ADVANCING EARTH AND SPACE SCIENCES

JGR Atmospheres

RESEARCH ARTICLE

10.1029/2025JD045557

Development and Evaluation of an Urban Atmospheric Inverse Modeling Framework for the Washington, DC and Baltimore, MD Metropolitan Area: Initial Results From an Inert Tracer Case Study

Miguel Cahuich-López[1,2], Christopher P. Loughner[1], Fong Ngan[1,2], Anna Karion[3], Lei Hu[4], Israel Lopez-Coto[3,5], Kimberly Mueller[3], Julia K. Marrs[3,6], Arlyn Andrews[4,7], John Miller[4], Brian C. McDonald[8], Colin Harkins[8,9], Congmeng Lyu[8,9], Meng Li[8,9], Kevin R. Gurney[10], Mark Cohen[1], Howard Diamond[1], Ariel Stein[1], and James Whetstone[3]

[1]Air Resources Laboratory, National Oceanic and Atmospheric Administration, College Park, MD, USA, [2]Earth System Science Interdisciplinary Center/Cooperative Institute for Satellite Earth System Studies, University of Maryland, College Park, MD, USA, [3]National Institute of Standards and Technology, Gaithersburg, MD, USA, [4]Global Monitoring Laboratory, National Oceanic and Atmospheric Administration, Boulder, CO, USA, [5]School of Marine and Atmospheric Sciences, Stony Brook University, Stony Brook, NY, USA, [6]National Environmental Satellite, Data, and Information Service, National Oceanic and Atmospheric Administration, College Park, MD, USA, [7]Now at Former Feds and Friends, LLC, Nederland, CO, USA, [8]Chemical Sciences Laboratory, National Oceanic and Atmospheric Administration, Boulder, CO, USA, [9]Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, CO, USA, [10]Northern Arizona University, Flagstaff, AZ, USA

Key Points:

- An urban inverse modeling system is developed to constrain hourly 1-km surface fluxes of an inert tracer in the DCBA for January 2019
- System incorporates enhanced high-resolution HYSPLIT backward dispersion simulations and explicit spatiotemporal covariance of prior errors
- Mean flux and total emissions are significantly improved at city and grid-cell scales; flux variations improve only at the city scale

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

M. Cahuich-López,
mcahuich@umd.edu;
miguel.cahuich@noaa.gov

Citation:

Cahuich-López, M., Loughner, C. P., Ngan, F., Karion, A., Hu, L., Lopez-Coto, I., et al. (2026). Development and evaluation of an urban atmospheric inverse modeling framework for the Washington, DC and Baltimore, MD Metropolitan Area: Initial results from an inert tracer case study. *Journal of Geophysical Research: Atmospheres*, *131*, e2025JD045557. https://doi.org/10.1029/2025JD045557

Received 3 OCT 2025
Accepted 24 JUN 2026

Author Contributions:

Conceptualization: Miguel Cahuich-López, Christopher P. Loughner, Fong Ngan, Anna Karion, Lei Hu, Israel Lopez-Coto, Kimberly Mueller, Arlyn Andrews, John Miller, Brian C. McDonald, Colin Harkins, Mark Cohen, Howard Diamond, Ariel Stein, James Whetstone
Data curation: Miguel Cahuich-López, Christopher P. Loughner, Fong Ngan

Abstract Accurate quantification of hazardous material releases in cities is critical for emergency managers aiming to mitigate health impacts and mortality in the population. Here, a study on the development and evaluation of a high-resolution (1-km, 1-hr) urban-scale atmospheric inverse modeling system for the Washington, DC, and Baltimore, MD, metropolitan area (DCBA) is presented to advance surface flux quantification in urban and industrial areas where the potential for a hazardous release is high. The system employs an inert tracer as a proof of concept for January 2019 and builds upon HYSPLIT atmospheric transport and dispersion simulations driven by high-resolution Weather Research and Forecasting (WRF) model simulations that ingest urban meteorological observations, coupled with the CarbonTracker-Lagrange (CT-L) inverse model tailored to account for hourly spatiotemporal fluxes. The system assimilates tower-based tracer observations from the National Institute of Standards and Technology (NIST) Northeast Corridor Urban Test Bed and utilizes the 1-km Vulcan and 4-km GRA2PES bottom-up inventories as prior knowledge of the tracer fluxes. Numerical experiments using synthetic and actual data reveal that the system significantly improves the quantification of flux strength (mean flux) and total mass released across the urban cores of Washington, DC, and Baltimore, MD. Hourly flux variations are fairly well resolved when originating from a broad territory (city scale), while the accuracy of source retrieval improves when flux estimates are aggregated to coarser spatial resolutions (e.g., 4 km). This work has important implications for estimating hazardous releases from urban areas for emergency response and hazard analysis.

Plain Language Summary Releases of hazardous materials into the atmosphere (emissions) during emergency or accidental events in urban and industrial areas pose a threat to city populations. Exposure to these harmful materials is primarily mediated by urban meteorology (e.g., wind), which determines how these substances spread from release sources to surrounding areas. Therefore, it is essential to have tools that estimate emission intensity, total emissions, temporal variation, and source location. This study presents and evaluates a model that infers these emission characteristics from downwind atmospheric observations in DCBA during January 2019. This type of approach, known as inverse modeling, employs a statistical technique to adjust a prior assumption about emissions to make it more consistent with observations. Our model is suitable for this purpose because its representation of the underlying urban atmospheric dispersion is sufficiently accurate, and it achieves a more robust tuning of prior errors, which is necessary for this statistical technique.

Formal analysis: Miguel Cahuich-López, Christopher P. Loughner, Fong Ngan
Funding acquisition: Howard Diamond, Ariel Stein
Investigation: Miguel Cahuich-López, Christopher P. Loughner, Fong Ngan
Methodology: Miguel Cahuich-López, Christopher P. Loughner, Fong Ngan, Anna Karion, Lei Hu, Israel Lopez-Coto, Kimberly Mueller, Arlyn Andrews, John Miller
Project administration: Mark Cohen, Howard Diamond, Ariel Stein, James Whetstone
Resources: Christopher P. Loughner, Anna Karion, Julia K. Marrs, Brian C. McDonald, Colin Harkins, Congmeng Lyu, Meng Li, Kevin R. Gurney, Mark Cohen, Howard Diamond, Ariel Stein
Software: Miguel Cahuich-López, Christopher P. Loughner, Fong Ngan, Anna Karion, Lei Hu, Julia K. Marrs
Supervision: Christopher P. Loughner, Mark Cohen
Validation: Miguel Cahuich-López, Christopher P. Loughner, Fong Ngan, Anna Karion
Visualization: Miguel Cahuich-López, Mark Cohen
Writing – original draft: Miguel Cahuich-López
Writing – review & editing: Miguel Cahuich-López, Christopher P. Loughner, Fong Ngan, Anna Karion, Lei Hu, Israel Lopez-Coto, Kimberly Mueller, Julia K. Marrs, Arlyn Andrews, John Miller, Brian C. McDonald, Colin Harkins, Congmeng Lyu, Meng Li, Kevin R. Gurney, Mark Cohen, Howard Diamond, Ariel Stein, James Whetstone

1. Introduction

Accurate quantification of the sources, transport, and fate of atmospheric hazardous materials is fundamental for effective emergency management, especially in urban and industrial areas where the potential for a hazardous release is higher. Stakeholders often use so-called bottom-up methods to estimate the release of these materials, based on emission factors multiplied by activity data (Fong et al., 2014). This type of release accounting often lacks the spatial and temporal information critical for a rapid and effective emergency response. It has also been shown that these types of methods can provide inconsistencies (Gurney et al., 2021). Therefore, complementary methods that provide estimates with higher resolution and lower latency, such as top-down inversion techniques, are essential for first responders and emergency managers seeking to develop targeted strategies for mitigating the impact of a hazardous release and protecting public health (Singh et al., 2015).

Urban-scale top-down atmospheric Bayesian inversions offer a different perspective for urban and industrial flux quantification (i.e., emissions) (e.g., Breón et al., 2015; Ghosh et al., 2021; Hajny et al., 2022; Huang et al., 2019; Kort et al., 2013; Kunik et al., 2019; Lauvaux & Davis, 2014; Lauvaux et al., 2016; Lopez-Coto et al., 2017; Vardag & Maiwald, 2024; Wu et al., 2018). They rely on direct atmospheric concentration measurements of hazardous materials based on satellites, aircraft, mobile platforms, or fixed surface locations, spatially explicit bottom-up emissions inventories as prior information, along with atmospheric transport models (e.g., Gourdji et al., 2018; Mallia et al., 2020; Nesser et al., 2024; Semerjian et al., 2024). Even though top-down atmospheric inversions have inherent uncertainties, they can provide an independent check of bottom-up methods on shorter time frames (e.g., Brioude et al., 2011; Feng et al., 2023; Hu et al., 2015, 2025; Itahashi et al., 2019; Nüß et al., 2025; Qu et al., 2022).

Top-down inversion methods for urban areas are most directly applicable and have primarily been applied to trace gases that are long-lived relative to the time scale of urban atmospheric transport and dispersion (e.g., CO, CH_4, and CO_2) so that the significant added complexity introduced by atmospheric transformations can be avoided. Atmospheric inversions of inert trace gases can uncover temporal patterns, trends, and anomalies in emissions at urban scales (e.g., Brioude et al., 2011; Karion, Ghosh, et al., 2023; Lopez-Coto, Ren, et al., 2020; Lopez-Coto et al., 2022; Pitt et al., 2022; Yadav et al., 2021; Yumimoto et al., 2014). Berchet et al. (2021) demonstrated with synthetic data that analytical inversions are the most accurate method for linear tracer problems, followed by variational and ensemble techniques.

Atmospheric transport models are used within trace gas inversion systems to link observations with surface fluxes. Commonly, backward in time atmospheric dispersion simulations are performed originating at the point of the observation in order to calculate all of the possible locations where surface fluxes may have influenced the measurement, which is the footprint of the measurement. The use of the Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) model in these inversion methods has been advanced by the incorporation of the Stochastic Time-Inverted Lagrangian Transport (STILT) model dispersion algorithms into the HYSPLIT model, allowing these STILT features to stay up to date (Loughner et al., 2021). In another study in Washington, DC, it was found that the use of local meteorological data to enhance weather models through observational nudging is essential to improve the accuracy of atmospheric transport and dispersion forecasts with the HYSPLIT model (Lichiheb et al., 2024).

In addition, previous urban-scale research argues that inversion performance is limited by both temporal resolution and temporal correlations of prior errors. Bayesian inversions typically focus on constraining time-averaged emissions with the goal of deriving temporal patterns of posterior fluxes not driven by changes in the prior (e.g., Ghosh et al., 2021; Karion, Ghosh, et al., 2023; Lauvaux et al., 2016; Lopez-Coto, Ren, et al., 2020; Lopez-Coto et al., 2017; Nalini et al., 2022; Pitt et al., 2022; Yadav et al., 2021). Nevertheless, Super et al. (2021) show that expected concentrations of inert trace gases used by the inversion algorithm (i.e., modeled observations computed by convolving footprints with the prior to map surface fluxes into observational space) are affected when realistic temporal variability is introduced into the prior, causing significant differences in results. Moreover, Kunik et al. (2019) demonstrated that proper characterization of temporal correlations of prior errors, along with spatial correlations, is crucial for obtaining robust inverse fluxes.

For these reasons, further research on urban inverse modeling is required. The tracer experiment perspective can be particularly helpful because the performance of atmospheric transport models can be estimated (Zannetti, 2013), and thus, inverse modeling. A requirement for tracers is that they must be relatively innocuous to human health (Rappolt & Kerrin, 2010). Due to this safety consideration—combined with the relatively high density of CO_2 monitoring networks—this compound has been employed as a "tracer of opportunity," spanning both global (e.g., Law et al., 2008) and urban-scale applications (e.g., Lopez-Coto, Hicks, et al., 2020; Martin et al., 2019). In the Washington, DC, and Baltimore, MD, Metropolitan Area (DCBA), Martin et al. (2019) used WRF-Chem simulations to show that wintertime modeled observation error relative to measurements is dominated by errors in model meteorology rather than by emission inventory errors, reinforcing the implications identified by Lichiheb et al. (2024).

This study evaluates a high-resolution (1 km, 1 hr) urban-scale inverse modeling system with an inert tracer (CO_2) for DCBA in January 2019, based on both synthetic and actual data numerical experiments. It differs from previous work for several key reasons: (a) the backbone of the system is the HYSPLIT model driven by high-resolution (1-km) WRF simulations nudged with urban meteorological observations collected across DCBA, which performs in reanalysis mode to generates the footprints; (b) the system adopts an hourly flux retrieval resolution within the CarbonTracker-Lagrange (CT-L) analytical inverse model (Hu et al., 2019; NOAA Global Monitoring Laboratory, 2019); (c) the data assimilation formulation explicitly represents temporal correlations in prior emission errors in DCBA, in addition to spatial correlations; and (d) the system uses CO_2 as a tracer of opportunity to demonstrate its capability to infer emissions in complex urban environments, thereby informing subsequent system developments to emergency response preparedness. The degree to which our inversion system captures the flux characteristics of tracer emissions in DCBA provides a benchmark for the accuracy of urban atmospheric modeling—including the application of the same underlying models for emergency response, such as the urban atmospheric transport framework (i.e., HYSPLIT configuration, meteorological nudging, and the spatial configuration of the observational network). Our premise is that before integrating the nonlinearities inherent in the chaotic nature of emergencies (e.g., chemical transformations), an urban inversion system must first undergo rigorous validation in an inert tracer environment. Consequently, the dense CO_2 observation network, the ubiquity of CO_2 emissions throughout DCBA, and the combination of other factors in January 2019 provide a unique tracer of opportunity for this case study.

The structure of this paper is as follows: Section 2 describes the methodology, including the experimental setup and the several components that are integrated in the system, Section 3 presents a detailed evaluation of the capabilities and limitations of the urban-scale inverse modeling system, and the conclusions of this study are presented in Section 4.

2. Materials and Methods

2.1. Experimental Design

The urban inverse modeling system has been built upon Observing System Simulation Experiments (OSSEs) and actual data inversions assimilating CO_2 atmospheric concentration measurements collected on 3–29 January 2019, in an urban network of towers in DCBA that forms part of the NIST Northeast Corridor Urban Test Bed (NIST-NEC-DCBA) (Karion et al., 2020). The selected time period (January 2019) is methodologically advantageous for several reasons: (a) DCBA has an existing dense measurement tower network that includes background sites; (b) biogenic fluxes, which are known to be challenging to model, are small compared to anthropogenic CO_2 fluxes during winter (Winbourne et al., 2022; Yadav et al., 2021), so CO_2 atmospheric concentrations are primarily shaped by urban meteorology, that is, while emissions themselves vary mainly with anthropogenic activity, meteorology determines how those emissions accumulate and are observed in urban CO_2 concentrations; (c) there are other bottom-up and top-down estimates for January 2019; and (d) while accidental releases can occur in any season, this case study offers an opportunity to evaluate the system during winter, when more typical temperature inversions lead to greater accumulation of pollutants at the surface, thus representing a worst-case scenario from an emergency response preparedness perspective.

Here, the current configuration of the urban inverse modeling system is examined using four baseline numerical experiments, summarized in Table 1 and briefly outlined below. Tests 01-a and 02-a assimilate actual measurements of the atmospheric tracer, which differ in the emission data products used to provide the prior fluxes. Test 01-a uses the Vulcan v3.0 data set (Gurney et al., 2020a, 2020b) for January 2015, which is a time-varying (1-hr) high-resolution (1-km) prior that is able to spatially resolve localized and linear sources. In contrast, test 02-a uses a preview release of the GReenhouse gas And Air Pollutants Emissions System (GRA2PES) for January 2019, developed by the NOAA Chemical Sciences Laboratory (CSL) and NIST (Lyu et al., 2025;

Table 1
Summary of Baseline Inverse Modeling Experiments Carried Out in This Work (Same = Identical to Test 01-a) (Detailed Metadata on Input Data Sets Utilized in the Modeling Setup is Presented in Table S5 in Supporting Information S1)

Test ID	Observational data	Prior fluxes	Atmospheric transport model	Data assimilation framework	Temporal covariance of prior errors	Retrieval resolution
01-a	Actual measurements	1-km Vulcan (Prior 1)	1-km WRF-HYSPLIT adjusted with urban meteorological observations	CT-L v0.2.0	Weekly-cosinusoidal	1-km, 1-hr
02-a	Actual measurements	4-km GRA2PES (Prior 2)	Same	Same	Same	Same
01-s	1-km ACES-derived synthetic data	1-km Vulcan (Prior 1)	Same	Same	Same	Same
02-s	1-km ACES-derived synthetic data	4-km GRA2PES (Prior 2)	Same	Same	Same	Same

NOAA Chemical Sciences Laboratory & National Institute of Standards and Technology, 2023), which is a time-varying (1-hr) 4-km resolution prior currently provided for three representative days of week in each month (weekdays, Saturdays and Sundays). The objective of test 02-a is to evaluate the combined effect of using (a) a coarser prior and (b) a distinct inventory (GRA2PES) developed using alternative methods and data sources. Tests 01-s and 02-s are configuration-mirrored experiments to, respectively, tests 01-a and 02-a (i.e., same atmospheric transport, observation noise, priors, error covariance, domain, and grid) that assimilate synthetic data generated from the Anthropogenic Carbon Emission System version 2 (ACES; Gately & Hutyra, 2022) for January 2017, which is another time-varying high-resolution product (1-km, 1-hr) with data closer to our study period (January 2019). The OSSEs serve for two purposes: (a) to provide insights into the skill of the system to detect the underlying tracer flux characteristics under controlled settings given the prior specification; and (b) to serve as benchmarks for evaluating the responses observed in the experiments conducted with actual data. Accordingly, the information provided by both tests 01-s and 02-s, and derived sensitivity inversions, is required to obtain a comprehensive understanding of the system performance (Michalak et al., 2017).

