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.

Major Comments

[R-01] *The authors claim that existing urban-scale inversions in China are heavily restricted to Observing System Simulation Experiments. This is factually incorrect. Numerous real-data inversion studies utilizing satellite and surface monitor networks have been published over major Chinese conglomerates. The literature review is inadequate.*

[A-01] Thank you for identifying the ambiguity in our literature review. The original statement did not adequately distinguish OSSEs from urban CO_2 inversions based on real atmospheric observations and therefore described the status of real-data inversion studies in China too broadly.

We have expanded the literature review to include studies that constrain urban and regional fossil-fuel CO_2 emissions using ground-based observations and regional carbon assimilation systems in China, including applications in Beijing and the Yangtze River Delta (Liu et al., 2024; Zhang et al., 2024). We have also cited international urban inversion studies using surface networks, ground-based column measurements, and satellite XCO_2 observations (Lian et al., 2022; Ohyama et al., 2023; Hamilton et al., 2024).

The revised manuscript now distinguishes the potential demonstrated by satellite-based OSSEs (Ye et al., 2020) from inversions conducted with real observations. The research gap has accordingly been defined more specifically: real-data frameworks that directly assimilate satellite XCO_2 observations to estimate spatially resolved urban CO_2 fluxes remain limited, particularly in the systematic assessment of inversion parameters, prior-flux uncertainties, and sparse satellite sampling.

Lines 95-99: "In China, real-data inversion studies have used surface observations to estimate urban and regional fossil-fuel CO_2 emissions (Liu et al., 2024; Zhang et al., 2024). Internationally, surface and ground-based column observations have supported urban inversions over Paris and Tokyo (Lian et al., 2022; Ohyama et al., 2023), while satellite XCO_2 retrievals have been assimilated over Los Angeles and the San Francisco Bay Area (Hamilton et al., 2024). Although OSSEs have demonstrated the potential of satellite observations to constrain urban fossil-fuel CO_2 emissions (Ye et al., 2020), high-resolution frameworks that directly assimilate satellite XCO_2 observations remain relatively limited in Chinese cities, particularly with respect to their systematic evaluation under sparse real satellite sampling. Therefore, developing and evaluating satellite-compatible urban CO_2 inversion frameworks under real observational conditions remains important."

[R-02] *Carbon flux inversion is a highly researched field, and the core models employed in this study are well-established, widely used community tools. Consequently, the specific scientific novelty of this manuscript remains unclear. The authors must explicitly articulate their original contributions, clearly detailing any methodological improvements, algorithmic advancements, or specific code developments they have made to these existing models.*

[A-02] We sincerely thank the reviewer for prompting us to clarify the specific contributions of FEISSO-Carbon. FLEXINVERT+ is an established Bayesian inversion framework and its current official implementation supports satellite observations, including OCO-2 XCO_2 (https://flexinvert.nilu.no/). FEISSO-Carbon builds on this framework but adopts a different implementation for satellite-based urban CO_2 flux inversion.

First, FEISSO-Carbon implements the satellite observation interface as an external module without modifying the core FLEXPART transport code. The module processes the satellite a priori CO_2 profile, averaging kernel, and pressure weighting function and combines them with the modeled vertical CO_2 contributions through column convolution. This modular structure maintains compatibility with the original FLEXPART code and allows other satellite products to be incorporated by adapting their retrieval inputs and convolution procedures. Second, we revised the calculation of the spatial and temporal correlation matrices in FLEXINVERT+ to reduce the computational burden associated with the large state vectors of kilometre-scale urban inversions while maintaining consistency with the original correlation formulation. FEISSO-Carbon further integrates multi-source prior fluxes within a nested global-to-urban configuration to represent spatially heterogeneous urban emissions and contributions from outside the urban inversion domain. These developments constitute the main methodological, computational, and code-level contributions of this study.

We have revised the Introduction and Sect. 2.1 to describe these contributions and distinguish them more clearly from the established capabilities of FLEXPART and FLEXINVERT+.

1.Introduction (Lines 125 – 129):

"In this study, we developed FEISSO-Carbon (Flux and Emission Inversion System for Satellite Observation), a modular extension of FLEXPART–FLEXINVERT+ for satellite-based urban CO_2 flux inversion. The system integrates a satellite-compatible column convolution operator without modifying the core FLEXPART code and reduces the computational burden of correlation-matrix calculations, thereby supporting efficient adaptation to satellite XCO_2 observations and spatially resolved urban flux estimation."

2.1 FEISSO-Carbon (Lines 141–148):

"FEISSO-Carbon was developed as a modular extension of the FLEXPART–FLEXINVERT+ Lagrangian–Bayesian inversion framework. An external satellite observation interface processes retrieval-specific information and links satellite XCO_2 observations to the modeled source–receptor relationships without modifying the core FLEXPART transport code. The column convolution operator combines the modeled vertical CO_2 contributions with the satellite a priori CO_2 profile, averaging kernel, and pressure weighting function to obtain satellite-consistent XCO_2 values. This modular implementation allows other satellite products to be incorporated by adapting their retrieval inputs and convolution procedures. The calculation of the spatial and temporal correlation matrices in FLEXINVERT+ was also revised to reduce the computational burden of kilometre-scale inversion grids while maintaining consistency with the original correlation structure. Based on these developments, FEISSO-Carbon integrates multi-source prior fluxes within a nested global-to-urban configuration. The modeled XCO_2 is separated into contributions from surface fluxes within the urban inversion domain, surface fluxes outside this domain, and the long-term atmospheric background. The following sections describe the transport simulation, prior fluxes, observation operator, and Bayesian inversion procedure."

[A-03] *In the methodological description, the authors artificially separate their "Bayesian framework" from EnKF and 4D-Var methods. Fundamentally, both EnKF and 4D-Var are mathematical implementations within the Bayesian inversion paradigm. The authors should avoid conceptual confusion and refrain from overstating the superiority of their method by mischaracterizing established assimilation frameworks.*

[R-03] Thank you for pointing this out. We have revised the Introduction to clarify their conceptual relationship.

Lines 112-124:

"Different mathematical implementations within the Bayesian inference framework are used to estimate posterior CO_2 fluxes. CarbonTracker (Peters et al., 2007), COLA (Liu et al., 2022), and GCAS v2 (Jiang et al., 2021) employ the ensemble Kalman filter, whereas CAMS (Chevallier et al., 2005; Remaud et al., 2018), THU (Kong et al., 2022), and GONGGA (Jin et al., 2023) use four-dimensional variational assimilation. Although these approaches differ in their mathematical formulations and uncertainty representations, they all combine prior information with observational constraints within a Bayesian framework. Bayesian synthesis inversion provides another implementation in which posterior CO_2 fluxes are estimated by combining prior flux information, observational constraints, and their associated uncertainties (Thompson and Stohl, 2014; Kunik et al., 2019; Nickless et al., 2019; Roten et al., 2023; Hamilton et al., 2024)."

[A-04] *The authors state that they used the 0.25° NCEP FNL operational analysis data to drive FLEXPART, while extracting SRRs at a refined spatial resolution of 0.1°. From a fluid dynamics perspective, this is highly problematic. Smooth 0.25° meteorological fields cannot resolve urban-scale localized circulations or complex micro-topography. This configuration is merely a high-frequency mathematical resampling (interpolation) of air parcel trajectories rather than a representation of high-resolution physical transport. Additionally, paradoxically, the authors state later that they ran the WRF model to simulate high-resolution (0.1°) hourly meteorological fields to drive their biospheric model (VPRM). Why was this fine-scale WRF meteorological output not directly utilized to drive FLEXPART?*

[R-04] We sincerely thank the reviewer for highlighting the need to distinguish meteorological forcing resolution from SRR output and flux inversion resolutions. The original description did not make this distinction sufficiently clear.

FEISSO-Carbon uses 0.25° NCEP FNL fields to drive FLEXPART and represent atmospheric transport across the global-to-urban domain. The 0.01° nested urban grid defines the spatial discretization of the SRRs and flux state vector, rather than the effective resolution of the meteorological forcing. The SRRs are calculated from the residence times of backward particles within the 0.01° output cells. Particle trajectories are driven by the resolved NCEP winds, while unresolved turbulent dispersion and convection are represented by the FLEXPART parameterizations (Pisso et al., 2019). The 0.01° output grid therefore provides spatially resolved source–receptor sensitivities under transport constrained by the 0.25° meteorological fields. It does not add meteorological information, but neither is it obtained by mathematically resampling a coarser SRR field.

