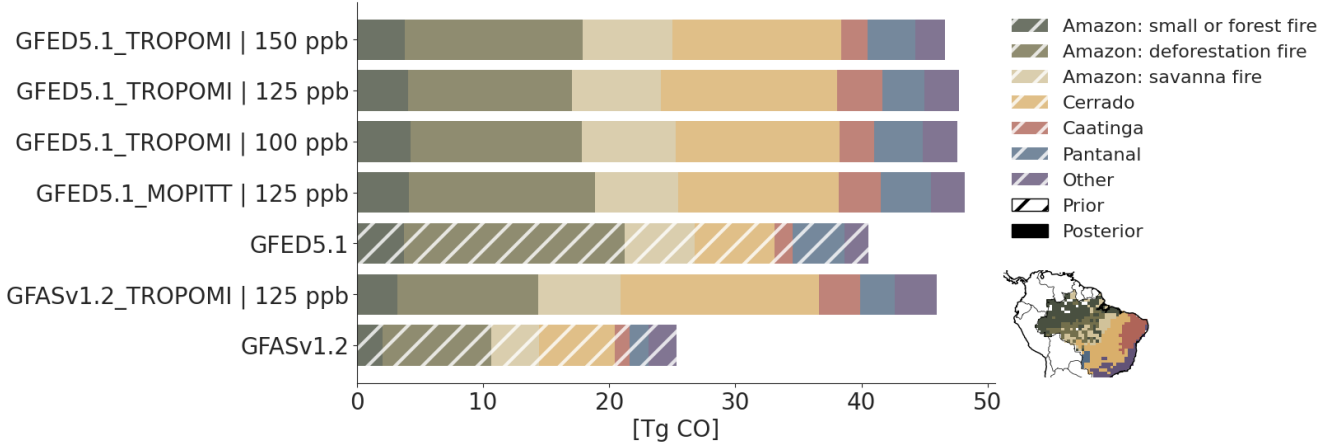


Figure 2: Prior (hatched) and posterior (solid) CO fire emissions summed over the simulation period (Jul 22 – Nov 30 2019 for Brazil, stacked by biome (indicated in the inset map). The figure is an extension of manuscript Fig. 4 with two sensitivity inversions using data selection thresholds of 100 and 150 ppb.

3. Observation Error Inflation Strategy

TROPOMI errors are inflated (transport error doubled + 4 ppb added) to achieve $\chi^2 \approx 1$. While common in data assimilation, further clarification is needed: Show χ^2 distributions before and after inflation (not only the mean). Explain how the “ $\times 2 + 4$ ppb” parameters were chosen. Provide sensitivity tests to alternative inflation factors. Importantly, TROPOMI native uncertainty (~ 3 ppb) is much smaller than MOPITT (~ 7 ppb). Inflation changes their relative weighting. Please discuss whether the reported convergence between TROPOMI and MOPITT inversions is partly influenced by this weighting adjustment.

We understand the reviewer’s reasoning regarding the potential impact of observation weighting. Isolating the impact of the observation error covariance choices (R) on flux convergence would require an extensive series of new inversions for both TROPOMI and MOPITT wherein we set different R choices, even then it would remain challenging to isolate the impact of R from intrinsic differences in the observations sets themselves on the flux convergence. Nonetheless, we agree that it could be a valuable exercise for future studies. For now, we clarify below why inflating R for the TROPOMI inversion was a necessary and logical step to facilitate a fair comparison with the MOPITT inversion in the context of this study.

An inversion must be balanced in terms of the actual model performance relative to observations ($y - H(x)$) and the anticipated errors ($HPH^T + R$). A poor balance could lead to over/under-fitting, make the inversion extra sensitive to noise in the obser-

variations, and possibly even lead to unstable solutions. Through either P or R the balance between prior weight vs. observation weight can be influenced.

We would like to underline that the R choice is not entirely free. Importantly, we need to consider here that R , besides instrument precision, must account also for forward model transport errors and the capacity of these models to represent the observations. As noted in Section 2.5.4 of the manuscript, TROPOMI and MOPITT "see" the atmosphere differently. TROPOMI's greater sensitivity to the lower troposphere exposes it to different transport modelling challenges that MOPITT largely avoids. Specifically, boundary-layer mixing dynamics, plume dispersion closer to the source, and turbulence-induced variability are features that our $1^\circ \times 1^\circ$ transport model (TM5) may not fully resolve or that are simply more difficult to capture. Assigning TROPOMI superobservation errors only, would underestimate the true model-data mismatch because it does not account for the aforementioned extra challenges of representing these observations.

Elaborating on Section 2.5.4, our inflation strategy consisted of:

1. *A uniform baseline structural error (+4 ppb)*: to account for the bulk of representation errors associated with boundary-layer dynamics across the entire domain.
2. *A plume dispersion error ($\times 2$ column gradient)*: this term scales dynamically, following the simulated column gradients, to represent plume dispersion uncertainty.

