Response to RC2:

This study analyzes CO_2 and CO concentration variations over 1 year and 9 months at three sites in Guangzhou—a coastal megacity—examining their relationship with land–sea breezes. Using backward trajectory footprint modeling, it quantifies fossil fuel and biogenic contributions. Given the scarcity of direct CO_2 observations in this region, these findings offer valuable insights into carbon sources/sinks in coastal southern Chinese megacities. The methodology is overall sound, and the work merits publication in Atmospheric Chemistry and Physics. However, the introduction of research background, uncertainty analysis in source partition, the robustness of the results, and the depth of discussion could be further improved:

Response:

We thank the reviewer for the thorough and constructive assessment of our manuscript. We are grateful for the reviewer's supportive comments and the encouraging recommendation for publication. We have carefully considered all suggestions and revised the manuscript accordingly to address the main concerns raised. In particular, we have:

1) strengthened the introduction and research motivation;
2) clarified and expanded the treatment of uncertainty in the source-partitioning analysis;
3) improved the robustness presentation and deepened the interpretation and discussion of the results.

We believe these revisions have substantially improved the clarity, rigor, and overall quality of the manuscript. Below we provide point-by-point responses and indicate the exact locations where the corresponding revisions were implemented in the revised manuscript. The revised text has been highlighted in the manuscript for ease of reference.

Specific comments:

Introduction Section: The introduction could benefit from restructuring to enhance its logical flow. The rationale for reporting urban-scale CO_2 concentrations—particularly the need to clarify carbon sources and sinks—should be more explicitly emphasized. While the study relies heavily on the CO_2/CO ratio and footprint modeling, there is limited discussion on the inherent uncertainties of these methods or how they compare with alternative approaches, such as 3-D atmospheric inversions. Introducing these

methodological considerations would strengthen the foundation for the work.

Response:

We appreciate the reviewer's helpful suggestion. In the revised Introduction, we implemented a targeted restructuring to improve the logical flow by (i) strengthening the rationale for using urban-scale atmospheric CO_2 concentrations while emphasizing the source–sink attribution challenge in coastal megacities, (ii) adding a concise methodological context (including 3-D inversions) to clarify how our approach complements existing methods while acknowledging their respective sensitivities/limitations, (iii) explicitly stating the inherent uncertainties of the CO_2/CO (R_{CO}) approach and footprint-informed transport, and (iv) adjusting the placement of our framework so that the "gap–site–objectives" narrative is more cohesive.

1) Clearer rationale for reporting urban-scale CO_2 concentrations and need for source–sink clarification. We added a clearer motivation that concentration observations provide an integrated constraint (anthropogenic emissions + biogenic exchange + transport) but inherently introduce attribution ambiguity. We highlighted why this is especially acute in coastal settings due to marine background inflow and diurnal transport/boundary-layer reversals (Introduction; Lines 40–52). For consistency, we also adjusted the Abstract opening to foreground the mitigation relevance and the coastal attribution challenge (Lines 15–17).

2) Comparison with alternative approaches, including 3-D inversions. We added a structured paragraph describing how 3-D inversions can estimate city-scale emissions, while acknowledging practical limitations such as sub-grid representativeness errors and sensitivities to biogenic priors. Accordingly, we position our observation-driven framework as a process-level complement to these inversion approaches (Introduction; Lines 54–71; Lines 119–125). This complementarity is reiterated in the Conclusions (Implications): while our framework does not produce posterior flux fields as in a formal Bayesian inversion, it provides an observation-driven tool for rapid process attribution and consistency checking in coastal urban carbon monitoring and mitigation assessment (Lines 770–773).

3) Explicit uncertainties in CO_2/CO (R_{CO}) and footprint modeling. We explicitly acknowledged that CO-based inference carries uncertainties (e.g., variable R_{CO} and chemical processing) and that footprint estimates are sensitive to meteorological forcing. we also clarified our evaluation strategy (Introduction; Lines 125–132). These limitations are further described in the Methods uncertainty section (Sect. 2.5.2; Lines 312–315) and summarized in the Conclusions (Lines 756–757).

4) Improved narrative flow. To enhance logical progression, we repositioned the description of our framework to follow the explicit gap statements and the justification of Guangzhou as a "living laboratory". This ensures the "solution" is introduced only after the challenges and site context are established (Introduction; Lines 107–132).

