New York City greenhouse gas emissions estimated with inverse modelling of aircraft measurements

Joseph R. Pitt[*,1,§], Israel Lopez-Coto[2,1], Kristian D. Hajny[1,3], Jay Tomlin[3], Robert Kaeser[3], Thilina Jayarathne[3,†], Brian H. Stirm[4], Cody R. Floerchinger[5], Christopher P. Loughner[6], Conor K. Gately[5,7,‡], Lucy R. Hutyra[7], Kevin R. Gurney[8], Geoffrey S. Roest[8], Jianming Liang[9], Sharon Gourdji[2], Anna Karion[2], James R. Whetstone[2], Paul B. Shepson*[*,1,3]

[1]School of Marine and Atmospheric Sciences, Stony Brook University, Stony Brook, NY 11794, USA

[2]National Institute of Standards and Technology, Gaithersburg, MD 20899, USA

[3]Department of Chemistry, Purdue University, West Lafayette, IN 47907, USA

[4]School of Aviation and Transportation Technology, Purdue University, West Lafayette, IN 47906, USA

[5]Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA 02138, USA

[6]Air Resources Laboratory, NOAA, College Park, MD 20740, USA

[7]Department of Earth and Environment, Boston University, Boston, MA 02215, USA

[8]School of Informatics, Computing and Cyber Systems, Northern Arizona University, Flagstaff, AZ 86011, USA

[9]Environmental Systems Research Institute, Redlands, CA 92373, USA

[§]Current address: School of Chemistry, University of Bristol, Bristol, BS8 1TS, UK

[†]Current address: Bristol Myers Squibb, New Brunswick, NJ 08901, USA

[‡]Current address: Metropolitan Area Planning Council, Boston, MA 02111, USA

*Corresponding authors: joseph.pitt@bristol.ac.uk; paul.shepson@stonybrook.edu

Abstract

Cities are greenhouse gas emission hotspots, making them targets for emission reduction policies. Effective emission reduction policies must be supported by accurate and transparent emissions accounting. Top-down approaches to emissions estimation, based on atmospheric greenhouse gas measurements, are an important and complementary tool to assess, improve and update the emission inventories on which policy decisions are based and assessed.

In this study we present results from nine research flights measuring CO_2 and CH_4 around New York City during the non-growing seasons of 2018 to 2020. We used an ensemble of dispersion model runs in a Bayesian inverse modelling framework to derive campaign-average posterior emission estimates for the New York-Newark, N.J. urban area of (125 ± 39) kmol CO_2 s^{-1} and (0.62 ± 0.19) kmol CH_4 s^{-1} (reported as mean ± 1σ variability across the nine flights). We also derived emission estimates of (45 ± 18) kmol CO_2 s^{-1} and (0.20 ± 0.07) kmol CH_4 s^{-1} for the five boroughs of New York City. These emission rates, among the first top-down estimates for New York City, are consistent with inventory estimates for CO_2 but are 2.4 times larger than the gridded EPA CH_4 inventory, consistent with previous work suggesting CH_4 emissions from cities throughout the northeast U.S. are currently underestimated.

Keywords: urban emissions; greenhouse gas emissions; methane; carbon dioxide; Bayesian inverse modelling; New York City

Introduction

Cities are large sources of greenhouse gas emissions, with an estimated 36.8% of total global emissions coming from within urban extents in the year 2000 (Marcotullio et al., 2013). In North America this urban share of greenhouse gas emissions was estimated to be even larger (49.2%; Marcotullio et al., 2013).

To reduce emissions and mitigate climate change, many cities are implementing new policy practices and regulations (Trencher et al., 2016; Sethi et al., 2020). New York City has legally binding emission reduction targets of 40% by 2030 and 80% by 2050, relative to 2005 levels (New York City Mayor's Office of Sustainability, 2016). As part of the legislative effort to meet these targets, the city council passed the Climate Mobilization Act, which introduces emission limits on large and medium-sized buildings under Local Law 97 (Climate Mobilization Act, 2019).

Progress towards emission targets and the efficacy of emission reduction policies can be assessed through the compilation of city-specific inventories, often referred to as self-reported inventories. These inventories estimate greenhouse gas emissions using a bottom-up approach, where the total emission from each source category is calculated by multiplying activity data (e.g., fuel sales) by an emission factor (e.g., greenhouse gas emissions per unit of fuel sold). Self-reported inventories typically provide emission estimates at the whole-city level, broken down by emission source. These estimates can include direct emissions within the city limits (scope 1 emissions), emissions associated with electricity used within the city but generated elsewhere (scope 2 emissions) and other indirect emissions that occur elsewhere (e.g., landfill CH_4) as a result of activity (e.g., waste production) within the city (scope 3 emissions; Nangini et al., 2019). The New York City emissions inventory (MacWhinney and Barnett, 2019) is compiled using a mixed-scope approach known as the city-induced framework (Fong et al., 2014), which is designed to account for

emissions attributable to activities within the city limits, regardless of emission location. This mixed-scope approach includes some emissions from each of scope 1, 2 and 3, but does not include all the emissions from any single scope. Quantifying indirect emissions is important from a policy perspective, but it is challenging to assess mixed-scope inventories with atmospheric measurements, which are only sensitive to scope 1 emissions.

Spatially-resolved inventories of scope 1 emissions have been compiled at the national level, either using bottom-up approaches (Gately and Hutyra, 2017; Gurney et al., 2020) or using proxy data to spatially disaggregate national emission totals (Maasakkers et al., 2016; Oda et al., 2018; Janssens-Maenhout et al., 2019). These can be used to assess and improve the self-reported inventories or adopted as the primary method by which cities tailor and assess their emission reduction policies (Hutyra et al., 2014; Gurney and Shepson, 2021). A recent study by Gurney et al. (2021) compared scope 1 emissions from the national Vulcan inventory with self-reported inventories from 48 U.S. cities and found that most cities underreported emissions. The spatial information in gridded national inventories allows them to be combined with atmospheric transport models to estimate atmospheric mole fractions, allowing direct comparison to measurements and thus facilitating top-down approaches to estimating emissions.

Many top-down studies have used tower-based measurements of greenhouse gas mole fractions in an inverse modelling approach to estimate urban emissions (Bréon et al., 2015; Lamb et al., 2016; Lauvaux et al., 2016; 2020; Staufer et al., 2016; Deng et al., 2017; Huang et al., 2019; Nickless et al., 2019; Yadav et al., 2019). Tower-based sites allow for continuous year-round observations, enabling top-down fluxes to be estimated at annual timescales. On the other hand, an individual tower has limited spatial sensitivity, so multiple towers may be required to provide the necessary spatial coverage to accurately estimate whole-city emissions. Aircraft-based

measurements cannot provide the continuous temporal coverage of tower observations. However, by sampling a wide area in a short period of time, well-tailored flight tracks enable aircraft to provide useful snapshot estimates of whole-city emissions, with the ready ability to define and change the area of study.

Aircraft measurements can be used in an inverse modelling framework to estimate urban emissions. This approach incorporates information regarding the prior fluxes, transport model and measurements, including the error covariance structure for each of these elements. Several previous studies have applied such a framework to estimate urban fluxes using aircraft measurements (Brioude et al., 2011; 2013; Pisso et al., 2019; Lopez-Coto et al., 2020). However, this approach has yet to be applied to estimate New York City greenhouse gas emissions. In this study we use measurements from nine research flights (during the non-growing season) and an ensemble of dispersion model runs in an inverse modelling approach to estimate carbon dioxide and methane emission rates for the New York-Newark, N.J. urban area and the five boroughs of New York City.

Methods

Aircraft measurements

Measurements were made on board the Purdue University Airborne Laboratory for Atmospheric Research (ALAR), a modified Beechcraft Duchess. Full details of the aircraft payload and sampling configuration are provided by Cambaliza et al. (2014). Carbon dioxide (CO_2) and methane (CH_4) mole fractions were measured using a Picarro Cavity Ringdown Spectrometer (Crosson, 2008), calibrated in flight using three calibration cylinders traceable to the National Oceanic and Atmospheric Administration (NOAA) reference scales for CO_2 (X2007; Tans et al.,

2017) and CH_4 (X2004A; Dlugokencky et al., 2005). The cylinders were prepared, filled and certified by NOAA; details are provided in SI Table S1.1.

