

RESEARCH ARTICLE

Assessing the bias and uncertainties in the aircraft mass balance technique for the determination of carbon dioxide emission rates

Kristian D. Hajny^{1,2,*} , David R. Lyon³, Austin Armstrong¹, Cody R. Floerchinger⁴, Thilina Jayarathne^{1,5}, Robert Kaeser¹, Tegan Lavoie¹, Olivia E. Salmon¹, Brian H. Stirm⁶, Andrew A. Stuff¹, Jay M. Tomlin¹, Bernard Wulle¹, Israel Lopez-Coto^{2,7}, and Paul B. Shepson^{1,2}

Urban areas are the major sources of greenhouse gas emissions but also leaders in emission reduction efforts. Appropriate techniques to quantify emissions and any potential reductions over time are necessary to effectively inform these mitigation efforts. The aircraft mass balance experiment (MBE) is an established technique used for such a purpose. In this work, we use a series of 55 MBEs downwind of power plants to assess the technique's bias and precision. In addition, we investigate what factors drive the absolute error, determined as the absolute difference between observed and reported emission rates, in individual experiments using multilinear regressions. Power plants are required to monitor their carbon dioxide emissions with an hourly resolution, and these publicly available reported emissions can be directly compared to the mass balance estimates as a pseudo-known release. To quantify the bias we calculated the mean error, which was $10 \pm 240 \text{ Mg}\cdot\text{h}^{-1}$ (1σ), regressed mass balance emission rates against reported emission rates to yield a slope of 0.967 ± 0.062 , and compared the sum across all mass balance emission rates, $31,000 \pm 1,000 \text{ Mg}\cdot\text{h}^{-1}$, to the sum across all reported emissions, $30,660 \pm 740 \text{ Mg}\cdot\text{h}^{-1}$. All three of these approaches suggest no systematic bias. Then to quantify the precision for individual determinations we calculated the slope of a regression between the standard deviation across repeated MBEs and the corresponding average emission rate, which is $30.7\% \pm 6.7\%$. The main drivers of the absolute error were sparse sampling of the plume, poor horizontal and vertical mixing of the plume, and smaller signal-to-noise ratios. Quantifying the capabilities of this technique provides context for previous analyses and allows stakeholders and researchers to make informed decisions when choosing quantification methods. Identifying the factors that drive the absolute error also allows us to adjust flight design to minimize it and potentially improve uncertainty estimates.

Keywords: Emission quantification, Airborne greenhouse gas measurements, Mass balance experiment

1. Introduction

The Biden administration has set national goals to reach a 50% reduction in greenhouse gases (GHG) by 2030 and

net zero by 2050, consistent with the Paris Agreement (Whitehouse.gov, 2021). This undertaking will require efforts at multiple scales, from the national to the state to the city. Urban areas alone account for approximately 45% of carbon dioxide (CO_2) emissions across the contiguous United States according to the Vulcan gridded emissions data product (Gurney et al., 2020; U.S. Census Bureau, 2021), and this percentage is expected to increase in the coming years as populations continue urbanizing (International Energy Agency, 2013; United Nations Department of Economic and Social Affairs Population Division, 2019). Accordingly, many cities have passed their own legislation and goals for specific GHG reductions (C40 Cities, 2022). However, high-precision measurement techniques are necessary to quantify emissions reductions to evaluate progress on this legislation and to identify effective mitigation strategies. We have conducted a large series of flights to assess the aircraft mass balance

¹ Department of Chemistry, Purdue University, West Lafayette, IN, USA

² School of Marine and Atmospheric Sciences, Stony Brook University, Stony Brook, NY, USA

³ Environmental Defense Fund, Austin, TX, USA

⁴ Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA, USA

⁵ Current address: Bristol Myers Squibb, New Brunswick, NJ, USA

⁶ School of Aviation and Transportation Technology, Purdue University, West Lafayette, IN, USA

⁷ Special Programs Office, National Institute of Standards and Technology, Gaithersburg, MD, USA

* Corresponding author:

Email: kristian.hajny@stonybrook.edu

experiment's (MBE) ability to address this need. This technique has been widely used in point source (cf. Carotenuto et al., 2018; Guha et al., 2020; and many others), regional (cf. Alfieri et al., 2010; Karion et al., 2015; and many others), and urban emission quantification studies (cf. Gioli et al., 2014; Klausner et al., 2020; and many others).

This work focuses on a data set of 55 MBEs from 26 flights at 19 different power plants conducted between 2016 and 2020, including some reanalyzed flights from previous publications (Hajny et al., 2019). This work takes advantage of this large data set to assess the aircraft MBE technique given the availability of power plant CO₂ emissions data that are estimated using continuous emissions monitoring systems (CEMS) and reported to the Environmental Protection Agency (U.S. Environmental Protection Agency [U.S. EPA], 2022). Here we quantify the bias and precision of the method, and we investigate the flight conditions that most significantly affect the performance of the MBE approach using multilinear regressions. It is important to clarify that because these analyses are based solely on point source MBEs, the results apply at the large emissions point source scale and may not translate perfectly to area sources where estimating the background is challenging and plumes are more difficult to define. Additionally, unlike typical methane (CH₄) point sources, these are hot buoyant plumes from an elevated stack and so these results may not be perfectly applicable to other types of point sources. However, this is an important step in better understanding the performance of the mass balance technique, which can help lead to more general results on other types of sources in the future.

Aircraft MBE techniques vary slightly across research groups, but few have directly assessed the capabilities of their technique against known emissions. Conley et al. (2017) and Ahn et al. (2020) are two of the few studies that have directly assessed the method precision and/or bias against known emissions. Both studies used a similar approach to calculate the bias of the quantification method, relying on power plant CEMS data (U.S. EPA, 2022) as the known emission rate and, thus, have the same limitations regarding area sources with a more complex background. Erland et al. (2022) also showed that, at least for the Conley et al. (2017) and the Environment and Climate Change Canada's top-down emission rate retrieval algorithm, if the underlying assumptions of the method are met, then the emission estimates are not overly sensitive to the choice in MBE algorithm. In our case, we used a very large data set and took the analysis a step further by examining the variables that contribute to scatter in the agreement between CEMS and MBE emission rates to assess variables that can be controlled to reduce the absolute error or to potentially help quantify it. Given the widespread application of the MBE technique, improving our understanding of these variables can provide important context for previous experiments while improving the experiment design of future works.

2. Materials and methods

2.1. Instrumentation

All flights were conducted using Purdue's Airborne Laboratory for Atmospheric Research (ALAR; Mays et al., 2009; Cambaliza et al., 2014; Cambaliza et al., 2015; Heimbürger et al., 2017; Lavoie et al., 2017; Hajny et al., 2019) which is a modified twin-engine Beechcraft Duchess with a typical airspeed of 66 m·s⁻¹. ALAR is equipped with a best air turbulence probe for 50-Hz high-precision 3D wind field measurements (Garman et al., 2006; Garman et al., 2008), a GPS/INS system reported at 50 Hz, 2 temperature probes reported at 50 Hz with measurement frequencies >1 Hz, and a Picarro cavity ring down spectrometer designed for 0.5 Hz measurements of CO₂, CH₄, and H₂O (Crosson, 2008). As the Picarro data are needed for analyses and have the slowest response time, all other relevant variables are downsampled to the Picarro's frequency.

Picarro measurements are reported as dry air mole fractions, in units of micromoles per mole of dry air, or parts per million (ppm). This work focuses exclusively on the CO₂ measurements because CEMS data do not include CH₄ and previous analyses focused on quantifying the CH₄ emissions from the subsets of these flights (Hajny et al., 2019). The Picarro operating range is between 0 and 1,000 ppm, with a guaranteed specifications range of 300–700 ppm and a flushing time of approximately 1.5 s (Picarro, Inc., 2021; Picarro, Inc., personal communication, 03/01/2023). Flights typically included 2 in-air three-point calibrations with National Oceanic and Atmospheric Administration-certified standard cylinders (WMO-CO₂-X2007) ranging from approximately 360 to approximately 450 parts per million CO₂ (Dlugokencky et al., 2005; Tans et al., 2017; Hall et al., 2021). The cylinder values, certification dates, and so on are provided in Supporting Information (SI) Table S1. The maximum measured concentration during the MBEs ranges from 395 ppm to 940 ppm with up to 11.5% of the data for a given MBE going above the range of the three-point calibration used. Picarro has stated in personal communication that the instrument is highly linear within its 1,000 ppm operating range, so we do not treat measurements above our calibration range differently (Picarro, Inc., personal communication, 03/01/2023).