We further clarify that the WRF fields used in this study have a horizontal resolution of 0.01°, rather than 0.1° as stated in the comment. These simulations were configured to provide the surface meteorological variables required by VPRM and were not used as transport forcing. Using kilometre-scale WRF fields for particle transport would require the FLEXPART-WRF interface and consistent configuration of the boundary-layer and turbulence treatments in the two models (Brioude et al., 2013). At this resolution, part of the boundary-layer turbulence may be explicitly resolved by WRF, and applying unadjusted FLEXPART stochastic parameterizations can double count some turbulent motions (Katharopoulos et al., 2022). A FLEXPART-WRF application would therefore require separate configuration and evaluation rather than direct substitution of the existing WRF output. In addition, extending kilometre-scale meteorological forcing across the full transport domain would considerably increase the computational demand.

We have revised Sect. 2.1.1 to clarify the respective resolutions and roles of the NCEP FNL forcing, SRR output, flux state vector, and WRF simulations. Statements implying that atmospheric transport is physically resolved at 0.01° have been corrected, and the limitation associated with the current meteorological forcing has been added.

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."

[A-05] *The extraction of background mixing ratios from three specific upper atmospheric pressure levels (down to 48.282 hPa, well into the stratosphere) lacks robust physical grounds for boundary layer inversions. More critically, the GEOS-Chem simulation period utilized for background fields exhibits a multi-year temporal mismatch with the study's actual assimilation window (2023).*

Atmospheric CO2 possesses strong interannual growth trends and pronounced vertical gradients. Forcing an unadjusted, historical, high-altitude background into a 2023 boundary-layer inversion will introduce severe, systematic biases into the posterior flux estimates. Standard protocol requires sampling the 3D background fields at the exact spatio-temporal coordinates where the backward trajectories terminate. Why was this bypassed?

[R-05] We sincerely thank the reviewer for requesting clarification of the background CO_2 representation, particularly its vertical and temporal consistency. FEISSO-Carbon follows the nested inversion framework of Thompson and Stohl (2014), in which the modeled atmospheric CO_2 concentration is separated into three components: short-term contributions from surface fluxes within the urban inversion domain, short-term contributions from surface fluxes outside this domain, and a long-term global background component. The first two components represent the influence of recent surface fluxes calculated from FLEXPART source–receptor relationships, whereas the third represents the large-scale atmospheric CO_2 state. It therefore differs conceptually from a boundary concentration obtained by sampling a three-dimensional CO_2 field at individual trajectory termination points.

The long-term background was derived from GEOS-Chem CO_2 mixing ratios at 510.475, 139.115, and 48.282 hPa, following the background estimation strategy of Thompson and Stohl (2014). These free-tropospheric and upper-atmospheric levels were selected to reduce the direct influence of recent surface emissions and characterize the vertical structure of the large-scale background. The GEOS-Chem simulations were conducted continuously over a multi-year period that includes the study year, and the background fields used for the inversions correspond to the 2023 assimilation periods. The AirCore observations from 2012–2021 were used only to evaluate and calibrate the simulated vertical CO_2 structure; they did not replace the time-matched GEOS-Chem background fields.

We have revised Sect. 2.1.4 to define the long-term background component more clearly and explain the selection of the pressure levels, the role of the AirCore observations, and the temporal consistency of the GEOS-Chem simulations.

2.1.4 Correction of background CO_2 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."

[A-06] *Furthermore, please justify the scientific meaning of sub-pixel inversion, where the observation resolution is coarser than the emission grid. The 1-km spatial structures in the posterior fields are overwhelmingly dictated by the prior spatial patterns rather than any gradient constraints provided by the satellite.*

[R-06] Thank you for pointing this out. The 0.01° grid in FEISSO-Carbon defines the spatial discretization of the flux state vector and should not be interpreted as the intrinsic resolution of the OCO-2 observations or the independently resolved spatial scale of the posterior fluxes. As in other satellite-based inversion systems, observations are not available for every inversion grid cell. The effective spatial constraint is determined jointly by the OCO-2 sampling, atmospheric transport sensitivities, and the prior error covariance structure.

OCO-2 observations are not redistributed to provide direct constraints on individual emission grid cells. Instead, each XCO_2 observation is linked to multiple upwind grid cells through the FLEXPART-derived source–receptor relationships (SRRs). The posterior fluxes are estimated by combining these transport-weighted observational constraints with the prior fluxes and their specified uncertainties. Thus, the 0.01° grid enables the representation of heterogeneous surface fluxes in the state vector, while the extent to which individual grid cells can be distinguished depends on the available observational information.

The inversion does not use spatial gradients in XCO_2 independently of atmospheric transport to adjust the posterior fluxes. The fine-scale posterior structure remains influenced by the prior spatial distribution, particularly where observational constraints are limited, whereas its emission magnitudes are updated through the transport-mediated influence of the assimilated XCO_2 observations. This treatment is consistent with recent urban inversion studies that use fine flux grids and introduce observational information through transport-derived Jacobian matrices or SRRs (Sim and Jeong, 2025).

We have revised the manuscript to distinguish the flux state-vector resolution from the effective observational resolution and to explain how OCO-2 observations constrain the posterior fluxes through FLEXPART-derived SRRs.

Lines 178-179:

"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."

[A-07] *Given that ODIAC natively provides high-resolution monthly global gridded emissions, why did the authors choose to execute a complicated spatial redistribution of EDGAR totals instead of directly deploying ODIAC as the prior emissions in the nested domain?*

[R-07] Thank you for pointing this out. The prior for the nested urban inversion domain was constructed by separating the spatial distribution from the domain-level emission magnitude. The normalized ODIAC distribution was used to allocate emissions among the 0.01° grid cells because it provides spatially explicit information based on point-source locations and spatial proxies, including nighttime lights. The EDGAR estimate was used to specify the total emissions over the inversion domain. The ODIAC spatial weights were then scaled to preserve this EDGAR domain total. This construction does not assume that either inventory is generally more accurate. Instead, it assigns distinct roles to two independently developed datasets: ODIAC defines the fine-scale spatial pattern, whereas EDGAR provides the domain-level magnitude used in this study. The resulting prior retains the spatial heterogeneity represented by ODIAC without deriving both its spatial distribution and total magnitude from the same inventory.

We have revised the manuscript to describe this scaling procedure and the respective roles of ODIAC and EDGAR more explicitly.

Lines 252-254:

"Within the nested urban inversion domain, anthropogenic prior emissions were constructed on a $0.01^{\circ} \times 0.01^{\circ}$ grid by combining the regional emission estimates from EDGAR with the fine-scale spatial patterns from ODIAC. ODIAC represents large point sources using power plant location data and distributes other fossil-fuel emissions using spatial proxy information (Oda et al., 2018). To retain the spatial heterogeneity represented by ODIAC while maintaining consistency with the regional emission estimates from EDGAR, the EDGAR emissions were redistributed on the $0.01^{\circ} \times 0.01^{\circ}$ urban inversion grid according to the ODIAC spatial patterns."

[A-08] *The prior flux error (FlxErr/GlobErr) was set at ~10%. This severely violates emissions inventory uncertainty standards. National-scale total inventory uncertainties frequently exceed 10%. At a hyper-local 1-km grid scale, uncertainties routinely exceed 100% or 200% due to point-source misallocations. What is the statistical or physical justification for a 10% grid-level uncertainty?*

[R-08] Thank you for pointing this out. The 10% value assigned to FlxErr in this study does not represent an estimate of the statistical uncertainty of individual 0.01° emission grid cells, for which inventory and spatial-allocation uncertainties may be considerably larger.

In FLEXINVERT (Thompson and Stohl, 2014), FlxErr specifies the initial relative error distribution used to construct the prior error covariance matrix. The matrix is subsequently scaled according to GlobErr, which defines the prescribed uncertainty in the total flux over the inversion domain, while SigLand and SigTime determine the spatial and temporal error correlations. The effective prior uncertainty used in the inversion therefore depends jointly on FlxErr, GlobErr, the prior flux distribution, and the specified correlation structure. FlxErr = 10% was used as the reference value for the sensitivity experiments rather than as a measured grid-scale uncertainty. The experiments varied FlxErr from 1% to 30% and GlobErr from 1% to 25% of the annual inversion-domain emissions to examine their respective influences on the posterior fluxes. The reference configuration should therefore not be interpreted as assigning a 10% uncertainty to each emission grid cell or as establishing 10% as a universally optimal value.

We have revised the manuscript to distinguish FlxErr from grid-scale inventory uncertainty and to explain the respective roles of FlxErr, GlobErr, SigLand, and SigTime in constructing the prior error covariance matrix.

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."