Three Picarro analysers were used over the course of the campaign: two different G2301-f analysers and one G2401-m analyser. The two G2301-f analysers were operated in low-flow mode, with each species measured at approximately 1.2 s intervals on one analyser and 1.4 s intervals on the other. The G2401-m analyser took a measurement of each species at approximately 2.3 s intervals. These analysers exhibited typical precisions of 0.2 ppm (μmol mol^{-1}) for CO_2 and 3 ppb (nmol mol^{-1}) for CH_4, based on analysis of data gathered during in-flight and ground-based calibrations. A linear calibration curve was derived for each flight day using in-flight measurements of all three calibrations cylinders on board the aircraft. In all cases we obtained r^2 values greater than 0.999, demonstrating good instrument linearity.

Flights were performed on nine separate days between November 2018 and March 2020, with all flights taking place in either November, February or March. Flight tracks are shown in Figure S4.1, which also gives the dates and times of the flights. The total flight duration for each day (as measured from the first usable measurement to the final usable measurement) varied between 2.5 hours to 5.5 hours.

Transport modelling

The surface influence, or footprint, of the sampled air was determined using HYSPLIT v5.0.0 (Hybrid Single Particle Lagrangian Integrated Trajectory Model; Draxler and Hess, 1998; Stein et al., 2015; Loughner et al., 2021), a Lagrangian particle dispersion model developed by the NOAA Air Resources Laboratory (ARL). This version of HYSPLIT incorporates elements of STILT (Stochastic Time-Inverted Lagrangian Transport Model; Gerbig et al., 2003; Lin et al., 2003; Fasoli et al., 2018), a related model that branched off an earlier version of HYSPLIT. This

combines features from STILT, including options for a different vertical turbulence scheme that prevents well-mixed particles from accumulating in low-turbulence regions (Thomson et al., 1997), a varying vertical Lagrangian timescale (Hanna, 1982) and an additional boundary layer turbulence parameterisation (Hanna, 1982), with recent model development and bug fixes within HYSPLIT (Loughner et al., 2021).

Each of our HYSPLIT runs was repeated using four different choices of input meteorology and two different turbulence parameterisations (Hanna, 1982; Kantha and Clayson, 2000) giving us an eight-member ensemble of dispersion model runs (hereafter referred to as the transport model ensemble). The input meteorology was taken from four publicly available archived products: the European Centre for Medium Range Weather Forecasts Fifth Reanalysis (ERA5), the NOAA Global Forecast System (GFS) model, the NOAA North American Mesoscale Forecast System (NAM) model, and the NOAA High-Resolution Rapid Refresh (HRRR) model (see SI section S2 for more details on each product).

A separate HYSPLIT model run was conducted for every minute of each flight that included aircraft sampling within the boundary layer. The mean measured mole fraction for each minute was calculated as an average over all measurement points where the aircraft was within the boundary layer, with the corresponding model particle release consisting of ~1000 particles distributed equally across these measurement locations (covering a horizontal distance of ~ 3 km to 4 km at typical aircraft speeds). In this way we ensured that the model run was representative of the measurements we included in our average.

Tiered multi-resolution Bayesian inversion framework

Fluxes were optimised separately for each flight using a Bayesian inversion framework, after first subtracting the background mole fraction from the aircraft measurements (where the

154 background is defined as the mole fraction of CO_2 or CH_4 within the measured air prior to it
155 entering the domain; see SI section S9 for details). This process consisted of multiple stages:

156 1) Optimise fluxes on a large domain at coarse spatial resolution (0.08º × 0.08º).
157 2) Use the posterior flux map from step 1) to estimate spatial variability in the mole fractions
158 of CO_2 and CH_4 flowing into a smaller, high resolution (0.02º × 0.02º) domain, nested within
159 the large domain and centred on New York City.
160 3) Optimise fluxes on the smaller, high-resolution domain, taking into account the background
161 variability calculated in step 2).

162 The large domain (henceforth referred to as the d01 domain) was bounded by 34.4ºN, 44.4ºN,
163 83.7ºW, 69.7ºW, while the smaller nested domain (henceforth the d03 domain) was bounded by
164 39.2ºN, 42.0ºN, 75.7ºW, 72.1ºW (see SI Figure S3.1 for map). This approach takes account of
165 upwind sources that influenced our measurements and provides an optimised representation of
166 their influence. This allows us to better isolate the specific contribution from the area of interest
167 (e.g., the New York-Newark urban area) by providing an optimised background for each minute
168 of the flight.

169 A gridded footprint (representing the influence of surface fluxes on the sampled air) was
170 calculated on both the d01 and d03 domains for each minute of the flight. Modelled mole fraction
171 enhancements (in units of μmol mol^{-1}; commonly referred to as ppm) were derived by multiplying
172 these gridded footprints (in units of ppm (μmol m^{-2} s^{-1})$^{-1}$) by surface fluxes (in units of μmol m^{-2}
173 s^{-1}) of CO_2 and CH_4. The inventories used to provide prior fluxes are described in the following
174 section. For both the initial d01 inversion and the nested d03 inversion, optimised posterior fluxes
175 were obtained by minimising the cost function $J(\boldsymbol{x})$:

176 $$J(\boldsymbol{x}) = \frac{1}{2}[(\boldsymbol{x}-\boldsymbol{x_b})^T\mathbf{P_b^{-1}}(\boldsymbol{x}-\boldsymbol{x_b}) + (\mathbf{H}\boldsymbol{x}-\boldsymbol{y})^T\mathbf{R^{-1}}(\mathbf{H}\boldsymbol{x}-\boldsymbol{y})] \quad (1)$$

This cost function assumes normally distributed uncertainties and has the following analytical solution (Enting, 2002; Tarantola, 2005):

$$\boldsymbol{x_a} = \boldsymbol{x_b} + \mathbf{P_b H^T}(\mathbf{H P_b H^T} + \mathbf{R})^{-1}(\boldsymbol{y} - \mathbf{H}\boldsymbol{x_b}) \quad (2)$$

Here $\boldsymbol{x}$ is a vector containing the fluxes that are to be optimised, $\boldsymbol{x_a}$ contains the posterior (optimised) fluxes, $\boldsymbol{x_b}$ contains prior fluxes derived from an emissions inventory and $\mathbf{H}$ is a matrix containing the modelled footprint for each minute of the flight. The vector $\boldsymbol{y}$ contains the measured mole fraction enhancements for each minute of the flight, relative to a background value representing the mole fraction present in the sampled air when it entered the domain. For the d01 inversion, this background value was taken to be constant, while for the d03 inversion a spatially-varying background was calculated based on the posterior fluxes within the d01 domain (see SI section S9 for more details, including details of a sensitivity analysis to determine the impact of the nested inversion approach). The error covariance matrices $\mathbf{P_b}$ and $\mathbf{R}$ represent uncertainty in the prior fluxes and model-measurement mismatch respectively.

Error covariances

Following previous studies (Lauvaux et al., 2016; 2020; Lopez-Coto et al., 2017; 2020), the off-diagonal components of the error covariance matrix $\mathbf{P_b}$ were populated using an exponential covariance model (in space). The diagonal elements of $\mathbf{P_b}$ were populated on the assumption that the 1σ flux uncertainty in each grid cell equalled 100% of the prior flux within that grid cell, following the values used by Lopez-Coto et al. (2017; 2020). For comparison, Andres et al. (2016) reported 2σ grid-cell-level uncertainties for a global CO_2 inventory ranging from 4% to more than 190%, averaging about 120%. Gately and Hutyra (2017) reported 50% to 250% differences between ACES (Anthropogenic Carbon Emissions System; version 1) and three global disaggregated CO_2 inventories at city scales, with grid-cell-level differences of over 100% for half

of the grid cells (urban and rural) in the northeast U.S. For CH_4 there are no good estimates of grid-cell-level uncertainty, but considering recent work indicating inventories significantly underestimate urban CH_4 emissions, and the lack of knowledge about CH_4 sources at urban scales, we can only assume that the relative uncertainties are at least as large as for CO_2. A lower limit uncertainty in each grid cell was set to 1 μmol m^{-2} s^{-1} and 1 nmol m^{-2} s^{-1} for CO_2 and CH_4 respectively, allowing some correction in grid cells with very low prior fluxes.