Section 2.5: The a priori emission inventory utilized in the study should be better specified, which inventory? What spatial resolution for what year?

Response:

Thank you for the comment. We agree that the wording in Sect. 2.5 could be read as implying that an a priori emissions inventory is required. We therefore revised Sect. 2.5 to state explicitly that our observation-driven framework does not assimilate or rely on any a priori inventory to derive CO_2tot/CO_2ff/CO_2bio; it uses only concentration enhancements at the receptor sites together with transport-model footprints (Sect. 2.5; Lines 246–253). We also clarify that emission inventories are mentioned only for contextual description (Sect. 2.1) and for an independent bottom-up comparison (Sect. 3.5), not as priors or constraints in the partitioning.

Because inventories are cited elsewhere in the manuscript, we now explicitly specify the requested details—dataset name, year, and spatial resolution—at the points where each inventory is introduced: the contextual description in Sect. 2.1 (Lines 158–159) and the independent bottom-up comparison in Sect. 3.5 (Lines 672–676). In addition, to ensure consistent messaging and avoid any remaining ambiguity, we made brief alignment edits in the Abstract (Lines 17–20) and Conclusions (Lines 718–721) noting that the partitioning is performed without assimilating emission inventories.

Line 219: for footprint simulation, using 500 particles over a 72-hour simulation period appears rather small. In this setting, the particle count per time step and per grid point is extremely low, potentially introducing substantial uncertainty in footprint estimation. It would be valuable to test and report the sensitivity of results to a larger number of particles (1000-10000) to ensure robustness.

Response:

Thank you for highlighting the need to document sensitivity to STILT particle number. We agree that particle number can influence footprint estimates and thus inferred fluxes. In response, we added a targeted winter-afternoon sensitivity analysis at PY to assess robustness to STILT setup choices. Starting

from our baseline configuration (500 particles, 0.08° grid, 72 h backward), we increased particle number to 1000 and 2000 particles and, in parallel, tested two additional key settings (grid spacing: 0.05° and 0.10°; backward duration: 96 h and 120 h). For each variant, we recomputed footprints, reran the full flux-estimation workflow, and compared CO_2tot/CO_2ff/CO_2bio against the baseline using paired daily afternoon means (12:00–16:00 LT), reporting effect sizes and 95 % confidence intervals (CIs) as primary quantities.

The test design is described in Sect. 2.5.1 (Lines 294–302), the statistical treatment in Sect. 2.5.2 (Lines 333–335), and results in Sect. 3.5 (Lines 641–664) and summarized in Fig. 11 (new) and Tables S6–S7 (new).

The results show that inferred winter fluxes are robust to increased particle number. Relative to the baseline, increasing particles to 1000/2000 changes CO_2ff by −0.56 %/−0.24 % and CO_2tot by −0.52 %/−0.31 % (paired daily afternoons), indicating near-converged behavior for winter afternoons. This is supported by extremely high baseline–variant correlations ($r \geq 0.999$) and small RMSE values (0.28–0.45 μmol m^{-2} s^{-1}; Fig. 11; Table S6). CO_2bio shows similarly robust behavior: because wintertime CO_2bio is near zero at PY, we assess it in absolute terms, and the test–baseline differences remain small with 95 % CIs generally spanning zero (Table S7). Across all tested settings, changes remain small; only the intentionally coarser 0.10° grid yields a statistically detectable but still minor decrease (CO_2ff: −1.47 %, $p = 0.0269$; CO_2tot: −1.40 %, $p = 0.0164$), while all other variants show changes ≤1.34 % with 95 % CIs spanning zero ($p \geq 0.083$). The across-run day-by-day ensemble spread is also tightly bounded (median 0.20–0.21 μmol m^{-2} s^{-1}; median CV ≈ 1.8 %), and paired-day scatter remains close to 1:1. Overall, these additions indicate that the inferred winter fluxes are not artifacts of particle number or reasonable transport-parameter choices. Given the near-converged behavior at 1000–2000 particles, extending to much larger particle numbers is unlikely to materially change the inferred fluxes within our uncertainty context and would substantially increase computational cost. We also made corresponding wording updates in the Abstract and Conclusions to reflect this robustness check (Lines 27–28; Lines 743–746).