2.2. Power plant sampling and the MBE method

We sampled 19 power plants over the period 2016–2020, covering a range of power plant emission rates in the Eastern United States. Details for each power plant are provided in SI Table S2. This list includes power plants discussed in previous works (Lavoie et al., 2017; Hajny et al., 2019) using the same identifiers. Those works chose to keep facilities anonymous, given facility identities do not affect the interpretation or results, and as such we do the same. Flights included between 1 and 4 MBEs at progressively increased downwind distances to quantify the emission rate. MBEs that were performed back-to-back can be considered repeat experiments if the CEMS emission rates are assumed to be constant, which we confirm with CEMS data, allowing us to use them to assess the method's

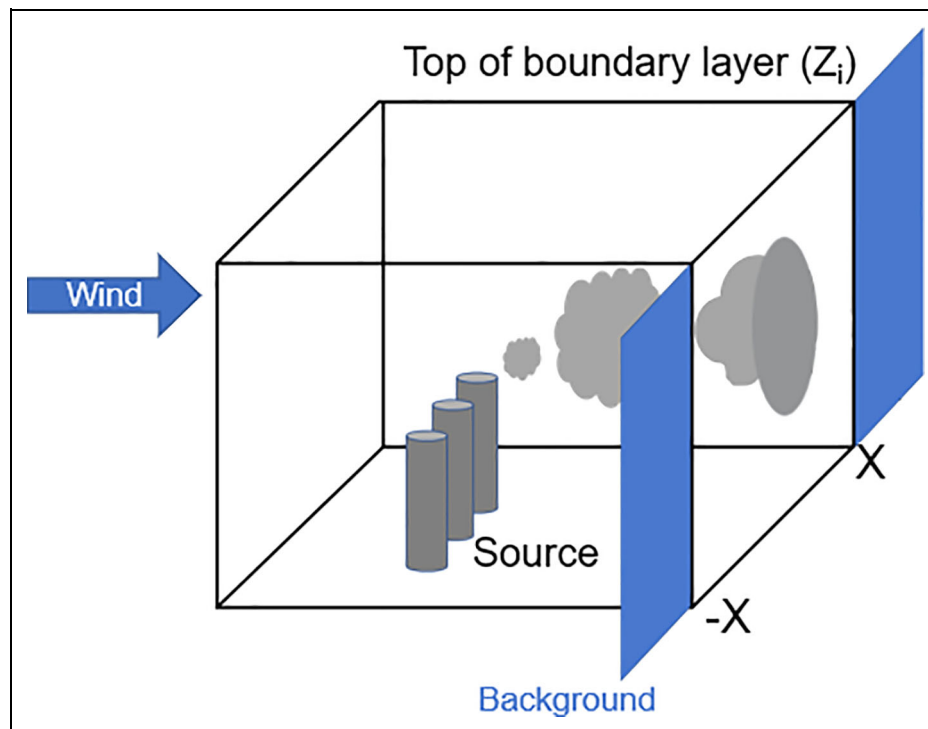


Figure 1. Conceptual diagram of a mass balance experiment. Flights involve multiple passes across the downwind plane from as low as is safe to the top of the plume. The plume is limited to the bounds of $-x$ to x and the background is then defined using data on the edges, shown here in blue.

precision. We note that our flight patterns are not Lagrangian as this is not only logistically impractical but also unnecessary since the emission rate is considered constant on the timescale of repeated MBE flight legs.

MBEs were typically conducted by flying at multiple altitudes, in approximately 250 ft intervals, at a fixed distance downwind of the site from as low to the ground as is safe to the highest altitude where enhanced CO_2 concentrations could be seen, often with a vertical spiral through the center of the plume. Each MBE took an average of 42 ± 17 min to complete, varying primarily based on the interval between passes and the width and height of the plume. The flights were designed so that these passes were approximately perpendicular to the mean wind direction. The MBEs were typically within 5 km of the power plant to measure the CH_4 as well given the much smaller CH_4 signal, but some, particularly those with more than 2 MBEs, went farther downwind, up to 13 km. These CH_4 measurements were the focus of previous works (Hajny et al., 2019). These downwind distances were chosen to provide an approximately Gaussian measurement and minimize sampling errors, that is, not close enough to inadequately sample the plume or risk signals above the operating range of the instrument, but not far enough to decrease the signal-to-noise (S/N) ratio too drastically.

This series of passes creates a theoretical 2D plane downwind of the power plant and we measure the flux through this plane. A conceptual diagram of this is shown in **Figure 1** and an example flight is shown in **Figure 2**. Transects extended sufficiently beyond the plume so that CO_2 concentrations returned to background values. An

ordinary least squares (OLS) regression was fit through the background values on either end of the plume to estimate the background concentrations for each transect (Heimburger et al., 2017; Lavoie et al., 2017; Ren et al., 2018; Hajny et al., 2019; Ahn et al., 2020). This estimate of the background is, on average, within 1.5 ppm of a background estimated using upwind data given how close the upwind and downwind data are in time and space, which is quite small in comparison to the measured enhancements that are often ≥ 100 ppm. We choose to use a background based on an OLS regression through data on either end of the plume in the downwind transect data to be consistent with previous publications using ALAR (Heimburger et al., 2017; Lavoie et al., 2017; Hajny et al., 2019). These 0.5-Hz pointwise concentration enhancements were then converted to pointwise fluxes according to Equation 1:

$$\text{Flux}_i = \frac{([C]_i - [C]_{\text{BG},i}) \times U_i \times \cos(\theta)_i \times P_i}{T_i \times R}. \quad (1)$$

Here $[C]_i$ and $[C]_{\text{BG},i}$ represent the measured and background concentrations of CO_2 respectively ($\text{mol}_c \cdot \text{mol}_{\text{air}}^{-1}$), U_i represents the measured horizontal wind speed ($\text{m} \cdot \text{s}^{-1}$), $\cos(\theta)_i$ is a unitless correction term ranging between 0 and 1 to account for cross winds with θ representing the angle between the wind direction and the transects, P_i represents the measured pressure (atm), T_i represents the measured temperature (Kelvin), and R represents the ideal gas constant ($\text{atm} \cdot \text{m}^3 \cdot \text{mol}_{\text{air}}^{-1} \cdot \text{K}^{-1}$). The subscript i indicates the point number, as the flux is calculated pointwise at 0.5 Hz using pointwise measurements for each variable. As

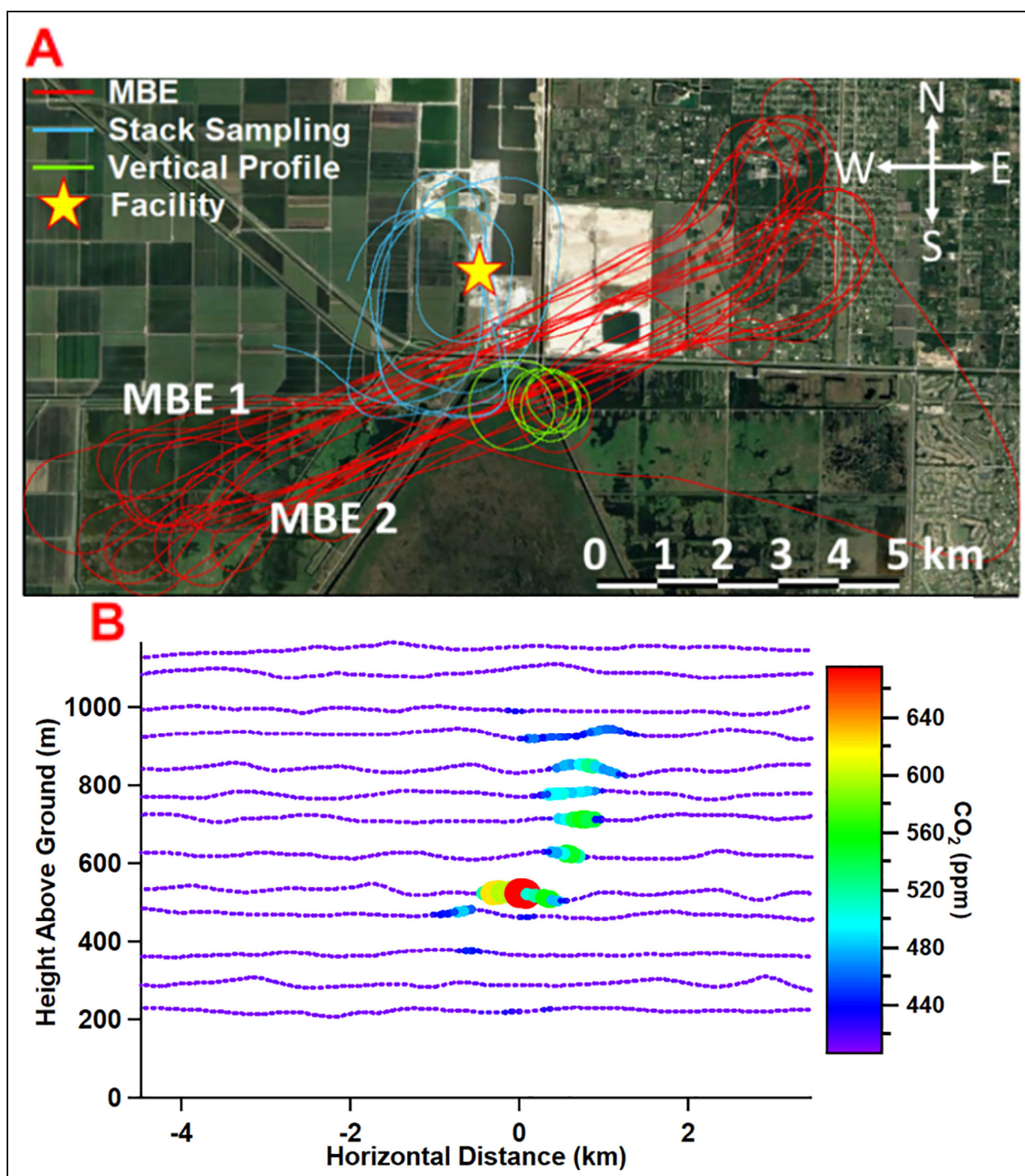


Figure 2. Example mass balance experiment (MBE). (A) Example MBE flight with 2 MBEs with key features labeled. (B) Downwind curtain of MBE 1 from A colored and size scaled by the concentration of carbon dioxide (ppm).

flights were designed to be perpendicular to the predicted wind direction, the average $\cos(\theta)$ term is 0.918 ± 0.074 , near the perfectly perpendicular value of 1. To reduce the noise in the 50-Hz meteorological variables (temperature, pressure, and wind speed) we also apply a 10-s rolling average to these data before downsampling them to 0.5 Hz to calculate fluxes. To calculate an emission rate from these pointwise fluxes we used Equation 2:

$$\text{MBE emission rate} = \int_0^{Z_i} \int_{-x}^x \text{Flux}_i \, dx \, dz. \quad (2)$$

We first interpolated the fluxes to a 100 m (horizontal) $\times$ 10 m (vertical) grid via Kriging (Chu, 2004; Cambaliza et al., 2014; Heimbürger et al., 2017). The emission rate ($\text{mol}\cdot\text{s}^{-1}$) was then calculated by integrating these pixels across the horizontal ($-x$ to x) and vertical (0 to the top of the plume/convective boundary layer [CBL], Z_i) bounds of the plume. As we were unable to measure concentrations down to the surface and often saw enhancements at the lowest flown altitude, fluxes below the lowest pass had to be approximated. For 4 MBEs we were also unable to fly above the plume due to overcast conditions forcing us to

approximate emissions above the plume up to the estimated cloud base, an estimate of the top of the CBL. This approach assumes that the flux through this downwind plane is approximately constant over the course of an experiment. As the sampling time is longer than the turbulence timescale and we are assuming a constant source, the composite plume created by interpolating fluxes across all transects is a representation of the average advected plume. Clearly, dense sampling of the plume will help reduce the potential sampling error in this approach, as is determined and discussed later in Section 3.2. All MBE emission rates, uncertainties, distances downwind, and other relevant data are provided in SI Table S3.