[A-09] *Within a 10-day window, valid OCO-2 overpasses occur only 1 to 3 times over a specific city. Given that urban carbon emissions and boundary layer heights undergo intense diurnal and hourly cycles, back-calculating a 10-day total budget from 1-3 instantaneous snapshot measurements is statistically unrepresentative and highly susceptible to random meteorological perturbations on the overpass days.*
[R-09] We agree with your point of view. It should be noted that the purpose of these experiments in our study is to demonstrate the application of FEISSO-Carbon to real satellite XCO_2 observations and assess the responses of the inversion to key parameter settings. Because temporally resolved emission inventories were unavailable for the three cities, the annual anthropogenic emission totals were proportionally scaled to 10 days to provide a reference of comparable magnitude. This scaling assumes temporally uniform emissions and is used only for contextual comparison; it does not constitute a temporal validation of the posterior estimates or imply that the selected periods represent long-term emission characteristics. Furthermore, the current assimilation of passive remote sensing observations is unable to take into account the diurnal variations of CO_2 emission fluxes. We are also seeking assistance from lidar observations.
More frequent observations would provide stronger constraints on short-term emission variability and improve the temporal representativeness of future inversions. We have revised the manuscript to clarify the rationale for the 10-day window, the role of the scaled annual inventories, and the temporal scope of the reported posterior emissions.

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."

[A-10] *a. In Figure 4b (Chengdu), as ObsErr or GlobErr increases, the total posterior emission exhibits a non-monotonic decreasing then increasing behavior. In a Bayesian inversion, such a parabolic response to error scaling is highly irregular. Across the entire sensitivity suite, the posterior fluxes exhibit massive volatility under different parameter combinations. Rather than demonstrating system stability, this proves that the inversion system lacks statistical robustness and is highly un-robust.*
b. Regarding L410, under sparse data constraints, increasing SigTime forces the system to aggressively fit the sparse observations across adjacent days. Figure 4 shows that the posterior

totals never stabilize or converge. This confirms that the system is operating in an unsaturated, unstable regime.

[R-10] We sincerely thank the reviewer for requesting further interpretation of the stronger parameter sensitivity observed for Chengdu.

(a) **We believe that the phenomenon you mentioned is caused by insufficient observations in the Chengdu area, rather than the inversion system being highly unreliable.** Chengdu was constrained by only one OCO-2 overpass during the selected assimilation period, compared with three overpasses for Weifang and Xining. Under this limited observational coverage, changes in ObsErr and GlobErr more strongly alter the relative contributions of the observational and prior constraints. Increasing ObsErr reduces the weight assigned to the observations and shifts the posterior toward the prior. Changes in GlobErr modify the allowable adjustment of the inversion-domain total flux and can redistribute posterior adjustments among grid cells. Their combined effects can therefore produce non-monotonic changes in the domain total. The inflection points observed for Chengdu are interpreted as transitions in the balance between these constraints, rather than as universally optimal parameter values.

(b) SigTime specifies the temporal correlation structure of the prior error covariance and determines how flux adjustments are distributed across the assimilation period. Increasing SigTime can extend the temporal correlation of these adjustments but does not provide additional independent observations. The Chengdu results therefore show that modifying the temporal covariance can continue to change the posterior allocation under sparse observational coverage, while providing limited additional observational constraint. The transport analysis in Sect. 3.4 further indicates that the basin topography and weak atmospheric mixing in Chengdu prolong the influence of local surface fluxes on the observations, which contributes to its longer temporal response.

To ensure a consistent comparison among the three cities, the same 10-day inversion window was adopted throughout this study. Although observational coverage differs among the three cities, the sensitivity experiments consistently demonstrate physically interpretable responses to parameter variations. The Chengdu results indicate that improving observational coverage would be more effective than further increasing the temporal correlation length for enhancing inversion robustness under observation-limited conditions.

Following the reviewer's suggestion, we have revised the manuscript to clarify that the non-monotonic responses of ObsErr and GlobErr arise from the changing balance between observational and prior constraints, whereas the continued response to increasing SigTime reflects limited observational coverage rather than instability of the inversion framework.

Lines 395-413:

"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. In Xining, an excessively large GlobErr allows unrealistically large flux adjustments, resulting in overcorrection of the posterior emissions.

The responses to SigTime are generally consistent among the three cities. Increasing SigTime extends the temporal correlation of the prior error covariance and allows the influence of individual observations to propagate over a longer period. In Weifang and Xining, where observational coverage is relatively sufficient, the influence of SigTime gradually approaches saturation. In Chengdu, however, the posterior emissions continue to respond to increasing SigTime because the limited number of observations within the fixed inversion window prevents the inversion from reaching a comparable level of constraint. These results indicate that improving observational coverage would be more effective than further increasing the temporal correlation length in observation-limited regions."

[A-11] *Page 18, the authors explicitly state that "Chengdu is a net sink" based on the flux distributions. However, Figure 4b clearly displays that all sensitivity runs for Chengdu's 10-day total emissions are entirely positive (ranging from 0.4 to 1.2 Mt), representing a net carbon source. This is a severe contradiction between the text and the figures.*

[R-11] We sincerely thank the reviewer for identifying the inconsistent interpretation of Figs. 4 and 5. Figure 4 reports the posterior emissions integrated within the administrative boundary of each city, which provides the spatial basis for comparison with the corresponding local emission inventory. Figure 5 instead shows the frequency distribution of grid-cell posterior fluxes across the entire nested urban inversion domain, including surrounding vegetated and mountainous areas. The two figures therefore differ in both their spatial domains and statistical quantities.

The frequency of positive and negative grid-cell fluxes in Fig. 5 characterizes the distribution of flux values but does not directly determine the area-integrated carbon balance. Accordingly, the previous statement that "Chengdu is a net sink" was not supported by this histogram and has been removed. The revised text describes Fig. 5 only as the frequency distribution of posterior fluxes within the Chengdu urban inversion domain.

We have revised the relevant text and figure captions to specify the spatial and statistical domains represented in Figs. 4 and 5 and to ensure consistent interpretation of the results.

Lines 380-383:

"Sensitivity experiments were conducted using a fixed 10-day inversion window for Weifang, Chengdu, and Xining. The total emissions shown in Fig. 4 were obtained by integrating the posterior fluxes within the administrative boundary of each city to facilitate comparison with the available local emission inventory. Figure 4 illustrates the sensitivity of the total posterior CO_2 emissions to the five key inversion parameters described in Sect. 2.3. For each parameter, the six sensitivity experiments (Exp1-Exp6) are arranged from left to right along the x-axis."

Lines 415-418:

"A systematic sensitivity analysis was conducted to assess the effects of the five inversion parameters described in Sect. 2.3 on the posterior CO_2 flux distributions in Weifang, Chengdu, and Xining. Figure 5 presents the flux frequency distributions obtained under different ObsErr, GlobErr, FlxErr, SigLand, and SigTime settings. The histograms were calculated over the entire nested urban inversion domain rather than within the administrative boundaries. The red, blue, and black curves represent Exp. 1, Exp. 6, and the baseline configuration, respectively.”

Lines 425-435:

“The Weifang urban inversion domain exhibits a clear bimodal distribution, with positive fluxes occurring more frequently than negative fluxes. The positive-flux peak lies between 10^{-8} and 10^{-7} kg m^{-2} s^{-1}, whereas the negative-flux peak lies between -10^{-8} and -10^{-9} kg m^{-2} s^{-1}. The predominance of positive fluxes and the relatively high frequency of strong-emission grid cells indicate that anthropogenic emissions dominate this domain. In the Chengdu urban inversion domain, negative fluxes occur more frequently, with a negative-flux peak near -10^{-8} and -10^{-9} kg m^{-2} s^{-1} and a positive-flux peak between 10^{-9} and 10^{-8} kg m^{-2} s^{-1}. Most grid cells are concentrated near zero flux, whereas strong-emission grid cells occur less frequently. In the Xining urban inversion domain, both positive and negative fluxes are comparatively weak. The positive-flux peak occurs near 10^{-8} kg m^{-2} s^{-1}, whereas the negative-flux peak occurs near -10^{-12} kg m^{-2} s^{-1}.”