To address the reviewer's comment, we present sensitivity tests with various additive structural errors and a larger multiplicative dispersion error for R and illustrate how these could impact the distribution of χ_{inno}^2 that represents the actual vs. observation error balance (Fig. 3). Note that we take the prior models for "MOPITT" and "With inflation" from the GFED5.1.MOPITT and GFED5.1.TROPOMI inversions respectively. The additional sensitivity tests are calculated from the GFED5.1.TROPOMI prior model and not from new inversions. The Figure demonstrates that R inflation ("With inflation"; median 1.15 ppb) yields a more balanced χ_{inno}^2 than without ("Without inflation"; median 3.12 ppb). The remaining tests show the χ_{inno}^2 distributions for the individual components and that the combination of both is more effective at bringing the χ_{inno}^2 of the bulk towards one.

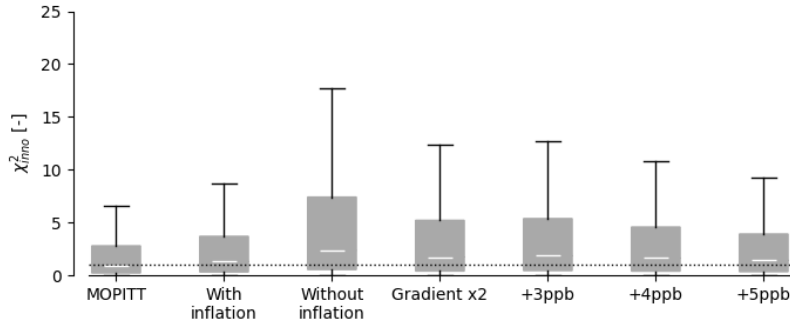


Figure 3: χ_{inno}^2 distributions comparison. The χ_{inno}^2 values were calculated using the prior model from the GFED5.1.MOPITT inversion (left-most box) and the GFED5.1.TROPOMI inversion (remaining boxes).

4. CO-to-C Conversion Assumption

CO emissions are converted to carbon emissions using a fixed CO/CO₂ ratio. However, emission factors vary by fuel type and biome. Please provide sensitivity analysis (e.g., $\pm 20\%$ in CO/CO₂ ratio). Quantify the resulting uncertainty in carbon estimates. Clarify the limitations of using CO-only inversion for total carbon quantification.

Thank you for this helpful suggestion. We re-calculated the total carbon emission (here: C from CO and CO₂) using 20% lower and higher $\frac{CO_2}{CO}$ and updated manuscript by including the following (lines 492-498):

To include the uncertainty arising from the assumed prior emission factor ratios, we performed a sensitivity analysis perturbing the assumed $\frac{CO_2}{CO}$ ratios by $\pm 20\%$ (Fig. A7). This widens the estimates by ~ 55 TgC in both directions and results in a range of 216–334 TgC. The downside of CO-only inversions for total carbon emission quantification is that we must rely on the accuracy of the emission factors that vary by fuel type, fuel moisture content, and meteorological conditions.

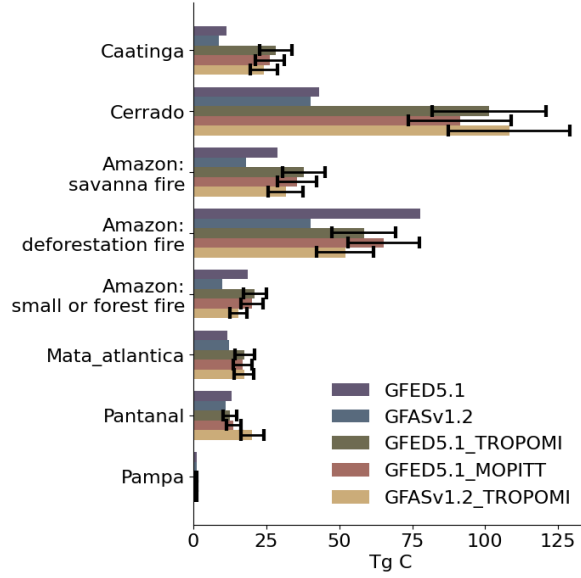


Figure 4: Carbon emissions (C from CO and CO₂) per biome. Posterior carbon emissions were inferred from CO using the prior inventory’s $\frac{\text{CO}_2}{\text{CO}}$ emission factor ratios at the location/time of the emission. The error bars represent the min/max range for a hypothetical 20% error margin on the CO₂/CO emission factor ratios. This Figure was added to the manuscript in Appendix A7.)

5. Underestimation over Cerrado/Caatinga

The manuscript reports posterior emissions nearly twice the inventory values in this region. While fuel pool explanations are discussed, the attribution remains qualitative.

Please consider: Showing spatial posterior/prior ratios. Providing innovation statistics by biome. Clarifying whether transport redistribution contributes to the regional signal.

Thank you for these helpful suggestions. We agree that the attribution of the large emission increments over the Cerrado and Caatinga remains largely qualitative. While providing a full quantitative attribution falls outside the scope of this manuscript, we have expanded our discussion (Section 4.2) building on your suggestion.

To address whether the emission increases over the Cerrado and Caatinga could be an artifact of transport redistribution (e.g., compensating for errors originating in other regions) we evaluated the spatial cross-correlations in our posterior emission ensemble. Figure 5 demonstrates that the posterior emissions for the Cerrado generally have very weak correlations with other biomes, confirming that the inversion is capable of disentangling these regional signals from one another.