2.3. Uncertainties

MBE uncertainties were calculated as in Hajny et al. (2019). The uncertainties of the MBE emission rate are calculated as the uncertainties associated with the point-wise fluxes, the plume capture (i.e., extrapolation), and the Kriging interpolation combined as the square root of the sum of relative uncertainties squared, as shown in Equation 3. Only the plume capture uncertainty differs from the method described in Hajny et al. (2019):

$$\frac{\delta_{\text{emissionrate}}}{\text{emissionrate}} = \sqrt{\left(\frac{\delta_{\text{Flux}}}{\text{Flux}}\right)^2 + \left(\frac{\delta_{\text{PlumeCapture}}}{\text{PlumeCapture}}\right)^2 + \left(\frac{\delta_{\text{Krig}}}{\text{Krig}}\right)^2}. \quad (3)$$

The Kriging interpolation uncertainty is calculated as the average relative percentage difference between the results of 2 different Kriging methods (15%) for a subset of flights (Hajny et al., 2019). The plume capture uncertainty in this work is defined as the standard deviation across 3 separate extrapolation techniques relative to the average across them. Lastly, the flux uncertainty requires several steps to propagate the uncertainty in each measured quantity.

We use the average of 3 extrapolation approaches to estimate near-surface advected fluxes: Kriging with a synthetic transect (at bottom and, when needed, top), a 3-pass average, and a plume average. The synthetic transect approach involves creating a copy of the lowest flown transect at ground level before Kriging to inform the interpolation down to this altitude to avoid inappropriate extrapolation to the surface level. This can also be done at the estimated top of the boundary layer if that top altitude is not accessible by the aircraft. The 3-pass and plume average approaches take fluxes post-Kriging and define fluxes below the lowest flown transect as the average of either the lowest 3 transects or of the entire plume. The 3-pass average is relative to the synthetic transect as both approaches use the information we have available to inform concentrations at the lowest altitudes. The plume average is similar to assuming the plume is well mixed in the CBL, which we note is unlikely given our small distance from the power plants. In SI Figure S1 we show an example set of Kriged results calculated via the 3 methods.

The uncertainty associated with incomplete plume capture was calculated as the standard deviation across these 3 approaches. MBEs where the plume was completely

captured within the transects are instead reported as the emission rate without extrapolation and thus have a zero uncertainty component from incomplete plume capture. The highest flown transect, or the estimated boundary layer height if we were unable to fly above the plume, averaged 880 ± 230 m for MBEs that completely capture the plume ($N = 9$) while the lowest transect averaged 190 ± 110 m. For MBEs that did not fully capture the vertical extent of the plume ($N = 46$), the average altitudes are 990 ± 250 m and 221 ± 98 m, respectively. SI Figure S2 shows the impact that emissions below the lowest flown transect have by comparing the MBE emission rate to the same without extrapolation. Specifically, on average, the results are 28% lower than the MBE emission rates when ignoring cases with complete plume capture, implying that, on average, 28% of the flux exists between the lowest transect and the surface. However, this is not itself an uncertainty, merely the fraction of the total estimated using our approach. The uncertainty, estimated as the variability across different extrapolation techniques, is generally lower.

The flux uncertainty is the propagation of the uncertainty of each term in Equation 1, as shown in Equation 4:

$$\frac{\delta_{\text{Flux}}}{\text{Flux}} = \sqrt{\left(\frac{\delta_{U,i}}{U_{\perp,i}}\right)^2 + \left(\frac{\delta_{P,i}}{P_i}\right)^2 + \left(\frac{\delta_{T,i}}{T_i}\right)^2 + \left(\frac{\sqrt{\delta_{\text{cal}}^2 + \delta_{\text{BG}}^2}}{\text{Enhancement}}\right)^2}. \quad (4)$$

The relative uncertainty in the perpendicular component of the wind speed $\left(\frac{\delta_{U,i}}{U_{\perp,i}}\right)$ is based on the reported horizontal wind uncertainty of $0.4 \text{ m}\cdot\text{s}^{-1}$ from Garman (2009). We do not explicitly incorporate an uncertainty due to the $\cos(\theta)$ term, indicating any nonperfect orientation of the downwind transects relative to the wind direction. The relative uncertainty in pressure $\left(\frac{\delta_{P,i}}{P_i}\right)$ is calculated as the relative difference between the measured pressure and the barometric pressure calculated using the most recent surface pressure measurements from a nearby airport. This is biased high as the barometric law does not have to precisely agree with the measured pressure at altitude and some sites may not have an airport as nearby as others. The relative uncertainty in temperature $\left(\frac{\delta_{T,i}}{T_i}\right)$ is calculated as the relative difference between 2 separate measurement systems on ALAR, a thermistor bead installed in the best air turbulence probe on the nose of the aircraft and a thermocouple system installed under the best air turbulence probe. For a small number of flights this was not possible as the thermistor bead was malfunctioning, so in these cases we use the campaign average instead. Both pressure and temperature uncertainties are generally quite small ($<5\%$) relative to wind and enhancement uncertainties which will generally dominate the flux uncertainty.

The uncertainty of the background is calculated as the standard error of the linear fit used to define the background. This approach does not accurately capture the uncertainty due to the choice of background or method of defining the background, it solely describes the

agreement between the fit used and the measured background data. For reference, the average absolute difference between the average upwind concentration and edge backgrounds for the 40 MBEs with significant upwind data was only 1.33 ± 0.27 ppm. This estimate is also biased high, especially for flights with multiple MBEs, as flights typically only had a single upwind pass at 1 altitude. This difference in background definition is negligible by comparison to the enhancements that were often over 100 ppm, although there are other methods one could propose to define a background value. As each transect has a separate background, this must be done separately for each transect. To convert to a pointwise background uncertainty, we then divide these transect-scale background uncertainties by the pointwise background values. Finally to convert from a relative background uncertainty to the absolute background component of the enhancement's uncertainty we multiply by the pointwise enhancements.

The total pointwise absolute uncertainty in the enhancement is then calculated as the combination of the calibration and background uncertainties ($\sqrt{\delta_{\text{cal}}^2 + \delta_{\text{BG}}^2}$). The uncertainty in the calibration is simply the propagated uncertainty of the slope and intercept of the calibration curve. To calculate the average relative uncertainty in the enhancement for the entire flight ($\frac{\sqrt{\delta_{\text{cal}}^2 + \delta_{\text{BG}}^2}}{\text{Enhancement}}$) we combine the pointwise absolute uncertainties as the square root sum of squares for each transect separately and divide by the transect's integrated enhancement. This is converted to a flight average by squaring the relative uncertainty for each transect, averaging the relative variances together, then taking the square root of the average variance to derive a flight-averaged relative uncertainty in the enhancement, ($\frac{\sqrt{\delta_{\text{cal}}^2 + \delta_{\text{BG}}^2}}{\text{Enhancement}}$). To address the fact that transects with near zero enhancement would have a relative enhancement uncertainty that approaches infinity, we only consider transects that contribute at least 5% to the total enhancement for the MBE in this calculation.

2.4. CEMS data

The CEMS data are publicly available at an hourly resolution from the clean air markets program data (U.S. EPA, 2022), although there may be subhourly variation in these emission rates that would also affect measurements. To directly relate these data to the flights, which varied in duration and typically covered multiple hours (e.g., 11:30 AM–12:30 PM), the hourly averaged CEMS CO₂ emission rates were combined in a time-weighted average. We also subtracted the time it took air to travel from the power plant to the MBE location, or transit time, from the actual MBE flight times. The transit time was calculated as the distance from the power plant to the plume (m) divided by the average wind speed ($\text{m}\cdot\text{s}^{-1}$) during the MBE. For example, the CEMS emission rate for an MBE from 1:00 PM to 1:50 PM with a 10-min transit time would be $(10 \times [12:00 \text{ PM CEMS}] + 40 \times [1:00 \text{ PM CEMS}])/50$. We combine the

uncertainty in the concentration and flow rate estimates in quadrature to estimate CEMS emission rate uncertainties at 14% (Peischl et al., 2010). In general, the facilities studied are all large enough so that the CEMS emission rates do not vary significantly during daytime hours, as shown for 3 example flights in SI Figure S3. We studied a wide range of power plants in this work, with CEMS emissions ranging from $77 \text{ Mg}\cdot\text{h}^{-1}$ to $2,260 \text{ Mg}\cdot\text{h}^{-1}$ for the 55 MBEs. For comparison, Ahn et al. (2020) had a range of CEMS emissions from $220 \text{ Mg}\cdot\text{h}^{-1}$ to $1,030 \text{ Mg}\cdot\text{h}^{-1}$ for their 16 MBEs, and Conley et al. (2017) had a range of CEMS emissions from $99 \text{ Mg}\cdot\text{h}^{-1}$ to $1,290 \text{ Mg}\cdot\text{h}^{-1}$ for their 5 flights.

