

Response to Referee Comments (RC1)

“Urban atmospheric CO₂ plumes – Part 2: Representation of fine-scale structures”

Manuscript egusphere-2026-2581 — Response to Referee #2 (RC1)

We thank the referee for a thorough and constructive review. We accept the central message of the review: the manuscript, particularly the CO₂/XCO₂ results, discussion and conclusions (Sects. 3.4–3.9, 4, 5), was too long, contained repetitive and sometimes unsupported statements, and did not consistently link CO₂ performance to meteorological performance. Below, each comment is summarized (in italics) and followed by our response, in the order of the review: general comments (including the modeling-approach and CO₂-analysis points), specific comments and technical corrections. Changes in the revised manuscript are shown in red italics.

Legend

Implemented — the change described is present in the revised manuscript.

Numbering: line numbers in the comments refer to the discussion paper (egusphere-2026-2581). Section numbers are unchanged. Figure and table numbers in our responses refer to the revised manuscript unless stated otherwise; main changes in numbering: the emission map and station map (former Fig. 2) and the PSD figure (former Fig. 3) moved to Appendices A and B, so former Figs. 4–12 and 14 are now Figs. 2–11; the E_{sym} maps (former Figs. 13 and 15) are now in Appendices I and J; a new Table 3 gives the CO₂ error statistics, and the former Table 3 is now Table 4.

General Comments

General Comment 1 — Manuscript length, repetition, and lack of focus in the CO₂ analysis (Sects. 3.4–3.9, 4, 5)

Comment: *This study investigates the performance of weather and CO₂ simulations based on the WRF model with a large dataset of meteorological and CO₂ observations, focused on the influence of horizontal model resolution. The manuscript is much too long and for many sections it is unclear what the key messages are, making it very hard to follow; many sections contain repetitive statements and concluding statements not always central to the topic. This is a problem mostly for the CO₂ analysis (Sects. 3.4–3.9, 4, 5), less so for the meteorology sections (3.1–3.3). The referee suggests a fresh start of the CO₂ results, discussion and conclusions, organised around: (1) identifying the performance metrics relevant to inverse modeling (mean CO₂ bias was omitted; absolute rather than relative differences were emphasised); (2) evaluating meteorology and CO₂ consistently so CO₂ performance can be linked to meteorology performance; (3) presenting only the results needed to demonstrate the key messages, since most of the qualitative Sects. 3.4–3.7 could be removed; (4) thoroughly linking CO₂ performance to meteorology performance (currently PBLH is over-cited, wind is neglected, as are flux and boundary-condition accuracy); and (5) keeping the inverse-modeling goal in view throughout. The referee estimates the points could be made in about half the length and offers to review a revised, significantly shorter version.*

Response: We have followed this five-point strategy. (1) Mean bias is now reported and discussed for all CO₂ comparisons: a new Table 3 gives R, MBE, MAE and RMSE against the Picarro and mid-cost networks, and an MBE row has been added to the native-versus-remapped table (now Table 4). Relative differences (E_{sym}) are discussed alongside the absolute metrics, including their different seasonal and diurnal behaviour and its relevance for flux estimation (Sect. 3.9.1, Discussion). (2) Wherever a CO₂ or XCO₂ result is interpreted, the corresponding wind-speed, wind-direction and PBLH results for the same resolution and season are now referenced (Sects. 3.6, 3.8.1, 3.9.1, 3.9.2); where a link cannot be established with our analyses, we say so rather than assert a mechanism. (3) The qualitative Sects. 3.4–3.7 were shortened substantially: non-quantified case-study narration was removed or moved to a clearly labelled illustrative appendix (Appendix E), Sect. 3.5 was rewritten around the condensed paragraph proposed by the referee, and the remaining text of Sects. 3.6–3.7 now gives numbers (peak values, bias magnitudes, timing errors). Sects. 3.8–3.9 were shortened by removing repeated conclusions and now quote the numbers from the tables. (4) The Discussion was rewritten and now links the CO₂ results to wind, vertical mixing, PBLH, and emission and boundary-condition uncertainty. (5) Statements that did not serve this goal, including unsupported causal claims and terminology misuses identified in the specific comments (“aggregation error”, “propagation”, the recommendation to abandon high-

resolution wind modelling), were removed or reworded. Overall, the Results and the Discussion/Conclusion are shorter than in the submitted version (part of the removed material now in appendices).

Status: Implemented

General Comment 2 — Nudging asymmetry between the mesoscale and LES domains (a major point)

Comment: A key difference between the simulations is that the mesoscale (900 m) domain was nudged towards observations using FDDA, while the LES domains (300 m, 100 m) were not. The referee suspects this “free-running WRF” configuration of the LES domains could explain the higher wind speed and direction errors there relative to 900 m, and could matter for CO₂ performance as well; this is neither acknowledged nor discussed in the submitted manuscript, which instead recommends abandoning high-resolution wind simulations in the conclusions, a recommendation the referee does not agree with given this confound, and instead attributes performance differences to other factors (PBLH), risking misinterpretation. The referee understands that nudging was deliberately withheld in the LES domains to avoid distorting resolved turbulence, but is unsure this concern is fully justified, since observation nudging can be set to act smoothly over wider areas; alternatively, nudging restricted to the driving meteorology above the PBL would avoid disturbing boundary-layer turbulence. Absent new simulations, the referee asks that the nudging asymmetry be incorporated into the interpretation of meteorology and CO₂ results throughout.

Response: We agree that this is a genuine limitation of the submitted study. FDDA observation nudging was applied to D01-D03 (including the 900 m domain) but not to the LES domains D04 (300 m) and D05 (100 m), and the submitted manuscript neither flagged this as a possible explanation of the wind (and hence CO₂) differences nor took it into account in its conclusions. Our choice was deliberate: nudging tendencies act directly on the resolved wind, temperature and moisture fields, and inside an explicit-LES domain they would damp or distort the turbulent eddies that LES is run to represent. The referee’s suggestion of nudging toward the driving analysis restricted to levels above the PBL would avoid most of this problem, but it would require a new set of multi-week LES simulations that was not feasible within this revision. We have therefore followed the referee’s alternative and incorporated the nudging asymmetry into the interpretation throughout: it is described in Sect. 2.1; the wind results (Sects. 3.1-3.2) are presented as the combined effect of resolution and nudging, citing Ho et al. (2024); the Discussion and Conclusion state that the 900 m wind advantage cannot be attributed to resolution alone and should not be read as evidence that high-resolution meteorology is intrinsically less accurate. The recommendation to abandon high-resolution wind simulations has been removed and replaced by a recommendation to repeat the comparison with a consistent nudging strategy across domains (Conclusion, future work).

Status: Implemented

General Comment 3 — The 300 m simulation and the Shin-Hong PBL scheme

Comment: Doubrawa and Muñoz-Esparza (2020) found that 333 m is too coarse for an LES simulation and recommend the Shin-Hong PBL scheme instead. Here, the 300 m simulation performs worst against observations, suggesting their finding may apply here and that 300 m could perform better run with the Shin-Hong scheme rather than in LES mode.

Response: We agree that this reference is directly relevant to the 300 m degradation and consistent with the gray-zone (“terra incognita”) regime. We did not test the Shin-Hong scheme at 300 m in this study. We now cite Doubrawa and Muñoz-Esparza (2020) in Sects. 3.1 and 4 as a likely explanation, and recommend testing a scale-aware PBL scheme (e.g. Shin-Hong) at 300 m as future work (Conclusion).

Status: Implemented

General Comment 4 — Temporal variability of anthropogenic CO₂ emissions

Comment: Apparently, the anthropogenic CO₂ emissions used as input were constant in time. In reality, emissions vary in time, which could contribute to model-observation biases; temporal profiles (diurnal, weekly, seasonal) approximating this variability are available in the literature and should be taken into account.