We note two exceptions to the weak correlation: First, a temporary strong correlation with the Caatinga occurs in mid-October. This coincides with a shift from the typical south-east trade winds to more easterly winds which likely causes the Caatinga air masses to mix with Cerrado to the extent that the inversion cannot separate the signals. Second, we observe occasional anti-correlations with savanna-type fires within the Amazon biome. This is an encouraging result: while our prior covariance matrix (P) structurally links these fires to the rest of the Amazon biome, the assimilation of TROPOMI X_{CO} seems to provide sufficient independent signal to break this prior assumption, which dynamically links their behaviour with Cerrado fires instead.

Finally, per the reviewer’s suggestion, we computed the residuals and grouped these by biome (Table 2). We note that these should be interpreted with some caution, as grouping by underlying biome borders neglects transport. The table demonstrates that assimilation of TROPOMI reduces the mean absolute error across all major biomes, including a robust reduction over the Cerrado and Caatinga. This puts confidence in the inversion at the biome-level.

Table 2: Innovation statistics grouped by biome over assimilated observations. Including the mean absolute error (MAE) and mean uncertainty reduction (MUR) which is calculated as $\frac{1}{N} \sum_{i=1}^n \left(1 - \frac{\sigma_i^{posterior}}{\sigma_i^{prior}}\right)$, with σ representing the standard deviation over the respective ensemble of realisations for each X_{CO} sample. The nature prior represents a single tracer forward run from the long-window solution (at $1^\circ \times 1^\circ$), whereas the cycle prior is influenced by all previous short window cycles through mean state propagation. Based on the GFED5.1-TROPOMI inversion.

Biome	n ($\times 1000$)	MUR [-] (w.r.t. prior)	MAE [ppb]		
			(nature) Prior	(cycle) Prior	Posterior
Amazon	43.7	0.89	13.0	12.6	7.8
Caatinga	0.3	0.65	14.3	9.4	6.4
Cerrado	14.9	0.80	15.9	10.6	7.0
Atlantic Forest	3.4	0.81	12.6	11.0	8.0
Pantanal	2.4	0.82	19.7	16.7	10.7

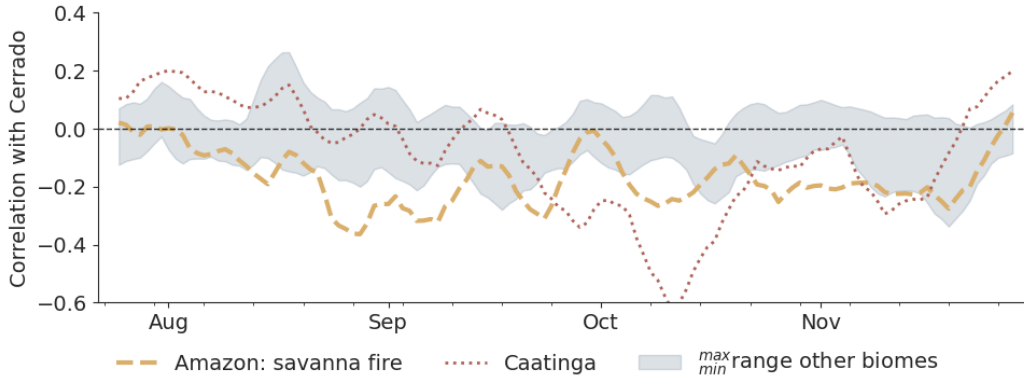


Figure 5: Six-day rolling average correlation between ensemble of posterior Cerrado emissions and that of other biomes. The shaded region represents the minimum-maximum range of correlations across other biomes/fire types in Brazil (incl. Amazon: other dominant fire type, Amazon: deforestation fire, Pantanal, Atlantic forest). Based on the GFED5.1-TROPOMI inversion.

Minor Comments

1. Clarify how the two-month MOPITT data gap in 2019 affects results.

We have added the following to Section 3.3 of the manuscript

Including this unconstrained period into our final multi-inversion posterior mean pulls the ensemble mean estimate toward the GFED5.1 baseline by up to 2.7 TgCO (difference between GFED5.1 and the mean of the two inventories for this period).

2. State earlier in the methods that the system cannot constrain super-local plumes due to resolution limits.

We included the following cautionary note in methods Section 2.4:

Note that our inversion system is not designed to resolve emissions from individual fire plumes, amongst others due to the $1^\circ \times 1^\circ$ resolution of our transport model and scaling factors, as well as biome level error correlations.

3. Clarify the origin of NMVOC and CH_4 pre-scaling factors.

The pre-scaling factors for secondary CO production from NMVOC and CH_4 oxidation were partly chosen to align our prior global budgets with established top-down estimates (e.g., Hooghiemstra et al., 2012; Zheng et al., 2019; Naus et al., 2022) and partly informed by visual inspection of prior and posterior CO mole fractions from initial forward runs. We have clarified this in the manuscript (Appendix A1).

4. In the conclusions, better define the applicable scale (regional totals vs. event-scale emissions) to avoid overgeneralization.

We appreciate the suggestion and revised the conclusion accordingly.