2.5. Statistical methods

Given the variety of statistical methods used in this work, this section aims to describe them all in a concise, complete manner so that the remainder of the text can focus on the results of these analyses. All uncertainties in this work are presented as ± 1 standard deviation. Throughout the work we round uncertainties to 2 significant digits and round measured values to the same number of decimal places as per the International Organization for Standardization recommendation (Joint Committee for Guides in Metrology, 2008). For example, as stated in the abstract, the sum across all CEMS emission rates is $30,656 \pm 741 \text{ Mg}\cdot\text{h}^{-1}$, which is rounded to $30,660 \pm 740 \text{ Mg}\cdot\text{h}^{-1}$. All of the methods discussed other than precision are considering each MBE as a separate independent estimate ($N = 55$). As discussed in more detail below, the precision is evaluated by treating back-to-back MBEs at a power plant as repeat experiments; thus, all repeat experiments in a given flight are used together to calculate a single data point for precision calculations ($N = 22$). Lastly, to define the terms as used in this work: The absolute error is estimated as the absolute difference between the MBE and CEMS emission rates given the CEMS is being treated as a known pseudo-controlled release. Uncertainty refers to either propagated quantitative uncertainties or the standard deviation of a measurement across a data set as appropriate. And precision refers to the repeatability of a technique, which we measure by treating back-to-back MBEs as repeat measurements.

To quantify the bias of the MBE method we used a York regression (York, 1968; Wehr and Saleska, 2017) between observed (MBE) and reported (CEMS) emission rates, the mean error (ME), and the difference between the summed CEMS and MBE emission rates across all flights. A York regression is a type 2 regression that accounts for uncertainties in both axes. The ME is calculated according to Equation 5:

$$\text{ME} = \sum (\text{MBE}_i - \text{CEMS}_i) / N, \quad (5)$$

where MBE_i is the MBE emission rate for a given experiment, CEMS_i is the corresponding CEMS emission rate, and N is the total number of MBEs.

Many flights included between 2 and 4 back-to-back MBEs at the same power plant. With verification from CEMS that the emission rate remained approximately

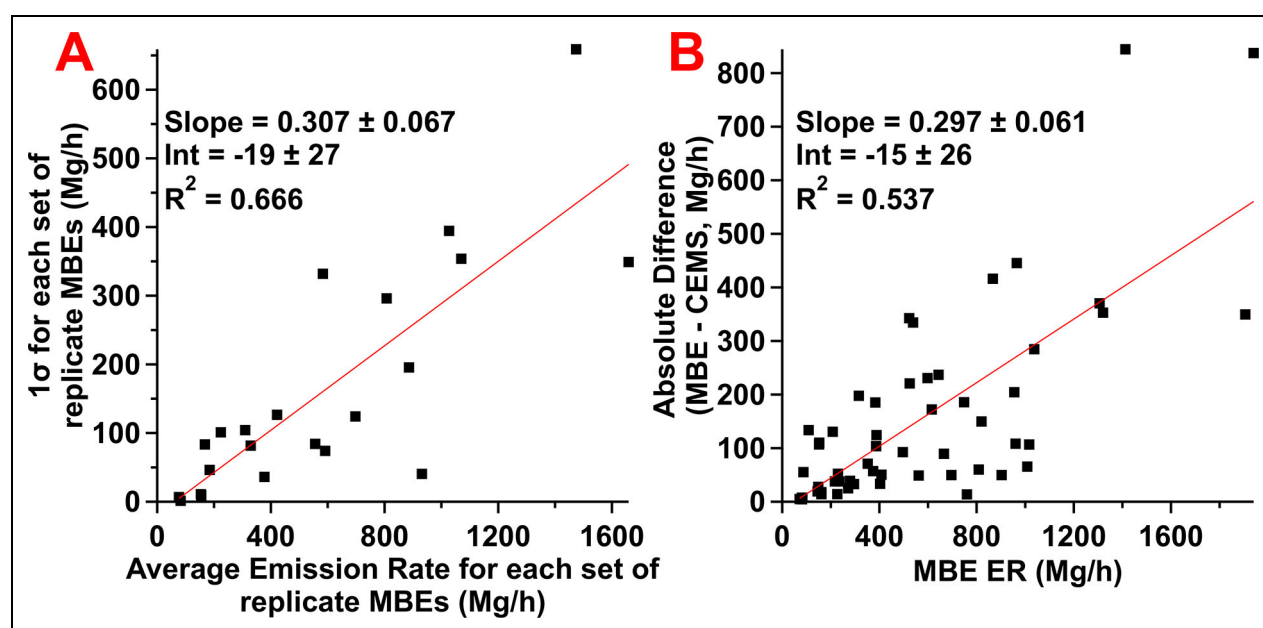


Figure 3. Precision and error regressions. (A) The absolute standard deviation across the replicate mass balance experiments (MBEs) plotted against the average emission rate across those replicate MBEs in $\text{Mg} \cdot \text{h}^{-1}$ ($N = 22$). (B) The absolute difference between MBE and CEMS emission rates plotted against the MBE emission rate, both in $\text{Mg} \cdot \text{h}^{-1}$, for all MBE estimations ($N = 55$). All uncertainties for both panels are calculated using bootstraps with 10,000 runs.

constant across these MBEs, these can be seen as repeat experiments. Although the increasing distance from the power plant across replicate MBEs could have some impact, we saw no significant linear relationship between transit time and the absolute difference between MBE and CEMS emission rates (see SI Figure S4), which suggests this changing distance should be a small effect. Transit time was calculated as the distance from the power plant to the plume (m) divided by the average wind speed ($\text{m} \cdot \text{s}^{-1}$) during the MBE. This was investigated rather than distance alone as both wind speed and distance affect plume dispersion, which would impact our S/N ratio. We calculate the precision of a single MBE as the slope of an OLS regression of the standard deviation of the emission rate across replicate MBEs against the average MBE emission rate across them, as shown in **Figure 3A**. For both panels of **Figure 3** we are calculating an OLS regression rather than a York regression as we are using uncertainty estimates as our y -axes. We have no uncertainty estimates of the y -axes and the uncertainty in the y is likely significantly larger than the x , which is treated similar to a known standard.

We quantified the degree of change in the CEMS emission rate to verify that the CEMS emission rate is approximately constant across back-to-back MBEs for the precision calculation. We did so by calculating the percentage difference in the CEMS emission rate across the time frame of the repeat MBEs. The percentage difference is calculated as the difference in the CEMS emission rate from 1 MBE to the next divided by the average, multiplied by 100 to convert to a percentage. For the 4 flights that had 3 or 4 repeat MBEs instead of 2, the percentage difference was calculated as the standard deviation of the CEMS emission rate divided by the average CEMS emission

rate across all MBEs, multiplied by 100 to convert to a percentage.

Lastly, to better understand the drivers of the absolute error, we use a multilinear regression of the absolute difference between observed (MBE) and reported (CEMS) emission rates using as covariates only quantities experimentally determined during the flights. As with the bias calculations, this absolute difference is a measure of the absolute error in the MBE emission rate given the CEMS is being treated as a known pseudo-controlled release. Every multilinear regression included an MBE emission rate term (MBE emission rate, MBE emission rate without extrapolation, MBE emission rate 1σ , or a scaled MBE term $\sqrt{\text{MBE}^3 / \text{total enhancement across all transects}}$), but otherwise every combination of the 33 covariates and their inverses was considered. Any multilinear fit with >4 parameters introduced additional collinearity (high correlations between parameters) had parameters that were not statistically significant, or did not improve the R^2 , so only 4 parameter multilinear regressions are discussed. All 33 covariates are listed in SI Table S4, while the best performing fits and their parameters are discussed in Section 3.2.

3. Results and discussion

3.1. Bias

All of our methods for assessing the bias in the MBE method suggest that there is no systematic bias in the approach. The York regression shown in **Figure 4** for the 55 MBEs has a calculated slope of 0.967 ± 0.062 with an intercept of $-12 \pm 12 \text{ Mg} \cdot \text{h}^{-1}$ with uncertainties here being calculated using a bootstrap with 10,000 runs. This slope is within 1σ of 1, suggesting that there is no statistically significant bias in the MBE results overall. We also