The off-diagonal elements of the prior flux error covariance matrix, representing the correlation between prior flux errors in different grid cells, were calculated according to an exponential decay model based on the distance between grid cells (Lauvaux et al., 2012; 2016; 2020; Lopez-Coto et al., 2017; 2020). The correlation length was set to 10 km following Lopez-Coto et al. (2017) who found this value to be appropriate for studies at urban scales. A variogram analysis of the spread of the prior flux ensemble indicated that the exponential covariance model and the correlation length used here are appropriate for both species in both domains (see SI section S11). The same approach to constructing $\mathbf{P_b}$ was used for both the d01 and d03 inversions.

A double exponential covariance model (in space and time) was selected for the model-measurement mismatch error covariance matrix **R**, with a correlation length of 1 km and a correlation time of 1 hour based on the short correlation length and time scales reported for atmospheric trace gases in urban environments (Shusterman et al., 2018; Turner et al., 2020). The diagonal elements of **R** represent the combined uncertainty due to random error in the measurements, model and background for each minute of the flight. This was calculated by combining background uncertainty (see below) with the variance across the eight transport model ensemble members for a given minute and the variance in measured mole fraction during that minute. In all cases, low limits for

the uncertainty of 0.2 ppm (μmol mol^{-1}) and 3 ppb (nmol mol^{-1}) were used for CO_2 and CH_4 respectively, based on an analysis of instrument precision from calibration data.

The background uncertainty for the d01 inversion was derived by calculating the respective variances in measured and modelled mole fractions across the background data points, then summing these variances (see SI section S9 for the definition of the background points). For the d03 inversion these terms were combined with the variance of the spatially-varying component of the background (i.e., the outside contribution derived from the posterior d01 fluxes), taken across the eight transport model ensemble members.

Prior fluxes

Separate inversions were performed using three different CO_2 flux priors and four different CH_4 flux priors. For CO_2 we used fluxes from Vulcan v3.0 (Gurney et al., 2020), ACES v2.0 (Gately and Hutyra, 2017) and EDGAR v5.0 (Emission Database for Global Atmospheric Research; European Commission Joint Research Centre (JRC)/Netherlands Environmental Assessment Agency (PBL), 2019; Crippa et al., 2020). For CH_4 we used fluxes from EDGAR v5.0, EDGAR v4.2 (European Commission Joint Research Centre (JRC)/Netherlands Environmental Assessment Agency (PBL), 2011) and the gridded Environmental Protection Agency (GEPA) inventory (Maasakkers et al., 2016). Because EDGAR v4.2 uses a population proxy to distribute fluxes, it contains much larger emissions from urban areas compared to EDGAR v5.0 and the GEPA inventory (conversely it is known to underestimate emissions in oil and natural gas production regions). To bridge this large emissions gap between inventories, we also used a composite CH_4 prior containing the average of the GEPA fluxes and the EDGAR v4.2 fluxes in each grid cell.

In all cases we used annual average emissions for the most recent year available: 2015 for Vulcan, 2017 for ACES, 2018 for EDGAR v5.0 CO_2, 2015 for EDGAR v5.0 CH_4, 2012 for the

GEPA, and 2008 for EDGAR v4.2. The use of annual average fluxes (even for those inventories for which hourly fluxes are available) was motivated by the fact that we did not have inventory fluxes for the actual flight days (which were more recent than the latest available inventory data). Furthermore, holding the prior emission rates constant for each flight ensures that flight-to-flight differences in posterior emission rates are not a consequence of prior temporal variability. While prior constraints are important for solving ill-posed problems, it is known that the choice of prior flux map can influence the posterior fluxes (Lauvaux et al., 2020). By using multiple priors for each gas we reduce the dependency of our mean posterior fluxes on one specific prior.

A conservative method was used to regrid each prior to the extent and resolution of the d01 and d03 domains (Zhuang, 2020). Canadian emissions for the U.S.-specific priors (Vulcan, ACES and GEPA) were taken from EDGAR v5.0. A more detailed discussion of prior fluxes is given in SI section S3.

Sensitivity analysis

In addition to our base case ensemble of inversions, we performed a sensitivity analysis to better understand the impact of different background choices and prior error covariance parameters in our posterior results, as well as to quantify uncertainties in the methodology. A full description of the sensitivity test cases is provided in SI sections S10 and S11.

Results and discussion

Total posterior emission rates were calculated for both the five boroughs of New York City (NYC) and the wider New York-Newark urban area (see SI section S3 for area definitions). The wider urban area better represents the area of sensitivity for our flights (see SI Figure S4.1 for footprint maps), therefore these wider urban area emission totals are the focus of analysis in this

section. Emission rate estimates for the five boroughs of NYC are also presented due to the high policy relevance of this area.

Urban area posterior emission rates

Figure 1 shows total emission rates within the New York-Newark urban area boundary, broken down by flight, prior and transport model. The mean posterior emission rate for CO_2 was (125 ± 39) kmol s^{-1}, where the reported uncertainty represents the 1σ variability across the nine flights (using ensemble-average totals for each flight). Also shown (in purple) are inventory emissions from the hourly ACES inventory and hourly Vulcan inventory, calculated using only dates and hours that were representative of our flights (see SI section S5 for details). The mean value of these representative inventory emissions was (145 ± 21) kmol s^{-1} for ACES (15.5% larger than our mean posterior estimate) and (124 ± 20) kmol s^{-1} for Vulcan (0.9% lower than our mean posterior estimate). The inventory variability reported here represents 1σ across 45 representative days in each inventory. Both our mean posterior emission rate and these representative inventory emission rates are larger than the corresponding annual average emission rates (116 kmol s^{-1} for ACES and 105 kmol s^{-1} for Vulcan) that we used as priors. While we do not expect these representative inventory values (from previous years) to agree perfectly with our mean posterior emissions rates, the agreement shown here is well within the observed flight-to-flight variability of our estimates as well as the expected daily variability of the inventories.

