We sincerely appreciate the editor and the reviewer for their careful evaluation of our manuscript and for providing constructive and insightful comments. These comments have been very helpful in improving the quality, clarity, and scientific rigor of our work. We have carefully considered each comment and have provided a detailed response to address the reviewer’s concerns.

For clarity, the reviewer’s comments are presented in *italic* font, while our responses are provided in regular font. The reviewer’s comments and the corresponding author responses are numbered sequentially as follows:

- **[R-01]** indicates Reviewer’s comment 1.
- **[A-01]** indicates Authors’ response 1.

To facilitate the evaluation of the revisions, we have adopted the following notation throughout this response letter. Figures that appear in the revised manuscript are referred to as “Figure 1”, “Figure 2”, etc. Figures that are provided only in this response letter are denoted as “R1”, “R2”, etc., while figures included in the Supplementary Material are denoted as “S1”, “S2”, etc. For ease of reference, the textual changes made in the revised manuscript are highlighted in blue font within this response letter.

We hope that our responses and the corresponding revisions adequately address the reviewer’s comments and further improve the quality and presentation of the manuscript.

[R-01] *The current title emphasizes the "impact of satellite observations", which suggests that the manuscript evaluates whether satellite observations improve inversion performance. However, the main contributions are actually development of the FEISSO framework, sensitivity analysis of inversion parameters, and recommended parameter settings for urban inversions. The title should better reflect these methodological contributions.*

[A-01] Thank you for pointing out this helpful suggestion. We agree that the original title placed undue emphasis on the “impact of satellite observations” and could therefore imply that the primary objective was to quantify the improvement in inversion performance attributable to satellite observations. The main contributions of this study are instead the development of the FEISSO-Carbon inversion framework, the implementation of a satellite-compatible observation operator for assimilating OCO-2 XCO_2 observations, and the systematic sensitivity analysis of key inversion parameters for urban-scale applications. We have therefore revised the title to:

"FEISSO-Carbon: A Lagrangian–Bayesian inversion framework for high-resolution urban CO_2 flux inversion using OCO-2 XCO_2 observations."

The revised title more accurately reflects the methodological focus and scope of the study.

[R-02] *The posterior emissions are primarily evaluated against EDGAR and local emission inventories. However, EDGAR (combined with ODIAC spatial allocation) is also used to construct the prior emissions. Consequently, agreement between the posterior estimates and EDGAR cannot be regarded as an independent validation. More importantly, the Chengdu posterior emissions are approximately 190% of the EDGAR estimate, which differs substantially from the prior inventory, while the abstract still concludes that the inversion shows "overall consistency". This statement appears overstated, particularly under complex terrain and sparse observational coverage. I strongly recommend including independent validation datasets, such as ground-based CO_2 or XCO_2 observations, urban CO_2 monitoring stations, or at least cross-validation using withheld OCO-2 observations. Without independent evaluation, it remains difficult to assess whether the* *inversion genuinely improves emission estimates.*

[A-02] Thank you for this important suggestion. Following the reviewer's recommendation, we conducted a 10-fold cross-validation for Weifang, Chengdu, and Xining. In each fold, 90 % of the available OCO-2 observations were assimilated, while the remaining 10 % were withheld for evaluation. The cross-validated simulations were evaluated using Bias, MAE, and RMSE.

Relative to the prior simulations, the cross-validated MAE decreased by 9.6 %, 46.0 %, and 15.3 % in Weifang, Chengdu, and Xining, respectively, while the RMSE decreased by 7.9 %, 43.2 %, and 20.2 %. The magnitude of the bias was also reduced in Weifang and Xining and remained small in Chengdu. These results demonstrate improved agreement with the withheld XCO_2 observations across all three cities. The observed range of improvement is consistent with previous urban CO_2 inversion studies, where residual errors may also reflect uncertainties in atmospheric transport and background representation (Ohyama et al., 2023; Sim and Jeong, 2025). We have added the cross-validation procedure to the Methods, presented the results in the new Table 5, and revised Sect. 4.2 accordingly. This analysis provides an observation-based assessment of the out-of-sample performance of FEISSO-Carbon and complements the comparisons with the available emission inventories.

Lines 566:

Table 5 Statistical evaluation of prior, posterior, and cross-validated XCO_2 simulations against OCO-2 observations for the three study cities.

City	BIAS(ppm)			MAE(ppm)			RMSE(ppm)		
	prior	posterior	cross-validation	prior	posterior	cross-validation	prior	posterior	cross-validation
Weifang	-0.6059	-0.3886	-0.5729	0.8034	0.6836	0.7263	0.9698	0.8447	0.8929
Chengdu	0.0244	-0.5848	0.0914	1.5425	1.0902	0.8336	2.0031	1.6021	1.1372
Xining	0.3163	0.2716	0.2177	1.0301	1.0038	0.8722	1.3489	1.3196	1.077

"A 10-fold cross-validation was subsequently conducted by assimilating 90 % of the available observations in each fold and evaluating the simulations against the remaining 10 %. Relative to the prior simulations, the cross-validated MAE decreased by 9.6 %, 46.0 %, and 15.3 % in Weifang, Chengdu, and Xining, respectively, while the RMSE decreased by 7.9 %, 43.2 %, and 20.2 %. The magnitude of the bias was also reduced in Weifang and Xining. In Chengdu, it increased slightly from 0.0244 to 0.0914 ppm but remained small relative to the MAE and RMSE. Overall, the consistent reductions in MAE and RMSE indicate improved predictive consistency for the observations withheld from the corresponding inversion folds. The more modest reductions in Weifang are consistent with previous urban CO_2 inversion studies, in which residual errors were also influenced by uncertainties in atmospheric transport and background representation (Ohyama et al., 2023; Sim and Jeong, 2025). Together, these results provide an observation-based assessment of the out-of-sample performance of FEISSO-Carbon and complement the contextual comparisons with the available emission inventories."