Response: On checking this against our own modeling framework (shared with Part 1 of this study), we can clarify that this was a wording ambiguity in the submitted text rather than an actual simplification in the simulations: as in Part 1, anthropogenic CO₂ emissions are not time-invariant. Annual emissions are mapped onto the WRF-Chem grid and temporally disaggregated to hourly values using TNO/Airparif sector-specific profiles (diurnal, weekly and monthly/seasonal), consistent with the literature-based temporal-profile approach the referee recommends. We have reworded the Model Input Data section to state this explicitly, so it can no longer be read as a single time-invariant scaling factor. The genuine, more limited caveat, stated identically in Part 1, is that these sector-level profiles are applied uniformly across all road segments within a sector and do not distinguish inbound/outbound or radial/orbital traffic direction, which can introduce phase errors of up to ~1 h at stations dominated by a strongly directional road (e.g. the périphérique in the Paris case), though the impact is expected to be smaller elsewhere; nor do these profiles capture facility-specific operational variability (e.g. a point source temporarily offline for maintenance).

Status: Implemented

General Comment 5 — Night-time XCO₂ and TCCON averaging kernels

Comment: It is unclear how model XCO₂ at night was obtained. The text mentions TCCON averaging kernels, but these are not defined for night-time observations. If anything, averaging kernels of an active sensor such as MERLIN would make sense for night-time results.

Response: This is a genuine methodological gap in the original submission that was not adequately documented. On review, the night-time XCO₂ fields used in the native-versus-remapped comparison (Sect. 3.9.2) were computed as a simple pressure-weighted vertical average of the WRF CO₂ profile, without an averaging-kernel, since no TCCON retrieval (and hence no kernel) exists at night; only the explicit TCCON model-observation comparison (Sect. 3.7), which is restricted to daytime, used the TCCON kernel and prior. The original manuscript did not state this distinction. We have corrected the Methodology to state the two distinct XCO₂ definitions explicitly, and note, as the referee suggests, that a genuine night-time column comparison against real retrievals would require an active-sensor instrument (e.g. MERLIN, DQ-1) with its own averaging kernel, which is beyond the scope of the present ground-based TCCON evaluation.

Status: Implemented

General Comment 6 — Repeated “winter higher than summer because of lower PBLH” statement

Comment: The topic of the study is the impact of resolution on performance, but the CO₂/XCO₂ texts often compare summer against winter, and several sections conclude that concentrations are higher in winter because of lower PBLH. This does not indicate which simulation performs best and should be stated once.

Response: Agreed. The winter/summer contrast in concentration amplitude is now stated once, in the Discussion (third paragraph of Sect. 4), where it is contrasted with the opposite seasonal behaviour of the relative differences. In the Results, the seasonal comparison is limited to what is needed to compare resolutions.

Status: Implemented

General Comment 7 — Wind, flux and boundary-condition differences neglected in favour of PBLH

Comment: A major shortcoming of the CO₂ discussions is that they often fail to account for wind speed differences among resolutions, seasons and time of day, as well as flux and time-varying boundary-condition differences (at least biogenic fluxes must have a temporal resolution); instead only PBLH is often cited, risking misinterpretation.

Response: Agreed. Wind is now considered wherever CO₂ differences are interpreted: the larger wind-direction errors of the LES runs in the CO₂ time series (Sect. 3.6); the wind-speed biases in relation to the CO₂ mean biases and PDFs (Sect. 3.8.1); wind speed and vertical transport, in addition to PBLH, in the day/night differences between native and remapped fields (Sect. 3.9.1) and for XCO₂ (Sect. 3.9.2). The emission inventory (sector-level temporal profiles, surface release) and the CAMS boundary conditions, identical at all resolutions, are now listed as candidate contributors to the common low CO₂ bias (Sect. 3.8.1, Discussion).

Status: Implemented

General Comment 8 — Emphasis on relative metrics (E_{sym} , mean bias in %) over absolute concentration differences

Comment: The manuscript repeatedly states that shallow boundary layers lead to higher discrepancies because of higher signals. In the referee's view, relative error is more informative because it relates more directly to flux-estimation error than concentrations themselves; a relative metric such as E_{sym} or mean bias in percent would be more informative and should be stressed more. There is some E_{sym} analysis, but only to compare model resolutions to each other, not models to data.

Response: We agree and have adjusted the framing (Sect. 3.9, Discussion). Regarding E_{sym} : in our study it compares native and remapped simulations with each other. No independent observations exist at the remapped scale, so this particular comparison cannot be made against data; this is now stated explicitly in Sect. 2.4. For the model-observation comparison, we added the mean bias (new Table 3), which gives an observation-referenced view of the systematic error relevant for flux estimation. We also revised our interpretation of the relative metric. The referee is right (see also comments L720-723 and L799-801) that a weaker signal does not by itself produce a larger E_{sym} if the differences scale with the signal. The Results now state that absolute differences are larger in winter and at night, whereas E_{sym} is similar by day and by night and about twice as large in summer. The inter-simulation differences therefore decrease less than the enhancement itself from winter to summer. As flux errors scale with relative errors, the summer case is not less demanding than the winter case (Sect. 3.9.1, Discussion).

Status: Implemented

General Comment 9 — Mean bias not discussed despite its relevance to flux-estimation error

Comment: The accuracy of flux estimation is the key motivation for this study, but mean biases are not discussed for CO₂/XCO₂, even though several figures indicate mean biases both between models and data (Fig. 9) and among model resolutions (Figs. 12, 14). Mean bias should be analysed to enhance the value of the study.

Response: Agreed. A new table (Table 3) gives R, MBE, MAE and RMSE of CO₂ against the Picarro and mid-cost networks by resolution and season, analogous to the meteorological Table 2. It shows a negative mean bias at all resolutions (-1.7 to -6.6 ppmv), discussed in Sect. 3.8.1 and in the Discussion and Conclusion. An MBE row (native minus remapped) was also added to the native-versus-remapped table (now Table 4); it quantifies the day/night sign change of the difference between native and remapped fields discussed in Sect. 3.9.1.

Status: Implemented

General Comment 10 — Missed opportunity to link CO₂ PDF performance to wind/PBLH performance

Comment: The authors miss the opportunity to link modeled meteorology performance to modeled CO₂ performance. Suggestion: investigate the relationship between the CO₂ PDF and wind/PBLH performance. The PDF peak is too low compared to observations (less so for 900 m in winter, Fig. 10a-d); wind is always overestimated

except for 900 m in winter (Table 2), perhaps this contributes to the winter CO₂ underestimation? This does not translate to summer, where PDF differences among simulations are small despite similar wind bias differences, perhaps because PBLH biases differ by season? At the one station shown, 900 m PBLH is always higher than 300/100 m (Fig. 5), which would predict underestimated enhancements at 900 m, contrary to Fig. 10; perhaps broadening at lower resolution counteracts this by increasing the chance of intersecting point-source plumes.

Response: We thank the referee for this worked example, which we followed. Using the new Table 3, we find that in winter the configuration with the largest wind-speed overestimation (300 m, +1.7 m/s) has the most negative CO₂ bias in both networks, while the nearly unbiased 900 m winds go with the PDF mode closest to the observed one. This is consistent with the referee's suggestion of stronger ventilation in the LES runs. As the referee anticipated, the link does not hold in summer, where CO₂ biases are almost identical across resolutions despite a similar contrast in wind bias. On PBLH: the 900 m run has the highest PBLH, which alone would give lower 900 m concentrations. The cross-sections (the referee's comment on L352-392) show that near-source plumes remain much shallower at 900 m than in the LES, i.e. weaker near-source vertical transport by the MYJ closure than by resolved updrafts. Together with the weaker 900 m winds, this may compensate the PBLH effect; the broader 900 m plumes, which intersect more stations (as the referee suggests), may also contribute. The same combination explains why the native 900 m field is higher than the remapped LES fields in winter daytime but lower at night (Sect. 3.9.1). These links are stated in Sects. 3.8.1, 3.9.1 and the Discussion as consistent, qualitative explanations; we note explicitly that we did not quantify the relative contributions of wind, vertical mixing and PBLH.

Status: Implemented

General Comment 11 — Sections 3.4–3.7: qualitative, repetitive, and not always about the paper's actual topic

Comment: Sections 3.4–3.7 contain qualitative case-study descriptions but lack quantitative analysis; 3.4 and 3.7 in particular are often vague and the referee does not agree with all statements. They are not always about simulation performance but often about differences between cases (summer/winter), which contributes little to the manuscript's actual topic. The referee's specific comments often ask for numbers, but suggests not expanding these sections to answer them; rather, remove sections that only repeat conclusions, or significantly shorten and refocus them on quantitative analysis.