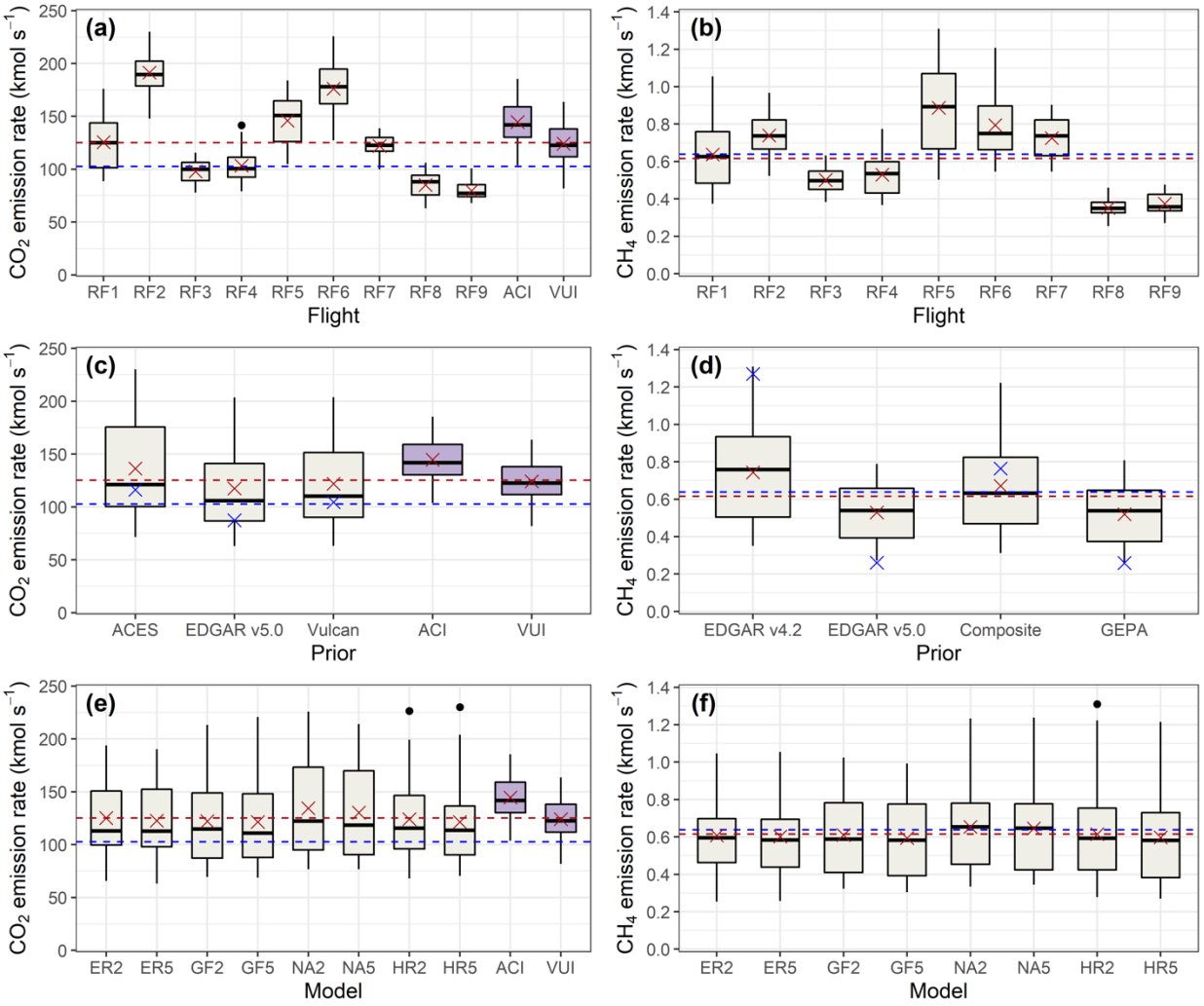

Figure 1. Boxplots of the total posterior emission rates for the New York-Newark urban area. Results are shown for CO_2 (a, c, e) and CH_4 (b, d, f), grouped by flight (a, b), by prior (c, d) and by transport model (e, f). Mean posterior emission rates are represented by red crosses and panels (c) and (d) also show prior values as blue crosses. The dashed lines represent the mean posterior (red) and prior (blue) emission rates taken over all flights, model ensemble members and priors. In the CO_2 panels, representative inventory estimates are also shown (ACI is ACES inventory and VUI is Vulcan inventory). In the transport model panels (e, f) ER is ERA5, GF is GFS, HR is HRRR and NA is NAM while the 2 and 5 respectively represent the Kantha and Clayson (2000) and Hanna (1982) turbulence parameterisations. The boxes extend between the upper and lower quartiles, with the median values shown as solid horizontal black bars. The whiskers extend to the highest and lowest data points within 1.5 times the interquartile range of the upper and lower quartiles, respectively. All data outside these whiskers are shown as individual points.

The posterior CO_2 emission estimates represent total emissions and are influenced by any contribution from biospheric respiration and photosynthesis, while the inventories only include anthropogenic fluxes. However, the flights in this study were conducted during the non-growing season and the estimated biospheric contribution was very small, ~ 2% on average (see SI section S6 for details).

The posterior CH_4 emission rates calculated for the New York-Newark urban area using all four priors are much larger than emission rates from the GEPA and EDGAR v5.0 inventories, but much smaller than the emission rate from EDGAR v4.2. The mean posterior CH_4 emission rate was (0.62 ± 0.19) kmol s^{-1} (1σ variability across flights). This mean posterior total is 2.4 times larger than the annual value from the GEPA inventory.

The temporal variability of urban CH_4 emissions on diurnal, weekly and seasonal timescales is poorly understood. None of the CH_4 inventories used in this study provide emission estimates on hourly timescales, so it is not possible to repeat the calculation of representative inventory totals presented above for CO_2. Recent studies in Los Angeles (Yadav et al., 2019) and Washington, DC-Baltimore (Huang et al., 2019) found large seasonal cycles in urban CH_4 emissions. Floerchinger et al. (2021) observed significant seasonal variability in the biogenic fraction of CH_4 emissions for several cities, including New York. More flights are required to provide the annual coverage necessary to confirm if strong seasonality also exists in total CH_4 emissions for the New York-Newark urban area.

It is important to emphasise that our mean posterior emission rates (for both CO_2 and CH_4) do not represent estimates of annual average emissions, but can be regarded as the current best top-down estimate for the non-growing-season, daytime value. However, the finding that emissions are underestimated in the GEPA inventory is likely to hold at the annual timescale, as emissions

during much of the rest of the year would have to be almost zero in order to offset the 2.4 times larger CH_4 emissions observed during November, February and March in this study.

Sources of variability and uncertainty analysis

The flight-to-flight variability in urban area posterior emission rate was 31% for both CO_2 and CH_4 (1σ, relative to the mean posterior emission rate). These values for relative flight-to-flight variability are similar to those reported by Lopez-Coto et al. (2020) for the Washington, DC-Baltimore urban area. Table 1 gives the 1σ spread in posterior emission rate across base case ensemble members (transport models and priors) and sensitivity test ensemble members (prior error covariance parameters and background definitions – see SI for more details), including the average spread for a single flight and the spread in campaign-average emissions. For both species, flight-to-flight variability in posterior emission rate was larger than the combined spread across the ensembles (calculated by adding individual ensemble spreads in quadrature). It is noteworthy that the combined ensemble spread for CH_4 is much smaller than the difference between the mean posterior emission rate and the inventory values.

Table 1: 1σ spread in posterior emission rate for the New York-Newark urban area across each ensemble, relative to the base case ensemble mean.

	Single-flight ensemble spread		**Campaign-average ensemble spread**	
Ensemble	**CO_2 (%)**	**CH_4 (%)**	**CO_2 (%)**	**CH_4 (%)**
Transport model	13	12	3.8	3.6
Prior	8.2	18	7.9	18
Prior flux covariance parameters	4.6	4.2	3.7	2.8
Background	4.6	8.6	4.3	7.7
Combined	16	24	10	20

To provide a rough estimate of expected variability in true CO_2 emissions from flight to flight, we calculated the variance in inventory emissions for the urban area using dates and hours corresponding to each flight (see SI section S5 for details). The standard deviation in inventory emissions across 45 representative days in ACES was equal to 15% of the average emissions across these days, while for Vulcan this value was 16% (boxplots showing these representative emissions for both inventories are shown in purple in Figure 1).

There are multiple potential reasons why the flight-to-flight variability (31%) of our posterior emissions is larger than that calculated using representative inventory emissions. Differences in sampling pattern for each flight (shown in SI Figure S4.1) mean that different parts of the urban area were sampled on different flights. Because the measurements are not sensitive to the whole urban area for each flight, emissions in unsampled areas default to the emission rate of the prior. In addition, different areas were sampled at different times within each flight, which for sources with large hourly variability in emission rate can result in apparent flight-to-flight variability in posterior emissions. Thus, this irregular sampling combined with large spatio-temporal variability in fluxes can become aliased as apparent temporal variability, as demonstrated by Lopez-Coto et al. (2020). In addition, for a given ensemble (e.g., transport or background), errors in the ensemble mean can differ in magnitude and direction for each flight, thus increasing the observed flight-to-flight variability.

Finally, the variability in inventory emissions may not be truly representative of real emission variability because the temporal emission profiles in the inventories are partly based on interpolation and/or downscaling using proxy data (resulting in smoothed temporal emission profiles). It is also worth considering that inventory emissions are for previous years, which may have had lower variability than the flight days, and that the flights were conducted in two different

362 winters (2018 to 2019 and 2019 to 2020) while the inventories are only for a single year each, so
363 potential interannual variability is not accounted for.

364 Quantifying the accuracy of the emission rate estimates is challenging since the true value is not
365 known. On one hand, the combined campaign-average ensemble spreads (10% for CO_2, 20% for
366 CH_4) could be larger than the uncertainty of the mean posterior emission rate, because this mean
367 value represents an average over the base case ensemble members. Conversely, the potential
368 methodological sources of flight-to-flight variability discussed above (e.g., emission aliasing) are
369 not accounted for in the campaign-average ensemble spread, but nonetheless contribute towards
370 the overall uncertainty. In general we focus on flight-to-flight variability when reporting results,
371 rather than the campaign-average ensemble spread, but it is important to note that the 1σ flight-to-
372 flight variability includes both real variability in emissions as well as methodological uncertainty.