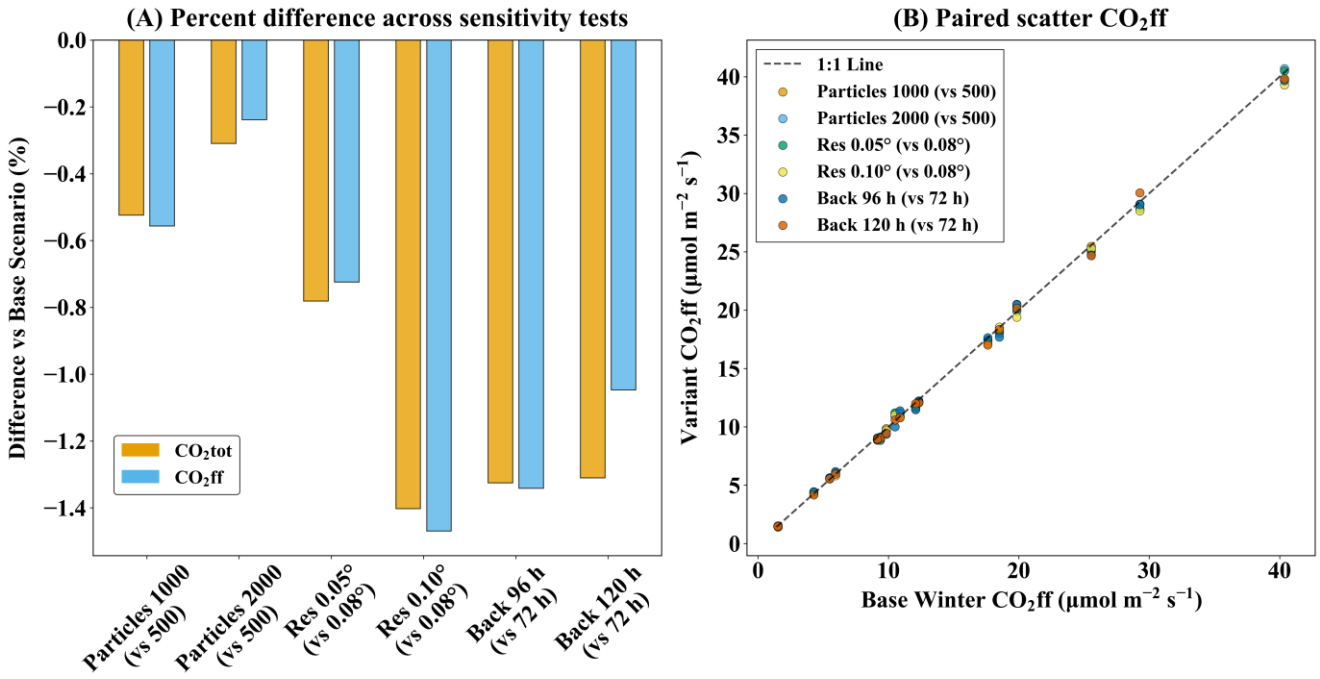

Figure 11. STILT parameter sensitivity at PY (winter). Panel A: mean percent difference (variant − baseline) of inferred fluxes relative to the winter baseline (500 particles, 0.08°, 72 h), computed from paired daily afternoon means (12:00–16:00 LT; n = 18); Δ% = (variant − base)/base × 100; negative values indicate lower than baseline. Panel B: paired scatter of CO_2ff (µmol m^{-2} s^{-1}) from each variant versus the baseline for the same days; solid line is 1:1 (y = x).

Table S6. Wintertime (12:00–16:00 LT) paired-day sensitivity of PY inferred fluxes to STILT parameter choices (n = 18). Variants (particle number, grid spacing, backward duration) are compared with the baseline (500 particles, 72 h, 0.08°). Metrics report effect size (pct_diff_% and 95 % CI), day-to-day consistency (Pearson r), RMSE (µmol m^{-2} s^{-1}), and detectability (paired t-test p value). Upper block: CO_2ff; lower block: CO_2tot. Across the baseline plus six variants, the day-by-day ensemble spread—computed as the standard deviation across the seven runs for each day and then summarized by the median—was 0.20–0.21 µmol m^{-2} s^{-1} (median CV ≈ 1.8%).