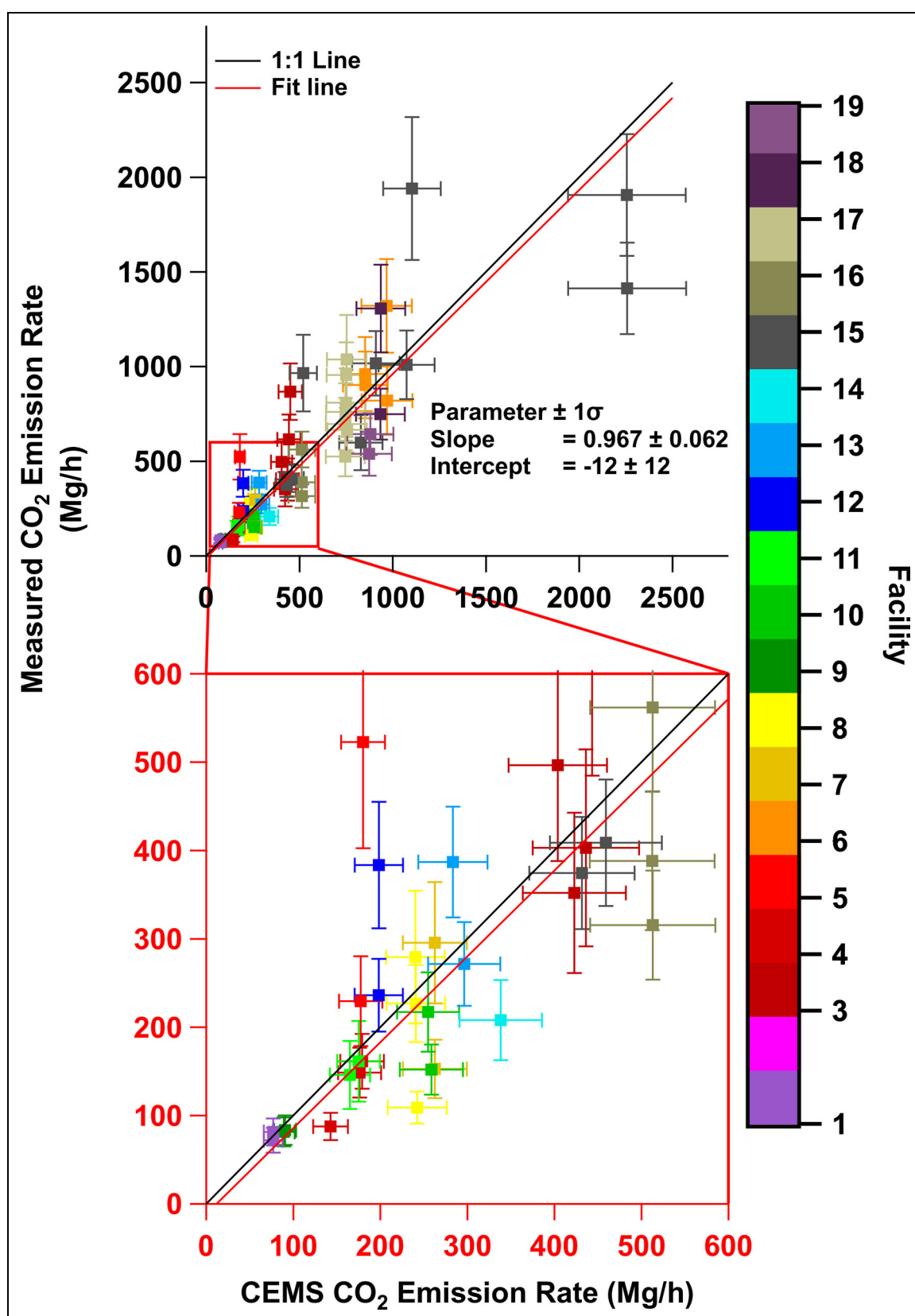


Figure 4. Emission rate comparison. Measured emission rates plotted against continuous emission monitoring systems emission rates, both in $\text{Mg}\cdot\text{h}^{-1}$, with an inset of data between 0 $\text{Mg}\cdot\text{h}^{-1}$ and 600 $\text{Mg}\cdot\text{h}^{-1}$ to more clearly see the large number of facilities in that emission rate range. All data presented are also in SI Table S3.

calculated the ME to be $10 \pm 240 \text{ Mg}\cdot\text{h}^{-1}$, meaning it is not statistically significantly different from 0. These results are consistent with previous works using smaller data sets

and somewhat different MBE methods. Ahn et al. (2020) calculated an ME of $-20 \pm 160 \text{ Mg}\cdot\text{h}^{-1}$ from 16 MBEs, and Conley et al. (2017) calculated an ME of -52 ± 75

$\text{Mg}\cdot\text{h}^{-1}$ from 5 flights. In this study, the CEMS emissions ranged from $77 \text{ Mg}\cdot\text{h}^{-1}$ to $2,260 \text{ Mg}\cdot\text{h}^{-1}$; in Ahn et al. (2020), they ranged from $220 \text{ Mg}\cdot\text{h}^{-1}$ to $1,030 \text{ Mg}\cdot\text{h}^{-1}$; and in Conley et al. (2017), they ranged from $99 \text{ Mg}\cdot\text{h}^{-1}$ to $1,290 \text{ Mg}\cdot\text{h}^{-1}$. If we limit our study to the same ranges of CEMS emissions used in these previous works, our ME is $10 \pm 190 \text{ Mg}\cdot\text{h}^{-1}$ and $30 \pm 210 \text{ Mg}\cdot\text{h}^{-1}$ for the range in Ahn et al. (2020) and Conley et al. (2017), respectively. All of these works, including ours, also include individual cases with much larger differences between MBE and CEMS emission rates. Lastly, if we sum all MBE emission rates, we calculate a total of $31,000 \pm 1,000 \text{ Mg}\cdot\text{h}^{-1}$, and if we sum all CEMS emission rates, we calculate a total of $30,660 \pm 740 \text{ Mg}\cdot\text{h}^{-1}$. As these 2 values are well within their respective uncertainties, it also supports the claim that there is no statistically significant bias in the MBE results overall. As all 3 metrics indicate the MBE emission rate is unbiased, our extrapolation method of estimating near-surface advected fluxes is likely unbiased as well. With our large data set of MBEs across a variety of meteorological conditions in this work, we can further investigate the potential causes of this variability in performance.

3.2. Explaining variability in performance

The precision across all 22 flights with repeat MBEs (a total of 49 MBEs) is calculated here using a regression of the variability across MBEs against the average MBE emission rate as shown in **Figure 3A** and discussed in more detail in Section 2.5. We find the precision for individual determinations of the emission rate to be $30.7\% \pm 6.7\%$ with the uncertainty here being calculated using a bootstrap with 10,000 runs. The intercept of this fit is $-19\% \pm 27\%$, which is not statistically significantly different from zero. We also investigated the average ratio of the variability across MBEs and the average MBE emission rate (i.e., the average ratio between y and x values in **Figure 3A**). This results in a precision estimate of $25\% \pm 16\%$. These results are not statistically significantly different, but we report here the more conservative slope-based estimate given the much narrower uncertainty range. A small number of flights showed larger percentage differences in the hourly CEMS emission rates during the MBE time frame, but excluding them had little impact on the results unless using a strict percentage difference threshold, which significantly reduces the number of usable flights (precision = $30.2\% \pm 5.0\%$ for $N = 18$ flights with $<5\%$ difference, precision = $21.6\% \pm 5.7\%$ for $N = 12$ flights with $<1\%$ difference). To recall, the uncertainty in the CEMS estimates themselves is 14% , so restricting ourselves to flights with a percentage difference in CEMS that is much less than 14% is not necessarily meaningful.

It is clear in **Figure 3B** that the absolute difference between MBE and CEMS emission rates is correlated to the MBE emission rate, but there is a large degree of spread. The OLS fit has an $R^2 = 0.537$, suggesting only approximately half of the variability in the absolute difference between MBE and CEMS emission rates can be explained by the emission rate alone. To further

understand the dominant factors affecting the absolute difference between MBE and CEMS emission rates, we calculated a multilinear regression. The 4 parameter multilinear regression that performed best (as judged by R^2) while avoiding high collinearity (defined here as >0.7 correlation between individual parameters; Farrar and Glauber, 1967) and also avoiding a statistically significant intercept is shown in Equation 6:

$$y = (0.525 \pm 0.064)x_1 - (3.9 \pm 1.7)x_2 - (0.068 \pm 0.028)x_3 + (520 \pm 200)x_4 + 20 \pm 43. \quad (6)$$

Here y is the absolute difference between MBE and CEMS emission rates, x_1 is the MBE emission rate calculated without extrapolating emissions to the surface, x_2 is the percentage of measured data that is >1 ppm above background, x_3 is the number of data points within a box encompassing the plume across the transects, and x_4 is the reciprocal of the average enhancement >1 ppm above background. This combination of terms has an R^2 of 0.594, no collinearity larger than 0.6, and all 4 terms (excluding the intercept) are statistically significant at the 99% level. This equation can be useful to estimate the uncertainty of a given MBE, as all terms can be directly calculated from the data.

Understanding the individual terms in this regression can potentially help us reduce uncertainty through better flight design. x_1 (MBE emission rate without extrapolation) is highly correlated to the MBE emission rate and performs slightly better than the MBE emission rate when combined with other terms as the extrapolation itself requires assumptions that introduce additional variability. As shown in **Figure 3B**, the emission rate explains a large fraction of the variability. x_2 (the percentage of measurements >1 ppm above background) and x_3 (the number of data points within a box encompassing the plume across the transects) provide similar insights. The more plume data sampled, both in absolute terms (x_3) and percentage terms (x_2), the smaller the absolute difference between MBE and CEMS emission rates. These point source flights generally had large amounts of background data relative to plume data, so the percentage of measurements >1 ppm above background ranges from 2.6% to 48%. Lastly, x_4 (the reciprocal of the average enhancement >1 ppm above background) tells us that the larger the average signal across the flight, the smaller the absolute difference between MBE and CEMS emission rates. This would be impacted by both larger maximum measured signals and more plume data being sampled and can be interpreted as a combination of the previous insights. The effects must be balanced as a larger emission rate results in a larger absolute (not relative) difference between MBE and CEMS emission rates, while more plume data result in a smaller absolute difference between MBE and CEMS emission rates.