373 **NYC posterior emission rates**

374 We also calculated mean posterior emission rates for the five boroughs of NYC: (45 ± 18) kmol
375 CO_2 s^{-1} and (0.20 ± 0.07) kmol CH_4 s^{-1} (1σ variability across flights). As was the case for the
376 urban area totals, inventory CO_2 emission rates calculated for NYC using dates and hours
377 representative of our flights are consistent with our posterior values. The representative NYC
378 emission rate for ACES was (45 ± 9) kmol CO_2 s^{-1}, while for Vulcan it was (52 ± 10) kmol CO_2
379 s^{-1}. The fact that the mean NYC posterior emission rate for CO_2 is in line with expectations based
380 on representative inventory emissions (within 15% for both inventories) is an encouraging sign
381 that the inversion is able to estimate emissions at spatial scales smaller than the wider urban area.
382 Further details of the spatial structure of prior and posterior fluxes in central NYC are given in SI
383 section S8.

It is difficult to compare our posterior emission rates directly to the NYC self-reported inventory (SRI), because that inventory includes a mixture of scope 1, 2 and 3 emissions and does not provide scope 1 totals for each emitted species. The 2019 SRI total for CH_4 is 0.21 kmol s^{-1} (New York City Mayor's Office of Sustainability, 2020), but this is dominated by waste emissions (0.17 kmol s^{-1}), which are mainly scope 3 (as landfill waste is exported from the city; MacWhinney and Klagsbald, 2017). The SRI CH_4 emission estimate for natural gas distribution leakage is a scope 1 value; for 2019 this was calculated to be 0.041 kmol s^{-1} (21% of our posterior emission total). Considering that previous studies (Plant et al., 2019; Floerchinger et al., 2021) have shown that the majority of New York CH_4 emissions are from thermogenic (i.e. natural gas) sources, it is likely that the SRI underestimates scope 1 thermogenic emissions. More measurements of NYC C_2H_6 emissions could enable a direct quantitative assessment of thermogenic SRI CH_4 emissions. CO_2 emissions in the NYC SRI are also comprised of multiple scopes, but a recent study (Gurney et al., 2021) has shown that total emissions from scope-1-only source categories in the NYC SRI are lower than the corresponding total Vulcan emissions for these categories by 19%.

Posterior−prior spatial differences

The spatial distributions of the difference between posterior and prior fluxes are shown in Figure 2. Prior emissions were adjusted upwards throughout the urban area for Edgar v5.0 CO_2, Edgar v5.0 CH_4 and GEPA CH_4. Conversely the large prior CH_4 emissions in Edgar v4.2 (distributed according to population) were adjusted downwards throughout the urban area. Posterior fluxes estimated using the Vulcan and ACES CO_2 priors showed an upwards adjustment throughout most of the urban area, but a downwards adjustment in central New York City (specifically in the southern part of Manhattan – see SI Figure S8.1 for a zoomed in map). The spread across posterior fluxes using three different priors for this area is smaller than the spread across the prior fluxes

themselves. However, further research is required, including higher resolution transport models, to understand whether small scale spatial features observed in these posterior emission maps reflect real spatial patterns in emissions or artefacts of the relatively coarse resolution of the transport models used (see SI section S2 for more details regarding the transport model configuration). New York City is a very complex land/urban/water interface and increased resolution would allow for better representation of smaller spatial features that could alter the local circulations.

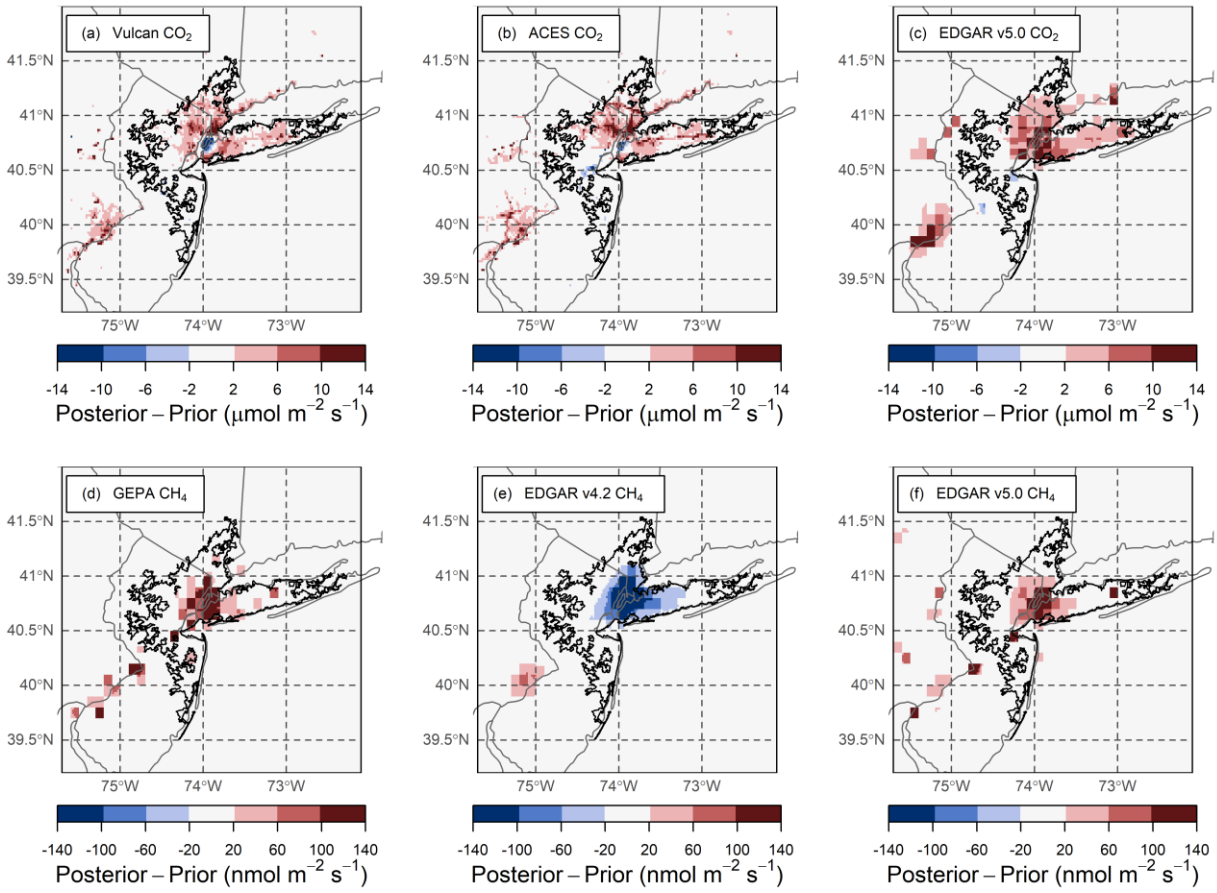

Figure 2. Maps showing the campaign-average difference between posterior and prior fluxes. Results are shown for CO_2 (a–c) and CH_4 (d–f). The New York-Newark urban area boundary is shown in black. Colour scale is saturated at the shown limits.

Emission ratio

The average posterior CH_4:CO_2 emission ratio for the New York-Newark urban area is (4.9 ± 0.7) nmol μmol^{-1} (1σ variability across flights), and for New York City it is (4.4 ± 1.1) nmol μmol^{-1}. As for the posterior CO_2 and CH_4 emission rate estimates, this ratio should be interpreted as an average over the dates and times (daytime, non-growing season) of our flights, not an annually representative value. Nonetheless, it is interesting to compare our emission ratio against enhancement ratios for the New York-Newark urban area reported by previous studies. Plant et al. (2019) reported an average CH_4:CO_2 enhancement ratio of 7.2 (+1.4/−1.0) nmol μmol^{-1} (95% confidence interval), taken over 10 flight days during April and May 2018. A lower enhancement ratio of 3.7 (+0.3/−0.2) nmol μmol^{-1} (95% confidence interval) was reported by Floerchinger et al. (2021) for a single ("winter") flight in March 2018 (see SI Figure S7.1 for a comparison plot).