Response: Agreed (see General Comment 1, point 3, and the specific comments on Sects. 3.4-3.7 below). Rather than answering each descriptive point in place, we shortened these sections, replaced qualitative statements by numbers where the figures support them, and moved the remaining illustrative case description to Appendix E.

Status: Implemented

General Comment 12 — Sections 3.8 (statistics) and 3.9 (aggregation): figures good, text long and sometimes inaccurate

Comment: For Sect. 3.8 (statistical analyses) and 3.9 (aggregation), the figures are informative but the text is lengthy, sometimes vague, and does not always accurately describe what is shown in the figures. The referee makes concrete suggestions below, but overall the text could and should be much shorter.

Response: Agreed. Both sections were shortened and now quote numbers from Tables 3 and 4. Each inaccuracy identified by the referee (figure-reference mix-ups, the numeric mismatch between text and Table 3, internally inconsistent statements) is corrected; see the specific comments below. The figure-reference mix-ups were caused by duplicate internal labels of the CO₂ and XCO₂ relative-error maps, which has been fixed.

Status: Implemented

General Comment 13 — Section 3.9's "impact of aggregation" framing is unclear

Comment: In Sect. 3.9, it is unclear what is meant by “impact of aggregation”, since the section compares simulations of different native resolutions, with all their own performance differences, that happen to be aggregated to enable the comparison. Studying the impact of aggregation in the referee’s sense would compare the performance of a native versus an aggregated version of the same simulation against data, though the referee is unsure this would be very informative, expecting the differences among simulations to be similar regardless of aggregation.

Response: Agreed; this is a framing and terminology issue that recurs through Sect. 3.9 and the Discussion/Conclusion (see also the specific comment on “aggregation error”, L355). We renamed the section (“Comparison of native 900 m and remapped high-resolution (LES) fields”) and rewrote its opening paragraph to state explicitly that it evaluates whether the native 900 m simulation is a dynamically consistent proxy for a spatially averaged high-resolution simulation, relevant when comparing model output to coarser satellite footprints, combining any effect of spatial averaging with all the other native-simulation differences (wind, PBLH, numerical diffusion, nudging) documented earlier in the paper, rather than presenting it as an isolated aggregation-effect study.

Status: Implemented

Specific Comments

Line numbers refer to the submitted preprint (egusphere-2026-2581).

Comment 1

Comment: L47f (Abstract): What motivates the aggregation analysis? Why average a high-resolution run when a coarser one is cheaper?

Response: We agree that the original phrasing implied a foregone conclusion. The motivation is not to propose averaged LES output as a cheaper alternative, but to test whether a native kilometre-scale simulation is an adequate proxy for a high-resolution simulation averaged over a coarser footprint, an assumption implicit whenever coarse-grid transport is compared with footprint-averaged satellite observations. This is now stated in a new paragraph of the Introduction and at the start of Sect. 3.9.

Status: Implemented

Comment 2

Comment: Fig. 1b: Stations near the D05 boundary may see a lower effective resolution because turbulence needs to build up from the inflow edge (cell-perturbation method, Muñoz-Esparza et al. 2018 style).

Response: We agree. No turbulence-generation technique (e.g. cell perturbation) is applied at the lateral boundaries of our LES domains, so stations close to the D05 inflow boundary may effectively be simulated at a coarser resolution than 100 m. A sentence stating this, with references to Mirocha et al. (2014) and Muñoz-Esparza et al. (2014, 2018), was added to Sect. 2.1. Because the inflow boundary and the development distance depend on wind direction, we state the caveat generally rather than flagging individual stations.

Status: Implemented

Comment 3

Comment: L101: D02 and D03 likely also fall in the convective gray zone, not just D01; provide a reference for the gray-zone resolution range.

Response: Agreed. Sentence revised to state that D01–D03 (8.1, 2.7, 0.9 km) all lie within the commonly cited convective gray zone ($\sim 1\text{--}10$ km), with a supporting citation added.

Status: Implemented

Comment 4

Comment: L105f: Please justify why FDDA/observation nudging was not used in the LES domains; observation nudging can be applied smoothly over larger distances. This is a major shortcoming if left undiscussed.

Response: See our response to General Comment 2. We now state explicitly that the nudging asymmetry (FDDA in D01–D03, free-running D04–D05) may contribute to the wind differences attributed to resolution, and we no longer present resolution as the only explanatory factor. The justification for not nudging inside the LES domains and its consequences are given in a new paragraph of Sect. 2.1, and the interpretation is revised in Sects. 3.1–3.2, the Discussion and the Conclusion.

Status: Implemented

Comment 5

Comment: L125–129: Statements (and Sects. 3.4–3.8) neglect plume displacement due to ignoring the vertical distribution of sources (Brunner et al., 2019).

Response: Agreed this is a relevant additional source of near-surface bias beyond the emission-height simplification already discussed. Added one sentence in Sect. 2.2 citing Brunner et al. (2019) alongside the existing discussion of surface-released emissions.

Status: Implemented

Comment 6

Comment: L140ff: State the number of observation locations for all datasets, not just weather stations.

Response: Agreed. Sect. 2.3 now gives the number of observation sites for each dataset (weather stations, Doppler wind LIDARs, ceilometers, Picarro and mid-cost stations, TCCON) and their temporal resolution.

Status: Implemented

Comment 7

Comment: L140ff (data selection): State whether all 24 h of CO₂ data were used and whether any quality screening was applied.

Response: All hours of the day passing the data providers' quality-control flags were used, with no additional selection. This is now stated in Sect. 2.3.

Status: Implemented

Comment 8

Comment: L184f: Reword “at face value” to something like “do not necessarily indicate model bias”; also, should the two PBLH methods not agree for a well-developed boundary layer?

Response: We accept the wording suggestion. On the substantive point: the two methods can disagree even for a well-mixed layer because the aerosol-based method (STRATfinder) detects the top of the aerosol-laden layer, which can

include a residual layer or entrainment zone, whereas the bulk Richardson number detects the top of the actively turbulent layer. The revised Sect. 2.4 states that model-observation PBLH differences therefore do not necessarily indicate model bias and explains why the two methods can diverge.

Status: Implemented

Comment 9

Comment: L279-281 / PSD: PSD (a variability measure) does not match observations in summer; please explain the summer over-amplification, since an underestimation (as in winter) would be expected from coarser-vs-finer arguments.

Response: We do not have a confirmed mechanism for the summer PSD overestimation at 100 m and 300 m. The revised Sect. 3.1 states the result, notes that it is not explained by the resolved-versus-parameterized turbulence argument, mentions the referee's hypothesis of a summer bias common to all configurations (consistent with the expected ordering at high frequency), and leaves it as an open question. The link to LES intermittency has been removed.

Status: Implemented

Comment 10

Comment: L282ff / nudging (Sect. 3.2): The wind profile discussion omits that D03 was nudged while D04/D05 were free-running, possibly the real explanation for the 900 m wind advantage, not resolution per se (cf. Ho et al., 2024).

Response: Agreed (see General Comment 2). Sects. 3.1 and 3.2 and the Discussion now state that the 900 m domain is nudged while the LES domains are free-running, and cite Ho et al. (2024) as evidence that nudging toward reanalysis or observations generally improves WRF winds relative to free-running configurations; the 900 m wind advantage is therefore not attributed to resolution alone.

Status: Implemented

Comment 11

Comment: L282ff (summer profiles): Only winter LIDAR wind profiles were shown; what about summer?

Response: Unfortunately, no Doppler wind LIDAR observations are available for the summer period (01-20 June 2024) at any of the four sites used (PACHEM, PALUPD, PAJUSS, PAARBO), so a summer profile evaluation cannot be made. This is now stated in Sect. 3.2 and Appendix C, and no summer ranking of wind profiles is claimed.

Status: Implemented

Comment 12

Comment: L316ff (PBLH stations): Sect. 3.3 claims four stations but only discusses one (QUALAIR) in the diurnal cycle; what about the other three, including MEUD which may behave differently?