metric	comparison	pct_diff_%	pearson_r	rmse	p_value	ci95_lo	ci95_hi
PY_CO_2ff	Back 120 h vs Base (72 h)	-1.05	0.9994	0.38	0.1136	-0.3216	0.0376
	Back 96 h vs Base (72 h)	-1.34	0.9993	0.45	0.0831	-0.3904	0.0265
	Particles 1000 vs Base (500)	-0.56	0.9996	0.33	0.3410	-0.2380	0.0871
	Particles 2000 vs Base (500)	-0.24	0.9995	0.31	0.6738	-0.1916	0.1269
	Res 0.05° vs Base (0.08°)	-0.72	0.9996	0.31	0.1917	-0.2506	0.0542
	Res 0.10° vs Base (0.08°)	-1.47	0.9997	0.39	0.0269	-0.3729	-0.0257
PY_CO_2tot	Back 120 h vs Base (72 h)	-1.31	0.9992	0.44	0.0862	-0.3819	0.0280
	Back 96 h vs Base (72 h)	-1.33	0.9992	0.45	0.0916	-0.3900	0.0322

	Particles 1000 vs Base (500)	-0.52	0.9996	0.28	0.2983	-0.2098	0.0683
	Particles 2000 vs Base (500)	-0.31	0.9995	0.29	0.5502	-0.1864	0.1029
	Res 0.05° vs Base (0.08°)	-0.78	0.9996	0.31	0.1527	-0.2541	0.0431
	Res 0.10° vs Base (0.08°)	-1.40	0.9997	0.35	0.0164	-0.3394	-0.0393

Table S7. Wintertime (12:00–16:00 LT) paired-day sensitivity of PY CO_2bio inferred fluxes to STILT parameter choices (n = 18). Variants (particle number, grid spacing, backward duration) are compared with the baseline (500 particles, 72 h, 0.08°). We report the absolute paired-day test–baseline difference, defined as ΔCO_2bio = CO_2bio(variant) − CO_2bio(baseline), summarized by the paired-day mean (Δ; μmol m^{-2} s^{-1}) and its 95% confidence interval (CI95_lo, CI95_hi; μmol m^{-2} s^{-1}). Because wintertime CO_2bio at PY is close to zero, percent differences are not shown. Across-run daily spread of CO_2bio—defined as the day-by-day standard deviation across the baseline and all variants—has median 0.045 and IQR 0.016–0.067 μmol m^{-2} s^{-1}.

metric	comparison	mean (ΔCO_2bio)	ci95_lo	ci95_hi
PY_CO_2bio	Back 120 h vs Base (72 h)	−0.035	−0.094	0.024
	Back 96 h vs Base (72 h)	0.003	−0.057	0.063
	Particles 1000 vs Base (500)	0.005	−0.031	0.041
	Particles 2000 vs Base (500)	−0.009	−0.052	0.033
	Res 0.05° vs Base (0.08°)	−0.007	−0.047	0.033
	Res 0.10° vs Base (0.08°)	0.010	−0.043	0.063

Line 225: The statement that "footprint uncertainties are neglected under the assumption of unbiased atmospheric transport" seems to lack justification. Could the authors provide references supporting this choice or acknowledge its limitations?

Response:

We appreciate the reviewer's careful comment and agree that atmospheric transport (footprints) is an important uncertainty source and cannot be assumed error-free. Our original wording was not intended to claim "unbiased transport" but rather to indicate that we do not explicitly propagate footprint/transport uncertainty within the analytical error-propagation formulas used for Eqs. (12)–(14). We have revised the text to avoid any implication of unbiased transport, to clarify that transport uncertainty is not explicitly quantified in the analytical uncertainty terms, and to explicitly acknowledge this as a limitation.

Within a feasible scope, we additionally assess transport-model setup sensitivity by introducing a winter paired-day STILT sensitivity analysis at PY. In this test, we perturb key STILT configuration choices (particle number, grid resolution, and backward duration), recompute footprints, rerun the full flux-estimation workflow, and evaluate baseline–variant differences using effect sizes (percent differences), correlation (Pearson r), and paired statistical tests; the quantitative outcomes are reported in Sect. 3.5 and Fig. 11.