Plant et al. (2019) also estimated a CH_4 emission rate for the urban area, which they calculated as the product of their CH_4:CO_2 enhancement ratio and an inventory estimate of CO_2 emissions. They derived a total CH_4 emission rate that was 2.7 times larger than the GEPA inventory value, in relatively close agreement to the results of our study (2.4 times larger than the GEPA inventory). At first sight, this appears at odds with the fact that the Plant et al. (2019) enhancement ratio is significantly larger than our estimated emission ratio. This apparent discrepancy is reconciled by considering that the CO_2 inventory emission rate used by Plant et al. (2019) is lower than our mean posterior CO_2 emissions. This is not unexpected, as CO_2 emission inventories suggest emissions during April and May are typically lower than emissions during November, February and March (as shown in SI Figure S5.1).

Conclusions

Our posterior emission rate estimates show good agreement with representative values from the ACES and Vulcan CO_2 inventories. In contrast, none of the available CH_4 inventories fell within a factor of 2 of our posterior estimates. It is of particular policy relevance that our posterior CH_4 emission rate estimate for the New York-Newark urban area was 2.4 times larger than the GEPA inventory. Taken together with previous studies in other cities, there is strong evidence that the GEPA inventory widely underestimates urban CH_4 emissions.

Underestimation of urban CH_4 emissions could impede urban emission reduction policies if cities neglect to address important (but difficult to quantify) emission sources such as natural gas leaks. Accurate, spatially-resolved bottom-up estimates of U.S. CH_4 emissions would allow cities to develop better informed emission reduction policies, but uncertainty in the magnitude and location of key sources (e.g., emissions from natural gas distribution) makes it difficult to compile such an inventory with the required accuracy.

Aircraft measurements enable top-down estimates of urban greenhouse gas emissions at spatial scales representative of the whole urban area. To track changes in annual emissions, it is necessary to design aircraft sampling campaigns to include repeat flights covering weekly and seasonal timescales of emission variability. In addition, as suggested by Lopez-Coto et al. (2020), increasing the number of simultaneous measurements might at least partially alleviate the flight-to-flight variability caused by irregular sampling. In the case of New York City, increasing the seasonal coverage of flights to better quantify the poorly understood seasonal cycle of CH_4 emissions is a particular priority. Regular aircraft sampling could be especially valuable in cases such as New York that lack an extensive network of urban tower measurements.

Inverse modelling can play a key role going forward by providing independent quantification of emissions in near-real time. It can be used to evaluate emission estimates in the past (if measurements exist) and it can be used to update inventories and tailor policy in the present. There is an inherent positive feedback loop between top-down inverse modelling emission estimates and inventory development. Top-down emission estimates can be used to improve the models and data on which the inventories rely, resulting in better inventory products. These improved inventories then provide more accurate priors for subsequent inverse modelling studies, resulting in improved emission estimates and reduced posterior uncertainties. Together, top-down and bottom-up emission estimation approaches can effectively guide emission reduction policy in urban areas.

References

Andres, RJ, Boden, TA, Higdon, DM. 2016. Gridded uncertainty in fossil fuel carbon dioxide emission maps, a CDIAC example. *Atmospheric Chemistry and Physics* **16**(23): 14979–14995. DOI: http://dx.doi.org/10.5194/acp-16-14979-2016.

Bréon, FM, Broquet, G, Puygrenier, V, Chevallier, F, Xueref-Remy, I, Ramonet, M, Dieudonné, E, Lopez, M, Schmidt, M, Perrussel, O, Ciais, P. 2015. An attempt at estimating Paris area CO2 emissions from atmospheric concentration measurements. *Atmospheric Chemistry and Physics* **15**(4): 1707–1724. DOI: http://dx.doi.org/10.5194/acp-15-1707-2015.

Brioude, J, Angevine, WM, Ahmadov, R, Kim, S-W, Evan, S, McKeen, SA, Hsie, E-Y, Frost, GJ, Neuman, JA, Pollack, IB, Peischl, J, Ryerson, TB, Holloway, J, Brown, SS, Nowak, JB, Roberts, JM, Wofsy, SC, Santoni, GW, Oda, T, Trainer, M. 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 CO2 and their impacts. *Atmospheric*

Chemistry and Physics **13**(7): 3661–3677. DOI: http://dx.doi.org/10.5194/acp-13-3661-2013.

Brioude, J, Kim, S-W, Angevine, WM, Frost, GJ, Lee, S-H, McKeen, SA, Trainer, M, Fehsenfeld, FC, Holloway, JS, Ryerson, TB, Williams, EJ, Petron, G, Fast, JD. 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): D20305. DOI: http://dx.doi.org/10.1029/2011JD016215.

Cambaliza, MOL, Shepson, PB, Caulton, DR, Stirm, B, Samarov, D, Gurney, KR, Turnbull, J, Davis, KJ, Possolo, A, Karion, A, Sweeney, C, Moser, B, Hendricks, A, Lauvaux, T, Mays, K, Whetstone, J, Huang, J, Razlivanov, I, Miles, NL, Richardson, SJ. 2014. Assessment of uncertainties of an aircraft-based mass balance approach for quantifying urban greenhouse gas emissions. *Atmospheric Chemistry and Physics* **14**(17): 9029–9050. DOI: http://dx.doi.org/10.5194/acp-14-9029-2014.

Climate Mobilization Act, Local Law 97. 2019. New York City Council. Available at https://www1.nyc.gov/assets/buildings/local_laws/ll97of2019.pdf. Accessed September 23, 2021.

Crippa, M, Solazzo, E, Huang, G, Guizzardi, D, Koffi, E, Muntean, M, Schieberle, C, Friedrich, R, Janssens-Maenhout, G. 2020. High resolution temporal profiles in the Emissions Database for Global Atmospheric Research. *Scientific Data* **7**(1): 121. DOI: http://dx.doi.org/10.1038/s41597-020-0462-2.

Crosson, ER. 2008. A cavity ring-down analyzer for measuring atmospheric levels of methane, carbon dioxide, and water vapor. *Applied Physics B: Lasers and Optics* **92**(3): 403–408. DOI: http://dx.doi.org/10.1007/s00340-008-3135-y.

Deng, A, Lauvaux, T, Davis, KJ, Gaudet, BJ, Miles, N, Richardson, SJ, Wu, K, Sarmiento, DP,

Hardesty, RM, Bonin, TA, Brewer, WA, Gurney, KR. 2017. Toward reduced transport errors in a high resolution urban CO2 inversion system. *Elementa: Science of the Anthropocene* **5**: 20. DOI: http://dx.doi.org/10.1525/elementa.133.

Dlugokencky, EJ, Myers, RC, Lang, PM, Masarie, KA, Crotwell, AM, Thoning, KW, Hall, BD, Elkins, JW, Steele, LP. 2005. Conversion of NOAA atmospheric dry air CH4 mole fractions to a gravimetrically prepared standard scale. *Journal of Geophysical Research* **110**(D18): D18306. DOI: http://dx.doi.org/10.1029/2005JD006035.

Draxler, RR, Hess, GD. 1998. An overview of the HYSPLIT_4 modelling system for trajectories, dispersion, and deposition. *Australian Meteorological Magazine* **47**: 295–308. Available at https://www.arl.noaa.gov/documents/reports/MetMag.pdf. Accessed September 23, 2021.

Enting, IG. 2002. *Inverse problems in atmospheric constituent transport*. Cambridge, UK: Cambridge University Press. DOI: http://dx.doi.org/10.1017/CBO9780511535741.

European Commission Joint Research Centre (JRC)/Netherlands Environmental Assessment Agency (PBL). 2011. Emission Database for Global Atmospheric Research (EDGAR), release version 4.2 [dataset]. Available at https://edgar.jrc.ec.europa.eu/overview.php?v=42. Accessed October 2, 2020.

European Commission Joint Research Centre (JRC)/Netherlands Environmental Assessment Agency (PBL). 2019. Emission Database for Global Atmospheric Research (EDGAR), release version 5.0 [dataset]. Available at https://edgar.jrc.ec.europa.eu/overview.php?v=50_GHG. Accessed October 2, 2020.

Fasoli, B, Lin, JC, Bowling, DR, Mitchell, L, Mendoza, D. 2018. Simulating atmospheric tracer concentrations for spatially distributed receptors: updates to the Stochastic Time-Inverted Lagrangian Transport model's R interface (STILT-R version 2). *Geoscientific Model*

Development **11**(7): 2813–2824. DOI: http://dx.doi.org/10.5194/gmd-11-2813-2018.