[A-12] *Page 24, the authors claim that Chengdu's complex topography causes the forward model to overestimate atmospheric CO2 columns. In data assimilation logic, to compensate for simulated concentration overestimation, the inversion must decrease the emissions. Yet, Figure 7b shows that FEISSO's optimized emissions for Chengdu are the highest among all inventories and statistical totals. This is also a contradiction. Page 25, the authors claim that Chengdu's emissions are insensitive to observations. However, the results in Figure 4b indicate that varying the observational or prior error bounds alters the posterior total emissions by more than a factor of two.*

[R-12] We sincerely thank the reviewer for these careful observations. Regarding the first point, the complex topography of the Sichuan Basin and the associated weak atmospheric dispersion increase the persistence of locally accumulated CO_2. These conditions affect the sensitivity of atmospheric CO_2 concentrations to changes in surface emissions and, consequently, the observational constraints on posterior flux adjustments. They do not independently determine the direction or magnitude of these adjustments, which are jointly influenced by the prior emissions, atmospheric transport sensitivities, observational information, and their associated uncertainties. We have therefore removed the statement that atmospheric CO_2 concentrations are “overestimated” and revised the discussion to describe the effects of Chengdu’s transport conditions more precisely.

Regarding the second point, the Chengdu inversion is not insensitive to observations. As shown in Fig. 4b, the posterior total emissions vary substantially with ObsErr and GlobErr. Our intended point was that, despite variations in posterior emission totals among the sensitivity experiments, the broad spatial distribution of the major emission hotspots and their responses to parameter changes remain generally consistent. We have revised the text to distinguish between variations in posterior emission magnitude and the consistency of the broad spatial response patterns.

The corresponding discussion has been revised to clarify these two points.

Lines 583-584:

“The resulting persistence of locally accumulated CO_2 may weaken the atmospheric concentration response to changes in surface emissions, thereby limiting the observational constraint on posterior flux adjustments over emission hotspot regions.”

Lines 598-600:

“Nevertheless, the sensitivity experiments show that the major posterior spatial patterns and their general responses to parameter variations remain broadly consistent across the tested configurations, although the estimated total emissions vary.”

[R-13] *The column-averaged XCO2 is a highly blended, composite signal of background air, anthropogenic plumes, and biospheric net ecosystem exchange. Without auxiliary chemical tracers (e.g., CO, NO2) or isotopic measurements, please explain how the assimilation framework cleanly separates fossil fuel signals from biospheric signals from a single integrated XCO2 value.*

[A-13] We sincerely thank the reviewer for this important comment regarding the treatment of fossil-fuel and biospheric contributions to satellite XCO_2 observations. Column-averaged XCO_2 represents an integrated atmospheric signal influenced by background CO_2, fossil-fuel emissions, and terrestrial ecosystem exchange. These components are therefore represented separately within the FEISSO-Carbon forward modeling and inversion framework.

Terrestrial ecosystem exchange is prescribed as an independent prior flux component, whereas fossil-fuel emissions are treated as the target fluxes to be optimized. The atmospheric transport model simulates the contributions of the background atmosphere, terrestrial ecosystem exchange, and fossil-fuel emissions to XCO_2. The posterior fossil-fuel emissions are then estimated through Bayesian assimilation of the observed XCO_2, with the other components explicitly accounted for in the forward simulation. This treatment is consistent with previous satellite-based urban CO_2 inversion studies that account for biospheric fluxes and atmospheric background when estimating anthropogenic emissions from OCO-2 observations (Ye et al., 2020; Lei et al., 2021; Zhang et al., 2025).

We have revised the manuscript to clarify that the reported fossil-fuel CO_2 emissions represent the optimized anthropogenic flux component obtained after accounting for terrestrial ecosystem exchange and atmospheric background within the inversion framework, rather than a component directly separated from individual satellite observations.

Lines 567-569:

"Figure 7 presents the posterior fossil fuel CO_2 emissions obtained by optimizing the anthropogenic flux component while explicitly accounting for terrestrial ecosystem exchange within the inversion framework, together with comparisons against the prior inventories, local statistical data, and the corresponding spatial distributions."

[R-14] *In Section 4.2, the authors list a set of recommended parameters in Table 4. Simply observing highly volatile sensitivity plots is entirely insufficient to define these parameters as optimal. The authors conducted no statistical consistency tests (e.g., Chi-square) to verify the posterior error allocations, nor did they use independent data for cross-validation. Furthermore, Equation 7 for calculating globerr on Page 22 appears entirely arbitrary, lacking any derivation, physical scaling arguments, or literature backing.*

[A-14] We sincerely thank the reviewer for this valuable comment regarding the determination of inversion parameters and the specification of prior uncertainties.

Regarding the first point, we agree that the term "recommended parameters" used in the original manuscript may imply that these parameter values were obtained through a formal optimization procedure. This was not our intention. The parameter configurations listed in Table 4 were selected as reference parameter configurations based on a comprehensive evaluation of parameter sensitivity, the physical interpretability of the parameter responses, and the agreement between posterior simulations and satellite observations. We have therefore revised the manuscript accordingly to avoid overinterpreting these parameter settings as mathematically optimal solutions.

Regarding the second point, following the Bayesian formulation implemented in FLEXINVERT (Thompson and Stohl, 2014), we additionally evaluated the selected reference parameter configurations using normalized χ^2 statistics in the observation space. The χ^2 values were calculated from the model–observation differences and the total observation-space errors reported by the FLEXINVERT outputs. The relative improvement was quantified using the ratio $\chi^2_{prior}/\chi^2_{posterior}$, where values greater than one indicate improved consistency between the posterior simulations and the satellite observations.

Table S2. Normalized χ^2 statistics of prior and posterior simulations ($\chi^2_{prior}/\chi^2_{posterior}$) for all sensitivity experiments.

City	(ppm)	Exp 1	Exp 2	Exp 3	Exp 4	Exp 5	Exp 6
Weifang	ObsErr	1.4752	1.3916	1.3006	1.1526	1.0922	1.0420
	GlobErr	1.0042	1.0877	1.2004	1.3006	1.3722	1.4268
	FlxErr	1.3009	1.3011	1.3006	1.2994	1.2975	1.2895
	SigLand	1.3117	1.3097	1.3037	1.3006	1.2968	1.2865
	SigTime	1.3202	1.3190	1.3006	1.2915	1.2834	1.2750
Chengdu	ObsErr	1.1747	1.1688	1.1633	1.1429	1.1141	1.0660
	GlobErr	1.0312	1.1478	1.1633	1.1720	1.1804	1.1879
	FlxErr	1.1729	1.1725	1.1720	1.1712	1.1700	1.1671
	SigLand	1.1793	1.1760	1.1737	1.1720	1.1707	1.1666
	SigTime	1.1763	1.1751	1.1720	1.1707	1.1696	1.1685
Xining	ObsErr	1.1241	1.0823	1.0480	1.0139	1.0064	1.0023
	GlobErr	1.0006	1.0139	1.0480	1.0880	1.1243	1.1537
	FlxErr	1.0481	1.0481	1.0480	1.0480	1.0479	1.0478
	SigLand	1.0428	1.0454	1.0470	1.0480	1.0490	1.0515
	SigTime	1.0419	1.0450	1.0480	1.0491	1.0499	1.0506

Note: Each value represents the ratio of normalized χ^2 statistics between the prior and posterior simulations (χ^2prior/χ^2posterior). Values larger than 1 indicate improved consistency after inversion.

As summarized in Table S2, all selected parameter configurations produced $\chi^2_{prior}/\chi^2_{posterior}$ values greater than one, indicating that the posterior simulations consistently reduced the observation–model mismatch. We emphasize that χ^2 was introduced as a posterior consistency diagnostic rather than an optimization criterion for parameter selection.

Regarding Eq. (7), we agree with your assessment that the dimensional basis and rationale for the original formulation were not sufficiently explained. We have therefore reformulated the calculation to distinguish the absolute inversion-domain total flux uncertainty, GlobErr, from its dimensionless relative uncertainty. The revised equations also define the interannual change in CO_2 emission intensity and the relative difference between the prior inventory and the statistical emission estimate. The scaling coefficient γ is now explicitly defined according to the confidence assigned to the prior inventory and was set to 2 in this study. We have revised Sect. 4.1 to provide the equations, variable definitions, units, data sources, and rationale for using absolute values. We have also replaced "recommended parameters" with "selected reference parameter configurations" and evaluated the selected configurations using normalized χ^2 diagnostics.

Lines 561-563:

"To further investigate how geographical differences and stages of urban development affect CO_2 flux inversion characteristics, reference parameter configurations for Weifang, Chengdu, and Xining were selected based on a comprehensive evaluation of parameter sensitivities, physical interpretation of parameter effects, and posterior–observation consistency (Sects. 3 and 4.1), as summarized in Table 4. Their performance was further evaluated using normalized χ^2 statistics (Table S2). The ratio of prior to posterior normalized χ^2 values was used to quantify the improvement, with ratios greater than one indicating reduced observation–model mismatch after inversion. For all three cities, the χ^2 ratios consistently exceeded one under different parameter configurations, demonstrating improved agreement between the posterior simulations and the satellite XCO_2 observations."