However, this is only the best regression and there are several that perform similarly well that we can investigate. **Figure 5** shows the parameters present in 5 similarly well-performing regressions (including the best-performing regression). Only regressions that meet the following

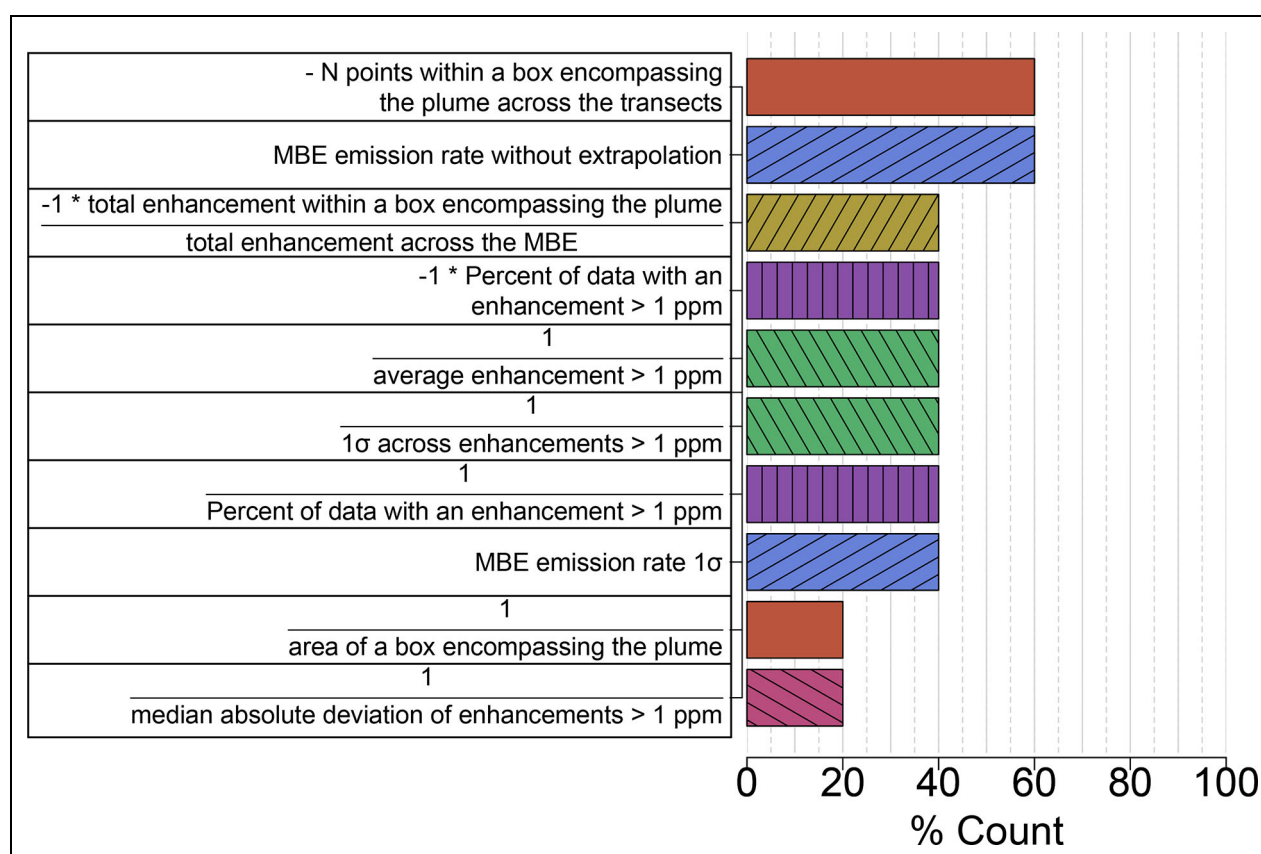


Figure 5. Multilinear regression barplot. The percentage of 4-parameter regressions that include each parameter. Since there are 4 parameters in each regression, the total across all parameters is 400%. Colors and shading lines are used to differentiate groups of parameters that are highly correlated ($P \geq 0.7$) as they cannot be used in the same regression but provide similar information to the multilinear regression. There are a total of 5 regressions used.

criteria are considered: they have an $R^2 > 0.55$, they have low collinearity (Pearson correlation < 0.7), they have an intercept that is not statistically significantly different from zero, and all 4 parameters are statistically significant at the 95% level. Additionally, 5 regressions were excluded because they included both the reciprocal of the average enhancement > 1 ppm above background and 1σ of the enhancement > 1 ppm above background in the same regression, giving relationships that did not make physical sense. Although the average and 1σ of the enhancement > 1 ppm above background and their inverses are correlated, they are not correlated above a Pearson correlation threshold of 0.7 if only 1 of the 2 variables is inverted. We also excluded 5 regressions that included conflicting relationships, each with 2 different source sensitivity terms with opposite signs. Finally, we excluded 1 regression that met these criteria yet still included a spurious relationship that did not make physical sense and had the opposite sign as other fits (the 1σ of the enhancement > 1 ppm above background). Parameters in **Figure 5** are organized into groups that represent similar concepts and are highly correlated (Pearson correlation > 0.7). Although the percentage of measurements > 1 ppm above background is not quite correlated with its own inverse at this 0.7 threshold, we still consider these terms as 1 group.

First, we would like to briefly discuss a few of the parameters tested that were not found in any of the

best-fitting regressions shown in **Figure 5**. Given that, on average, 28% of the total flux across the downwind plane exists in the near surface, below flown transects, one may expect variables related to the extrapolation to be potentially significant sources of absolute error. We tested 3 parameters focused on the impact of our extrapolation: the minimum height above ground during the lowest flown transect (m), the total extrapolated flux below the lowest flown transect after interpolation ($\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$), and the maximum enhancement in the lowest flown transect (ppm). However, none of these were present in the 5 best-performing regressions. This suggests that our method of extrapolation represents the near-surface advected fluxes well enough that other sources affect the absolute error more significantly.

As mentioned in Section 2.5, every regression was calculated with 1 MBE emission rate term (MBE emission rate, MBE emission rate without extrapolation [x_1], MBE emission rate 1σ , or a scaled MBE term $\sqrt{\text{MBE}^3/\text{total enhancement across all transects}}$) so this group is in all of the regressions. The group including the number of data points within a box encompassing the plume across the transects (x_3) and the reciprocal of the area of a box encompassing the plume is present in 80% of the regressions. The group of the reciprocal of the percentage of measurements > 1 ppm above background (x_2) and its negative inverse are also present in 80% of regressions.

The total enhancement within a box encompassing the plume relative to the total enhancement across the MBE is present in 40% of the regressions and provides another metric for the relative amount of plume data. As discussed earlier in this section, the more plume data in an MBE whether measured in absolute or relative terms, the smaller the absolute difference between MBE and CEMS emission rates. However, the area of a box encompassing the plume is not only an indirect measure of the amount of data in the plume, it is also a measure of how well mixed the plume has become as increased mixing would lead to a broader plume that's more completely mixed vertically and is thus more likely to be sampled in any flight transects. This study has many MBEs conducted close to the facility as many were part of a previous study investigating CH₄ emissions which had a much smaller S/N ratio (Hajny et al., 2019). The plume transit time from facility to transects ranges from 2 min to 39 min with a median of 7.5 min.

Lastly, the correlated terms of the reciprocal of the average enhancement >1 ppm above background (x_4) and the reciprocal of the standard deviation across all enhancements >1 ppm above background are present in 80% of the regressions. However, the 1 regression that does not have these terms instead has the reciprocal of the median absolute deviation across enhancements >1 ppm. Although calculated differently, this provides similar information to the regression as the reciprocal of the standard deviation across all enhancements >1 ppm above background. As we have both the average and the variability, these can be interpreted to be somewhat of a balance between the emission rate and the amount of plume data.

Given the relative importance of these terms, we can use them to make some suggestions about ideal flight plans that would minimize the potential impact on the absolute error. To summarize the previous paragraphs, from **Figure 5** and Equation 6, we have identified that the absolute difference between MBE and CEMS emission rates is smaller for flights with a larger fraction of data within the plume, broader well-mixed plumes, a larger emission rate, and larger enhancements in general. As such, the highest quality MBEs should be conducted under relatively turbulent mixing conditions (high insolation) that balance the increased mixing with the decreased enhancements and prioritize dense sampling of the plume.

4. Conclusions

MBEs are a valuable tool for quantifying emissions, both for research and validating efforts to reduce emissions. This work adds to the existing literature (Cambaliza et al., 2014; Conley et al., 2017; Heimbürger et al., 2017; Ahn et al., 2020) by assessing the bias and uncertainties of the aircraft MBE. There is not a statistically significant difference between the MBE and CEMS emission rates on average, but rather scatter about the 1:1 line. The precision of the method from multiple cases of replicate observations was 30.7%, which represents the precision for an individual MBE determination. Since there is no observed bias, we assume that the calculated precision is a reasonable estimate of the overall quantitative uncertainty in typical,

individual MBE determinations. As such, this work finds no bias in the aircraft MBE method and a moderate level of uncertainty similar to other top-down atmospheric science quantification methods.

For example, Yuan et al. (2015) estimated the propagated uncertainty for eddy covariance emission estimates of CH₄ at approximately 21% for 2 shale gas field flights. They also saw that their estimates were within uncertainty of MBE emission rates estimated using the same flight data, which had slightly higher uncertainties of approximately 35% (Peischl et al., 2015). Varon et al. (2018) estimated that satellite remote sensing approaches to quantify CH₄ point sources would be associated with between 15% and 65% uncertainty, depending on factors such as wind speed, and Duren et al. (2019) saw an average CH₄ point source emission rate uncertainty of 30% using an airborne remote sensing technique in California. Although these uncertainties are for a different trace gas, they are in large part caused by similar meteorological and background uncertainties and are provided here merely for comparison to our reported precision. Our estimate is also consistent with the lower limit of approximately 30% presented in Angevine et al. (2020). In their work, real flight data from 4 flights were simulated using dispersion modeling, allowing the authors to investigate some sources of uncertainty in an MBE under controlled simulated conditions. The uncertainty of longer term emission estimates could be reduced by both performing multiple MBEs and integrating with other empirical data such as site-level measurements (Heimbürger et al., 2017; Lopez-Coto et al., 2020).