Floerchinger, C, Shepson, PB, Hajny, K, Daube, BC, Stirm, BH, Sweeney, C, Wofsy, SC. 2021. Relative flux measurements of biogenic and natural gas-derived methane for seven U.S. cities. *Elementa: Science of the Anthropocene* **9**(1): 1–15. DOI: http://dx.doi.org/10.1525/elementa.2021.000119.

Fong, WK, Sotos, M, Doust, M, Schultz, S, Marques, A, Deng-Beck, C. 2014. Global protocol for community-scale greenhouse gas emission inventories. Washington DC, USA: World Resources Institute, C40 Cities, ICLEI. Available at https://ghgprotocol.org/sites/default/files/standards/GHGP_GPC_0.pdf. Accessed September 23, 2021.

Gately, CK, Hutyra, LR. 2017. Large uncertainties in urban-scale carbon emissions. *Journal of Geophysical Research: Atmospheres* **122**(20): 11242–11260. DOI: http://dx.doi.org/10.1002/2017JD027359.

Gerbig, C, Lin, JC, Wofsy, SC, Daube, BC, Andrews, AE, Stephens, BB, Bakwin, PS, Grainger, CA. 2003. Toward constraining regional-scale fluxes of CO2 with atmospheric observations over a continent: 2. Analysis of COBRA data using a receptor-oriented framework. *Journal of Geophysical Research: Atmospheres* **108**(D24): 4757. DOI: http://dx.doi.org/10.1029/2003JD003770.

Gurney, K, Shepson, P. 2021. Opinion: The power and promise of improved climate data infrastructure. *Proceedings of the National Academy of Sciences* **118**(35): e2114115118. DOI: http://dx.doi.org/10.1073/pnas.2114115118.

Gurney, KR, Liang, J, Patarasuk, R, Song, Y, Huang, J, Roest, G. 2020. The Vulcan version 3.0 high-resolution fossil fuel CO2 emissions for the United States. *Journal of Geophysical*

Research: Atmospheres **125**(19). DOI: http://dx.doi.org/10.1029/2020JD032974.

Gurney, KR, 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): 553. DOI: http://dx.doi.org/10.1038/s41467-020-20871-0.

Hanna, SR. 1982. Applications in air pollution modelling, in Nieuwstadt, FT, Van Dop, H eds. *Atmospheric turbulence and air pollution modelling*. Dordrecht, The Netherlands: Reidel. 275–310.

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

Hutyra, LR, Duren, R, Gurney, KR, Grimm, N, Kort, EA, Larson, E, Shrestha, G. 2014. Urbanization and the carbon cycle: Current capabilities and research outlook from the natural sciences perspective. *Earth's Future* **2**(10): 473–495. DOI: http://dx.doi.org/10.1002/2014EF000255.

Janssens-Maenhout, G, Crippa, M, Guizzardi, D, Muntean, M, Schaaf, E, Dentener, F, Bergamaschi, P, Pagliari, V, Olivier, JGJ, Peters, JAHW, van Aardenne, JA, Monni, S, Doering, U, Petrescu, AMR, Solazzo, E, Oreggioni, GD. 2019. EDGAR v4.3.2 Global Atlas of the three major greenhouse gas emissions for the period 1970–2012. *Earth System Science Data* **11**(3): 959–1002. DOI: http://dx.doi.org/10.5194/essd-11-959-2019.

Kantha, LH, Clayson, CA. 2000. *Small scale processes in geophysical fluid flows*. San Diego: Academic Press.

Lamb, BK, Cambaliza, MOL, Davis, KJ, Edburg, SL, Ferrara, TW, Floerchinger, C, Heimburger,

AMF, Herndon, S, Lauvaux, T, Lavoie, T, Lyon, DR, Miles, N, Prasad, KR, Richardson, S, Roscioli, JR, Salmon, OE, Shepson, PB, Stirm, BH, Whetstone, J. 2016. Direct and indirect measurements and modeling of methane emissions in Indianapolis, Indiana. *Environmental Science & Technology* **50**(16): 8910–8917. DOI: http://dx.doi.org/10.1021/acs.est.6b01198.

Lauvaux, T, Gurney, KR, Miles, NL, Davis, KJ, Richardson, SJ, Deng, A, Nathan, BJ, Oda, T, Wang, JA, Hutyra, L, Turnbull, J. 2020. Policy-relevant assessment of urban CO2 emissions. *Environmental Science & Technology* **54**(16): 10237–10245. DOI: http://dx.doi.org/10.1021/acs.est.0c00343.

Lauvaux, T, Miles, NL, Deng, A, Richardson, SJ, Cambaliza, MO, Davis, KJ, Gaudet, B, Gurney, KR, Huang, J, O'Keefe, D, Song, Y, Karion, A, Oda, T, Patarasuk, R, Razlivanov, I, Sarmiento, D, Shepson, P, Sweeney, C, Turnbull, J, Wu, K. 2016. High-resolution atmospheric inversion of urban CO2 emissions during the dormant season of the Indianapolis Flux Experiment (INFLUX). *Journal of Geophysical Research: Atmospheres* **121**(10): 5213–5236. DOI: http://dx.doi.org/10.1002/2015JD024473.

Lauvaux, T, Schuh, AE, Uliasz, M, Richardson, S, Miles, N, Andrews, AE, Sweeney, C, Diaz, LI, Martins, D, Shepson, PB, Davis, KJ. 2012. Constraining the CO2 budget of the corn belt: exploring uncertainties from the assumptions in a mesoscale inverse system. *Atmospheric Chemistry and Physics* **12**(1): 337–354. DOI: http://dx.doi.org/10.5194/acp-12-337-2012.

Lin, JC, Gerbig, C, Wofsy, SC, Andrews, AE, Daube, BC, Davis, KJ, Grainger, CA. 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**(D16): 4493. DOI: http://dx.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. DOI: http://dx.doi.org/10.1007/s00376-017-6094-6.

Lopez-Coto, I, Ren, X, Salmon, OE, Karion, A, Shepson, PB, Dickerson, RR, Stein, A, Prasad, K, Whetstone, JR. 2020. Wintertime CO2, CH4, and CO emissions estimation for the Washington, DC–Baltimore metropolitan area using an inverse modeling technique. *Environmental Science & Technology* **54**(5): 2606–2614. DOI: http://dx.doi.org/10.1021/acs.est.9b06619.

Loughner, CP, Fasoli, B, Stein, AF, Lin, JC. 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. DOI: http://dx.doi.org/10.1175/JAMC-D-20-0158.1.

Maasakkers, JD, Jacob, DJ, Sulprizio, MP, Turner, AJ, Weitz, M, Wirth, T, Hight, C, DeFigueiredo, M, Desai, M, Schmeltz, R, Hockstad, L, Bloom, AA, Bowman, KW, Jeong, S, Fischer, ML. 2016. Gridded national inventory of U.S. methane emissions. *Environmental Science & Technology* **50**(23): 13123–13133. DOI: http://dx.doi.org/10.1021/acs.est.6b02878.

MacWhinney, R, Barnett, M. 2019. Inventory of New York City greenhouse gas emissions in 2017. New York, USA: New York City Mayor's Office of Sustainability. Available at https://www1.nyc.gov/assets/sustainability/downloads/pdf/GHG_Inventory_2017.pdf. Accessed September 23, 2021.

MacWhinney, R, Klagsbald, O. 2017. Inventory of New York City greenhouse gas emissions in

2016. New York, USA: New York City Mayor's Office of Sustainability. Available at https://www1.nyc.gov/assets/sustainability/downloads/pdf/publications/GHG%20Inventory%20Report%20Emission%20Year%202016.pdf. Accessed September 23, 2021.

Marcotullio, PJ, Sarzynski, A, Albrecht, J, Schulz, N, Garcia, J. 2013. The geography of global urban greenhouse gas emissions: an exploratory analysis. *Climatic Change* **121**(4): 621–634. DOI: http://dx.doi.org/10.1007/s10584-013-0977-z.

Nangini, C, Peregon, A, Ciais, P, Weddige, U, Vogel, F, Wang, J, Bréon, F-M, Bachra, S, Wang, Y, Gurney, K, Yamagata, Y, Appleby, K, Telahoun, S, Canadell, JG, Grübler, A, Dhakal, S, Creutzig, F. 2019. A global dataset of CO2 emissions and ancillary data related to emissions for 343 cities. *Scientific Data* **6**(1): 180280. DOI: http://dx.doi.org/10.1038/sdata.2018.280.