[R-03] *The study is based on very limited OCO-2 observations. Within the selected 10-day assimilation window. Weifang uses three satellite overpasses. Xining uses three overpasses. Chengdu relies on only one OCO-2 track. A 10-day inversion based on such limited observations cannot necessarily represent monthly, seasonal, or annual urban emissions. The limitations associated with observational coverage should be discussed more thoroughly. Therefore, conclusions regarding the scalability of FEISSO and its applicability for policy evaluation appear* *insufficiently supported.*

[A-03] We thank the reviewer for highlighting this important limitation. We agree that the limited OCO-2 sampling within the selected 10-day periods, comprising three overpasses for Weifang and Xining and only one for Chengdu, provides limited observational constraints on the posterior emissions and restricts their temporal representativeness.

The 10-day periods were selected to evaluate the implementation of FEISSO-Carbon and the sensitivity of key inversion parameters under the available satellite sampling, rather than to characterize monthly, seasonal, or annual urban emissions. The posterior estimates therefore represent only the selected assimilation periods. Because the available local statistical inventories provide annual totals, these totals were proportionally scaled to 10-day equivalents solely for comparison on a common timescale.

In response to this comment, we have expanded the discussion of observational coverage and revised the Abstract, Discussion, and Conclusions to remove unsupported claims regarding the scalability of FEISSO-Carbon and its current applicability to policy evaluation. The revised manuscript now frames the study as a methodological evaluation based on selected short-term periods with limited OCO-2 sampling. We also clarify that evaluating longer-term urban emissions will require denser and more frequent observations, inversions covering multiple periods, complementary observing platforms, and further independent validation.

Lines 318:

"All experiments were conducted using a fixed 10-day inversion window, which balanced OCO-2 observation availability with the temporal consistency of the atmospheric transport simulations. This experimental design was used to evaluate the implementation of FEISSO-Carbon and its sensitivity to key inversion parameters using real satellite observations. The corresponding results are presented in Sect. 3."

5 Conclusion (Lines 680-688):

"Using the selected reference configurations, FEISSO-Carbon estimated 10-day anthropogenic CO_2 emissions of 3.09×10^6, 2.19×10^6, and 0.57×10^6 t CO_2 for Weifang, Chengdu, and Xining, respectively. The posterior estimates for Weifang and Xining were comparable in magnitude to the corresponding 10-day values derived from EDGAR and local statistical inventories. The 10-fold cross-validation further showed improved agreement between the simulated and withheld OCO-2 XCO_2 observations across all three cities. The hotspot analysis for Weifang also demonstrated that, when point-source locations are adequately represented in the prior emissions, FEISSO-Carbon can better characterize localized industrial emission hotspots and adjust their emission magnitudes under satellite observational constraints.

Overall, this study establishes FEISSO-Carbon as a Lagrangian–Bayesian framework for assimilating satellite XCO_2 observations in spatially resolved urban CO_2 flux inversions and provides practical guidance for parameter specification under different urban and observational conditions. The current applications are limited to selected 10-day periods, and their temporal representativeness is constrained by the spatial and temporal coverage of OCO-2. Denser satellite sampling, higher revisit frequency, improved retrieval accuracy, and complementary observations would support applications over longer periods and a wider range of urban regions."

Abstract. (Lines 31-33):

"FEISSO-Carbon provides a satellite-compatible framework for spatially resolved urban CO_2 flux inversion. With the increasing availability of high-frequency satellite observations and complementary monitoring networks, this framework has the potential to support more frequent and comprehensive assessments of urban carbon emissions in the future."

[R-04] *Background concentrations are derived from GEOS-Chem and corrected using AirCore observations, with three pressure levels (510, 139, and 48 hPa) used to define the background. However, the manuscript does not explain why these pressure levels are universally appropriate for different cities, seasons, and topographic conditions. No sensitivity analysis is performed to quantify how background uncertainty propagates into the posterior emissions. Since background uncertainty can be comparable to the urban enhancement signal in XCO_2 inversions, this represents one of the dominant uncertainty sources and deserves much more detailed discussion.*

[A-04] Thank you for raising this important point. We agree that the representation of background CO_2 is important in urban XCO_2 inversions because the emission-related enhancement is superimposed on the much larger atmospheric CO_2 baseline.

In FEISSO-Carbon, the background component represents the long-lived atmospheric CO_2 baseline that is separated from the short-term concentration variations caused by surface fluxes during the backward simulation period. This definition follows the background formulation of FLEXINVERT (Thompson and Stohl, 2014; Thompson et al., 2015) and differs from the initial and lateral boundary conditions used to introduce large-scale atmospheric CO_2 into Eulerian regional models (He et al., 2018; Chen et al., 2023).