Technical / Presentation Issues

1.The “thick semi-transparent black line” in Figure 1 is difficult to identify; please adjust visibility.

We removed it to improve the figure’s readability, assuming that the term *arc of deforestation* is generally known.

2.Provide bias and RMSE metrics for ATTO validation.

We have updated manuscript Figure 3 to include the requested metrics and added the following to Section 3.1.2:

Although the short-window inversion yields improved validation scores for most cycles (Fig. 3), from time to time, deviations from this pattern are to be expected for a non-assimilated site, and they underline the importance of aggregating posterior results before interpretation. In particular, we find a local reduction in performance in the last cycle which can be attributed to Ensemble Kalman Smoother (EnKS) edge effects, where the final cycle is constrained by only a limited set of observations. When combined with state propagation and a poorly represented fire in close proximity to the ATTO site, this can affect performance metrics at a local scale. However, the influence of these edge effects on domain total emissions remains minimal, as is evident from Fig. 5.

3.Add y-axis labels in Figure 5 for clarity.

Thank you for the suggestion. Note that the y-axes are shared across the ridges and the ridges are equally spaced by $0.3 \text{ TgCO day}^{-1}$ to help interpretation. We have clarified this in the caption of the Figure.

References:

Hooghiemstra, P. B., M. C. Krol, P. Bergamaschi, et al. “Comparing Optimized CO Emission Estimates Using MOPITT or NOAA Surface Network Observations.” *Journal of Geophysical Research: Atmospheres* 117, no. D6 (2012). <https://doi.org/10.1029/2011JD017043>

Naus, S., L. G. Domingues, M. Krol, et al. “Sixteen Years of MOPITT Satellite Data Strongly Constrain Amazon CO Fire Emissions.” *Atmospheric Chemistry and Physics* 22, no. 22 (2022): 14735–50. <https://doi.org/10.5194/acp-22-14735-2022>

Zheng, B., F. Chevallier, Y. Yin, et al. “Global Atmospheric Carbon Monoxide Budget 2000–2017 Inferred from Multi-Species Atmospheric Inversions.” *Earth System Science Data* 11, no. 3 (2019): 1411–36. <https://doi.org/10.5194/essd-11-1411-2019>

Reviewer:

Constraining fire carbon emissions is essential for understanding the carbon budget and carbon cycle. This study presents satellite-based estimates of CO emissions from fires in Brazil using inverse modeling. Different inversion configurations lead to consistent posterior estimates, demonstrating the robustness of the framework. The manuscript is generally well-structured and clearly written. This study demonstrates physical consistency of fire emission monitoring through an innovative two-step inversion framework (long-window and short-window inversion), multi-source satellite data fusion, dynamic error modeling, and detailed biome-level analysis. It not only provides reliable emission estimates for the 2019 fires in Brazil but also offers important methodological references for the design of future fire monitoring systems, the improvement of emission inventories, and carbon cycle research. The manuscript is valuable and suitable for publication after moderate revisions.

We thank the reviewer for their thorough and encouraging assessment of the manuscript. We have carefully addressed the suggested revisions below (in blue text).

1. The entire framework hinges on the premise that the long-window (LW) inversion can effectively optimize non-fire sources (NMVOC oxidation, CH₄ oxidation, anthropogenic) to provide a pristine background for the short-window (SW) inversion, which then exclusively optimizes fire emissions. There is a risk of "aliasing," where fire signals not fully captured by the prior fire inventory could be absorbed by the LW-optimized terms (e.g., through adjustments to NMVOC production), or conversely, that residual errors in the LW background could be erroneously attributed to fires in the SW step. The authors should include a dedicated sensitivity analysis or a set of diagnostics to validate this decoupling.

Thank you for highlighting this important point. Indeed, source aliasing is always a risk in CO inversions. One that we, like any other inversion systems, cannot fully prevent. However, we explicitly designed our framework to mitigate this risk. Common inverse modelling practice to cover such risks, is to assign the prior uncertainties to the dominant term, and to make smart use of signals in observations. This is exactly what we try to do, and by splitting the spatio-temporal scales of our inverse problem into a LW and SW component, we can do this more strategically. We couple high spatio-temporal resolution CO residuals directly to fires, and the low/slow/smooth residuals to the "other" background terms, as successfully demonstrated also by Naus et al. (2022) who applied a similar split in a 4D-var setup".

Because the LW inversion structurally lacks the temporal and spatial freedom to fit sharp local fire enhancements, its ability to erroneously absorb daily fire variability into background NMVOC or CH₄ corrections is limited. Even if some of the weaker fire signals are aliased in the long-window, it is not a major concern since what we care about most is that the long-window step supplies mostly unbiased CO-levels in inflowing airmasses during the 2019 fire season. We extensively validated this, for instance in Figure 2, but also Figure A3 in the appendix demonstrates that the LW posterior residual distributions at Atlantic inflow stations (ASC, NAT) are narrow and more zero-centred compared to the prior.

2. What is the practical benefit of the LW inversion in this application? While correcting background concentrations is conceptually beneficial, Fig. 2b shows substantial background bias distributed across the entire domain rather than localized in fire regions, which may limit the effectiveness of the LW inversion. Please further illustrate the prior residuals before applying the LW inversion to clarify its contribution.