Response: Agreed. The winter diurnal PBLH cycles at the three remaining D03 stations (Hotel des Villes, Chemin Vert Bobigny, Meudon) are now shown in Appendix C (Fig. C2). They broadly follow the QUALAIR pattern except at Meudon at night, where 900 m matches the observations while 100 m and 300 m underestimate them, as the referee suspected; this is stated in Sect. 3.3. No summer PBLH observations are available at these three stations.

Status: Implemented

Comment 13

Comment: L316ff (concluding statement): Section 3.3 lacks a concluding statement on which resolution performs best for PBLH, especially since summer appears inconsistent.

Response: Agreed. A concluding statement was added at the end of the diurnal-cycle paragraph of Sect. 3.3: LES improves winter PBLH, particularly at night and across the urban core, whereas no configuration is consistently best in summer and none reproduces the full afternoon boundary-layer depth.

Status: Implemented

Comment 14

Comment: L342f: The claim that 900 m overestimates PBLH “along most of the transect” looks more like 3 of 7 stations to the referee.

Response: We re-examined the figure and agree with the referee's reading. The sentence now states that 900 m overestimates PBLH at the three stations closest to the urban core (by about 100 m), whereas at the other stations the simulations and observations differ much less.

Status: Implemented

Comment 15

Comment: L352-392 (Fig. 6 vertical extent): Please comment on the striking difference in plume depth: the 900 m plume looks much shallower than 300/100 m, suggesting a large vertical-mixing difference between configurations.

Response: We agree that this is an important difference and thank the referee for pointing it out. Sect. 3.4 now states that near-source plumes remain within the lowest model levels at 900 m but are lifted over several hundred metres by the LES, and that this is not explained by the diagnosed PBLH (similar or larger at 900 m). It indicates much weaker near-source vertical transport by the MYJ closure at 900 m than by resolved updrafts in the LES. We did not quantify plume depth, but we use this result to help interpret the higher daytime near-surface concentrations of the 900 m run (Sect. 3.9.1; see also General Comment 10).

Status: Implemented

Comment 16

Comment: L355: “Aggregation error” is used incorrectly here, in the community (e.g. Ciais et al., 2010) it refers to coarse state-vector errors in flux inversion, not to what is described (which is a representation/representativeness error).

Response: Accepted. “Aggregation error” has been replaced by “representation error”, with Ciais et al. (2010) cited at first use (Sect. 2.4), and the term is no longer used elsewhere (see also L795, L805f, L837-839). Sect. 3.9 was renamed “Comparison of native 900 m and remapped high-resolution (LES) fields” to avoid implying that an isolated aggregation effect is measured (see L578ff).

Status: Implemented

Comment 17

Comment: L361-370: Shorten this paragraph on the transect definition to the essential point (it crosses point sources/plumes).

Response: Accepted. The transect definition is reduced to the essential point (it crosses the main point sources and is oriented along the mean wind), and the transect coordinates and station distances were moved to Appendix D.

Status: Implemented

Comment 18

Comment: L377-382 (cross-section interpretation, itemised): The cross-section (Fig. 6c,d) interpretation is largely a summer-vs-winter comparison, not a resolution comparison, uses non-matching colour scales, and several individual statements ("strengthens gradients", "accentuates differences", "more vertically homogeneous", "confined") are judged vague, unsupported by a matched dynamic range, or simply not followed by the referee. Suggests quantification (matched colour scales, histograms of within-PBL concentration, height containing X% of plume mass) or removal.

Response: We agree that this paragraph mixed a seasonal comparison with the resolution comparison, and that several statements were not supported quantitatively. Following the referee's advice not to expand it point by point, we kept in Sect. 3.4 only the resolution-relevant statement (the much shallower near-source plumes at 900 m; see L352-392) and moved a shortened, explicitly qualitative description of the two cases to Appendix E. The disputed statements ("accentuates differences", "more vertically homogeneous", "confined near the emission source") were removed.

Status: Implemented

Comment 19

Comment: L406f: Clarify: the delayed peak at high resolution can also occur if the source grid cell begins later along the transect, not only because the transect misses the maximum grid cell.

Response: Accepted; both mechanisms are relevant and are now both stated. Sentence expanded to include this alternative explanation.

Status: Implemented

Comment 20

Comment: L413-415: Name the specific stations affected by the grid-cell-alignment issue (likely HBO, TF1, and EIF).

Response: Confirmed; EIF is in the plume but far from the source, HBO and TF1 are close to the source but in a neighbouring cell. Station names added explicitly.

Status: Implemented

Comment 21

Comment: L393-441 (Sect. 3.5 length): Section 3.5 is too long; move transect coordinates to the previous section; the referee proposes a specific condensed replacement paragraph.

Response: Accepted. Sect. 3.5 was rewritten around the condensed paragraph proposed by the referee, with numbers added for the transect maxima (about 590 ppmv at 100 m against 525-550 ppmv at 900 m in winter, 510 against 445 ppmv in summer) and the station names (HBO, TF1 and EIF in winter; NEY in summer). The transect coordinates and station distances were moved to Appendix D. The section is now one paragraph.

Status: Implemented

Comment 22

Comment: L451-458: This paragraph repeats conclusions already stated ("consistent with the plume and transect-based...") and adds little.

Response: Agreed; removed. Paragraph deleted from Sect. 3.6; the single new piece of information (station-specific timing of events) retained as one sentence.

Status: Implemented

Comment 23

Comment: L463-465: Add that wind direction performance is worse at 300/100 m than at 900 m as part of the explanation for CO₂ time-series mismatches.

Response: Accepted (see also General Comment 7). Sect. 3.6 now states that wind-direction errors are larger at 300 m and 100 m than at 900 m (MAE 26-33° against 18-27°, Table 2) and that this is enough to displace a plume laterally by a grid cell or more a few kilometres downwind.

Status: Implemented

Comment 24

Comment: L471f: Be specific about “certain conditions” (e.g. “depending on plume displacement”); consider that spurious modelled peaks with no observed counterpart could reflect a source that was offline (e.g. maintenance) in reality.

Response: We accept the wording fix: “under certain conditions” is replaced by “depending on plume displacement relative to the observation site”. On source outages: our inventory has no facility-level operational records at hourly resolution, so we can neither confirm nor exclude this. Sect. 3.6 now mentions the use of sector-level rather than facility-specific temporal profiles for point sources, and a temporarily offline source, as untested explanations for simulated peaks without an observed counterpart.

Status: Implemented

Comment 25

Comment: L473-498 (Sect. 3.7 length/quality): Section 3.7 has little quantitative content, is often vague, does not state which resolution performs best, and could largely stand on its concluding lines alone (491–498). Referee gives several detailed sub-comments (L482, L483, L486f, L486-488, L489f) but explicitly says not all need to be answered individually if the section is rewritten and shortened.

Response: We rewrote Sect. 3.7 rather than patching it. The revised text gives the bias and timing errors for each case day and states which resolution performs best, and includes the corrections raised by the referee: on 11 January, the morning observations are reproduced well but the XCO₂ rise around 12:00 UTC is mistimed by about half an hour; on 16 January, the simulations match the observations at 11:00 and 14:30-15:00 UTC but miss the observed peak in between; this column peak occurs several hours after the near-surface CO₂ maximum at JUS (Fig. 6a), so the two underestimations may not reflect the same signal; in summer, the simulations overestimate XCO₂ (by up to about 1 ppm on 7 June) and miss the two isolated high values of 9 June. We also note that model-observation differences of this size are not negligible for inversions and exceed the differences between resolutions.

Status: Implemented

Comment 26

Comment: L509 / L512: State the actual peak concentration for the observations and the 900 m simulation (not just the 300/100 m values), so magnitudes are comparable.

Response: Accepted. The winter PDF modes are now given for the observations and all resolutions (about 440 ppmv for Picarro and 450 ppmv for mid-cost in both the observations and the 900 m run, against 432 and 437 ppmv at 300 m and 100 m), as well as the summer modes (Sect. 3.8.1).