The pressure levels of 510.475, 139.115, and 48.282 hPa were adopted from the FLEXINVERT background sampling strategy. Rather than defining universal boundaries for individual cities, seasons, or topographic settings, these levels provide vertical sampling coordinates in the middle and upper atmosphere, where short-term variations associated with surface emissions are substantially weaker. Figure R1 shows that the pronounced variability in the lower-atmospheric GEOS-Chem CO_2 profiles decreases with altitude. Although the sampling levels are fixed, the corresponding CO_2 mixing ratios vary spatially and temporally in the continuous GEOS-Chem simulations.

To improve the accuracy of the background fields, the GEOS-Chem profiles were calibrated using AirCore profiles from NOAA GML covering January 2012 to July 2021. The mean GEOS-Chem–AirCore differences at 510.475, 139.115, and 48.282 hPa were 0.34, 0.69, and 3.5 ppm, respectively, and were corrected in the background profiles. This observational calibration reduces systematic biases while retaining the spatial and temporal variability represented by GEOS-Chem. Although a separate propagation experiment for the remaining background uncertainty was not included in the present parameter sensitivity analysis, the multi-year simulations, vertical-profile evaluation, and AirCore calibration provide observational constraints on the background fields used in the inversion. We have revised Sect. 2.1.4 to clarify this background definition, the rationale for the selected pressure levels, and the calibration procedure.

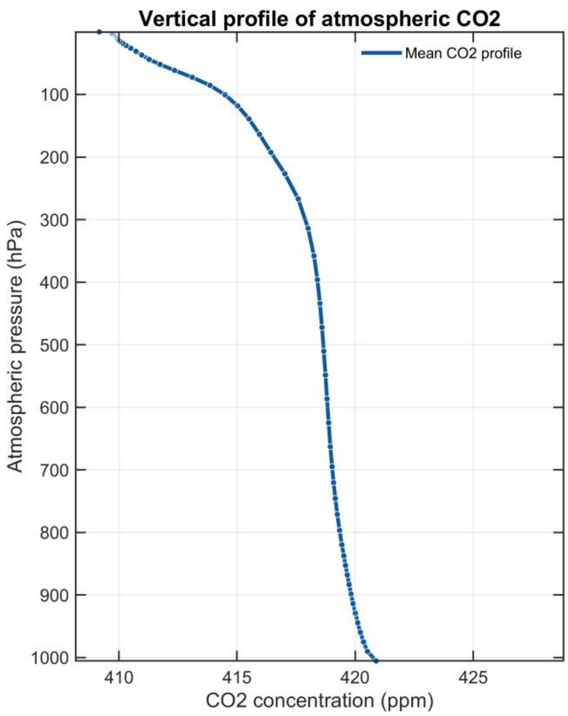

Figure R1. Mean vertical profile of GEOS-Chem-derived background CO_2 concentrations

2.1.4 Correction of background CO2 concentration fields (Lines 233-245):

"Since Lagrangian atmospheric transport simulations trace air parcels backward over a finite period of several days to weeks, the modeled atmospheric CO_2 concentration is separated into recent contributions from surface fluxes and a large-scale background component. Following FLEXPART–FLEXINVERT+, the concentration comprises (1) recent contributions from emissions within the urban inversion domain, (2) recent contributions from emissions outside the urban inversion domain, and (3) a large-scale background component representing atmospheric CO_2 variations beyond the timescale of the backward simulation. Unlike the initial and lateral boundary conditions used in Eulerian regional models, this background component represents the large-scale atmospheric CO_2 contribution separated from short-term surface-flux contributions.

The three-dimensional background CO_2 fields were obtained from continuous multi-year GEOS-Chem simulations and calibrated using 84 AirCore vertical profiles from NOAA GML (https://gml.noaa.gov/ccgg/aircore/) collected between January 2012 and July 2021. Following the FLEXPART–FLEXINVERT+ background sampling strategy (Thompson and Stohl, 2014; Thompson et al., 2015), CO_2 mixing ratios were extracted at 510.475, 139.115, and 48.282 hPa. These levels extend from the middle troposphere to the lower stratosphere, where short-term variations associated with surface emissions are weaker, while retaining the spatial and temporal variability represented by GEOS-Chem. The AirCore calibration further reduces systematic biases in the simulated vertical CO_2 profiles."

[R-05] *The manuscript identifies a turning point around ObsErr ≈ 1.5 ppm and GlobErr ≈ 8 Mt yr^{-1}, but does not explain the physical mechanisms responsible for this behavior.*

[A-05] Thank you for pointing out this insightful comment. We agree that the turning points identified in the sensitivity experiments require further explanation. These turning points reflect changes in the balance between observational constraints and prior information within the Bayesian inversion framework, particularly under the limited OCO-2 sampling conditions in

Chengdu.

Decreasing ObsErr initially increases the weight assigned to the OCO-2 observations and allows greater adjustment of the posterior emissions from the prior emissions. At lower ObsErr values, however, further increasing the observational weight does not increase the information content of the available observations. Consequently, the spatial adjustment of the posterior emissions remains strongly influenced by the prior error covariance structure, producing the observed change in the response of the aggregated emissions. Conversely, increasing ObsErr weakens the observational constraint and causes the posterior emissions to approach the prior emissions.