Finally, we now state explicitly that residual transport biases (e.g., winds and boundary-layer mixing) remain unquantified and may bias the inferred fluxes, thereby contributing to inventory–observation differences when benchmarking against bottom-up inventories, alongside emission-inventory uncertainty and the representativeness mismatch between footprint-weighted enhancements and unweighted inventory means. These revisions are implemented in Sect. 2.5.2 (Lines 306–312). This limitation is further discussed in Sect. 3.5 (Lines 668–670; Lines 678–680; Lines 686–696) and summarized in the Conclusions (Lines 754–756).

Fig 3A: Consider revising the x-axis label to "January–December" for clarity.

Response:

We appreciate the reviewer's suggestion. We have revised this (formerly Fig. 3) by replacing the numeric month ticks ("01–12") with month abbreviations ("Jan–Dec") to improve readability and clarity (Line 398).

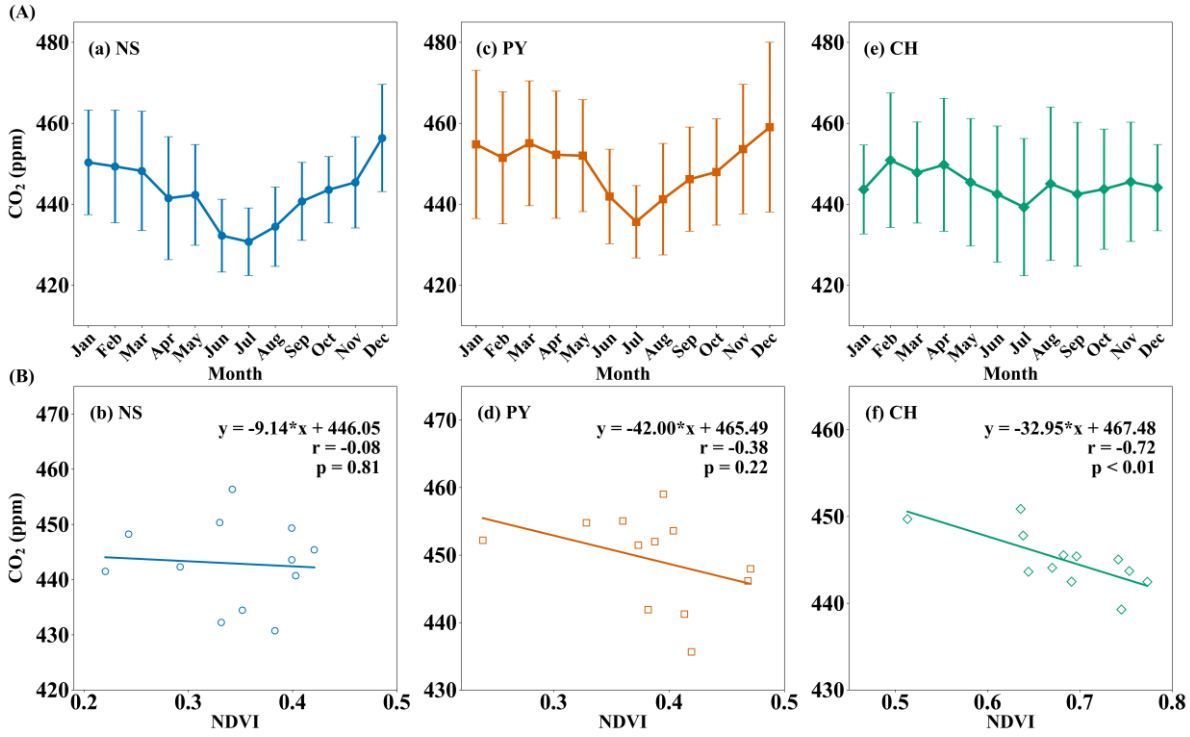

Figure 4. (A) Variations in monthly mean CO_2 concentrations and (B) their correlations with Normalized Difference Vegetation Index (NDVI) for the (a–b) NS, (c–d) PY, and (e–f) CH stations. Error bars indicate ± 1 standard deviation (SD).

Lines 255–265: The urban–suburban differences highlighted here are insightful. Adding a figure to illustrate this spatial comparison would be valuable. Discussion of seasonal variations in this gradient would further enrich the analysis.