New York City Mayor's Office of Sustainability. 2016. New York City's roadmap to 80 x 50. Available at https://www1.nyc.gov/assets/sustainability/downloads/pdf/publications/New%20York%20City%27s%20Roadmap%20to%2080%20x%2050_20160926_FOR%20WEB.pdf. Accessed September 23, 2021.

New York City Mayor's Office of Sustainability. 2020. Inventory of New York City Greenhouse Gas Emissions [dataset]. Available at https://nyc-ghg-inventory.cusp.nyu.edu/. Accessed April 6, 2021.

Nickless, A, Rayner, PJ, Scholes, RJ, Engelbrecht, F, Erni, B. 2019. An atmospheric inversion over the city of Cape Town: sensitivity analyses. *Atmospheric Chemistry and Physics* **19**(11): 7789–7816. DOI: http://dx.doi.org/10.5194/acp-19-7789-2019.

Oda, T, Maksyutov, S, Andres, RJ. 2018. The Open-source Data Inventory for Anthropogenic CO2 , version 2016 (ODIAC2016): a global monthly fossil fuel CO2 gridded emissions data

product for tracer transport simulations and surface flux inversions. *Earth System Science Data* **10**(1): 87–107. DOI: http://dx.doi.org/10.5194/essd-10-87-2018.

Pisso, I, Patra, P, Takigawa, M, Machida, T, Matsueda, H, Sawa, Y. 2019. Assessing Lagrangian inverse modelling of urban anthropogenic CO2 fluxes using in situ aircraft and ground-based measurements in the Tokyo area. *Carbon Balance and Management* **14**(1): 6. DOI: http://dx.doi.org/10.1186/s13021-019-0118-8.

Plant, G, Kort, EA, Floerchinger, C, Gvakharia, A, Vimont, I, Sweeney, C. 2019. Large fugitive methane emissions from urban centers along the U.S. east coast. *Geophysical Research Letters* **46**(14): 8500–8507. DOI: http://dx.doi.org/10.1029/2019GL082635.

Sethi, M, Lamb, W, Minx, J, Creutzig, F. 2020. Climate change mitigation in cities: a systematic scoping of case studies. *Environmental Research Letters* **15**(9): 093008. DOI: http://dx.doi.org/10.1088/1748-9326/ab99ff.

Shusterman, AA, Kim, J, Lieschke, KJ, Newman, C, Wooldridge, PJ, Cohen, RC. 2018. Observing local CO2 sources using low-cost, near-surface urban monitors. *Atmospheric Chemistry and Physics* **18**(18): 13773–13785. DOI: http://dx.doi.org/10.5194/acp-18-13773-2018.

Staufer, J, Broquet, G, Bréon, FM, Puygrenier, V, Chevallier, F, Xueref-Rémy, I, Dieudonné, E, Lopez, M, Schmidt, M, Ramonet, M, Perrussel, O, Lac, C, Wu, L, Ciais, P. 2016. The first 1-year-long estimate of the Paris region fossil fuel CO2 emissions based on atmospheric inversion. *Atmospheric Chemistry and Physics* **16**(22): 14703–14726. DOI: http://dx.doi.org/10.5194/acp-16-14703-2016.

Stein, AF, Draxler, RR, Rolph, GD, Stunder, BJB, Cohen, MD, Ngan, F. 2015. NOAA's HYSPLIT atmospheric transport and dispersion modeling system. *Bulletin of the American Meteorological Society* **96**(12): 2059–2077. DOI: http://dx.doi.org/10.1175/BAMS-D-14-

671 00110.1.
672 Tans, PP, Crotwell, AM, Thoning, KW. 2017. Abundances of isotopologues and calibration of
673 CO2 greenhouse gas measurements. *Atmospheric Measurement Techniques* **10**(7): 2669–
674 2685. DOI: http://dx.doi.org/10.5194/amt-10-2669-2017.
675 Tarantola, A. 2005. *Inverse problem theory and methods for model parameter estimation*.
676 Philadelphia, USA: Society for Industrial and Applied Mathematics.
677 Thomson, DJ, Physick, WL, Maryon, RH. 1997. Treatment of Interfaces in Random Walk
678 Dispersion Models. *Journal of Applied Meteorology* **36**(9): 1284–1295. DOI:
679 http://dx.doi.org/10.1175/1520-0450(1997)036<1284:TOIIRW>2.0.CO;2.
680 Trencher, G, Castán Broto, V, Takagi, T, Sprigings, Z, Nishida, Y, Yarime, M. 2016. Innovative
681 policy practices to advance building energy efficiency and retrofitting: approaches, impacts
682 and challenges in ten C40 cities. *Environmental Science & Policy* **66**: 353–365. DOI:
683 http://dx.doi.org/10.1016/j.envsci.2016.06.021.
684 Turner, AJ, Kim, J, Fitzmaurice, H, Newman, C, Worthington, K, Chan, K, Wooldridge, PJ,
685 Köehler, P, Frankenberg, C, Cohen, RC. 2020. Observed Impacts of COVID-19 on Urban
686 CO2 Emissions. *Geophysical Research Letters* **47**(22): 1–6. DOI:
687 http://dx.doi.org/10.1029/2020GL090037.
688 Yadav, V, Duren, R, Mueller, K, Verhulst, KR, Nehrkorn, T, Kim, J, Weiss, RF, Keeling, R,
689 Sander, S, Fischer, ML, Newman, S, Falk, M, Kuwayama, T, Hopkins, F, Rafiq, T,
690 Whetstone, J, Miller, C. 2019. Spatio-temporally Resolved Methane Fluxes From the Los
691 Angeles Megacity. *Journal of Geophysical Research: Atmospheres* **124**(9): 5131–5148. DOI:
692 http://dx.doi.org/10.1029/2018JD030062.
693 Zhuang, J. 2020. xESMF: Universal Regridder for Geospatial Data. DOI:

http://dx.doi.org/10.5281/zenodo.1134365.

Contributions

Contributed to conception and design: JRP, ILC, KDH, CRF, AK, JRW, PBS

Contributed to acquisition of data: JRP, ILC, KDH, JT, RK, TJ, BHS, CPL, CKG, LRH, KRG, GSR, JL, SG, PBS

Contributed to analysis and interpretation of data: JRP, ILC, KDH, CRF, CKG, GSR, SG, AK, PBS

Drafted and/or revised the article: JRP, ILC, KDH, CPL, LRH, GSR, SG, AK, PBS

Approved the submitted version for publication: JRP, ILC, KDH, JT, RK, TJ, BHS, CRF, CPL, CKG, LRH, KRG, GSR, JL, SG, AK, JRW, PBS

Acknowledgments

The authors thank FAA air traffic control in and around NYC for their support with our flight patterns and the Jonathan Amy Facility for Chemical Instrumentation at Purdue University for continued assistance in maintaining equipment in ALAR. We thank Róisín Commane for the loan of her Picarro analyser during the November 2019 flights.

Certain commercial equipment, instruments, or materials are identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by NIST nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

Funding information

We thank the National Institute of Standards and Technology (NIST) for financial support of this work, via Awards No. 70NANB16H262, 70NANB19H167 and 70NANB21H021. K. R. Gurney acknowledges the support from the National Aeronautics and Space Administration (NASA) Grant NNX14AJ20G and the NASA Carbon Monitoring System program: “Understanding User Needs for Carbon Monitoring Information” (subcontract 1491755).

Competing interests

The authors declare that no competing interests exist. Paul B. Shepson is an Associate Editor at Elementa. He was not involved in the review process of this article.

Supplemental material

Appendix A. Further details regarding dispersion model configuration, domains, footprints, inventory emissions, biospheric influence, emission ratios, posterior spatial distribution, inverse modelling configuration and sensitivity tests. (PDF)

Appendix B. Prior and posterior emission rates from individual ensemble members, and representative inventory emission rates. (XLSX)

Data accessibility statement

Aircraft measurement data used in this study are available from the Harvard Dataverse at https://doi.org/10.7910/DVN/UECQXX.