GlobErr represents the inversion-domain total flux uncertainty. Increasing GlobErr initially relaxes the constraint on the total flux adjustment and allows the posterior emissions to deviate further from the prior emissions. Beyond the turning point, further increases in GlobErr expand the allowable adjustment without providing additional observational information. The posterior response therefore remains strongly influenced by the prior error covariance structure, resulting in the observed non-monotonic variation.

We have added this explanation to the revised manuscript.

Lines 395-403:

"In Chengdu, the limited OCO-2 sampling during the selected inversion period weakens the observational constraints on the inversion. Persistent cloud cover limits the number of valid retrievals, while complex terrain and meteorological conditions affect atmospheric transport. The posterior emissions consequently exhibit non-monotonic responses to ObsErr and GlobErr, with turning points near 1.5 ppm and 8 Mt yr^{-1}, respectively. Decreasing ObsErr initially strengthens the observational constraint and allows greater adjustment from the prior emissions, whereas increasing GlobErr relaxes the constraint imposed by the inversion-domain total flux uncertainty. However, excessively small ObsErr or excessively large GlobErr values cannot be effectively supported by the limited observational information, causing the posterior solution to become increasingly influenced by prior information. Conversely, increasing ObsErr or decreasing GlobErr strengthens the relative influence of the prior estimate. The non-monotonic responses therefore reflect changes in the balance between observational constraints and prior information. "

[R-06] *The opposite responses observed in Weifang and Xining require further explanation. Section 3.2 states that Weifang and Xining exhibit opposite responses to changes in ObsErr and GlobErr, yet the underlying physical reasons are not discussed. A mechanistic explanation, considering emission patterns, atmospheric transport, and inversion constraints, would improve the interpretation.*

[A-06] Thank you for pointing out this valuable comment. We agree that the opposite responses observed in Weifang and Xining require further explanation.

Within the Bayesian inversion, decreasing ObsErr increases the weight assigned to the OCO-2 observations, while increasing GlobErr relaxes the constraint imposed by the inversion-domain total flux uncertainty. The direction of the resulting adjustment depends on whether the OCO-2 constraints support an increase or decrease relative to the prior emissions. The prior spatial emission patterns and city-specific source – receptor relationships determine how these observational constraints are propagated to the posterior emissions.

In Weifang, the OCO-2 constraints support an upward adjustment from the prior emissions. Therefore, decreasing ObsErr or increasing GlobErr leads to higher posterior emissions. In Xining, the OCO-2 constraints support a downward adjustment from the prior emissions, resulting in lower posterior emissions under the same parameter changes. The opposite responses therefore arise from the different directions of the observation-driven adjustments relative to the prior emissions.

We have revised Sect. 3.2 to clarify this mechanism.

Lines 391-395:

"In Weifang and Xining, the posterior emissions exhibit opposite responses to changes in ObsErr and GlobErr. Decreasing ObsErr strengthens the observational constraint, while increasing GlobErr relaxes the constraint imposed by the inversion-domain total flux uncertainty. Through the prior spatial emission patterns and city-specific source–receptor relationships, the OCO-2 constraints support an upward adjustment from the prior emissions in Weifang but a downward adjustment in Xining, resulting in opposite responses of the posterior emissions."

[R-07] *The proposed "seesaw effect" is interesting but insufficiently explained. The authors should further explain why the seesaw effect occurs, under what atmospheric or inversion conditions it appears, and whether it is expected to be generally applicable or only specific to the selected case studies.*

[A-07] Thank you for pointing out this insightful comment. We agree that the mechanism, occurrence conditions, and applicability of the "seesaw effect" require further clarification.

The "seesaw effect" refers to the compensating redistribution of posterior fluxes among different flux magnitude ranges when inversion parameters alter the balance between observational constraints and prior information. Decreasing ObsErr strengthens the influence of the OCO-2 observations, while increasing GlobErr relaxes the constraint imposed by the inversion-domain total flux uncertainty. Because OCO-2 XCO_2 observations constrain integrated atmospheric enhancements rather than individual flux grid cells directly, the posterior flux adjustments are spatially distributed through the source–receptor relationships and the prior error covariance structure. Consequently, increases in some flux ranges are accompanied by reductions in others, producing the observed seesaw pattern.

The three cities show distinct manifestations of this redistribution. Weifang is a plain city dominated by anthropogenic emissions and has greater OCO-2 coverage than Chengdu. As ObsErr decreases and GlobErr increases, the frequency of high-emission fluxes (10^{-5}–10^{-4} kg m^{-2} s^{-1}) increases, while that of low-emission fluxes (10^{-13}–10^{-11} kg m^{-2} s^{-1}) decreases. Chengdu is located in a basin with complex terrain and has only one OCO-2 overpass during the selected period. Its redistribution mainly occurs in the high-emission range (approximately 10^{-5} kg m^{-2} s^{-1}) and the weak-flux range ($<10^{-10}$ kg m^{-2} s^{-1}), with a more variable response under limited observational constraints. Xining is a high-altitude valley city with weaker anthropogenic emissions and a greater contribution from terrestrial ecosystem fluxes. Changes in ObsErr and GlobErr have little influence on its peak emission frequency but more strongly modify the distribution of CO_2 uptake fluxes.