Response:

We appreciate the reviewer's constructive comment. To better visualize the urban–suburban spatial contrast and its seasonality, we added a new figure (Fig. 3) and expanded the discussion in Sect. 3.1 (Lines 355–357; Lines 368–385). Figure 3 summarizes the annual and seasonal mean gradients PY–NS and PY–CH (error bars: ±1 SE). The PY–NS gradient remains positive year-round but is smallest in winter (2.23 ppm), consistent with prevailing northerlies elevating CO_2 at the coastal NS site (Fig. 2). In contrast, PY–CH is more strongly seasonally modulated, peaking in winter (8.83 ppm) but reversing sign in summer (−2.86 ppm), consistent with more frequent southerly (marine-influenced) ventilation at PY (SW/SE sectors in Fig. 2) and seasonally enhanced CO_2 at CH linked to transport, biogenic and boundary-layer processes. Overall, these results highlight a seasonally displaced apparent CO_2 "dome" in this coastal setting, where the network maximum can occur outside the urban core; this coastal-specific feature becomes evident only when the spatial gradient is examined seasonally.

In addition, we integrated the gradient results into the broader mechanistic interpretation through targeted cross-references. In Sect. 3.1.1 (Lines 408–410), we clarify that the month-to-month south–north CO_2 pattern is consistent with the seasonal gradients in Fig. 3, including the winter weakening of the urban–coastal contrast and the summertime displacement of the urban–suburban gradient. We also added a cross-reference in Sect. 3.3 (Lines 536–538) linking the suburban summertime CO_2 maximum to the combined influence of coastal transport, biogenic exchange, and boundary-layer mixing, consistent with the seasonal gradient shift summarized in Fig. 3. For consistency, we aligned the Abstract and Conclusions with this new result (Lines 20–23; Lines 727–731).

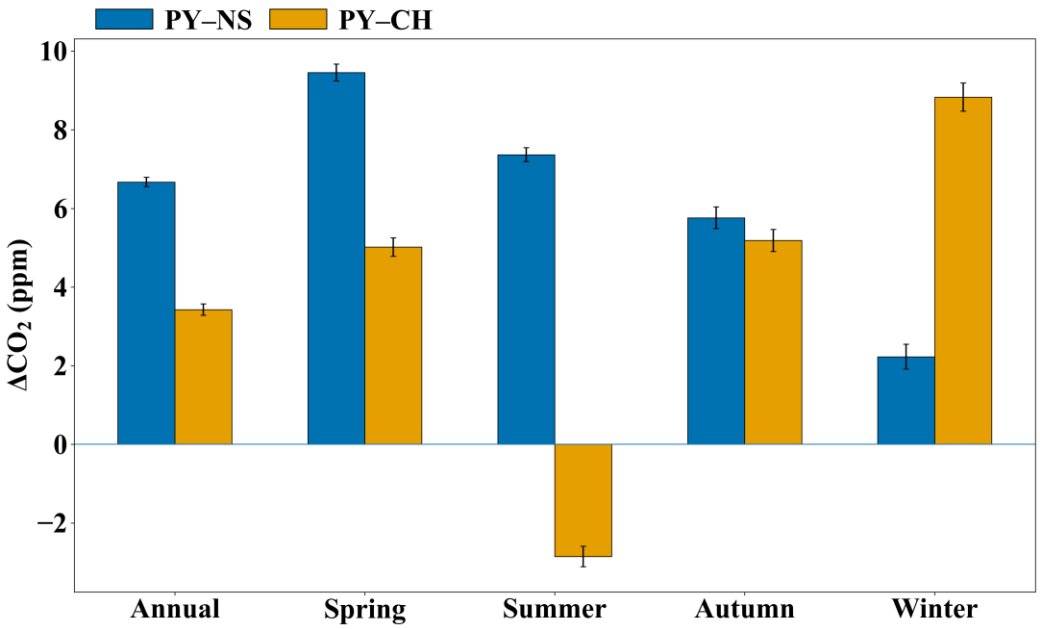

Figure 3. Urban–suburban/coastal CO_2 gradients in Guangzhou. Annual and seasonal mean concentration differences (ΔCO_2, ppm) between the urban site (PY) and the coastal site (NS) (PY–NS; blue bars) and between PY and the suburban forest site (CH) (PY–CH; orange bars). Seasons are defined as Spring (Mar–May), Summer (Jun–Aug), Autumn (Sep–Nov), and Winter (Dec–Feb). Error bars denote ± 1 standard error (SE).

Line 337: "with coastal > urban > suburban impacts" is not clear, please rephrase.