Lines 538-559:

"The sensitivity analysis identifies observation error (ObsErr) and inversion-domain total flux uncertainty (GlobErr) as the parameters with the strongest influence on total posterior emissions. ObsErr represents the uncertainty assigned to the OCO-2 XCO_2 observations and was specified according to the reported uncertainty of the satellite product. GlobErr represents the uncertainty in the total annual CO_2 flux over the inversion domain and is used to scale the prior error covariance matrix. Because the prior inventory and inversion correspond to different years, GlobErr was estimated from the interannual change in emission intensity and the difference between the prior inventory and the available statistical emission estimate:

$$GlobErr = E_{prior} u_{Glob} \tag{7}$$

$$u_{Glob} = \gamma \ \left| g_{EI} \right| + \left| d_{prior-stat} \right| \tag{8}$$

$$g_{EI} = \frac{EI_t - EI_{t-1}}{EI_{t-1}}, \quad EI_t = \frac{E_t}{GDP_t} \tag{9}$$

$$d_{prior-stat} = \frac{E_{prior} - E_{stat}}{E_{stat}} \tag{10}$$

where GlobErr is the uncertainty in the total annual CO_2 flux over the inversion domain and is calculated using Eq. (7), while E_{prior} is the annual total CO_2 emission from the prior inventory. The term u_{Glob} is the relative inversion-domain total flux uncertainty calculated using Eq. (8), and γ is a scaling coefficient specified according to the confidence assigned to the prior inventory. The term g_{EI} is the relative change in CO_2 emission intensity between the inversion year and the preceding year and is calculated using Eq. (9), where EI_t is the CO_2 emission intensity in year t, defined as the ratio of annual CO_2 emissions E_t to gross domestic product GDP_t. The term $d_{prior-stat}$ is the relative difference between E_{prior} and the statistical annual emission estimate E_{stat} for the same year and is calculated using Eq. (10).

The coefficient γ determines the contribution of the interannual emission-intensity change to u_{Glob} and is specified according to the confidence assigned to the prior inventory. In this study, γ was set to 2 to provide a conservative allowance for the temporal uncertainty of the prior emissions. Absolute values were used because deviations in either direction contribute to the total flux uncertainty.

Using GDP data from the China Urban Statistical Yearbook and CO_2 emissions from EDGAR, the emission-intensity growth rates from 2022 to 2023 were estimated as −1.4 %, −0.45 %, and 4.85 % for Weifang, Chengdu, and Xining, respectively. The relative differences between the 2022 prior inventory and the available statistical emission estimates were −13.9 % for Weifang and 1.5 % for Xining. The resulting GlobErr values were 16.7 % and 11.2 % of the annual prior emissions for Weifang and Xining, respectively. Because a comparable statistical emission estimate was unavailable for Chengdu, GlobErr was set to 15 % of its annual prior emissions within the range examined in the sensitivity analysis.”

[R-15] *The posterior emission fields undergo absolutely no independent evaluation against third-party observations (e.g., surface networks, eddy covariance flux towers, or aircraft profiles). The authors argue that because their optimized inventory aligns closer with the "local inventory," the system performance is improved. This is a circular argument: what is the spatial structure of this local inventory, and how is its accuracy established? If the local inventory contains inherent systematic biases, this closer agreement would actually indicate a failure of the inversion.*

[A-15] Thank you for this important comment. We agree that closer agreement with a local inventory alone cannot demonstrate improved inversion performance, because bottom-up inventories contain uncertainties in activity data, emission factors, temporal scaling, and spatial allocation. The local inventory used in this study was constructed from locally reported emission information and spatially explicit point-source records. We have clarified its construction and now use it only as an independent bottom-up reference for contextual comparison, rather than as validation of the posterior emissions.

In the absence of surface networks, eddy covariance measurements, or aircraft observations for the three study cities, we evaluated the inversion in observation space by comparing the simulated XCO_2 with OCO-2 observations. This approach is commonly used in atmospheric CO_2 inversion studies, with changes in Bias, MAE, and RMSE providing a measure of the agreement between simulated and observed atmospheric CO_2 (Jiang et al., 2021; Zhang et al., 2023; Wang et al., 2025). As reported in the new Table 5, the posterior simulations reduce Bias, MAE, and RMSE relative to the prior simulations for all three cities, indicating improved atmospheric consistency after assimilation.

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

To further evaluate performance using observations excluded from the inversion, we conducted a 10-fold cross-validation. In each fold, 90 % of the available OCO-2 observations were assimilated, while the remaining 10 % were withheld exclusively for evaluation. Relative to the prior simulations, the cross-validated MAE decreased by 9.6 %, 46.0 %, and 15.3 % for 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 for Weifang and Xining and remained small for Chengdu. These results show improved agreement with the withheld XCO_2 observations across all three cities and provide an observation-based assessment of the out-of-sample performance of FEISSO-Carbon.

Accordingly, we have added the cross-validation procedure to the Methods, reported the prior, posterior, and cross-validated evaluation results in the new Table 5, and revised Sect. 4.2 to clarify the respective roles of the atmospheric evaluation and inventory comparisons. The posterior and cross-validated XCO_2 results are now used to assess inversion performance, whereas comparisons with EDGAR and the local inventories are presented only as contextual comparisons of emission magnitude.

Lines 566:

"The prior and posterior XCO_2 simulations were further evaluated against the OCO-2 observations using Bias, mean absolute error (MAE), and root mean square error (RMSE; Table 5). Compared with the prior simulations, the posterior MAE decreased by 14.9 %, 29.3 %, and 2.6 % in Weifang, Chengdu, and Xining, respectively, while the corresponding RMSE decreased by 12.9 %, 20.0 %, and 2.2 %. The magnitude of the Bias decreased in Weifang and Xining, whereas the posterior Bias in Chengdu became more negative despite the reductions in MAE and RMSE. This indicates that the near-zero prior Bias in Chengdu partly resulted from the cancellation of positive and negative residuals rather than uniformly small model–observation discrepancies.

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."

Lines 588-589:

Figure 7. Comparison of posterior anthropogenic CO_2 emissions estimated by FEISSO-Carbon with the EDGAR inventory and local statistical emission estimates (left), together with the corresponding spatial distribution of the posterior anthropogenic CO_2 emissions (right).

[R-16] *I strongly recommend that the authors include spatial maps of the prior emissions and the posterior-minus-prior spatial increments alongside the bar charts in Figure 7. This will allow readers to visually inspect where and how the data assimilation system adjusts the emission structures.*

[A-16] Thank you for this constructive suggestion. We agree that presenting the spatial distributions of both the prior emissions and the posterior-minus-prior increments provides a more intuitive illustration of how the assimilation system modifies the emission field and helps readers better understand the spatial characteristics of the inversion adjustments.

Following the reviewer's suggestion, we have generated maps of the prior emission fields together with the corresponding posterior-minus-prior spatial increments for all three cities. These figures clearly illustrate the locations and magnitudes of the emission adjustments introduced by the assimilation system and complement the quantitative comparisons presented in Fig. 7.

To maintain the readability and conciseness of the main manuscript while providing the complete set of spatial diagnostics, these additional maps have been included in the Appendix (Fig. S1). Appropriate references to these figures have also been added to the main text.

Lines 569:

"The corresponding prior emission fields and posterior-minus-prior increment maps are presented in Appendix Fig. S1 to further illustrate the spatial adjustments introduced by the assimilation system.”