To guide the reader, this chapter proceeds as follows: Sections 2.2 and 2.3 describe, respectively, the model domain and system framework utilized in the actual data numerical experiments. The inversion system framework description is followed by dedicated subsections for describing the system components. Next, in Section 2.4, it is explicated how the synthetic data numerical experiments are constructed. Finally, the metrics used to estimate the model performance are described in Section 2.5.

2.2. Model Domain

A map of the DCBA model domain is presented in Figure 1, which includes the locations of the observational network of NIST-NEC-DCBA. The green circle markers indicate the locations of the NIST towers used in the inversions, the blue triangles denote the NIST sites used to provide the boundary (background) concentrations, and the white square markers represent the meteorological data network incorporated into the high-resolution meteorological modeling. The NIST sites (green circles and blue triangles) also provide meteorological observations that are used within the meteorological model. With respect to the modeling domain, the inner white box in Figure 1 indicates the boundaries of the region (south: 38.4°; north: 39.6°; west: −77.8°; east: −76.2°), where hourly fluxes are constrained on a regular latitude-longitude grid with a nominal resolution of 1 km (i.e., 0.01° latitude by 0.01° longitude resolution). The high-resolution (1 km) meteorological modeling domain covers the outer white box. As shown in Figure 1, the NIST towers used in the inversions are located in a high- and medium-intensity developed urban area according to the National Land Cover Database (U.S. Geological Survey, 2024). A subsequent analysis of the average afternoon HYSPLIT-calculated surface sensitivities across the domain in January 2019 (Figure S1 in Supporting Information S1) reveals that the urban corridor can be effectively observed by the tower network, as the surface sensitivity is higher than in peripheral areas. This region has been referred to as the *central subdomain*, as it is used to evaluate the skill of the system in the urban corridor.

2.3. Inversion System Framework

The data assimilation technique of CT-L is analytical and computes the algebraic solution of the Gaussian Bayesian inverse problem outlined by Tarantola (2005) using the algorithms described by Yadav and Michalak (2013) to reduce the computational cost of performing large linear space–time inverse problems (details in

AGU ADVANCING EARTH AND SPACE SCIENCES
Journal of Geophysical Research: Atmospheres
10.1029/2025JD045557

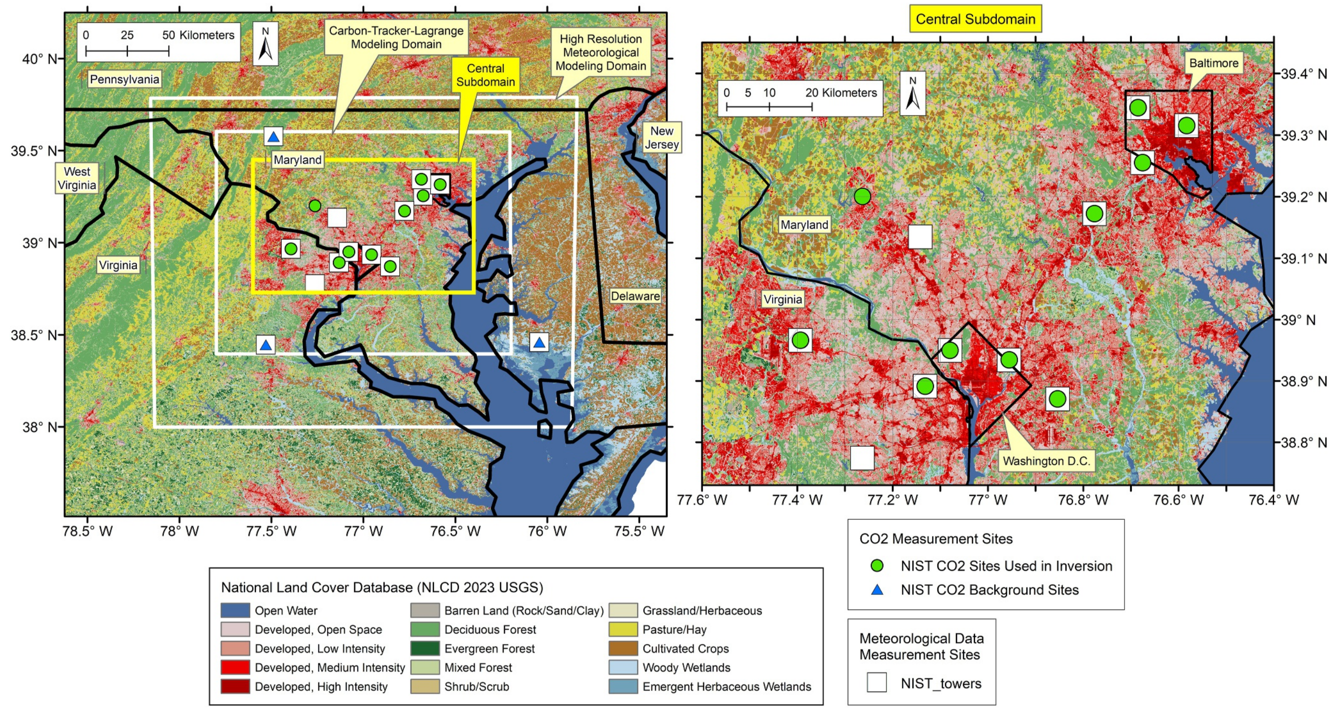

Figure 1. Observational network and domain of the urban inverse modeling system in the DCBA. The yellow box delimits the central subdomain region, with an enlarged view in the right panel, where the surface sensitivities associated with the NIST-NEC-DCBA tower network are higher.

Text S1 in Supporting Information S1). This unusual feature of CT-L allows calculating the exact mathematical solution to the optimal value ($\hat{\mathbf{s}}$) and its uncertainty ($\mathbf{V}_{\hat{s}}$), unlike ensemble or variational approaches, which are given by,

$$\hat{\mathbf{s}} = \mathbf{s}_p + (\mathbf{HQ})^T \left(\mathbf{HQH}^T + \mathbf{R}\right)^{-1} (\mathbf{z} - \mathbf{Hs}_p), \text{and} \tag{1}$$

$$\mathbf{V}_{\hat{s}} = \mathbf{Q} - (\mathbf{HQ})^T \left(\mathbf{HQH}^T + \mathbf{R}\right)^{-1} \mathbf{HQ}, \tag{2}$$

where $\hat{\mathbf{s}}$ ($m \times 1$) is the posterior best estimate of the unknown true fluxes with units of μmol/m^2/s, $\mathbf{s}_p$ ($m \times 1$) is the prior estimate of $\hat{\mathbf{s}}$, $\mathbf{z}$ ($n \times 1$) is a vector of observed tracer enhancements as given by Equation 3 in units of μmol/mol or ppm, $\mathbf{H}$ ($n \times m$) is a matrix of HYSPLIT-calculated footprint values that describes the relationship between the observable tracer and the fluxes, $\mathbf{Q}$ ($m \times m$) is a covariance matrix of prior errors describing the expected deviations between the prior and true fluxes, and $\mathbf{R}$ ($n \times n$) is a covariance matrix of errors associated with the measurements and modeling, also known as model-data mismatch errors. The product of the second and third terms in Equations 1 and 2, $(\mathbf{HQ})^T \left(\mathbf{HQH}^T + \mathbf{R}\right)^{-1}$, corresponds to the Kalman gain matrix, which determines the weight given to the observed data relative to the prior state (Berchet et al., 2021).

In the model domain, an assimilation time interval of 27 days (3–29 January 2019) is adopted, incorporating the available afternoon (12:00–17:00 ET) tracer enhancements. Observations falling outside the 12:00–17:00 ET time window are excluded from the inversion although footprints associated with the assimilated observations are not restricted exclusively to the afternoon hours. Consequently, we use the available footprints to constrain $\mathbf{s}_p$, but limited here the analysis of flux retrievals to the afternoon, when there is generally more confidence in atmospheric transport modeling, that is, when the planetary boundary layer is higher and well-mixed (e.g., Breón et al., 2015). A preliminary assessment of the system showed a decreasing retrieval skill as time goes backward—from the afternoon to morning, early morning, and night— with residual retrieval at early morning and practically no retrieval at night (Cahuich-López et al., 2024). This pattern is expected because we are constraining our fluxes only with afternoon data; thus, the surface sensitivities associated with these observations decrease as temporal lag increases.

In the context of matrix dimensions, our inverse problem has n (953 observations) and m ($p \times r$) dimensions, where $p = 648$ time steps, and $r = 19{,}200$ grid cells (i.e., 120×160 grid cells) which leads to 12,441,600 unknowns ($648 \times 19{,}200$). The otherwise ill-posed problem is regularized through the inclusion of the prior information and its associated error covariance matrix, which incorporates spatial and temporal correlations (described in Section 2.3.3). This structure allows to spread out observational information in space and time and reduces the degrees of freedom in the model space (Bannister, 2008; Hu et al., 2015; Kountouris et al., 2015; Singh et al., 2015), thereby meaningfully constraining such a large parameter space primarily in the afternoon time interval.

2.3.1. Atmospheric Transport Model and Its Drivers

2.3.1.1. High-Resolution Meteorological Modeling

In this study, meteorological fields are generated utilizing the Weather Research and Forecasting (WRF; version 4.5.2) model (Powers et al., 2017), configured with three nested domains at horizontal grid spacings of 9, 3, and 1 km. The innermost 1-km domain encompassed the DCBA area (see Figure 1) and provided meteorological inputs for HYSPLIT backward dispersion simulations. The vertical grid consisted of 55 layers, with the lowest mid-layer height at ∼15 m AGL and the model top set at ∼16,000 m AGL (100 hPa). Simulations are initialized by the Global Data Assimilation System (GDAS; National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce, 2015) in 0.25-degree spatial resolution and available every 6 hr, while the output from the coarser domains provides the initial and boundary conditions for the inner domains. Each WRF run spans 30 hr, including a 6-hr spin-up period starting at 18:00 UTC the previous day.

WRF physics schemes (Skamarock et al., 2021) include the Rapid Radiative Transfer Model for radiation, WSM6 for microphysics, the Grell 3D ensemble for the sub-grid cloud, the Noah land surface model, the MM5 Monin-Obukhov surface scheme, and the Shin-Hong scheme for the planetary boundary layer (PBL) parameterization. To improve the accuracy of meteorological fields for dispersion modeling, four-dimensional data assimilation (FDDA; so-called "nudging"; Reen, 2016) is applied using analysis nudging for the outer domains and observational nudging for the innermost domain with u- and v-wind measurements from NIST-NEC-DCBA towers. Observational nudging settings include: a 1-hr observational time window, a wind nudging coefficient of 6.4×10^{-3} s, and a 100-km horizontal radius of influence.

Statistical metrics were computed using NIST tower observations in the DCBA domain (shown in Table S1 in Supporting Information S1) to evaluate the WRF meteorological fields later used for dispersion simulations. The mean absolute error (MAE) was in the range of 0.849°C–1.138°C for temperature, 0.326 ms^{-1} to 0.597 ms^{-1} for wind speed, and 5.332°–20.729° for wind direction. We also computed the statistics using surface observations from three major airports in the DCBA region that were not included in the nudging process (shown in Table S2 in Supporting Information S1). The maximum MAE values from these independent sites were 1.556°C for temperature, 1.386 ms^{-1} for wind speed, and 25.982° for wind direction. These results fall within the statistical benchmark established for air quality studies by Emery and Tai (2001), which recommend MAEs < 2 K for temperature, <2 ms^{-1} for wind speed, and <30° for wind direction. Furthermore, our evaluation aligns with Deng et al. (2017), who reported comparable MAEs (2.2°C for temperature, 0.8 ms^{-1} for wind speed, and 19° for wind direction) for WRF simulations used to drive an urban CO_2 inversion system, demonstrating that the WRF meteorological fields in this study meet established performance standards for urban modeling applications.

2.3.1.2. Lagrangian Particle Dispersion Modeling (HYSPLIT)

The WRF simulations described above are used to drive HYSPLIT (Draxler & Hess, 1998; Stein et al., 2015) backward dispersion simulations to obtain footprints for each measurement. Simulations are initiated at each location every hour and run for 12 hr backward in time, with 1,000 particles released throughout the first hour of each simulation to correspond with the hourly averaged measurement. The simulation is run with vertical velocity computed from the horizontal velocity divergence, rapid convection when CAPE exceeded 500 J/kg (Draxler & Hess, 1997), boundary layer stability derived from heat and momentum fluxes (Draxler & Hess, 1997), the Beljaars and Holtslag (1991) turbulence scheme for stable surface conditions and the Betchov and Yaglom (1971) turbulence scheme for unstable conditions (Draxler & Hess, 1997), boundary layer height taken directly from the input meteorology with a minimum boundary layer height of 150 m, and the STILT dispersion scheme (Lin et al., 2003; Loughner et al., 2021). The minimum boundary layer height is applied at all locations and all times in order to prevent the output vertical resolution from being too fine, which can result in noisy results. The output vertical resolution of the calculated footprint is 0.5× the boundary layer height.

2.3.2. Observational Data and Error Covariance Specification

The NIST urban testbed system is a set of dense surface tower networks that observe inert trace gas atmospheric concentrations (details in Text S2 in Supporting Information S1). One testbed is located in Indianapolis, another in Los Angeles, and the third is situated in the Northeast Corridor, with a special focus on the DCBA. For our first attempt with this urban system, we are focused on the NIST-NEC-DCBA measurements.

Hourly urban atmospheric enhancements of the inert tracer are given by (Yadav et al., 2021),

$$\mathbf{z} = \mathbf{z}_{\text{meas}} - \mathbf{z}_{\text{back}} - \mathbf{z}_{\text{bio}}, \tag{3}$$

where $\mathbf{z}_{\text{meas}}$ represents atmospheric concentration measurements of CO_2 from NIST-NEC-DCBA, $\mathbf{z}_{\text{back}}$ the background calculated concentrations, and $\mathbf{z}_{\text{bio}}$ atmospheric CO_2 enhancements associated with biosphere activity. $\mathbf{z}_{\text{back}}$ is estimated from mean afternoon values derived from the background towers located at the outskirts of the domain, similarly to Karion et al. (2021) after subtracting the in-domain contribution estimates (details in Text S3 in Supporting Information S1). $\mathbf{z}_{\text{bio}}$ consists of modeled atmospheric enhancements generated from Vegetation Photosynthesis and Respiration Model (VPRM) fluxes from Gourdji et al. (2022) and Gourdji and Marrs (2021). Fluxes are regridded on the model domain and convolved with the HYSPLIT footprints calculated for the system. Given that the case study is conducted during the winter, we have assumed that the biospheric fluxes are perfectly well known (details in Text S4 in Supporting Information S1). Therefore, the resultant values from Equation 3 represent tracer enhancements associated with local urban and industrial sources, and these are assimilated by the inversion system.

The model-data mismatch covariance matrix (**R**) represents the uncertainty of the observational constraint, which includes measurement uncertainties and modeling that relates surface fluxes to observational signals uncertainties (Kountouris et al., 2018). Here, **R** is expressed as a diagonal matrix, as in previous urban inversion studies of inert trace gases (e.g., Gourdji et al., 2018; Hajny et al., 2022; Huang et al., 2019; Karion, Ghosh, et al., 2023; Kort et al., 2013; Lopez-Coto et al., 2017; Vardag & Maiwald, 2024), where diagonal elements (variances) have four independent contributions based on Karion, Ghosh, et al. (2023) (details in Text S6.1 in Supporting Information S1),

$$\sigma_r^2 = \frac{(\text{OBS}_{\text{sd}})^2}{N} + (\text{OBS}_{\text{unc}})^2 + (\text{bg})^2 + \text{var}(H_{\text{sp}}), \tag{4}$$

where the first term is the standard deviation of the 1-min observational means within each hour, divided by the number of minutes sampled (N); the second term is the measurement uncertainty; the third term represents the uncertainty in the background CO_2 estimation method; and the fourth term represents the atmospheric transport modeling uncertainty estimated through the intermodel variance of simulated CO_2 atmospheric enhancements, as in previous works (e.g., Engelen et al., 2002; Hajny et al., 2022; Lopez-Coto, Ren, et al., 2020; Pitt et al., 2022). For the latter, HYSPLIT is driven by two different high-resolution meteorological models: an in-house developed research model (i.e., 1-km nudged WRF) and an operational model (3-km HRRR), whose resulting footprints are convolved with the prior to obtain the modeled enhancements.