The reviewer remarks that the LW inversion should serve a more important purpose than just providing background CO for our study. Indeed, it does. The global CO budget is known to consist of many source and sink terms that vary in space and time over the past decade. No regional inversion can ignore these global fluxes and their trends and the best way to represent them is through inverse modelling, to ensure consistency with the flask records of observed CO mole fractions. Our group has applied this global inverse approach for nearly two decades (since Meirink et al. (2008) for CH₄ and Hooghiemstra et al. (2011) for CO). With the expansion of satellite data, higher spatio-temporal information became available and we targeted these scales in Naus et al. (2022) with a combination of 3-daily (fast, fires) to 3-monthly (slow, other sources) statevectors. The LW inversion with the EnKF presents a different way of solving the latter budget once again and in that sense is an optimal solution in its own, to be compared to earlier global CO budgets (as we did in Appendix A1). Its benefit is the computational efficiency through the fact that we can start (in parallel) SW optimizations from them, and the capacity to see yearly repeating patterns of flux updates over this period. That minimizes the risk of aliasing these slowly varying (annual, seasonal) changes into the SW inversions.

The requested prior and posterior statistics of our LW inversion to flask CO observations are provided in the Table below for a subset of tropical sites. They are comparable to those reported in Naus et al. (2022) and Hooghiemstra et al. (2012), as well as other global CO studies. Since both our posterior LW fluxes and posterior LW CO residuals are as expected, and within reported

numbers from earlier studies, we did not elaborate on them in the manuscript. This choice, along with a brief representation of the benefit of the LW system was added to Appendix A1, and we thank the reviewer for bringing the interest in this part of our solution to our attention.

Table 1: Long window prior and posterior mean mole fraction residuals ($y - H(x)$) and uncertainty reduction (UR) for the tropical stations. The UR is calculated as $\frac{1}{N} \sum_{i=1}^n \left(1 - \frac{\sigma_i^{\text{posterior}}}{\sigma_i^{\text{prior}}}\right)$, with σ representing the standard deviation over the respective ensemble of realisations for each CO sample.

Station	2014-2020			July-November 2019		
	Prior [ppb]	Posterior [ppb]	UR [-]	Prior [ppb]	Posterior [ppb]	UR [-]
ASC	13.33	4.91	0.78	12.10	2.30	0.75
ATT	10.63	7.70	0.77	14.94	9.12	0.71
CHR	-5.57	-6.75	0.76	-2.30	-5.52	0.73
CVO	-11.83	4.41	0.82	-0.22	12.32	0.79
GMI	-10.44	-1.60	0.80	-5.94	-3.98	0.76
MEX	-14.25	-4.67	0.80	-5.44	2.93	0.78
MLO	-8.23	2.81	0.83	-5.48	0.46	0.79
NAT	7.26	-0.12	0.78	2.06	-5.89	0.75
RPB	-2.66	8.22	0.80	-0.14	9.19	0.77
SEY	-2.22	-4.03	0.79	6.09	-1.12	0.75
SMO	2.19	-2.44	0.77	1.19	-2.76	0.74

3. MOPITT data are missing between 26 June and 24 August (L203), covering roughly one quarter of the inversion period. Since Fig. 5 shows that deforestation fire emissions peak in mid-August, this gap may introduce non-negligible bias. Notably, the posterior mean is averaged across three inversions, whereas GFED5.1.MOPITT represents the prior before 24 August only. This could shift the posterior mean of deforestation fires toward the GFED5.1 prior (see Fig. A4). Please clarify this issue.

Thank you for pointing this out. This is indeed correct: in the absence of MOPITT observations between 26 June and 24 August, the GFED5.1.MOPITT inversion defaults to the unscaled GFED5.1 prior and because our reported posterior mean does not include an unconstrained GFAS v1.2 inversion, the posterior mean pulls the ensemble mean slightly toward the GFED5.1 baseline during the peak fire season.

To evaluate the magnitude of this effect, we compared the Brazil total prior emissions during this specific data gap, which account for 13.6 TgCO in GFED5.1 and 8.1 TgCO in GFAS v1.2. The mean of these two priors (10.9 TgCO) is 2.7 TgCO lower than GFED5.1 alone. If we hypothetically substitute the neutral GFED/GFAS prior mean for the GFED5.1.MOPITT inversion during the missing data period (or alternatively ignore this period altogether), the total seasonal posterior estimates across the three inversion configurations actually converge *even more tightly*. Therefore, while this data gap introduces some weighting toward the GFED5.1 prior, it does not change our conclusions.

We added the following to Section 3.3 (~line 344):

Including this unconstrained period into our final multi-inversion posterior mean pulls the ensemble mean estimate toward the GFED5.1 baseline by up to 2.7 TgCO (difference between GFED5.1 and the mean of the two inventories for this period).

4. 2.5.1: How are observations within the 3-day assimilation window treated, given that data are available at roughly three time steps?

To clarify our approach: we did not average or aggregate the observations over the three-day window. Instead, all individual satellite overpasses occurring within the three-day cycle are assimilated simultaneously in a single batch optimization step. Specifically, for each individual satellite superobservation or retrieval, we sample an ensemble of mole fraction and pressure profiles from the TM5 transport model at the time and location of the overpass, apply averaging kernels to calculate the X_{CO} (Section 2.3), and assign R (Section 2.5.4).