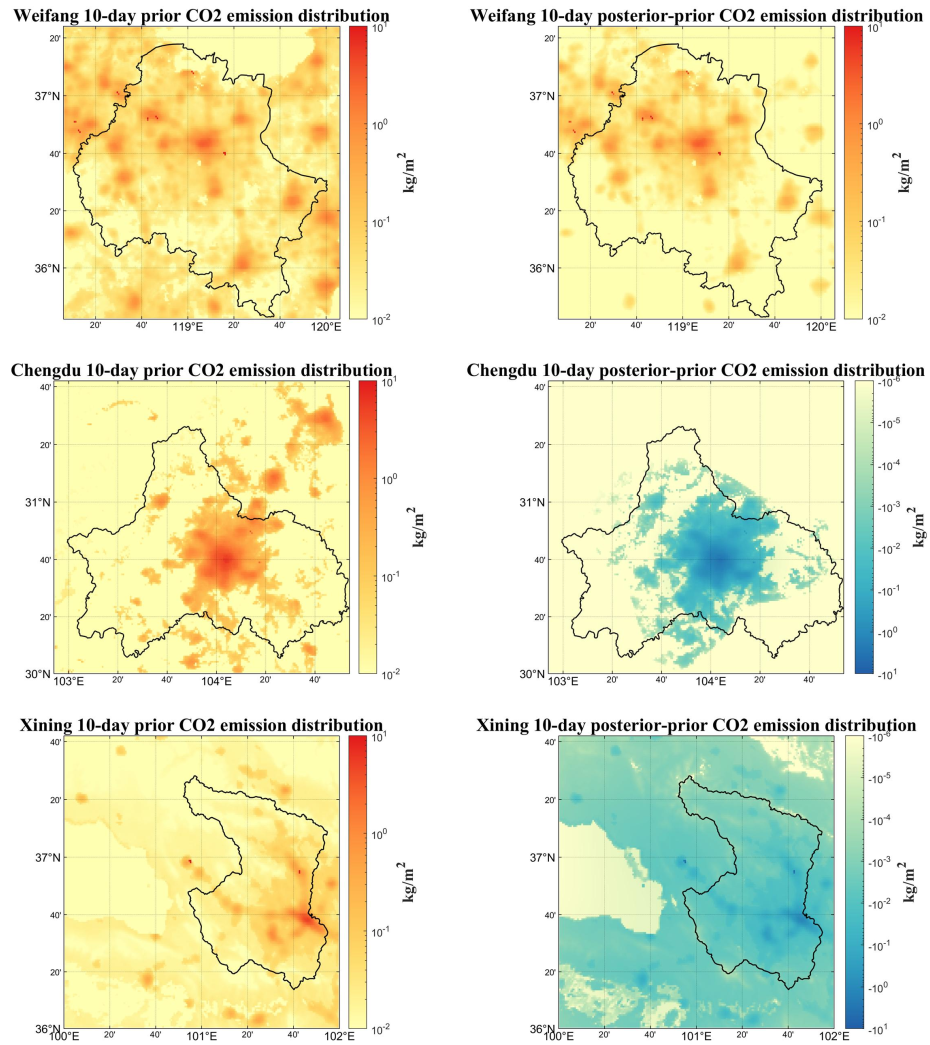

Fig. S1 Prior fossil-fuel CO_2 emission fields and posterior-minus-prior emission increments for Weifang, Chengdu, and Xining.

Technical and Minor Comments

[R-1] *The conceptual architecture in Figure 1 contains a topological error: the "Satellite-based remote sensing observations" box has no direct arrow pointing to the "CO2flux distribution inversion" module.*

[A-1] Thank you for pointing out this oversight in the conceptual framework. We agree that the original Figure 1 did not explicitly show the direct connection between the "Satellite-based remote sensing observations" module and the "CO_2 flux distribution inversion" module, which could lead to ambiguity regarding the information flow within the FEISSO-Carbon framework.

Following the reviewer's suggestion, we have revised Figure 1 by adding the missing arrow from "Satellite-based remote sensing observations" to "CO_2 flux distribution inversion", thereby providing a more accurate representation of the workflow and the role of satellite observations in the inversion process.

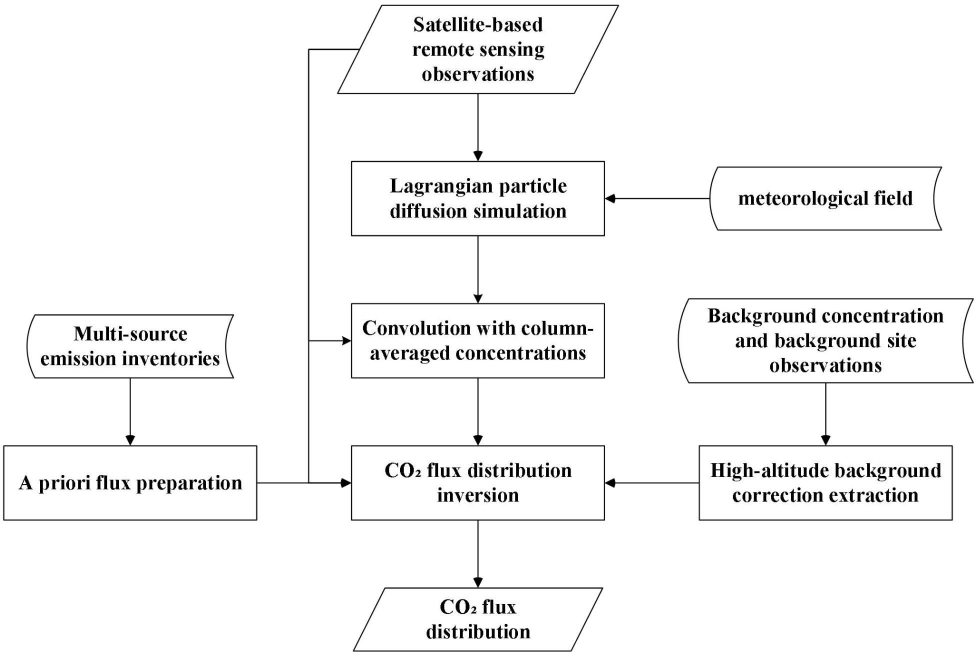

Figure 1: Schematic Structure of the FEISSO-Carbon Inversion System

[R-2] *The geographical boundaries and spatial extent of the "outer domain" are never specified. Please specify the exact boundaries of this outer domain.*

[A-2] Thank you for pointing out this lack of clarity. We agree that the geographical extent of the outer domain was not explicitly specified in the original manuscript, which could lead to ambiguity regarding the nested inversion framework.

As discussed in our responses to the major comments concerning the atmospheric transport framework and background concentration treatment, the outer domain in FEISSO-Carbon corresponds to the global atmospheric domain. It is used to represent large-scale atmospheric transport and background CO_2 contributions outside the nested urban inversion domain, while the nested domain is responsible for resolving high-resolution urban CO_2 fluxes.

Following the reviewer's suggestion, we have revised the manuscript to explicitly define the outer domain as the global domain in the methodology section and have clarified its role within the nested inversion framework.

Lines 314-317:

"To evaluate the sensitivity of FEISSO-Carbon to meteorological field resolution and key inversion parameters, a series of sensitivity experiments was conducted using a nested configuration consisting of a global outer domain and a high-resolution urban inversion domain. Meteorological fields for the global outer domain were tested at resolutions of 1° × 1° and 0.25° × 0.25° to examine the influence of meteorological field resolution on large-scale atmospheric transport, while the configuration of the urban inversion domain remained unchanged. "

[R-3] *Page 21, the authors state that local emissions dominate the satellite signal. This reflects a fundamental misconception of satellite remote sensing physics. The absolute magnitude of an XCO2 measurement is overwhelmingly (>99%) dominated by the global background concentration. Local emissions merely induce minute enhancements (typically 0.5 to 3 ppm) in the residual increment signal, not the total satellite signal.*

[A-3] Thank you for pointing out this important distinction regarding the physical interpretation of satellite XCO_2 observations. We agree that our original wording was not sufficiently precise and could be interpreted as implying that local emissions dominate the absolute satellite XCO_2 signal. This was not our intended meaning.

Figure 6 analyzes the normalized temporal contributions of different emission components to the modeled XCO_2 within the inversion framework, rather than the absolute magnitude of the observed XCO_2. Our objective was to investigate the temporal response characteristics of emissions originating from different source regions, rather than to compare their absolute contributions to the total atmospheric CO_2 column.

Following the reviewer's suggestion, we have revised the terminology throughout this section to better reflect the physical meaning of the analysis. Specifically, expressions such as "local emissions dominate the satellite signal" and "the satellite signal is mainly affected by local emissions" have been replaced with more precise descriptions, including "the contribution from emissions within the urban inversion domain to the modeled XCO_2" and "the local-domain contribution". These revisions clarify that Figure 6 describes the temporal evolution of modeled emission contributions within the inversion framework, rather than the composition of the absolute satellite XCO_2 observations.

[R-4] *The term "global" is loosely defined throughout the text. Please standardize the terminology.*

[A-4] Thank you for this helpful suggestion. We agree that the term "global" was used in several different contexts throughout the original manuscript, which could lead to ambiguity.

Following the reviewer's suggestion, we have carefully standardized the terminology throughout the manuscript according to the specific physical meanings in each context. The terms "global-scale", "global CO_2v2014 product", and "global outer domain" have been retained where they explicitly refer to the global spatial domain or established dataset names. In contrast, expressions such as "global emissions" and "global contributions" in the temporal contribution analysis have been replaced with "contributions from emissions outside the urban inversion domain", which is subsequently abbreviated as "outer-domain contributions". Correspondingly, "contributions from emissions within the urban inversion domain" are abbreviated as "local-domain contributions" after their first appearance.