The approach implemented in Equation 4 results in a mean combined standard uncertainty (i.e., the average of individual $\sqrt{\sigma_r^2}$ values across each NIST-NEC-DCBA site) ranging from 3.84 to 5.49 ppm, while the signal-to-noise ratio ranges from 0.53 to 1.3, depending on the NIST site and prior fluxes used to calculate the spread of simulated CO_2 atmospheric concentrations. The signal-to-noise ratio per site is defined here as the ratio between the mean value of the atmospheric enhancement and the mean combined standard uncertainty.

2.3.3. Prior Emissions and Error Covariance Specification

Another key input to the Bayesian inversion system is a priori information of the urban and industrial emissions for the region that is improved to a posteriori when the tracer enhancements are assimilated (Singh et al., 2015). As mentioned earlier, the prior information is provided by two different emission inventories resolved spatially and temporally which differ in their native spatial resolution, methods and data sources: 1-km hourly Vulcan v3.0

over January 2015, regridded on the domain and shifted to match the day of the week in January 2019 (prior 1); and an early release of the 4-km hourly GRA2PES product over January 2019, regridded on the domain as well (prior 2) (details in Text S5 in Supporting Information S1).

The prior error covariance matrix (**Q**), which represents the uncertainty in the prior fluxes, is constructed using a decaying exponential function in space and time as follows,

$$\mathbf{Q} = \left(\sigma\sigma^{T}\right) \times (\mathbf{D} \otimes \mathbf{E}), \tag{5}$$

where σ ($m \times 1$) is a vector of time-varying prior uncertainties, $\times$ denotes element-by-element multiplication, $\otimes$ denotes the Kronecker product, and **D** ($p \times p$) and **E** ($r \times r$) are square and symmetric matrices that describe temporal and spatial isotropic correlations of the prior errors. Given the available information about the uncertainties of the priors (Gurney et al., 2020a, 2020b), the construction of **Q** for this case study involves prescribing $1\sigma = 50\%$ the prior flux magnitude at the grid cell level (details in Text S6.2 in Supporting Information S1). This uncertainty value is also considered appropriate to minimize flux overfitting at the pixel level based on preliminary work.

The temporal and spatial correlation patterns needed to construct **D** and **E** in Equation 5 are estimated, respectively, based on autocorrelation and semivariogram analyses of the hourly afternoon prior errors (details in Text S6.2 in Supporting Information S1). Errors for this purpose are estimated by two methods: (a) 1σ, and (b) based on a surrogate state, utilized as the "true" fluxes (1-km, 1-hr ACES January 2017; Gately & Hutyra, 2022) to measure errors according to Bannister (2008). The resultant autocorrelograms show that afternoon temporal correlations at the metropolitan scale have a distinctive shape, characterized by a consistent weekly-cosinusoidal pattern of the error correlations in DCBA from both 1σ and the errors derived from a surrogate state (Figure S5 in Supporting Information S1). This persistent pattern suggests that the uncertainty of afternoon urban emissions on a given day of the week (e.g., Thursday) can predict, to some extent, the uncertainty of urban emissions on the afternoon of the same day the following week, as well as 2 and 3 weeks after that day. Therefore, to represent the weekly cycle of the uncertainty described by the autocorrelograms, the temporal correlations of prior errors in this study are given by,

$$\mathbf{D} = \left[\left(\exp\left(-\frac{X_\tau}{l_\tau}\right)\right)\left(\frac{\cos(2\pi X_T F) + 1}{2}\right)\right], \tag{6}$$

where F is a weekly frequency in cycles per day ($F = 0.143$), X_τ is the lag time expressed in days, and $l_\tau =$ 14 days. Correlations are assumed only across days at the same time of the day (i.e., intraday correlations are ignored and given a value of zero). With respect to spatial correlations, semivariogram analyses indicated a correlation length of 10 km. This parameter ranges from 6 to 25 km, although 10 km is the most typical quantity (e.g., Ghosh et al., 2021; Kunik et al., 2019; Lopez-Coto, Ren, et al., 2020; Lopez-Coto et al., 2017; Pitt et al., 2022).

2.4. Observing System Simulation Experiments

Tests 01-s and 02-s are built with the same general settings as the actual data tests 01-a and 02-a, including assimilation time interval (3–29 January 2019), prior fluxes, atmospheric transport model, domain, and grid resolution (Table 1). Likewise, only afternoon tracer observations (12:00–17:00 ET) are assimilated. For the OSSEs, synthetic true observations for the tower sites are derived by transporting surface fluxes in the DCBA domain from the ACES emission data product for January 2017 using our HYSPLIT calculated footprints (**H**). Again, ACES is a high-resolution (1-km, 1-hr) CO_2 emissions data product, similar to Vulcan v3.0, but with data available closer to the study period. Figure 2 presents the spatial and temporal patterns of the reference flux scenarios adopted in the OSSEs, including the synthetic true and priors: 1-km Vulcan v3.0 January 2015 (prior 1/tests 01-s) and 4-km GRA2PES January 2019 (prior 2/tests 02-s). The Vulcan 2015 fluxes are significantly larger than those in the truth and GRA2PES inventories (Figures 2a–2c). This is especially true on weekdays, where those differences can reach up to 100% of the truth (i.e., ∼4 μmol/m^2/s) in terms of average domain flux (Figure 2d).

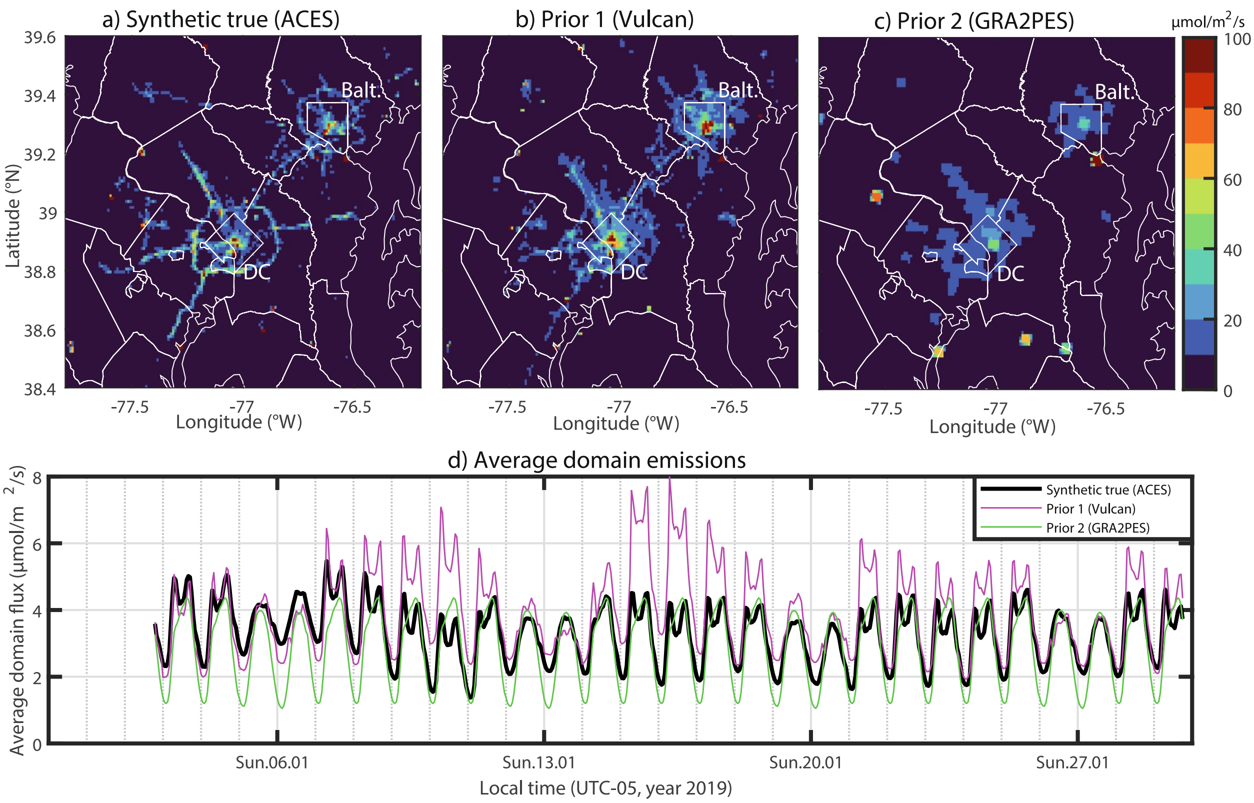

Figure 2. Reference flux scenarios adopted in the synthetic data numerical experiments 01-s and 02-s. The figure displays maps of the averaged fluxes of (a) synthetic true (1-km ACES), (b) prior 1 (1-km Vulcan), and (c) prior 2 (4-km GRA2PES) for 3–29 January 2019; and (d) the time series of average domain flux related to each reference scenario. The colorbar in maps is truncated.

To mimic the errors associated with the NIST-NEC-DCBA tower network and their effects on the flux retrievals, synthetic observations are perturbed with Gaussian noise scaled proportionally to the mean combined standard uncertainty of the actual tracer enhancements. We used their site-dependent signal-to-noise ratios (Tables S3 and S4 in Supporting Information S1), analogously to Hu et al. (2015). Furthermore, the magnitude of the Gaussian noise is expressed as variance in the diagonal **R** matrix to maintain consistency (e.g., Berchet et al., 2021; Lopez-Coto et al., 2017) and **Q** is treated with the same assumptions as the prior error covariance matrix in the actual data inversions. In this way, OSSEs use analogous parameterizations for the error covariance operators.

However, we note that the baseline OSSEs do not explicitly represent incoming airflow (background) or biogenic contributions, and that the same atmospheric transport is used in both inversions and the synthetic data generation. Furthermore, given that the background was not considered, the reference experiments assimilate all the afternoon data. Nonetheless, because both background and biogenic contributions impact the signal-to-noise ratio, the procedure used to construct the synthetic observations indirectly reflects their effects. Details on the OSSE setup are described in Text S7 in Supporting Information S1.

2.5. Metrics and Assessment Criteria

OSSEs are adopted as a diagnostic tool at both the aggregate (city) scale and grid scale levels. At the grid scale level, we use metrics such as Normalized Mean Bias (NMB) and Mean Absolute Percentage Error (MAPE) to assess model bias and accuracy (Boylan & Russell, 2006; Morley et al., 2018),

$$\mathrm{NMB}_j = \frac{\sum_{k=1}^{C}\left(x_{0,k} - x_{t,k}\right)}{\sum_{k=1}^{C} x_{t,k}} \cdot 100, \text{and} \tag{7}$$

$$\mathrm{MAPE}_j = \frac{1}{C}\sum_{k=1}^{C}\left|\frac{x_{0,k} - x_{t,k}}{x_{t,k}}\right| \cdot 100, \tag{8}$$

where $x_{0,k}$ and $x_{t,k}$ are, respectively, the prediction (from the prior or posterior state) and true flux values from ACES at the grid cell j, and C is the number of afternoon flux time steps.

The statistical self-consistency of the inversion system is evaluated utilizing the reduced chi-square value (χ^2_{red}) under the assumption of Gaussian and unbiased errors (Michalak et al., 2017; Tarantola, 2005; Yadav et al., 2019) as follows,

$$\chi^2 = (\mathbf{z} - \mathbf{H}\hat{\mathbf{s}})^{\mathrm{T}}\, \mathbf{R}^{-1}\, (\mathbf{z} - \mathbf{H}\hat{\mathbf{s}}) + (\hat{\mathbf{s}} - \mathbf{s}_{\mathrm{p}})^{\mathrm{T}}\, \mathbf{Q}^{-1}\, (\hat{\mathbf{s}} - \mathbf{s}_{\mathrm{p}}), \text{and} \tag{9}$$

$$\chi^2_{\mathrm{red}} = \frac{\chi^2}{v}, \tag{10}$$

where v is the number of degrees of freedom (i.e., the number of observations), and the other terms have been defined previously. Since computing the solution is numerically expensive, we use the methods described by Kunik et al. (2019) (details in Text S8 in Supporting Information S1). $\chi^2_{\mathrm{red}} \approx 1$ is generally perceived as an indicator of correct assumptions made in **Q** and **R**, as it indicates a close alignment of the prescribed errors with model residuals (e.g., Karion, Ghosh, et al., 2023; Kunik et al., 2019; Remaud et al., 2022; Yadav et al., 2019, 2021).

Other metrics involved comparing the prior and posterior analytical uncertainties to calculate the uncertainty reduction to provide insights into the increase in confidence in the inverse fluxes (details in Text S9 in Supporting Information S1). Furthermore, correlation, Root Mean Square Error (RMSE), and bias of modeled enhancements allow for verifying whether the posterior fluxes capture the data (e.g., Kunik et al., 2019; Yadav et al., 2021).

During the preparation of this manuscript, Grammarly, Microsoft Copilot, and Gemini were utilized as editing tools to improve the readability and flow of the text. The authors carefully reviewed and edited the outputs and take full responsibility for the content.

3. Top-Down Trace Gas Flux Quantification in DCBA (January 2019): Results and Discussion

3.1. Analysis and Interpretation of OSSEs

3.1.1. OSSEs Performance at the City-Scale Level

In this and the following subsections, the results of the baseline experiments are examined for the 12:00–17:00 ET emission interval. Figure 3 shows domain scale top-down (posterior) estimates, compared to the truth and prior. The left panel (Figure 3a) presents the central tendency indicators and variability of the total emission rate (i.e., aggregated flux values across all pixels within the domain area at each hourly time step). The truth exhibits 73.2 kmol/s ± 8.1 kmol/s (mean ±1 standard deviation) compared to prior 1/tests 01-s and prior 2/tests 02-s that show, respectively, 87.6 kmol/s ± 15.4 kmol/s and 76.3 kmol/s ± 3.4 kmol/s. In contrast, the posterior estimates reveal, respectively, 74.8 kmol/s ± 12.3 kmol/s and 72.4 kmol/s ± 4.3 kmol/s, with the mean of both posteriors of 73.6 kmol/s ± 9.3 kmol/s, as shown by the horizontal dashed line in Figure 3a. Clearly, posterior estimates reflect the constraining data, as they are more closely aligned with the truth. Initially, the estimates from the prior 1/tests 01-s showed the greatest bias and dispersion relative to the truth (Figure 3a, second box). The system adjusted the variability of the total emission rate—that from prior 1 has decreased (Figure 3a, fourth box) and that from prior 2 has increased (Figure 3a, fifth box)—such that the posterior data are in greater agreement with the true values. Therefore, this analysis demonstrates that the model reduces bias and adjusts the prior dispersion toward more realistic values.

However, while both OSSEs have used the same observations and prior error spatial and temporal correlation scales, the variability associated with the posterior estimates was not identical (Figure 3a). This is expected and is primarily attributed to the inherent characteristics of the prior flux fields and the grid cell level uncertainty (1σ)

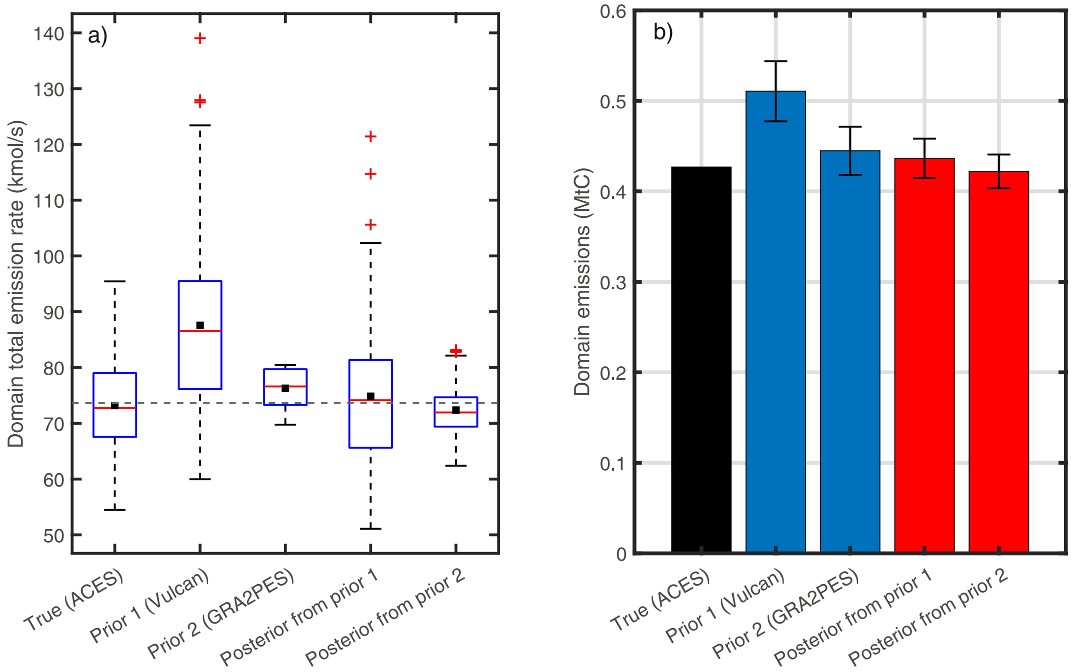

Figure 3. Posterior flux estimates derived from tests 01-s and 02-s, with comparisons made against the synthetic true and priors to evaluate model performance on a metropolitan scale. (a) Boxplot of total domain CO_2 emission rate and (b) bar chart of cumulative CO_2 emissions (in carbon equivalents) for 3–29 January 2019, during afternoon hours (12:00–17:00 ET). In panel (a), the bottom and top box edges indicate the 25th and 75th percentiles; the whiskers extend above and below the box up to 1.5 times the interquartile range; the + markers indicate the outliers; the red line represents the median; the black square markers the mean; and the horizontal dotted line the mean of both posteriors. The error bars in panel (b) represent the uncertainty of aggregated emissions (details in Text S9 in Supporting Information S1), shown as 95% confidence intervals.

assumed for the priors, corresponding to 50% of the grid-cell level prior fluxes in the baseline inversions, as this provides inversions with leeway to adjust the fluxes (e.g., Pitt et al., 2022; Yadav et al., 2021). Since the uncertainty is linked to the prior structure, it differs for each case in terms of granularity and magnitude.