Status: Implemented

Comment 27

Comment: L514f / L533-534 (PDF logic): “Finer resolution suppresses high-mixing-ratio episodes” (linear-scale discussion) contradicts the later, correct statement that the 100 m run best reproduces the high tail (log-scale discussion); the “narrower distributions reflect preservation of localized emission maxima” framing is also judged internally inconsistent, narrower central distributions more likely indicate a low mean bias in enhancements.

Response: Agreed, this was an internal inconsistency. The narrower central PDFs at 300/100 m reflect a low bias in typical (non-peak) enhancements, which is a separate feature from the improved representation of rare high-concentration events seen in the log-scale tails. Both statements corrected and reconciled; the summer summary sentence replaced with the referee’s suggested wording (pdf maxima too low and too narrow, little inter-resolution difference except at rare high-concentration events, which finer resolutions capture better).

Status: Implemented

Comment 28

Comment: L516-519: The text does not note that the simulated PDFs are skewed low while the minima match observations, implying a mean bias in the enhancements. Could this be linked to PBLH, wind, or the emission inventory?

Response: Agreed. The revised Sect. 3.8.1 states that the simulated distributions are shifted toward lower values while their lower bounds match the observations, i.e. the enhancement is underestimated, which the new Table 3 confirms (negative mean bias at all resolutions). We discuss the candidate causes: the wind-speed overestimation of the LES runs, which is partly consistent with the winter bias ranking but not with the summer results; PBLH and near-source vertical mixing; and uncertainties in the emission inventory and boundary conditions, which are common to all runs. We state that these factors could not be disentangled with the present experiments.

Status: Implemented

Comment 29

Comment: L526: State the total number of observations, to put the 5×10^{-4} probability-density threshold in context.

Response: Accepted. The total number of station-hours of each PDF sample is now given in Sect. 3.8.1, and the number of stations per network in Sect. 2.3.

Status: Implemented

Comment 30

Comment: L526-528: The proposed explanation (winter stable-boundary-layer accumulation) does not match the figure, since the 100 m tail overestimation is more prominent in summer (Fig. 10f,h) than winter (10e,g).

Response: Agreed; the original text mis-stated which season the effect is strongest in. Sentence corrected to state the spurious high-tail peaks are more prominent in summer; the stable-boundary-layer winter mechanism is removed as the stated cause, and the phenomenon is now flagged as not fully explained.

Status: Implemented

Comment 31

Comment: L532: This summary sentence itself shows that not all of the preceding analyses need to be kept in the paper.

Response: Agreed, taken as direct support for our overall restructuring (see General response). Sentence removed along with the surrounding redundant summary paragraph.

Status: Implemented

Comment 32

Comment: L535 / L535-537: “Central tendency captured regardless of grid spacing” is vague/too optimistic; the claim that the 100 m spurious peaks occur in “both seasons” is not supported by Fig. 10 (winter tail is captured well).

Response: Accepted; both corrected together with L514f/L533-534 above. Summer paragraph replaced with the referee's suggested sentence; “both seasons” corrected to “summer” for the spurious-peak statement.

Status: Implemented

Comment 33

Comment: L504-539 (Sect. 3.8.1 overall): The section does not note the systematic underestimation visible in Fig. 10a-d; recommends a Table-2-style summary table for CO₂/XCO₂ bias, and asks whether the smaller 900 m summer underestimation links to meteorological performance.

Response: Agreed on all three points. The new Table 3 gives R, MBE, MAE and RMSE per network, resolution and season; the systematic underestimation is stated explicitly in Sect. 3.8.1; and the smaller winter underestimation at 900 m is discussed in relation to the wind-speed biases (see General Comment 10). The summer mean biases are, however, almost identical across resolutions (-4.1 to -4.6 ppmv for Picarro), so we do not find a clear summer link to meteorological performance.

Status: Implemented

Comment 34

Comment: L552: Quantify “higher-resolution simulations improve correlations and reduce RMSD” in summer, the effect looks small in Fig. 11 and its significance is unclear.

Response: Accepted. The text now gives the numbers from Table 3: for Picarro in summer, R increases from 0.58 at 900 m to 0.68-0.70 at 300 m and 100 m and RMSE decreases from 9.1 to 8.4-8.6 ppmv. We state that significance was not tested and describe the improvement as modest.

Status: Implemented

Comment 35

Comment: L552-555: Be specific about how “occasional overestimations” influence the Taylor-diagram statistics (e.g. “reduce” correlation, not just “influence” it).

Response: Accepted. The sentence now states that the occasional overestimated peaks at high resolution reduce correlation and increase RMSE, even as the standard deviation moves closer to the observed one.

Status: Implemented

Comment 36

Comment: L556 / L556-557: *State the actual similarity for the mid-cost network before describing the differences; the explanation given for the “broader spread” (stations within Paris) is judged inconsistent with the Picarro-network explanation, where within-Paris stations cluster tightly.*

Response: Agreed the explanation was inconsistent. The broader mid-cost spread more likely reflects the larger number and greater siting heterogeneity (rooftop heights, local exposure) of mid-cost sensors rather than their location within Paris. Sentence revised to state the similarity to Picarro results first, and the spread explanation corrected to cite sensor siting heterogeneity rather than intra-Paris location.

Status: Implemented

Comment 37

Comment: L563: *“Particularly in summer” should be “only in summer”.*

Response: Accepted. Wording corrected.

Status: Implemented

Comment 38

Comment: L564f: *Clarify “occasionally slightly larger RMSD” with actual numbers/stations, and note that the winter degradation and summer improvement appear similar in magnitude in Fig. 11c, contrary to the text’s implication.*

Response: Accepted. The paragraph now gives the numbers from Table 3 and, as the referee noted, states that at the mid-cost stations the winter degradation at high resolution is of similar size to the summer improvement: R decreases from 0.58 (900 m) to 0.42-0.49 in winter and increases from 0.52 to 0.63 in summer, while RMSE is larger at 300 m and 100 m in winter (25.4-27.8 against 22.6 ppmv) and similar in summer (12.2-12.7 ppmv).

Status: Implemented

Comment 39

Comment: L540-577 (Sect. 3.8 overall): *The section is an improvement over 3.4–3.7 but is still too long and mostly repeats earlier conclusions; recommends shortening.*

Response: Accepted. Sect. 3.8.2 shortened, focusing on the new (non-repeated) information: the double-penalty pattern and the urban/non-urban station grouping.

Status: Implemented

Comment 40

Comment: L578ff (Sect. 3.9 scope/title): *The section title (“impact of aggregation”) does not match its content, which compares native-resolution simulations (with all their dynamical differences) after remapping to a common*

grid, it does not isolate an aggregation effect. The referee is unsure the isolated aggregation effect (native vs. its own aggregate) would even be informative.

Response: Agreed; this is the same terminology/framing issue as L355 and recurs through Sect. 3.9 and the Discussion/Conclusion. We renamed the section and reframed its purpose: it evaluates whether a native 900 m simulation is a dynamically consistent proxy for a spatially averaged high-resolution simulation, not the isolated effect of averaging on a single simulation. Sect. 3.9 title changed to “Comparison of native 900~m and remapped high-resolution (LES) fields”; opening paragraph rewritten to state this scope explicitly; “aggregation error”/“effects of aggregation” replaced by “representation difference between native and remapped fields” throughout the section, Discussion and Conclusion.

Status: Implemented

Comment 41

Comment: L583: “Near-surface numerical noise” is not a recognised term; what is meant?

Response: The intended meaning was noise associated with the first model level immediately above the surface (e.g., surface-layer parameterization artifacts), not a distinct physical phenomenon. Phrase replaced with a specific description (“artifacts of the surface-layer parameterization at the first model level”).

Status: Implemented

Comment 42

Comment: L597-604: The claimed consistency with a “well-mixed” daytime boundary layer with comparable PBLH does not explain Fig. 12a: the 900 m simulation has the higher PBLH, which should dilute the same sources over a larger volume and give lower mixing ratios than at 300/100 m (as recognized for night-time in L606-608), yet in daytime the opposite is the case. Why?