The optimizer then ingests this entire set (three-day ensemble) of model-observation pairs to derive one single set of posterior flux scaling factors per three days. These optimized scaling factors are then applied uniformly across the underlying prior fire

emissions for that specific three-day block.

5. Line 248: An outlier threshold of three is standard in many inversion systems. However, for rapidly varying fire emissions, large model-data mismatches at individual timesteps may result in the rejection of valid observations. The missing peaks in Fig. 3 suggest this may be occurring. Please discuss the potential impact of the filtering strategy.

We agree with the reviewer that a strict outlier threshold risks rejecting valid (high-concentration) observations, particularly for rapidly varying fire plumes. However, in our framework, this risk is substantially mitigated because our assigned model-data mismatch uncertainty (R) is dynamic (Sect. 2.5.4). Because we scale R dynamically with the simulated local spatial gradient, the total model-data mismatch (MDM) is proportionally larger around plumes. Consequently, the absolute rejection threshold (for residuals $3 \times \text{MDM}$) widens for most fire plumes, which prevents valid high-concentration observations from being rejected. As shown in Figure 1, the overall rejection rate remains relatively low during active fire cycles with over 1,000 selected samples, typically well below 10%.

Nevertheless, it is not unlikely that we reject some valid observations. We can speculate whether this systematically biased our emission estimates by evaluating the prior residuals of the rejected observations. We find that these are binomially distributed around zero, with a median of -0.25 ppb, which is not suggestive of a systematic bias in our peak emission estimates. We agree that exploring more sophisticated outlier detection methods that better account for the dynamic nature of fire plumes are worth investigating.

We have added the following paragraph to the manuscript’s discussion Section 4.3:

In addition, we acknowledge that our $3 \times R$ rejection strategy, while standard practice, may have led to the exclusion of some valid high-concentration observations. As a result, the most extreme fire enhancements that were simulated and not observed, or vice versa, may be partially filtered out. The prior residuals of these rejected samples do not suggest that we systematically under- or overestimate the posterior fire emissions. Nonetheless, we may have rejected some valid data and we recommend future studies to consider exploring more sophisticated, possibly non-Gaussian outlier detection methods that are better suited to process fire plumes.

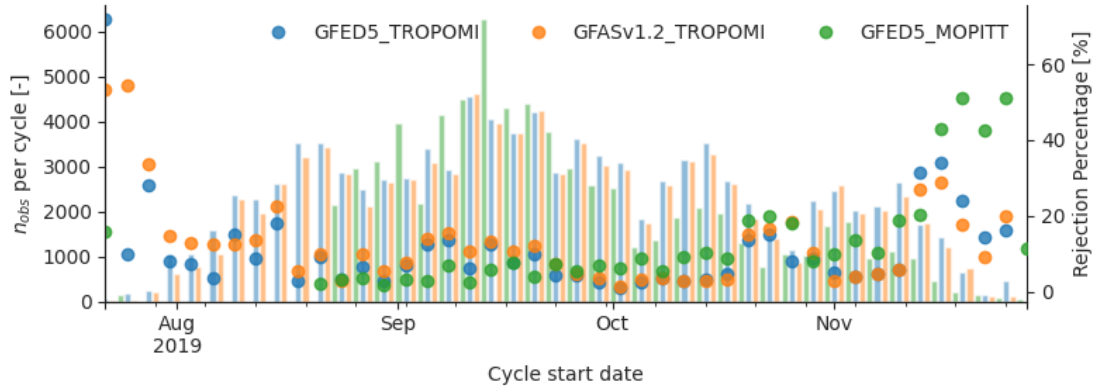


Figure 1: Total number of observations (left axis, bars) per short-window cycle and the percentage of rejected observations (right axis, dots).

6. It would be helpful to include the spatial pattern of prior emissions to facilitate comparison between emission and concentration patterns.

Thank you for the suggestion. We have added the spatial pattern of the prior emissions (GFED5.1) as panel to manuscript Figure 2.

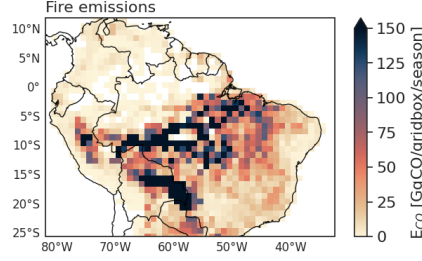


Figure 2: Included as panel to manuscript Figure 2

7. Lines 404–405: Fig. 2d does not present an observation-simulation comparison. Fig. 2e shows residuals, but residuals over B1 remain substantial. Please clarify or correct the description.

Thank you for noticing this typo, we referred to the wrong sub-figure indeed. We would like to clarify that the main inflow into the domain is represented by box B2 more than B1 during the main fire period. Besides, during the main fire period, the residuals remain relatively small (between -5 and 5 ppb). We revised the text in Section 4.2 to clarify this point and now refer to the correct sub figure.