We have also determined several parameters that explain up to 59% of the variability in the magnitude of the absolute error for individual flights. The fraction of data within the plume in both percentage and absolute terms, the average and variability of the plume signals, the magnitude of the emission rate, and the degree of mixing all help to explain the variability in the absolute error. This information can be used to estimate the uncertainty of MBEs since all terms derive directly from measurements, and it can also inform flight design to minimize the absolute errors. Our flights covered a range of wind speeds from approximately 3 m·s⁻¹ to 13 m·s⁻¹, with 1 flight having a much higher 21 m·s⁻¹ wind speed, and the distance from the source ranged from 0.7 km to 13.3 km. Given these ranges, the most important parameters that are actionable suggest that the highest quality MBEs should be conducted under turbulent/convective mixing conditions (balanced by avoiding loss of S/N) such that the plumes are broader and more vertically mixed and prioritize dense sampling of the plume. Of course, this mixing must also be balanced as the average and variability of measured signals are also important covariates and increasing the mixing decreases the S/N ratio. While the outcomes of this study do not directly apply to urban areas or non-buoyant plumes, this is an important step toward better understanding the performance of the MBE technique, and the conclusions may still help to design better flight approaches for such more complex emission sources.

Data accessibility statement

Aircraft measurement data used in this study are available from the Harvard Dataverse at <https://doi.org/10.7910/DVN/7EWN9E>.

Supplemental files

The supplemental files for this article can be found as follows:

Text S1. Calibration cylinder details.

Text S2. Defining parameters describing the facilities studied.

Figure S1. Example of extrapolation methods.

Figure S2. Impact of extrapolation on emission rate.

Figure S3. Example CEMS emission rates.

Figure S4. Relationship between MBE and CEMS plotted against transit time.

Table S1. Details of calibration cylinders used.

Table S2. Key parameters describing the facilities studied.

Table S3. Key parameters describing individual MBEs.

Table S4. All covariate descriptions.

Acknowledgments

The authors thank air traffic control for their support with our various flight patterns and the Jonathan Amy Facility for Chemical Instrumentation at Purdue University for continued assistance in maintaining equipment in Airborne Laboratory for Atmospheric Research.

Funding

The authors thank the Environmental Defense Fund and the National Institute of Standards and Technology (NIST; NIST Award No. 70NANB19H167) for support of this work.

Competing interests

The authors have no competing interests, as defined by *Elementa*, that might be perceived to influence the research presented in this manuscript. Paul B. Shepson is an Associate Editor at *Elementa*. He was not involved in the review process of this article.

Disclaimer

Certain commercial equipment, instruments, or materials are identified in this article 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.

Author contributions

Contributed to flight measurements and flight design: All authors.

Contributed to analysis and interpretation of data: KDH, IL-C, PBS.

Drafted and/or revised the article: KDH, IL-C, PBS.

Approved the submitted version for publication: All authors.

References

- Ahn, DY, Hansford, JR, Howe, ST, Ren, XR, Salawitch, RJ, Zeng, N, Cohen, MD, Stunder, B, Salmon, OE, Shepson, PB, Gurney, KR, Oda, T, Lopez-Coto, I, Whetstone, J, Dickerson, RR. 2020. Fluxes of atmospheric greenhouse-gases in Maryland (FLAGG-MD): Emissions of carbon dioxide in the Baltimore, MD-Washington, D.C. Area. *Journal of Geophysical Research: Atmospheres* **125**(9): e2019JD032004. DOI: <http://dx.doi.org/10.1029/2019JD032004>.
- Alfieri, S, Amato, U, Carfora, M, Esposito, M, Magliulo, V. 2010. Quantifying trace gas emissions from composite landscapes: A mass-budget approach with aircraft measurements. *Atmospheric Environment* **44**(15): 1866–1876. DOI: <http://dx.doi.org/10.1016/j.atmosenv.2010.02.026>.
- Angevine, WM, Peischl, J, Crawford, A, Loughner, CP, Pollack, IB, Thompson, CR. 2020. Errors in top-down estimates of emissions using a known source. *Atmospheric Chemistry and Physics* **20**(20): 11855–11868. DOI: <http://dx.doi.org/10.5194/acp-20-11855-2020>.
- C40 Cities. 2022. C40 Cities are making faster progress on the climate crisis than most others, but it's not enough. Available at <https://www.c40.org/news/c40-cities-are-making-faster-progress-on-the-climate-crisis-than-most-others-but-its-not-enough/>. Accessed January 16, 2023.
- Cambaliza, MOL, Shepson, PB, Bogner, J, Caulton, DR, Stirr, B, Sweeney, C, Montzka, SA, Gurney, KR, Spokas, K, Salmon, OE, Lavoie, TN, Hendricks, A, Mays, K, Turnbull, J, Miller, BR, Lauvaux, T, Davis, K, Karion, A, Moser, B, Miller, C, Obermeyer, C, Whetstone, J, Prasad, K, Miles, N, Richardson, S. 2015. Quantification and source apportionment of the methane emission flux from the city of Indianapolis. *Elementa: Science of the Anthropocene* **3**: 000037. DOI: <http://dx.doi.org/10.12952/journal.elementa.000037>.
- Cambaliza, MOL, Shepson, PB, Caulton, DR, Stirr, 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>.
- Carotenuto, F, Gualtieri, G, Miglietta, F, Riccio, A, Toscano, P, Wohlfahrt, G, Gioli, B. 2018. Industrial point source CO₂ emission strength estimation with aircraft measurements and dispersion modelling. *Environmental Monitoring and Assessment* **190**(3): 165. DOI: <http://dx.doi.org/10.1007/s10661-018-6531-8>.
- Chu, D. 2004. The GLOBEC Kriging Software Package—EasyKrig3. Available at <http://globec.who.edu/>

- software/kriging/easy_krig/easy_krig.html. Accessed August 16, 2018.
- Conley, S, Faloona, I, Mehrotra, S, Suard, M, Lenschow, DH, Sweeney, C, Herndon, S, Schwietzke, S, Pétron, G, Pifer, J, Kort, EA, Schnell, R.** 2017. Application of Gauss's theorem to quantify localized surface emissions from airborne measurements of wind and trace gases. *Atmospheric Measurement Techniques* **10**(9): 3345–3358. DOI: <http://dx.doi.org/10.5194/amt-10-3345-2017>.
- Crosson, ER.** 2008. A cavity ring-down analyzer for measuring atmospheric levels of methane, carbon dioxide, and water vapor. *Applied Physics B* **92**(3): 403–408. DOI: <http://dx.doi.org/10.1007/s00340-008-3135-y>.
- Dlugokencky, EJ, Myers, RC, Lang, PM, Masarie, KA, Croftwell, 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: Atmospheres* **110**(D18). DOI: <http://dx.doi.org/10.1029/2005JD006035>.
- Duren, RM, Thorpe, AK, Foster, KT, Rafiq, T, Hopkins, FM, Yadav, V, Bue, BD, Thompson, DR, Conley, S, Colombi, NK, Frankenberg, C, McCubbin, IB, Eastwood, ML, Falk, M, Herner, JD, Croes, BE, Green, RO, Miller, CE.** 2019. California's methane super-emitters. *Nature* **575**(7781): 180–184. DOI: <http://dx.doi.org/10.1038/s41586-019-1720-3>.
- Erland, BM, Adams, C, Darlington, A, Smith, ML, Thorpe, AK, Wentworth, GR, Conley, S, Liggio, J, Li, SM, Miller, CE, Gamon, JA.** 2022. Comparing airborne algorithms for greenhouse gas flux measurements over the Alberta oil sands. *Atmospheric Measurement Techniques* **15**(19): 5841–5859. DOI: <http://dx.doi.org/10.5194/amt-15-5841-2022>.
- Farrar, DE, Glauber, RR.** 1967. Multicollinearity in regression analysis: The problem revisited. *The Review of Economics and Statistics* **49**(1): 92–107. DOI: <http://dx.doi.org/10.2307/1937887>.
- Garman, KE.** 2009. Precision of airborne wind measurement for atmospheric flight research [PhD thesis]. West Lafayette, IN: Purdue University. Available at <https://search.proquest.com/docview/304985680/abstract/A225868A7B254C47PQ/1>. Accessed February 19, 2019.
- Garman, KE, Hill, KA, Wyss, P, Carlsen, M, Zimmerman, JR, Stirm, BH, Carney, TQ, Santini, R, Shepson, PB.** 2006. An airborne and wind tunnel evaluation of a wind turbulence measurement system for aircraft-based flux measurements. *Journal of Atmospheric and Oceanic Technology* **23**(12): 1696–1708. DOI: <http://dx.doi.org/10.1175/JTECH1940.1>.
- Garman, KE, Wyss, P, Carlsen, M, Zimmerman, JR, Stirm, BH, Carney, TQ, Santini, R, Shepson, PB.** 2008. The contribution of variability of lift-induced upwash to the uncertainty in vertical winds determined from an aircraft platform. *Boundary-Layer Meteorology* **126**(3): 461–476. DOI: <http://dx.doi.org/10.1007/s10546-007-9237-y>.
- Gioli, B, Carfora, MF, Magliulo, V, Metallo, MC, Poli, AA, Toscano, P, Miglietta, F.** 2014. Aircraft mass budgeting to measure CO₂ emissions of Rome, Italy. *Environmental Monitoring and Assessment* **186**(4): 2053–2066. DOI: <http://dx.doi.org/10.1007/s10661-013-3517-4>.
- Guha, A, Newman, S, Fairley, D, Dinh, TM, Duca, L, Conley, SC, Smith, ML, Thorpe, AK, Duren, RM, Cusworth, DH, Foster, KT, Fischer, ML, Jeong, S, Yesiller, N, Hanson, JL, Martien, PT.** 2020. Assessment of regional methane emission inventories through airborne quantification in the San Francisco Bay Area. *Environmental Science & Technology* **54**(15): 9254–9264. DOI: <http://dx.doi.org/10.1021/acs.est.0c01212>.
- 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): e2020JD032974. DOI: <http://dx.doi.org/10.1029/2020JD032974>.
- Hajny, KD, Salmon, OE, Rudek, J, Lyon, DR, Stuff, AA, Stirm, BH, Kaeser, R, Floerchinger, CR, Conley, S, Smith, ML, Shepson, PB.** 2019. Observations of methane emissions from natural gas-fired power plants. *Environmental Science & Technology* **53**(15): 8976–8984. DOI: <http://dx.doi.org/10.1021/acs.est.9b01875>.
- Hall, BD, Croftwell, AM, Kitzis, DR, Mefford, T, Miller, BR, Schibig, MF, Tans, PP.** 2021. Revision of the World Meteorological Organization Global Atmosphere Watch (WMO/GAW) CO₂ calibration scale. *Atmospheric Measurement Techniques* **14**(4): 3015–3032. DOI: <http://dx.doi.org/10.5194/amt-14-3015-2021>.
- Heimbürger, AMF, Harvey, RM, Shepson, PB, Stirm, BH, Gore, C, Turnbull, J, Cambaliza, MOL, Salmon, OE, Kerlo, AEM, Lavoie, TN, Davis, KJ, Lauvaux, T, Karion, A, Sweeney, C, Brewer, WA, Hardesty, RM, Gurney, KR.** 2017. Assessing the optimized precision of the aircraft mass balance method for measurement of urban greenhouse gas emission rates through averaging. *Elementa: Science of the Anthropocene* **5**: 26. DOI: <http://dx.doi.org/10.1525/elementa.134>.
- International Energy Agency.** 2013. *World energy outlook 2013*. Paris, France: OECD Publishing.
- Joint Committee for Guides in Metrology.** 2008. Joint Committee for Guides in Metrology 100:2008(E). Available at <https://www.iso.org/sites/JCGM/GUM/JCGM100/C045315e.html/C045315e.html?csnumber=50461>. Accessed January 17, 2023.
- Karion, A, Sweeney, C, Kort, EA, Shepson, PB, Brewer, A, Cambaliza, M, Conley, SA, Davis, K, Deng, A, Hardesty, M, Herndon, SC, Lauvaux, T, Lavoie, T, Lyon, D, Newberger, T, Pétron, G, Rella, C, Smith, M, Wolter, S, Yacovitch, TI, Tans, P.** 2015. Aircraft-based estimate of total methane emissions from the