Response: The referee is correct: in winter daytime the native 900 m field is higher than the remapped LES fields over most of Paris (domain-mean MBE +4.8 to +6.0 ppmv, Table 4), although the 900 m run has the highest PBLH. The submitted text did not describe this correctly. The revised Sect. 3.9.1 states the result and notes that PBLH alone would predict the opposite. It then points to two documented differences acting in the other direction: the LES runs overestimate wind speed (+0.9 to +1.7 m/s in winter, while 900 m is unbiased), and their resolved updrafts lift near-source plumes much higher than the 900 m closure (Sect. 3.4). At night the sign reverses (MBE -2.5 to -3.9 ppmv), consistent with the shallower LES boundary layer. We state that the relative contributions of wind, vertical mixing and PBLH were not quantified.

Status: Implemented

Comment 43

Comment: L609-613: This explanation (better-resolved turbulence “limiting vertical mixing”) is judged inaccurate/circular and mostly repeats the previous paragraph.

Response: Agreed; removed. Sentences deleted; the single new claim (nocturnal CO₂ accumulation is larger at finer resolution) is kept as a one-line factual statement without the disputed mechanism.

Status: Implemented

Comment 44

Comment: L613-615 / L615-617: These sentences either speculate without confirming (whether CO₂ is trapped below 35 m) or restate points made “numerous times already” and are unclear.

Response: Agreed; the speculative sentence is removed rather than left unconfirmed, and the repeated sentence is deleted. Both passages removed from Sect. 3.9.1.

Status: Implemented

Comment 45

Comment: L625f: Define “modest” quantitatively; note that relative differences (E_{sym}) are similar day/night even though absolute concentration differences are not, relevant because flux-estimation error tracks the relative metric.

Response: Accepted. The paragraph now quantifies the absolute differences (e.g. winter RMSE of 11.0-11.6 ppmv by day and 15.5-16.6 ppmv at night) and states that E_{sym} is only slightly lower at night than during the day in both seasons (e.g. 28-34 % against 28-29 % in winter). Because flux-estimation errors scale with relative errors, the larger absolute night-time differences do not imply larger flux errors at night (Sect. 3.9.1, Discussion).

Status: Implemented

Comment 46

Comment: Table 3: Include mean bias, in addition to RMSE/MAE/ E_{sym} /R.

Response: Accepted. A mean bias row (MBE, native 900 m minus remapped field, in ppmv) was added for each tracer and period in the native-versus-remapped table (now Table 4).

Status: Implemented

Comment 47

Comment: L626: “Confinement” and “amplification” are used inconsistently/ambiguously throughout the manuscript; please use more precise language.

Response: Accepted. Global pass through the manuscript replacing “confinement” with either “shallower boundary layer” or “smaller horizontal extent” depending on the intended meaning, and “amplification/amplified” with an explicit comparison (e.g. “higher concentrations due to the shallower PBLH”) or simply “larger”.

Status: Implemented

Comment 48

Comment: L627, 638, 644: Figure references appear swapped: “Fig. 15” should be “Fig. 13” in these instances (surface CO_2 , not XCO_2).

Response: Confirmed and corrected; see also L645 below for the related, more serious mix-up. Figure callouts corrected throughout Sect. 3.9.1.

Status: Implemented

Comment 49

Comment: L640-643: The explanation (shallower nocturnal PBL → “broader domain-wide differences” in E_{sym}) is judged inconsistent with the same PBLH argument being used to explain larger night-time absolute differences, since night-time E_{sym} is in fact smaller than daytime; wind speed/direction differences are not considered.

Response: Agreed. The two arguments are now separated: the absolute night-time differences are larger, whereas E_{sym} changes little between day and night because the differences and the night-time enhancement increase together (Sect. 3.9.1). The spatial pattern of E_{sym} is no longer attributed to PBLH, and wind-direction differences between the native runs are mentioned as a factor in the night-time pattern near point sources.

Status: Implemented

Comment 50

Comment: L644: Replace “seasonally amplified” with “larger”.

Response: Accepted. Wording corrected.

Status: Implemented

Comment 51

Comment: L645: The text says surface CO_2 E_{sym} reaches “up to 70%” while consistently citing Fig. 15 (XCO_2) rather than Fig. 13 (surface CO_2); Table 3 in fact gives 88.4% for surface CO_2 and 73.18% for XCO_2 , the latter being much closer to the quoted “70%” figure, suggesting the text describes XCO_2 while claiming to describe surface CO_2 .

Response: Confirmed as an error: the paragraph mixed up the two tracers. The revised Sect. 3.9.1 refers only to near-surface CO_2 and its E_{sym} maps (Appendix I), and Sect. 3.9.2 only to XCO_2 (Appendix J). We now quote domain means from Table 4 (e.g. 56-61 % for near-surface CO_2 in summer) rather than grid-cell maxima.

Status: Implemented

Comment 52

Comment: L646: What is “background mixing”? Perhaps “vertical mixing” is meant.

Response: The phrase has been removed in the rewritten paragraph.

Status: Implemented

Comment 53

Comment: L648-650: This statement (aggregation “partially smooths extreme local maxima” under “certain stable conditions”) is judged unsupported, since the analysis does not separate aggregation effects from native-simulation dynamical differences, and the terms “extreme local maxima” and “certain stable conditions” are unclear.

Response: Agreed; removed rather than clarified, since we cannot isolate the claimed mechanism with the present analysis. Sentence deleted from Sect. 3.9.1.

Status: Implemented

Comment 54

Comment: L661: What is “coarser mixing”, and why was it not offered as an explanation earlier (L373)?

Response: “Coarser mixing” was meant as the same numerical-diffusion and spatial-averaging argument used for L373. The phrase has been removed, and the differences between native and remapped fields are now attributed consistently to differences in wind, vertical mixing, PBLH and numerical diffusion between the native simulations (Sects. 3.9.1-3.9.2).

Status: Implemented

Comment 55

Comment: L669ff: How was night-time XCO_2 computed, given that TCCON averaging kernels are not defined without sunlight? State that night-time XCO_2 is more relevant to active sensors (e.g. MERLIN, DQ-1).

Response: This was insufficiently documented in the submitted manuscript. The TCCON averaging kernel and prior were applied only in the model-TCCON comparison (Sect. 3.7), which is restricted to daytime. For the native-versus-remapped comparison (Sect. 3.9.2), day and night, no averaging kernel was applied: model XCO_2 is a simple pressure-weighted column average of the WRF CO_2 profile. Both definitions are now stated explicitly in Sect. 2.4, together with the note that a night-time comparison with column retrievals would require an active-sensor instrument (e.g. MERLIN, DQ-1) with its own averaging kernel.

Status: Implemented

Comment 56

Comment: L672: The reported correlations (~ 0.8) are not obviously “good agreement” given the reported 20–60% E_{sym} , which matters more for flux estimation.

Response: Agreed; the framing was too optimistic relative to the relative-error metric that is more directly relevant to the paper’s stated motivation. “Good overall agreement” replaced with a more neutral statement pairing the correlation with the corresponding E_{sym} range.

Status: Implemented

Comment 57

Comment: L673f: “Stronger sensitivity to boundary-layer stability” should not exclude wind as a contributing factor.

Response: Accepted. The sentence was removed. Wind is now discussed together with boundary-layer depth as a factor in the differences between native and remapped fields (Sect. 3.9.1) and in the day/night XCO_2 contrast (Sect. 3.9.2).

Status: Implemented

Comment 58

Comment: L680ff (averaging kernel shape): For TCCON, the averaging-kernel shape depends on solar zenith angle (Laughner et al., 2024); kernels are undefined at night, what kernel was actually used here?

Response: Addressed together with L669ff: no averaging kernel was applied in the native-versus-remapped XCO_2 comparison, and the sentence on averaging-kernel sensitivity was removed from Sect. 3.9.2. Sect. 2.4 now notes that the TCCON kernels used for the daytime TCCON comparison depend on solar zenith angle (Laughner et al., 2024).

Status: Implemented

Comment 59

Comment: L680ff (wind speed factor): The paragraph on day/night XCO_2 differences does not consider wind speed, even though wind speed is typically lower at night, which would work against (not for) the stated explanation.

Response: Agreed; we removed the implication that PBLH alone explains the day/night contrast and now note that the wind-speed argument would work in the opposite direction, leaving the net day/night driver only partially explained. Paragraph revised accordingly; explanation qualified rather than asserted.