The occurrence of compensating redistribution across these contrasting urban environments suggests that the seesaw effect is not restricted to a single city. Its specific manifestation depends on the local flux composition, source–receptor relationships, observational coverage, and prior error covariance structure. We have expanded Sect. 3.3 to clarify this mechanism and the differences among the three cities.

Lines 442-479:

"Among the five inversion parameters, ObsErr and GlobErr exert the strongest influence on the 10-day cumulative posterior CO_2 emissions, consistent with the results in Sect. 3.2. Decreasing ObsErr strengthens the constraint provided by the OCO-2 observations, whereas increasing GlobErr relaxes the constraint imposed by the inversion-domain total flux uncertainty. Because OCO-2 XCO_2 observations constrain integrated atmospheric CO_2 signals rather than individual flux grid cells, posterior adjustments are distributed spatially through the source–receptor relationships and prior error covariance structure. Increases in some flux ranges are therefore accompanied by reductions in others. This compensating redistribution is referred to here as the "seesaw effect".

The seesaw effect differs among the three cities. In Weifang, a plain city dominated by anthropogenic emissions, decreasing ObsErr and increasing GlobErr increase the frequency of high-emission fluxes (10^{-5}–10^{-4} kg m^{-2} s^{-1}) and decrease the frequency of low-emission fluxes (10^{-13}–10^{-11} kg m^{-2} s^{-1}). In Chengdu, a basin city with complex terrain and limited OCO-2 sampling, the redistribution occurs mainly in the high-emission range (approximately 10^{-5} kg m^{-2} s^{-1}) and the weak-flux range ($<10^{-10}$ kg m^{-2} s^{-1}), and its response is more variable than that in Weifang. In Xining, a high-altitude valley city with weaker anthropogenic emissions and a greater contribution from terrestrial ecosystem fluxes, the peak frequency of positive fluxes changes little, whereas the uptake flux distribution is more strongly modified. The compensating redistribution observed in all three cities indicates a common adjustment behavior within the inversion framework, although the affected flux ranges and magnitude of redistribution vary with the local flux composition, source–receptor relationships, observational coverage, and prior error covariance structure.

The remaining parameters produce different spatial adjustments. FlxErr mainly affects weak-emission fluxes, particularly those below 10^{-9} kg m^{-2} s^{-1}. In Weifang, increasing FlxErr decreases the frequency of fluxes near 10^{-11} kg m^{-2} s^{-1} and increases the frequency of lower fluxes between 10^{-13} and 10^{-12} kg m^{-2} s^{-1}, indicating redistribution among weak-emission flux ranges. SigLand affects the spatial smoothness of the posterior fluxes through the land spatial correlation structure. In Weifang, increasing SigLand gradually decreases the frequency of high-emission fluxes. In Chengdu, a shorter SigLand of 10 km shifts part of the distribution from weak uptake (-10^{-11} to -10^{-15} kg m^{-2} s^{-1}) to weak emissions (10^{-16}–10^{-14} kg m^{-2} s^{-1}). SigTime affects the temporal allocation of posterior emissions by modifying the temporal correlation structure of the prior flux error covariance. Its influence on the spatial flux distribution is relatively limited in Weifang and Xining but more pronounced in Chengdu, where the available OCO-2 observations are limited to a single overpass."

[R-08] *Important FLEXPART configuration details are missing, including the number of released particles, backward integration time, and the vertical release heights. Since these settings directly affect the source-receptor relationships, they should be explicitly documented to ensure reproducibility.*

[A-08] Thank you for pointing out the missing FLEXPART configuration details. We agree that the particle number, backward integration time, and vertical release levels should be explicitly documented because these settings affect the calculated source–receptor relationships (SRRs).

We have added the complete FLEXPART configuration to Sect. S1 and Table S1 of the

Supplement. For each OCO-2 observation, the FLEXPART simulation was integrated backward for 7 days. A total of 5000 particles were released from each of the 20 vertical levels corresponding to the OCO-2 retrieval pressure grid, resulting in 100,000 particles per observation. The vertical releases were specified according to the pressure levels of the corresponding OCO-2 observation rather than a common set of fixed geometric heights. The vertically resolved SRRs were subsequently combined using the OCO-2 averaging kernel and pressure weighting function. A reference to these settings has also been added to Sect. 2.1 of the revised manuscript.

Lines 171-172:

"Detailed FLEXPART settings are provided in Sect. S1 and Table S1 of the Supplement."

S1 FLEXPART configuration

FLEXPART version 10.4 was operated in backward mode to calculate the source–receptor relationships (SRRs) between each OCO-2 observation and the upwind surface fluxes. Each simulation was integrated backward for 7 days from the time and horizontal location of the corresponding OCO-2 observation.