To assess the total effect over time, Figure 3b compares the sum of total emission rates over the month for the truth, prior, and posterior scenarios in carbon mass units (i.e., cumulative monthly emissions). The associated uncertainty is expressed as 95% confidence intervals (CI). From this assessment, it follows that cumulative emissions arising from the synthetic true are 0.427 megatons of carbon (MtC), and prior estimates exhibit an overestimation compared to the truth, reporting emissions of 0.511 MtC ± 0.033 MtC, 95% CI (0.445 MtC ± 0.027 MtC, 95% CI) from the prior 1/Test 01-s (prior 2/Test 02-s) (Figure 3b). It is of particular interest to note that the prior 1 does not match the truth within the confidence intervals. The system adjusted these estimates downward with narrower confidence intervals, corresponding to 0.436 MtC ± 0.022 MtC, 95% CI (0.422 MtC ± 0.019 MtC, 95% CI) for the posterior 1/Test 01-s (posterior 2/Test 02-s) (Figure 3b). This led to posterior cumulative emissions that agreed with the truth within 2.3% (1.1%) within the confidence intervals, indicating that the model leads to a stable solution and is reliable for total mass accounting of the emitted tracer across DCBA.

While these baseline results demonstrate the overall stability of the system, to further examine the convergence of total estimates toward the observational constraints, we performed additional sensitivity inversions by setting the prior uncertainty (1σ) to 100% and 200% of the prior fluxes (details in Text S10 in Supporting Information S1). These additional analyses reinforce the idea that the system is driven by a top-down signal, as increasing the error in the prior leads to greater reductions in uncertainty. However, for the cases evaluated here, we observed that sensitivity inversions of the Test 01-s succeeded in reducing discrepancies relative to the true value at the domain level (i.e., by 0.6%–1%), whereas sensitivity inversions of the Test 02-s did not (i.e., by 2.1%–3.5%). Nevertheless, a closer examination at the subdomain level in the Washington, D.C., and Baltimore, MD, areas reveals that this is a consequence of the physical adjustments made by the system to correct spatial error cancellations in

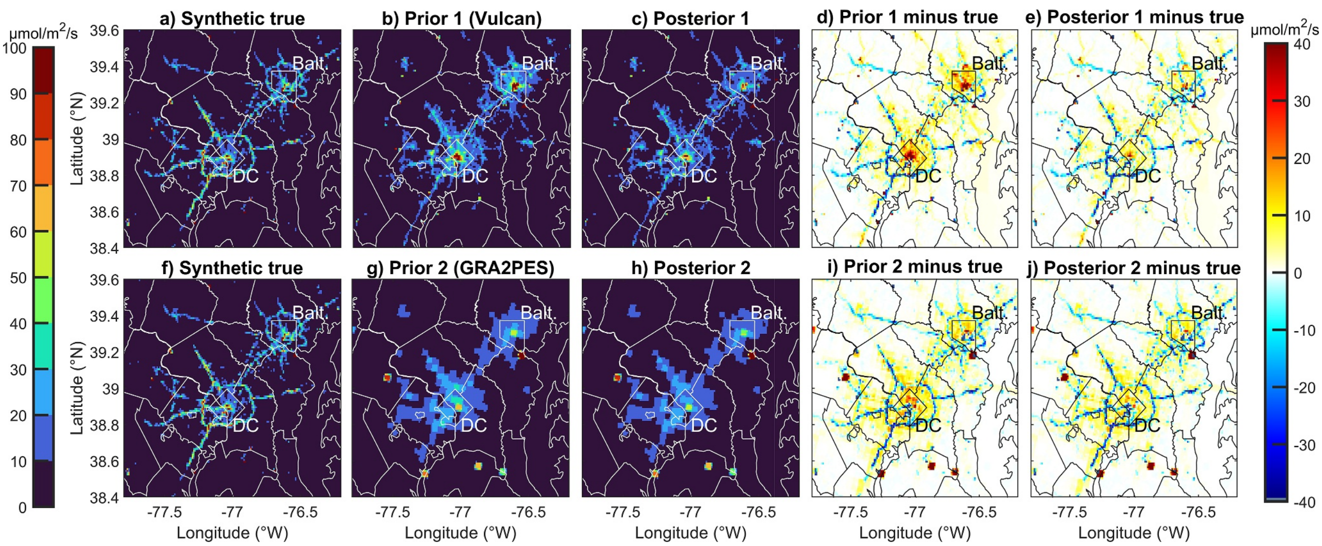

Figure 4. Comparison of synthetic true, prior, and posterior flux estimates at the pixel level. The top panel presents test 01-s-derived estimates, showing afternoon-averaged (12:00–17:00 ET) flux maps of (a) synthetic true, (b) prior, (c) posterior, and residuals: (d) prior minus true and (e) posterior minus true. The bottom panel (f–j) presents the same analyses for test 02-s-derived estimates. Both colorbars are truncated.

the prior inventory and converge toward the true value in those specific regions, where the uncertainty reduction is even more pronounced (e.g., up to 66.7% in Baltimore, and 64.2% in DC) due to the concentration of the observational network.

3.1.2. OSSEs Performance at the Grid-Scale Level

Figure 4 presents 1-km resolution flux maps averaged over the afternoon (12:00–17:00 ET) for the synthetic true, prior, and posterior scenarios, as well as average emission residuals (prior minus true, and posterior minus true). The upper row shows test 01-s-derived estimates (Figures 4a–4e), while the bottom row shows test 02-s estimates (Figures 4f–4j). By examining the first row, particularly the last two columns (Figures 4d and 4e), it is visibly discernible that the inversion system yielded a noticeable reduction in residuals across multiple pixels, especially in the Washington DC and Baltimore MD urban cores [i.e., mean emissions differences from ∼35 μmol/m^2/s (prior minus true; Figure 4d) to ∼10 μmol/m2/s (posterior minus true; Figure 4e)]. This is related to the fact that the Vulcan data set is a high-resolution prior that contains more information about fine-scale features, which are often co-located with the synthetic true features from ACES, thereby leading to the correct assignment of high prior uncertainty values to those pixels and greater flexibility for the inversion to adjust them. For the case of the prior 2/test 02-s (bottom row; Figures 4f–4j), the differences between the spatial patterns of the original and final emission residuals were less pronounced (Figures 4i and 4j). The main factor contributing to the observed result was the limited spatial granularity of the coarser 4-km-resolution GRA2PES, which led to the loss of fine-scale features and hindered accurate uncertainty assignment.

The NMB was calculated to quantify the model bias at the pixel level, and Figures 5a–5d present the spatial distribution of the estimated afternoon NMB (12:00–17:00 ET) in the central subdomain area. By comparing Figures 5a and 5b (test 01-s) and Figures 5c and 5d (test 02-s), it is confirmed that the test 01-s-derived estimates achieved the largest bias reduction: NMB medians shifted from 89%, 58%, and 70% in the prior state (Figure 5a) to 29%, 25%, and 54% in the posterior state (Figure 5b), respectively, for the areas within the beltways surrounding Baltimore, MD [delimited here by (−76.5° longitude; 39.38° latitude; −76.74° longitude; 39.19° latitude)], Washington, DC (−76.82° longitude; 39.02° latitude; −77.18° longitude; 38.78° latitude), and the entire central subdomain region. This NMB behavior indicates that the use of a high-resolution prior (1-km Vulcan) yields a significant reduction in systematic errors in hotspot areas, where higher prior uncertainty values coincide with a greater density of the NIST tower network (i.e., Baltimore and DC; see Figure 1).

In terms of MAPE, Figures 5e–5h reveal that test 01-s-derived estimates were more accurate. MAPE medians shifted from 94%, 74%, and 82% in the prior state (Figure 5e) to 54%, 52%, and 69% in the posterior state

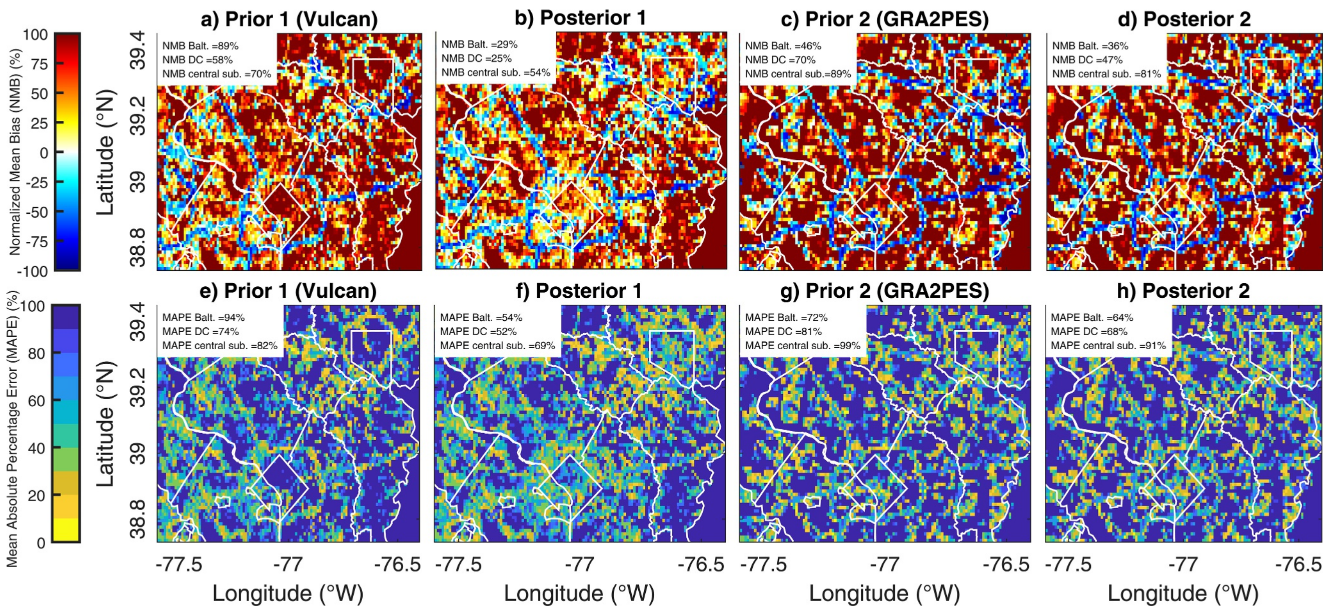

Figure 5. Maps of tests 01-s and 02-s-derived NMB and MAPE in the afternoon time window (12:00–17:00 ET) in the central subdomain region. The top row presents the NMB values for prior 1 (1 km Vulcan; a), posterior 1 (b), prior 2 (4 km GRA2PES; c), and posterior 2 (d). Bottom row (panels e–h) shows the MAPE values for the same flux scenarios. The median values for the areas within the beltways surrounding Baltimore, MD, and Washington, DC, are indicated in each panel, as well as the median for all the central subdomain areas. Both colorbars are truncated.

(Figure 5f), respectively, for the areas within the beltways surrounding Baltimore, MD, Washington, DC, and the entire central subdomain region. The metrics obtained were less ideal than the NMB, suggesting that the system struggles to accurately assign hourly flux peaks, since MAPE penalizes overprediction more severely (Morley et al., 2018). This finding is not unexpected, given the sparsity and relatively high noise of the observations used to pinpoint where tracer releases are occurring. To the best of our knowledge, flux accuracy metrics are not typically reported at the pixel level in urban inversions and other evaluations based on aggregated scales, or the downwind observations themselves are preferred (Michalak et al., 2017). Regardless, synthetic data inversions with Vulcan are found to capture temporal variability (approximately half in DC and Baltimore), which can be reconstructed more accurately when flux data are aggregated over the domain area, as explained in Figure 3a.

We performed an additional analysis with fluxes regridded at a coarser spatial resolution of 4 km to align with the native scale of prior 2 (Figure S2 in Supporting Information S1). By regridding to a 4-km nominal resolution, both systematic and overall discrepancies were further reduced. Although inverse fluxes derived from test 01-s still achieved the best performance, we estimate that median NMB values derived from test 02-s more closely aligned with them (denoted below in parentheses): −1% (6%); 2% (20%); and 39% (45%) for the areas within the beltways surrounding Baltimore, MD, Washington DC, and the entire central subdomain region, respectively. Regarding the MAPE, the corresponding median values were 31% (40%), 36% (37%), and 52% (58%) for test 01-s (test 02-s). This evidence indicates that posterior fluxes become less biased and noisy, suggesting that solving for fluxes at a 1 km grid resolution, followed by a regridding to a coarser spatial resolution (4 km for the cases analyzed here), provides a means to improve total mass counting and attribution at the individual grid cell level, particularly within the Baltimore and DC beltways. The cause of this behavior lies not only in the aggregation of fluxes but also, significantly, in atmospheric transport, since the signal observed at the measurement sites is amplified when the mass originating from individual 1 km cells is aggregated into 4 km (16 cells). Consequently, the observation network exhibits greater representativeness regarding fluxes aggregated at this spatial resolution.

To summarize, the overall performance of the inverse model at the pixel level is fairly good. Systematic errors at the pixel level are significantly reduced in hotspot regions, primarily when the system uses a 1-km resolution prior. This means that granular details of the urban corridor can be more effectively resolved when the prior estimates contain a similar resolution to the truth. This idea is consistent with Lopez-Coto, Ren, et al. (2020), who found that fine-scale features are only resolved in the modeling results when high-resolution data products are used as priors. McDonald et al. (2014) similarly found that transport infrastructure cannot be resolved in emission inventories until getting a resolution of 1 km or finer. Finally, there is still room for improvement in terms of hourly flux variation at the 1-km pixel level.

3.1.3. Statistical Self-Consistency of OSSEs

In the inversion, the prior and observational error covariance operators (**Q** and **R** in Equation 1) weight the solution, assigning greater weight to the prior or the observations depending on the associated uncertainty. Therefore, it is essential to employ tools to assess whether the uncertainties required by the inversion technique have been correctly quantified. In this work, we computed the reduced chi-square (χ^2_{red}) on the model residuals arising from each OSSE to gain insights into the statistical consistency of the inverse fluxes. Consequently, the reduced chi-square values derived were 1 and 0.99, respectively, for the tests 01-s and 02-s; that is, in both cases, a value close to or equal to the ideal value of one was achieved (e.g., Karion, Ghosh, et al., 2023; Kunik et al., 2019; Remaud et al., 2022; Yadav et al., 2019, 2021). These results suggest that the model fit is reasonable, given that the model residuals are consistent with the prescribed errors at the collective level. This indicates that the assumptions made to construct both **Q** and **R** are acceptable from a joint perspective.

3.2. Analysis and Interpretation of Inversions From Observational Data

Grounded in OSSE-derived confidence, we now examine the outcomes derived from the actual measurements. The system configuration and assimilation setup from the OSSEs were preserved, with only some parameters pertinent to the actual data inversions being changed. The constraining atmospheric enhancements over DCBA were calculated by subtracting estimates of the background and biogenic contributions from the observed mole fractions at the towers, as explained in Equation 3. Regarding the background calculation procedure, out of 1,600 afternoon concentration measurements that were made during the time period of 1 January–1 February 2019, 347 (21.69%) were discarded due to either particles not exiting the domain or exiting toward the northeast quadrant where there is no background site. Of those, 1 (0.06%) was due to particles not exiting the domain (i.e., a 12-hr HYSPLIT simulation was not long enough for the particles to exit), 341 (21.31%) were due to particles exiting toward the northeast, and 5 (0.31%) were due to a combination of particles not exiting the domain and particles exiting toward the northeast.