Status: Implemented

Comment 60

Comment: L720-723: It is unclear why weaker XCO₂ signal (summer) should mechanically produce larger E_{sym}; larger transport differences among the simulations in summer are a plausible alternative, but this was not tested; background variability is not the explanation since background is subtracted.

Response: We agree that our original explanation was not correct. As the referee points out, if the differences between simulations scaled with the signal, a weaker summer signal would not change E_{sym}. The larger summer E_{sym} therefore means that the inter-simulation differences decrease less than the enhancement itself. We did not compare the meteorology of the native simulations with each other, so we cannot say whether transport differences are larger in summer; we mention as a possible contributing factor that the summer enhancement, reduced by deeper mixing and partly offset by biogenic uptake, is often close to zero, where any relative metric becomes large. The revised Sects. 3.9.1-3.9.2 state this without the overstated causal claim.

Status: Implemented

Comment 61

Comment: L741f: The double-penalty effect is not specific to the gray zone; it should worsen at even finer resolutions, and fully resolving eddies would not, by itself, improve point comparisons because of turbulence's random nature.

Response: Agreed; the gray-zone framing of the double-penalty argument was incorrect. Sentence corrected to state that the double-penalty effect is a general property of high-resolution deterministic verification, not specific to the gray zone, and that further increasing resolution does not remove it.

Status: Implemented

Comment 62

Comment: L751-756: These lines imply model differences are larger in winter than summer; true for absolute concentrations but the opposite for relative differences (E_{sym}), which matter more for flux estimation.

Response: Accepted; this seasonal-framing inconsistency recurs (cf. L625f) and is corrected globally. Paragraph revised to state both the absolute (winter > summer) and relative (summer > winter) seasonal contrasts explicitly, rather than only the absolute one.

Status: Implemented

Comment 63

Comment: L756-757: The Mitchell et al. (2018) citation is presented as establishing a finding (shallow PBL enhances surface-released pollutant concentrations) that was well known long before.

Response: Agreed; the citation is retained as an example of urban CO₂-specific evidence but the wording no longer implies it established the general finding. Sentence reworded to remove the implication of novelty.

Status: Implemented

Comment 64

Comment: L758-762: The paragraph implies the PBLH-methodology difference could affect results beyond the PBLH comparison itself (winds, CO₂); PBLH is a diagnostic in WRF and should not feed back on other fields; also these lines break the surrounding argument.

Response: Agreed on both points: PBLH is a diagnostic in our WRF configuration and does not affect the simulated wind or CO₂ fields. The caveat was moved out of the Discussion to the evaluation strategy (Sect. 2.4), where it is explicitly limited to the PBLH comparison, and Sect. 3.3 refers to it.

Status: Implemented

Comment 65

Comment: L756-764: The claim that “benefits of high-resolution modelling are strongest under stable conditions” is judged unsupported and in tension with Sects. 3.8.1 (900 m PDF often closer to obs. in winter) and 3.8.2 (RMSD similar across resolutions in winter, improving at high resolution mainly in summer); Sect. 3.9 compares simulations to each other, not to data, so supports no performance conclusion.

Response: Agreed; this was an overreaching synthesis statement not supported by the paper's own results. Sentence/claim removed from the Discussion; replaced by a more limited, directly supported statement (LES improves representation of variability and rare high-concentration events in both seasons; it does not consistently improve bulk error metrics, and the picture differs between CO₂/XCO₂ statistics on the one hand and Sect. 3.9's cross-resolution comparison, which is not a performance comparison, on the other).

Status: Implemented

Comment 66

Comment: L772: “May partly reflect representativeness effects rather than purely model deficiencies” is unclear; suggests removing, as the key point was already made in the previous sentence.

Response: Accepted. Sentence removed.

Status: Implemented

Comment 67

Comment: L773-778: This passage (“trade-off between simulated variability and conventional error metrics...”) is hard to follow and largely repeats earlier statements; referee proposes a specific replacement sentence.

Response: Accepted; the referee's proposed sentence was adopted almost verbatim (Discussion, fourth paragraph): “Comparisons with the ground-based observation networks show that better-matching simulated variability does not necessarily translate into better average agreement with the data (e.g. RMSE).”

Status: Implemented

Comment 68

Comment: L785f: “Resolution effects propagate vertically through the atmospheric column” overstates what was shown; the study did not demonstrate a propagation mechanism from the surface upward.

Response: Accepted. Reworded to: “resolution differences also affect column-averaged CO₂ (XCO₂)”, removing the causal “propagation” language.

Status: Implemented

Comment 69

Comment: L795: Same misuse of “aggregation error” as at L355.

Response: Corrected together with the global terminology fix (see L355). See L355 change entry.

Status: Implemented

Comment 70

Comment: L797-799: “Sub-grid plume variability... cannot be fully represented by simple spatial averaging” is unclear; the referee attributes the differences instead to differing winds, PBL heights and numerical diffusion across the native simulations.

Response: Agreed; “sub-grid plume variability” was an imprecise catch-all. The referee's list (wind, PBLH, numerical diffusion) is a more accurate and already-supported description of the drivers. Sentence replaced with

an explicit statement that the persistent differences are attributable to differing simulated winds, PBLH, and numerical diffusion between the native configurations, not to an intrinsic “sub-grid variability” concept.

Status: Implemented

Comment 71

Comment: L799-801: Disagrees with the explanation that weaker summer plume amplitude directly “translates into” larger E_{sym} ; suggests larger summer inter-simulation transport differences as an alternative, but notes this was not tested since meteorology differences among simulations were only evaluated against data, not against each other.

Response: Agreed; see our response to L720-723. We no longer assert a causal chain, and we state explicitly that the meteorology of the simulations was compared only with observations, not between simulations, so that larger summer transport differences remain possible but untested. Comparing the simulated meteorology between resolutions is listed as a direction for future work (Conclusion, item iv).

Status: Implemented

Comment 72

Comment: L801-802: Repetition of a previous statement.

Response: Accepted; removed. Sentence deleted.

Status: Implemented

Comment 73

Comment: L805f: “Significant representation errors” is used incorrectly (see L355); moreover, the claim that aggregating without accounting for sub-grid plume variability introduces such errors was not actually tested (native vs. aggregated performance against data was not compared), and “sub-grid plume variability” is again unclear.

Response: Agreed on all points; see the L355 terminology fix and the L797-799 rewording. We removed the untested causal claim about footprint aggregation. Passage revised to state only what was shown (native and remapped high-resolution fields differ non-negligibly from the native 900 m simulation) without the untested claim about the consequence for footprint-matched satellite comparisons, which is now phrased as a hypothesis for future work rather than a finding.

Status: Implemented

Comment 74

Comment: L826f: What exactly are the “careful considerations” regarding spatial resolution, observational representativeness and scale mismatch?

Response: This was too vague to be actionable. Sentence replaced with 2–3 concrete recommendations (e.g., match model and footprint scale before comparing to satellite retrievals; report relative as well as absolute error metrics; document nudging/physics differences between resolutions explicitly when interpreting resolution sensitivity).

Status: Implemented

Comment 75

Comment: L827: What does “interaction between model resolution and observational footprint” mean?

Response: Reworded together with L826f above to be concrete. See L826f change entry.

Status: Implemented

Comment 76

Comment: L834: *“May amplify biases related to PBL dynamics” is unclear.*

Response: The sentence was removed. The relationship between PBLH and near-surface concentrations is now discussed explicitly, together with wind and vertical mixing, in Sect. 3.9.1 and the Discussion.

Status: Implemented

Comment 77

Comment: L837-839: *“Differences... are not primarily due to model physics, but rather to aggregation errors...” is not supported by the analyses, and the aggregation-error concept is again used incorrectly.*

Response: Agreed; this unsupported and terminologically incorrect claim was removed (see General Comment 2 and our response to L355). The Discussion and Conclusion now attribute the resolution-dependent differences to a combination of numerical diffusion, wind, vertical mixing and boundary-layer representation, and the nudging asymmetry between the mesoscale and LES domains, and state that these were not disentangled in this study.