- Barnett Shale region. *Environmental Science & Technology* **49**(13): 8124–8131. DOI: <http://dx.doi.org/10.1021/acs.est.5b00217>.
- Klausner, T, Mertens, M, Huntrieser, H, Galkowski, M, Kuhlmann, G, Baumann, R, Fiehn, A, Jöckel, P, Pühl, M, Roiger, A.** 2020. Urban greenhouse gas emissions from the Berlin area: A case study using airborne CO₂ and CH₄ in situ observations in summer 2018. *Elementa: Science of the Anthropocene* **8**: 15. DOI: <http://dx.doi.org/10.1525/elementa.411>.
- Lavoie, TN, Shepson, PB, Gore, CA, Stirm, BH, Kaeser, R, Wulle, B, Lyon, D, Rudek, J.** 2017. Assessing the methane emissions from natural gas-fired power plants and oil refineries. *Environmental Science & Technology* **51**(6): 3373–3381. DOI: <http://dx.doi.org/10.1021/acs.est.6b05531>.
- 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>.
- Mays, KL, Shepson, PB, Stirm, BH, Karion, A, Sweeney, C, Gurney, KR.** 2009. Aircraft-based measurements of the carbon footprint of Indianapolis. *Environmental Science & Technology* **43**(20): 7816–7823. DOI: <http://dx.doi.org/10.1021/es901326b>.
- Peischl, J, Ryerson, TB, Aikin, KC, de Gouw, JA, Gilman, JB, Holloway, JS, Lerner, BM, Nadkarni, R, Newman, JA, Nowak, JB, Trainer, M, Warneke, C, Parrish, DD.** 2015. Quantifying atmospheric methane emissions from the Haynesville, Fayetteville, and northeastern Marcellus shale gas production regions. *Journal of Geophysical Research: Atmospheres* **120**(5): 2119–2139. DOI: <http://dx.doi.org/10.1002/2014JD022697>.
- Peischl, J, Ryerson, TB, Holloway, JS, Parrish, DD, Trainer, M, Frost, GJ, Aikin, KC, Brown, SS, Dubé, WP, Stark, H, Fehsenfeld, FC.** 2010. A top-down analysis of emissions from selected Texas power plants during TexAQS 2000 and 2006. *Journal of Geophysical Research: Atmospheres* **115**(D16). DOI: <http://dx.doi.org/10.1029/2009JD013527>.
- Picarro, Inc.** 2021. G2301 Analyzer Datasheet | Picarro. Available at https://www.picarro.com/support/library/documents/g2301_analyzer_datasheet. Accessed January 9, 2023.
- Ren, X, Salmon, OE, Hansford, JR, Ahn, D, Hall, D, Benish, SE, Stratton, PR, He, H, Sahu, S, Grimes, C, Heimbürger, AMF, Martin, CR, Cohen, MD, Stunder, B, Salawitch, RJ, Ehrman, SH, Shepson, PB, Dickerson, RR.** 2018. Methane emissions from the Baltimore-Washington area based on airborne observations: Comparison to emissions inventories. *Journal of Geophysical Research: Atmospheres* **123**(16): 8869–8882. DOI: <http://dx.doi.org/10.1029/2018JD028851>.
- Tans, PP, Crotwell, AM, Thoning, KW.** 2017. Abundances of isotopologues and calibration of CO₂ greenhouse gas measurements. *Atmospheric Measurement Techniques* **10**(7): 2669–2685. DOI: <http://dx.doi.org/10.5194/amt-10-2669-2017>.
- United Nations Department of Economic and Social Affairs Population Division.** 2019. *World urbanization prospects: The 2018 revision*. New York, NY: United Nations.
- U.S. Census Bureau.** 2021. TIGER/line shapefiles from the geography program. Available at <https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html>. Accessed February 18, 2021.
- U.S. Environmental Protection Agency.** 2022. Clean Air Markets Program Data. Available at <https://campd.epa.gov/>. Accessed June 29, 2022.
- Varon, DJ, Jacob, DJ, McKeever, J, Jervis, D, Durak, BOA, Xia, Y, Huang, Y.** 2018. Quantifying methane point sources from fine-scale satellite observations of atmospheric methane plumes. *Atmospheric Measurement Techniques* **11**(10): 5673–5686. DOI: <http://dx.doi.org/10.5194/amt-11-5673-2018>.
- Wehr, R, Saleska, SR.** 2017. The long-solved problem of the best-fit straight line: Application to isotopic mixing lines. *Biogeosciences* **14**(1): 17–29. DOI: <http://dx.doi.org/10.5194/bg-14-17-2017>.
- Whitehouse.gov.** 2021 Apr 22. FACT SHEET: President Biden sets 2030 greenhouse gas pollution reduction target aimed at creating good-paying union jobs and securing U.S. leadership on clean energy technologies. Available at <https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/27/executive-order-on-tackling-the-climate-crisis-at-home-and-abroad/>. Accessed September 17, 2021.
- York, D.** 1968. Least squares fitting of a straight line with correlated errors. *Earth and Planetary Science Letters* **5**: 320–324. DOI: [http://dx.doi.org/10.1016/S0012-821X\(68\)80059-7](http://dx.doi.org/10.1016/S0012-821X(68)80059-7).
- Yuan, B, Kaser, L, Karl, T, Graus, M, Peischl, J, Campos, TL, Shertz, S, Apel, EC, Hornbrook, RS, Hills, A, Gilman, JB, Lerner, BM, Warneke, C, Flocke, FM, Ryerson, TB, Guenther, AB, de Gouw, JA.** 2015. Airborne flux measurements of methane and volatile organic compounds over the Haynesville and Marcellus shale gas production regions. *Journal of Geophysical Research: Atmospheres* **120**(12): 6271–6289. DOI: <http://dx.doi.org/10.1002/2015JD023242>.

How to cite this article: Hajny, KD, Lyon, DR, Armstrong, A, Floerchinger, CR, Jayarathne, T, Kaeser, R, Lavoie, T, Salmon, OE, Stirm, BH, Stuff, AA, Tomlin, JM, Wulle, B, Lopez-Coto, I, Shepson, PB. 2023. Assessing the bias and uncertainties in the aircraft mass balance technique for the determination of carbon dioxide emission rates. *Elementa: Sciences of the Anthropocene* 11(1). DOI: <https://doi.org/10.1525/elementa.2022.00135>

Domain Editor-in-Chief: Detlev Helmig, Boulder AIR LLC, Boulder, CO, USA

Guest Editor: Frank Flocke, NCAR Earth System Laboratory, Boulder, CO, USA

Knowledge Domain: Atmospheric Science

Published: June 28, 2023 **Accepted:** May 11, 2023 **Submitted:** November 2, 2022

Copyright: © 2023 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.



Elem Sci Anth is a peer-reviewed open access journal published by University of California Press.

OPEN ACCESS The Open Access icon, which is a stylized circular logo with a central figure and radiating lines.