8. The inversion domain extends beyond Brazil (Fig. 1), yet only Brazilian results are discussed. Given the significant CO fire signals in surrounding regions (Fig. 2a), how do emission adjustments in neighboring countries influence the Brazilian estimates? We thank the reviewer for raising this point. We focused our narrative on Brazil because most of the 2019 fire CO emissions were emitted here. Specifically, Brazilian fires accounted for 73% of the total (statevector) domain emissions in the GFED5.1 prior and 82% in our GFED5.1-TROPOMI posterior (see also manuscript Fig. 4 and Fig. 3 added below).

To test whether emission adjustments in neighbouring countries artificially inflate or deflate the Brazilian estimates, we analysed the posterior temporal correlations between the total non-Brazilian emissions and the emissions from individual Brazilian biomes (Fig. 4).

The results do not show strong (anti-) correlations between non-Brazilian regions and the Brazilian biomes. This lack of correlation demonstrates that observations we assimilate in the short-window inversion can disentangle the different signals. In other words, the emission adjustments in Brazil do not merely act as a spatial compensation mechanism for neighbouring countries.

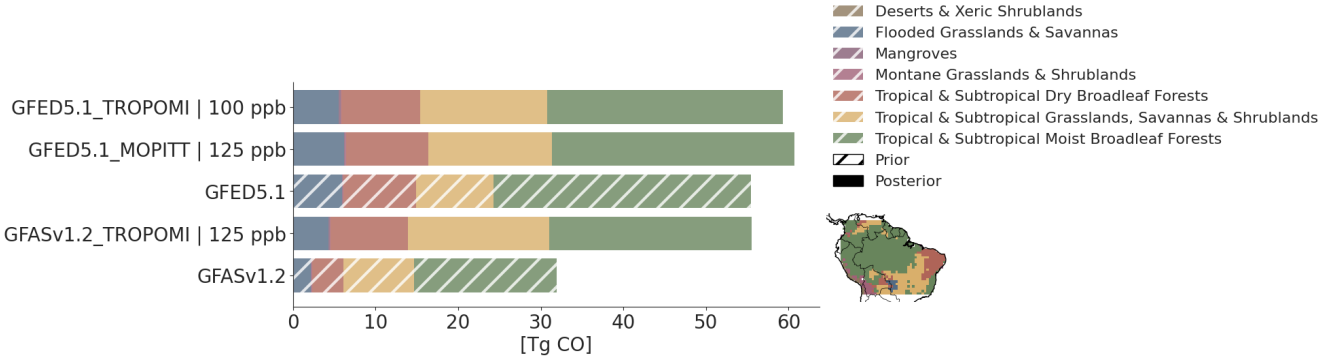


Figure 3: Statevector domain total emissions according to GFED5.1, GFAS v1.2 and the posteriors, following manuscript Fig. 1 definitions.

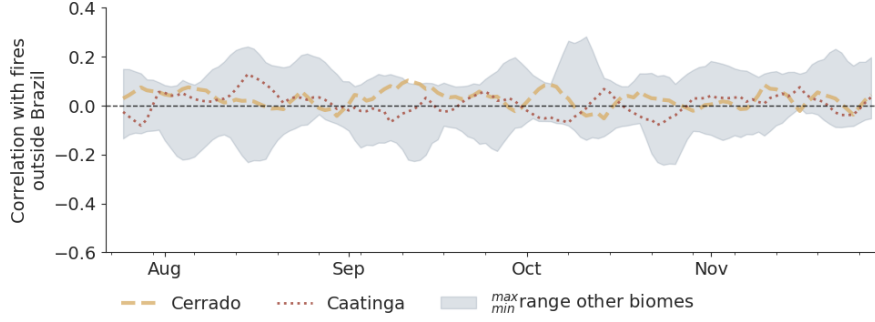


Figure 4: Six-day rolling average correlation between the ensemble of posterior emissions for fires outside Brazil with other biomes. The shaded region represents the minimum-maximum range of correlations across the set of not explicitly indicated biomes (labelled as "other biomes"). Rest-of-domain represents the fires outside Brazil, but inside the statevector domain (visualised in Fig. 1 of the manuscript). Based on the GFED5.1-TROPOMI inversion.

9. Relatedly, could biases in emissions from surrounding biomes contribute to the inferred increases in the Cerrado and Caatinga? For example, Fig. 2b shows substantial underestimation over deforestation and Pantanal regions, which would likely drive emission increases during inversion. Although a biome-dependent covariance structure may limit such cross-region influence, the Amazon savanna fire region is not separated in Fig. 1.

To test the impact of this, we repeated the exercise we did for the previous comment, but this time with respect to the Cerrado emissions. The figure demonstrates generally weak correlations with other biomes, including the Amazon and Pantanal fires (Fig. 5). This suggests that observations can disentangle the emission signals from the different regions well and the inferred increases in the Cerrado are not driven by biases in emissions from surrounding biomes.