Status: Implemented

Comment 78

Comment: L840-842: *Sentence meaning unclear (“inability of coarse grids to aggregate high-resolution emission variability...”).*

Response: Agreed; removed along with L837-839 as part of the same unsupported argument. Sentence removed.

Status: Implemented

Comment 79

Comment: L846: *“Excessive surface CO₂ accumulation” at night is not clearly shown; Fig. 10 rather suggests an overall underestimation (not distinguishing day/night).*

Response: Agreed that the Conclusion overstated this. The revised Conclusion distinguishes two separate results: the model-observation comparison, where all resolutions underestimate CO₂ (Table 3), and the comparison between simulations, where the LES runs give higher night-time concentrations than the 900 m run (Sect. 3.9.1).

Status: Implemented

Comment 80

Comment: L848 (turbulence schemes / vertical profile data): *Which results show that turbulence schemes need improvement? What is “vertical profile data”, wind LIDAR used for nudging in the 900 m domain? If so, this implicitly acknowledges the nudging asymmetry.*

Response: Agreed; the sentence alluded to the nudging issue without stating it. It was removed and replaced by an explicit statement of the nudging asymmetry (Discussion, first paragraph; Conclusion, item 3).

Status: Implemented

Comment 81

Comment: L848 (vertical/horizontal coupling): *“Vertical and horizontal resolution effects are tightly coupled” is not supported, the study did not vary vertical resolution.*

Response: Accepted; removed. Sentence deleted.

Status: Implemented

Comment 82

Comment: L849 (propagation): “Errors in near-surface transport propagate to the total column XCO_2 ” overstates what was shown; no propagation mechanism was demonstrated.

Response: Accepted; corrected together with L785f. Reworded to state that near-surface and column-integrated CO_2 show related, but not causally demonstrated, resolution-dependent differences.

Status: Implemented

Comment 83

Comment: L851f (sensor/grid alignment): “Model–observation agreement depends on how sensor locations align with grid cells” was not directly investigated in this study.

Response: We partly disagree: the dependence on station position relative to the plume is shown for the transects (Sect. 3.5) and the time series (Sect. 3.6), although not as a general, quantified relationship across the whole network. The statement is kept but restricted to these results (Discussion, fourth paragraph).

Status: Implemented

Comment 84

Comment: L858-862 / L865f: Disagrees with the recommendation of a hybrid coarse-meteorology/fine- CO_2 -transport approach, since the 900 m wind advantage may stem from nudging rather than an intrinsic limitation of 100/300 m meteorology.

Response: Agreed; we no longer recommend against high-resolution wind modelling in general. The Conclusion is revised to state that the observed 900 m wind advantage is confounded with the nudging asymmetry, and that a fair comparison of resolution effects on meteorology would require nudging (or an equivalent boundary-layer-respecting technique, e.g. nudging to the driving analysis above the PBL) applied consistently across resolutions, which is left for follow-on work. Recommendation for a coarse-meteorology/fine- CO_2 hybrid removed from the Conclusion; replaced with a recommendation to test consistent nudging across resolutions in future work, explicitly citing the nudging asymmetry as an unresolved confound in the present study.

Status: Implemented

Technical Corrections

All technical/wording corrections below are accepted as suggested unless noted otherwise.

Comment 85

Comment: L10: Add “[compared to the native 900 m run]” for clarity.

Response: Accepted. The abstract was rewritten; the corresponding sentence now states explicitly that the LES fields remapped onto the 900 m grid are compared with the native 900 m simulation.

Status: Implemented

Comment 86

Comment: L11-12: Specify what the quoted ranges mean (not obviously min–max).

Response: Accepted. The abstract now states that the ranges are domain-mean values of E_{sym} , covering day and night and both LES resolutions (e.g. 28–34 % in winter and 56–61 % in summer for near-surface CO_2).

Status: Implemented

Comment 87

Comment: L21: “([e.g.,] Ye et al., 2020)” – fix citation formatting.

Response: Accepted. “e.g.,” added inside the citation.

Status: Implemented

Comment 88

Comment: L24: “([e.g.,] Che et al., 2024)” – fix citation formatting.

Response: Accepted. “e.g.,” added inside the citation.

Status: Implemented

Comment 89

Comment: L32ff: “Large Eddy simulation” singular usage feels off; prefer plural “LES”/“LES models”.

Response: Accepted. Usage corrected throughout for grammatical number where flagged.

Status: Implemented

Comment 90

Comment: L37f, L76f: Clarify how the WRF modification used here (Lauvaux et al., 2012) differs from WRF-GHG (Beck et al., 2011).

Response: Accepted. Sect. 2.1 now states that CO₂ is transported as a passive tracer using the greenhouse-gas tracer option of WRF-Chem (WRF-GHG; Beck et al., 2011) as configured for urban applications by Lauvaux et al. (2012), and the Introduction cites the tracer implementations (Ahmadov et al., 2007; Pillai et al., 2010, 2011; Beck et al., 2011; Lauvaux et al., 2012).

Status: Implemented

Comment 91

Comment: L63: Clarify what “intensive” periods means.

Response: Accepted. Reworded to “intensive observation periods, during which the complete multi-instrument network was operating” (Sect. 2).

Status: Implemented

Comment 92

Comment: L69: Clarify “complex land–atmosphere interactions”.

Response: Accepted. This sentence was removed when Sect. 2 was condensed;

Status: Implemented

Comment 93

Comment: L70: “Representative” of what?

Response: Accepted. This sentence was removed when Sect. 2 was condensed.

Status: Implemented

Comment 94

Comment: L72: “Seasonal variability” → “temporal” or “spatio-temporal” variability fits better.

Response: Accepted. This sentence was removed when Sect. 2 was condensed.

Status: Implemented

Comment 95

Comment: L115: Fix reference "(Agency, 2024)".

Response: Accepted. Corrected to European Space Agency

Status: Implemented

Comment 96

Comment: L134: Fix reference "(Eric, 2021)" → "(Vermote, 2021)".

Response: Accepted. As above, the author field has been corrected to Vermote.

Status: Implemented

Comment 97

Comment: L168: "Vertical distribution of CO₂" → "vertical average of CO₂".

Response: Accepted. The sentence was rephrased; it now states that TCCON integrates CO₂ over the full atmospheric column.

Status: Implemented

Comment 98

Comment: L210: "Stratified" → "divided".

Response: Accepted. The sentence was removed in the condensed Sect. 2.4.

Status: Implemented

Comment 99

Comment: L247: "Previous study" → "A previous study".

Response: Accepted. The sentence was rewritten and now cites Udina et al. (2020) directly.

Status: Implemented

Comment 100

Comment: L262: "Verification metrics" → just "metrics".

Response: Accepted. Wording simplified.

Status: Implemented

Comment 101

Comment: L265: Clarify "second level timescales".

Response: Accepted. The sentence now states that, given the 10-min data resolution, the analysed frequencies span sub-hourly to multi-hour variability rather than microscale turbulence.

Status: Implemented

Comment 102

Comment: Fig. 5d: Add "winter" to the panel title.

Response: Accepted. Caption/panel label updated.

Status: Implemented

Comment 103

Comment: L355 (*plume integrity*): Clarify what “*plume integrity*” means.

Response: Accepted. Reworded to “sharp concentration gradients”, consistent with usage elsewhere.

Status: Implemented

Comment 104

Comment: L361: “*The transect was defined...*” → “*The transects were defined...*” (*plural*).

Response: Accepted. The sentence was removed when the transect description was condensed (Sect. 3.4 and Appendix D).

Status: Implemented

Comment 105

Comment: L448: “*Figures 8c and 8c*” → “*Figures 8c and 8d*”.

Response: Accepted. The sentence was removed in the shortened Sect. 3.6.

Status: Implemented

Comment 106

Comment: L458: *Missing full stop.*

Response: Accepted. Punctuation fixed.

Status: Implemented

Comment 107

Comment: L509: “*Distributions*” → “*distribution*”.

Response: Accepted. Grammar corrected.

Status: Implemented

Comment 108

Comment: L520: “*The differences between model resolutions [at high mixing ratios] become more pronounced*” – clarify, since differences are in fact less pronounced near the distribution centre.

Response: Accepted. Clause added to scope the statement to the high-mixing-ratio tail specifically.

Status: Implemented