To quantify the impact of missing background information on constraining fluxes and, consequently, uncertainty reduction, we compared the baseline synthetic data experiments 01-s and 02-s with two different sensitivity cases. The baseline experiments used a configuration file that incorporated all afternoon measurement data within the assimilation time interval (i.e., 3–29 January 2019) while the sensitivity cases used a configuration file that did not include measurement data due to missing background information (i.e., sensitivity inversions B1 and B2; see Table S6 in Supporting Information S1). These additional analyses, as expected, showed a smaller uncertainty reduction compared to the baseline OSSEs, although the differences were not substantial (i.e., 29.8%–34.6% for baseline OSSEs; 26.1%–30.7% for sensitivity runs), while the posterior cumulative emissions continued to overlap with the true value within the 95% CI (Figure S10 in Supporting Information S1). Therefore, this behavior suggests that the inversion system remains reliable for counting the tracer emissions at the city-scale level despite the data wasted/unused due to the lack of background information.

3.2.1. Emission Rates and Monthly Cumulative Emissions

Afternoon (12:00-17:00 ET) total emission rates and monthly cumulative emissions in DCBA are presented in Figure 6, with comparisons made against the bottom-up (prior) emission estimates. The boxplots in the left panel (Figure 6a) show that posterior estimates for the total emission rate are 62.7 kmol/s ± 11.7 kmol/s (mean ±1 standard deviation) and 60.1 kmol/s ± 8.4 kmol/s compared to, respectively, 87.6 kmol/s ± 15.4 kmol/s (prior 1/tests 01-a) and 76.3 kmol/s ± 3.4 kmol/s (prior 2/tests 02-a). Both posterior distributions have shifted downward with respect to the prior states and are centered very close to 61.4 kmol/s, the mean emission rate of both posteriors. The system also adjusted the variability, notably for the prior 2 (GRA2PES 2019; Figure 6a, fourth box), where the spread was increased in the posterior estimates, allowing for a closer overlap between the two posterior distributions. Similar behavior was detected in the OSSEs, where the system also adjusted the variability in the posteriors to account for the synthetic true flux spread. Therefore, we can interpret this behavior as the system capturing the actual variability of the fluxes in DCBA.

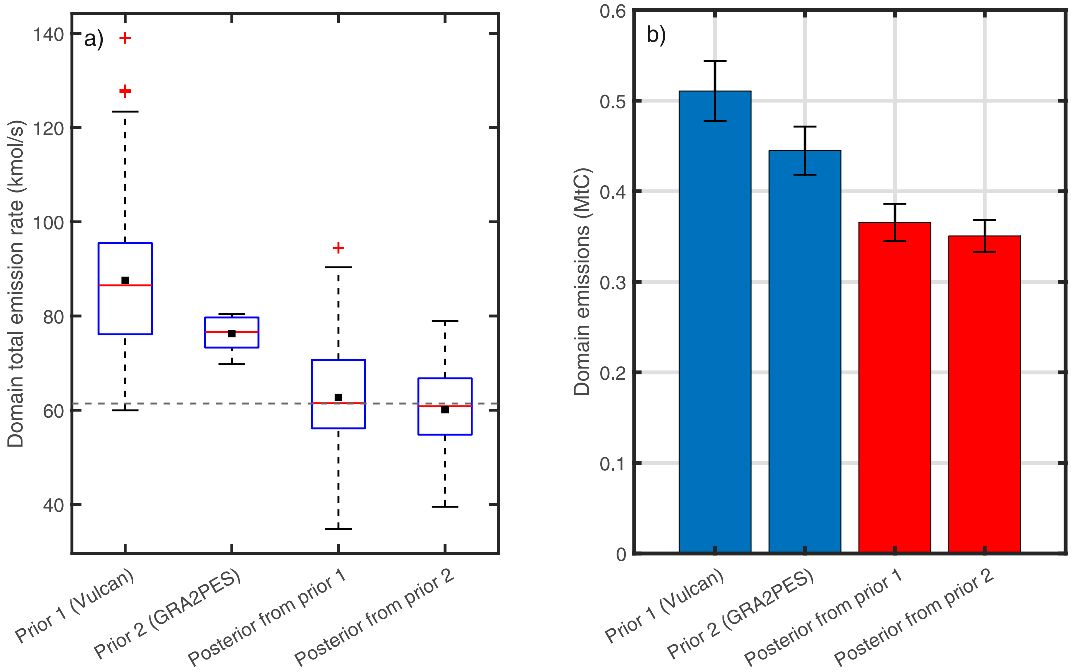

Figure 6. Same as Figure 3, but derived from actual data numerical experiments 01-a and 02-a.

Estimates of afternoon monthly emissions in DCBA from the inversion system are presented in Figure 6b. This analysis reveals 0.366 MtC ± 0.021 MtC, 95% CI (0.351 MtC ± 0.017 MtC, 95% CI) for the posterior 1/tests 01-a (posterior 2/tests 02-a), and, again, 0.511 MtC ± 0.033 MtC, 95% CI (0.445 MtC ± 0.027 MtC, 95% CI) for the prior 1/test 01-s (prior 2/test 02-s). As in the OSSEs, both posteriors agree within narrower confidence intervals, which indicates that the inversion system not only detects lower emissions for January 2019 but also increases the accuracy of the total monthly emissions in DCBA. The uncertainty reduction achieved by both synthetic and actual data inversions is discussed further in Section 3.2.3. Concerning the existing discrepancy between the bottom-up and top-down approaches, we deduce that the mismatch with Vulcan 2015 is more significant as it reports a 4-year lagged emission period (January 2015) compared to the atmospheric observations from which emissions are inferred (January 2019).

The emission rate estimates for January 2019 can be compared with Lopez-Coto, Ren, et al. (2020), who used two airborne platforms that flew simultaneously over DCBA during 5 days in February 2016, mainly during the afternoon hours. A mean domain emission rate of 87 kmol/s ± 28 kmol/s for DCBA was estimated, higher, but still consistent within one standard deviation with the top-down estimate derived by our inversion system (i.e., 61.4 kmol/s ± 10.2 kmol/s). Interestingly, the mean domain emission rate derived from Lopez-Coto, Ren, et al. (2020) is more closely aligned with the bottom-up estimates from our high-resolution prior, Vulcan January 2015 (i.e., 87.6 kmol/s ± 15.4 kmol/s; Figure 6a). This is likely related to real temporal variability or trends in urban and industrial emissions, as the study period of Lopez-Coto, Ren, et al. (2020) is closer in time to Vulcan January 2015 than the period of this study, where the time lag is approximately 3 years.

An additional earlier study by Yadav et al. (2021) obtained top-down estimates of CO_2 for January 2019 by adopting as a priori a temporally smoothed version of Vulcan over 2015. While Yadav et al. (2021) also reported lower domain-monthly emissions estimates for January 2019 than Vulcan, the two top-down estimates do not coincide within the 95% CI. According to Yadav et al. (2021), the mean domain emission rate estimated is 2.093 ktC/hr ± 0.322 ktC/hr, 95% CI, compared to 2.711 ktC/hr ± 0.152 ktC/hr, 95% CI, obtained during the afternoon in the present study. Some of the key technical aspects that may explain this variation in results include differences in the atmospheric transport model choice—HYSPLIT-WRF_1km nudged in this study versus STILT-NAMS_12km in Yadav et al. (2021)—that can contribute to a divergence in model outputs, as demonstrated by Munassar et al. (2023). Yadav et al. (2021) also assumed a larger prior uncertainty (i.e., 100% of the prior flux magnitude, compared to 50% in this study), which allows for greater leeway in adjusting the prior.

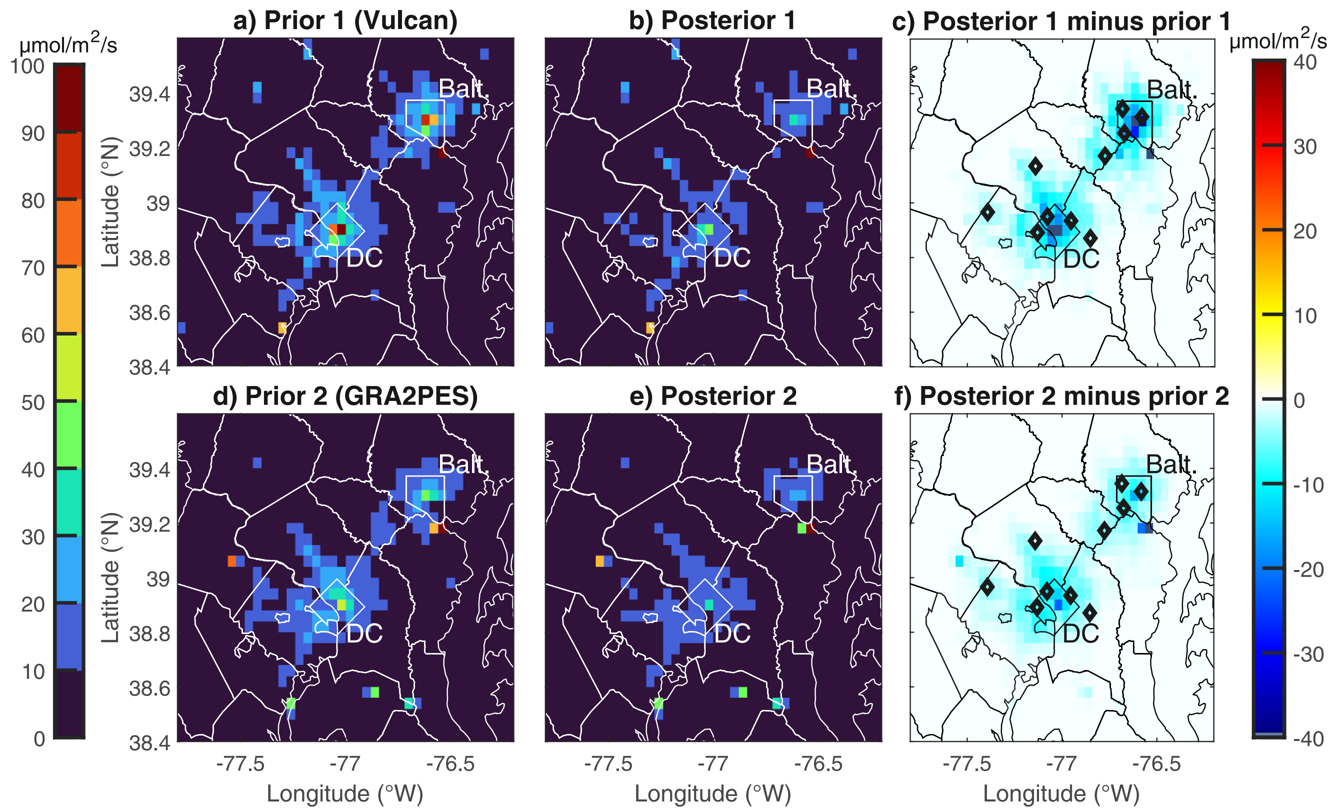

Figure 7. Top-down derived afternoon-averaged (12:00–17:00 ET) flux maps at 4 km resolution for 3–29 January 2019. The top row shows estimates from test 01-a, including (a) prior, (b) posterior, and (c) posterior minus prior. The bottom row (d–f) presents the same estimates derived from test 02-a. Both colorbars are truncated, and diamond markers in the third column indicate the location of the NIST-NEC-DCBA tower network.

Another key difference is that while both inversions used afternoon observations only, Yadav et al. (2021) solved for 4-day average fluxes rather than hourly fluxes, which allowed the system to adjust the full-day emission. In contrast, the inversion system presented here generally allows the fluxes during periods unconstrained by observations to revert to the prior. Specifically, early morning and nighttime periods tend to lack available footprints, which causes a residual retrieval in the early morning and practically no retrieval at night.

3.2.2. Spatial Distribution of Releases

Figure 7 displays afternoon (12:00–17:00 ET) average flux maps with data regridded at a nominal resolution of 4 km to provide insight into spatial patterns during January 2019. From Subsection 3.1.1, it was learned from synthetic data that larger bias and accuracy corrections can be achieved through a 4-km regridding for both tests 01-s and 02-s. Therefore, we chose to assess the spatial distribution of the releases derived from actual observations at a spatial resolution of 4 km (the native resolution of GRA2PES). In terms of the change from prior to posterior state, areas with large adjustments are revealed within the beltways surrounding Washington, DC, and Baltimore, MD. This is especially true in the test 01-a-derived posterior, where spatial features of areas emitting more than 20 μmol/m^2/s have decreased within the political boundaries of Washington, DC, and Baltimore, MD (indicated by light blue, green, and red tones in Figure 7b). This pattern also occurs across clusters of pixels both on and off the beltways.

For the case of test 02-a, which uses as prior information a contemporaneous bottom-up inventory to the observations (GRA2PES 2019), the difference in spatial patterns between the prior and posterior average maps is less pronounced (Figures 7d and 7e). However, the top-down estimates show that this inventory was overestimating by ∼10 μmol/m^2/s along several sections of the urban corridor, as evidenced by spatial features emitting 10 to 19 μmol/m^2/s (blue tones) and 20 to 29 μmol/m^2/s (light blue tones) that are reduced in size. Consequently, both top-down estimates broadly agree on the average flux patterns, especially those in the heart of

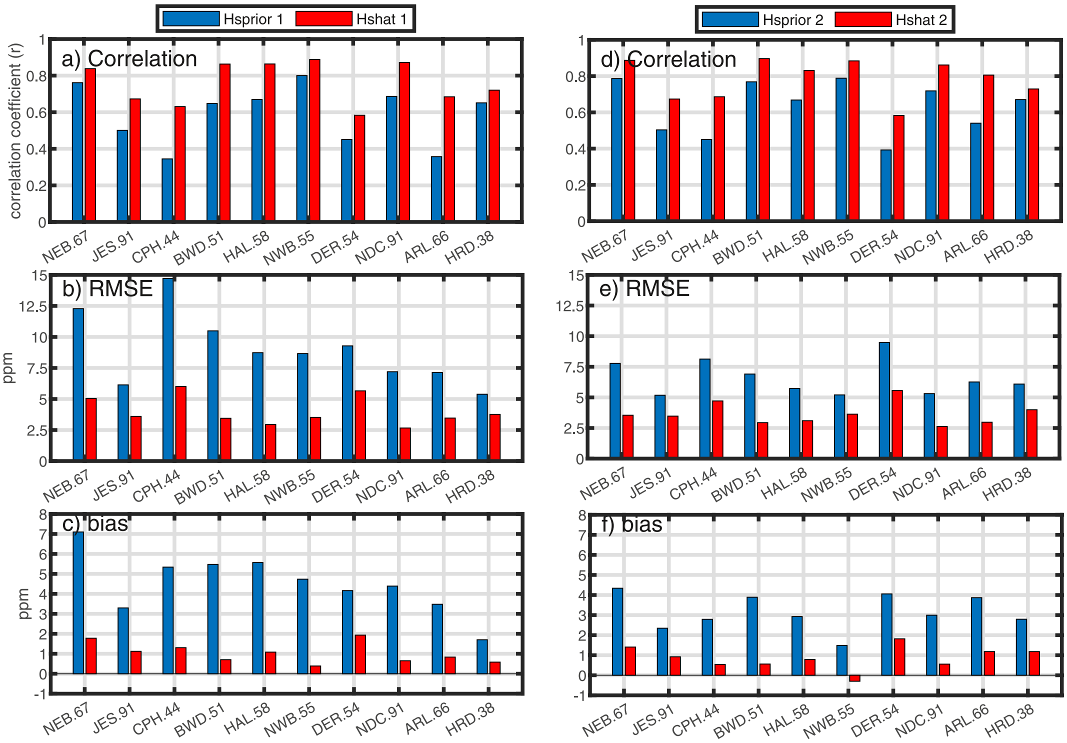

Figure 8. Statistical evaluation of model-observation agreement for the inert tracer (CO_2) at the NIST-NEC-DCBA tower sites. The left column (a–c) presents statistics obtained from test 01-a, including (a) correlation coefficient, (b) Root Mean Squared Error, and (c) bias between observed and modeled concentrations (in μmol/mol, or ppm) from both prior ($\mathbf{Hs}_p$; indicated by blue bars) and posterior fluxes ($\mathbf{H\hat{s}}$; indicated by red bars). The right column (d–f) presents the same analysis from test 02-a.

urban centers, which are dominated by features emitting 10 to 19 μmol/m^2/s (blue tones; Figures 7b and 7e). These results suggest that the use of the atmospheric tracer improved the estimation of the geographic distribution of emission sources, as observed in previous urban-scale studies (Brioude et al., 2013; Lopez-Coto, Ren, et al., 2020). Most of the corrections are negative and concentrated where greatest measurement coverage coexists with higher prior uncertainty values, as shown by the mean residual maps (posterior minus prior) presented in Figures 7c and 7f. Nevertheless, even for some pixels that are located apart from the NIST-NEC-DCBA tower network (e.g., Brandon Shores power plant) both analyses reveal significant corrections (i.e., more negative than −40 μmol/m^2/s) due to their large values of prior fluxes.