In addition, the term "global uncertainty" has been replaced throughout the manuscript with "inversion-domain total flux uncertainty" to better represent the physical meaning of the GlobErr parameter. These revisions establish a consistent terminology throughout the manuscript and eliminate potential ambiguity associated with the term "global".

[R-5] *Figure 5 Legibility: The text labels and axes in the flux frequency distribution histograms are extremely small and overlap with each other, rendering the plots completely unreadable*

[A-5] Thank you for pointing out this presentation issue. We agree that the text labels and axis annotations in the original Figure 5 were too small, which reduced the readability of the flux frequency distribution histograms.

Following the reviewer's suggestion, we have revised Figure 5 by increasing the font sizes of the axis labels, tick labels, legends, and annotations. We also optimized the figure layout and spacing between subplots to improve the overall readability and visual clarity of the histograms.

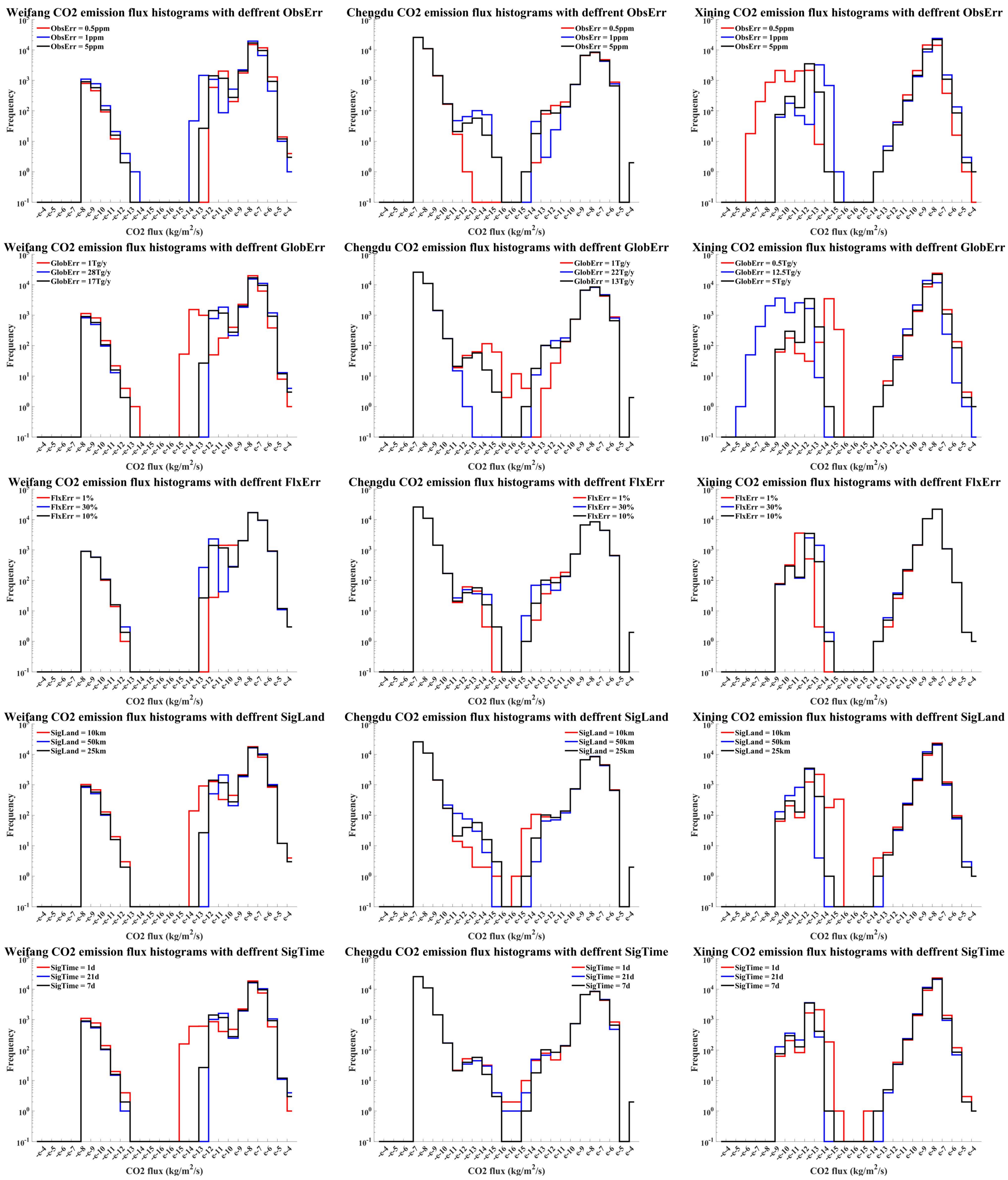

Figure 2: Histograms of CO_2 flux distributions for different parameter configurations in Weifang, Chengdu, and Xining (left to right): effects of ObsErr, GlobErr, FlxErr, SigLand and SigTime (top to bottom). Red, blue, and black lines represent Groups 1, 6, and the standard case, respectively.

[R-6] *On Page 23, the biospheric contribution percentages are written as "-5.26%"?*

[A-6] Thank you for pointing out this ambiguity. We agree that expressing the terrestrial ecosystem contribution as negative percentages may be misleading because the negative values represent the offsetting effect of biospheric CO_2 uptake rather than negative "contributions" in a literal sense.

Following the reviewer's suggestion, we have revised the corresponding sentence to explicitly distinguish between offsetting and positive contributions. Specifically, the negative percentages are now described as the proportion of fossil-fuel emissions offset by terrestrial ecosystem uptake, whereas the positive percentage indicates an additional contribution from the terrestrial ecosystem to the total CO_2 flux. This revision provides a clearer physical interpretation of the biospheric contribution.

Lines 566-567:

"The terrestrial ecosystem exerted contrasting influences on the prior CO_2 fluxes among the three cities. In Weifang and Chengdu, terrestrial ecosystem uptake offset fossil-fuel CO_2 emissions by 5.26% and 111%, respectively, whereas in Xining the terrestrial ecosystem contributed an additional 11.3% to the total CO_2 flux."

[R-7] *The definition of "super observations" in Section 2.2.2 violates basic geometric logic. The physical footprint of an OCO-2 single footprint is already coarser than the target inversion grid cell. Statistically, it is impossible to have multiple independent satellite observations residing inside a single 1-km grid cell to perform an outlier filtering and extract a median. This is simple resampling/data filtering rather than true super observations which requires multi-source aggregation for noise reduction.*

[A-7] Thank you for this insightful comment. The term "super observation" did not accurately describe the observation preprocessing applied in the present study.

FEISSO includes a general preprocessing procedure for aggregating multiple observations located within the same inversion grid cell, which is applicable to datasets with sufficiently dense spatial sampling. For the OCO-2 observations used here, the footprint is generally larger than the 0.01° inversion grid, and multiple independent observations rarely fall within the same grid cell during the selected assimilation periods. Therefore, this aggregation procedure was not applied in the present inversion, and each grid cell typically contains at most one valid OCO-2 observation after observation-to-grid assignment.

We have revised the manuscript to remove the term "super observation" from the description of the present application. The aggregation procedure is now described only as a general observation preprocessing option implemented in FEISSO, with its applicability determined by the spatial sampling density of the observational dataset.

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."

Critical Data Quality Issue

[R-1] *The authors deposited their demonstration dataset on Zenodo. However, an inspection of the repository reveals a substantial presence of extreme outliers and physically impossible values. The gridded spatial distributions exhibit structural chaotic noise. The authors must meticulously audit and correct their data files immediately.*

[A-1] Thank you for raising this important data quality concern. We agree that rigorous quality control and transparent documentation of publicly released data are essential for reproducible research.

Following the reviewer's comment, we carefully re-examined the deposited flux outputs and verified the physical consistency of the released data. We would like to clarify that the publicly available files contain daily gridded CO_2 flux fields rather than directly aggregated emission totals. The flux values are expressed in units of kg m^{-2} s^{-1}, representing surface flux densities. Therefore, localized high values in the gridded flux fields should be interpreted together with their corresponding spatial extent and integration period, rather than directly as city-scale emission totals. After the additional quality assessment, we confirmed that the posterior flux values remain within physically reasonable ranges. Specifically, the posterior fluxes ranged from 1.2990×10^{-13} to 5.6144×10^{-4} kg m^{-2} s^{-1} in Weifang, 1.2990×10^{-17} to 3.1107×10^{-4} kg m^{-2} s^{-1} in Chengdu, and 1.8454×10^{-16} to 1.4871×10^{-4} kg m^{-2} s^{-1} in Xining, with corresponding mean values of 1.6248×10^{-7}, 1.6763×10^{-7}, and 3.3624×10^{-8} kg m^{-2} s^{-1}, respectively.