Note the two exceptions: 1) a temporary strong positive correlation with the Caatinga occurred in mid-October. This coincides with a shift from the typical south-east trade winds to more easterly winds, likely causing Caatinga air masses to mix with the Cerrado to an extent that limits the inversion's ability to separate the signals. 2) We observe occasional anti-correlations with savanna-type fires from within the Amazon biome. This is actually encouraging, since our prior covariance matrix ($\mathbf{P}$) structurally links these fires to the rest of the Amazon biome, but the assimilation of X_{CO} provides sufficient independent information to break this prior assumption, which dynamically links their behaviour with the Cerrado fires instead.

Based on the reviewer's points 8 and 9, we elaborated Section 4.2 and included Appendix A6.

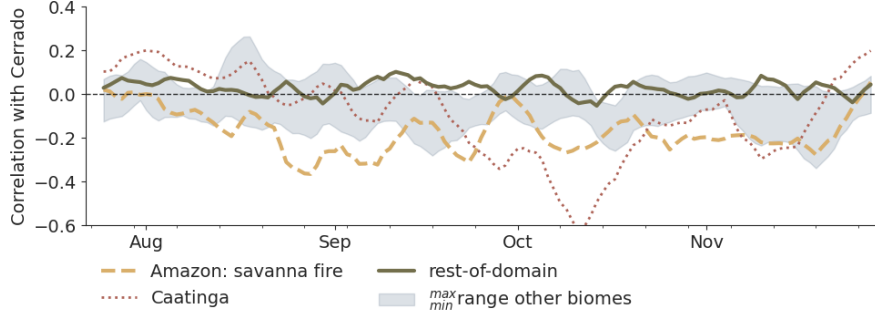


Figure 5: Six-day rolling average correlation between the ensemble of posterior emissions for Cerrado fires with other biomes. The shaded region represents the minimum-maximum range of correlations across the set of not explicitly indicated biomes (labelled as "other biomes"). Rest-of-domain represents the fires outside Brazil, but inside the statevector domain (visualised in Fig. 1 of the manuscript). Based on the GFED5.1-TROPOMI inversion.

10. The authors do a good job explaining the inflation of the model-data mismatch to achieve a χ^2 near 1. It would be useful to state the mean χ^2 value afterinflation, not just that it "approached one." Was it 1.05? 0.98? Providing the number adds

precision.

Thank you for the suggestion, we have included it in the text and updated our text **as follows**:

*After inflation, the **bulk of the χ^2_{inno} approached one (mean 3.75; median 1.51)**, indicating a well-balanced and statistically consistent error characterisation.*

Note that we cannot calculate the exact mean χ^2_{inno} from before inflation without starting a full formal simulation, because each cycle influences the next (o.a. through mean state propagation, Section 2.4). However, neglecting the sequential nature of the EnKS and taking the prior model of the current inversion (in this example GFED5.1.TROPOMI), we can approximate it, which gives χ^2_{inno} mean and median values of 8.77 and 3.12 respectively (before inflation).

11. Author presents a clear mass balance equation. I may miss it somewhere, the cost of function and configuration of the prior error covariance matrix B is not clearly described. Given that the specification of B can significantly influence the posterior results, the authors should provide a detailed explanation of its construction (e.g., variances and auto-correlations).

We did not include an explicit cost function description, as we did not deviate from the core system described in Peters et al. (2005); we have added an explicit reference indicating that the cost function formulation can be found there [added to manuscript Section 2.4 Line 152]. Regarding the construction of the prior error covariance matrix of the short-window inversion, we describe this in detail in Section 2.4 of the manuscript. We appreciate that the reviewer points out the variance, which we indeed forgot to explicitly mention [added to manuscript Section 2.4 at Line 160]. For the long-window inversion, please find the covariance description in Appendix Table A1.

References:

Hooghiemstra, P. B., M. C. Krol, J. F. Meirink, et al. “Optimizing Global CO Emission Estimates Using a Four-Dimensional Variational Data Assimilation System and Surface Network Observations.” *Atmospheric Chemistry and Physics* 11, no. 10 (2011): 4705–23. <https://doi.org/10.5194/acp-11-4705-2011>

Hooghiemstra, P. B., M. C. Krol, P. Bergamaschi, et al. “Comparing Optimized CO Emission Estimates Using MOPITT or NOAA Surface Network Observations.” *Journal of Geophysical Research: Atmospheres* 117, no. D6 (2012). <https://doi.org/10.1029/2011JD017043>

Meirink, J. F., P. Bergamaschi, and M. C. Krol. “Four-Dimensional Variational Data Assimilation for Inverse Modelling of Atmospheric Methane Emissions: Method and Comparison with Synthesis Inversion.” *Atmospheric Chemistry and Physics* 8, no. 21 (2008): 6341–53. <https://doi.org/10.5194/acp-8-6341-2008>.

Naus, S., L. G. Domingues, M. Krol, et al. “Sixteen Years of MOPITT Satellite Data Strongly Constrain Amazon CO Fire Emissions.” *Atmospheric Chemistry and Physics* 22, no. 22 (2022): 14735–50. <https://doi.org/10.5194/acp-22-14735-2022>

Peters, W., J. B. Miller, J. Whitaker, et al. “An Ensemble Data Assimilation System to Estimate CO₂ Surface Fluxes from Atmospheric Trace Gas Observations.” *Journal of Geophysical Research: Atmospheres* 110, no. D24 (2005). <https://doi.org/10.1029/2005JD006157>