3.2.3. Diagnostic Analysis and OSSEs Results Integration

Figure 8 presents the agreement between the model and observations. To generate the model forecasts, the HYSPLIT-calculated footprints associated with each NIST-NEC-DCBA site were convolved with the surface fluxes associated with the prior ($\mathbf{Hs}_\mathrm{p}$) and posterior ($\mathbf{H\hat{s}}$) estimates. The output of this procedure is modeled atmospheric dry mole fractions of CO_2 (in units of μmol/mol, or ppm), which are then evaluated in terms of correlation coefficient, RMSE, and bias (calculated as the mean modeled enhancement minus the mean observed enhancement). Given this, the most significant changes were revealed for test 01-a, where the metrics for $\mathbf{H\hat{s}}$ showed a better agreement with observations than those corresponding to $\mathbf{Hs}_\mathrm{p}$ (Figures 8a–8c). The mean correlation coefficient across all sites was increased from 0.59 to 0.76, the mean RMSE was decreased from 9 to 4.01 ppm, and the mean bias was reduced from 4.52 to 1.04 ppm, respectively. These average values represent the general tendencies across sites, but we can highlight that for sites such as NWB.55 and NEB.67 over Baltimore, the prior bias was very significant at 4.73 and 7.09 ppm, respectively, and this was reduced by the posterior to 0.39 and 1.78 ppm, respectively (Figure 8c). For the case of the test 02-a (Figures 8d–8f), a better agreement with the observations for all NIST sites was also achieved. The mean correlation across sites was updated from 0.63 to 0.78, mean RMSE from 6.61 to 3.65 ppm, and mean bias from 3.15 to 0.86 ppm. Therefore, these analyses demonstrate that the system is effectively using the atmospheric tracer to adjust the prior estimates, and the remaining mismatch shown by the statistical metrics for $\mathbf{H\hat{s}}$ is determined by the balance between prior and

10.1029/2025JD045557

model-data uncertainties. The reduced chi-square statistic applied to our model residuals was 0.78 for both tests 01-a and 02-a. Both values are slightly less than one, but still close, suggesting that the prescribed uncertainties remain valid and that the system is statistically self-consistent.

The monthly mean afternoon (12:00–17:00 ET) uncertainty reduction across the domain (UR_{domain}) was 38.2% (test 01-a) and 34.4% (test 02-a). For the analogous synthetic data inversions, UR_{domain} was 34.6% (test 01-s) and 29.8% (test 02-s). In contrast, Lopez-Coto et al. (2017) estimated an average uncertainty reduction of 28.2%–30.5% in February 2013 for the urban land use in DCBA, depending on the compactness of the NIST-NEC-DCBA monitoring network within a synthetic data framework. Yadav et al. (2021) calculated a monthly mean uncertainty reduction close to 20% in DCBA for January 2019. Lauvaux et al. (2016) estimated an uncertainty reduction of the order of 30% for the Indianapolis urban area. While the domain uncertainty reductions achieved by our system are comparable to previous studies, it is important to clarify that the cited metrics—and those derived in this study—are not strictly equivalent. Our reported values pertain to monthly emissions across the entire domain and explicitly include covariances, following Kountouris et al. (2015) and Thompson et al. (2011) (details in Text S9 in Supporting Information S1). In contrast, Lopez-Coto et al. (2017) and Lauvaux et al. (2016) observed, respectively, the average and approximate pixel values for urban land use areas. Yadav et al. (2021) estimated the mean for the entire DCBA domain, likely without considering covariances, as they cited Deng et al. (2007). To illustrate this discrepancy, when we recalculated our uncertainty reduction following Deng et al. (2007) methodology, our estimates decreased to 16.6%–11.8% (18.1%–11.4%) for the actual (synthetic) data experiments. This significant variation shows the impact of not accounting for covariances when characterizing the uncertainty reduction for a study region.

Our analyses indicated that actual data inversions incorporated relatively more information from the observational data network compared to synthetic data inversions. Additional analyses of the spatial patterns of uncertainty reduction reinforced this systematic reduction in uncertainty (Text S12 in Supporting Information S1). The differences between the two approaches are partially explained by the construction of the synthetic atmospheric enhancements themselves, where, for convenience, certain simplifications were assumed, consisting of scaling the synthetic noise magnitude using the monthly signal-to-noise ratio of the actual observations. While this approach captures the typical magnitude of errors, the error structure in real-world observations is more complex, which can lead to further uncertainty reduction. That is, the uncertainty of the data has a time-varying atmospheric transport modeling component that sometimes imposes low reliability and, at other times, high confidence for a given atmospheric enhancement, depending on the divergence of the two HYSPLIT backward dispersion simulations used to calculate it.

3.3. Discussion and Outlook

We have evaluated a high-resolution urban inverse modeling system in DCBA during January 2019, using an inert tracer (CO_2) as a proof of concept. The system was first examined in a configuration-mirrored OSSE framework that replicates the signal-to-noise ratio of the actual data monitoring network, atmospheric transport, assimilation interval, prior fluxes, error covariance parameterizations, domain, and grid. Synthetic data experiments showed that the system can determine the underlying flux characteristics of the tracer with acceptable accuracy as measured by total emission rate, cumulative monthly emissions and pixel-level diagnostics (i.e., emission residuals, NMB, MAPE), mainly at the areas within the beltways surrounding Baltimore, MD, and Washington, DC. With increased confidence in the reliability of the fluxes derived from the system, downwind actual observations of the tracer in January 2019 revealed lower total emission rates and monthly cumulative emissions than the two bottom-up inventories adopted as priors. These top-down flux estimates for January 2019, along with the associated uncertainty reduction, align with previous top-down wintertime estimates, although some variations occur, and we identify factors that may explain them. This is especially true compared with Yadav et al. (2021), who also reported lower top-down emission estimates than Vulcan January 2015 at DCBA. Their estimates do not overlap within the 95% CI, suggesting variations resulting from the configuration of each inverse modeling system. However, the modeled enhancements from posterior fluxes consistently showed higher correlation and lower systematic and random errors than the priors, reinforcing the idea that our system is effectively linking atmospheric tracer concentration measurements to urban anthropogenic surface fluxes in DCBA.

10.1029/2025JD045557

The system utilized footprints calculated by a 1-km WRF-HYSPLIT simulation nudged with urban meteorological observations collected at multiple sites across DCBA. These high-resolution enhanced source-receptor sensitivities were utilized to constrain hourly 1-km surface fluxes, incorporating spatiotemporal correlations in **Q** to propagate the observational information. For Bayesian urban-scale inverse modeling of inert trace gases, high-resolution retrieval resolution is rare but includes 1-km, 6-hr (Breón et al., 2015; Kunik et al., 2019; Lian et al., 2022; Mallia et al., 2020); 4-km, 2-hr (Brioude et al., 2011); and 1-km, 1-hr (Wu et al., 2018). Among these studies, only Wu et al. (2018) utilized a customized WRF simulation to drive a Lagrangian Particle Dispersion Model, while only Kunik et al. (2019) explicitly approximated spatiotemporal correlations using a decaying exponential model with ordinary spatial and temporal e-folding lengths, both within an OSSE framework. Our parametrization of **Q** differs from Kunik et al. (2019) in the treatment of temporal correlations, in which, given the hourly flux resolution we aimed for in our system, the correlation structure of afternoon prior uncertainties in DCBA was estimated to follow a weekly cosinusoidal pattern, and we imposed a correlation function to represent this periodic behavior in **Q**. The application of this parameter, along with the others mentioned, was evaluated here on both synthetic and actual data frameworks.

We carried out a synthetic numerical experiment as a diagnostic tool but also performed an actual data analysis that can only succeed if the underlying urban atmospheric transport and dispersion modeling, as well as the tuning of prior error statistics, are sufficiently accurate (e.g., Martin et al., 2019). However, we must acknowledge that the overall uncertainty in the dispersion simulations, is determined by, but not limited to, the meteorological inputs. Future work should quantify total uncertainty in dispersion simulations by explicitly addressing other factors such as modeled mixing processes using ensembles (e.g., Ngan et al., 2019), and its impact on the flux retrievals. Other major contributors to the overall uncertainty budget include the inversion method and setup, as well as how the background concentration is determined.

The limits on the flux quantification that our approach can achieve were also determined. The inversions yielded a stable solution when fluxes were aggregated at the domain scale, both in synthetic and actual data frameworks. To further test this interpretation, we performed additional sensitivity inversions at coarser spatial aggregations of our 1 km spatial resolution prior Vulcan (i.e., 10 km and domain-level) with uncertainty set at 50% (baseline), 100%, and 200% of the prior flux magnitude (details in Text S11 in Supporting Information S1). This analysis suggests that there are no statistically significant differences in afternoon cumulative emissions and afternoon mean emission rates derived at 10 km resolution compared to the baseline experiments obtained at 1 km resolution (i.e., cumulative emissions within 95% posterior CIs; mean rates agree within one posterior standard deviation). However, the truth monthly emissions may not fall within the confidence intervals when using the baseline prior uncertainty (i.e., 50% of the prior fluxes). In the most extreme case, when prior spatial information is lacking (i.e., a spatially uniform or "flat" prior), total emissions reliability can only be achieved in the central region of the subdomain (delimited in Figure 1), provided that the prior uncertainty is increased up to 200% of the prior fluxes. Therefore, these additional analyses indicate that, when omitting prior information about small-scale features at much coarser spatial resolutions than 1–4 km, confidence can only reasonably be placed at the subdomain level, subject to the condition that the prior uncertainty is scaled as needed.

This study also aimed to estimate the model performance at the pixel level and found that baseline inversions at such a fine resolution of 1 km still depend heavily on prior distributions, as shown by Lopez-Coto, Ren, et al. (2020) and McDonald et al. (2014). In our OSSEs, a high-resolution (1-km) prior (prior 1/test 01-s) allowed us to retrieve fairly well the average and total flux estimates over pixels in the DC (median posterior NMB 25%; median prior NMB 58%) and Baltimore (median posterior NMB 29%; median prior NMB 89%) urban cores. However, a more modest bias reduction was observed when the prior information was provided by a different coarser 4-km emission data product (prior 2/test 02-s). A post-processing step consisting of regridding to a coarser spatial resolution (4 km) was necessary to improve the mean and cumulative flux estimates, suggesting that retrieval errors are partially explained by spatial resolution. The remaining discrepancies arise in part from the alternative methods/data sources utilized to develop the GRA2PES inventory. Furthermore, our analysis indicates that retrieving hourly temporal variations at the 1-km level is difficult, even with our high-resolution prior, which again is ameliorated when fluxes are aggregated at a 4-km level (i.e., test 01-s: DC median posterior MAPE 36%; Baltimore median posterior MAPE 31%). Therefore, the spatial inference of emissions at a resolution of 1–4 km does not constitute a robust solution and should be interpreted with caution. Singh et al. (2015) argue that the accuracy of source retrieval is limited by the imbalance between the number of observations and the degrees of freedom in the model space, the lack of priori information, measurement and model errors, and the limited and sparse nature of observations.

Regarding methodological assumptions and limitations, the primary weakness is the prior uncertainty estimation approach, which assigned a 50% prior flux value at the grid cell level for the baseline inversions based on available information about the uncertainty of the emission inventories considered here (Gurney et al., 2020a, 2020b). Since this estimation is intrinsically linked to the prior flux structure, our posterior estimates reflect significant spatial adjustments only in cells with high emissions, that is, a greater flexibility for the inversion over those pixels. This configuration makes the system vulnerable to missing large sources when they are located in pixels where little emissions are allocated in the prior. Furthermore, while tracer concentrations in winter are primarily influenced by meteorology, the flexibility of the inversion to adjust fluxes based on the magnitude of prior emissions means that our solution is limited by the accuracy of the prior in estimating winter anthropogenic emissions, which may lead to biases in the flux corrections. Future research to address this limitation includes setting a floor of prior uncertainty to give inversions more leeway to adjust anywhere (e.g., Lopez-Coto, Ren, et al., 2020); adjusting the uncertainty in known large sources and neighboring pixels (e.g., Yadav et al., 2021); and infer the prior uncertainties from the observations themselves (e.g., Hu et al., 2015). On the other hand, estimates of prior flux residuals can be improved by comparing prior fluxes with in situ measurements of real-world fluxes, such as power-plant data reported in the USA EPA Continuous Emission Monitoring System (https://campd.epa.gov/) or eddy covariance measurements (e.g., Wesloh et al., 2024), whenever possible. This will allow for a more realistic estimate of prior flux residuals, thereby improving our understanding of the spatial and temporal correlation structures of prior flux errors.

A limitation in the OSSE-based assessment of the system that must be acknowledged is that the same atmospheric transport and dispersion scheme is used in both synthetic data generation and inversions, leading to inadequate representation of the effects of systematic atmospheric transport model biases and potentially overestimating performance. Future OSSEs should adopt "bad transport" approaches, in which a different atmospheric transport model is used in synthetic data generation (e.g., Chevallier et al., 2010). Additionally, although the studies of Hegarty et al. (2013) and Selvaratnam et al. (2023) suggest that using different meteorological input data within the same Lagrangian Particle Dispersion Model is the primary way to reveal uncertainties in atmospheric transport (i.e., 1 km nudged WRF and 3 km HRRR), uncorrelated transport model errors are assumed and do not consider their potential coupling with prior errors. Future studies should explore HYSPLIT-based ensemble inversion frameworks driven by different meteorological models with different internal parameterizations (e.g., mixing processes) to address these limitations, by sampling uncertainties in both the prior and the transport representation, recognizing that the true magnitude and structure of these errors and their interactions are not precisely known.

Despite the above limitations, the evidence provided by this case study demonstrates that the urban-scale inverse modeling system can improve wintertime emission data in DCBA with CO_2 as a tracer, particularly in terms of the flux strength (mean flux) and total mass released across the urban cores of DC and Baltimore. These metrics are fundamental for assessing accidental urban/area-source releases in emergency response scenarios. Consequently, the main long-term implication of this study is that the modeling framework informs the design of subsequent system developments that potentially incorporate chemical transformations. This is especially true for the urban atmospheric transport framework (i.e., HYSPLIT configuration, meteorological nudging, and the spatial configuration of the observational network), which will enable estimation of urban- and neighborhood-scale releases of more reactive or short-lived substances associated with emergency response. Furthermore, while acknowledging that analytical solvers are typically employed for inverse linear problems (Berchet et al., 2021), the methodological approach used here to characterize model-data and prior uncertainty structure in this complex urban environment may prove useful for integration into more complex numerical solvers.

In the short term, we plan a summertime case study with the same inert tracer to adapt the inversion framework to capture summer physics and emissions. Based on this wintertime baseline, tailored configurations and parameterizations will be implemented—especially for boundary-layer representation, boundary conditions, and the quantification of biogenic fluxes—that will allow constraining the tracer fluxes over a full calendar year. Other upcoming initiatives include reducing uncertainties in nighttime and morning urban atmospheric transport modeling, as well as incorporating a second function to model intraday prior error temporal correlations. These approaches would allow for the assimilation of additional data beyond afternoon observations, as implemented here and in previous urban-scale studies (e.g., Breón et al., 2015; Karion, Ghosh, et al., 2023; Kunik et al., 2019; Lopez-Coto, Ren, et al., 2020; Wu et al., 2018; Yadav et al., 2021), and testing alternative prior error structures to evaluate their effects on the inverse modeling, including the accuracy of source retrieval.

4. Conclusions

This study developed and evaluated a high-resolution (1-km, 1-hr) urban-scale atmospheric inverse modeling system for the Washington, DC, and Baltimore, MD, metropolitan area (DCBA) using an inert tracer (CO_2) in January 2019 as a proof of concept. Observing System Simulation Experiments (OSSEs) and actual data inversions demonstrated that the system reliably quantifies urban surface fluxes. Flux strength, temporal variations, and cumulative emissions at the city scale are significantly improved. At the source (grid-cell) scale, OSSEs indicated that a high-resolution prior (1-km Vulcan) improved the estimation of flux strength and total emissions within the Baltimore, MD, and Washington, DC, beltways. In contrast, when the system was challenged to resolve the truth at finer scales, spatial regridding to a coarser resolution (e.g., 4 km) was necessary to improve these estimates. The system consistently detected lower emission estimates in DCBA than the bottom-up inventories (priors) for January 2019 (i.e., 0.366 MtC ± 0.021 MtC, 95% CI; and 0.351 MtC ± 0.017 MtC, 95% CI), aligning more closely with observed tracer enhancements and previous wintertime top-down estimations. Therefore, because inert trace gases ubiquitously emitted across cities are among the most challenging targets for top-down atmospheric inversions, this study concludes that the proposed framework can contribute to the development of future systems aimed at emergency-response preparedness.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Availability Statement

The data, models, and software used in this study are publicly available. The input data sets are as follows: The Vulcan v3.0 emission data product from Gurney et al. (2020b), utilized as the 1-km prior, is available at https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1810. The preview release of the GRA2PES emission data set for January 2019 used as the 4-km prior was provided by the NOAA Chemical Sciences Laboratory & National Institute of Standards and Technology (2023). It is listed on this website at the bottom in a section called beta version: https://csl.noaa.gov/groups/csl4/gra2pes/datasets/. The ACES emission data product version 2 from Gately and Hutyra (2022), which was utilized as the synthetic true fluxes in the OSSEs, can be accessed at https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1943. NIST's tower-based atmospheric CO_2 concentration measurements (Karion, DiGangi, et al., 2023) and VPRM model outputs (Gourdji & Marrs, 2021) for the DCBA were downloaded from the NIST Public Data Repository at https://data.nist.gov/od/id/mds2-3012 and https://data.nist.gov/od/id/mds2-2382, respectively. The most recent version of the CO_2 concentration data is now at https://doi.org/10.18434/mds2-3765. Regarding the modeling tools, the HYSPLIT model version 5.3.0 was used to calculate the footprints (NOAA Air Resources Laboratory, 2025). Model is available on the NOAA Air Resources Laboratory via its website: https://www.ready.noaa.gov/HYSPLIT.php. Inverse fluxes were derived using a modified version of the CarbonTracker-Lagrange (CT-L) regional inverse modeling framework version 0.2.0 developed by NOAA Global Monitoring Laboratory (2019). The original CT-L Python source code is available from the NOAA Global Monitoring Laboratory at https://gml.noaa.gov/aftp/products/carbontracker/lagrange/src/. The modeling code and raw data supporting the conclusions of this paper will be made available by the authors, without undue reservation. Figures in the manuscript were made with MATLAB version R2024a (The MathWorks Inc, 2024), available under the MATLAB license at https://www.mathworks.com/downloads.