To represent the vertical sensitivity of the satellite column observation, particles were released from 20 vertical levels corresponding to the pressure grid of the OCO-2 XCO_2 retrieval. Because the retrieval pressure grid is determined for each OCO-2 observation, the vertical release levels were specified in pressure coordinates rather than as a common set of fixed geometric heights. A total of 5000 particles were released from each vertical level, resulting in 100,000 particles for each OCO-2 observation. FLEXPART calculated an SRR separately for each vertical level. These vertically resolved SRRs were subsequently combined using the OCO-2 averaging kernel and pressure weighting function to construct the satellite-compatible column SRR used in FEISSO-Carbon. The principal FLEXPART settings are summarized in Table S1.

Table S1. FLEXPART configuration used to calculate the SSRs for OCO-2 observations.

Configuration item	Setting
FLEXPART version	10.4
Simulation direction	Backward
Receptor location	Location of each OCO-2 observation
Receptor time	Observation time of each OCO-2 observation
Backward integration time	7 days
Number of vertical release levels	20
Vertical coordinate	Pressure coordinates corresponding to the OCO-2 retrieval pressure grid
Particles released per vertical level	5000
Total particles per observation	100,000
SRR calculation	Calculated separately for each vertical level
Column integration	OCO-2 averaging kernel and pressure weighting function

[R-09] *The manuscript concludes that high-resolution meteorological fields are critical for urban CO_2 source attribution. However, FLEXPART is still driven by meteorological fields at 0.25° resolution. Interpolating these fields onto a 0.01° inversion grid does not fundamentally increase the effective meteorological resolution. Therefore, the statement should be moderated, particularly for kilometer-scale urban inversions where sub-kilometer meteorological variability is important.*

[A-09] Thank you for pointing this out. We have revised the corresponding statements to distinguish the resolution of the meteorological forcing from that of the inversion grid. The revised manuscript clarifies that the 0.01° grid defines the spatial discretization of the flux state vector, whereas atmospheric transport is driven by the 0.25° meteorological fields. Accordingly, we have moderated the conclusion regarding the role of meteorological resolution in kilometre-scale urban inversions.

2.1.1 FLEXPART (Lines 165-179):

"FEISSO-Carbon uses FLEXPART version 10.4 (https://www.flexpart.eu/) to perform backward Lagrangian particle simulations and calculate source–receptor relationships (SRRs), which describe the sensitivity of atmospheric CO_2 observations to upwind surface fluxes (Thomson, 1987; Pisso et al., 2019). The simulations were driven by NCEP FNL meteorological fields at 0.25° × 0.25° resolution. Particle trajectories and residence times were calculated by accounting for advection, turbulent diffusion, and convection, and the resulting surface sensitivities were accumulated on the 0.01° × 0.01° urban inversion grid to construct the SRRs (Stohl, 1998; Stohl et al., 2005; Pisso et al., 2019). Detailed FLEXPART settings are provided in Sect. S1 and Table S1 of the Supplement.

Following the FLEXPART–FLEXINVERT+ nested inversion framework, the modeled atmospheric CO_2 concentration consists of contributions from the nested urban inversion domain, the global outer domain, and the long-term background component. The outer domain represents large-scale atmospheric transport and contributions from outside the urban inversion domain, while the nested SRRs link each OCO-2 XCO_2 observation to multiple upwind flux grid cells. Posterior fluxes were optimized on the 0.01° × 0.01° urban inversion grid, which defines the spatial discretization of the optimized flux state vector."

[R-10] *The manuscript states that "A super observation is defined as the median of all observations within one standard deviation in each 1 km × 1 km grid cell." However, the native OCO-2 sounding footprint (~2.25 km × 1.3 km) is larger than the aggregation grid. Please clarify how individual soundings are assigned to the 1 km grid cells.*

[A-10] We thank the reviewer for this important comment. We agree that the original description did not clearly explain how OCO-2 soundings were assigned to the 0.01° × 0.01° aggregation grid. Each OCO-2 sounding was assigned to the grid cell containing the center coordinate of its retrieved XCO_2 footprint. The footprint itself was neither subdivided among multiple grid cells nor spatially interpolated onto the inversion grid. This grid assignment was used only to identify soundings for the super-observation procedure; the actual sounding geolocation was retained for the atmospheric transport simulation.

When multiple valid soundings were assigned to the same grid cell, observations outside one standard deviation were excluded and the median of the remaining observations was used as the super observation. When a grid cell contained only one valid sounding, that sounding was retained without aggregation. Because most grid cells in the selected OCO-2 datasets contained only one valid sounding, the super-observation procedure had a limited effect on the number of assimilated observations.

We have revised the manuscript to clarify the center-based assignment procedure and to state explicitly that the OCO-2 footprints were not subdivided into the 0.01° inversion grid cells.

Lines 307-311:

"Each valid OCO-2 sounding was assigned to the 0.01° × 0.01° inversion grid cell containing the center coordinate of its retrieval footprint. When multiple soundings occurred within the same grid cell, soundings with XCO_2 values within one standard deviation of the cell mean were retained, and their median was used as the super observation. A single sounding within a grid cell was retained directly. As most grid cells contained only one valid sounding, super-observation aggregation was limited under the OCO-2 sampling conditions of this study. After preprocessing, 829, 311, and 676 OCO-2 observations were assimilated for Weifang (35.7–37.4° N, 118.1–120.1° E), Chengdu (30.05–31.26° N, 102.54–104.53° E), and Xining (36.27–36.56° N, 101.18–101.52° E), respectively."