Furthermore, after integrating the gridded flux fields over the corresponding spatial domains and time periods, the resulting city-scale CO_2 emissions (Mt CO_2) are consistent with the inventory-based estimates and local statistical data presented in the manuscript. Therefore, the apparent extreme values identified by the reviewer are related to the interpretation of high-resolution flux densities rather than physically unrealistic emission estimates.

We also rechecked the spatial distributions of the released flux fields. The observed spatial variability reflects the high-resolution posterior adjustments generated by the inversion system rather than structural numerical noise.

In addition, following the suggestion of the handling editor, Juan A. Añel, we have updated the FEISSO-Carbon code repository on Zenodo to ensure compliance with the GMD Code and Data Policy. The updated code release provides a clearer description of the model workflow, data processing procedures, and computational implementation required to reproduce the inversion experiments.

References

Brioude, J., Arnold, D., Stohl, A., Cassiani, M., Morton, D., Seibert, P., Angevine, W., Evan, S., Dingwell, A., Fast, J. D., Easter, R. C., Pisso, I., Burkhart, J., and Wotawa, G.: The Lagrangian particle dispersion model FLEXPART-WRF version 3.1, Geosci. Model Dev., 6, 1889–1904, 10.5194/gmd-6-1889-2013, 2013.

Chevallier, F., Fisher, M., Peylin, P., Serrar, S., Bousquet, P., Bréon, F. M., Chedin, A., and Ciais, P.: Inferring CO2 sources and sinks from satellite observations: Method and application to TOVS data, Journal of Geophysical Research: Atmospheres, 110, 2005.

Hamilton, S. D., Wu, D., Johnson, M. S., Turner, A. J., Fischer, M. L., Dadheech, N., and Jeong, S.: Estimating Carbon Dioxide Emissions in Two California Cities Using Bayesian Inversion and Satellite Measurements, Geophysical Research Letters, 51, e2024GL111150, https://doi.org/10.1029/2024GL111150, 2024.

Jiang, F., Wang, H., Chen, J. M., Ju, W., Tian, X., Feng, S., Li, G., Chen, Z., Zhang, S., and Lu, X.: Regional CO 2 fluxes from 2010 to 2015 inferred from GOSAT XCO 2 retrievals using a new version of the Global Carbon Assimilation System, Atmospheric Chemistry and Physics, 21, 1963–1985, 2021.

Jin, Z., Wang, T., Zhang, H., Wang, Y., Ding, J., and Tian, X.: Constraint of satellite CO2 retrieval on the global carbon cycle from a Chinese atmospheric inversion system, Science China Earth Sciences, 66, 609–618, 2023.

Katharopoulos, I., Brunner, D., Emmenegger, L., Leuenberger, M., and Henne, S.: Lagrangian Particle Dispersion Models in the Grey Zone of Turbulence: Adaptations to FLEXPART-COSMO for Simulations at 1 km Grid Resolution, Boundary-Layer Meteorology, 185, 129–160, 10.1007/s10546-022-00728-3, 2022.

Kong, Y., Zheng, B., Zhang, Q., and He, K.: Global and regional carbon budget for 2015–2020 inferred from OCO-2 based on an ensemble Kalman filter coupled with GEOS-Chem, Atmos. Chem. Phys., 22, 10769–10788, 10.5194/acp-22-10769-2022, 2022.

Kunik, L., Mallia, D. V., Gurney, K. R., Mendoza, D. L., Oda, T., and Lin, J. C.: Bayesian inverse estimation of urban CO2 emissions: Results from a synthetic data simulation over Salt Lake City, UT, Elem Sci Anth, 7, 36, 2019.

Lei, R., Feng, S., Danjou, A., Broquet, G., Wu, D., Lin, J. C., O'Dell, C. W., and Lauvaux, T.: Fossil fuel CO2 emissions over metropolitan areas from space: A multi-model analysis of OCO-2 data over Lahore, Pakistan, Remote Sensing of Environment, 264, 112625, https://doi.org/10.1016/j.rse.2021.112625, 2021.

Lian, J., Lauvaux, T., Utard, H., Bréon, F.-M., Broquet, G., Ramonet, M., Laurent, O., Albarus, I., Cucchi, K., and Ciais, P.: Assessing the Effectiveness of an Urban CO2 Monitoring Network over the Paris Region through the COVID-19 Lockdown Natural Experiment, Environmental Science & Technology, 56, 2153–2162, 10.1021/acs.est.1c04973, 2022.

Liu, Z., Zeng, N., Liu, Y., Kalnay, E., Asrar, G., Wu, B., Cai, Q., Liu, D., and Han, P.: Improving the joint estimation of CO2 and surface carbon fluxes using a constrained ensemble Kalman filter in COLA (v1.0), Geosci. Model Dev., 15, 5511–5528, 10.5194/gmd-15-5511-2022, 2022.

Liu, Z., Zeng, N., Han, P., Cai, Q., Yao, B., Wang, P., Quan, W., Tang, W., Liang, Z., Liu, Y., and Cory, M.: Detection of Chinese Spring Festival in Beijing using in-situ CO2 observations and atmospheric inversion, Atmospheric Environment, 325, 120446, https://doi.org/10.1016/j.atmosenv.2024.120446, 2024.

Nickless, A., Rayner, P. J., Scholes, R. J., Engelbrecht, F., and Erni, B.: An atmospheric inversion over the city of Cape Town: sensitivity analyses, Atmospheric Chemistry and Physics, 19, 7789–7816, 2019.

Oda, T., Maksyutov, S., and Andres, R. J.: The Open-source Data Inventory for Anthropogenic CO 2, version 2016 (ODIAC2016): a global monthly fossil fuel CO 2 gridded emissions data product for tracer transport simulations and surface flux inversions, Earth System Science Data, 10, 87–107, 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.

Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., Miller, J. B., Bruhwiler, L. M., Pétron, G., and Hirsch, A. I.: An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker, Proceedings of the National Academy of Sciences, 104, 18925–18930, 2007.

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.

Remaud, M., Chevallier, F., Cozic, A., Lin, X., and Bousquet, P.: On the impact of recent developments of the LMDz atmospheric general circulation model on the simulation of CO2 transport, Geosci. Model Dev., 11, 4489–4513, 10.5194/gmd-11-4489-2018, 2018.

Roten, D., Lin, J. C., Das, S., and Kort, E. A.: Constraining Sector-Specific CO2 Fluxes Using Space-Based XCO2 Observations Over the Los Angeles Basin, Geophysical Research Letters, 50, e2023GL104376, https://doi.org/10.1029/2023GL104376, 2023.

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.

Wang, Y., Cheng, Y., Mai, B., Dai, T., Deng, X., Deng, T., Zhao, X., Diao, Y., Xia, F., Liang, M., Li, Y., and Zhu, Y.: Near real-time inversion of high-resolution anthropogenic carbon emissions in the Pearl River Delta region based on the four-dimensional local ensemble transform Kalman filter, EGUsphere, 2025, 1–32, 10.5194/egusphere-2025-6272, 2025.

Ye, X., Lauvaux, T., Kort, E. A., Oda, T., Feng, S., Lin, J. C., Yang, E. G., and Wu, D.: Constraining fossil fuel CO2 emissions from urban area using OCO-2 observations of total column CO2, Journal of Geophysical Research: Atmospheres, 125, e2019JD030528, 2020.

Zhang, L., Jiang, F., Mao, Y., Lv, G., Wang, H., Feng, S., and Ju, W.: A Top-Down Method for Estimating Regional Fossil Fuel Carbon Emissions Based on Satellite XCO2 Retrievals, Remote Sensing, 17, 447, 2025.

Zhang, L., Jiang, F., He, W., Wu, M., Wang, J., Ju, W., Wang, H., Zhang, Y., Sitch, S., Walker, A. P., Yue, X., Feng, S., Jia, M., and Chen, J. M.: A Robust Estimate of Continental-Scale Terrestrial Carbon Sinks Using GOSAT XCO2 Retrievals, Geophysical Research Letters, 50, e2023GL102815, https://doi.org/10.1029/2023GL102815, 2023.

Zhang, Z., Feng, S., Chen, Y., Liu, Q., Ju, W., Xiao, W., Huang, C., Wang, Y., Wang, H., Jia, M., Wang, X., and Jiang, F.: Development of a regional carbon assimilation system and its application for estimating fossil fuel carbon emissions in the Yangtze River Delta, China, Science of The Total Environment, 957, 177720, https://doi.org/10.1016/j.scitotenv.2024.177720, 2024.