Response:

We thank the reviewer for pointing out this unclear phrasing. Here "coastal > urban > suburban impacts" refers to the magnitude of the SLBD-related CO_2 reduction (i.e., the ventilation/dispersion effect) across the three sites. We have rephrased the sentence to explicitly state that the SLBD-related CO_2 decrease is largest at the coastal site, intermediate at the urban site, and smallest at the suburban site (NS > PY > CH). We have revised this in Section 3.2 (Lines 481–484).

Line 425-429. This sentence is a little bit long and difficult to understand. Please rephrase and clarify.

Response:

We agree and have rewritten the sentence into several shorter sentences to improve clarity and readability. We have revised this in Section 3.5 (Lines 601–608).

Section 3.5: The reliability of the source analysis needs further strengthening. The current results are highly dependent on the accuracy of the prior emission inventory. However, significant discrepancies exist in fossil fuel emission estimates across different inventories, and emissions vary considerably

between years. Furthermore, the biogenic contribution is derived by subtracting the fossil fuel estimates. Therefore, it is strongly recommended that the authors assess the impact of using different inventories on the conclusions.

Response:

Thanks for this important point. We agree that bottom-up fossil-fuel CO_2 inventories can differ substantially across products and years. We therefore clarify that our CO_2ff/CO_2bio partitioning is inventory-independent: CO_2tot is inferred from observed concentration enhancements together with footprint-informed transport diagnostics, CO_2ff is derived from observed CO enhancements using the PY-specific R_{CO} relationship (determined from observed $\Delta CO/\Delta CO_2$ regression slopes), and CO_2bio is diagnosed as the residual (CO_2bio = CO_2tot − CO_2ff). Emission inventories are not used as priors or constraints in the partitioning, but only for independent context and benchmarking. Consequently, inventory choice does not affect the inferred CO_2tot/CO_2ff/CO_2bio or our main conclusions; it only affects the bottom-up benchmark range used for contextual comparison.

To address the reviewer's recommendation, we added a like-for-like cross-inventory benchmarking using same-year (2023) inventories at 0.1° × 0.1°, temporally disaggregated with the same hourly profiles, and compared winter-afternoon (12:00–16:00 LT) mean emissions aggregated over the winter footprint-defined sensitivity region (Fig. 12). This yields a wide inter-inventory spread (e.g., EDGAR vs. MEIC), which we now interpret explicitly as a plausibility envelope rather than a one-to-one validation target; the spread likely reflects differences in spatial disaggregation and point-source treatment that can be amplified at sub-provincial scales (Sect. 3.5; Lines 671–680).

We further clarify why inventory-based means and our observation-based estimates are not expected to match one-to-one: our CO_2ff reflects a footprint-weighted receptor enhancement at PY (after background removal), whereas inventories provide unweighted gridded emission fields from which a domain-mean flux is computed over the sensitivity region. Because STILT footprints impose heterogeneous, transport-dependent weights and only grid cells with substantial influence contribute materially to the receptor signal (Fig. 12), the footprint-weighted effective mean can differ from the unweighted inventory mean, depending on the overlap between footprint patterns and spatially heterogeneous emissions (including hotspots) (Sect. 3.5; Lines 686–699). Regarding the residual formulation, we clarify that CO_2bio is computed as (CO_2tot − CO_2ff), but CO_2ff is not taken from inventories; it is inferred from observed CO using the site-specific R_{CO} relationship. We explicitly

acknowledge that CO_2bio inherits uncertainty from both CO_2tot and CO_2ff (Sect. 2.5.2; Lines 315–318; Sect. 3.5; Lines 616–623) and report bootstrap 95 % CIs for the summertime biogenic offset (Sect. 3.5; Lines 705–708). We further place the summertime offset in the context of independent coastal-urban estimates from the literature as an external consistency check (Sect. 3.5; Lines 709–715). As an additional robustness check, paired-day STILT configuration sensitivity tests show only small changes (≤1.47 %) in inferred winter-afternoon CO_2tot/CO_2ff across reasonable transport choices (Fig. 11; Tables S6–S7), and measurement/background-selection uncertainties contribute only a small fraction of the winter-afternoon CO_2ff error budget (~3 %) (Sect. 3.5; Lines 641–668). These reliability assessments are summarized in the Conclusions (Lines 743–751).