[R-11] *The manuscript concludes that Flux Err has relatively limited influence. However, the tested range of 1–30% appears relatively narrow for high-resolution urban inventories, especially considering the large uncertainties in spatial allocation. Additional experiments using larger prior uncertainties (e.g., 50–100%), would provide stronger support for this conclusion.*

[A-11] Thank you for pointing out this valuable comment. We agree that our original description of FlxErr as a prior flux uncertainty was not sufficiently precise and could be interpreted as representing the overall uncertainty of the gridded emission inventory. In FLEXINVERT, however, FlxErr is a parameter used in constructing the initial prior error covariance matrix rather than a direct estimate of the uncertainty of individual emission grid cells or of the inventory as a whole. The effective prior uncertainty used in the inversion is jointly determined by FlxErr, GlobErr, the prior flux distribution, and the prescribed spatial and temporal correlation structure.

The tested range of FlxErr (1–30%) was therefore designed to evaluate the sensitivity of the inversion to this covariance-construction parameter, rather than to represent the full range of uncertainties in high-resolution urban emission inventories. We have revised the manuscript to clarify this definition and removed wording that could imply that FlxErr directly represents grid-scale inventory uncertainty. We have also limited the corresponding conclusion to state that, within the tested range, variations in FlxErr have a relatively limited influence on the posterior emissions compared with ObsErr and GlobErr.

Lines 324-335:

"In the system parameter sensitivity analysis, five key inversion parameters were investigated: the observation error (ObsErr), inversion-domain total flux uncertainty (GlobErr), prior flux error parameter (FlxErr), land spatial correlation length (SigLand), and temporal correlation length (SigTime). Each parameter was varied individually while the remaining parameters were held at their baseline values. The tested ranges were 0.5–5 ppm for ObsErr, 1%–25% of annual city-level CO_2 emissions for GlobErr, 1%–30% for FlxErr, 10–50 km for SigLand, and 1–21 d for SigTime (Table 2).

Within the FLEXINVERT+ framework, FlxErr is used to construct the initial prior error covariance matrix, whereas GlobErr specifies the inversion-domain total flux uncertainty used to scale this matrix (Thompson and Stohl, 2014). The tested FlxErr range therefore evaluates the sensitivity of the inversion to this parameter in the prior error covariance construction rather than representing the full uncertainty range of high-resolution emission inventories. The baseline values were ObsErr = 1 ppm, FlxErr = 10%, SigLand = 25 km, and SigTime = 7 d. GlobErr was determined according to the city-specific emission characteristics."

[R-12] *As currently presented, Table 2 gives the impression that only six experiments were* *conducted while simultaneously varying all five parameters. However, the manuscript later indicates that each parameter was perturbed independently while all remaining parameters were fixed. I recommend revising the table layout or adding an explicit note to avoid misunderstanding.*

[A-12] Thank you for identifying this potential ambiguity in the presentation of Table 2. We agree that the original table layout could be misinterpreted as indicating that all five parameters were simultaneously varied in each experiment.

In the sensitivity analysis, each inversion parameter was perturbed independently, while all other parameters were fixed at their baseline values. The experiments listed in Table 2 represent the tested ranges for each individual parameter rather than combined parameter configurations.

Following the reviewer's suggestion, we have revised Table 2 to clearly separate the sensitivity experiments for different parameters and added an explanatory note indicating that each parameter was varied independently. The revised table now avoids the impression that Exp 1–6 correspond to simultaneous changes in all inversion parameters.

Table 2 Experimental Design Table for System Parameter Sensitivity Analysis

Spatial resolution of meteorological fields		1°×1°	0.25°×0.25°						
			baseline	Exp 1	Exp 2	Exp 3	Exp 4	Exp 5	Exp 6
ObsErr (ppm)		1.0	**1.0**	0.5	0.7	1.0	2.0	3.0	5.0
GlobErr(% of a priori annual total emissions)	Weifang	15%	**15%**	1%	5%	10%	15%	20%	25%
	Chengdu	15%	**15%**	1%	5%	10%	15%	20%	25%
	Xining	10%	**10%**	1%	5%	10%	15%	20%	25%
FlxErr		10%	**10%**	1%	5%	10%	15%	20%	30%
SigLand(km)		25.0	**25.0**	10.0	15.0	20.0	25.0	30.0	50.0
SigTime(day)		7	**7**	1	3	7	10	14	21

Note: Each parameter was varied independently, while all other parameters were fixed at their baseline values.

Lines 332:

"Each parameter was varied individually while the remaining parameters were held at their baseline values. The tested ranges were 0.5–5 ppm for ObsErr, 1%–25% of annual city-level CO_2 emissions for GlobErr, 1%–30% for FlxErr, 10–50 km for SigLand, and 1–21 d for SigTime (Table 2)."