Acknowledgments

Computational resources were provided by the NOAA Air Resources Laboratory. This study was supported by NOAA Grants NA24NESX432C0001 and NA19NES4320002 (Cooperative Institute for Satellite Earth System Studies—CISESS) at the University of Maryland/ESSIC. Support for NIST authors was provided by the NIST Greenhouse Gas Measurements Program. Certain commercial software and instruments are identified in this paper to foster understanding. Such identification does not imply recommendation or endorsement by NIST, nor does it imply that the software or instruments identified are necessarily the best available for the purpose. The scientific results and conclusions, as well as any views or opinions expressed herein, are those of the authors and do not necessarily reflect those of NOAA's Oceanic and Atmospheric Research (OAR) or the Department of Commerce. The authors gratefully acknowledge the anonymous reviewers for their comments, which contributed to improving the presentation and content of this study.

References

Bannister, R. N. (2008). A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characteristics and measurements of forecast error covariances. *Quarterly Journal of the Royal Meteorological Society*, *134*(637), 1951–1970. https://doi.org/10.1002/QJ.339

Beljaars, A. C. M., & Holtslag, A. A. M. (1991). Flux parameterization over land surfaces for atmospheric models. *Journal of Applied Meteorology and Climatology*, *30*(3), 327–341. https://doi.org/10.1175/1520-0450(1991)030<0327:FPOLSF>2.0.CO;2

Berchet, A., Sollum, E., Thompson, R. L., Pison, I., Thanwerdas, J., Broquet, G., et al. (2021). The community inversion framework v1.0: A unified system for atmospheric inversion studies. *Geoscientific Model Development*, *14*(8), 5331–5354. https://doi.org/10.5194/GMD-14-5331-2021

Betchov, R., & Yaglom, A. M. (1971). Comments on the theory of similarity as applied to turbulence in an unstably stratified fluid. *Atmospheric and Oceanic Physics*, *7*, 829–832.

Boylan, J. W., & Russell, A. G. (2006). PM and light extinction model performance metrics, goals, and criteria for three-dimensional air quality models. *Atmospheric Environment*, *40*(26), 4946–4959. https://doi.org/10.1016/j.atmosenv.2005.09.087

Breón, F. M., Broquet, G., Puygrenier, V., Chevallier, F., Xueref-Remy, I., Ramonet, M., et al. (2015). An attempt at estimating Paris area CO_2 emissions from atmospheric concentration measurements. *Atmospheric Chemistry and Physics*, *15*(4), 1707–1724. https://doi.org/10.5194/ACP-15-1707-2015

Brioude, J., Angevine, W. M., Ahmadov, R., Kim, S. W., Evan, S., McKeen, S. A., et al. (2013). Top-down estimate of surface flux in the Los Angeles Basin using a mesoscale inverse modeling technique: Assessing anthropogenic emissions of CO, NOx and CO_2 and their impacts. *Atmospheric Chemistry and Physics*, *13*(7), 3661–3677. https://doi.org/10.5194/ACP-13-3661-2013

Brioude, J., Kim, S. W., Angevine, W. M., Frost, G. J., Lee, S. H., McKeen, S. A., et al. (2011). Top-down estimate of anthropogenic emission inventories and their interannual variability in Houston using a mesoscale inverse modeling technique. *Journal of Geophysical Research*, *116*(D20), 20305. https://doi.org/10.1029/2011JD016215

Cahuich-López, M. A., Loughner, C. P., Cohen, M. D., Zinn, S., Ren, X., Luke, W., et al. (2024). Synthetic data experiments of a measurement-modeling greenhouse gas emissions estimation system for the Washington, DC and Baltimore, MD Metropolitan Area. In *104th Annual AMS Meeting 2024* (Vol. 104).436445.

Chevallier, F., Feng, L., Bösch, H., Palmer, P. I., & Rayner, P. J. (2010). On the impact of transport model errors for the estimation of CO_2 surface fluxes from GOSAT observations. *Geophysical Research Letters*, *37*(21). https://doi.org/10.1029/2010GL044652;REQUESTEDJOURNAL:JOURNAL:19448007;WGROUP:STRING:PUBLICATION

Deng, A., Lauvaux, T., Davis, K. J., Gaudet, B. J., Miles, N., Richardson, S. J., et al. (2017). Toward reduced transport errors in a high resolution urban CO_2 inversion system. *Elementa*, *5*. https://doi.org/10.1525/ELEMENTA.133/112393

Deng, F., Chen, J. M., Ishizawa, M., Yuen, C. W., Mo, G., Higuchi, K., et al. (2007). Global monthly CO_2 flux inversion with a focus over North America. *Tellus Series B Chemical and Physical Meteorology*, *59*(2), 179–190. https://doi.org/10.1111/J.1600-0889.2006.00235.X

Draxler, R. R., & Hess, G. D. (1997). Description of the HYSPLIT4 modeling system. In *NOAA Technical Memorandum. ERL ARL-224*. Retrieved from https://repository.library.noaa.gov/view/noaa/31133

Draxler, R. R., & Hess, G. D. (1998). An overview of the HYSPLIT_4 modelling system for trajectories. *Australian Meteorological Magazine*, *47*(4), 295–308. https://doi.org/10.1071/es98032

Emery, C., & Tai, E. (2001). Enhanced meteorological modeling and performance evaluation for two Texas ozone episodes. Retrieved from https://api.semanticscholar.org/CorpusID:127579774

Engelen, R. J., Denning, A. S., & Gurney, K. R. (2002). On error estimation in atmospheric CO_2 inversions. *Journal of Geophysical Research*, *107*(D22), ACL10-1–ACL10-13. https://doi.org/10.1029/2002JD002195

Feng, X., Zhang, X., & Wang, J. (2023). Update of SO_2 emission inventory in the Megacity of Chongqing, China by inverse modeling. *Atmospheric Environment*, *294*, 119519. https://doi.org/10.1016/J.ATMOSENV.2022.119519

Fong, W. K., Sotos, M., Schultz, S., Deng-Beck, C., Marques, A., & Doust, M. (2014). Global protocol for community-scale greenhouse gas emission inventories (GPC): An accounting and reporting standard for cities. Retrieved from https://ghgprotocol.org/sites/default/files/ghgp/standards/GHGP_GPC_0.pdf

Gately, C., & Hutyra, L. R. (2022). Anthropogenic carbon emission system, 2012-2017 (Version 2) [Dataset]. *ORNL DAAC*. https://doi.org/10.3334/ORNLDAAC/1943

Ghosh, S., Mueller, K., Prasad, K., & Whetstone, J. (2021). Accounting for transport error in inversions: An urban synthetic data experiment. *Earth and Space Science*, *8*(7), e2020EA001272. https://doi.org/10.1029/2020EA001272

Gourdji, S., & Marrs, J. (2021). Biospheric CO_2 surface flux estimates from the Vegetation Photosynthesis & Respiration Model (VPRM) in the eastern USA and Canada: November 2016 to December 2021 (Version 1.3) [Dataset]. *National Institute of Standards and Technology*. https://doi.org/10.18434/mds2-2382

Gourdji, S. M., Karion, A., Lopez-Coto, I., Ghosh, S., Mueller, K. L., Zhou, Y., et al. (2022). A modified vegetation photosynthesis and respiration model (VPRM) for the Eastern USA and Canada, evaluated with comparison to atmospheric observations and other biospheric models. *Journal of Geophysical Research: Biogeosciences*, *127*(1), e2021JG006290. https://doi.org/10.1029/2021JG006290

Gourdji, S. M., Yadav, V., Karion, A., Mueller, K. L., Conley, S., Ryerson, T., et al. (2018). Reducing errors in aircraft atmospheric inversion estimates of point-source emissions: The Aliso Canyon natural gas leak as a natural tracer experiment. *Environmental Research Letters*, *13*(4), 045003. https://doi.org/10.1088/1748-9326/AAB049

Gurney, K. R., Liang, J., Patarasuk, R., Song, Y., Huang, J., & Roest, G. (2020a). The Vulcan version 3.0 high-resolution fossil fuel CO_2 emissions for the United States. *Journal of Geophysical Research: Atmospheres*, *125*(19), e2020JD032974. https://doi.org/10.1029/2020JD032974

Gurney, K. R., Liang, J., Patarasuk, R., Song, Y., Huang, J., & Roest, G. (2020b). Vulcan: High-resolution hourly fossil fuel CO_2 emissions in USA, 2010-2015 (Version 3) [Dataset]. *ORNL DAAC*. https://doi.org/10.3334/ORNLDAAC/1810

Gurney, K. R., Liang, J., Roest, G., Song, Y., Mueller, K., & Lauvaux, T. (2021). Under-reporting of greenhouse gas emissions in U.S. cities. *Nature Communications*, *12*(1), 1–7. https://doi.org/10.1038/s41467-020-20871-0

Hajny, K. D., Floerchinger, C. R., Lopez-Coto, I., Pitt, J. R., Gately, C. K., Gurney, K. R., et al. (2022). A spatially explicit inventory scaling approach to estimate urban CO_2 emissions. *Elementa: Science of the Anthropocene*, *10*(1), 00121. https://doi.org/10.1525/elementa.2021.00121

Hegarty, J., Draxler, R. R., Stein, A. F., Brioude, J., Mountain, M., Eluszkiewicz, J., et al. (2013). Evaluation of Lagrangian particle dispersion models with measurements from controlled tracer releases. *Journal of Applied Meteorology and Climatology*, *52*(12), 2623–2637. https://doi.org/10.1175/JAMC-D-13-0125.1

Hu, L., Andrews, A. E., Montzka, S. A., Miller, S. M., Bruhwiler, L., Oh, Y., et al. (2025). An unexpected seasonal cycle in U.S. oil and gas methane emissions. *Environmental Science and Technology*, *59*(20), 9968–9979. https://doi.org/10.1021/acs.est.4c14090

Hu, L., Andrews, A. E., Thoning, K. W., Sweeney, C., Miller, J. B., Michalak, A. M., et al. (2019). Enhanced North American carbon uptake associated with El Niño. *Science Advances*, *5*(6), eaaw0076. https://doi.org/10.1126/sciadv.aaw0076

Hu, L., Montzka, S. A., Miller, J. B., Andrews, A. E., Lehman, S. J., Miller, B. R., et al. (2015). U.S. emissions of HFC-134a derived for 2008–2012 from an extensive flask-air sampling network. *Journal of Geophysical Research: Atmospheres*, *120*(2), 801–825. https://doi.org/10.1002/2014JD022617

Huang, Y., Kort, E. A., Gourdji, S., Karion, A., Mueller, K., & Ware, J. (2019). Seasonally resolved excess urban methane emissions from the Baltimore/Washington, DC Metropolitan Region. *Environmental Science and Technology*, *53*(19), 11285–11293. https://doi.org/10.1021/acs.est.9b02782

Itahashi, S., Yumimoto, K., Kurokawa, J. I., Morino, Y., Nagashima, T., Miyazaki, K., et al. (2019). Inverse estimation of NOx emissions over China and India 2005–2016: Contrasting recent trends and future perspectives. *Environmental Research Letters*, *14*(12), 124020. https://doi.org/10.1088/1748-9326/AB4D7F

Karion, A., Callahan, W., Stock, M., Prinzivalli, S. R., Verhulst, K., Kim, J., et al. (2020). Greenhouse gas observations from the Northeast Corridor tower network. *Earth System Science Data*, *12*(1), 699–717. https://doi.org/10.5194/ESSD-12-699-2020

Karion, A., DiGangi, E., Prinzivalli, S., Draper, C., Baldelli, S., Fain, C., et al. (2023). Observations of carbon dioxide (CO_2), methane (CH_4), and carbon monoxide (CO) mole fractions from the NIST Northeast Corridor urban testbed (Version 1.1) [Dataset]. *National Institute of Standards and Technology*. https://doi.org/10.18434/mds2-3012

Karion, A., Ghosh, S., Lopez-Coto, I., Mueller, K., Gourdji, S., Pitt, J., & Whetstone, J. (2023). Methane emissions show recent decline but strong seasonality in two US Northeastern cities. *Environmental Science and Technology*, *57*(48), 19565–19574. https://doi.org/10.1021/acs.est.3c05050

Karion, A., Lopez-Coto, I., Gourdji, S. M., Mueller, K., Ghosh, S., Callahan, W., et al. (2021). Background conditions for an urban greenhouse gas network in the Washington, DC, and Baltimore metropolitan region. *Atmospheric Chemistry and Physics*, *21*(8), 6257–6273. https://doi.org/10.5194/ACP-21-6257-2021

Kort, E. A., Angevine, W. M., Duren, R., & Miller, C. E. (2013). Surface observations for monitoring urban fossil fuel CO_2 emissions: Minimum site location requirements for the Los Angeles megacity. *Journal of Geophysical Research: Atmospheres*, *118*(3), 1577–1584. https://doi.org/10.1002/JGRD.50135

Kountouris, P., Gerbig, C., Rödenbeck, C., Karstens, U., Frank Koch, T., & Heimann, M. (2018). Technical Note: Atmospheric CO_2 inversions on the mesoscale using data-driven prior uncertainties: Methodology and system evaluation. *Atmospheric Chemistry and Physics*, *18*(4), 3027–3045. https://doi.org/10.5194/ACP-18-3027-2018

Kountouris, P., Gerbig, C., Totsche, K. U., Dolman, A. J., A. Meesters, A. G. C., Broquet, G., et al. (2015). An objective prior error quantification for regional atmospheric inverse applications. *Biogeosciences*, *12*(24), 7403–7421. https://doi.org/10.5194/BG-12-7403-2015

Kunik, L., Mallia, D. V., Gurney, K. R., Mendoza, D. L., Oda, T., & Lin, J. C. (2019). Bayesian inverse estimation of urban CO_2 emissions: Results from a synthetic data simulation over Salt Lake City, UT. *Elementa: Science of the Anthropocene*, *7*, 36. https://doi.org/10.1525/elementa.375

Lauvaux, T., & Davis, K. J. (2014). Planetary boundary layer errors in mesoscale inversions of column-integrated CO_2 measurements. *Journal of Geophysical Research: Atmospheres*, *119*(2), 490–508. https://doi.org/10.1002/2013JD020175

Lauvaux, T., Miles, N. L., Deng, A., Richardson, S. J., Cambaliza, M. O., Davis, K. J., et al. (2016). High-resolution atmospheric inversion of urban CO_2 emissions during the dormant season of the Indianapolis flux experiment (INFLUX). *Journal of Geophysical Research: Atmospheres*, *121*(10), 5213–5236. https://doi.org/10.1002/2015JD024473

Law, R. M., Peters, W., Rödenbeck, C., Aulagnier, C., Baker, I., Bergmann, D. J., et al. (2008). TransCom model simulations of hourly atmospheric CO_2: Experimental overview and diurnal cycle results for 2002. *Global Biogeochemical Cycles*, *22*(3), 17. https://doi.org/10.1029/2007gb003050

Lian, J., Lauvaux, T., Utard, H., Bréon, F. M., Broquet, G., Ramonet, M., et al. (2022). Assessing the effectiveness of an urban CO_2 monitoring network over the Paris Region through the COVID-19 lockdown natural experiment. *Environmental Science & Technology*, *56*(4), 2153–2162. https://doi.org/10.1021/ACS.EST.1C04973

Lichiheb, N., Ngan, F., & Cohen, M. (2024). Improving the atmospheric dispersion forecasts over Washington, D.C. using UrbanNet observations: A study with HYSPLIT model. *Urban Climate*, *55*, 101948. https://doi.org/10.1016/J.UCLIM.2024.101948

Lin, J. C., Gerbig, C., Wofsy, S. C., Andrews, A. E., Daube, B. C., Davis, K. J., & Grainger, C. A. (2003). A near-field tool for simulating the upstream influence of atmospheric observations: The Stochastic Time-Inverted Lagrangian Transport (STILT) model. *Journal of Geophysical Research*, *108*(16), 4493. https://doi.org/10.1029/2002JD003161