[R-13] *The manuscript attributes hotspot displacement primarily to ODIAC's nighttime-light-based spatial allocation.In fact, ODIAC does not allocate all emissions using nighttime lights. Large power plants are represented using dedicated global power plant databases (e.g., CARMA), where locations are generally well constrained. Hotspot displacement originates from inaccuracies in the spatial structure of the prior inventory, while nighttime-light allocation is only one contributing factor. In addition, Bayesian inversion mainly adjusts emission magnitudes rather than relocating emission hotspots, allowing prior spatial biases to propagate into posterior estimates.*

[A-13] Thank you for pointing out this important clarification. We agree that the original manuscript overly attributed the displacement of emission hotspots to the nighttime-light-based spatial allocation in ODIAC. ODIAC uses dedicated point-source databases to locate large power plants, whereas nighttime-light data are primarily used to spatially distribute non-point-source emissions. Nighttime-light-based allocation is therefore only one possible source of uncertainty in the prior spatial emission pattern.

We have revised the manuscript to attribute the observed hotspot displacement more generally to inaccuracies in the spatial structure of the prior emission inventory. We have also clarified that the Bayesian inversion updates gridded emission magnitudes according to the observational constraints and prior error covariance structure but does not explicitly relocate emission sources. Consequently, spatial biases in the prior inventory may be retained in the posterior flux distribution. The corresponding discussion has been revised in Sect. 4.3.

Lines 618-622:

"These discrepancies are mainly associated with inaccuracies in the spatial structure of the prior emissions. The ODIAC spatial distribution combines point-source locations from dedicated databases for large power plants with proxy-based allocation for other fossil-fuel emissions. For emissions distributed using nighttime-light data, the allocation may not fully represent the locations and relative magnitudes of industrial sources and may therefore contribute to the displacement of emission hotspots. Because the Bayesian inversion primarily adjusts gridded emission magnitudes according to the observational constraints and prior error covariance structure rather than independently relocating emission sources, spatial biases in the prior emissions may propagate into the posterior flux distribution. "

References

Chen, H. W., Zhang, F., Lauvaux, T., Scholze, M., Davis, K. J., and Alley, R. B.: Regional CO2 Inversion Through Ensemble-Based Simultaneous State and Parameter Estimation: TRACE Framework and Controlled Experiments, Journal of Advances in Modeling Earth Systems, 15, e2022MS003208, https://doi.org/10.1029/2022MS003208, 2023.

He, W., van der Velde, I. R., Andrews, A. E., Sweeney, C., Miller, J., Tans, P., van der Laan-Luijkx, I. T., Nehrkorn, T., Mountain, M., Ju, W., Peters, W., and Chen, H.: CTDAS-Lagrange v1.0: a high-resolution data assimilation system for regional carbon dioxide observations, Geosci. Model Dev., 11, 3515–3536, 10.5194/gmd-11-3515-2018, 2018.

Ohyama, H., Frey, M. M., Morino, I., Shiomi, K., Nishihashi, M., Miyauchi, T., Yamada, H., Saito, M., Wakasa, M., Blumenstock, T., and Hase, F.: Anthropogenic CO2 emission estimates in the Tokyo metropolitan area from ground-based CO2 column observations, Atmos. Chem. Phys., 23, 15097–15119, 10.5194/acp-23-15097-2023, 2023.

Pisso, I., Sollum, E., Grythe, H., Kristiansen, N. I., Cassiani, M., Eckhardt, S., Arnold, D., Morton, D., Thompson, R. L., Groot Zwaaftink, C. D., Evangeliou, N., Sodemann, H., Haimberger, L., Henne, S., Brunner, D., Burkhart, J. F., Fouilloux, A., Brioude, J., Philipp, A., Seibert, P., and Stohl, A.: The Lagrangian particle dispersion model FLEXPART version 10.4, Geosci. Model Dev., 12, 4955–4997, 10.5194/gmd-12-4955-2019, 2019.

Sim, S. and Jeong, S.: Constraining urban fossil fuel CO2 emissions in Seoul using combined ground and satellite observations with Bayesian inverse modelling, Atmos. Chem. Phys., 25, 18509–18526, 10.5194/acp-25-18509-2025, 2025.

Stohl, A.: Computation, accuracy and applications of trajectories—A review and bibliography, Atmospheric Environment, 32, 947–966, https://doi.org/10.1016/S1352-2310(97)00457-3, 1998.

Stohl, A., Forster, C., Frank, A., Seibert, P., and Wotawa, G.: Technical note: The Lagrangian particle dispersion model FLEXPART version 6.2, Atmos. Chem. Phys., 5, 2461–2474, 10.5194/acp-5-2461-2005, 2005.

Thompson, R. L. and Stohl, A.: FLEXINVERT: an atmospheric Bayesian inversion framework for determining surface fluxes of trace species using an optimized grid, Geosci. Model Dev., 7, 2223–2242, 10.5194/gmd-7-2223-2014, 2014.

Thompson, R. L., Stohl, A., Zhou, L. X., Dlugokencky, E., Fukuyama, Y., Tohjima, Y., Kim, S.-Y., Lee, H., Nisbet, E. G., Fisher, R. E., Lowry, D., Weiss, R. F., Prinn, R. G., O'Doherty, S., Young, D., and White, J. W. C.: Methane emissions in East Asia for 2000–2011 estimated using an atmospheric Bayesian inversion, Journal of Geophysical Research: Atmospheres, 120, 4352–4369, https://doi.org/10.1002/2014JD022394, 2015.

Thomson, D. J.: Criteria for the selection of stochastic models of particle trajectories in turbulent flows, Journal of Fluid Mechanics, 180, 529–556, 10.1017/S0022112087001940, 1987.