Lopez-Coto, I., Ghosh, S., Prasad, K., & Whetstone, J. (2017). Tower-based greenhouse gas measurement network design—The National Institute of Standards and Technology North East Corridor Testbed. *Advances in Atmospheric Sciences*, *34*(9), 1095–1105. https://doi.org/10.1007/s00376-017-6094-6

Lopez-Coto, I., Hicks, M., Karion, A., Sakai, R. K., Demoz, B., Prasad, K., et al. (2020). Assessment of planetary boundary layer parameterizations and urban heat Island comparison: Impacts and implications for tracer transport. *Journal of Applied Meteorology and Climatology*, *59*(10), 1637–1653. https://doi.org/10.1175/JAMC-D-19-0168.1

Lopez-Coto, I., Ren, X., Karion, A., McKain, K., Sweeney, C., Dickerson, R. R., et al. (2022). Carbon monoxide emissions from the Washington, DC, and Baltimore metropolitan area: Recent trend and COVID-19 Anomaly. *Environmental Science and Technology*, *56*(4), 2172–2180. https://doi.org/10.1021/acs.est.1c06288

Lopez-Coto, I., Ren, X., Salmon, O. E., Karion, A., Shepson, P. B., Dickerson, R. R., et al. (2020). Wintertime CO_2, CH_4, and CO emissions estimation for the Washington, DC-Baltimore metropolitan area using an inverse modeling technique. *Environmental Science and Technology*, *54*(5), 2606–2614. https://doi.org/10.1021/acs.est.9b06619

Loughner, C. P., Fasoli, B., Stein, A. F., & Lin, J. C. (2021). Incorporating features from the stochastic time-inverted Lagrangian transport (STILT) model into the hybrid single-particle Lagrangian integrated trajectory (HYSPLIT) model: A unified dispersion model for time-forward and time-reversed applications. *Journal of Applied Meteorology and Climatology*, *60*(6), 799–810. https://doi.org/10.1175/JAMC-D-20-0158.1

Lyu, C., Harkins, C., Li, M., Mueller, K., Prothero, J., Verreyken, B., et al. (2025). Development of the United States GReenhouse gas and Air Pollutants Emissions System (GRA2PES). *Journal of Geophysical Research: Atmospheres*, *130*(20), e2025JD043597. https://doi.org/10.1029/2025JD043597

Mallia, D. V., Mitchell, L. E., Kunik, L., Fasoli, B., Bares, R., Gurney, K. R., et al. (2020). Constraining urban CO_2 Emissions using mobile observations from a light rail public transit platform. *Environmental Science and Technology*, *54*(24), 15613–15621. https://doi.org/10.1021/acs.est.0c04388

Martin, C. R., Zeng, N., Karion, A., Mueller, K., Ghosh, S., Lopez-Coto, I., et al. (2019). Investigating sources of variability and error in simulations of carbon dioxide in an urban region. *Atmospheric Environment*, *199*, 55–69. https://doi.org/10.1016/J.ATMOSENV.2018.11.013

McDonald, B. C., McBride, Z. C., Martin, E. W., & Harley, R. A. (2014). High-resolution mapping of motor vehicle carbon dioxide emissions. *Journal of Geophysical Research: Atmospheres*, *119*(9), 5283–5298. https://doi.org/10.1002/2013JD021219

Michalak, A. M., Randazzo, N. A., & Chevallier, F. (2017). Diagnostic methods for atmospheric inversions of long-lived greenhouse gases. *Atmospheric Chemistry and Physics*, *17*(12), 7405–7421. https://doi.org/10.5194/ACP-17-7405-2017

Morley, S. K., Brito, T. V., & Welling, D. T. (2018). Measures of model performance based on the log accuracy ratio. *Space Weather*, *16*(1), 69–88. https://doi.org/10.1002/2017SW001669

Munassar, S., Monteil, G., Scholze, M., Karstens, U., Rödenbeck, C., Koch, F. T., et al. (2023). Why do inverse models disagree? A case study with two European CO_2 inversions. *Atmospheric Chemistry and Physics*, *23*(4), 2813–2828. https://doi.org/10.5194/ACP-23-2813-2023

Nalini, K., Lauvaux, T., Abdallah, C., Lian, J., Ciais, P., Utard, H., et al. (2022). High-resolution Lagrangian inverse modeling of CO_2 emissions over the Paris Region during the first 2020 lockdown period. *Journal of Geophysical Research: Atmospheres*, *127*(14), e2021JD036032. https://doi.org/10.1029/2021JD036032

National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce. (2015). *NCEP GDAS/FNL 0.25 degree global tropospheric analyses and forecast grids*. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. UCAR/NCAR—Research Data Archive. https://doi.org/10.5065/D65Q4T4Z

Nesser, H., Jacob, D. J., Maasakkers, J. D., Lorente, A., Chen, Z., Lu, X., et al. (2024). High-resolution US methane emissions inferred from an inversion of 2019 TROPOMI satellite data: Contributions from individual states, urban areas, and landfills. *Atmospheric Chemistry and Physics*, *24*(8), 5069–5091. https://doi.org/10.5194/acp-24-5069-2024

Ngan, F., Loughner, C. P., & Stein, A. (2019). The evaluation of mixing methods in HYSPLIT using measurements from controlled tracer experiments. *Atmospheric Environment*, *219*, 117043. https://doi.org/10.1016/J.ATMOSENV.2019.117043

NOAA Global Monitoring Laboratory. (2019). CarbonTracker Lagrange (v0.2.0) [Software]. *NOAA Global Monitoring Laboratory*. Retrieved from https://gml.noaa.gov/ccgg/carbontracker-lagrange/doc/index.html

NOAA Air Resources Laboratory. (2025). Hybrid Single Particle Lagrangian Integrated Trajectory Model (Version 5.3.0) [Software]. *NOAA Air Resources Laboratory*. Retrieved from https://www.ready.noaa.gov/HYSPLIT.php

NOAA Chemical Sciences Laboratory & National Institute of Standards and Technology. (2023). GRA2PES emission dataset (January 2019 Beta Version) [Dataset]. *NOAA CSL*. Retrieved from https://csl.noaa.gov/groups/csl4/gra2pes/datasets/

Nüß, J. R., Daskalakis, N., Piwowarczyk, F. G., Gkouvousis, A., Schneising, O., Buchwitz, M., et al. (2025). Top-down CO emission estimates using TROPOMI CO data in the TM5-4DVAR (r1258) inverse modeling suit. *Geoscientific Model Development*, *18*(10), 2861–2890. https://doi.org/10.5194/GMD-18-2861-2025

Pitt, J. R., Lopez-Coto, I., Hajny, K. D., Tomlin, J., Kaeser, R., Jayarathne, T., et al. (2022). New York City greenhouse gas emissions estimated with inverse modeling of aircraft measurements. *Elementa: Science of the Anthropocene*, *10*(1), 00082. https://doi.org/10.1525/elementa.2021.00082

Powers, J. G., Klemp, J. B., Skamarock, W. C., Davis, C. A., Dudhia, J., Gill, D. O., et al. (2017). The weather research and forecasting model: Overview, system efforts, and future directions. *Bulletin of the American Meteorological Society*, *98*(8), 1717–1737. https://doi.org/10.1175/BAMS-D-15-00308.1

Qu, Z., Henze, D. K., Worden, H. M., Jiang, Z., Gaubert, B., Theys, N., & Wang, W. (2022). Sector-Based top-down estimates of NOx, SO_2, and CO emissions in east Asia. *Geophysical Research Letters*, *49*(2), e2021GL096009. https://doi.org/10.1029/2021GL096009

Rappolt, T. J., & Kerrin, S. L. (2010). Measurement of atmospheric dispersion using gaseous tracers. In P. Zannetti (Ed.), *Air Quality Modeling—Theories, methodologies, computational techniques, and available databases and software, Vol. IV—Advances and Updates (25)*. The EnviroComp Institute and The Air & Waste Management Association. Retrieved from http://www.envirocomp.org/

Reen, B. P. (2016). A brief guide to observational nudging in WRF.

Remaud, M., Chevallier, F., Maignan, F., Belviso, S., Berchet, A., Parouffe, A., et al. (2022). Plant gross primary production, plant respiration and carbonyl sulfide emissions over the globe inferred by atmospheric inverse modelling. *Atmospheric Chemistry and Physics*, *22*(4), 2525–2552. https://doi.org/10.5194/ACP-22-2525-2022

Selvaratnam, V., Thomson, D. J., Webster, H. N., Selvaratnam, V., Thomson, D. J., & Webster, H. N. (2023). Validation of the atmospheric dispersion model NAME against long-range tracer release experiments. *Journal of Applied Meteorology and Climatology*, *62*(9), 1165–1174. https://doi.org/10.1175/JAMC-D-23-0021.1

Semerjian, H. G., Wong, T. M., & Whetstone, J. R. (2024). *NIST greenhouse gas measurements program: A decade of critical accomplishments* Technical Note (NIST TN). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.TN.2291

Singh, S. K., Sharan, M., & Issartel, J. P. (2015). Inverse modelling methods for identifying unknown releases in emergency scenarios: An overview. *International Journal of Environment and Pollution*, *57*(1–2), 68–91. https://doi.org/10.1504/IJEP.2015.072121

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Liu, Z., Berner, J., et al. (2021). *A description of the advanced research WRF model version 4* (Vol. 145). National Center for Atmospheric Research. Number 145. https://doi.org/10.5065/1dfh-6p97

Stein, A. F., Draxler, R. R., Rolph, G. D., Stunder, B. J. B., Cohen, M. D., & Ngan, F. (2015). NOAA's HYSPLIT atmospheric transport and dispersion modeling system. *Bulletin of the American Meteorological Society*, *96*(12), 2059–2077. https://doi.org/10.1175/BAMS-D-14-00110.1

Super, I., Dellaert, S. N. C., Tokaya, J. P., & Schaap, M. (2021). The impact of temporal variability in prior emissions on the optimization of urban anthropogenic emissions of CO_2, CH_4 and CO using in-situ observations. *Atmospheric Environment X*, *11*, 100119. https://doi.org/10.1016/J.AEAOA.2021.100119

Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*. Society for Industrial and Applied Mathematics.

The MathWorks Inc. (2024). MATLAB (version R2024a) [Software]. *The MathWorks, Inc*. Retrieved from https://www.mathworks.com/downloads

Thompson, R. L., Gerbig, C., & Rödenbeck, C. (2011). A Bayesian inversion estimate of N_2O emissions for western and central Europe and the assessment of aggregation errors. *Atmospheric Chemistry and Physics*, *11*(7), 3443–3458. https://doi.org/10.5194/ACP-11-3443-2011

U.S. Geological Survey. (2024). *Annual NLCD collection 1 science products* (Vol. 2024). U.S. Geological Survey data release. https://doi.org/10.5066/P94UXNTS

Vardag, S. N., & Maiwald, R. (2024). Optimising urban measurement networks for CO_2 flux estimation: A high-resolution observing system simulation experiment using GRAMM/GRAL. *Geoscientific Model Development*, *17*(4), 1885–1902. https://doi.org/10.5194/GMD-17-1885-2024

Wesloh, D., Keller, K., Feng, S., Lauvaux, T., & Davis, K. J. (2024). Temporal error correlations in a terrestrial carbon cycle model derived by comparison to carbon dioxide eddy covariance flux tower measurements. *Journal of Geophysical Research: Biogeosciences*, *129*(1), e2023JG007526. https://doi.org/10.1029/2023JG007526;CTYPE:STRING:JOURNAL

Winbourne, J. B., Smith, I. A., Stoynova, H., Kohler, C., Gately, C. K., Logan, B. A., et al. (2022). Quantification of urban Forest and Grassland carbon fluxes using field measurements and a satellite-based model in Washington DC/Baltimore area. *Journal of Geophysical Research: Biogeosciences*, *127*(1), e2021JG006568. https://doi.org/10.1029/2021JG006568

Wu, K., Lauvaux, T., Davis, K. J., Deng, A., Coto, I. L., Gurney, K. R., & Patarasuk, R. (2018). Joint inverse estimation of fossil fuel and biogenic CO_2 fluxes in an urban environment: An observing system simulation experiment to assess the impact of multiple uncertainties. *Elementa: Science of the Anthropocene*, *6*, 17. https://doi.org/10.1525/elementa.138

Yadav, V., Duren, R., Mueller, K., Verhulst, K. R., Nehrkorn, T., Kim, J., et al. (2019). Spatio-temporally resolved methane fluxes from the Los Angeles megacity. *Journal of Geophysical Research: Atmospheres*, *124*(9), 5131–5148. https://doi.org/10.1029/2018JD030062

Yadav, V., Ghosh, S., Mueller, K., Karion, A., Roest, G., Gourdji, S. M., et al. (2021). The impact of COVID-19 on CO_2 emissions in the Los Angeles and Washington DC/Baltimore metropolitan areas. *Geophysical Research Letters*, *48*(11), e2021GL092744. https://doi.org/10.1029/2021GL092744

Yadav, V., & Michalak, A. M. (2013). Improving computational efficiency in large linear inverse problems: An example from carbon dioxide flux estimation. *Geoscientific Model Development*, *6*(3), 583–590. https://doi.org/10.5194/GMD-6-583-2013

Yumimoto, K., Uno, I., & Itahashi, S. (2014). Long-term inverse modeling of Chinese CO emission from satellite observations. *Environmental Pollution*, *195*, 308–318. https://doi.org/10.1016/J.ENVPOL.2014.07.026

Zannetti, P. (2013). *Air pollution modeling: Theories, computational methods and available software*. Springer Science & Business Media.

References From the Supporting Information

Andres, R. J., Boden, T. A., & Higdon, D. M. (2016). Gridded uncertainty in fossil fuel carbon dioxide emission maps, a CDIAC example. *Atmospheric Chemistry and Physics*, *16*(23), 14979–14995. https://doi.org/10.5194/ACP-16-14979-2016

Basu, S., Lehman, S. J., Miller, J. B., Andrews, A. E., Sweeney, C., Gurney, K. R., et al. (2020). Estimating US fossil fuel CO_2 emissions from measurements of 14C in atmospheric CO_2. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(24), 13300–13307. https://doi.org/10.1073/pnas.1919032117

Cahuich-López, M. A., Mariño-Tapia, I., Souza, A. J., Gold-Bouchot, G., Cohen, M., & Lozano, D. V. (2020). Spatial and temporal variability of sea breezes and synoptic influences over the surface wind field of the Yucatán Peninsula. *Atmósfera*, *33*(2), 123–142. https://doi.org/10.20937/ATM.52713

Dewitz, J. (2020). *National Land Cover Database (NLCD) 2016 Products (ver. 3.0, November 2023)*. U.S. Geological Survey data release. https://doi.org/10.5066/P96HHBIE

Dowell, D. C., Alexander, C. R., James, E. P., Weygandt, S. S., Benjamin, S. G., Manikin, G. S., et al. (2022). The high-resolution rapid refresh (HRRR): An hourly updating convection-allowing forecast model. Part I: Motivation and system description. *Weather and Forecasting*, *37*(8), 1371–1395. https://doi.org/10.1175/WAF-D-21-0151.1

Harkins, C., McDonald, B. C., Henze, D. K., & Wiedinmyer, C. (2021). A fuel-based method for updating mobile source emissions during the COVID-19 pandemic. *Environmental Research Letters*, *16*(6), 065018. https://doi.org/10.1088/1748-9326/AC0660

Ide, K., Courtier, P., Ghil, M., & Lorenc, A. C. (1997). Unified notation for data assimilation: Operational, sequential and variational (gtSpecial IssueltData Assimilation in Aeteology and oceanography: Theory and Practice). *Journal of the Meteorological Society of Japan. Series II*, *75*(1B), 181–189. https://doi.org/10.2151/JMSJ1965.75.1B_181

Lyu, C., Harkins, C., Li, M., McDonald, B., Prothero, J., & Mueller, K. (2024). *The U.S. greenhouse gas and air pollutant emissions system (GRA2PES)*. National Institute of Standards and Technology. https://doi.org/10.18434/mds2-3520

McDonald, B. C., McKeen, S. A., Cui, Y. Y., Ahmadov, R., Kim, S. W., Frost, G. J., et al. (2018). Modeling ozone in the Eastern U.S. using a fuel-based mobile source emissions inventory. *Environmental Science and Technology*, *52*(13), 7360–7370. https://doi.org/10.1021/acs.est.8b00778

Mueller, K., Yadav, V., Lopez-Coto, I., Karion, A., Gourdji, S., Martin, C., & Whetstone, J. (2018). Siting background towers to characterize incoming air for urban greenhouse gas estimation: A case study in the Washington, DC/Baltimore area. *Journal of Geophysical Research: Atmospheres*, *123*(5), 2910–2926. https://doi.org/10.1002/2017JD027364

NOAA Chemical Sciences Laboratory & National Institute of Standards and Technology. (2023). GReenhouse gas and Air Pollutants Emissions System (GRA2PES). Retrieved from https://csl.noaa.gov/groups/csl4/gra2pes/whitepaper.pdf